

# **Probabilistic Forecasting for On-line Operation of Urban Drainage Systems**

Roland Löwe



Kongens Lyngby 2014  
PHD-2014-324

Technical University of Denmark  
Department of Applied Mathematics and Computer Science  
Building 303B, DK-2800 Kongens Lyngby, Denmark  
Phone +45 45253031  
[compute@compute.dtu.dk](mailto:compute@compute.dtu.dk)  
[www.compute.dtu.dk](http://www.compute.dtu.dk)  
PHD-2014-324  
ISSN: 0909-3192

# Summary (English)

---

This thesis deals with the generation of probabilistic forecasts in urban hydrology. In particular, we focus on the case of runoff forecasting for real-time control (RTC) on horizons of up to two hours.

For the generation of probabilistic on-line runoff forecasts, we apply the stochastic grey-box model approach. Building on previous work concerning the development of conceptual stochastic rainfall-runoff model structures, we

- investigate approaches for the calibration of model parameters that tune the models for multistep predictions,
- develop an approach for generating probabilistic multistep predictions of runoff volume in an on-line setting,
- develop a new approach for dynamically modelling runoff forecast uncertainty.

We investigate how rainfall inputs can be optimally combined for runoff forecasting with stochastic grey-box models and what effect different types of radar rainfall measurements and forecasts have on on-line runoff forecast quality.

Finally, we implement the stochastic grey-box model approach in a real-world real-time control (RTC) setup and study how RTC can benefit from a dynamic quantification of runoff forecast uncertainty.



# Summary (Danish)

---

Denne afhandling omhandler metoder for beregning af probabilistiske prognoser for afløbssystemer. Vi fokuserer især på flowprognoser for horisonter op til 2 timer med henblik på styring af afløbssystemer.

Vi bruger stokastiske grey-box modeller for at generere probabilistiske online flow prognoser. Tidligere forskning har beskæftiget sig med udviklingen af stokastiske, konceptuelle model strukturer for afløbssystemer. Med udgangspunkt heri fokuserer denne afhandling på at:

- udvikle nye metoder for parameter kalibrering som fokuserer på modellens evne til beregning af prognoser af flow på langtidshorisonter,
- udvikle nye metoder for at generere probabilistiske prognoser af strømingsvolumen på langtidshorisonter til online formål,
- udvikle nye metoder for en dynamisk beskrivelse af prognoseusikkerheden.

Vi undersøger hvorledes forskellige typer regn målinger bedst kan kombineres når formålet er at lave probabilistiske afløbsprognoser med grey-box modeller, samt hvilken type regn måling og -prognose er mest velegnet som grundlag for afløbsprognoser.

Endelig implementerer vi stokastiske grey-box modeller i en online styring og tester hvordan en styring af afløbssystemer kan profitere fra en dynamisk beskrivelse af prognose usikkerheden.



# Preface

---

This thesis was prepared at the department of Applied Mathematics and Computer Science (DTU Compute) at the Technical University of Denmark in partial fulfilment of the requirements for acquiring a Ph.D. degree.

The thesis deals with the development of methods for probabilistic forecasting of runoffs from urban drainage systems and their application for real-time control.

The thesis consists of a summary report and 7 papers, documenting the work carried out during the period between January 2011 and January 2014. Three of these papers are published or accepted in international peer-reviewed journals, two papers will be submitted shortly after the hand-in of this thesis and two papers are under preparation.

København, 28-February-2014

Roland Löwe



# Acknowledgements

---

This thesis was prepared during an intense period of 3 years and 2 months of Ph.D. study at the Department of Applied Mathematics and Computer Science at the Technical University of Denmark. It is the outcome of an inspiring research environment, support from my supervisors and a friendly and supportive atmosphere, both at the institute and amongst the project partners.

The thesis is a part of the Storm- and Wastewater Informatics (SWI) project and was financed by the Danish Council for Strategic Research, Programme Commission on Sustainable Energy and Environment, the Technical University of Denmark and the utility companies HOFOR (former Københavns Energi), Aarhus Vand, Lynettefællesskabet and Spildevandscenter Avedøre and involved significant contributions from Krüger AS and DHI. I further acknowledge the UrbanWaterTech and ITMAN graduate schools for the possibility of interacting with other researchers at scientific seminars.

I wish to thank my supervisors Henrik Madsen (DTU Compute) and Peter Steen Mikkelsen (DTU Environment) for their support during this period, for the always open doors, for the freedom to try new ideas and for the patience with a newcomer in this field. In retrospect, I was lucky to receive important guidance methodologically and in the choice of collaborators.

Amongst our collaborators I want to thank Søren Thorndahl and Michael Rasmussen (both Department of Civil Engineering, Aalborg University) for providing radar rainfall data and extensive support for the interpretation of results and for discussions with reviewers in paper B. I am very grateful to Luca Vezzaro (DTU Environment and Krüger AS) for his endurance, the many discussions and

feedbacks during our collaboration on papers E and F over a period of almost 2 years. Finally, I would like to thank Morten Grum (Krüger AS) for always pushing this work towards the problems relevant in practice and Rune Juhl (DTU Compute) for the discussions on parameter estimation and his improvements in the *CTSM* software.

A different but no less important aspect is the very friendly atmosphere at the former Time Series (now Dynamical Systems) group at DTU Compute, which made it very much enjoyable for me to go to work. I wish to name Anne Katrine Duun-Christensen and Julija Tastu with whom I was lucky to share the office for several years and with whom I had many discussions and talks.

Finally, I wish to very much thank my girlfriend Rasa Narkevičiūtė for her patience with the increasingly frequent long nights over the last year, for moral support and for proof-reading this thesis.

# List of Publications

---

## Papers Included in the Thesis

- A Roland Löwe, Peter Steen Mikkelsen, Michael Rasmussen, Henrik Madsen (2013). "State-space adjustment of radar rainfall and skill score evaluation of stochastic volume forecasts in urban drainage systems", *Water Science and Technology*, 68(3): 584–590, 2013.
- B Roland Löwe, Søren Thorndahl, Peter Steen Mikkelsen, Michael Rasmussen, Henrik Madsen (2014). "Probabilistic online runoff forecasting for urban catchments using inputs from rain gauges as well as statically and dynamically adjusted weather radar", accepted by *Journal of Hydrology*, 2014.
- C Roland Löwe, Peter Steen Mikkelsen, Henrik Madsen (2014). "Stochastic rainfall-runoff forecasting: parameter estimation, multi-step prediction, and evaluation of overflow risk", *Stochastic Environmental Research and Risk Assessment*, 28: 505–516, 2014.
- D Dario Del Giudice, Roland Löwe, Henrik Madsen, Peter Steen Mikkelsen, Jörg Rieckermann (2014). "Comparing two stochastic approaches to predict urban rainfall-runoff with explicit consideration of model bias", in preparation for *Water Resources Research*, 2014.
- E Roland Löwe, Luca Vezzaro, Peter Steen Mikkelsen, Morten Grum, Henrik Madsen (2014). "Investigating the use of probabilistic forecasts for RTC of urban drainage systems - A Layout for Probabilistic Online Forecasting of Sewer Flows", in preparation for *Environmental Modelling and Software*, 2014.

- F Luca Vezzaro, Roland Löwe, Henrik Madsen, Morten Grum, Peter Steen Mikkelsen (2014). "Investigating the use of probabilistic forecasts for RTC of urban drainage systems – Full scale testing in the city of Copenhagen, Denmark", in preparation for *Environmental Modelling and Software*, 2014.
- G Roland Löwe, Rune Juhl, Peter Steen Mikkelsen, Henrik Madsen (2014). "Forecasting Operational Runoff Forecast Uncertainties - State, Rainfall and Error Dependencies", in preparation for *Hydrology and Earth System Science Discussions*.

## Other Publications not Included

The following publications have also been prepared during the course of the Ph.D. study. They are omitted in this thesis because they either resulted from work not directly related to the thesis or overlap with contents in the above publications.

- H Katja Seggelke, Roland Löwe, Thomas Beeneken, Lothar Fuchs (2013). "Implementation of an integrated real-time control system of sewer system and waste water treatment plant in the city of Wilhelmshaven", *Urban Water Journal*, 10, 330–341, 2013.
- I Roland Löwe, Peter Steen Mikkelsen, Henrik Madsen (2012). "Forecast generation for real-time control of urban drainage systems using greybox modelling and radar rainfall", in *Proceedings of the 10th International Conference on Hydroinformatics*, Hamburg, Germany, 2012.
- J Roland Löwe, Peter Steen Mikkelsen, Michael Rasmussen, Henrik Madsen (2012). "State-space adjustment of radar rainfall and stochastic flow forecasting for use in real-time control of urban drainage systems", in *Proceedings of the 9th International Conference on Urban Drainage Modeling*, Belgrade, Serbia, 2012.
- K Luca Vezzaro, Roland Löwe, Henrik Madsen, Morten Grum, Peter Steen Mikkelsen (2013). "Investigating the use of stochastic forecast for RTC of urban drainage systems", in *Proceedings of NOVATECH 2013*, Lyon, France, 2013.
- L Roland Löwe, Henrik Madsen, Peter Steen Mikkelsen (2013). "Stokastiske prognoser for afløb og real tids styring", *EVA : Erfaringsudveksling i vandmiljøteknikken*, 26(2), 11–18, ISSN 1901-3663, 2013.

# Glossary

---

Abbreviation	Explanation
ACF	autocorrelation function
ARIL	average interval length
ANN	artificial neural network
CCF	cross correlation function
CRPS	continuous ranked probability score
CSO	combined sewer overflow
CTSM	Continuous Time Stochastic Modeling software
DORA	dynamic overflow risk assessment
EKF	extended Kalman filter
iid	independent and identically distributed
LAWR	local area weather radar
ML	maximum likelihood
MAP	maximum a posteriori
NSE	Nash-Sutcliffe efficiency index
ODE	ordinary differential equation
QPE	quantitative precipitation estimate
QPF	quantitative precipitation forecast
PI	Persistence Index
RTC	real-time control
SC	Interval Score
SDE	stochastic differential equation
SS	skill score (generic)
SWI	Storm and Wastewater Informatics project
WWTP	wastewater treatment plant



# Contents

---

<b>Summary (English)</b>	i
<b>Summary (Danish)</b>	iii
<b>Preface</b>	v
<b>Acknowledgements</b>	vii
<b>List of Publications</b>	ix
<b>Glossary</b>	xi
<b>I Summary Report</b>	1
<b>1 Introduction</b>	3
1.1 Models in Urban Hydrology and their Objectives . . . . .	3
1.2 Forecasting and Real-time control of Urban Drainage Systems . . . . .	5
1.3 Objective of the PhD project . . . . .	7
1.4 Thesis Outline . . . . .	8
<b>2 Verification of Probabilistic Forecasts</b>	11
2.1 Scoring Rules for Point Forecast Quality . . . . .	12
2.2 Elements of Probabilistic Forecast Quality . . . . .	13
2.2.1 Calibration . . . . .	14
2.2.2 Sharpness . . . . .	16
2.2.3 Scoring Rules for Probabilistic Forecast Quality . . . . .	17

<b>3 Grey-box Modelling of Sewer Flows</b>	<b>21</b>
3.1 Introduction to Stochastic Grey-box Models . . . . .	21
3.2 Rainfall Runoff Modeling with Stochastic Grey-box Models . . . . .	22
3.2.1 The Linear Reservoir Cascade . . . . .	22
3.2.2 The Linear Reservoir Cascade as Stochastic Grey-box Model	24
3.3 Model Structures for Uncertainty in Stochastic Grey-box models	24
3.4 From State-Dependent Diffusion to Constant Diffusion - Lamperti Transformations . . . . .	27
3.4.1 The General Lamperti Transformation . . . . .	28
3.4.2 Lamperti Transformation for the Input Dependent Uncertainty Description . . . . .	29
3.4.3 Overview of Applied Lamperti Transformations . . . . .	30
3.5 The Relation between Lamperti and Data Transformations . . . . .	32
3.6 Parameter Estimation in Stochastic Grey-box Models . . . . .	34
3.6.1 Maximum Likelihood Estimation . . . . .	34
3.6.2 Maximum A Posteriori Estimation . . . . .	36
3.6.3 Estimation Based on Multi-step Forecast Verification . . . . .	37
3.6.4 Numerical Optimisation . . . . .	39
3.6.5 Comparing Estimation Approaches for a Sample Catchment	39
3.7 Simulation of Stochastic Differential Equations . . . . .	42
3.8 Generating Multi-step Probabilistic Forecasts from Stochastic Grey-box models . . . . .	43
3.9 Relation to Other Uncertainty Techniques Applied in Hydrology	44
<b>4 Rainfall Inputs for On-line Runoff Forecasting</b>	<b>47</b>
4.1 Requirements to Rainfall Input for On-line Runoff Forecasting . . . . .	47
4.2 Raingauge vs. Radar Rainfall Input . . . . .	48
4.2.1 Advantages and Disadvantages . . . . .	48
4.2.2 Merging Different Rainfall Inputs . . . . .	52
4.3 The Effect of Rainfall Forecasts on Runoff Forecast Quality - An Example . . . . .	53
4.3.1 Problem Description . . . . .	53
4.3.2 Results and Discussion . . . . .	54
4.3.3 Conclusion . . . . .	55
<b>5 Probabilistic Runoff Forecasts in Practice</b>	<b>59</b>
5.1 Real-Time Control under Uncertainty - the DORA algorithm . . . . .	59
5.2 Current Practical Implementation for Probabilistic Runoff Forecasting . . . . .	61
5.3 Control Results Using Probabilistic Forecasts . . . . .	61
<b>6 Conclusions</b>	<b>65</b>
<b>7 Outlook</b>	<b>71</b>

---

Bibliography	75
<b>II Papers</b>	<b>85</b>
A State-space adjustment of radar rainfall and skill score evaluation of stochastic volume forecasts in urban drainage systems	87
B Probabilistic online runoff forecasting for urban catchments using inputs from rain gauges as well as statically and dynamically adjusted weather radar	101
C Stochastic rainfall-runoff forecasting: parameter estimation, multi-step prediction, and evaluation of overflow risk	155
D Comparing two stochastic approaches to predict urban rainfall-runoff with explicit consideration of model bias	185
E Investigating the use of probabilistic forecasts for RTC of urban drainage systems - A Layout for Probabilistic Online Forecasting of Sewer Flows	211
F Investigating the use of probabilistic forecasts for RTC of urban drainage systems – Full scale testing in the city of Copenhagen, Denmark	261
G Forecasting Operational Runoff Forecast Uncertainties - State, Rainfall and Error Dependencies	295



## Part I

# Summary Report



## CHAPTER 1

# Introduction

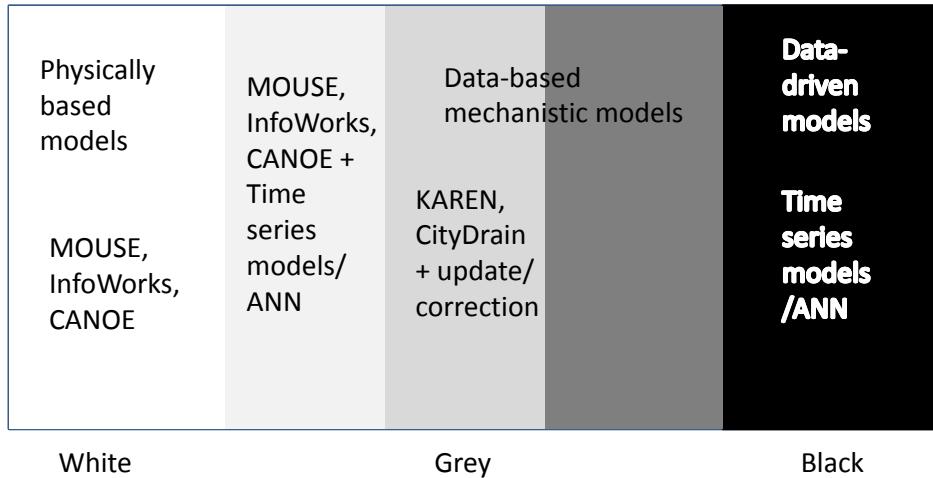
---

## 1.1 Models in Urban Hydrology and their Objectives

In general, models are created with different complexities and based on different principles. In hydrology, it is reasonable to classify models according to their complexity. For example, this approach can be found in [Ref96] for deterministic hydrologic models and, more generally, in [Lju10]. [Bre12] classifies urban hydrological models according to this approach (figure 1.1).

The exact classification varies among authors. Yet, the general line of thought is that white-box models are based on first order principles which describe the actual physical process, while black-box models are purely data-based and typically do not allow for a physical interpretation of model structure and parameters. Grey-box models cover the area between the two extremes.

Other than depicted in Figure 1.1, models applied in urban hydrology are generally not white. Even rather detailed model structures as MOUSE ([DHI03]) or WEST ([VMA<sup>+</sup>03]) build on conceptual assumptions for example in the runoff formation and transformation processes or in the description of treatment processes.



**Figure 1.1:** Classification of urban hydrological models according to their structural complexity (from [Bre12])

The task of the hydrologist is thus, to identify a model that makes optimal use of the available data and is suitable for a given purpose rather than aiming at a perfect physical process description.

Modelling tasks in urban hydrology can be separated into off-line and on-line applications. Off-line applications typically focus on design or analysis of the system and comprise decision horizons of several years. They very often include 'What if?' tasks, leading to an analysis of effects resulting from an anticipated change in the system. In this field, complex model structures have their main range of application as they permit for the use of a maximum of physical (static) information such as pipe diameters, the location of sealed area in a catchment or the dimensions of treatment facilities.

On-line applications, on the other hand, are related to supervision of the system and decision making over short horizons in the range from minutes to days. Computational speed of the models is often relevant in this area, limiting the use of complex model structures. In turn, some current measurement information about the system is typically available and should be used in the model. In this case, the use of statistical techniques such as auto-calibration or state updating is typically simplified by simple model structures. Examples of on-line modelling tasks in urban hydrology are

- forecasting runoff or treatment capacity for real-time control ([VG14])
- software sensors deriving information about the system from an indirect measurement ([DDR13])
- automated cleaning and correction of measurements ([ATCV13])

Most of these tasks profit from model structures which are simple, and thus numerically and statistically easy to handle, but have a physical interpretation. Such a structure allows for the automatic estimation of parameters from measurements as well as for algorithms to update model states and parameters. At the same time, physical information can be included using for example prior parameter estimates. We denote such models as grey-box models.

The thesis is centred around the application of a special class of grey-box models, so called stochastic grey-box models (see chapter 3), for runoff forecasting in urban hydrology. Generally, every model that is fitted to data may be considered as a stochastic model, but the different approaches distinguish themselves in whether and how they explicitly account for the time-dynamic and autocorrelated model error structures. One of the major advantages of the stochastic grey-box modelling approach is that it allows us to describe these error structures time-dynamically in a way which is suitable for on-line purposes.

## 1.2 Forecasting and Real-time control of Urban Drainage Systems

An important area of application for on-line models is real-time control of urban drainage systems. Following the definition of [SCC<sup>+</sup>04], we denote as real-time control the operation in real time of actuators in the system based on the monitoring of process variables.

The most common objective for real-time control is to use existing drainage infrastructure efficiently. This permits the operators to reduce environmental (and economic) impacts (such as combined sewer overflows, sludge escape from the wastewater treatment plant, energy consumption or use of treatment chemicals) with moderate investments into the infrastructure.

While existing implementations have focused largely on the reduction of combined sewer overflows and thus the reduction of investments into storage basins (for example [FPM11, FB05, SLBF13, SGT<sup>+</sup>13]), foreseen challenges such as

the inclusion of water systems into smart cities ([RTSS13, MB10]) or the efficient utilisation of storage space in sustainable urban drainage are likely to contribute to a development where real-time operation of drainage systems becomes a standard rather than an exception.

Implementations of real-time control systems on the global or system ([SCC<sup>+</sup>04]) level can coarsely be distinguished into strategies that were developed off-line and strategies where an optimal control decision is determined on-line ([SBB02]). Off-line strategies can be seen as the traditional approach to real-time control and have, as a result of their robustness and usually easy interpretation, been implemented in a multitude of urban water systems. Typically such strategies are implemented in the form of 'IF-THEN' rules, possibly supplemented by fuzzy systems. Examples of such implementations can be found in [FB05] and [SLBF13]. With suitable decision rules, such systems can also make use of for example precipitation forecasts ([MMI05]).

An on-line optimisation strategy dynamically determines the actuator settings on-line based on (model) forecasts of future system states (such as basin fillings) and a model-based evaluation of the effects of different actuator set points. Forecasts are required in such a system to gain information about parameters such as expected future runoff, treatment capacity or energy consumption.

This type of optimisation approach is often considered problematic because the underlying forecast models need to be simple enough to run in an optimization routine but at the same time provide realistic forecasts of the system states ([BS05]). Yet, these systems are appealing due to the objectivity of the derived control decisions, the use of forecast information and thus a potentially higher efficiency as well as the rather simple adaptability of the system to new requirements. Recently developed schemes have successfully combined on-line optimization and forecast models for the operation of drainage systems ([PCL<sup>+</sup>05, PCR<sup>+</sup>09]) and partly combined these schemes with decision rules to determine boundary conditions ([GTC<sup>+</sup>11]).

New developments also account for the uncertainty of forecasts during decision making in the control setup ([VG14]). This is an important step due to high uncertainty related to rainfall measurements and forecasts, and the uncertainty inherent in the model structures. However, a missing link is the development of models that can provide probabilistic forecasts (which quantify forecast uncertainty) in an on-line setting over a multitude of horizons from 10 minutes to 12 hours. Stochastic grey-box models may provide such a means as suggested by [BMPN00, BTM<sup>+</sup>11, CNH96, TBM<sup>+</sup>12].

## 1.3 Objective of the PhD project

The aim of this PhD project is to develop probabilistic forecast models that can be applied on-line in urban hydrology. Requirements for such models are:

- Forecast models need to be fast in order to be applicable in on-line routines.
- Forecast models need to provide accurate forecasts that are reliable and sharp (see Section 2.2)
- Forecast models need to provide probabilistic forecasts over a multitude of forecast horizons and account for correlation between the different horizons.

This thesis focuses on the generation of probabilistic runoff forecasts for horizons up to 120 minutes using stochastic grey-box models. However, the developed methods can and will also be applied to other problems such as forecasting pollutant loads and the capacity of the wastewater treatment plant. The thesis builds on and extends previous work in particular by [BTM<sup>+</sup>11] and [TBM<sup>+</sup>12]. It particularly focuses on practical applicability and aims to answer the following questions:

1. What type of rainfall inputs should be used for short-term runoff forecasting and, in particular, do we benefit from using quantitative precipitation estimates (QPE) from weather radar?
2. Do quantitative precipitation forecasts (QPF) provide benefits for short term runoff forecasts?
3. Do short-term runoff forecasts benefit from a combined rainfall input, making use of both rain gauge and radar rainfall measurements?
4. How can forecast models and parameters be identified in the context of noisy data and providing forecasts over a multitude of horizons?
5. How can dynamically changing forecast uncertainties be correctly captured in a probabilistic model structure?
6. How can probabilistic forecasts be generated for decision making in real-time control?
7. What effect does the consideration of forecast uncertainty have on the efficiency of real-time control schemes?

## 1.4 Thesis Outline

This thesis is structured as follows. Part I is a report which provides the background and introduces and summarizes the papers. Within this part, Chapter 2 gives an overview of the general requirements for probabilistic forecast and introduces ways to evaluate probabilistic forecast quality. The methods introduced in Chapter 2 are used in the later chapters and the articles.

Chapter 3 introduces the grey-box modelling approach. This approach is the basis for the generation of probabilistic runoff forecasts throughout the thesis. The chapter explains how to apply grey-box models for rainfall runoff modelling, how to estimate model parameters, how to generate probabilistic forecasts etc. and discuss how the approach relates to other uncertainty techniques applied in hydrology.

Chapter 4 discusses how and what rainfall data should be used for on-line runoff forecasting over short horizons. We focus in particular on the questions whether rain gauge or radar data should be applied and how the two data sources can be combined.

In Chapter 5 we discuss the practical implementation of probabilistic forecasts from stochastic grey-box models in real-time control and what implications forecast uncertainty has on decision making.

Finally, in Chapters 6 and 7 we conclude the thesis with reference to the objectives defined in section 1.3 and provide an outlook for future work.

Part II is a collection of publications and it contains the following papers.

**Paper A** is a journal article published in *Water Science and Technology*. It deals with the combination of radar and rain gauge measurements using a state-space modelling approach and the application of different rainfall data for probabilistic runoff forecasting.

**Paper B** is a journal article accepted by *Journal of Hydrology*. It assesses the impact on on-line runoff forecasting skill from time-constant and time-varying radar adjustment and investigates whether an improved spatial model resolution is desirable for on-line runoff forecasting.

**Paper C** is a journal article published in *Stochastic Environmental Research and Risk Assessment*. It develops a method for estimation of probabilistic on-line runoff forecasting models based on the skill of multi-step-ahead

forecasts and assesses the effect of forecast uncertainty on the expected risk of combined sewer overflow (CSO).

**Paper D** is a journal article in preparation for *Water Resources Research*. It compares the stochastic grey-box modeling approach with a Bayesian bias approach for uncertainty modelling in a common urban case study.

**Paper E** is a journal article in preparation for *Environmental Modelling and Software*. It describes the implementation of stochastic grey-box models for probabilistic runoff forecasting into an existing real-time control scheme and evaluates the forecast quality that can be obtained in six different subcatchments.

**Paper F** is a journal article in preparation for *Environmental Modelling and Software*. It evaluates the effect of runoff forecast uncertainty on decision making in real-time control.

**Paper G** is a manuscript in preparation for *Hydrology and Earth System Science Discussions*. It describes different model structures for describing forecast uncertainty in stochastic grey-box models and their effect on the calibration of multi-step probabilistic forecasts.



## CHAPTER 2

# Verification of Probabilistic Forecasts

---

As explained in Section 1.3 the thesis focuses on the generation of probabilistic forecasts for short horizons (<2 hours) in urban hydrology. The most common approach to forecasting in hydrology is still the generation of deterministic or point forecasts. In this case, the forecast quality can be measured in terms of the error of the forecasts as compared to observations. Different measures such as the root mean square error, the delay to peak or the volume error can be applied here.

For probabilistic forecasts we obtain not a single forecast value but probabilities that the forecasted quantity takes a given value. This can for example be expressed parametrically by generating a probability distribution of the forecasted quantity or non-parametrically by generating various realisations or ensembles of the forecasted quantity. In either case, we evaluate probabilistic forecast quality in terms of

- the location of the forecasted distribution with respect to the observations, quantified for example by the mean absolute error of the forecast median,
- the spread of the forecasted distribution, quantified for example by the width of prediction intervals,

- and, in the case of multivariate forecasts over multiple forecast horizons, the correlation between different horizons.

The purpose of this section is to introduce the measures applied for forecast verification in this work. For a good overview on measures that can be used for evaluating point forecast quality, we refer the reader to [BCG<sup>+</sup>13]. Scoring rules for probabilistic forecast evaluation are summarized by [GR07].

In this work we largely omit the verification of the correlation structure of forecasts generated for different forecast horizons as the forecasts are always generated as multi-step forecasts from a single model. A correlation between different horizons is thus implicit in the forecasting method.

## 2.1 Scoring Rules for Point Forecast Quality

Following a definition by [Wil11], a generic forecast skill  $SS$  can be defined as

$$SS = \frac{A - A_{ref}}{A_{perf} - A_{ref}}. \quad (2.1)$$

Here  $A$  is a measure of forecast accuracy,  $A_{ref}$  the accuracy obtained for a reference (or benchmark) forecast and  $A_{perf}$  the accuracy obtained for a perfect forecast. Positive values of  $SS$  indicate that the evaluated forecast performs better than the reference. The score values can range between negative infinity (if our forecast performs much worse than the reference) and 1 (if our forecast performs equivalently to the perfect forecast).

In hydrology, the most commonly applied forecast skill is the **Nash-Sutcliffe efficiency  $NSE$**  ([NS70]). For a point forecast  $\hat{y}_i$  which is compared to an observation  $y_i$  we obtain

$$NSE = \frac{\frac{1}{n} \sum (y_i - \hat{y}_i)^2 - \frac{1}{n} \sum (y_i - \bar{y})^2}{-\frac{1}{n} \sum (y_i - \bar{y})^2}. \quad (2.2)$$

The measure of forecast accuracy  $A$  in this case is the mean squared error. The mean squared error obtained for the perfect forecast is  $A_{perf} = 0$ . As reference forecast, the average  $\bar{y}$  over all observations  $y_i$  is applied. This, however, is a very weak reference which will lead to  $NSE$  values indicating positive forecast skill even for very badly performing models as is also argued by [SKZ<sup>+</sup>12].

We can define a more critical forecast skill by replacing the mean of the observations in equation 2.2 by the last observed value. This leads to the **persistence**

**index PI** ([BCG<sup>+</sup>13])

$$PI = \frac{\frac{1}{n} \sum (y_i - \hat{y}_i)^2 - \frac{1}{n} \sum (y_i - y_{i-1})^2}{-\frac{1}{n} \sum (y_i - y_{i-1})^2}. \quad (2.3)$$

The application of the persistence index is difficult in situations where the observations are very noisy. This problem can be solved by replacing the last known observation  $y_{i-1}$  in equation 2.3 by the result of an exponential smoothing  $y_{i-1}^S$  of the observations up to the last known observation:

$$y_{i-1}^S = \lambda \cdot y_{i-2}^S + (1 - \lambda) \cdot y_{i-1} \quad (2.4)$$

The parameter  $\lambda$  must be tuned in a calibration period. In this tuning, the squared forecast error of the reference forecast from exponential smoothing is minimized. We follow this approach in paper E and denote it as **smoothed persistence index**.

When dealing with probabilistic forecasts, we often evaluate a point forecast derived from the median of the probabilistic forecasts. We choose the median to reduce the influence from long tails of the distribution on the forecast error.

## 2.2 Elements of Probabilistic Forecast Quality

Generally, the aim of probabilistic forecasting is to maximize the sharpness of predictive distributions subject to calibration ([GR07, Pin07]). "Calibration refers to the statistical consistency between the distributional forecasts and the observations, and is a joint property of the forecasts and the events or values that materialize. Sharpness refers to the concentration of the predictive distributions and is a property of the forecasts only" ([GR07]).

The above definition implies that probabilistic forecasts should be evaluated by checking if they are probabilistically calibrated (see for example [Pin13]). This requires assessing over the whole distribution, whether the observed probabilities match the predicted (or nominal) ones. Subsequently the sharpness of the distributions should be assessed.

In the following section, we introduce the measures we have applied for the evaluation of calibration and sharpness of probabilistic forecasts in this thesis. In particular, the evaluation of probabilistic calibration over a whole distribution can be hard to communicate to practitioners. We have therefore in many cases used a simplification. We extracted the median from the probabilistic forecast

and evaluated the corresponding point forecast error. This will give an indication on how good the forecast model captures the physical behaviour of the considered system. To evaluate the calibration and sharpness of the distribution, we then focus on the reliability and width of for example 90% prediction intervals. In addition, a probabilistic score can be applied to evaluate the overall fit of the predicted distribution.

### 2.2.1 Calibration

A probabilistic forecast can be considered calibrated or reliable in a probabilistic sense, if the forecasted probability distribution matches the distribution of the observations of the considered variable. A simple way to evaluate reliability of a forecast is to analyse if the predicted (nominal) probabilities match the observed ones, i.e. whether, for example, for a given confidence level  $\alpha$  between 0 and 1, a prediction interval with coverage  $(1 - \alpha) \cdot 100\%$  indeed includes  $(1 - \alpha) \cdot 100\%$  of the observations.

We express **reliability**  $Rel$  for a confidence level  $\alpha$  as the portion of observations included in a  $(1 - \alpha) \cdot 100\%$  prediction interval. Initially, in paper C the inverse definition, describing reliability as the portion of observations not included in a  $(1 - \alpha) \cdot 100\%$  prediction interval was used. We switched to the former approach later as it is more intuitive.

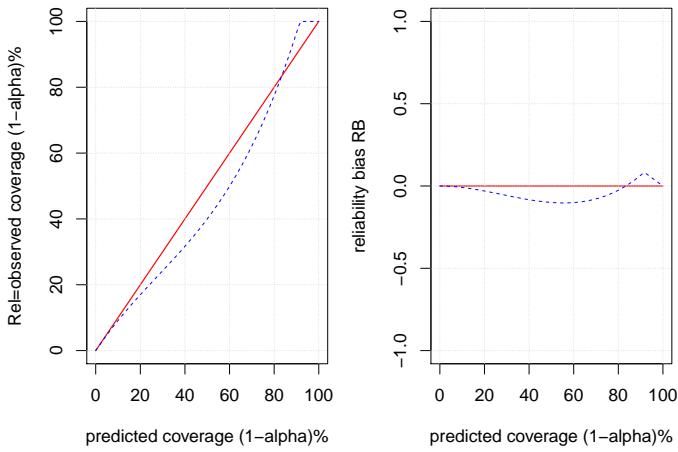
Following [TBM<sup>+</sup>12], a **reliability bias** for confidence level  $\alpha$  can be defined as

$$RB = Rel - (1 - \alpha) \quad (2.5)$$

and becomes negative if the forecasted distribution is unreliable (a  $(1 - \alpha) \cdot 100\%$  prediction interval includes less than  $(1 - \alpha) \cdot 100\%$  of the observations) and positive if it is overreliable (a  $(1 - \alpha) \cdot 100\%$  prediction interval includes more than  $(1 - \alpha) \cdot 100\%$  of the observations).

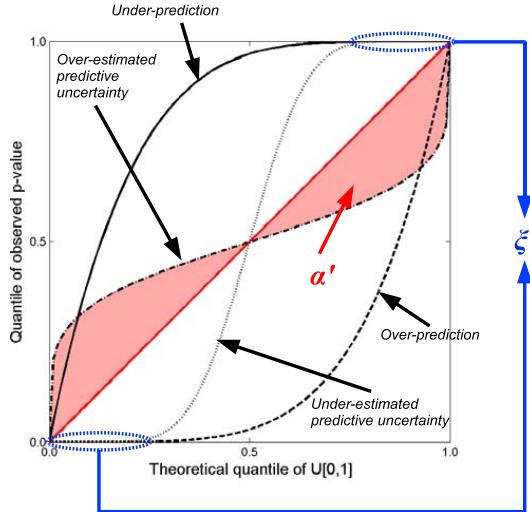
Reliability  $Rel$  can be analysed for different confidence levels and plotted in a reliability diagram ([MW77]). A perfectly reliable forecast will result in a straight line in this diagram. This is exemplified in Figure 2.1. If the reliability bias  $RB$  is plotted instead, a perfectly reliable forecast will yield a constant line at 0. The forecast evaluated in Figure 2.1 is unreliable for predicted coverages between 10 and 80% and overreliable for predicted coverages above 80%.

The reliability bias has an upper and a lower bound depending on the considered level  $(1 - \alpha)$ . A more general approach for the evaluation of predictive distributions is the **predictive QQ-plot** which is well described in [RKK<sup>+</sup>10]. From



**Figure 2.1:** Observed reliability Rel (left) and corresponding reliability bias (RB) for an example forecast (blue) and a perfect forecast (red)

the probabilistic predictions we can for each observation  $y_t$  derive the value of the cumulative predicted distribution  $F(y_t) = p(Y_t \leq y_t)$ . If the observations  $y_t$  are consistent with the predicted distributions, these p values follow a standard uniform distribution, in the interval  $[0, 1]$ . This can be illustrated graphically in a quantile-quantile plot (see Figure 2.2). A calibrated probabilistic forecast will lead to a straight line in this plot.



**Figure 2.2:** Predictive QQ-Plot comparing quantiles from 'p values' of observations and a standard uniform distribution  $U[0,1]$  (from [RKK<sup>+</sup>10])

### 2.2.2 Sharpness

Assuming that a probabilistic forecast is calibrated, a good forecast will require only a very narrow spread of the predictive distribution to capture the observations, while a bad forecast will yield a wide spread of the predictive distribution and thus only little information content to the decision maker.

In this work we assess the **sharpness** ( $Sh$ ) of probabilistic forecasts by measuring the average width of a  $(1 - \alpha) \cdot 100\%$  prediction interval. Defining  $\hat{u}_t$ ,  $\hat{l}_t$  as the upper and lower prediction bounds generated for such an interval at time step  $t$  and considering observations for  $N$  time steps, sharpness can be expressed as:

$$Sh = \frac{1}{N} \sum_{t=1}^N (\hat{u}_t - \hat{l}_t). \quad (2.6)$$

Lower sharpness values indicate a higher information content in the probabilistic forecast. The sharpness measure  $Sh$ , however, depends on the absolute value of the forecasted quantity. [JXZS10] therefore introduced the **average interval length ARIL** as a normalized measure by dividing the sharpness value of the

forecast generated at time step  $t$  by the corresponding observation  $y_t$ :

$$ARIL = \frac{1}{N} \sum_{t=1}^N \frac{\hat{u}_t - \hat{l}_t}{y_t}. \quad (2.7)$$

### 2.2.3 Scoring Rules for Probabilistic Forecast Quality

In practice, probabilistic forecasts will rarely be perfectly calibrated. A comparison of different forecasts must thus take both, the calibration and the sharpness of the forecasts into account.

There are different approaches to combine these two qualities into a single value. We consider two of these approaches in this work: the interval score  $SC$  focuses on the evaluation of prediction intervals and the continuous ranked probability score  $CRPS$  can be considered as a mean squared error measure of the whole predictive distribution.

For a single time step, the **interval score** of a  $(1-\alpha) \cdot 100\%$  prediction interval with upper bound  $u$  and lower bound  $l$  and the corresponding observation  $y$  is found as

$$SC^\alpha = u - l + \frac{2}{\alpha}(l - y) \cdot H(l - y) + \frac{2}{\alpha}(y - u) \cdot H(y - u) \quad (2.8)$$

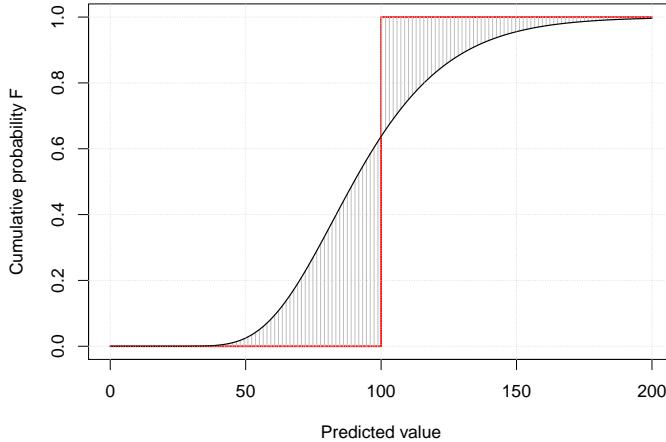
where  $H$  denotes the Heaviside function and takes the value of one if its argument is greater than zero and zero otherwise. The score for a whole time series is the average of the score values derived for the single time steps. This score evaluates the width of the prediction interval and adds a penalty for observations not included in the prediction interval. Smaller score values correspond to better forecasts.

For a single time step the **continuous ranked probability score** for a probabilistic forecast  $s$  with cumulative distribution  $F(s)$  and observation  $y$  is defined as

$$CRPS = - \int_{-\infty}^{\infty} (F(s) - H(s \geq y))^2 ds. \quad (2.9)$$

Again,  $H$  denotes the Heaviside function and the score for a whole time series is the average over the different time steps. This score can be depicted graphically as shown in Figure 2.3. The observation is represented as a stepwise cumulative distribution function (red). The  $CRPS$  measures the area between the observation and the predicted cumulative distribution function (black). Better

forecasts yield lower differences between the two distributions and thus lower score values.



**Figure 2.3:** Graphical interpretation of the CRPS - the score corresponds to the marked area between the observation (red) and the predictive distribution (black), both expressed as cumulative probability functions F

Analytical expressions for the *CRPS* are available if the distribution of the probabilistic forecast is known. However, in most cases we have used a scenario (or ensemble) approach for the generation of probabilistic forecasts. [Brö12] proposed an approach for the evaluation of the CRPS using ensembles which is based on approximating the cumulative distribution function of the forecast with a piecewise constant function.

Considering  $k = 1, \dots, K$  ensemble members with values  $e_k$ , we obtain

$$F(s) = \sum_{k=1}^K \omega_k H(s - e_k). \quad (2.10)$$

$H$  denotes the Heaviside function. The weights  $\omega_k$  are assumed  $> 0$  for all  $k$  and  $\sum_k \omega_k = 1$  and correspond to the probability that the forecasted value equals any ensemble member  $e_k$ . Thus  $\omega_k = \frac{1}{K}$ . This expression of  $F(s)$  can be inserted in equation 2.9 to evaluate the *CRPS* for an observation  $y$ .

Probabilistic scores can be normalized according to equation 2.1. The measure of accuracy  $A$  is then the CRPS, for example. In this case the score value  $A_{perf}$  for the perfect forecast is 0. A reference can be derived from the so-called "climatological" forecast. We use all the observations in the considered

dataset to derive an empirical distribution. This same "climatological" forecast distribution is then applied at every time step and the *CRPS* is derived for each observation. Averaging over all time steps, we obtain the reference score  $A_{ref}$ .



## CHAPTER 3

# Grey-box Modelling of Sewer Flows

---

### 3.1 Introduction to Stochastic Grey-box Models

The thesis focuses on the generation of probabilistic runoff forecasts using stochastic grey-box models (c.f. Section 1.3). As described in Section 1.1 we consider a grey-box model to be a (strongly) simplified representation of reality which can, nevertheless, be interpreted physically. In this work a stochastic grey-box model is defined as the implementation of such models in a state-space framework using stochastic differential equations (SDE). This terminology was largely developed at the Department for Applied Mathematics and Computer Science at the Technical University of Denmark and can be found in numerous previous works (for example [BNMP99, BTM<sup>+</sup>11, KMJ04, Møl10, NM06]).

Generally, such a model structure consists of a system of time-continuous state (equation 3.1) and time-discrete observation equations (equation 3.2). The former describe the evolution of the considered system and often may be considered the "actual model structure". The latter relate the system states (for example filling of virtual basins in a reservoir cascade) to actual observations (for

example flow measurements).

$$dX_t = \underbrace{f(X_t, u_t, t, \theta) dt}_{\text{Drift term}} + \underbrace{\sigma(X_t, u_t, t, \theta) d\omega_t}_{\text{Diffusion term}} \quad (3.1)$$

$$Y_k = h(X_k, u_k, t_k, \theta) + e_k \quad (3.2)$$

In equation 3.1  $X$  corresponds to a vector of (unobserved) system states,  $u$  to a vector of external forcings (model inputs),  $t$  to the considered time point and  $\theta$  to a vector of model parameters. Other than an ordinary differential equation (ODE), the system equations are divided into a (physical) drift term (expressed as function  $f$ ) and a (stochastic) diffusion term (expressed as function  $\sigma$ ). The latter can be used to model uncertainties resulting from insufficient model structures and uncertain inputs. The diffusion term is driven by a Wiener process with increments  $d\omega_t$ . These are normally distributed with mean 0 and variance  $dt$  (c.f. [Mad08] and paper D). It is important to note that drift and diffusion term represent a coupled system and that generally the expected value for the states  $X_t$  is not equal to the solution of the ODE represented by the drift term.

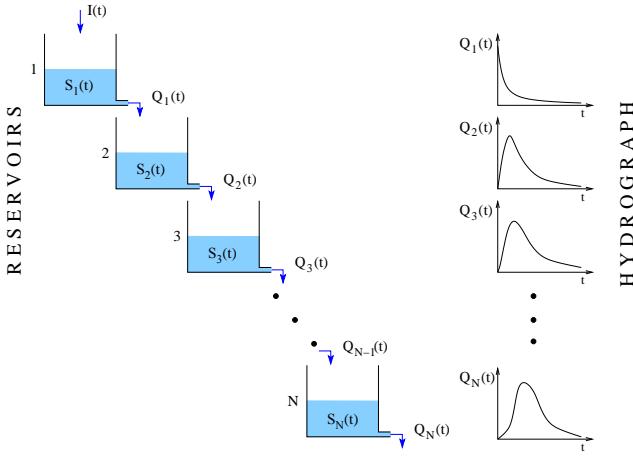
In equation 3.2 the system states for discrete time steps  $k$  are related to a vector of observations  $Y_k$  by the function  $h$ . The observations are assumed to be subject to a normally distributed error  $e_k$  with mean 0.

An extended Kalman filter (EKF) is applied to adjust the system states based on current observations  $Y_k$  of the system ([KMJ04]). The models applied in this thesis are implemented in the open source software *CTSM* ([JKB<sup>+</sup>13]) and the routines implemented in this software are applied mainly for filtering and (with exceptions) for parameter estimation.

## 3.2 Rainfall Runoff Modeling with Stochastic Grey-box Models

### 3.2.1 The Linear Reservoir Cascade

In this work, we exclusively apply a linear reservoir cascade considering either two or three reservoirs for runoff forecasting. This approach is conceptually simple, in fact, in most cases structural model deficiencies can be observed. However, it also allows for the identification of problems arising in parameter estimation and for modeling forecast uncertainties in the non-ideal case where



**Figure 3.1:** A system of  $N$  linear reservoirs. Corresponding hydrographs for the output  $Q_n(t)$  are shown to the right (from [Tho11]).

the model does not perfectly describe the data. We consider this case relevant for practical applications.

A sketch of such a model layout is shown in Figure 3.1. We refer to [CMM88] for a detailed explanation of the reservoir cascade. The principal assumption is that runoff hydrographs can be described by routing the rainfall input through a series of virtual reservoirs where the outflow  $Q_{i,t}$  at time  $t$  from a reservoir depends on the storage level  $S_{i,t}$  in the reservoir and on a time constant  $K$ .

$$Q_{i,t} = \frac{1}{K} \cdot S_{i,t} \quad (3.3)$$

Rainfall input is considered only to the first reservoir and, considering an effective catchment area  $A$  and rainfall input  $P$ , we obtain a system of differential equations describing the runoff routing through the reservoir cascade:

$$d \begin{bmatrix} S_{1,t} \\ S_{2,t} \\ \vdots \\ S_{N,t} \end{bmatrix} = \begin{bmatrix} A \cdot P_t - \frac{1}{K} S_{1,t} \\ \frac{1}{K} S_{1,t} - \frac{1}{K} S_{2,t} \\ \vdots \\ \frac{1}{K} S_{N-1,t} - \frac{1}{K} S_{N,t} \end{bmatrix} dt \quad (3.4)$$

where the outflow hydrograph is found as

$$Q_t = \frac{1}{K} \cdot S_{N,t} \quad (3.5)$$

This model structure can be considered a grey-box model as it does not rely on first order principles but the model parameters effective catchment area  $A$  and time constant  $K$  can be interpreted physically.

### 3.2.2 The Linear Reservoir Cascade as Stochastic Grey-box Model

The model structure described above can easily be cast into a state-space layout and thus be applied as a stochastic grey-box model. This approach was adapted to modeling urban drainage systems by [BTM<sup>+</sup>11]. Considering  $N = 2$  virtual reservoirs, we obtain the state equations

$$d \begin{bmatrix} S_{1,t} \\ S_{2,t} \end{bmatrix} = \begin{bmatrix} A \cdot P_t + a_0 - \frac{1}{K} S_{1,t} \\ \frac{1}{K} S_{1,t} - \frac{1}{K} S_{2,t} \end{bmatrix} dt + \begin{bmatrix} g_1(S, u, t, \theta) & 0 \\ 0 & g_2(S, u, t, \theta) \end{bmatrix} d\omega_t. \quad (3.6)$$

Here, the dry weather flow  $a_0$  is considered as input to the first model states to avoid state values approaching 0 which can be problematic in the case of state dependent noise descriptions.  $g_1$  and  $g_2$  symbolize generic functions for scaling the variance of the diffusion.

The observation equation for time step  $k$  in hours  $h$  is found as

$$Y_k = Q_k = \frac{1}{K} S_{2,k} + D_k + e_k \quad (3.7)$$

where the dry weather variation  $D_k$  is described by a set of trigonometric functions

$$D_k = \sum_i^2 (s_i \sin \frac{i 2\pi k}{24h} + c_i \cos \frac{i 2\pi k}{24h}) \quad (3.8)$$

A transformation may be applied to the observations which will be discussed in Section 3.5. Depending on the catchment characteristics, more detailed model structures may be applied as discussed by [Bre12].

## 3.3 Model Structures for Uncertainty in Stochastic Grey-box models

### *Constant Diffusion*

The diffusion term in equation 3.6 can take different structures. In the simplest case, a constant diffusion term is applied:

$$g_i = \sigma_i \quad (3.9)$$

where the parameter  $\sigma_i$  for the i-th state is determined as part of the parameter estimation routine.

#### *State Dependent Diffusion*

[Bre12] and [TBM<sup>+</sup>12] have focused on diffusion structures that depend on the current state of the model. Suggested candidates were

1. A direct linear dependency on the state

$$g_i = \sigma_i \cdot S_i \quad (3.10)$$

2. An exponentially scaled dependency on the state with scaling parameter  $\gamma$

$$g_i = \sigma_i \cdot S_i^\gamma \quad (3.11)$$

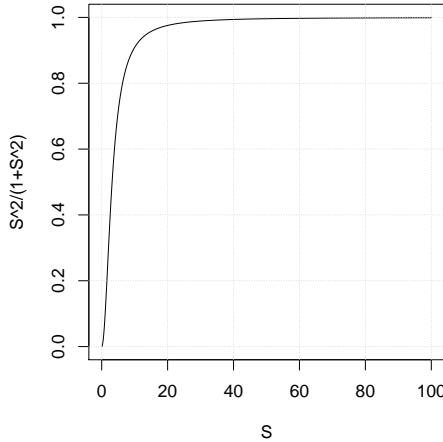
3. A logistic dependency on the state that reduces the variance if the state approaches an anticipated maximal level  $S_{max}$ . This approach corresponds to the case of sewer flows reaching the maximal pipe capacity.

$$g_i = \sigma_i \cdot S_i \cdot (S_{i,max} - S_i) \quad (3.12)$$

The reasoning behind these approaches is that the uncertainty of flow predictions is high during rain events and low in dry weather situations. Even the simple linear state dependency has proven to be a strong benchmark and hard to outperform by other diffusion structures. In addition, these approaches have the desirable property of scaling the diffusion to 0 if the states approach very small values so that negative forecasts are avoided. The state dependent diffusion term was therefore applied in several works (papers B,C and E).

Nevertheless, the state-dependent approaches are problematic for two reasons:

- Forecast uncertainty for dry weather periods and rain events is not addressed separately and lumped into the same parameters. This can lead to an underestimation of forecast uncertainties during rain periods if extended dry weather periods are included in the parameter estimation (papers B and E).
- Forecast uncertainty depends on the forecasted states themselves. This implies that, for example, an underestimation of observed flows by the forecast (in particular at the beginning of rain events) will also lead to an underestimation of forecast uncertainty (and vice versa). In addition, forecast uncertainty at the end of a rainfall event is typically small, although flows in the sewer system are still high. The state dependent diffusion structure will not be able to capture this behaviour.



**Figure 3.2:** Fraction  $\frac{S^2}{10+S^2}$  for different state values  $S$ . Realistic state values in the considered catchments are  $> 300m^3$ .

#### *Input Dependent Diffusion*

To overcome the above limitations, we propose a diffusion structure which describes forecast uncertainty as the sum of a constant dry weather uncertainty  $\sigma_{i,1}$  and an external forcing  $F$  scaled by a constant parameter  $\sigma_{i,2}$ . It is desirable to avoid negative state forecasts to ensure stability of the model. Therefore, we introduce a rational state dependency as shown in equation 3.13.

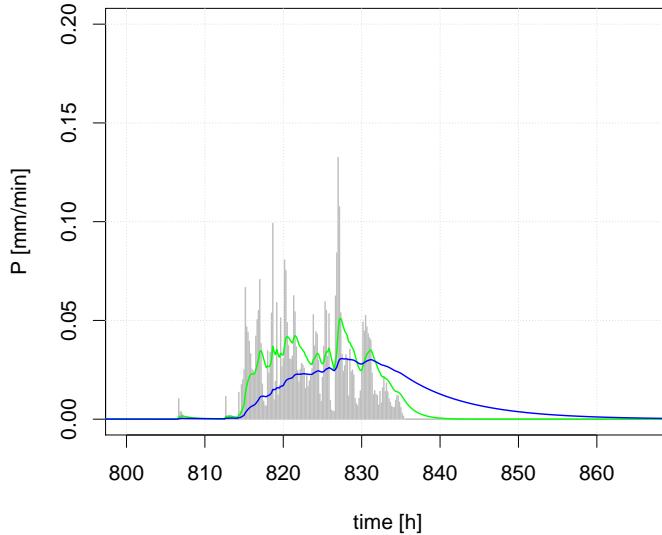
$$g_i = (\sigma_{i,1} + \sigma_{i,2} \cdot F) \frac{S_i^2}{c + S_i^2} \quad (3.13)$$

The constant  $c$  can be selected by the modeller. The fraction in equation 3.13 should ideally be one for any realistically anticipated state values and approach 0 as  $S_i$  approaches 0 (Figure 3.2). We here use a value of  $c = 1$ .

The diffusion structure proposed in equation 3.13 resembles the formulation of model bias used to describe simulation uncertainty in [DHS<sup>+</sup>13]. The time lagged rainfall input  $P_{k-l}$  is applied as external forcing  $F$  by the authors. The rainfall input, however, exhibits strong relative variations at irregular patterns. The diffusion scaling then becomes similarly irregular. Consequently, we apply a smoothed and time lagged rainfall input instead.

$$F_k = \lambda \cdot F_{k-1} + (1 - \lambda) P_{k-l} \quad (3.14)$$

The smoothing parameter  $\lambda \in [0, 1]$  and the time lag  $l \in [0, 1, \dots, \infty)$  are identified as part of the parameter estimation procedure. Figure 3.3 depicts the effect of smoothening rainfall observations using different smoothing parameters.



**Figure 3.3:** Rainfall observations in 10min resolution (grey) together with exponentially smoothed rainfall according to equation 3.14 with  $l = 0$  and  $\lambda = 0.9$  (green) and  $\lambda = 0.98$  (blue)

Instead of the rainfall input, we can also use a smoothed version of the observed model errors as external forcing for the diffusion term in equation 3.13. If we consider innovations (or one-step-ahead prediction errors)  $\epsilon_k$  for a forecast created at time step  $k - 1$ , we obtain

$$F_k = \lambda \cdot F_{k-1} + (1 - \lambda)\epsilon_k^2 \quad (3.15)$$

This may be interpreted as a model for generalized autoregressive conditional heteroskedasticity (GARCH) as applied in Econometrics ([Bol86]). This kind of external forcing for the diffusion term is expected to perform well in capturing the forecast uncertainty over short horizons. Over longer forecast horizons, the performance of this approach is questionable as the future model error is unknown at the time of forecast generation (paper G).

## 3.4 From State-Dependent Diffusion to Constant Diffusion - Lamperti Transformations

All structures of the diffusion term discussed in Section 3.3 include a dependency on the state value itself. Such structures are difficult to handle in *CTSM* as the

extended Kalman Filter requires higher order terms ([Ves98]). We can avoid this problem by applying so-called Lamperti transformations that move the state dependency from the diffusion term of the SDE to the drift term. We obtain a set of SDE's with a more complicated (and typically highly non-linear drift term) and a state-independent (and ideally constant) diffusion term.

As discussed by [Møl10] and [Iac08], the application of such transformations is also advisable for the simulation of SDE's. State dependent diffusions can, together with the drift term, put restrictions on the state space. Considering for example equation 3.6 with a linearly state-dependent diffusion, limits the state-space to positive values for  $S_1$  and  $S_2$ . When simulating the untransformed equation it is numerically not guaranteed that this limitation is obeyed. However, after an appropriate transformation the process lives on the entire real axis and numerical problems on the boundary of the domain are avoided ([Møl10]).

[Møl10] discusses the derivation of Lamperti transformations in detail. In this work, we give only the principle and summarize the results for the linear and the exponentially scaled state-dependent diffusion terms described in Section 3.3. In Section 3.3 we derive the transformation for the input dependent diffusion.

### 3.4.1 The General Lamperti Transformation

Consider the stochastic differential equation for the  $i$ -th state of a model with state vector  $X$

$$dX_{i,t} = f_i(X_t, u_t, t, \theta)dt + \sigma_i(X_t, u_t, t, \theta)d\omega_t = f_i(\cdot)dt + \sigma_i(\cdot)d\omega_t. \quad (3.16)$$

Now consider a transformed state  $Z_{i,t} = \phi(X_{i,t})$ . Assuming a diffusion process which does not depend on other model states, we can derive a SDE for the transformed state using Itô's lemma ([Øks98])

$$dZ_i = \left( \frac{\partial \phi}{\partial t} + f_i(\cdot) \frac{\partial \phi}{\partial X_i} + \frac{\sigma_i^2(\cdot)}{2} \frac{\partial^2 \phi}{\partial X_i^2} \right) dt + \sigma_i(\cdot) \frac{\partial \phi}{\partial X_i} d\omega_i \quad (3.17)$$

To obtain a transformed process with constant diffusion, the task is thus to define a transformation such that

$$\frac{1}{\sigma_i(\cdot)} = \frac{\partial \phi}{\partial X_i} \quad (3.18)$$

### 3.4.2 Lamperti Transformation for the Input Dependent Uncertainty Description

We consider the stochastic differential equation

$$dX_{i,t} = f_i(X_t, u_t, t, \theta)dt + \left( \varphi(F) \frac{X_{i,t}^2}{c + X_{i,t}^2} \right) d\omega_t \quad (3.19)$$

where  $\varphi$  is a function depending on an external forcing  $F$  (c.f. equation 3.13) and  $c$  is some constant.

As shown in equation 3.18 we find a transformed state  $Z_i = \phi(X_i)$  from

$$\frac{X_i^2 + c}{\varphi(F) \cdot X_i^2} = \frac{\partial \phi}{\partial X_i}. \quad (3.20)$$

Integrating equation 3.20 results in

$$Z_i = \frac{1}{\varphi(F)} \left( X_i - \frac{c}{X_i} \right). \quad (3.21)$$

The corresponding backtransformation is found as

$$X_i = \frac{\varphi(F) \cdot Z_i}{2} + \sqrt{\frac{(\varphi(F) \cdot Z_i)^2}{4} + c}. \quad (3.22)$$

Note that we only consider the positive root of the quadratic equation as the untransformed state  $X_i$  must be positive.

We find

$$\frac{\partial \phi}{\partial t} = 0, \quad \frac{\partial \phi}{\partial X_i} = \frac{X_i^2 + c}{\varphi(F) \cdot X_i^2}, \quad \frac{\partial^2 \phi}{\partial X_i^2} = -\frac{2c}{\varphi(F) \cdot X_i^3} \quad (3.23)$$

and apply Itô's lemma ([Øks98]) to obtain the transformed state equation:

$$dZ_i = \left( 0 + f_i(\cdot) \frac{c + X_i^2}{\varphi(F) \cdot X_i^2} - \frac{c \cdot \varphi(F) \cdot X_i}{(c + X_i^2)^2} \right) dt + 1 \cdot d\omega_i \quad (3.24)$$

We substitute  $X_i$  by equation 3.22 and use this new state equation for parameter estimation and simulation of the stochastic model in the transformed space. Also in the observation equation, we apply the backtransformation 3.22. Clearly, the backtransformation guarantees positive state values  $X_i$ .

The disadvantage of applying the Lamperti transformation is that a simple linear model equation can become strongly nonlinear in the transformed space and thus difficult to handle numerically.

### 3.4.3 Overview of Applied Lamperti Transformations

Table 3.1 gives an overview of the transformed state equations resulting for the diffusion structures applied in this work. The linear and exponentially scaled state dependencies are discussed in [Møl10] and [BTM<sup>+</sup>11].

**Table 3.1:** Diffusion structures, state transformations and transformed state equations for selected diffusion types

Structure of diffusion term $\sigma_i$	Backtransformation from transformed state $Z_i$	Transformed state equation
$\sigma_1 \cdot X_i$	$X_i = e^{\sigma_1 \cdot Z_i}$	$dZ_{i,t} = \left( \frac{f_i(X_t, u_t, t, \theta)}{\sigma_1 \cdot X_{i,t}} - \frac{1}{2}\sigma_1 \right) dt + d\omega_{i,t}$
$\sigma_1 \cdot X_i^\gamma$	$X_i = (\sigma_1 \cdot (1-\gamma) \cdot Z_i)^{\frac{1}{1-\gamma}}$	$dZ_{i,t} = \left( \frac{f_i(X_t, u_t, t, \theta)}{\sigma_1 \cdot X_{i,t}^\gamma} - \frac{1}{2}\sigma\gamma X^{\gamma-1} \right) dt + d\omega_{i,t}$
$\varphi \cdot \frac{X_i^2}{c+X_i^2}$	$X_i = \frac{\varphi \cdot Z_i}{2} + \sqrt{\frac{(\varphi \cdot Z_i)^2}{4} + c}$	$dZ_i = \left( f_i(X_t, u_t, t, \theta) \frac{c+X_i^2}{\varphi \cdot X_i^2} - \frac{c \cdot \varphi \cdot X_i}{(c+X_i^2)^2} \right) dt + d\omega_{i,t}$

### 3.5 The Relation between Lamperti and Data Transformations

Several options for transforming the flow observations are discussed in [BTM<sup>+</sup>11] and [BMM12]. In [BTM<sup>+</sup>11] a logarithmic transformation of the observations is suggested, based on the consideration that measurement uncertainty increases with the flow rate. The logarithmic transformation accounts for this behaviour as it leads to a multiplicative behaviour of the observation error.

We propose a different reasoning here. The parameter estimation procedure (c.f. Section 3.6.1) and the extended Kalman filter implemented in *CTSM* both assume that the innovations or one-step ahead predictions  $\hat{Y}_{k|k-1}$  are normally distributed with covariance matrix  $\sum_{k|k-1}^{yy}$ . Assuming that, as a result of the Lamperti transform, the state prediction  $\hat{Z}_{k|k-1}$  from equation 3.17 is normally distributed, the form of the observation equation (in our case linear, see equation 3.7) and the backtransformation (table 3.1) define which transformation should be applied to the observations. The exact same transformation will be applied also to the observation equation, and we need to obtain approximate normality for the resulting forecast  $\hat{Y}_{k|k-1}$  (c.f. Section 3.6.1).

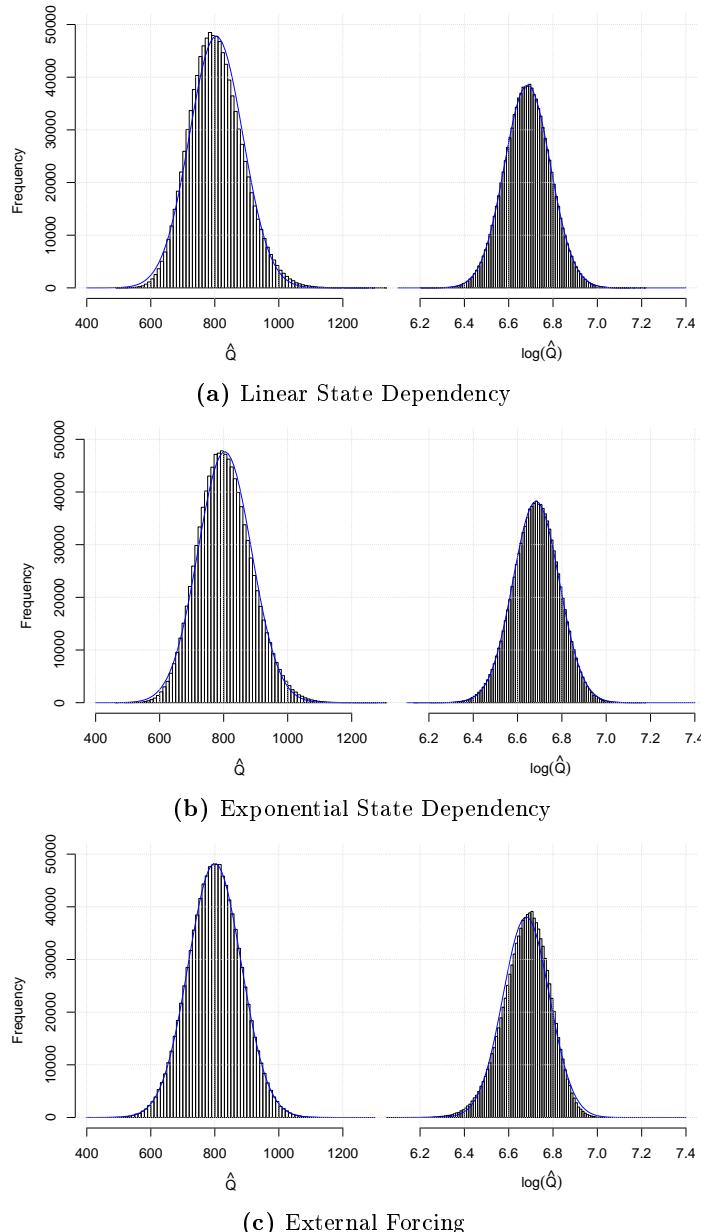
To illustrate this problem, we perform an experiment by sampling from an assumed normal distribution for  $Z$ . We select a time constant  $K = 5\frac{1}{h}$  and mean and variance for the (imagined) state  $Z$  in such a way, that the flow prediction

$$\hat{Q} = \frac{1}{K} \cdot \hat{X} = \frac{1}{K} \cdot \phi^{-1}(\hat{Z}) \quad (3.25)$$

resulting from the backtransformed state  $X$  has a mean of  $800m^3/h$  and a width of  $320m^3/h$  of the 95% prediction interval.

We use a sample size of  $10^6$  values for  $Z$ . Considering the different variants of  $\phi$  resulting from different Lamperti transformations, we assess whether  $\hat{Q}$  and its log-transformed pendant  $\log(\hat{Q})$  can be assumed normally distributed. Figure 3.4 shows the corresponding histograms.

We see from figures 3.4a and 3.4b that for the linear state dependency and the exponentially scaled state dependency (we consider  $\gamma = 0.8$ ), a log transform of the observation equation (and thus also the observations) will lead to a predictive distribution which is closer to normality. This is a shortcoming of papers B and E where no transformation was applied to the observations. We can, however, also notice in figures 3.4a and 3.4b that the distribution for  $\hat{Q}$  is only slightly skewed, if the standard deviation is not too big. This is also the case in the articles B and E.



**Figure 3.4:** Histograms of flow predictions  $\hat{Q}$  and log-transformed flow predictions  $\log(\hat{Q})$  considering different Lamperti transformations for the normally distributed state  $Z$

For the new state dependency that was introduced for the consideration of external forcings (we consider  $c = 1$  and  $\varphi = 1$ ), we obtain the reverse picture (Figure 3.4c). The predictive distribution for  $\hat{Q}$  is almost perfectly normal, while the distribution for  $\log(\hat{Q})$  is slightly skewed. Again the deviation is limited, but no transformation should be applied to the observation equation (and the observations) in combination with this Lamperti transformation.

We conclude that the choice of transformation for the observations should be based on considerations concerning the assumptions on normality for the Kalman filtering and parameter estimation procedures rather than temporal variations of the measurement error.

Finally, the choice of data transformation can also be affected by the quality of the observed data. Negative flow observations (observed in paper E, for instance) make it impossible to use a logarithmic transformation unless some smoothing is applied to the data which in turn introduces time lags into the observations.

## 3.6 Parameter Estimation in Stochastic Grey-box Models

### 3.6.1 Maximum Likelihood Estimation

The parameter estimation method implemented in *CTSM* is based on the maximum likelihood principle ([Mad08, Paw01]). This method is comprehensively documented in [KM03] and [KMJ04]. For time series data, the likelihood function is commonly expressed as a product of one step ahead conditional densities. Most commonly when working with *CTSM*, we use a frequentist approach where the only aim is to find the best performing parameter set. We do not consider parameter uncertainty when generating probabilistic forecasts. The underlying assumption is that commonly sufficient data is available to identify the correct set of parameters. However, it is also possible to consider prior information in the parameter estimation procedure (see Section 3.6.2).

The likelihood for a parameter set  $\theta$  given a set of  $l$ -dimensional observations  $y_i \in \mathbf{R}^l$

$$Y_n = [y_n, y_{n-1}, \dots, y_1, y_0] \quad (3.26)$$

can be expressed as

$$L(\theta, Y_n) = \left( \prod_{k=1}^n p(y_k | Y_{k-1}, \theta) \right) p(y_0 | \theta) \quad (3.27)$$

Assuming that the one-step ahead forecast errors (innovations)  $\epsilon_k = y_k - \hat{y}_{k|k-1}$  are iid and normally distributed with mean 0 and covariance matrix  $R_{k|k-1} = V(y_k | Y_{k-1}, \theta)$ , the likelihood function can be rewritten as

$$L(\theta, Y_n) = \left( \prod_{k=1}^n \frac{\exp\left(-\frac{1}{2}\epsilon_k^T R_{k|k-1}^{-1} \epsilon_k\right)}{\sqrt{\det(R_{k|k-1})}(\sqrt{2\pi})^l} \right) p(y_0 | \theta) \quad (3.28)$$

Conditioning on  $y_0$ , equation 3.28 can be solved as an optimization problem. The one-step ahead forecasts and hence the innovations are obtained through extended Kalman filtering. This implies that during parameter estimation every forecast step is preceded by a state updating to the current observation.

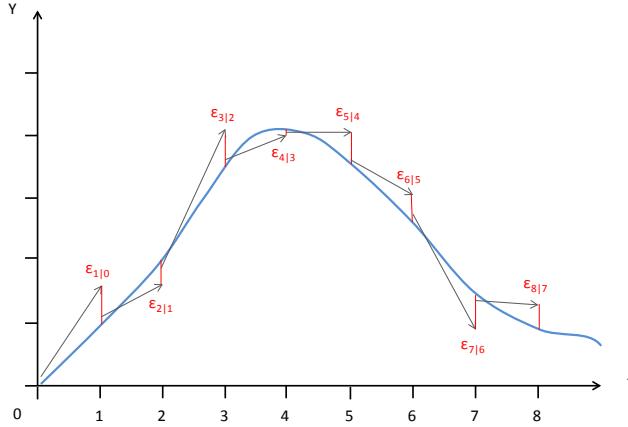
The updating ensures that the model follows the realisation of the stochastic process described by the observations. We can avoid biased parameter estimates resulting from an adjustment of the simulation model to a single realisation of the (stochastic) reality and obtain a better description of the short term dynamics. We assume a normal distribution of the innovations which is usually a realistic assumption.

The principle is illustrated in Figure 3.5. The model is updated to the observations at every time step (each forecast starts closer to the observation than the previous forecasted value). In simple terms, the model parameters are identified by minimizing the average of the innovations shown in red.

The downside of this approach is that occasionally the state updating becomes to be dominant in the parameter estimation. The observation noise is then estimated excessively small and the model states are perfectly updated to the flow observations at every time step while the likelihood values become less sensitive to the physical model parameters which are then difficult to identify.

Furthermore, the mean of the one-step ahead predictions is generally obtained by evaluating only the drift term of the SDE as we assume a locally linear behaviour. This is not necessarily equal to the behaviour of the actual stochastic process.

Nevertheless, the Maximum likelihood based estimation has proven to provide robust results in a multitude of applications (for example [ARM13, BMPN00, BTM<sup>+</sup>11, Møl10, BM11]).



**Figure 3.5:** Estimation principle for the Maximum Likelihood approach. The true flow (with observations at time point  $T = 1, \dots, 8$ ) is shown blue, arrows symbolize one step-ahead model forecasts starting from an updated model state, red lines innovations  $\epsilon$ .

### 3.6.2 Maximum A Posteriori Estimation

In the parameter estimation procedure provided by *CTSM* prior information can be considered by using the maximum a posteriori (MAP) estimate. The likelihood function is in this case extended with a penalty for the deviation of the parameter from its prior mean. Considering  $\theta \in \mathbb{R}^p$ ,  $\mu_\theta = E(\theta)$ ,  $\Sigma_\theta = V(\theta)$  and  $\epsilon_\theta = \theta - \mu_\theta$ , we obtain

$$L(\theta, Y_n) = \left( \prod_{k=1}^N \frac{\exp\left(-\frac{1}{2}\epsilon_k^T R_{k|k-1}^{-1} \epsilon_k\right)}{\sqrt{\det(R_{k|k-1})} (\sqrt{2\pi})^l} \right) p(y_0|\theta) \frac{\exp\left(-\frac{1}{2}\epsilon_\theta^T \Sigma_\theta \epsilon_\theta\right)}{\sqrt{\det(\Sigma_\theta)} (\sqrt{2\pi})^p} \quad (3.29)$$

Again, conditioning on  $y_0$ , the MAP estimate is obtained as ([KM03, KMJ04])

$$\hat{\theta} = \arg \min(-\ln(L(\theta|Y_n, y_0))). \quad (3.30)$$

Prior information can be assigned to selected parameters only by using non-informative priors with large standard deviations for the other parameters. This estimation approach can be used to avoid overly small estimates of the observation noise or if the data do not provide sufficient information to identify all the parameters.

### 3.6.3 Estimation Based on Multi-step Forecast Verification

The aim to apply the stochastic grey-box models for multi-step ahead forecasts motivated the search for a robust estimation method that allows us to identify a model which also captures the long-term dynamics of the system. This is not always guaranteed with the estimation approach described in Section 3.6.1. A first step in this direction was made in paper A and the estimation procedures were documented in paper C.

As we intend to use the runoff forecasting models for multi-step forecasts in a real-time control setting, the idea behind this approach is to estimate the model parameters based on the quality of multi-step forecasts rather than one-step ahead forecasts as described in Section 3.6.1. In this setting, multi-step forecasts are generated by repeated extended Kalman filtering without updating, as implemented in *CTSM* ([KM03]), and assumed normally distributed with variance  $R_{i+j|i}$ .

The objective function at time step  $i$  is found as a weighted average of the scores  $SC$  for flow prediction horizons  $i + j$  up to the maximal horizon  $i + k$

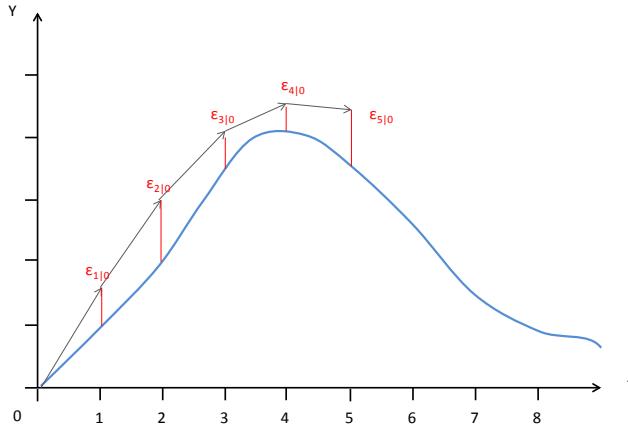
$$SC_i = \frac{1}{\sum_{j=1}^k (k - j + 1)} \left( \sum_{j=1}^k (k - j + 1) \cdot SC_{i,j} \right). \quad (3.31)$$

This is illustrated in Figure 3.6. Shorter horizons receive more weighting in the parameter estimation procedure, as the objective in practice is to generate forecasts of runoff volume for the different horizons  $i + j$  to  $i + k$ . The runoff volume for a given horizon is an integral over the flow predictions for this horizon and all previous horizons. Flow predictions for shorter horizons thus affect the runoff volume forecast for more horizons than the flow predictions for longer horizons and should be given higher weighting.

We apply the *CRPS* (Section 2.2.3) as the score function  $SC$  for the flow predictions. The approach is documented as "model D" in [LMM14]. Advantages of the approach are a more robust parameter identification and that the model is estimated according to the forecasting objective, as the model is estimated based on multi-step forecasts rather than one-step ahead forecasts. A comparison in Section 3.6.5 shows that we obtain lower point forecast errors with this estimation approach.

However, the approach also has several disadvantages.

- *First*, although practically appealing, parameter estimation based on a



**Figure 3.6:** Estimation principle for the forecast based approach. The true flow (with observations at time point  $T = 1, \dots, 8$ ) is shown blue, arrows symbolize model forecasts generated at  $t = 0$ . In this example we consider a horizon  $k = 5$ . The score value for the first time step  $SC_0$  is found by an evaluation of the probabilistic flow forecasts for every horizon  $j = 1, \dots, 5$

score function removes the theoretical support of the likelihood principle and thus provides a less well founded interpretability of prediction bounds, for example

- *Second*, the estimation approach described in [LMM14] relies on a normality assumption for multi-step predictions, which cannot necessarily be assumed on longer forecast horizons.
- *Third*, when generating multi-step forecasts using the EKF setup in *CTSM*, we do not actually simulate the stochastic process described by the SDE but only the drift term of the state equations ([KM03]). The expected value of the multi-step forecast used during parameter estimation and the expected value from an actual stochastic simulation of the SDE's (see Section 3.7) are therefore not necessarily the same.

The latter two issues can be avoided if multi-step ahead forecasts during parameter estimation are generated not by extended Kalman filtering but by an ensemble-based approach relying on a stochastic simulation of the SDE's as described in Section 3.7. Furthermore, as we observe an underestimation of forecast uncertainties by the models estimated using the forecast based approach

(papers B, C, E), we can replace the  $CRPS$  as a score function in equation 3.31 by another score function, for example, a Gaussian density.

### 3.6.4 Numerical Optimisation

All of the parameter estimation approaches described above rely on a numerical minimization of the objective function. We have in the first works applied Genetic algorithms ([Whi94], paper B) and the *DDS* algorithm ([TS07], papers B and E) because the methods are insensitive to missing values in the objective function evaluation. This behaviour was necessary because the previous version of *CTSM* would frequently produce errors for "random" parameter combinations. We have applied these algorithms repeatedly starting from the optimum of the previous optimization run to ensure that a parameter set near the true optimum was identified.

As a result of improvements in *CTSM* ([JKB<sup>+</sup>13]) the application of the *PORT* algorithm ([Gay90]) became possible in later works (papers D and G). In paper G we combine this algorithm in series with the *DDS* algorithm if integer variables need to be optimized.

### 3.6.5 Comparing Estimation Approaches for a Sample Catchment

Extending the discussion in paper C, in this section we compare the forecast quality for three of the catchments considered in paper E (Amager East - EAm, Colloseum - COL, Kløvermarken - Klo). We consider a cascade of three reservoirs with time-invariant parameters as shown in paper E. We consider point and probabilistic forecast quality for models estimated using the Maximum Likelihood approach (ML, Section 3.6.1) and the forecast based approach using the  $CRPS$  (Section 3.6.3). Forecast quality is compared on a 120min or 60 time step horizon for the four validation events described in paper E.

Figure 3.7 shows the point and probabilistic forecast skill for the different events and catchments while tables 3.2 and 3.3 depict the reliability *Rel* and *ARIL* (c.f. 2.2.1) values for 90% prediction intervals, respectively. Point forecast are derived using the median of the probabilistic multi-step forecasts (c.f. 3.8).

We see that the model estimated using the forecast based approach (Section 3.6.3) provides better point (*NSE*, *PI*) and probabilistic forecast skill ( $CRPS$ ) on the 60 time step horizon than the model estimated using the ML approach.

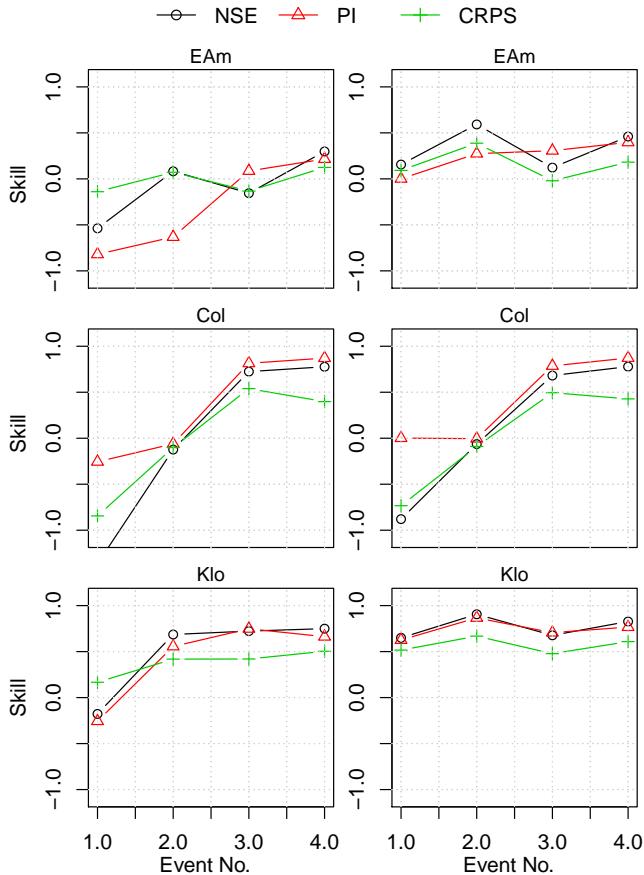
This is exactly the intended effect of performing parameter estimation based on forecasts. The forecast based estimation does, however, provide less reliable forecasts than the ML approach. The forecast uncertainty (table 3.3) is estimated smaller leading to smaller reliability values (Table 3.2). Also this tendency has been observed before (papers B, C, E).

**Table 3.2:** Reliability (*Rel*) of 90% prediction intervals on a 120min horizon for three catchments considered in paper E

Catchment	Estimation approach	Event 1	Event 2	Event 3	Event 4
EAm	<i>ML</i>	59%	76%	65%	49%
EAm	<i>CRPS</i>	60%	68%	72%	65%
COL	<i>ML</i>	26%	49%	76%	40%
COL	<i>CRPS</i>	40%	52%	88%	37%
Klo	<i>ML</i>	75%	79%	79%	65%
Klo	<i>CRPS</i>	87%	96%	89%	81%

**Table 3.3:** Average Interval Length (*ARIL*) of 90% prediction intervals on a 120min horizon for three catchments considered in paper E

Catchment	Estimation approach	Event 1	Event 2	Event 3	Event 4
EAm	<i>ML</i>	72%	73%	78%	57%
EAm	<i>CRPS</i>	118%	117%	117%	90%
COL	<i>ML</i>	55%	54%	83%	136%
COL	<i>CRPS</i>	84%	77%	117%	206%
Klo	<i>ML</i>	87%	78%	73%	82%
Klo	<i>CRPS</i>	251%	213%	182%	196%



**Figure 3.7:** Forecast skill (NSE, Persistence Index PI, normalized CRPS) in the EAm, Col and Klo catchments for the validation events in paper E for models estimated using the ML based estimation approach (left, c.f. Section 3.6.1) and the forecast based estimation approach (right, c.f. Section 3.6.3)

### 3.7 Simulation of Stochastic Differential Equations

Simulations of stochastic differential equations distinguish themselves from those of ODE's in that we need to consider a coupled drift and diffusion term system. There is no single solution for a SDE but only a set of realisations which as a whole have statistical properties describing the stochastic process. Simulation methods for SDE's are described by [KP99] and [Iac08]. In this work, we mainly apply such stochastic simulations for the generation of ensemble-based probabilistic forecasts (see Section 3.8).

The simplest simulation method for SDE's is the Euler-Maruyama scheme ([KP99]). The considered time steps  $\Delta t$  are discretized into increments of equal length  $h$ . The simulated state for the  $i$ -th increment  $\hat{X}_{t+h \cdot i|t}$  is then found as

$$\hat{X}_{t+h \cdot i|t} = \hat{X}_{t+h \cdot (i-1)|t} + f\left(\hat{X}_{t+h \cdot (i-1)|t}, u_{t+h \cdot i}, t + h \cdot i, \theta\right) \cdot h + \quad (3.32)$$

$$\sigma\left(\hat{X}_{t+h \cdot (i-1)|t}, u_{t+h \cdot i}, t + h \cdot i, \theta\right) \cdot \Delta W_{t+h \cdot (i-1)} \quad (3.33)$$

$\Delta W_{t+h \cdot (i-1)}$  is a realisation of an increment of the Wiener process which is equivalent to a random normal number with mean 0 and variance  $h$ . We apply the Euler-Maruyama scheme for the generation of probabilistic forecasts in papers E and F. A discretization step of  $h = \Delta t / 100$  is used for the simulations in these cases.

The Euler-Maruyama scheme is easily implemented but does have drawbacks in that very small increments  $h$  may be required to properly capture the physical and stochastic behaviour of the simulated process. We can avoid issues in modeling time-varying diffusions and make the state variables defined over the whole real domain by applying Lamperti transformations to our model (see Section 3.4). The resulting transformed state equations, however, are typically non-linear and stiff in some cases. Improved solution methods that are valid in such cases are implicit (such as the single step backward Euler method (SSBE, [HMS02]) - see paper G), account for the Jacobian of the drift term (such as the weak exponential scheme ([Mor05])) or apply predictor-corrector schemes ([BLP08]).

## 3.8 Generating Multi-step Probabilistic Forecasts from Stochastic Grey-box models

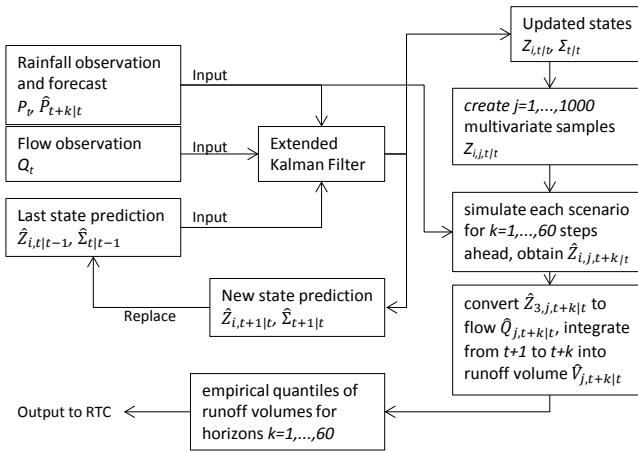
The eventual purpose of the considered stochastic grey-box models is to generate probabilistic forecasts of runoff volume for real-time control. The relevant forecast horizon depends on the current state of the system, in particular the current basin fillings. Flow forecasts must therefore be generated for a multitude of forecast horizons and runoff volumes must be derived from these.

In a probabilistic sense, we need to account for the correlation between the flow forecasts for different horizons. Furthermore, the runoff volume is a sum of the flow forecasts for different horizons. The distribution for this sum of random variables is not necessarily straight forward to derive. In paper C we have used a sampling approach while subsequently we have moved to an ensemble (or scenario) based approach (papers B, E and G) to derive the predicted distribution of runoff volume.

We aim to generate a probabilistic  $k$ -step ahead forecast of runoff volume  $\hat{V}_{t+k|t}$  starting at time step  $t$ . We follow the scheme shown in Figure 3.8:

1. Assuming the last flow observation  $Q_t$  being available at  $t$ , we start with the updated model states  $Z_{t,t}$  provided by the extended Kalman filter. We assume these to be multivariate normal with variance matrix  $\Sigma_{t|t}$  (this assumption is also made in the Kalman filtering).
2. We generate  $N$  multivariate samples from this distribution (using the R-package ([RCT13]) *MASS* [VR02]). These serve as a starting point for the simulations.
3. We use the Euler-Maruyama scheme as described in 3.7 to generate  $N$  simulations of the Lamperti-transformed state equations up to  $t + k$ .
4. We integrate each flow simulation into a runoff volume and obtain a sample of  $N$  runoff volumes  $\hat{V}_{t+k|t}$ .
5. We empirically describe the forecasted distribution using quantiles from the sample of runoff volumes.

The observation noise in the proposed scheme can be disregarded as it is not relevant for decision making in real-time control. Correlation between the runoff forecasts for different horizons does not need to be considered explicitly as it is provided by the scenario simulations (other than if we would consider different models for different forecast horizons, for example). Moreover, we simulate the



**Figure 3.8:** Ensemble approach for the generation of probabilistic forecasts of runoff volume  $\hat{V}_{t+k|t}$  for different forecast horizons  $k$

stochastic process described by the state equations 3.6 without any distributional assumptions.

However, the approach also has drawbacks. First, the computational effort is rather high as ensemble simulations need to be performed. Nevertheless, the approach is feasible and suitable for an on-line application as demonstrated in paper F. Second, the forecast approach is not consistent with the parameter estimation approaches described in Section 3.6. These rely on a forecast solely of the drift term of the model. We may consider this a drawback of the parameter estimation rather than the forecasting procedure which suggests further investigations in this direction.

### 3.9 Relation to Other Uncertainty Techniques Applied in Hydrology

A multitude of uncertainty techniques have been proposed in hydrology and other fields. These include Bayesian techniques, *GLUE* and other methods which are described below.

The grey-box modelling approach discussed in this work distinguishes itself in its very strong focus on on-line applications. It provides techniques for parameter estimation and data assimilation with very limited computational effort and

the ability to generate probabilistic forecasts and account for model deficiencies during parameter estimation as an inherent feature. Drawbacks include the assumption of normality in the filtering and parameter estimation procedures and the limit on model complexity which is imposed by the EKF and the requirement to solve a multivariate system of SDE's.

**Bayesian techniques** are very widely applied in hydrology. These mainly distinguish themselves from the grey-box approach in the explicit consideration of parameter uncertainty during parameter estimation and simulation, while the grey-box approach 'lumps' parameter, structural and input uncertainty into the diffusion term. The use of prior information is common in the Bayesian school while it is an exception in the grey-box approach. Mostly, Bayesian approaches have a strong focus on uncertainty analysis in an off-line setting and structural model identification. Commonly, they require Monte Carlo simulations during parameter estimation and simulation which can be seen as their main drawback.

Uncertainty formulations used in conjunction with Bayesian techniques in the literature show a wide range of complexities. In the most simplistic case, errors are lumped into a single output error term  $e$  which is assumed iid (and typically normally distributed). The observations  $Y$  are then obtained depending on some model function  $g$  as

$$Y_t = g(X_t, u_t, t, \theta) + e \quad (3.34)$$

More advanced approaches account for input uncertainty using rainfall multipliers ([KKFT06],[RKK<sup>+</sup>10], [SBK13]) and structural uncertainty using time varying parameters ([RKK<sup>+</sup>10]). A large number of parameters is used in these approaches which makes them unattractive for on-line modelling purposes. This drawback is removed by adding a time varying bias term  $D$  to equation 3.34 in [RS12]. The bias captures structural model uncertainties and input uncertainties. A comparison with the grey-box approach in paper D indicates good simulation performances of this approach. The computational effort, however, is challenging even in this case.

The **GLUE** approach ([Bev93]) is widely used for uncertainty estimation in urban hydrology ([TBJSJ08, DMK<sup>+</sup>12, VMDM13]). It is methodologically similar to the Bayesian approaches but does not restrict itself to the theoretical likelihood framework. Due to the similarity to the Bayesian methods, we do not discuss the method further here.

Other approaches combine Bayesian estimation techniques and data assimilation using ensemble ([VDG<sup>+</sup>05]) or particle ([MDS12, VTDS13]) filtering approaches. The ensemble / particle filtering makes these approaches more computationally demanding than the grey-box approach. However, they are closer to an on-line application than the approaches above as the model can recur-

sively be updated to new observations without considering the whole time series. In particular the approach presented by [VDG<sup>+</sup>05] is close to the grey-box approach in that it offers the possibility to use time-varying state noise descriptions and may be considered for the development of more complex conceptual model structures.

Other commonly applied uncertainty techniques focus on a **post-processing** of model results. The downside of these methods is that structural and input uncertainty is not explicitly accounted for during the identification of the actual model. On the other hand, these methods can perfectly be used in combination with complex physical models which is not true for the grey-box approach (unless a grey-box model is used as post-processing method rather than as forecasting model).

Among others, such approaches use time series models in combination with data assimilation in the physical model ([MS05]), Bayesian model averaging (BMA) ([HFZ13]) or non-parametric methods ([Pin07, PMN<sup>+</sup>09]). For the consideration of forecasts over multiple horizons either multivariate distributions (and suitable transformations of the model residuals) ([PMN<sup>+</sup>09, HFZ13]) or copulas ([TPM13]) must be considered to correctly capture correlation between different forecast horizons. As the correlation structure is likely to be time varying, it might be necessary to estimate it recursively as described in [PMN<sup>+</sup>09] and paper C. The feature of describing correlation between forecast horizons is inherent for approaches that generate probabilistic forecasts based on direct simulations of the model (such as the grey-box and the Bayesian approaches).

## CHAPTER 4

# Rainfall Inputs for On-line Runoff Forecasting

---

## 4.1 Requirements to Rainfall Input for On-line Runoff Forecasting

Runoff forecast models rely on rainfall observations and forecasts as input. This raises the issue of how good the rainfall input needs to be. Several criteria are important in this context:

- spatial resolution
- temporal resolution
- reliable measurement of the rainfall process
- reliable operation
- forecast quality

The criteria of spatial and temporal resolution have been intensely discussed in the literature. We have partly summarized this discussion in paper B. In

addition, [Sch91] gives guidelines on the resolution rainfall measurements should have for design and operational applications in urban hydrology.

Reliability of rainfall measurements may be considered the key criterion in terms of a forecast model based operation of urban drainage systems. We consider reliability both in terms of operational availability and accuracy of the available measurements and forecasts. This matter is discussed further in Section 4.2. Identifying periods with an acceptable quality of rainfall data was the main issue during the work on paper E.

Finally, it seems obvious that a runoff forecast in an on-line setting profits from using a rainfall forecast. However, due to the reaction time of a catchment, the expected future runoff from a catchment up to a certain horizon will be well determined by the measured rainfall. The extent of this effect depends on the characteristics of the catchment (size, shape, degree of sealing and location of sealed areas).

In Section 4.3 we describe an experiment comparing runoff forecast quality on a 2 hour horizon with perfect rainfall forecast and without rainfall forecast and conclude that in some cases reasonable runoff forecasts can actually be obtained without rainfall forecast. On the contrary, [TR13] analyse the runoff forecast quality that can be obtained using radar rainfall forecasts in a 80ha urban catchment and find that the runoff forecast skill is very limited for horizons exceeding 60 minutes. We are not aware of systematic investigations into situations in which rainfall forecasts are beneficial for short term on-line runoff forecasting (in terms of catchment characteristics and forecast horizons) and what quality of rainfall forecasts is required. We consider this an important item of future research.

## 4.2 Raingauge vs. Radar Rainfall Input

### 4.2.1 Advantages and Disadvantages

In this thesis we have applied rainfall inputs from rain gauge and C-band radar measurements for runoff forecasting up to horizons of 120 minutes. Rainfall inputs from X-band local area weather radars (LAWR) were applied at the start of the project but found to not provide useful input for runoff forecasting. New processing methodologies have been developed recently which are likely to lead to more reliable rainfall observations with LAWR ([NTR12, NJR13]).

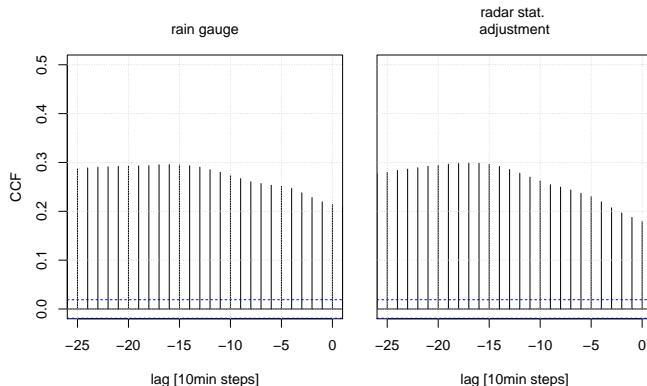
Below we discuss the advantages and disadvantages of rain gauge and radar rainfall measurements in terms of the criteria described in Section 4.1.

### Spatial Resolution

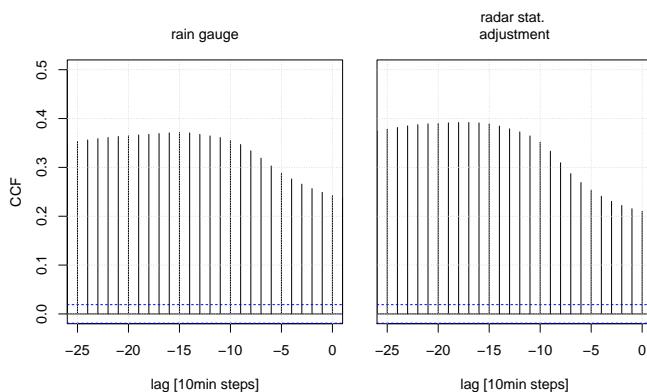
Depending on the characteristics of the rain event, rainfall can vary strongly in space. An improved spatial resolution may generally be considered an advantage of radar rainfall measurements as compared to rain gauges. The former will provide an average of the rain intensities present within a radar pixel while the latter provide only a point measurement. Having only a point measurement may lead to false assumptions about the average rain intensity over a catchment and time displaced rainfall observations ([BGM13]).

In paper B we demonstrate that radar rainfall measurement as input to the stochastic runoff forecasting models has the potential to provide improved runoff forecasts in comparison to rain gauges. It cannot be clearly determined from the results, whether the improvement results from an improved rainfall forecast or an improved representation of the areal rainfall by the radar. Considering the results in Section 4.3 that show that reasonable runoff forecasts on a 120min horizon can be obtained without rainfall forecast in several cases in the bigger catchments (750ha) and the fact that the catchments considered in paper B are even bigger (1300 and 3000ha), we see an indication that spatial resolution may have an important role.

Figures 4.1 and 4.2 support this assumption. The plots show that the mean areal catchment rainfall derived from the time-statically adjusted radar data considered in paper B yields a similar or better correlation to the measured runoff from the catchment than the mean areal rainfall derived from rain gauges. Notably, the highest correlation is identified for lags of approximately 150 minutes between rainfall and runoff measurements. This is longer than the considered forecast horizon of 100 minutes. This result remains valid after prewhitening of the time series ([MS05]), the peak in the cross correlation is then identified for lags between 100 and 140 minutes.



**Figure 4.1:** Cross correlation (CCF) between mean areal rainfall derived from rain gauges (left) and time-statically adjusted radar rainfall measurements (right) and runoff measurements in the Ballerup catchment (see paper B)



**Figure 4.2:** Cross correlation (CCF) between mean areal rainfall derived from rain gauges (left) and time-statically adjusted radar rainfall measurements (right) and runoff measurements in the Damhusåen catchment (see paper B)

### Temporal Resolution

A clear advantage for one of the measurement methods cannot be identified in terms of temporal resolution. The available resolutions depend in both cases on the applied measurement principles, processing algorithms and the agreed protocols for on-line data transfer. In this work we were largely bound to a temporal resolution of 10 minutes for the rainfall measurements, corresponding to the available resolution of the C-band radar measurements.

In [Sch91] a temporal resolution  $\Delta t$  for the rainfall data that depends on the concentration time  $t_c$  of the catchment as

$$\Delta t = \frac{1}{3}t_c \dots \frac{1}{5}t_c \quad (4.1)$$

is suggested. A coarse estimate for  $t_c$  in seconds can be found from the total catchment area  $A$  in  $m^2$  as  $t_c = \sqrt{A}$ . Applying this to the Ballerup and Damhusåen catchments considered in papers A, B, C, D and G, we see that a 10 minute resolution of rainfall measurements is sufficient to model runoff from these catchments. This is, however, not necessarily true for the smaller catchments considered in papers E and F.

### Reliable Measurement of the Rainfall Process

In a measuring sense, it is intuitive to assume the radar rainfall measurements less reliable for an on-line operation of drainage systems. There are issues in the conversion from the quantity measured by the radar (reflectivity for the C-band radar, drop counts for the LAWR) to rain intensities and in the attenuation of the signal (see for example [KVPT12, BK13, NJR13]).

However, as [BGM13] discusses, a time displacement of the rainfall measurement (which may, for example, result from measuring only at a point location) can have the same impact on runoff-forecasting models as a strong bias in the rainfall measurement. This issue needs to be considered when applying rain gauge observations as input to runoff forecasting models.

Moreover, in paper B we demonstrate that the rainfall input applied for runoff forecasting with automatically calibrated models does not need to resemble the observations on the ground. [TR13] evaluate the runoff forecast quality that can actually be obtained for a small catchment (80ha) using radar rainfall input and find reasonable runoff forecasts up to a horizon of 60 minutes.

### Operational Reliability

Operational reliability needs to be ensured for both rain gauge and radar measurements. [ML13] show that a gauge network can be operated reliably for on-line purposes also over long periods. It may be a benefit that the maintenance of the rain gauges is within the responsibilities of the operator of the

drainage network. This is quite often not the case for the weather radars. A careful processing of the radar rainfall measurements is required before using them as input for runoff forecasting models.

#### 4.2.2 Merging Different Rainfall Inputs

Radar and rain gauges give different perspectives on the rainfall process. While the radar provides the spatial distribution of the rainfall, there are problems with obtaining accurate rain intensities due to spatially and temporally varying drop size distributions and signal attenuation ([BK13]). Rain gauges, on the other hand, provide very local measurements of the rainfall process, which, depending on the density of the gauge network, may not be spatially representative.

A multitude of publications is available on how to combine radar and rain gauge measurements into an improved observation of the 'true' rainfall process. We refer to the overview provided by [GD09] and the introduction of paper B. Common merging approaches focus on the adjustment of the radar measurements to make them more "similar" to rain gauge observations.

In paper B we demonstrate that this may not be necessary for on-line runoff forecasting models that are calibrated using radar rainfall input. On the contrary, if the adjustment is performed in an unsuitable way, for example, with strong time variations in the adjustment factors, the resulting rainfall measurement may yield worse results as input for runoff forecasting than the original radar measurement (paper B).

We have evaluated the performance of a simple statistical combination method presented by [GHL02] in paper A. In this approach, a simple autoregressive state space model

$$X_t = A \cdot X_{t-1} + e_t \quad (4.2)$$

$$Y_t = C \cdot X_t + s_t \quad (4.3)$$

is created for the rainfall process  $X_t$  with radar and rain gauge observations  $Y_t$ , parameter matrices  $A$  and  $C$  and assumed normal error vectors  $e_t$  and  $s_t$  (see [GHL02]). The resulting combined rainfall measurement corresponds to the rainfall state  $X_t$ , which in paper A is gridded with a resolution of 2x2km. The different rainfall measurements  $Y_t$  are merged with the rainfall state  $X_t$  using an ordinary Kalman filter and the model parameters are estimated by maximizing the likelihood of obtaining the radar and rain gauge measurements  $Y_t$ .

This approach is very appealing because physical information about the rainfall process can, in principle, be incorporated in the model  $A$  and because any kind

of measurement providing information about rainfall can be included in the merging process given a suitable formulation of the observation equation in  $C$ .

The approach does, however, in the given form also have disadvantages in that we assume normality for the rainfall states  $X_t$  and observations  $Y_t$ . This assumption certainly does not hold for the rainfall process which is bounded at 0. Moreover, the Kalman filter algorithm cannot handle the consideration of areas corresponding to the full size of a C-band radar image. Such an image has an extent of 240x240 pixels and the resulting covariance matrix for  $e_t$  would be of dimension 57600x57600. Processing of the full radar image, however, is necessary to generate rainfall forecasts from the radar observations. This is also a disadvantage of the adjustment methodology presented by [Tod01] and [WOSo<sup>+</sup>13] which is based on the Kalman filter.

Finally, as a result of the discussion in paper B, we concluded that radar adjustment in particular, but also any procedure for merging different rainfall measurements in general, should focus on the final purpose of the resulting rainfall information. This may, for example, be the forecasting of runoff in an on-line setting. To avoid the above problems, we may apply geostatistical merging techniques as described by [BRH13, GD09]. We can then apply simple rainfall runoff models for one or several considered catchments and identify the parameters of the rainfall merging algorithm as a part of the overall parameter calibration for the rainfall runoff model by maximizing the likelihood of the the runoff observations. Such a setting could also be laid out with time varying parameters for the rainfall merging algorithm. These can be identified using a simple state space model layout and an (extended) Kalman filter (see, for example, [BDLY01]).

## **4.3 The Effect of Rainfall Forecasts on Runoff Forecast Quality - An Example**

### **4.3.1 Problem Description**

We have in Section 4.1 questioned whether rainfall forecasts always provide benefit for the generation of runoff forecasts for short horizons. This is underlined by the observation that runoff forecasts based on rain gauge measurements show a strong performance in paper B as compared to the forecasts using radar measurements and forecasts.

In this section we follow up on this discussion by analysing the performance of the stochastic runoff forecast models for the six subcatchments and 8 rain events

considered in paper E. We use the same dataset as in paper E and refer the reader to this article for a description of the catchments. Rainfall information is available from C-band radar measurements in a 10min resolution, while a 2min time step is considered for the runoff forecast models and flow data. We consider forecasts of runoff volume for a forecast horizon of 120min (or 60 time steps) (c.f. Section 3.8). Rain events 1 to 4 are used for model calibration, while events 5 to 8 are used for model validation only.

We consider two sets of models:

- (a) In the first set, future rainfall is assumed known during parameter estimation and forecast generation.
- (b) In the second set, future rainfall is assumed unknown during parameter estimation and forecast generation.

In set (b), future rainfall is generated by extrapolating a local linear model. The linear model is fitted to the rainfall observations over the last 120min (in a resolution of 10min) and then extrapolated over the next 120min (in a resolution of 2min).

### 4.3.2 Results and Discussion

Figure 4.3 shows the point and probabilistic forecast skill for the different events and catchments while tables 4.1 and 4.2 depict the reliability *Rel* and *ARIL* (c.f. 2.2.1) values for 90% prediction intervals, respectively. Point forecasts are derived using the median of the probabilistic multi-step forecasts (c.f. 3.8).

Considering the score values in Figure 4.3, we can see a difference in the forecast skills of the models with known and unknown rainfall input in all catchments except Str. This catchment is a special case due to time varying dry weather flows. These lead to parameter estimates with very large time constants  $K$  and, as a result, to runoff forecasts that are hardly affected by the rainfall input. If we allow for a time-varying dry weather flow parameter  $a_0$  in this model, the model using a known rainfall input performs clearly better than the one without (not shown).

In the other catchments we observe a strong loss of forecast skill in the EAm and WAm catchments when the future rainfall is considered unknown. In the Klo and Ler catchments (the biggest catchments considered), we observe similar forecast skill in the cases where rainfall input is known and unknown for some

events. Evaluating the *ARIL* values (Table 4.2), we see a clear increase of the forecast uncertainty in the EAm, COL and WAm catchments when the future rainfall is unknown, while the *ARIL* values only slightly increase in the Klo and Ler catchments. This behaviour fits with the point forecast performance of the models for the calibration events 1-4.

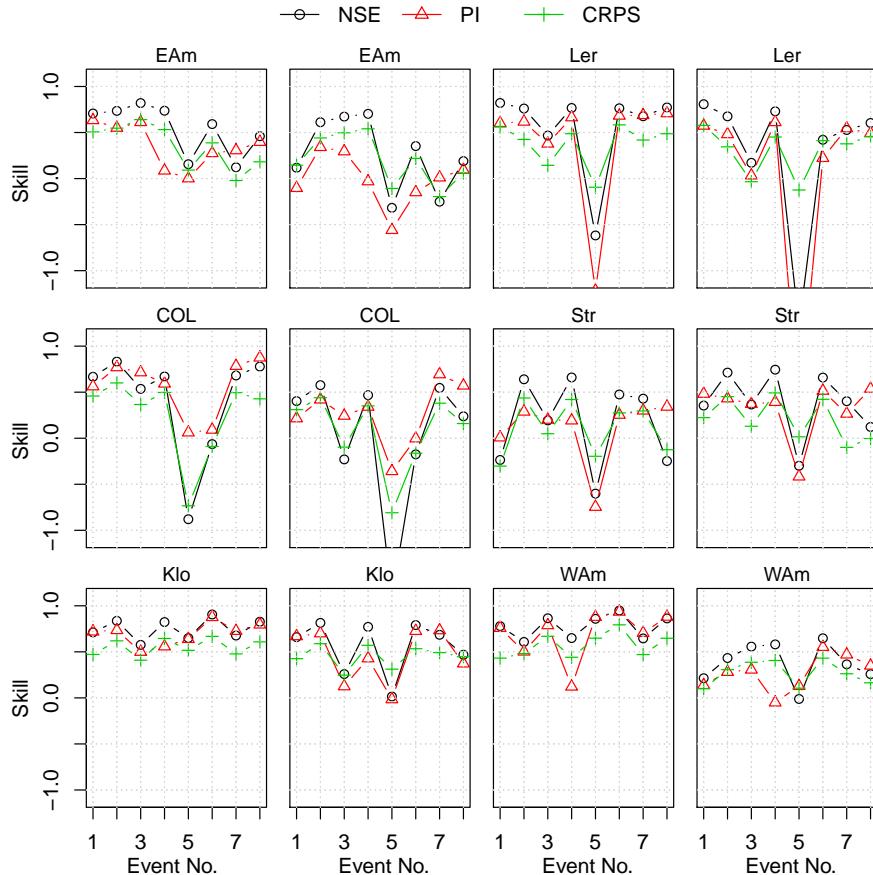
### 4.3.3 Conclusion

We conclude that the quality of rainfall forecasts has an influence on the runoff forecast quality in the considered urban catchments. This effect is particularly pronounced in the smaller EAm, COL and WAm catchments. A similar result is found for an 80ha catchment in [TR13], where runoff forecast quality diminishes with the quality of the rainfall forecasts when exceeding a horizon of 60min.

However, in the larger catchments (Klo and Ler) the result is less clear. As discussed in Section 4.1, a rainfall forecast may consequently not be necessary for the generation of runoff forecasts for certain combinations of forecast horizon and catchment size. This behaviour can have practical relevance in cases where short-term on-line rainfall forecasts in catchments are not available due to failures or missing installations.

**Table 4.1:** Reliability (*Rel*) of 90% prediction intervals on a 120min horizon for the catchments considered in paper E

Catch- ment	Rain input	Rain Event								Mean
		1	2	3	4	5	6	7	8	
EAm	Known	80%	72%	82%	73%	59%	76%	65%	49%	70%
EAm	Unknown	74%	77%	84%	81%	65%	72%	71%	65%	74%
COL	Known	78%	78%	68%	64%	26%	49%	76%	40%	60%
COL	Unknown	82%	82%	70%	59%	43%	50%	78%	51%	65%
Klo	Known	77%	72%	65%	81%	75%	79%	79%	65%	74%
Klo	Unknown	75%	74%	65%	74%	74%	68%	80%	62%	71%
Ler	Known	67%	46%	21%	73%	66%	77%	78%	55%	60%
Ler	Unknown	69%	46%	21%	81%	81%	86%	78%	68%	66%
Str	Known	62%	76%	53%	72%	63%	78%	67%	62%	66%
Str	Unknown	89%	64%	45%	71%	73%	70%	12%	52%	60%
WAm	Known	51%	76%	79%	61%	81%	98%	78%	72%	74%
WAm	Unknown	63%	67%	72%	74%	72%	81%	78%	63%	71%



**Figure 4.3:** Forecast skill (NSE, Persistence Index PI, normalized CRPS) in the catchments considered in paper E. Events 1-4 were used for calibration, events 5-8 for validation only. We consider perfectly known future rain inputs (columns 1 and 3) and unknown future rain inputs (columns 2 and 4)

**Table 4.2:** Average interval length (*ARIL*) of 90% prediction intervals on a 120min horizon for the catchments considered in paper E

Catch- ment	Rain input	Rain Event								Mean
		1	2	3	4	5	6	7	8	
EAm	Known	75%	87%	79%	70%	72%	73%	78%	57%	74%
	Unknown	109%	124%	114%	97%	105%	105%	115%	89%	107%
COL	Known	82%	103%	108%	128%	55%	54%	83%	136%	94%
	Unknown	107%	137%	150%	170%	75%	79%	113%	174%	125%
Klo	Known	84%	307%	110%	74%	87%	78%	73%	82%	112%
	Unknown	88%	310%	115%	77%	89%	80%	75%	87%	115%
Ler	Known	269%	427%	235%	237%	139%	167%	200%	540%	277%
	Unknown	279%	425%	283%	256%	146%	167%	222%	518%	287%
Str	Known	44%	51%	51%	47%	49%	47%	57%	49%	50%
	Unknown	46%	57%	59%	53%	50%	52%	74%	53%	55%
WAm	Known	47%	55%	54%	47%	52%	56%	64%	73%	56%
	Unknown	64%	72%	71%	64%	69%	72%	86%	93%	74%



## CHAPTER 5

# Probabilistic Runoff Forecasts in Practice

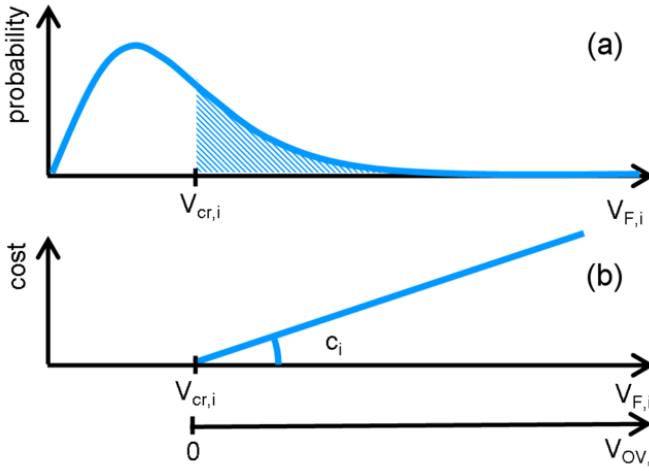
---

## 5.1 Real-Time Control under Uncertainty - the DORA algorithm

The final purpose of generating probabilistic rainfall runoff forecasts is to improve decision making in real-time control. In this work we have considered the *DORA* (dynamic overflow risk assessment) algorithm ([VG14]) for decision making under uncertainty as it was created by the project partners in the *SWI* (Storm and Wastewater Informatics) project and allowed for the direct application of the probabilistic forecast models in a real world setup.

The *DORA* algorithm tries to reduce the impact of combined sewer overflows (CSO). It adjusts outflows from control points as the free variables in an optimization setting. Its objective function is to minimize the CSO cost in the whole catchment over the considered forecast horizon  $T$ . CSO cost  $C_F$  for a single structure is determined as a function of the forecasted runoff volume  $V_F$  to the control point

$$C_F = \int_0^{\infty} C(V_F) \cdot p(V_F) dV_F, \quad (5.1)$$



**Figure 5.1:** Subfigure a: Probabilistic runoff volume forecast for a horizon  $T$  at structure  $i$  with critical volume  $V_{cr}$  (marks the runoff volume where the basin is filled and CSO starts), subfigure b: overflow cost  $c_i = C(V_{F,i})$  as a function of predicted volume  $V_{F,i}$  and resulting overflow volume  $V_{OV,i}$  (from [VG14])

where  $p(V_F)$  corresponds to the forecasted probability that a certain runoff volume will occur and  $C(V_F)$  to the overflow cost associated with this volume.

The cost function  $C(V_F)$  must be defined by the stakeholders. In this work it is a piecewise linear function of  $V_F$  which is equal to 0 if  $V_F$  does not induce overflow (Figure 5.1). The slope of the cost function determines how much weight an overflow structure receives in the control setup.

*DORA* is attractive as a control algorithm, because it allows to account for forecast uncertainty without exploding dimensionality in the optimization setting. It can account for any distributional shapes of the probabilistic forecasts by evaluating the integral in equation 5.1 either parametrically or non-parametrically. In paper F we have implemented an empirical evaluation of equation 5.1 using quantiles of the scenario based probabilistic forecasts described in Section 3.8 in a 2% resolution. Other criteria such as pollutant loads, surface flooding or energy consumption as a result of pumping or water treatment can be implemented in the same framework.

A major drawback of the current implementation of the *DORA* algorithm is that only a single forecast horizon  $T$  is considered in the decision making. Con-

## **5.2 Current Practical Implementation for Probabilistic Runoff Forecasting**

---

sequently we need to define this horizon at every control time step and the resulting outflows will correspond to an average setting over this horizon rather than the series of decisions which will be implemented in reality.

### **5.2 Current Practical Implementation for Probabilistic Runoff Forecasting**

The current practical implementation of *DORA* in Copenhagen uses runoff forecasts generated by conceptual models that are recalibrated in intervals of 10 minutes. This process is described in [LPB<sup>+</sup>14]. We have included this approach as a benchmark in papers E and F.

The runoff forecast uncertainty in this approach is described using a Gamma distribution. The mean of this distribution is found as

$$E[X] = k \cdot \theta, \quad (5.2)$$

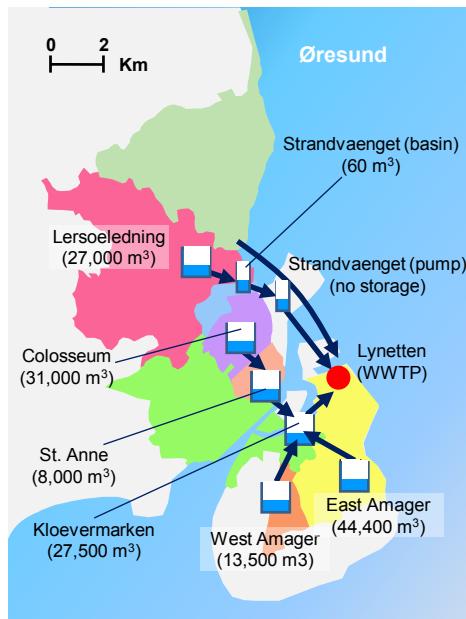
where  $k$  is the shape and  $\theta$  the scale parameter of the distribution. In the implementation, the mean of the distribution is assumed equal to the predicted runoff volume  $V_F$  for the considered horizon and we thus have  $\theta = V_F/k$ . The shape parameter is fixed to  $k = 3$ .

In paper C we demonstrate that this assumption of a fixed shape parameter leads to a strong overestimation of forecast uncertainty for a reasonably well performing forecast model.

### **5.3 Control Results Using Probabilistic Forecasts**

An experiment in paper C shows that a strong overestimation of forecast uncertainties for runoff volume will also lead to a strong overestimation of overflow risk in accordance with equation 5.1.

Neglecting forecast uncertainty in the computation of overflow risk, on the other hand, hardly affects the computed overflow risk as compared to an evaluation with a proper quantification of forecast uncertainty. The reason for this behaviour is the simplistic layout considered in the experiment. During the evaluation of predicted overflow risk over the time series, we mainly stay in the linear range of the cost functions while only very few data points are located at the break point from zero cost to non-zero cost (c.f. Figure 5.1).



**Figure 5.2:** Control points in the Lynetten catchment with available storage volume (from paper F)

A different behaviour can be expected for non-linear cost functions and when using probabilistic forecasts in an actual optimization setting for real time control. Due to the variety of different basin outflows considered during the optimization, the critical runoff volume  $V_{cr}$  that leads to CSO will move around on the x-axis at every time step and a proper quantification of runoff forecast uncertainty will affect decision making.

This was tested for the Lynetten catchment in Copenhagen in paper F. We implemented grey-box based runoff forecast models for the subcatchments EAm, Col, Klo, Ler, Str and WAm (Figure 5.2). A forecast model is in this case applied only for the specific subcatchment of a control point, while inflow from upstream control points is considered as a model input as it is determined by the control algorithm.

The results in the article confirm our hypothesis. Runoff forecasts based on grey-box models that account for uncertainty lead to a clear reduction of CSO volume compared to both the baseline and the grey-box based runoff forecasts where forecast uncertainty is neglected in the decision making. Similar to the results in paper C, an overestimation of runoff forecast uncertainty with the approach based on the Gamma distribution leads to a significantly worse performance in

decision making and increased CSO volumes.



## CHAPTER 6

# Conclusions

---

This thesis focuses on the development of methods for the generation of probabilistic runoff forecasts using stochastic grey-box models. We have made the step from considering this type of models for simulation studies towards an implementation that can actually be applied on-line. Forecast horizons of up to 120 minutes are considered in this work and the forecasts are applied in a real-time control context.

Using the questions posed in Section 1.3 as an outline, here we present the main conclusions from this work.

1. *What rainfall inputs should be used for short-term runoff forecasting and, in particular, do we benefit from using quantitative precipitation estimates (QPE) from weather radar?*

The results from paper B suggest that we can obtain better runoff forecasts when using radar rainfall measurements rather than rain gauge observations. It is not clear whether the improvement results from a better measurement of the rainfall process or an improved rainfall forecast provided by the radar. However, catchments with reduced areas of 1300 and 3000ha were considered. For such large catchment sizes, the influence of the rainfall forecasts on the runoff forecasts may be small on a forecast horizon of 2 hours. In addition, a correlation analysis in Section 4.2.1 indicates a better description of the mean areal rainfall by the radar measurements.

The operational reliability of radar rainfall measurements is problematic. Rain gauges, on the other hand, provide a mature method for measuring rainfall. Operators of urban drainage networks, weather services and authorities are therefore experienced in the operation of hardware and data analysis ([JRMM98]). Rain gauges can thus be considered a "robust" means of measuring rainfall. Our experience from practical implementation in papers E and F suggests that this is not necessarily the case for radar rainfall measurements in Denmark.

2. *Do quantitative precipitation forecasts (QPF) provide benefit for short term runoff forecasts?*

As discussed in Sections 4.1 and 4.3, the extent to which runoff forecasts can benefit from rainfall forecasts depends on the catchment characteristics and the considered forecast horizon. [TR13] demonstrate in an 80ha catchment that the benefits from radar rainfall forecasts on runoff forecasts diminish when exceeding horizons of 60 minutes. On the other hand, we show in Section 4.3 that in some cases we can obtain reasonable runoff forecasts without (or with very simple) rainfall forecasts.

We are not aware of systematic investigations that evaluate which forecast horizons and catchment characteristics require rainfall forecasts to generate good on-line runoff forecasts and consider this an interesting item of future work.

3. *Do short term runoff forecasts benefit from a combined rainfall input making use of both rain gauge and radar rainfall measurements?*

As discussed in paper B, numerous authors in the literature suggest that radar rainfall measurements should be adjusted to rain gauges before being used in (urban) hydrological applications. The validation of the adjustment methodology is, however, performed by comparing the adjusted radar data to rain gauge measurements or by using them as input for models that were calibrated using rain gauge measurements. The results in paper B indicate that if the adjustment is performed in an unsuitable way, the resulting rainfall measurement will yield worse results than using the non-adjusted radar rainfall data, even though the adjusted radar data have smaller bias when compared to the rain gauge observations.

Nevertheless, it seems somewhat natural that the consideration of additional rainfall information from different sources should result in an improved information about the rainfall process. We emphasize that the adjustment (or merging) of radar and rain gauge measurements should focus on the runoff forecasting purpose. This means that the adjustment methodologies should be calibrated using an objective function that is based on the runoff forecast skill that can be obtained with the adjusted rainfall input.

---

The investigation presented in paper A does not yet account for the above considerations. Nevertheless, it demonstrates that a merging of radar and rain gauge measurements yields an improved runoff forecasting skill. In this context, it is important to mention the aspect of operational reliability, as the merging of two different types of rainfall information will introduce redundancy in the forecast system and thus increase operational reliability.

*4. How can forecast models and parameters be identified in the context of noisy data and provide forecasts over a multitude of horizons?*

The probabilistic runoff forecasts generated by the stochastic grey-box models are required over a multitude of forecast horizons in the considered real time control applications. We suggest that the forecast models are estimated with this purpose in mind, minimizing the (weighted, probabilistic) forecast error of the multi-step predictions, rather than applying the (theoretically more sound) estimation method based on likelihood maximization which is implemented in the software framework *CTSM*.

We develop an estimation methodology based on minimisation of the continuous ranked probability score (*CRPS*) for multi-step ahead forecasts of runoff volume in paper C. The methodology currently has drawbacks as we assume normal distribution for the multi-step ahead forecasts and the resulting models underestimate forecast uncertainty. Nevertheless, we demonstrate in Section 3.6.5 that the new estimation method improves runoff forecast skill on a 2 hour (60 step) horizon.

An important conclusion from these considerations is that the model should ideally be estimated considering the intended application in mind. The state-of-the-art maximum likelihood approach that is usually applied for stochastic grey-box models will commonly identify models that perform well on short horizons. If we want to apply the models for multi-step predictions over longer horizons, different estimation approaches should be considered.

*5. How can dynamically changing forecast uncertainties be correctly captured in a probabilistic model structure?*

The results in paper G suggest that it is suitable to model runoff forecast uncertainty as a combination of a constant dry weather uncertainty and a dynamic uncertainty for rain periods which is scaled by a smoothed version of the rainfall input.

This type of model structure yields better results than the linear dependence of forecast uncertainty on the predicted model states which was suggested by [BTM<sup>+</sup>11]. In particular, we can avoid the mix up of dry and wet weather uncertainty which, for the linear state dependence, leads to very large forecast uncertainties in some of the catchments in paper G. We are furthermore able to render forecast uncertainty independently

from the forecasted runoff and can thus better capture forecast uncertainties occurring, for example, in the start of a rain event.

Nevertheless, the new approach cannot compensate for cases where the physical model structure is clearly insufficient for the data and we will still obtain unreliable forecasts in such cases.

*6. How can probabilistic forecasts be generated for decision making in real-time control?*

The thesis builds on the assumption that simple models that can be tuned to minimize the forecast error need to be used for forecasting in an on-line context and that we need to account for forecast uncertainty in decision making. Following this line of thought, we use stochastic grey-box models to generate probabilistic on-line forecasts.

We propose a scenario-based approach for the generation of multi-step probabilistic runoff forecasts. This approach is not bound to distributional assumptions. Moreover, it spares us the task of explicitly modelling the correlation between probabilistic runoff forecasts for different horizons, as it is inherent in the scenarios. Scenario simulations can be created from SDE's using the methods described in Section 3.7.

Scenario (or ensemble) forecasts also provide flexibility for the use of the probabilistic information in the real-time control algorithm. We have implemented a real-time control scheme which makes use of the quantiles of the probabilistic forecasts (see Section 5.1 and paper F).

*7. What effect does the consideration of forecast uncertainty have on the efficiency of real-time control schemes?*

We have evaluated the effect of runoff forecast uncertainty on the expected combined sewer overflow risk in paper C and on decision making in real-time control in paper F. From both applications we can conclude, that real-time control (in our case with respect to CSO) will benefit from a correct quantification of forecast uncertainty if non-linearities exist in the relation between the runoff forecast and the objective function of the control scheme.

In this thesis, we apply piecewise linear cost functions (see Section 5.1). These cost functions are zero if a forecasted runoff volume will not lead to overflow and increase linearly with the runoff volume otherwise. When the uncertainty of the forecasted runoff volume is strongly overestimated, the forecasted overflow risk (the objective function) will be affected by this non-linearity and forecasted too big. This was demonstrated in paper C and leads to suboptimal decision making in paper F.

When neglecting forecast uncertainty during decision making, the case is less obvious in paper C, because the estimation of overflow risk is only

for very few time steps affected by the non-linearity in the cost function according to Figure 5.1.

However, considering an actual real-time control setting based on an optimization of basin outflows, the position of the non-linearity in the cost function ( $V_{cr}$  in Figure 5.1) with respect to the forecasted runoff volume changes depending on the basin outflows defined by the controller. A correct (or reasonable) quantification of forecast uncertainty then becomes important. We can see this from the strong performance of the control scheme when applying stochastic grey-box models in paper F.



## CHAPTER 7

# Outlook

---

As an outcome of this thesis, we have established a framework of stochastic, conceptual models that can be applied for probabilistic runoff forecasting in an on-line context in urban hydrology. We have suggestions for how different sources of observations of rainfall should be used for runoff forecasting and how forecast uncertainties can be quantified reliably. Furthermore, we have obtained results that suggest that an appropriate quantification of forecast uncertainty has a positive impact on decision making in real-time control, while a large overestimation of forecast uncertainty will negatively impact real-time control schemes.

As a result of new developments in the field and of remaining deficiencies in the applied modelling approach, we suggest further research in the following directions:

### **Optimal combination of rainfall input, rainfall forecast and rainfall-runoff model**

For on-line runoff forecasting, rainfall input, rainfall forecast and runoff forecast model should be considered as a chain. First steps in this direction are taken by [TR13], for example.

The combination of different rainfall inputs (radar and rain gauge, for example) has the potential to improve runoff forecasts (see paper A and the introductory

discussion in paper B). The combination of these inputs, however, should be performed in such a way that runoff forecast quality is optimized.

This implies setting up a framework where adjustment procedures and runoff forecast models are tuned in one and the same calibration procedure. We are not aware of any such framework than can be applied operationally. The approach presented by [GHL02] and applied in paper A follows this line of thought but has limitations in the extent of datasets that can be considered.

### **When do rainfall forecasts improve on-line runoff forecasts**

The results in Section 4.3 suggest that we may in some cases be able to generate runoff forecasts without or with very limited information about the future rainfall. In the literature, discussions on the influence of the uncertainty of rainfall input on simulation quality are available (see the work by [BGM13], [LAP<sup>+</sup>14] and [SF86] for example). A structured discussion considering several catchments of different characteristics is missing in the literature on urban hydrology. Such a discussion should consider these parameters

- forecast horizon,
- catchment size,
- catchment characteristics (in particular reduced area).

For generality, simulation studies based on theoretical catchments can be considered, similar to the approach in [SF86]. We would expect to find that rainfall forecasts are not required for applications using runoff forecasts on very short horizons. Similarly, for large catchments we may find the forecast uncertainty related to weather models acceptable in order to, for example, coarsely determine process settings on the wastewater treatment plant.

### **Optimal estimation of probabilistic rainfall-runoff models**

If the stochastic grey-box models should be applied for generating runoff forecasts on longer forecast horizons, they should be calibrated in a way that reflects this application. We propose a corresponding estimation approach in paper C. This approach can be improved in the following ways:

- Parameter calibration is currently performed by generating multi-step predictions using the extended Kalman filtering approach implemented in *CTSM* and by minimizing the *CRPS* score averaged over the different forecast horizons. In several applications (papers B and E) we have observed that this approach leads to reasonable model performance in terms

---

of point forecast quality. Runoff forecast uncertainties, however, are somewhat underestimated. We suggest to consider whether this problem can be solved by considering different score functions (such as likelihood inspired scores or a combination of interval scores focusing on different quantiles).

- During parameter calibration, forecasts are generated through the extended Kalman filter implemented in *CTSM*. We should investigate if it is possible to (in a computationally feasible way) base the parameter estimation procedure on direct simulations of the SDE's as described in Section 3.8. Such an approach would not rely on the assumption of normality of the forecasts and give a better representation of the actual stochastic processes. Particle filtering approaches as described by [MDS12] may also be relevant in this context.

In addition, there is also an issue of what time resolution should be selected for the runoff forecast models. When considering longer time steps (and thus shorter forecast horizons in terms of time steps), the forecast error from the very simple model structures may in some cases be smaller. This issue was not investigated in this work.

### **Structural complexity**

The stochastic grey-box approach using *CTSM* has limitations in how complex the considered models can be. Discussions in the literature (for example [SF86, Bre12]) and the results in paper B suggest that somewhat more complex structures give benefit to short-term forecasts. In real-time control we currently apply a separate stochastic runoff forecast model for each control point. The forecast is then used as input for the decision making algorithm. It is very likely that the control schemes could benefit if we used a single stochastic model for the whole catchment which is then also used during decision making. Finally, new applications such as the development of on-line models for the capacity of the wastewater treatment plant also call for stochastic modelling techniques that allow for greater model complexity to represent the different processes.

The number of model states that can be considered in the stochastic grey-box approach is limited because an extended Kalman filter is applied. In this context we can consider different updating methods described, for example, by [Bor14] and [HVLKS11]. In particular, the combined Bayesian uncertainty description and data assimilation approach described by [VDG<sup>+</sup>05] is very related to the grey-box approach and should be tested with respect to the dynamic modelling of uncertainties and the generation of reliable forecasts on multi-step horizons.

In this context, Maximum a posteriori parameter estimation may be helpful for identifying parameters for more complex model structures.

### **Future applications**

From a practical viewpoint, a foreseen item of future research is the development of libraries of stochastic rainfall-runoff models in combination with automated routines for residual analysis. Multiple models of different complexity can then be tested for a given catchment and the best performing model structure can be selected for on-line applications.

In a qualitative sense, we should investigate model structures that can be used for on-line forecasting of water quality and treatment capacity of the WWTP. Such structures will allow for an integrated operation of the urban drainage system with respect to pollutant loads and thus the actual stress on natural water bodies. First steps were taken by [BNMP99], but generally the model structures in this area are extremely complex and not suitable for on-line purposes.

Furthermore, we expect the integrated operation of urban energy and water systems (Smart Cities) to become a major driver of the on-line operation of urban drainage systems. Such a combined operation offers strong economic incentives for the operators of drainage systems and environmental benefits for society. This calls for the development of new control strategies that account for incentives from the energy market and uncertainties in both forecasts of electricity prices (for example [JPN<sup>+</sup>13]) and loads from the drainage system, as well as model structures that can forecast the energy production and demand in the urban water cycle.

# Bibliography

---

- [ARM13] Philip Delff Andersen, Carsten Rode, and H Madsen. An arctic low-energy house as experimental setup for studies of heat dynamics of buildings. *Front. Archit. Res.*, 2(4):488–499, December 2013.
- [ATCV13] Janelcy Alferes, Sovanna Tik, John Copp, and P A Vanrolleghem. Advanced monitoring of water systems using in situ measurement stations: data validation and fault detection. *Water Sci. Technol.*, 68(5):1022–1030, January 2013.
- [BCG<sup>+</sup>13] Neil D Bennett, Barry F W Croke, Giorgio Guariso, Joseph H A Guillaume, Serena H Hamilton, Anthony J Jakeman, Stefano Marsili-Libelli, Lachlan T H Newham, J.P. Norton, Charles Perrin, Suzanne A Pierce, Barbara Robson, Ralf Seppelt, Alexey A Voinov, Brian D Fath, and Vazken Andreassian. Characterising performance of environmental models. *Environ. Model. Softw.*, 40:1–20, 2013.
- [BDLY01] P E Brown, P J Diggle, M E Lord, and P C Young. Space–time calibration of radar rainfall data. *J. R. Stat. Soc. Ser. C - Appl. Stat.*, 50(2):221–241, 2001.
- [Bev93] K J Beven. Prophecy, reality and uncertainty in distributed hydrological modelling. *Adv. Water Resour.*, 16(1):41–51, 1993.
- [BGM13] M Borup, M Grum, and P S Mikkelsen. Comparing the impact of time displaced and biased precipitation estimates for online updated urban runoff models. *Water Sci. Technol.*, 68(1):109–116, 2013.

- [BK13] A Berne and W F Krajewski. Radar for hydrology: Unfulfilled promise or unrecognized potential? *Adv. Water Resour.*, 51:357–366, January 2013.
- [BLP08] Nicola Bruti-Liberati and E Platen. Strong predictor-corrector Euler methods for stochastic differential equations. *Stochastics Dyn.*, 8(3):561–581, September 2008.
- [BM11] P Bacher and H Madsen. Identifying suitable models for the heat dynamics of buildings. *Energy Build.*, 43(7):1511–1522, July 2011.
- [BMMM12] A Breinholt, J K Møller, H Madsen, and P S Mikkelsen. A formal statistical approach to representing uncertainty in rainfall-runoff modelling with focus on residual analysis and probabilistic output evaluation – Distinguishing simulation and prediction. *J. Hydrol.*, 472–473:36–52, 2012.
- [BMPN00] H Bechmann, H Madsen, N K Poulsen, and M K Nielsen. Grey box modeling of first flush and incoming wastewater at a wastewater treatment plant. *Environmetrics*, 11(1):1–12, 2000.
- [BNMP99] H Bechmann, M K Nielsen, H Madsen, and N K Poulsen. Grey-box modelling of pollutant loads from a sewer system. *Urban Water*, 1(1):71–78, 1999.
- [Bol86] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *J. Econom.*, 31(3):307–327, April 1986.
- [Bor14] M Borup. *Real time updating in distributed urban rainfall runoff modelling*. Phd, Technical University of Denmark (DTU), 2014.
- [Bre12] A Breinholt. *Uncertainty in prediction and simulation of flow in sewer systems*. Phd, Technical University of Denmark, 2012.
- [BRH13] C. Berndt, E. Rabiei, and U. Haberlandt. Geostatistical merging of rain gauge and radar data for high temporal resolutions and various station density scenarios. *J. Hydrol.*, in press, October 2013.
- [Brö12] J Bröcker. Evaluating raw ensembles with the continuous ranked probability score. *Q. J. R. Meteorol. Soc.*, 138(667):1611–1617, 2012.
- [BS05] D Butler and M Schütze. Integrating simulation models with a view to optimal control of urban wastewater systems. *Vulnerability Water Qual. Intensively Dev. Urban Watersheds*, 20(4):415–426, 2005.

- [BTM<sup>+</sup>11] A Breinholt, F O Thordarson, J K Møller, M Grum, P S Mikkelsen, and H Madsen. Grey-box modelling of flow in sewer systems with state-dependent diffusion. *Environmetrics*, 22(8):946–961, 2011.
- [CMM88] V T Chow, D R Maidment, and L W Mays. *Applied hydrology*. McGraw-Hill, 1st edition, 1988.
- [CNH96] J Carstensen, M K Nielsen, and P Harremoës. Predictive Control of Sewer Systems by Means of Grey-box Models. *Water Sci Technol*, 34(3-4):189–194, 1996.
- [DDR13] D J Dürrenmatt, D Del Giudice, and J Rieckermann. Dynamic time warping improves sewer flow monitoring. *Water Res.*, 47:3803–3816, 2013.
- [DHI03] DHI. MOUSE User Guide and Reference Manual. *Agern Alle*, 11, 2003.
- [DHS<sup>+</sup>13] D Del Giudice, M Honti, A Scheidegger, C Albert, P Reichert, and J Rieckermann. Improving uncertainty estimation in urban hydrological modeling by statistically describing bias. *Hydrol. Earth Syst. Sci.*, 17:4209–4225, 2013.
- [DMK<sup>+</sup>12] C.B.S. Dotto, Giorgio Mannina, M. Kleidorfer, L Vezzaro, Malte Henrichs, D.T. McCarthy, Gabriele Freni, W Rauch, and A. Deletic. Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling. *Water Res.*, 46(8):2545–58, May 2012.
- [FB05] L Fuchs and T Beenken. Development and implementation of a real-time control strategy for the sewer system of the city of Vienna. *Water Sci. Technol.*, 52(5):187–194, 2005.
- [FPM11] O Fradet, M Pleau, and C Marcoux. Reducing CSOs and giving the river back to the public: innovative combined sewer overflow control and riverbanks restoration of the St. Charles River in Quebec City. *Water Sci. Technol.*, 63(2):331–338, 2011.
- [Gay90] D M Gay. Usage Summary for Selected Optimization Routines. Technical report, AT&T Bell Laboratories, Murray Hill, NJ, United States, 1990.
- [GD09] E Goudenhoofdt and L Delobbe. Evaluation of radar-gauge merging methods for quantitative precipitation estimates. *Hydrol. Earth Syst. Sci.*, 13(2):195–203, 2009.
- [GHL02] M Grum, P Harremoës, and J J Linde. Assimilating a multitude of rainfall and runoff data using a stochastic state space modeling approach. In *9th Int. Conf. Urban Drain.*, volume Portland,, 2002.

- [GR07] T Gneiting and A E Raftery. Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.*, 102(477):359–378, 2007.
- [GTC<sup>+</sup>11] M Grum, D Thornberg, M L Christensen, S A Shididi, and C Thirsing. Full-Scale Real Time Control Demonstration Project in Copenhagen’s Largest Urban Drainage Catchments. In *Proc. 12th Int. Conf. Urban Drain.*, Porto Alegre, Brazil, 2011.
- [HFZ13] S. Hemri, F. Fundel, and M. Zappa. Simultaneous calibration of ensemble river flow predictions over an entire range of lead times. *Water Resour. Res.*, 49:6744–6755, October 2013.
- [HMS02] D J Higham, X Mao, and A M Stuart. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. *SIAM J. Numer. Anal.*, 40(3):1041–1063, 2002.
- [HVLKS11] CJ Hutton, CS Vamvakeridou-Lyroudia, Z Kapelan, and DA Savic. Real-time modelling and Data Assimilation techniques for improving the accuracy of model predictions. *Prep. 2011.010 Deliv. D3.6.2*, page 83, 2011.
- [Iac08] S M Iacus. *Simulation and inference for stochastic differential equations : with R examples*. Springer Series in Statistics. Springer, Milan, Italy, 2008.
- [JKB<sup>+</sup>13] R Juhl, N R Kristensen, P Bacher, J K Møller, and H Madsen. CTSM-R User Guide, 2013.
- [JPN<sup>+</sup>13] Tryggvi Jónsson, Pierre Pinson, Henrik Aalborg Nielsen, Henrik Madsen, and Torben Skov Nielsen. Wind Power Predictions. *Ieee Trans. Sustain. Energy*, 4(1):210–218, 2013.
- [JRMM98] H K Jørgensen, S Rosenørn, H Madsen, and P S Mikkelsen. Quality control of rain data used for urban runoff systems. *Water Sci. Technol.*, 37(11):113–120, 1998.
- [JXZS10] X Jin, C Y Xu, Q Zhang, and V P Singh. Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model. *J. Hydrol.*, 383(3-4):147–155, 2010.
- [KKFT06] George Kuczera, Dmitri Kavetski, S Franks, and Mark Thyre. Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters. *J. Hydrol.*, 331(1):161–177, 2006.

- [KM03] N R Kristensen and H Madsen. Continuous Time Stochastic Modelling CTSM 2.3 - Mathematics Guide. Technical report, Technical University of Denmark (DTU), Lyngby, Denmark, 2003.
- [KMJ04] N R Kristensen, H Madsen, and S B Jørgensen. Parameter estimation in stochastic grey-box models. *Automatica*, 40(2):225–237, 2004.
- [KP99] P E Kloeden and E Platen. *Numerical Solution of Stochastic Differential Equations*, volume 3rd. Springer, Berlin-Heidelberg-New York, 1999.
- [KVPT12] S Kraemer, H R Verworn, A Pfister, and A Treis. Streckenintegrierte Regenmessung mit dem Mikrowellendämpfungs differenzverfahren und Quantifizierung von Radarsignal dämpfung. *Hydrol. und Wasserbewirtschaftung*, 56:59–77, 2012.
- [LAP<sup>+</sup>14] F. Lobligeois, V. Andréassian, C. Perrin, P. Tabary, and C. Loumagne. When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events. *Hydrol. Earth Syst. Sci.*, 18(2):575–594, February 2014.
- [Lju10] G M Ljung. Perspectives on system identification. *Annu. Rev. Control*, 34(1):1–12, April 2010.
- [LMM14] R Löwe, P S Mikkelsen, and H Madsen. Stochastic rainfall-runoff forecasting: parameter estimation, multi-step prediction, and evaluation of overflow risk. *Stoch. Environ. Res. Risk Assess.*, 28:505–516, 2014.
- [LPB<sup>+</sup>14] N S V Lund, J W Pedersen, M Borup, P S Mikkelsen, and M Grum. Auto-Calibration For Updating Of Linear Reservoir Models Used In Flow Forecasting Of Urban Runoff. *Hydrol. Process.*, (in preparation), 2014.
- [Mad08] H Madsen. *Time series analysis*. Chapman and Hall/CRC, Boca Raton, FL, 2008.
- [MB10] Christos K. Makropoulos and David Butler. Distributed Water Infrastructure for Sustainable Communities. *Water Resour. Manag.*, 24(11):2795–2816, January 2010.
- [MDS12] Hamid Moradkhani, C M DeChant, and S Sorooshian. Evolution of ensemble data assimilation for uncertainty quantification using the particle filter markov chain monte carlo method. *Water Resour. Res.*, 48(12), 2012.

- 
- [ML13] F Männig and M Lindenberg. Betriebserfahrungen mit der Abflusssteuerung des Dresdner Mischwassernetzes. *KA Korrespondenz Abwasser, Abfall*, 2013(12):1036–1043, 2013.
- [MMI05] M Maeda, H Mizushima, and K Ito. Development of the real-time control (RTC) system for Tokyo sewage system. *Water Sci. Technol.*, 51(2):213–220, 2005.
- [Møl10] J K Møller. *Stochastic State Space Modelling of Nonlinear systems*. Imm-phd-2010-246, Technical University of Denmark (DTU), 2010.
- [Mor05] C. M. Mora. Weak exponential schemes for stochastic differential equations with additive noise. *IMA J. Numer. Anal.*, 25(3):486–506, February 2005.
- [MS05] H Madsen and C Skotner. Adaptive state updating in real-time river flow forecasting—a combined filtering and error forecasting procedure. *J. Hydrol.*, 308(1–4):302–312, December 2005.
- [MW77] A H Murphy and R L Winkler. Reliability of Subjective Probability Forecasts of Precipitation and Temperature. *J. R. Stat. Soc. C (Applied Stat.)*, 26(1):41–47, 1977.
- [NJR13] J E Nielsen, N.E. Jensen, and M R Rasmussen. Calibrating LAWR weather radar using laser disdrometers. *Atmos. Res.*, 122:165–173, 2013.
- [NM06] H A Nielsen and H Madsen. Modelling the heat consumption in district heating systems using a grey-box approach. *Energy Build.*, 38(1):63–71, 2006.
- [NS70] J E Nash and J V Sutcliffe. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.*, 10(3):282–290, 1970.
- [NTR12] J E Nielsen, S Thorndahl, and M R Rasmussen. Evaluering af ny processeringsmetode til LAWR. Technical report, Department of Civil Engineering, Aalborg University, Aalborg, Denmark, 2012.
- [Øks98] B Øksendal. *Stochastic differential equations : An introduction with applications*. Springer, 6th edition, 1998.
- [Paw01] Yudi. Pawitan. *In all likelihood : Statistical modelling and inference using likelihood*. Clarendon Press, Cork, Ireland, 2001.
- [PCL<sup>+</sup>05] M Pleau, H Colas, P Lavalle, G Pelletier, and R Bonin. Global optimal real-time control of the Quebec urban drainage system. *Environ. Model. Softw.*, 20(4):401–413, 2005.

- [PCR<sup>+</sup>09] V Puig, G Cembrano, J Romera, J Quevedo, B Aznar, G Ramón, and J Cabot. Predictive optimal control of sewer networks using CORAL tool: application to Riera Blanca catchment in Barcelona. *Water Sci. Technol.*, 60(4):869–878, 2009.
- [Pin07] P Pinson. Non-parametric probabilistic forecasts of wind power: required properties and evaluation. *Wind Energy*, 10(6):497, 2007.
- [Pin13] P Pinson. Wind Energy : Forecasting Challenges for its Operational Management. *Stat. Sci.*, 2013.
- [PMN<sup>+</sup>09] P Pinson, H Madsen, H A Nielsen, G Papaefthymiou, and B Kloekl. From Probabilistic Forecasts to Statistical Scenarios of Short-term Wind Power Production. *Wind Energy*, 12(1):51–62, 2009.
- [RCT13] R-Core-Team. R: A Language and Environment for Statistical Computing, 2013.
- [Ref96] J C Refsgaard. Terminology, Modelling Protocol And Classification of Hydrological Model Codes. In Michael B. Abbott and Jens Christian Refsgaard, editors, *Distrib. Hydrol. Model.*, volume 22 of *Water Science and Technology Library*, pages 17–39. Springer Netherlands, Dordrecht, Netherlands, 1996.
- [RKK<sup>+</sup>10] B Renard, Dmitri Kavetski, George Kuczera, Mark Thyre, and S Franks. Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resour. Res.*, 46(5):W05521, 2010.
- [RS12] P Reichert and N Schuwirth. Linking statistical bias description to multiobjective model calibration. *Water Resour. Res.*, 48(9):W09543, 2012.
- [RTSS13] Helena M. Ramos, Charlotte Teyssier, Irene Samora, and Anton J. Schleiss. Energy recovery in SUDS towards smart water grids: A case study. *Energy Policy*, in press, November 2013.
- [SBB02] M Schütze, D Butler, and M B Beck. *Modelling, simulation and control of urban wastewater systems*. Springer, London, 2002.
- [SBK13] Siao Sun and Jean-Luc Bertrand-Krajewski. Separately accounting for uncertainties in rainfall and runoff: Calibration of event-based conceptual hydrological models in small urban catchments using Bayesian method. *Water Resour. Res.*, 49(9):5381–5394, 2013.
- [SCC<sup>+</sup>04] M Schütze, A Campisano, H Colas, W Schilling, and P A Vanrolleghem. Real time control of urban wastewater systems—where do we stand today? *J. Hydrol.*, 299(3):335–348, 2004.

- [Sch91] W Schilling. Rainfall data for urban hydrology: what do we need? *Atmos. Res.*, 27(1):5–21, 1991.
- [SF86] W Schilling and L Fuchs. Errors in stormwater modeling-a quantitative assessment. *J. Hydraul. Eng.*, 112(2):111–123, 1986.
- [SGT<sup>+</sup>13] A K Sharma, T Guildal, H A R Thomsen, P S Mikkelsen, and B N Jacobsen. Aeration tank settling and real time control as a tool to improve the hydraulic capacity and treatment efficiency during wet weather: results from 7 years' full-scale operational data. *Water Sci. Technol.*, 67(10):2169–2176, 2013.
- [SKZ<sup>+</sup>12] Michael Smith, Victor Koren, Ziya Zhang, Yu Zhang, Seann Reed, Zhengtao Cui, Fekadu Moreda, Brian a. Cosgrove, Naoki Mizukami, and Eric Anderson. Results of the DMIP 2 Oklahoma experiments. *J. Hydrol.*, 418–419:17–48, February 2012.
- [SLBF13] K Seggelke, R Löwe, T Beenken, and L Fuchs. Implementation of an integrated real-time control system of sewer system and waste water treatment plant in the city of Wilhelmshaven. *Urban Water J.*, 10(5):330–341, September 2013.
- [TBJSJ08] S Thorndahl, K J Beven, J B Jensen, and Kjeld Schaarup-Jensen. Event based uncertainty assessment in urban drainage modelling, applying the GLUE methodology. *J. Hydrol.*, 357(3):421–437, 2008.
- [TBM<sup>+</sup>12] F O Thordarson, A Breinholt, J K Møller, P S Mikkelsen, M Grum, and H Madsen. Uncertainty assessment of flow predictions in sewer systems using grey box models and skill score criterion. *Stoch. Environ. Res. Risk Assess.*, 26(8):1151–1162, 2012.
- [Tho11] F O Thordarson. *Grey Box Modelling of Hydrological Systems*. Phd, Technical University of Denmark (DTU), Kgs. Lyngby, Denmark, 2011.
- [Tod01] E Todini. A Bayesian technique for conditioning radar precipitation estimates to rain-gauge measurements. *Hydrol. Earth Syst. Sci.*, 5(2):187–199, 2001.
- [TPM13] Julija Tastu, P Pinson, and H Madsen. Space-time scenarios of wind power generation produced using a Gaussian copula with parametrized precision matrix. Technical report, Technical University of Denmark (DTU), Kgs. Lyngby, Denmark, 2013.
- [TR13] S Thorndahl and M R Rasmussen. Short-term forecasting of urban storm water runoff in real-time using extrapolated radar rainfall data. *J. Hydroinformatics*, 15(3):897–912, 2013.

- [TS07] B A Tolson and C A Shoemaker. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resour. Res.*, 43(1):W01413, 2007.
- [VDG<sup>+</sup>05] J A Vrugt, C G H Diks, Hoshin V Gupta, W Bouten, and J M Verstraten. Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. *Water Resour. Res.*, 41(1):W01017, 2005.
- [Ves98] M Vestergaard. *Nonlinear filtering in stochastic volatility models*. Master, Technical University of Denmark (DTU), Lyngby, Denmark, 1998.
- [VG14] L Vezzaro and M Grum. A generalized Dynamic Overflow Risk Assessment (DORA) for urban drainage Real Time Control. *J. Hydrol.*, submitted, 2014.
- [VMA<sup>+</sup>03] Henk Vanhooren, Jurgen Meirlaen, Youri Amerlinck, Filip Claeys, Hans Vangheluwe, and Peter A Vanrolleghem. WEST : modelling biological wastewater treatment. *J. Hydroinformatics*, 5(1):27–50, 2003.
- [VMDM13] L Vezzaro, P S Mikkelsen, A. Deletic, and D.T. McCarthy. Urban drainage models – simplifying uncertainty analysis for practitioners. *Water Sci. Technol.*, 68(10):2136, November 2013.
- [VR02] W N Venables and B D Ripley. *Modern Applied Statistics with S*, volume Fourth. Springer, New York, 2002.
- [VTDS13] J A Vrugt, C J F Ter Braak, C G H Diks, and G Schoups. Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: Theory, concepts and applications. *Adv. Water Resour.*, 51:457–478, January 2013.
- [Whi94] D Whitley. A genetic algorithm tutorial. *Stat. Comput.*, 4(2):65–85, 1994.
- [Wil11] D S Wilks. *Statistical methods in the atmospheric sciences*, volume 100. Academic press, Oxford, United Kingdom, 2011.
- [WOSo<sup>+</sup>13] L Wang, Susana Ochoa, Nuno Simões, C Onof, and C Maksimović. Radar-raingauge data combination techniques: a revision and analysis of their suitability for urban hydrology. *Water Sci. Technol.*, 68(4):737–747, 2013.



## Part II

## Papers



PAPER A

# State-space adjustment of radar rainfall and skill score evaluation of stochastic volume forecasts in urban drainage systems

---

**Authors:**

Roland Löwe, Peter Steen Mikkelsen, Michael Rasmussen, Henrik Madsen

**Published in:**

*Water Science and Technology*, 68(3): 584–590, 2013.



---

# State-space adjustment of radar rainfall and skill score evaluation of stochastic volume forecasts in urban drainage systems

Roland Löwe<sup>1</sup>, Peter Steen Mikkelsen<sup>2</sup>, Michael Rasmussen<sup>3</sup>, Henrik Madsen<sup>1</sup>

## Abstract

Merging of radar rainfall data with rain gauge measurements is a common approach to overcome problems in deriving rain intensities from radar measurements. We extend an existing approach for adjustment of C-band radar data using state-space models and use the resulting rainfall intensities as input for forecasting outflow from two catchments in the Copenhagen area. Stochastic grey-box models are applied to create the runoff forecasts, providing us with not only a point forecast but also a quantification of the forecast uncertainty. Evaluating the results, we can show that using the adjusted radar data improves runoff forecasts compared with using the original radar data and that rain gauge measurements as forecast input are also outperformed. Combining the data merging approach with short-term rainfall forecasting algorithms may result in further improved runoff forecasts that can be used in real time control.

---

<sup>1</sup>Department of Applied Mathematics and Computer Science, Technical University of Denmark, Matematiktorvet, B322, DK-2800 Kgs. Lyngby, Denmark

<sup>2</sup>Department of Environmental Engineering, Technical University of Denmark, Miljøvej, B115, DK-2800 Kgs. Lyngby, Denmark

<sup>3</sup>Department of Civil Engineering, Aalborg University, Sohngårdsholmsvej 57, DK-9000 Aalborg, Denmark

# **State-space adjustment of radar rainfall and skill score evaluation of stochastic volume forecasts in urban drainage systems**

Roland Löwe<sup>1</sup>, Peter Steen Mikkelsen<sup>2</sup>, Michael R. Rasmussen<sup>3</sup>, Henrik Madsen<sup>4</sup>

<sup>1</sup> Department of Applied Mathematics and Computer Science, Technical University of Denmark (DTU), Denmark, rolo@dtu.dk

<sup>2</sup> Department of Environmental Engineering, Technical University of Denmark (DTU), Denmark, psmi@env.dtu.dk

<sup>3</sup> Department of Civil Engineering, Aalborg University, Denmark, mr@civil.aau.dk

<sup>4</sup> Department of Applied Mathematics and Computer Science, Technical University of Denmark (DTU), Denmark, hmad@dtu.dk

## **ABSTRACT**

Merging of radar rainfall data with rain gauge measurements is a common approach to overcome problems in deriving rain intensities from radar measurements. We extend an existing approach for adjustment of C-band radar data using state-space models and use the resulting rainfall intensities as input for forecasting outflow from two catchments in the Copenhagen area. Stochastic greybox models are applied to create the runoff forecasts, providing us with not only a point forecast but also a quantification of the forecast uncertainty. Evaluating the results, we can show that using the adjusted radar data improves runoff forecasts compared to using the original radar data and that rain gauge measurements as forecast input are also outperformed. Combining the data merging approach with short term rainfall forecasting algorithms may result in further improved runoff forecasts that can be used in real time control.

## **KEYWORDS**

Flow forecast, greybox model, radar rainfall, state space model

## **1 INTRODUCTION**

Radar observations are increasingly used for measuring rainfall in urban areas. The good spatial coverage, however, comes along with problems in determining the rainfall intensity due to problems such as beam attenuation and the drop size dependency of the relation between reflectivity and rain intensity. Merging the radar measurements with gauge observations is a practitioners approach to this problem.

Classically, radar rainfall measurements are adjusted with mean field bias to reflect ground measurements as well as possible. Thorndahl et al. (2010) follow this approach in a two-step adjustment that is used operationally within the real time control framework in the Copenhagen area (Grum et al. (2011)). Uncertainties of the ground measurements are thereby neglected. Further, assumptions need to be made on how to apply rain gauge point measurements to the radar rainfall plane. Integrating gauge and radar rainfall measurements using state space models has been proposed by several authors in the past. Chumchean et al. (2006) and Costa and Alpuim (2009) use these techniques for temporal updating of the mean field bias. Brown et al. (2001) integrate spatial interaction into their model via a vector autoregressive process. Similarly Grum et al. (2002) construct a simple state space model that implicitly enables spatial interaction between the pixels and allows for the integration of a multitude of measurement types that can be related to the rainfall process.

We adopt this last approach due to its ability to incorporate spatial interaction and various measurement types and extend the uncertainty structure. The reconstruction of the rainfall process is then used to create stochastic runoff forecast from a simple grey-box model. We evaluate the quality of different forecasts using skill scores.

## 2 METHODOLOGY

### 2.1 Data and Catchments

We consider two catchments in the Copenhagen area. The Ballerup catchment has a total area of approx. 1300 ha. It is mainly laid out as a separate system but has a small combined part. The runoff in this area is strongly influenced by rainfall dependent infiltration.

The Damhusåen catchment is located close to Ballerup but drains to a different treatment plant. We consider the northern part of the catchment with a total area of approx. 3000 ha. The catchment is laid out as a combined sewer system and a multitude of CSO's are located in the area. Flow measurements are available from both catchments in 5 min resolution.

A C-band radar is operated by the Danish Meteorological Institute (DMI) in Stevns approx. 45 km south of the considered catchments. The spatial resolution of the radar pixels is 2x2 km. The provided radar data are rain intensities derived using the Marshall Palmer relationship, where the coefficients have been adjusted such that the average rainfall depth observed by the radar during the considered period matches selected gauge measurements (Thorndahl et al. (2010)). We denote these data ‘unadjusted radar data’. We consider an area of 9x11 pixels that covers the whole Copenhagen area (Figure 1).

Within the catchments online rain gauge measurements are available from the Danish SVK network (Jørgensen et al. (1998)). The gauges marked with grey circles in Figure 1 are used to adjust the radar measurements. Only few of the available gauges are used for this purpose as one objective for using radar rainfall data is to derive rain intensities from as few ground measurements as possible. To make results comparable, we use the same gauges that are used for radar adjustment in a real time control project in the Copenhagen area (Grum et al. (2011)). A reference simulation is performed where flow forecasts are generated using rain gauge measurements as an input. The gauges for these simulations were selected with respect to their location to the catchment as marked in Figure 1.

We have selected a 3-month period of measurements from 25/06/2010 until 29/09/2010 for this study. The period contains several summer storms that should be relevant for control applications in urban drainage systems. The radar data contain extensive gaps during smaller rain events from 27/08/2010 23:20 to 30/08/2010 12:50, 07/09/2010 08:20 to 08/09/2010 11:50 and 09/09/2010 09:20 to 14/09/2010 07:10.

A modelling time step of 10 min is adopted corresponding to the resolution of the provided radar measurements. The flow and rain gauge data are averaged to match this time step. Considering the size of the two catchments and the resulting concentration time of more than one hour, this resolution can be considered sufficient to correctly capture the runoff process (Schilling (1991)).

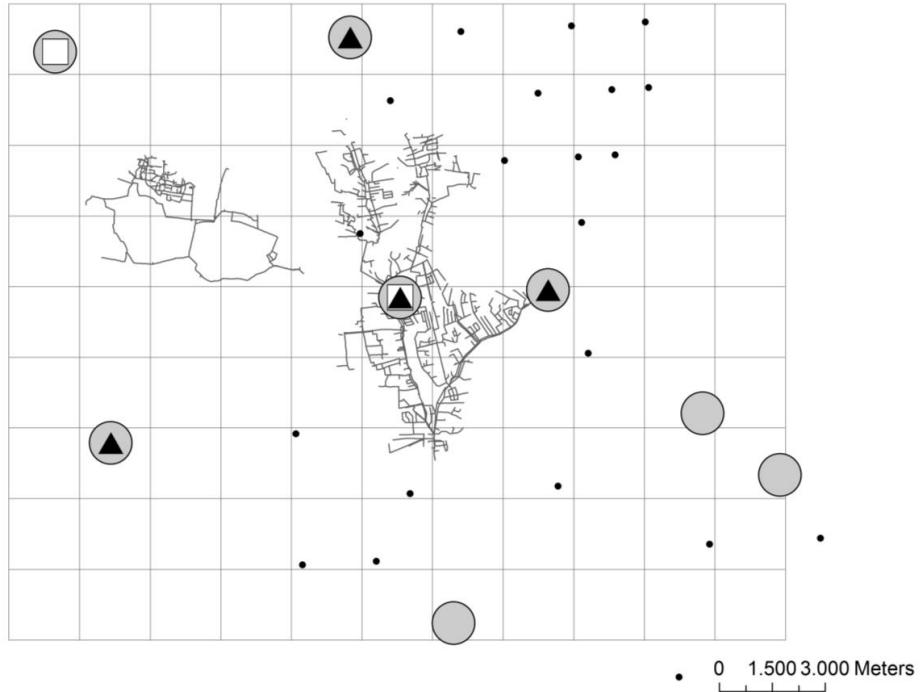


Figure 1. Considered area with C-band radar pixels, Ballerup (left) and Damhusåen (right) catchments. Rain gauges in the area (small dots), gauges used for radar adjustment (grey circles), gauges used as input for reference simulations in the Ballerup (white rectangles) and Damhusåen (black triangles) catchments.

## 2.2 Radar Adjustment

We investigate a state-space approach that has first been described in Grum et al. (2002) and Grum et al. (2005). We only give a brief summary here and describe parts that differ from the previous publication. The general setup is as follows:

- Create a linear model that predicts rainfall at the next time step for every pixel. Each pixel value is considered a state.
- Relate model state predictions and observed rainfall values from different sources in a set of observation equations
- Determine the adjusted rainfall values by “averaging” between model prediction and observation using a Kalman filtering approach. The filtering is performed based on the ratio of the uncertainties of model predictions and observations. We refer to Grum et al. (2002) for a simplified explanation of this approach.

In the model equations as well as in the uncertainty description for model predictions and observations, parameters need to be defined. These are estimated using a maximum likelihood routine, maximizing the probability of obtaining all the measured values included in the observation equations.

### 2.2.1 Model Setup

The adjusted rainfall depth in each radar pixel in the considered area is considered a state. The rain intensity of the current time step is predicted as a weighted average of the rain intensities in the 3x3 neighbourhood of the pixel at the previous time step.

$$X_{i,j,t} = \sum_{k=-1}^1 \sum_{l=-1}^1 \alpha_{k,l} X_{i+k, j+l, t-1} + e_{i,j,t} \quad (1)$$

$X_{i,j,t}$  refers to the adjusted rainfall value at pixel  $(i,j)$  in the radar matrix at time step  $t$ ,  $e_{i,j,t}$  to the corresponding gaussian prediction error with variance  $\sigma_x$  and  $\alpha_{k,l}$  to the weighting factors. The pixels adjacent to pixel  $(i,j)$  are indexed with  $k$  for row shifts and  $l$  for column shifts. We define

$$\alpha_{k,l} = \begin{cases} a & \text{central pixel, } k = l = 0 \\ \frac{1-a}{8} & \text{non-central pixel, } k \neq 0, l \neq 0 \end{cases} \quad (2)$$

i.e. the sum of the weighting factors is 1 and all non-central pixels in the 3x3 neighbourhood receive the same weighting. In matrix notation we have

$$X_t = AX_{t-1} + e_t \quad (3)$$

where  $X$  is a vector containing 99 rainfall state values corresponding to 9x11 pixels,  $A$  is a 99x99 weighting matrix performing the spatial averaging defined in (1) and  $e$  is a 99x1 vector of model errors with covariance matrix  $\Sigma_1$  with constant variance  $\sigma_x$  for all states on the diagonal and 0 on all off-diagonal elements, i.e. no correlation between the states. States and measurements are related in the observation equation (4):

$$Y_t = CX_t + s_t \quad (4)$$

The observation vector  $Y_t$  contains 99 non-adjusted measurements from all radar pixels and 8 rain gauge measurements. The 107x107 matrix  $C$  relates states and observations (see Grum et al. (2002)) and  $s$  is a 107x1 vector of observation errors with covariance matrix  $\Sigma_2$ .

## 2.2.2 Observation Error Covariance Structures

We investigate different structures of the observation error covariance matrix  $\Sigma_2$ . Model 1 (Eq. 5) includes constant variances  $\sigma_R$  and  $\sigma_G$  for radar and rain gauge observations, respectively. Spatial correlation between observations is not considered.

$$\Sigma_2 = \begin{bmatrix} \sigma_R & & & \\ & \ddots & & 0 \\ & & \sigma_R & \\ & & & \sigma_G \\ 0 & & & \\ & & & \ddots \\ & & & & \sigma_G \end{bmatrix} \quad (5)$$

Model 2 extends the above setup by considering correlation  $\rho_R$  only between neighbouring radar pixel observations. The correlation is found in the parameter estimation procedure.

Model 3 considers correlation for each radar pixel observation with all other pixels. The correlation is assumed to decay as a power function of distance between the pixels according to Eq. 6, where the distance  $D$  between pixels is defined in no. of pixels and parameters  $\rho_a$  and  $\rho_b$  are estimated from the variogram of the radar observations and fixed during the maximum likelihood estimation of the whole model. No correlation is considered for the rain gauge measurements.

$$\rho = \rho_a \cdot D^{\rho_b} \quad (6)$$

Model 4 is equivalent to model 1. However, we do in addition introduce an error marker. If a radar or rain gauge observation is missing, the corresponding variance is set to a large value and the correlation values are set to 0.

## 2.3 Stochastic Runoff Forecasting

The estimation of rainfall forecast models does not permit a direct evaluation of the quality of the adjusted radar data. We therefore generate runoff forecasts with different rainfall inputs and evaluate the forecast quality. We use stochastic greybox models for generating the forecasts with focus on wet weather periods as these are most relevant for real time control. When generating predictions we assumed a perfect rain forecast was available.

A lumped model consisting of a cascade of two reservoirs is applied for both catchments. The model setup and development is described in Breinholt et al. (2011) using the (smaller) Ballerup catchment as an example. For the (bigger) Damhusåen catchment better forecasts could most likely be obtained by applying a more elaborated model. However, here we are mainly interested in the effect of different rainfall inputs on the forecast quality, not the best forecasting model. We consider the following lumped model structure:

$$d\begin{bmatrix} S_{1,t} \\ S_{2,t} \end{bmatrix} = \begin{bmatrix} A \cdot P_t + a_0 - \frac{1}{K} S_{1,t} \\ \frac{1}{K} S_{1,t} - \frac{1}{K} S_{2,t} \end{bmatrix} dt + \begin{bmatrix} \sigma(S_{1,t}) \\ \sigma(S_{2,t}) \end{bmatrix} d\omega_t \quad (7)$$

$$\log(Q_k) = \log\left(\frac{1}{K} S_{2,k} + D_k\right) + e_k \quad (8)$$

Similarly to the rainfall model described above, the model is laid out as a state-space model where Eq. 7 is termed system or state equation and Eq. 8 observation equation.  $S_1$  and  $S_2$  correspond to the storage states,  $A$  to the impervious catchment area,  $P_t$  to the rain intensity,  $a_0$  to the mean dry weather flow and  $K$  to the travel time constant. The uncertainty of model predictions is captured by the Wiener process  $d\omega_t$  with incremental variance  $\sigma^2$ . The variance depends on the current state values, so a Lamperti transform is applied and the estimation performed with transformed states (Breinholt et al. (2011)). In Eq. 8  $Q$  corresponds to the observed flow values,  $D$  describes the variation of the dry weather flow using trigonometric functions and  $e$  corresponds to the observation error with standard deviation  $\sigma_e$ .

Differently from Breinholt et al. (2011) we do not estimate the model parameters based on one-step ahead flow forecasts. The runoff forecasts are intended to be used in a model predictive real time control setup (Grum et al. (2011)). The relevant decision variable in the setup is expected runoff volume over the prediction horizon. We therefore compute the expected flow values for the next 10 time steps (step length  $\Delta t=10\text{min}$ ) starting from time step  $k$  and integrate them to a predicted runoff volume (Eq. 9).

$$\hat{V}_k = \left( \sum_{i=1}^{10} \hat{Q}_{k+i} \right) \cdot \Delta t \quad (9)$$

The extended Kalman filter used in the modelling procedure also provides a variance for each predicted flow value. Assuming normal distribution, we derive a 95% prediction interval on the flow predictions for each horizon. Equivalent to Eq. 9, we integrate the upper and lower bounds for the different horizons.

$$\hat{V}_{k,up} = \left( \sum_{i=1}^{10} (\hat{Q}_{k+i} + n_{0.975} \cdot \sigma_{\hat{Q}_{k+i}}) \right) \cdot \Delta t \quad (10)$$

$$\hat{V}_{k,low} = \left( \sum_{i=1}^{10} (\hat{Q}_{k+i} - n_{0.975} \cdot \sigma_{\hat{Q}_{k+i}}) \right) \cdot \Delta t \quad (11)$$

In Eq. 10 and 11 indices *up* and *low* mark the upper and lower prediction bounds, respectively,  $n_{0.975}$  the 97.5% quantile of the standard normal distribution and  $\sigma_{\hat{Q}_{k+i}}$  the standard deviation of the flow prediction  $i$  steps into the future starting from time step  $k$ . Eq. 10 and 11 correspond to an average of the upper and lower flow prediction bounds, not to actual prediction bounds for the volume forecast. Though we obtain a simplified estimate for the uncertainty of volume predictions that we here apply to compare the quality of forecasts generated using different rainfall inputs.

Comparing the bounds derived in Eq. 10 and 11 to the observed runoff volume, we find the optimal model parameters by minimizing the skill score ( $S_k$ ) described in the next section. As wet weather periods are the main focus of real time control, only the model parameters relevant to runoff ( $A$ ,  $K$ , uncertainty parameters) are estimated and dry weather periods are excluded from the evaluation of the skill score function. The dry weather parameters ( $a_0$ ,  $D$ ) for the two catchments are estimated deterministically from a 14 day dry weather period at the beginning of the considered period and then fixed during estimation of the other model parameters.

The model parameters describing forecast uncertainty are only influenced by wet weather situations resulting in a good fit during wet weather (e.g. Figure 2) and an overestimation of uncertainty during dry weather. We consider this uncritical as we are here interested only in the forecast quality during rain events. In an online application, the model structure needs to be changed to properly distinguish between wet and dry weather periods.

## 2.4 Forecast Evaluation

When evaluating stochastic flow forecasts, we need to consider the quality of prediction intervals rather than just a mean squared error between prediction and observation. Criteria for forecast evaluation were proposed by Jin et al. (2010) and Thordarson et al. (2012):

- Reliability (*Rel*) – percentage of observations not contained in a 95 % prediction interval. A reliability of 5% means that we as intended include 95% of the observations in the interval.
- *ARIL* - average width of the 95 % prediction interval (=Sharpness *Sh*) relative to observation

- Skill score (*Sk*) 
$$Sk = Sh + \frac{2}{0.05 \cdot N} \sum_i (U_i + L_i)$$

where  $N$  is the number of wet weather observations,  $Sh$  is the average width of the 95 % prediction interval and  $U_i$  and  $L_i$  are the distances of the  $i$ -th observation from the upper / lower prediction interval (over-/ undershoots).  $U_i$  and  $L_i$  are 0 if the observation is contained in the prediction band.

We compute these criteria for a runoff volume prediction interval as described above. Only wet weather periods are considered in the computation of the evaluation criteria to obtain a more clear indication of forecast quality during rain events.

## 3 RESULTS

Table 1 shows the parameters derived for the radar adjustment models 1-4. The weighting factor  $a=0.2$  of the central pixel indicates that the model predictions include information from the whole 3x3 neighbourhood, rather than just the central pixel. The variance of the model predictions  $\sigma_x$  is generally estimated smaller than that of the observations  $\sigma_R$  and  $\sigma_G$ . When computing the adjusted rainfall values, the Kalman filter will therefore show a tendency to smoothen the observed values.

We further observe that including spatial correlation between the radar observations into the model (Models 2-3) increases the variance of the rain gauge observation errors. The correlation term reduces the weight of the single radar observation and allows for retrieving information from the gauges also if they are considered more uncertain. Similarly, the variance of the model prediction errors  $\sigma_x$  can be increased in this case as less weight is put on the observations.

Generally, the estimation of the state-space radar adjustment models using the described maximum likelihood approach has turned out problematic in application. Similar likelihood values may be obtained for rather different sets of parameters leading to poor identifiability of the models. The improved flow forecasts obtained with adjusted radar input as compared to e.g. rain gauge input (Table 2) indicate that we were able to identify reasonable parameter sets. Improved estimates can most likely be obtained if an objective function based on flow measurements is also used to find the parameters of the rainfall adjustment models.

Table 1. Parameter values for state-space radar adjustment models

Model	$a$	$\sigma_x$	$\sigma_R$	$\sigma_G$	$\rho_R$	$\rho_a$	$\rho_b$
Model 1	0.20	$1.23 \cdot 10^{-4}$	$8.12 \cdot 10^{-4}$	$9.12 \cdot 10^{-4}$	-	-	-
Model 2	0.20	$1.58 \cdot 10^{-4}$	$4.90 \cdot 10^{-4}$	$3.39 \cdot 10^{-1}$	0.22	-	-
Model 3	0.20	$1.51 \cdot 10^{-4}$	$8.12 \cdot 10^{-4}$	$3.80 \cdot 10^{-1}$	-	0.614	0.384
Model 4	0.20	$1.23 \cdot 10^{-4}$	$8.12 \cdot 10^{-4}$	$9.12 \cdot 10^{-4}$	-	-	-

Table 2. Forecast quality criteria for the two catchments with different rainfall inputs. Measures are given in  $\text{m}^3$  per 100 min and values are evaluated in wet weather periods only

Model Input	Ballerup catchment			Damhusåen catchment		
	<i>Rel</i>	<i>ARIL</i>	<i>Sk</i>	<i>Rel</i>	<i>ARIL</i>	<i>Sk</i>
Rain gauge	5%	65%	1466	4%	116%	11777
Radar no adjustment	5%	56%	1378	6%	95%	12283
Radar Model 1	5%	56%	1342	6%	90%	10975
Radar Model 2	5%	57%	1345	5%	92%	10721
Radar Model 3	5%	64%	1403	6%	93%	11265
Radar Model 4	5%	59%	1339	5%	94%	10479

Table 2 shows the results of the stochastic runoff volume forecasts generated using the different rainfall inputs. The prediction intervals in the Damhusåen catchment are generally wider than in the Ballerup catchment, indicating a too simple model structure for this catchment. Still, the simple model allows us to judge the quality of different rainfall inputs for flow forecasting. Comparing the volume forecast quality obtained with pure rain gauge and pure radar rainfall input to that obtained with adjusted radar rainfall input, we notice skill scores improved by 3-15%. The prediction intervals are generally narrower when using radar rainfall input compared to rain gauge input.

Including correlation into the covariance structure of the radar observations (Models 2 and 3) does not give clear improvements of the runoff forecasts. At this stage it is not possible to conclude if consideration of this effect actually has no significant effect on flow predictions. A better estimation method for the radar adjustment models may be able to exploit this effect better.

Figure 2 illustrates the effect of using an error marker in the adjustment of the radar data. The radar observations are missing for the small events between time steps 11500 and 11800. Using the error marker in Model 4, we are able to reconstruct rainfall values from the gauge observations. The runoff prediction of this model is consequently closer to the observation than that of Model 1 where forecast values increase not as a result of rainfall input but due to the adjustment of the runoff model to new (increasing) flow observations (state updating, c.f. Breinholt et al. (2011)).

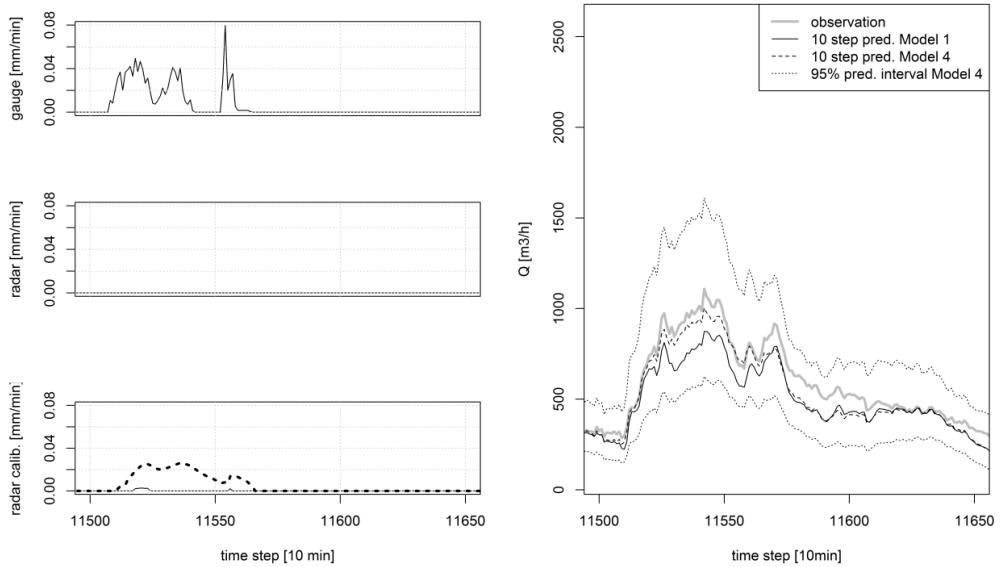


Figure 2. Left - mean area rain intensities for the Ballerup catchment from gauge measurements (top), non-adjusted radar measurements (centre, missing values in this case) and adjusted radar measurements (bottom) without (Model 1, full thin line) and with (Model 4, dotted bold) error marker

#### 4 CONCLUSIONS

We have evaluated the possibility of adjusting radar rainfall measurements with rain gauge measurements using state space models and evaluated the effect of the adjustment on runoff forecasts for two catchments generated from stochastic greybox models. With the adjusted radar data as input we obtain improved runoff forecasts as compared to using rain gauge or non-adjusted radar data as model input. Using an error marker allows to reconstruct adjusted rainfall values also if radar or some of the gauge observations are missing.

Despite the improved flow forecasts obtained with the adjusted radar data, we see several possibilities for improvements of the presented approach. Estimating parameters for the radar rainfall adjustment based on rainfall observations only has proven difficult. Better results can most likely be obtained by including a runoff prediction into the rainfall adjustment model and comparing predicted and observed runoff.

Further, we have considered an area of 9x11 C-band radar pixels in this study. This area is sufficient to cover the whole of Copenhagen. However, it is too small to generate short term rainfall forecasts from the radar. Preferably, the whole radar matrix of 240x240 pixels should be considered for this purpose corresponding to 57600 observations. Operating on variance matrices with 57600x57600 entries in the Kalman filtering procedure is impossible. A modified procedure that directly estimates the Kalman gain may be a possible solution to this problem.

With respect to the runoff forecasting models, improved forecasts for the bigger catchment can very likely be obtained by applying a more elaborated model structure that accounts e.g. for effects such as overflows. Further, improvements could be obtained by modelling prediction uncertainties depending on rainfall characteristics. These characteristics should aim at identifying convective events as these

imply the highest forecast uncertainties. Using these methods we aim at providing forecasts that clearly improve decision making in real time control of sewer networks.

## 5 ACKNOWLEDGEMENTS

This research is financially supported by the Danish Council for Strategic Research, programme Commission on Sustainable Energy and Environment through the Storm- and Wastewater Informatics (SWI) project. Catchment and flow data were kindly provided by Avedøre Wastewater Services and Copenhagen Energy. Radar data from the Danish Meteorological Institutes (DMI) C-Band radar at Stevns were kindly provided from a repository at Aalborg University. We thank Pjerre-Julien Trombe at DTU Compute for assistance and advice in handling radar data.

## 6 REFERENCES

- Breinholt A., Thordarson F.Ø., Møller J.K., Grum M., Mikkelsen P.S. and Madsen H. (2011). Grey-box modeling of flow in sewer systems with state-dependent diffusion. *Environmetrics*, 22(8), 946-961.
- Brown P.E., Diggle P.J., Lord M.E. and Young P.C. (2001). Space-time calibration of radar rainfall data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 50(2), 221-241.
- Chumchean S., Seed A. and Sharma A. (2006). Correcting of real-time radar rainfall bias using a Kalman filtering approach. *Journal of Hydrology*, 317, 123-137.
- Costa M. and Alpuim T. (2009). Adjustment of state space models in view of area rainfall estimation. *Environmetrics*, 22, 530-540.
- Grum M., Harremoes P. and Linde J.J. (2002). Assimilating a multitude of rainfall and runoff data using a stochastic state space modeling approach. In *Proceedings of the 9th International Conference on Urban Drainage*, 8–11 September 2002, Portland, Oregon USA.
- Grum, M., Kraemer, S., Verworn, H. and Redder, A. (2005). Combined use of point rain gauges, radar, microwave link and level measurements in urban hydrological modelling. *Atmospheric Research*, 77, 313-321.
- Grum M., Thornberg D., Christensen M.L., Shididi S.A. and Thirsing C. (2011). Full-scale real time control demonstration project in Copenhagen's largest urban drainage catchments. In *Proceedings of the 12th International Conference on Urban Drainage*, 11-16 September 2011, Porto Alegre, Brazil.
- Jin X., Xu C.-Y., Zhang Q. and Singh V. P. (2010). Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model. *Journal of Hydrology*, 383(3-4), 147-155.
- Jørgensen H.K., Rosenørn S., Madsen H. and Mikkelsen P.S. (1998). Quality control of rain data used for urban runoff systems. *Water Science and Technology*, 37(11), 113-120.
- Schilling, W. (1991). Rainfall data for urban hydrology: what do we need? *Atmospheric Research*, 27, 5-21.
- Thordarson F.Ø., Breinholt A., Møller J.K. ,Mikkelsen P.S., Grum M. and Madsen H. (2012). Evaluation of probabilistic flow predictions in sewer systems using grey box models and a skill score criterion. *Stochastic Environmental Research and Risk Assessment*, 26(8), 1151-1162.
- Thorndahl S., Rasmussen M.R., Neve S., Poulsen T.S. and Grum M. (2010). Vejrradarbaseret styring af spildevandsanlæg (Weather radar based control of wastewater systems). *DCE Technical Report No. 95, Aalborg University, Institut for Byggeri og Anlæg*, ISSN 1901-726X.



PAPER B

# Probabilistic online runoff forecasting for urban catchments using inputs from rain gauges as well as statically and dynamically adjusted weather radar

---

**Authors:**

Roland Löwe, Søren Thorndahl, Peter Steen Mikkelsen, Michael Rasmussen,  
Henrik Madsen

**Accepted by:**

*Journal of Hydrology*, 2014.



# Probabilistic online runoff forecasting for urban catchments using inputs from rain gauges as well as statically and dynamically adjusted weather radar

Roland Löwe<sup>1</sup>, Søren Thorndahl<sup>2</sup>, Peter Steen Mikkelsen<sup>3</sup>, Michael Rasmussen<sup>2</sup>, Henrik Madsen<sup>1</sup>

## Abstract

We investigate the application of rainfall observations and forecasts from rain gauges and weather radar as input to operational urban runoff forecasting models. We apply lumped rainfall runoff models implemented in a stochastic grey-box modelling framework. Different model structures are considered that account for the spatial distribution of rainfall in different degrees of detail.

Considering two urban example catchments, we show that statically adjusted radar rainfall input improves the quality of probabilistic runoff forecasts as compared to input based on rain gauge observations, although the characteristics of these radar measurements are rather different from those on the ground. Data driven runoff forecasting models can to some extent adapt to bias of the rainfall input by model parameter calibration and state-updating. More detailed structures in these models provide improved runoff forecasts compared to the structures considering mean areal rainfall only.

A time-dynamic adjustment of the radar data to rain gauge data provides improved rainfall forecasts when compared with rainfall observations on the ground. However, dynamic adjustment reduces the potential for creating runoff forecasts and in fact also leads to reduced cross correlation between radar rainfall and runoff measurements. We conclude that evaluating the performance of radar rainfall adjustment against rain gauges may not always be adequate and

---

<sup>1</sup>Department of Applied Mathematics and Computer Science, Technical University of Denmark, Matematiktorvet, B322, DK-2800 Kgs. Lyngby, Denmark

<sup>2</sup>Department of Civil Engineering, Aalborg University, Sohngårdsholmsvej 57, DK-9000 Aalborg, Denmark

<sup>3</sup>Department of Environmental Engineering, Technical University of Denmark, Miljøvej, B115, DK-2800 Kgs. Lyngby, Denmark

that adjustment procedure and online runoff forecasting should ideally be considered as one unit.

1 **Probabilistic online runoff forecasting for urban**  
2 **catchments using inputs from rain gauges as well as**  
3 **statically and dynamically adjusted weather radar**

4 Roland Löwe<sup>a,\*</sup>, Søren Thorndahl<sup>c</sup>, Peter Steen Mikkelsen<sup>b</sup>, Michael R. Rasmussen<sup>c</sup>,  
5 Henrik Madsen<sup>a</sup>

6

7 \*correspondence to:rolo@dtu.dk, +45-45253399

8 <sup>a</sup>Department of Applied Mathematics and Computer Science, Technical University of Denmark (DTU  
9 Compute), Matematiktorvet B322, Kgs. Lyngby, 2800, Denmark, rolo@dtu.dk, hmad@dtu.dk

10 <sup>b</sup>Department of Environmental Engineering, Technical University of Denmark (DTU Environment),  
11 Miljøvej B113, Kgs. Lyngby, 2800, Denmark, psmi@env.dtu.dk

12 <sup>c</sup>Department of Civil Engineering, Aalborg University, Sohngårdsholmsvej 57, Aalborg, 9000,  
13 Denmark, st@civil.aau.dk, mr@civil.aau.dk

14

15 *ABSTRACT*

16

17 We investigate the application of rainfall observations and forecasts from rain gauges  
18 and weather radar as input to operational urban runoff forecasting models. We apply  
19 lumped rainfall runoff models implemented in a stochastic grey-box modelling  
20 framework. Different model structures are considered that account for the spatial  
21 distribution of rainfall in different degrees of detail.

22

23 Considering two urban example catchments, we show that statically adjusted radar  
24 rainfall input improves the quality of probabilistic runoff forecasts as compared to

25 input based on rain gauge observations, although the characteristics of these radar  
26 measurements are rather different from those on the ground. Data driven runoff  
27 forecasting models can to some extent adapt to bias of the rainfall input by model  
28 parameter calibration and state-updating. More detailed structures in these models  
29 provide improved runoff forecasts compared to the structures considering mean areal  
30 rainfall only.

31

32 A time-dynamic adjustment of the radar data to rain gauge data provides improved  
33 rainfall forecasts when compared with rainfall observations on the ground. However,  
34 dynamic adjustment reduces the potential for creating runoff forecasts and in fact  
35 also leads to reduced cross correlation between radar rainfall and runoff  
36 measurements. We conclude that evaluating the performance of radar rainfall  
37 adjustment against rain gauges may not always be adequate and that adjustment  
38 procedure and online runoff forecasting should ideally be considered as one unit.

39

40 *KEYWORDS*

41 Stochastic grey-box model, radar rainfall, radar adjustment, probabilistic forecasting,  
42 real time control, urban hydrology

43

44 Received

45 2014/02/21

46

47

48    **1 INTRODUCTION**

49

50    Urban catchments are typically of a spatial extent where a homogeneous distribution  
51    of rainfall over the catchment cannot be assumed. This is one of the main drivers for  
52    developing real time control (RTC) setups for urban drainage systems. The load on  
53    the sewer network is higher in some places than in others, which results in an uneven  
54    use of the available storage capacities. This suboptimal load distribution can be  
55    improved by a dynamic operation of the network. As a result, combined sewer  
56    overflows can be reduced, for example.

57

58    Real time control systems are in operation in a multitude of urban catchments (Fuchs  
59    and Beenenken, 2005; Pleau et al., 2005; Sharma et al., 2012, Seggelke et al., 2013).  
60    Classically, decision making is done on the basis of offline knowledge about the  
61    system, for example in a framework of decision rules. More recent developments  
62    incorporate an online optimization of the system that accounts for runoff forecasts  
63    (Puig et al., 2009; Vezzaro and Grum, 2012). The control setup suggested in Vezzaro  
64    and Grum (2012) makes it possible to account for forecast uncertainties in the  
65    optimization and decision making process.

66

67    In a dynamic optimisation based real time control setup, simplified rainfall runoff  
68    models that lump a bigger part of the catchment are typically applied for forecasting  
69    over short horizons of a few hours as they are fast enough to generate forecasts  
70    within seconds to minutes (for example Pleau et al., 2001, Puig et al., 2009, Vezzaro  
71    and Grum (2012)). Using highly simplified models for forecasting is also common in

72 other fields like district heating (Nielsen and Madsen, 2006) or wind power  
73 forecasting (Giebel et al., 2011). Apart from being computationally efficient, lumped  
74 models make the application of statistical techniques such as state-updating and  
75 automated parameter calibration easier. Generating runoff forecasts in such an on-  
76 line setup is the case we consider here.

77

78 Generating runoff forecasts on-line requires rainfall inputs. For forecast horizons up  
79 to two hours, rainfall radars are currently the only means that provide the possibility  
80 to generate rainfall forecasts with a spatial and temporal resolution suitable for urban  
81 catchments. Examples of radar rainfall forecasting systems applied for quantitative  
82 online predictions in urban drainage systems are rare (Einfalt et al., 2004), but can  
83 for example be found in Einfalt et al. (1990), Kraemer et al. (2005) and Thorndahl  
84 and Rasmussen (2013).

85

86 Emmanuel et al. (2012a) discourage the direct application of the French operational  
87 weather radar product for quantitative purposes in urban hydrology. Similarly, other  
88 authors propose an adjustment of radar data to rain gauge measurements (Thorndahl  
89 et al., 2009; Villarini et al., 2010). Whereas the results of Villarini et al. (2010)  
90 suggest a constant bias between radar and rain gauge measurements during an event,  
91 other authors propose adjustment of radar measurements to gauge data also in the  
92 course of an event (Borup et al., 2009; Brown et al., 2001; Chumchean et al., 2006;  
93 Thorndahl et al., 2009, Wang et al., 2013, Wood et al., 2000). Gjertsen et al. (2003)  
94 and Goudensoofdt and Delobbe (2009) give overviews of different methods applied  
95 in Europe.

96

97 Radar adjustment is quite usually demonstrated to be beneficial by validating  
98 adjusted radar observations against rain gauge observations (Goudenhoofdt and  
99 Delobbe, 2009, Smith et al., 2007, Thorndahl et al., 2014, Wang et al., 2013) or by  
100 generating runoff forecasts from models that were statically calibrated using rain  
101 gauge input (Borup et al., 2009, Cole and Moore, 2008, Vieux and Bedient, 2004,  
102 Wang et al., 2013). The improvement in runoff forecasting performance may  
103 however be less clear for auto-calibrated online models that can dynamically adapt to  
104 observations as well as different rainfall inputs. In such cases the skill of different  
105 quantitative precipitation estimates to describe runoff should be assessed instead.  
106 Gourley and Vieux (2005) follow this thought on a 1200 km<sup>2</sup> catchment to compare  
107 results of spatially variable radar adjustments against mean field bias adjustment by  
108 evaluating hydrologic simulation results with different rainfall inputs and ensembles  
109 of different model parameters. They argue that rain gauge data may not be sufficient  
110 for the validation of quantitative precipitation estimates (QPE) as they are often used  
111 in the QPE algorithm itself, because rain gauge point measurements are often  
112 inaccurate and because there are issues of different scales between rain gauges and  
113 remotely sensed rainfall. The value of time varying radar adjustments for urban  
114 online runoff forecasting is in our view unclear.

115

116 A second issue in the generation of online runoff forecasts is the required spatial  
117 resolution of the rainfall input. A multitude of studies have been performed in  
118 hydrology as to what degree of spatial model resolution is appropriate. The results  
119 from the Distributed Model Intercomparison Project (Reed et al., 2004) show in a

120 non-urban context that conceptual models outperformed distributed models in the  
121 majority of cases. Das et al. (2008) give an overview of studies and find that  
122 generally, a higher spatial resolution does not necessarily lead to improved model  
123 performance. The authors conclude that a multitude of factors like scale of the  
124 catchment, physiographic characteristics or data availability influence model  
125 performance and that a lower, optimal limit of spatial resolution is to be expected  
126 because the model “represents spatial average behaviour”. This is underlined by  
127 results obtained by the authors in predicting river discharge from a 4000 km<sup>2</sup>  
128 catchment using different degrees of spatial resolution of model input data.

129

130 In urban hydrology, where catchment response is generally much faster than in  
131 natural catchments and data typically available in higher resolutions, Schilling (1984)  
132 and Schilling and Fuchs (1986) find that spatial rainfall variability is the key factor  
133 for the accuracy of simulations of urban runoff and that rainfall estimation errors are  
134 amplified by the rainfall runoff models. The authors suggest the use of high  
135 resolution rainfall data and simplified models for on-line operations. Using a  
136 hydrodynamic modelling setup for an 1100 ha catchment, Schellart et al. (2011)  
137 conclude that spatial resolution of inputs should be high (in their case 1 km<sup>2</sup>) in order  
138 to obtain a good representation of the observed flows in the sewer network. Finally,  
139 Berne et al. (2004) suggest a spatial rainfall resolution of 3 km for a 1000 ha  
140 catchment, while Emmanuel et al. (2012b) suggest 2.5 km resolution for a 600 ha  
141 catchment and Schilling (1991) suggests 1 km for on-line purposes. Studies in urban  
142 hydrology generally point in a direction where improved spatial resolution of rainfall  
143 inputs leads to improved model performance, a result which is less clear in modelling

144 of river flows as the spatial scales considered are much larger and data more scarce.  
145 We note that previous studies in urban hydrology focused on simulation, not on the  
146 case of on-line runoff forecasting with models that adapt to observations, although  
147 similar results may be expected.

148

149 Despite the above discussed results on model performance considering different  
150 spatial resolutions of rainfall inputs, a practitioners approach to building an on-line  
151 forecast model for real time control would often be to lump the catchment upstream  
152 from a control point. Practical experience suggests that the effect of this lumping on  
153 runoff simulation quality is limited (Achleitner et al., 2007; Grum et al., 2011; Wolfs  
154 et al., 2013). Similar to previous studies in natural catchments (Das et al., 2008), we  
155 therefore consider lumped models of different spatial resolutions for runoff  
156 forecasting in urban catchments over short horizons.

157

158 Finally, runoff forecasts generated by any model are uncertain due to uncertain  
159 measurements and forecasts of the rainfall input as well as an incomplete description  
160 of the reality by the model. Achleitner et al. (2009) and Thorndahl and Rasmussen  
161 (2013) evaluate the quality of urban runoff forecasts using radar rainfall input.  
162 Acceptable forecast errors could be obtained for forecast horizons of 90 and 60  
163 minutes, respectively. In an online setting, however, predicting also the uncertainty  
164 of runoff forecasts is of strong interest. The performance of lumped rainfall-runoff  
165 models in a stochastic grey-box layout was evaluated by Breinholt et al. (2011) and  
166 Thordarson et al. (2012) but rainfall input was assumed known. We here present an

167 evaluation of probabilistic runoff forecast quality that can be obtained in a realistic  
168 on-line setting.

169

170 Other approaches for modelling uncertainty in conceptual models exist and these  
171 apply Bayesian frameworks (Del Giudice et al., 2013, Kuczera et al. 2006, Renard et  
172 al., 2010), for example, GLUE (Breinholt et al., 2013, Dotto et al., 2012, Thorndahl  
173 et al., 2008) or simple output error methods (Breinholt et al., 2012). The approach  
174 presented here distinguishes itself in the explicit focus on forecasting over a  
175 multitude of horizons on a short time scale instead of describing simulation  
176 uncertainty and thus improving the capability of the model to describe reality. In  
177 addition, high computational efficiency is a focus of the presented approach.

178

179 In the following, the article first gives an introduction to the rainfall data considered  
180 as input for runoff forecasting in this study. Rainfall observations and forecasts from  
181 rain gauges and two types of C-band radar data are evaluated and compared. The  
182 types of weather radar data considered are

- 183 • temporally and spatially constant adjustment over the whole period (static  
184 adjustment)
- 185 • time-dynamic mean-field bias adjusted to rain gauge measurements in the  
186 course of an event, in addition to the static adjustment (dynamic adjustment).

187 The purpose of this evaluation is to demonstrate how the different rainfall  
188 measurements relate to each other and that the dynamic adjustment indeed makes the  
189 radar observations resemble the ground measurements more closely.

190

191 Subsequently, the different rainfall measurements and forecasts are considered as  
192 inputs for runoff forecasting. A quantification of probabilistic online runoff  
193 forecasting skill is provided on a 100 minute horizon. We evaluate if runoff forecasts  
194 can be improved by the different types of radar rainfall input and by an increased  
195 spatial resolution of the forecast model.

196

197 The article is structured as follows: section 2 describes the considered catchments,  
198 available rainfall measurements, the methodology for generating and evaluating  
199 stochastic runoff predictions, and the different model layouts considered. In section 3  
200 we compare the available rainfall measurements from gauges and radar in the area,  
201 and evaluate the runoff forecast quality obtained with different rainfall inputs and  
202 model layouts. Finally, in section 4 we conclude the article.

203

## 204 **2 MATERIAL AND METHODS**

### 205 **2.1 CATCHMENTS**

206

207 Two catchments in the Copenhagen area are considered in this study. The Ballerup  
208 catchment has a total area of approximately 1,300 ha. It is mainly laid out as a  
209 separate sewer system but has a small combined part and shows strong influences  
210 from rainfall-dependent infiltration and misconnection of surface runoff to sanitary  
211 sewers (Breinholt et al., 2013).

212

213 The Damhusåen catchment is located close to Ballerup but drains to a different  
214 treatment plant. We consider the northern part of the catchment with a total area of

215 approximately 3,000 ha. The catchment is laid out as a combined sewer system and  
216 consists of several subcatchments with a longest flow path of approximately 10 km.

217

218 An overview of the catchments can be seen in Figure 1. Flow measurements  
219 averaged over 5 min are available at the outlets of both catchments. Flow predictions  
220 are generated for both outlets and compared to the observations at 10 min resolution,  
221 where the measurements within an interval are averaged.

222

223

224 FIGURE 1 APPROX. HERE

225

226

227 **2.2 RAINFALL MEASUREMENTS AND FORECASTS USING GAUGES AND RADAR**

228 Observations from tipping bucket rain gauges from the Danish SVK network  
229 (Jørgensen et al., 1998) are available in the considered catchments. Rainfall  
230 measurements are available at 1 min intervals and averaged to 10 min time steps  
231 (equivalent to the temporal resolution of the radar data). In the rain gauge based  
232 forecast models we use 2 and 4 gauges as input for the Ballerup and Damhusåen  
233 catchments (Figure 1). The gauges are located within or close to the catchment  
234 borders.

235

236 Rainfall forecasts are generated from the gauge measurements using a local linear  
237 trend method. A trend line is fitted to the rain gauge intensities in the past 100 min  
238 and then extrapolated over the forecast horizon.

239

240 The Danish weather service operates a C-band radar in Stevns approx. 45 km south  
241 of the considered catchments (Gill et al., 2006). Measurements from this radar were  
242 made available for this study with a resolution of 10 min and 2x2 km<sup>2</sup>. Figure 1  
243 shows the location of the catchments within the utilized C-band radar pixels.

244

245 We apply radar rainfall forecasts with lead times up to 100 min generated by Aalborg  
246 University using the CO-TREC algorithm (Thorndahl and Rasmussen, 2013).

247 Corresponding to the available temporal resolution of the radar data, we apply all  
248 rainfall input data with a temporal resolution of 10 min. Considering the spatial  
249 extent of the catchments and concentration times  $t_c$  above 60 min, this resolution can  
250 be considered sufficient to capture the rainfall runoff process in the catchments.

251 Schilling (1991) suggests a temporal resolution of the rainfall data which is between  
252  $0.2t_c$  and  $0.33t_c$ .

253

### 254 **2.3 RADAR RAINFALL ADJUSTMENT**

255 C-band radar measurements are provided as reflectivities. A direct conversion to rain  
256 intensities is commonly considered problematic. A methodology to adjust the radar  
257 measurements to gauge observations has therefore been developed at Aalborg  
258 University and is applied here.

259

260 In the adjustment, the rain gauges marked in Figure 1 are used (SVK numbers 30252,  
261 30309, 30313, 30316, 30319, 30326, 30348 and 30386). The adjustment is  
262 performed with only 8 gauges distributed in the Copenhagen area, as one of the main

263 objectives for using radar rainfall measurements is to derive rain intensities using as  
264 small a number of ground measurements as possible.

265

266 In a first ‘static’ adjustment step, the coefficients in the reflectivity ( $Z$ ) – rain  
267 intensity ( $R$ ) relationship are adjusted for the whole data period (see Section 2.4).  
268 The rainfall depths from all rain events at all considered rain gauges are plotted  
269 against the rainfall depths derived from the radar observations in the corresponding  
270 pixels. The  $Z$ - $R$  coefficients are adjusted, such that the regression line between radar  
271 rainfall depths and rain gauge observations has slope 1 (Thorndahl et al., 2010). The  
272 resulting  $Z$ - $R$  relationship is used for deriving rain intensities over the whole data  
273 period.

$$Z = 50 \cdot R^{1.8} \quad (1)$$

274

275 In a second ‘dynamic’ adjustment step, the radar rain intensities are again adjusted,  
276 this time at every 10 min time step (Thorndahl et al., 2014). Considering the last 4  
277 observations, a spatially constant adjustment factor is derived, such that the radar  
278 measurements on average match the rain gauge measurements in the considered area.

279 This is a mean field bias adjustment in the sense of Goudenhoofdt and Delobbe  
280 (2009), however, with an adjustment window of 40 minutes instead of one day.

281

282 When generating forecasts, the time-dynamic adjustment factor is, over a period of  
283 120 minutes, linearly changed to 1 with increasing lead time. The linear transition  
284 towards zero-bias is performed because unrealistic and biased rainfall forecasts have

285 been observed on the longer lead times when forecasting with a time-dynamic  
286 adjustment factor based on only the past 40 minutes.

287

288 **2.4 DATA PERIOD**

289 We use a summer period of 2.5 months from 25/06/2010 until 6/09/2010 for  
290 generating probabilistic runoff forecasts. Figure 2 shows rain gauge and flow  
291 observations from the Ballerup catchment for this period. We can clearly identify the  
292 diurnal dry weather variations and a number of rain events that can be considered  
293 relevant for real time control purposes. The measurements contain no major gaps in  
294 this period.

295

296 FIGURE 2 APPROX. HERE

297

298

299 **2.5 STOCHASTIC FLOW FORECASTING**

300 **2.5.1 General Model Layout**

301 As mentioned before, we use stochastic grey-box models to generate flow forecasts  
302 for the catchments. In the basic setup we use a linear reservoir cascade of 2 storages  
303 with one rainfall input, implemented as stochastic differential equations in a state-  
304 space model layout (Breinholt et al., 2011). The model is at every time step updated  
305 to current flow observations using an extended Kalman filter (Kristensen et al.,  
306 2004).

307

308 This setup has been extensively tested for the Ballerup catchment but not for the  
 309 Damhusåen catchment. The model is obviously too simple, especially for the  
 310 (bigger) Damhusåen catchment. As we are mainly interested in investigating the  
 311 effects of different rainfall inputs on the forecasts, we still apply this most simple  
 312 setup. With respect to the magnitude of runoff forecast uncertainties, this could be  
 313 considered a ‘worst case scenario’.

314

$$d \begin{bmatrix} S_{1,t} \\ S_{2,t} \end{bmatrix} = \begin{bmatrix} A \cdot P + a_0 - \frac{1}{K} S_{1,t} \\ \frac{1}{K} S_{1,t} - \frac{1}{K} S_{2,t} \end{bmatrix} dt + \begin{bmatrix} \sigma_1 \cdot S_{1,t}^{\gamma_1} \\ \sigma_2 \cdot S_{2,t}^{\gamma_2} \end{bmatrix} d\omega_t \quad (2)$$

$$Q_k = \frac{1}{K} S_{2,k} + D_k + e_k \quad (3)$$

315

316 (2) is called the system equation, where  $S_1, S_2$  correspond to the storage states,  $A$  to  
 317 the impervious catchment area,  $P$  to the rain intensity,  $a_0$  to the mean dry weather  
 318 flow and  $K$  to the travel time constant. The uncertainty of model predictions is  
 319 described in the so-called diffusion term by a Wiener process  $\omega_t$ . The increments  $d\omega_t$   
 320 of this process are independent and normally distributed with a standard deviation  
 321 corresponding to the considered time interval  $dt$ .

322

323 The variance of the diffusion is here scaled dynamically depending on the current  
 324 model states  $S$  and a scaling factor  $\sigma$ . Such a scaling can be problematic for the  
 325 extended Kalman filtering. A Lamperti transform is therefore applied that removes  
 326 the state-dependency from the diffusion term and leads to a set of transformed drift

327 equations that equivalently describe the dynamics of the system, but have constant  
328 diffusion (Breinholt et al., 2011).

329

330 States and flow measurements are related in the observation equation (3).  $Q_k$   
331 corresponds to the observed flow values at times  $k$ ,  $D$  describes the variation of the  
332 dry weather flow using trigonometric functions and  $e$  corresponds to the observation  
333 error with standard deviation  $\sigma_e$ .

334

335 We refer to Kristensen et al. (2004) and Breinholt et al. (2011, 2012) for a detailed  
336 description of the modelling principles. We use the open-source software framework  
337 CTSM for the modelling process (Kristensen and Madsen, 2003).

338

### 339 **2.5.2 Stochastic Model Layout and Rainfall Inputs**

340 To investigate the influence of spatial resolution of rainfall inputs on the ability to  
341 create stochastic runoff forecasts, we consider the following model layouts:

- 342 • Area mean – the rainfall is assumed constant over the whole catchment and  
343 inputs from gauges or radar pixels are averaged (as shown in equation (2)).
- 344 • Integrated subcatchment – for radar inputs, the catchment is divided into  
345 subcatchments (Figure 1), an impervious area is estimated for every  
346 subcatchment, but only one storage cascade is used and all inputs are fed into  
347 the first storage. The same approach is applied for rain gauge input, but we  
348 estimate an effective area for every rain gauge and perform no assignment to  
349 subcatchments. This approach is applied for the (smaller and less complex)  
350 Ballerup catchment.

- Distributed subcatchments – every subcatchment has a cascade of 2 storages of its own and the outflows from the northern and eastern subcatchments are inputs to the western subcatchment. This approach is applied for the (bigger and more complex) Damhusåen catchment. In the simulation run with rain gauges, these are assigned to the closest subcatchment.

356

357 As a variety of rainfall inputs and model layouts are considered, in the following we  
358 denote the different simulation runs with a 3-letter identifier in accordance with  
359 Figure 3.

360

FIGURE 3 APPROX. HERE

362

363

364 2.6 PARAMETER ESTIMATION

Parameters for the proposed stochastic rainfall runoff models are estimated in an automated optimization routine. Most commonly this is done by maximizing the likelihood of one-step-ahead model predictions (Breinholt et al., 2011). In an online setup, however, the models are intended to provide multistep predictions. The model identified by minimizing the error of one-step-ahead predictions may not be the best model in terms of forecasting with longer lead times.

371

372 Further, if there is strong noise on the flow observations, the model may not be  
373 identifiable. The model setup includes a Kalman filtering procedure, which means  
374 that the model states are updated to follow the observations at each time step. If the

375 model is estimated on the basis of one-step-ahead predictions, there is a risk that the  
376 estimated model parameters simply optimize this state-updating and do not describe  
377 the physical behaviour of the system.

378

379 We therefore here apply a parameter estimation method that minimizes the error of  
380 the probabilistic multistep flow predictions (Löwe et al., 2014). The according  
381 criterion is the continuous ranked probability score (*CRPS*). At every time step, this  
382 score measures the squared difference between the cumulative distribution function  
383 (CDF) of the forecast and the CDF of the observation, where the latter is considered  
384 as a unit step at the observed value (Gneiting et al., 2005; Gneiting, 2007).

385

386 The dry weather parameters  $a_0$  and  $D$  of the model are assumed fixed and are  
387 estimated deterministically in a dry weather period of one week at the beginning of  
388 the considered time series. We apply a heuristic optimization algorithm described by  
389 Tolson and Shoemaker (2007) for automated parameter estimation.

390

## 391 **2.7 ON-LINE RUNOFF FORECAST GENERATION AND EVALUATION**

392 We evaluate the quality of probabilistic forecasts of runoff volume obtained from the  
393 different models. Runoff volumes are the relevant decision variable in a real time  
394 control setup for urban drainage systems as described e.g. by Vezzaro and Grum  
395 (2012). To obtain probabilistic predictions of runoff volume, we do at every time  
396 step generate 1000 realizations of multistep flow predictions from the model  
397 equations (2) using an Euler Maruyama scheme (Kloeden and Platen, 1999). We  
398 consider forecast horizons up to 10 steps or 100 min.

399

400 In this approach, forecast uncertainties are in the on-line setting only determined by  
401 the state uncertainties, not the observation uncertainties. This is reasonable as in a  
402 real time control scheme we are not interested in the observation uncertainty. The  
403 estimated observation uncertainties are furthermore small compared to the  
404 uncertainties of the model predictions (c.f. section 3.3).

405

406 Each of the 1000 multistep flow prediction scenarios can be integrated into a runoff  
407 volume prediction. We can then analyse the distribution of these values to obtain an  
408 empirical description of the predictive distribution of runoff volumes for each  
409 horizon. We evaluate the quality of the 10-step probabilistic runoff volume  
410 predictions as compared to the observed runoff volumes for this horizon. We  
411 consider the following criteria:

- 412 • Reliability (*Rel*) – percentage of observations included in a 90% prediction  
413 interval. Ideally, this value corresponds to 90%, higher values suggest an  
414 overfitted model, lower values an unreliable model.
- 415 • Average Interval Length (*ARIL*) – average width of the 90% prediction  
416 interval relative to the observations (Jin et al., 2010).
- 417 • Continuous ranked probability score (*CRPS*) – mean squared error of the  
418 predictive runoff volume distribution for a 10-step horizon. The best forecast  
419 minimizes this value (Gneiting, 2007).
- 420 • Root mean squared error (*RMSE*) between the 50% quantile of probabilistic  
421 runoff volume predictions and the corresponding observation.

422

423 **3 RESULTS AND DISCUSSION**

### 424 3.1 COMPARING RADAR AND RAIN GAUGE OBSERVATIONS AND FORECASTS

425 In a first step, the different rainfall observations are compared. The considered data  
426 period is split into rain events. Based on the spatially averaged rain gauge  
427 observations, it is assumed that a new event starts after 10 hours of dry weather.  
428 Rainfall intensities below 0.2mm/10min are considered dry weather and we only  
429 consider events with a total rainfall sum of at least 5 mm

430  
431 Based on the above considerations, 10 rain events are identified from the averaged  
432 rain gauge observations in the Ballerup catchment and used for comparison. Figure 4  
433 shows the total rainfall depth and the maximum intensity together with the duration  
434 of the events.

FIGURE 4 APPROX. HERE

437  
438 The effect of the dynamic radar adjustment clearly varies from event to event. Yet,  
439 on average, the root mean squared error (*RMSE*) between the total areal rainfall sums  
440 measured by radar and rain gauges is reduced from 9.8 mm with the static  
441 adjustment, to 7.3 mm for the dynamic adjustment. In the Damhusåen catchment (not  
442 shown) similar results are obtained with a reduction of the *RMSE* from 11.0 mm for  
443 the static adjustment to 6.7 mm for the dynamic adjustment.

445 Figure 5 supports the indications from the analysis of total rainfall sums. The  
446 dynamically adjusted radar observations seem to better capture the rainfall dynamics

447 observed on the ground. However, in both cases, the radar rainfall forecasts fail to  
448 predict the intense rainfall peak towards the end of the event. A delay is observed for  
449 the rainfall forecasts derived from the gauge measurements. This is induced by the  
450 forecast method which is based on extrapolating the observations of the last 100 min.

451

452 FIGURE 5 APPROX. HERE

453

454 Figure 6 shows the total rainfall sum for the different events derived from forecast  
455 values for a 2-step (20 min) and a 10-step (100 min) horizon. The simplistic forecast  
456 applied for the rain gauge data leads to a systematic overestimation of the total  
457 rainfall. The forecasts generated from the dynamically adjusted radar data are close  
458 to the rain gauge observations for the shorter horizon and approach the value for the  
459 statically adjusted data on the longer horizon. This is in accordance with the radar  
460 adjustment and forecasting methodology described in section 2.3.

461

462 FIGURE 6 APPROX. HERE

463

464 In Figure 4 one rain event can be identified (event 2) which is only present in the  
465 gauge measurements but not in the radar measurements for the Ballerup catchment.  
466 This deviation is a result of the gauges being located outside the catchment.

467

468 **3.2 CORRELATION BETWEEN RAINFALL AND RUNOFF OBSERVATIONS**

469  
470 Figure 7 shows the estimated cross correlation between catchment averaged rainfall  
471 observations and the measured runoff in the Ballerup catchment. For all rainfall

472 inputs, the highest cross correlation is identified for a lag of 16 time steps or 160  
473 minutes. The highest correlation between rainfall and runoff measurements is  
474 identified for statically adjusted radar measurements, and it is noticeably smaller for  
475 dynamically adjusted radar measurements. The same result is obtained in the  
476 Damhusåen catchment (not shown). It indicates that the type of time varying radar  
477 adjustment as described in section 2 may actually reduce the information about the  
478 runoff process that is contained in the radar rainfall time series. This is in spite of the  
479 fact that the time varying adjustment makes the radar data resemble the rainfall  
480 measurements on the ground more closely as described above.

481

482 FIGURE 7 APPROX. HERE

483

### 484 **3.3 PROBABILISTIC RUNOFF FORECASTING WITH DIFFERENT RAINFALL**

#### 485 **INPUTS**

486 We consider the quality of probabilistic runoff forecasts obtained using mean areal  
487 rainfall input derived from rain gauges and weather radar (model type a). Table 1  
488 shows the effective catchment area and the time constant estimated for the different  
489 models. We observe a tendency to estimate higher effective area values for the  
490 models with statically adjusted radar rainfall input. This is likely to be a result of the  
491 lower rain intensities in this type of input data (Figure 5).

492

493 Table 1. Estimated travel time constant ( $K$ ) and impervious catchment area ( $A$ ) for  
494 mean areal rainfall models (type a) with rain gauge (1) and statically (2) and

495 dynamically (3) adjusted radar input for Ballerup (B) and Damhusåen (D)

496 catchments.

Model	K [h]	A [ha]
B1a	4.62	70.6
B2a	4.50	74.4
B3a	3.79	61.4
D1a	2.01	278.3
D2a	4.45	392.1
D3a	4.66	253.4

497

498

499 Table 2 and Table 3 summarize the runoff forecast skill of the different models

500 averaged over all 10 events. We see that all models seem to be rather unreliable, with

501 only 51% to 72% of the observations included in a 90% prediction interval during

502 rain periods. Considering the whole data period, including dry weather periods, 84 to

503 92 % of the observations are included in a 90 % prediction interval (not shown).

504 During dry weather periods, the flows in the sewer system are low and follow the

505 well-defined diurnal cycle. The forecast error made by the runoff forecasting model

506 is thus much smaller than during rain events. The uncertainty description in the

507 model, however, accounts for dry and wet weather uncertainty in only one parameter.

508 Uncertainties during rain events are hence forecasted too small. A solution to this

509 problem could be to include a separate parameter for dry weather uncertainty in the

510 diffusion term of equation (2).

511

512 We further identify an insufficient quantification of forecast uncertainties, in

513 particular at the start of rain events (Figure 8). The reason is the state dependent

514 uncertainty description in the model, which only leads to high forecast uncertainties  
515 for high forecast values. Ideally, the forecast uncertainty should increase already at  
516 the start of the event. This may be achieved by conditioning the forecast uncertainty  
517 on the rainfall input instead of the state values, but it is not further investigated here.

518

519 The models with statically adjusted radar rainfall input (input type 2) perform best in  
520 both catchments and all model variations in matters of *RMSE*, whereas the models  
521 with dynamically adjusted radar rainfall input result in higher *RMSE* values (Table 2 and Table 3). The forecast uncertainties for the radar based models are in  
522 most cases estimated smaller. During rain periods, this leads to a more pronounced  
523 underestimation of forecast uncertainties, resulting in some cases in a lower  
524 probabilistic forecasting skill expressed as *CRPS*.

526

527 The better quality of runoff forecasts obtained with statically adjusted radar rainfall  
528 input as compared to dynamically adjusted radar rainfall input seems to somewhat  
529 contradict the results obtained by Borup et al. (2009). The authors showed that a  
530 dynamic calibration of X-band radar rainfall measurements results in better  
531 simulations of water levels at an overflow weir than a static calibration. Apart from  
532 using a different type of radar rainfall measurements in this work, we see the main  
533 reason for the differing results in the applied type of rainfall-runoff model and the  
534 way radar forecasts are generated. A distributed simulation model was applied in the  
535 work of Borup et al. (2009). These models are typically statically calibrated to reflect  
536 observations in the sewer system based on rain gauge input. The model parameters  
537 modified during the calibration (for example impervious area, surface roughness

538 values, pipe roughness values) do then also reflect the characteristics of rain gauge  
539 input (for example higher intensities as compared to radar rainfall measurements).  
540 Radar rainfall observations will consequently give better results the better they  
541 reflect the rainfall measurements on the ground, but different results may be obtained  
542 if the model is calibrated using radar rainfall input.

543

544 The rainfall runoff models applied in this work are data driven and fitted to the  
545 supplied input data. The rainfall input for this type of model may well be biased as  
546 compared to the “ground truth”, as the bias can be compensated for by different  
547 parameter estimates (for example the impervious catchment area) and by the state  
548 updating. The best runoff forecast will with this type of model be obtained with the  
549 rainfall input that has the highest ‘information content’ with respect to the runoff  
550 observations. This is the statically adjusted radar input in our case which is  
551 underlined by the fact that this type of rainfall measurement shows the highest cross  
552 correlation with the runoff time series.

553

554 More generally, it is interesting that the dynamically adjusted radar data appear to  
555 provide less information about the runoff time series than both, the rain gauge and  
556 the statically adjusted rainfall data. This is the case, not only for the radar rainfall  
557 forecasts, but also for the radar rainfall measurements (see the cross correlation  
558 function in Figure 7).

559

560 One likely reason is that the adjustment window of 40 minutes may be too short,  
561 leading to a nonlinear alteration of the radar data which cannot be compensated for

562 by the automatic calibration of the rainfall-runoff model. This effect may be  
563 amplified by the fact that radar rainfall measurements are made as ‘snapshots’ every  
564 10 minutes, while the rain gauge data used for radar adjustment are continuous over  
565 the interval. Recent works by Nielsen et al. (2014) and Thorndahl et al. (2014)  
566 suggested an interpolation of the radar data to a higher temporal resolution using an  
567 advective model and demonstrated that such processing reduces the bias as compared  
568 to rain gauge measurements. The effect of such interpolation schemes on on-line  
569 runoff forecasts needs to be investigated. In general, we suggest that the development  
570 of an adjustment methodology focuses not only on the deviation between radar  
571 rainfall estimates and rain gauges but also on the information content about the  
572 runoff time series.

573

574 Additionally, when generating rainfall forecasts, the bias between the dynamically  
575 adjusted radar forecast that is considered here and the observation on the ground  
576 changes linearly as a function of the forecast horizon. The reason is that especially in  
577 situations with sparse or very inhomogeneous rainfall within the radar range we risk  
578 adjusting the mean field based on very few radar-rain gauge pairs with very little  
579 observed rain. This might result in very small or very large adjustment factors.  
580 Applying these adjustment factors to the forecast has previously produced severe  
581 over- or underestimation of the forecasted rain. More rain gauges within the range of  
582 the radar might reduce the problem but would also reduce the added value of the  
583 radar. Generally, the non-constant bias in the rainfall forecasts introduces additional  
584 uncertainty in the runoff forecast. Improved results can likely be obtained if

585 replacing the simple linear change of the bias factor by time series models that are  
586 fitted in a way such that the runoff forecast error is minimized.

587

588 Table 2. Forecast evaluation for mean areal rainfall (type a) and integrated  
589 subcatchment (type b) models with different rainfall inputs for the Ballerup  
590 catchment (B). Values are based on predicted runoff volumes in m<sup>3</sup> over a prediction  
591 horizon of 100 min (10 time steps) and averaged over the considered rain events. We  
592 include RMSE values for 1-step and 10-step prediction horizons.

Model	Rel	ARIL	CRPS	RMSE 1	RMSE 10
	[%]	[%]		[m <sup>3</sup> ]	[m <sup>3</sup> ]
B1a	69%	30%	131.9	10.8	247.9
B1b	72%	30%	127.7	10.7	234.1
B2a	68%	29%	126.8	10.7	231.4
B2b	70%	30%	126.0	10.7	230.1
B3a	59%	23%	133.0	10.7	235.5
B3b	64%	27%	128.7	10.6	234.8

593

594 Table 3. Forecast evaluation for mean areal rainfall (type a) and distributed  
595 subcatchment (type c) models with different rainfall inputs for the Damhusåen  
596 catchment (D). Values are based on predicted runoff volumes in m<sup>3</sup> over a prediction  
597 horizon of 100 min (10 time steps) and averaged over the considered rain events. We  
598 include RMSE values for 1-step and 10-step prediction horizons.

Model	Rel	ARIL	CRPS	RMSE 1	RMSE 10
	[%]	[%]		[m <sup>3</sup> ]	[m <sup>3</sup> ]
D1a	66%	35%	1126.3	61.2	2864.1
D1c	51%	20%	1029.5	46.9	2112.9
D2a	53%	23%	1210.6	70.2	2330.2

D2c	51%	22%	962.6	45.1	1900.9
D3a	51%	21%	1262.5	70.9	2416.0
D3c	52%	21%	1133.4	47.3	2301.3

599

600 Evaluating the probabilistic runoff forecast skill obtained for the different events  
 601 (Figure 9), we see that the event with the highest volume and rain intensity (no. 6,  
 602 c.f. Figure 4 and Figure 6) also leads to rather high forecast errors. We cannot  
 603 identify a clear relation between event characteristics (Figure 4) and runoff forecast  
 604 qualities which may be due to the small number of events considered. For event 2 the  
 605 clearly lowest forecasting skill in the Ballerup catchment is observed when using rain  
 606 gauge input which is a result of the gauges being located outside the catchment as  
 607 discussed earlier.

608

609 FIGURE 8 APPROX. HERE

610

611 FIGURE 9 APPROX. HERE

612

613 **3.4 PROBABILISTIC RUNOFF FORECASTING WITH DIFFERENT SPATIAL  
 614 RESOLUTIONS**

615 Comparing model layouts that account for the spatial distribution of rainfall  
 616 observations in different degrees of detail, we can identify a trend that smaller  
 617 forecast errors are obtained with more complex model structures.  
 618 Table 2 compares the runoff forecasting skills in the Ballerup catchment. For all  
 619 rainfall inputs, slightly smaller *CRPS* and *RMSE* values are obtained on the 10-step  
 620 horizon for the integrated subcatchment approach (model type b). The estimation of a

621 separate effective area and using different rainfall inputs for each subcatchment,  
622 instead of averaging all inputs into a mean areal rainfall, consequently yields better  
623 results.

624

625 A similar trend can be observed in the Damhusåen catchment (Table 3). Accounting  
626 for the spatial rainfall distribution with a more complex model structure (distributed  
627 subcatchment approach - model type c) leads to a clear reduction in forecast error for  
628 all rainfall inputs. Following the discussion in Schilling and Fuchs (1986), this result  
629 was expected. Also with the slightly more complex model structure, models with  
630 statically adjusted radar input outperform those with rain gauge input and  
631 dynamically adjusted radar input (comparing models D1c, D2c and D3c in terms of  
632 *CRPS* and *RMSE*).

633

#### 634 **4 CONCLUSIONS**

635 The quality of probabilistic on-line runoff forecasts obtained with different types of  
636 rainfall input and different conceptual model layouts that account for the spatial  
637 distribution of rainfall in varying degrees of detail was analysed. Forecasts were  
638 generated for two urban catchments with forecast horizons of up to 100 min. A  
639 number of conclusions were identified with respect to the considerations described in  
640 section 1. These are summarized here.

641

642 1) The time-dynamic adjustment of radar observations to rain gauges that is applied  
643 here makes those data resemble the rain gauge observations more closely.

644

645 2) Radar rainfall observations and forecasts can improve the skill of probabilistic  
646 runoff forecasts compared with those based on rain gauges.

647

648 3) For all considered runoff model structures, the best results are obtained with radar  
649 input that is time-statically adjusted to rain gauge observations. The time varying  
650 (dynamic) adjustment of the radar data reduces the potential for creating runoff  
651 forecasts with the stochastic grey box models. In fact, also the cross correlation  
652 between radar rainfall and runoff measurements is reduced as a result of the time  
653 varying radar adjustment.

654

655 4) Rainfall inputs for conceptual, data-driven forecasting models need not be the  
656 same as the values observed by gauges on the ground. The model can to some extent  
657 adapt to the characteristics of the input series in the parameter estimation procedure  
658 and will give the best forecasts with the rainfall input that best explains the patterns  
659 in the flow observations. In this sense, the radar is likely to provide a better spatial  
660 representation of rainfall patterns which, although biased compared with the ground  
661 observations, leads to better runoff forecasts. It is, however, important that the bias of  
662 the radar observations is not altered in a non-constant fashion. The aim of the radar  
663 adjustment should in this context be to merge rainfall information from different  
664 sources in a statistically optimal way.

665

666 5) An evaluation of radar adjustment methodologies should not only focus on the  
667 comparison with rain gauge observations but also on the final purpose for the  
668 adjusted measurements. In our case, this was runoff forecasting with data-driven

669 models and the radar adjustment and the runoff forecasting models should  
670 consequently be considered as a chain and coordinated.

671

672 6) Generally, rainfall runoff forecasting models will yield best results if the applied  
673 rainfall input closely resembles the input used in model calibration. Distributed  
674 simulation models are typically calibrated to resemble observations in the sewer  
675 network based on rain gauge observations. Adjusting radar data to more closely  
676 resemble the observations of rain gauges will consequently improve the results  
677 obtained with these models. Any type of model calibrated using radar rainfall  
678 observations as input may, however, yield different results.

679

680 7) The probabilistic runoff forecasts obtained with the stochastic grey-box models  
681 improve if we account for the spatial distribution of rainfall in the model. The best  
682 forecasts in the Damhusåen catchment are obtained for the distributed subcatchment  
683 approach, i.e. when splitting the catchment into 3 subcatchments that are modelled  
684 by separate, connected reservoir cascades.

685

686 8) We can identify insufficiencies in the applied models. The uncertainty description  
687 based on the model states does not allow us to capture the high forecast uncertainty  
688 at the start of a rain event. An improved model layout should be obtained by making  
689 the model uncertainty depend on the rainfall input. Further, considering also dry  
690 weather periods during parameter estimation of the models leads to unacceptably  
691 small uncertainty estimates during rain events. Either, only periods with rainfall  
692 should be considered for parameter estimation, or the model structure should be

693 modified to allow for proper separation between forecast uncertainties during dry and  
694 wet weather.

695

696 **5 ACKNOWLEDGEMENTS**

697

698 This research has been financially supported by the Danish Council for Strategic  
699 Research, Programme Commission on Sustainable Energy and Environment through  
700 the Storm- and Wastewater Informatics (SWI) project. The catchment and flow data  
701 were kindly provided by Avedøre Wastewater Services and Copenhagen Utility  
702 Company (HOFOR). We thank the Danish Weather Service (DMI) for providing  
703 data from the C-Band radar at Stevns.

704

705 **6 REFERENCES**

706 Achleitner, S., Möderl, M., Rauch, W., 2007. CITY DRAIN © – An open source  
707 approach for simulation of integrated urban drainage systems. Environ. Model.  
708 Softw. 22, 1184–1195.

709

710 Achleitner, S., Fach, S., Einfalt, T., Rauch, W., 2009. Nowcasting of rainfall and of  
711 combined sewage flow in urban drainage systems. Water Sci Technol 59, 1145-1151.

712

713 Berne, A., Delrieu, G., Creutin, J.D., Obled, C., 2004. Temporal and spatial  
714 resolution of rainfall measurements required for urban hydrology. J Hydrol 299, 166-  
715 179.

716

- 717 Borup, M., Grum, M., Linde, J. J., Mikkelsen, P. S., 2009. Application of high  
718 resolution x-band radar data for urban runoff modelling: constant vs. dynamic  
719 calibration, 8th International Workshop on Precipitation in Urban Areas, 10-13  
720 December, 2009, St. Moritz, Switzerland.
- 721
- 722 Breinholt, A., Thordarson, F.O., Møller, J.K., Grum, M., Mikkelsen, P.S., Madsen,  
723 H., 2011. Grey-box modelling of flow in sewer systems with state-dependent  
724 diffusion. Environmetrics 22, 946-961.
- 725
- 726 Breinholt, A., Møller, J.K., Madsen, H., Mikkelsen, P.S., 2012. A formal statistical  
727 approach to representing uncertainty in rainfall-runoff modelling with focus on  
728 residual analysis and probabilistic output evaluation - distinguishing simulation and  
729 prediction, J Hydrol, 472-473, 36-52.
- 730
- 731 Breinholt, A., Grum, M., Madsen, H., Thordarson, F.Ö., Mikkelsen, P.S., 2013.  
732 Informal uncertainty analysis (GLUE) of continuous flow simulation in a hybrid  
733 sewer system with Infiltration Inflow – consistency of containment ratios in  
734 calibration and validation? Hydrol Earth Syst Sci. 17, 4159-4176
- 735
- 736 Brown, P.E., Diggle, P.J., Lord, M.E., Young, P.C., 2001. Space–time calibration of  
737 radar rainfall data. J Roy Stat Soc C-App 50, 221-241.
- 738
- 739 Cole, S.J., Moore, R.J., 2008. Hydrological modelling using raingauge- and radar-  
740 based estimators of areal rainfall. J Hydrol 358, 159-181.

- 741
- 742 Chumchean, S., Seed, A., Sharma, A., 2006. Correcting of real-time radar rainfall
- 743 bias using a Kalman filtering approach. *J Hydrol* 317, 123.
- 744
- 745 Das, T., Bárdossy, A., Zehe, E., He, Y., 2008. Comparison of conceptual model
- 746 performance using different representations of spatial variability. *J Hydrol* 356, 106–
- 747 118.
- 748
- 749 Del Giudice, D., Honti, M., Scheidegger, A., Albert, C., Reichert, P., Rieckermann,
- 750 J., 2013. Improving uncertainty estimation in urban hydrological modeling by
- 751 statistically describing bias. *Hydrol Earth Syst Sci Discuss*, 10, 5121–5167.
- 752
- 753 Dotto, C.B.S., Mannina, G., Kleidorfer, M., Vezzaro, L., Henrichs, M., McCarthy,
- 754 D.T., Freni, G., Rauch, W., Deletic, A., 2012. Comparison of different uncertainty
- 755 techniques in urban stormwater quantity and quality modelling. *Water Res* 46, 2545–
- 756 58.
- 757
- 758 Einfalt, T., Denoeux, T., Jacquet, G., 1990. A radar rainfall forecasting method
- 759 designed for hydrological purposes. *J Hydrol* 114, 229-244.
- 760
- 761 Einfalt, T., Arnbjerg-Nielsen, K., Golz, C., Jensen, N., Quirmbach, M., Vaes, G.,
- 762 Vieux, B., 2004. Towards a roadmap for use of radar rainfall data in urban drainage.
- 763 *J Hydrol* 299, 186-202.
- 764

- 765 Emmanuel, I., Andrieu, H., Tabary, P., 2012a. Evaluation of the new French  
766 operational weather radar product for the field of urban hydrology. Atmos Res 103,  
767 20-32.
- 768
- 769 Emmanuel, I., Andrieu, H., Leblois, E., Flahaut, B., 2012b. Temporal and spatial  
770 variability of rainfall at the urban hydrological scale. J Hydrol 430, 162-172.
- 771
- 772 Fuchs, L., Beeneken, T., 2005. Development and implementation of a real-time  
773 control strategy for the sewer system of the city of Vienna. Water Sci Technol 52,  
774 187-194.
- 775
- 776 Giebel, G., Brownsword, R., Karioniotakis, G., Denhard, M., Draxl, C., 2011. The  
777 State-Of-The-Art in Short-Term Prediction of Wind Power - A Literature Overview.  
778 FP6 ANEMOS.plus deliverable.
- 779
- 780 Gill, R. S., Overgaard, S., Bøvith, T., 2006. The Danish weather radar network, 4<sup>th</sup>  
781 European Conference on Radar in Meteorology and Hydrology, 18-22 September,  
782 2006, Barcelona, Spain.
- 783
- 784 Gjertsen, U., Salek, M., Michelson, D.B., 2003. Gauge-adjustment of radar-based  
785 precipitation estimates—a review. COST-717 Work. Doc. WDD 2.200310.
- 786

- 787 Gneiting, T., Raftery, A.E., Westveld III, A.H., Goldman, T., 2005. Calibrated  
788 probabilistic forecasting using ensemble model output statistics and minimum CRPS  
789 estimation. Mon Weather Rev 133, 1098-1118.
- 790
- 791 Gneiting, T., 2007. Strictly proper scoring rules, prediction, and estimation. J Am  
792 Stat Assoc 102, 359-378.
- 793
- 794 Goudenhoofdt, E., Delobbe, L., 2009. Evaluation of radar-gauge merging methods  
795 for quantitative precipitation estimates. Hydrol Earth Syst Sci 13, 195-203.
- 796
- 797 Gourley, J.J., Vieux, B.E., 2005. A Method for Evaluating the Accuracy of  
798 Quantitative Precipitation Estimates from a Hydrologic Modeling Perspective. J  
799 Hydrometeorol 6, 115–133.
- 800
- 801 Grum, M., Thornberg, D., Christensen, M.L., Shididi, S.A., Thirsing, C., 2011. Full-  
802 Scale Real Time Control Demonstration Project in Copenhagen's Largest Urban  
803 Drainage Catchments, Proceedings of the 12th International Conference on Urban  
804 Drainage, 11-16 Sep, Porto Alegre / Brazil.
- 805
- 806 Jin, X., Xu, C.Y., Zhang, Q., Singh, V., 2010. Parameter and modeling uncertainty  
807 simulated by GLUE and a formal Bayesian method for a conceptual hydrological  
808 model. J Hydrol 383, 147-155.
- 809

- 810 Jørgensen, H.K., Rosenørn, S., Madsen, H., Mikkelsen, P.S., 1998. Quality control of  
811 rain data used for urban runoff systems. Water Sci Technol 37(11), 113-120.
- 812
- 813 Kloeden, P.E., Platen, E., 1999. Numerical Solution of Stochastic Differential  
814 Equations. third ed., Springer, Berlin-Heidelberg-New York.
- 815
- 816 Kraemer, S., Grum, M., Verworn, H.R., Redder, A., 2005. Runoff modelling using  
817 radar data and flow measurements in a stochastic state space approach. Water Sci  
818 Technol 52, 1-8.
- 819
- 820 Kristensen N.R., Madsen H., 2003. Continuous time stochastic modelling CTSM -  
821 mathematics guide. Technical report, Technical University of Denmark, URL  
822 <http://CTSM.info>.
- 823
- 824 Kristensen, N.R., Madsen, H., Jørgensen, S.B., 2004. Parameter estimation in  
825 stochastic grey-box models. Automatica 40, 225–237.
- 826
- 827 Kuczera, G., Kavetski, D., Franks, S., Thyre, M., 2006. Towards a Bayesian total  
828 error analysis of conceptual rainfall-runoff models: Characterising model error using  
829 storm-dependent parameters, J Hydrol 331, 161-177.
- 830
- 831 Löwe, R., Mikkelsen, P., Madsen, H., 2014. Stochastic rainfall-runoff forecasting:  
832 parameter estimation, multi-step prediction, and evaluation of overflow risk. Stoch  
833 Environ Res Risk Assess 28, 505-516, doi 10.1007/s00477-013-0768-0.

- 834
- 835 Nielsen, H.A., Madsen, H., 2006. Modelling the heat consumption in district heating  
836 systems using a grey-box approach. Energy Build 38, 63–71.
- 837
- 838 Nielsen, J.E., Thorndahl, S., Rasmussen, M.R., 2014. A numerical method to  
839 generate high temporal resolution precipitation time series by combining weather  
840 radar measurements with a nowcast model. Atmos Res 138, 1–12, doi  
841 10.1016/j.atmosres.2013.10.015.
- 842
- 843 Pleau, M., Pelletier, G., Lavalle, P., Bonin, R., 2001. Global predictive real-time  
844 control of Quebec Urban Communitys westerly sewer network. Water Sci Technol  
845 43, 123-130.
- 846
- 847 Pleau, M., Colas, H., Lavallée, P., Pelletier, G., Bonin, R., 2005. Global optimal real-  
848 time control of the Quebec urban drainage system. Environ Modell Softw 20, 401-  
849 413.
- 850
- 851 Puig, V., Cembrano, G.G., Romera, J., Quevedo, J., Aznar, B., Ramón, G., Cabot, J.,  
852 2009. Predictive optimal control of sewer networks using CORAL tool: application  
853 to Riera Blanca catchment in Barcelona. Water Sci Technol 60, 869-878.
- 854
- 855 Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D., DMIP Participants,  
856 2004. Overall distributed model intercomparison project results. J Hydrol 298, 27-60.
- 857

- 858 Renard, B., Kavetski, D., Kuczera, G., Thyer, M., Franks, S., 2010: Understanding  
859 predictive uncertainty in hydrologic modeling: The challenge of identifying input  
860 and structural errors, Water Resour Res 46, W05521.
- 861
- 862 Schellart, A., Shepherd, W., Saul, A., 2011. Influence of rainfall estimation error and  
863 spatial variability on sewer flow prediction at a small urban scale. Adv Water Resour  
864 45, 65-75.
- 865
- 866 Schilling, W., 1984. Effect of spatial rainfall distribution on sewer flows. Water Sci  
867 Technol 16, 177-188.
- 868
- 869 Schilling, W., Fuchs, L., 1986. Errors in stormwater modeling-a quantitative  
870 assessment. J Hydraul Eng 112, 111-123.
- 871
- 872 Schilling, W., 1991. Rainfall data for urban hydrology: what do we need? Atmos Res  
873 27, 5-21.
- 874
- 875 Seggelke, K., Löwe, R., Beenenken, T., Fuchs, L., 2013. Implementation of an  
876 integrated real-time control system of sewer system and WWTP in the city of  
877 Wilhelmshaven. Urban Water J 10, 330-341.
- 878
- 879 Sharma, A.K., Guildal, T., Thomsen, H.A.R., Mikkelsen, P.S., Jacobsen, B.N., 2013.  
880 Aeration tank settling and real time control as a tool to improve the hydraulic

881 capacity and treatment efficiency during wet weather: results from 7 years' full-scale  
882 operational data. Water Sci Technol 67, 2169-2176.

883

884 Thordarson, F.O., Breinholt, A., Møller, J.K., Mikkelsen, P.S., Grum, M., Madsen,  
885 H., 2012. Uncertainty assessment of flow predictions in sewer systems using grey  
886 box models and skill score criterion. Stoch Env Res Risk A 26, 1151-1162.

887

888 Thorndahl, S., Beven, K.J., Jensen, J.B., Schaarup-Jensen, K., 2008. Event based  
889 uncertainty assessment in urban drainage modeling, applying the GLUE  
890 methodology, J Hydrol 357, 421-437.

891

892 Thorndahl, S., Rasmussen, M. R., Grum, M., Neve, S. L., 2009. Radar Based Flow  
893 and Water Level Forecasting in Sewer Systems: a Danish case study. 8th  
894 International Workshop on Precipitation in Urban Areas, 10-13 December, 2009, St.  
895 Moritz, Switzerland.

896

897 Thorndahl, S., Rasmussen, M.R., Neve, S., Poulsen, T.S., Grum, M., 2010.  
898 Vejrradarbaseret styring af spildevandsanlæg (in Danish). DCE Technical Report No.  
899 95, Aalborg Universitet. Institut for Byggeri og Anlæg, ISSN 1901-726X.

900

901 Thorndahl, S., Rasmussen, M.R., 2013. Short-term forecasting of urban storm water  
902 runoff in real-time using extrapolated radar rainfall data. J Hydroinf 15, 214-226.

903

- 904 Thorndahl, S., Nielsen, J.E., Rasmussen, M.R., 2014. Bias adjustment and advection  
905 interpolation of long-term high resolution radar rainfall series. *J Hydrol* 508, 214-  
906 226.
- 907
- 908 Tolson, B. A., Shoemaker, C. A., 2007. Dynamically dimensioned search algorithm  
909 for computationally efficient watershed model calibration, *Water Resour Res* 43,  
910 W01413.
- 911
- 912 Vezzaro, L., Grum, M., 2012. A generalized Dynamic Overflow Risk Assessment  
913 (DORA) for urban drainage RTC. 9th International Conference on Urban Drainage  
914 Modelling, 3-7 September, 2012, Belgrade, Serbia.
- 915
- 916 Vieux, B.E., Bedient, P.B., 2004. Assessing urban hydrologic prediction accuracy  
917 through event reconstruction. *J Hydrol* 299, 217–236.
- 918
- 919 Villarini, G., Smith, J.A., Lynn Baeck, M., Sturdevant-Rees, P., Krajewski, W.F.,  
920 2010. Radar analyses of extreme rainfall and flooding in urban drainage basins. *J*  
921 *Hydrol* 381, 266-286.
- 922
- 923 Wang, L., Ochoa, S., Simões, N., Onof, C., Maksimović, Č., 2013. Radar-raingauge  
924 data combination techniques: a revision and analysis of their suitability for urban  
925 hydrology. *Water Sci Technol* 68, 737-747.
- 926

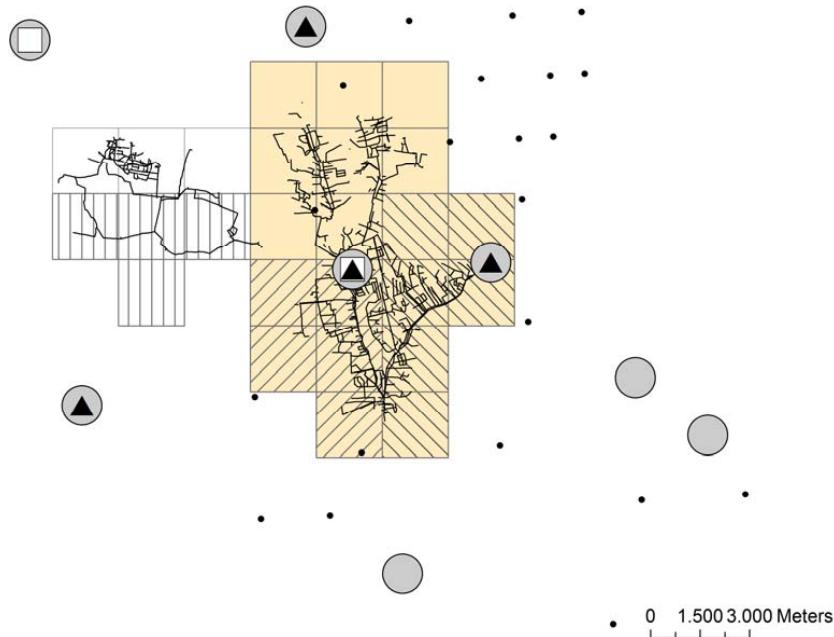
- 927 Wolfs, V., Villazon, M.F., Willems, P., 2013. Development of a semi-automated  
928 model identification and calibration tool for conceptual modelling of sewer systems.  
929 Water Sci. Technol. Sci. Technol. 68, 167–175.
- 930
- 931 Wood, S.J., Jones, D.A., Moore, R.J., 2000. Static and dynamic calibration of radar  
932 data for hydrological use. Hydrol Earth Syst Sci 4, 545–554.
- 933

934

935 **7 FIGURES**

936

937



938

939 Figure 1. Ballerup (left) and northern Damhusåen (right) catchments with C-band  
940 radar pixels (2x2km), location of rain gauges shown as dots (large grey circles – used  
941 in radar adjustment, white rectangles – used as input to Ballerup model, black  
942 triangles – used as input to Damhusåen model, small black dots – other gauges).  
943 Different radar pixel shadings correspond to different subcatchments (c.f. section  
944 2.5.2).

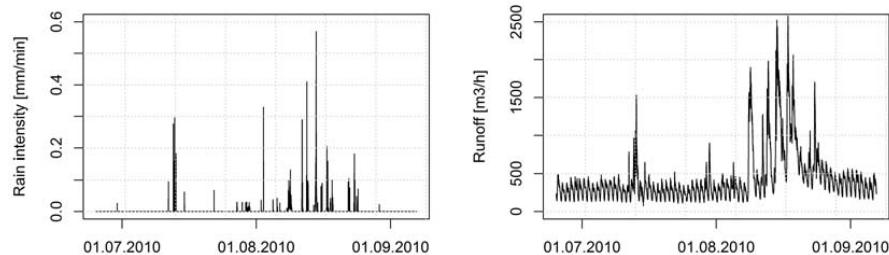
945

946

947

948

949



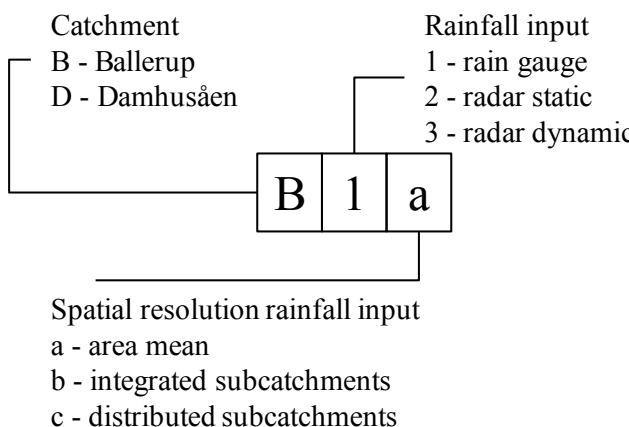
950

951

952 Figure 2. Areal mean of rain gauge observations and flow measurements for the  
953 Ballerup catchment in the estimation period.

954

955



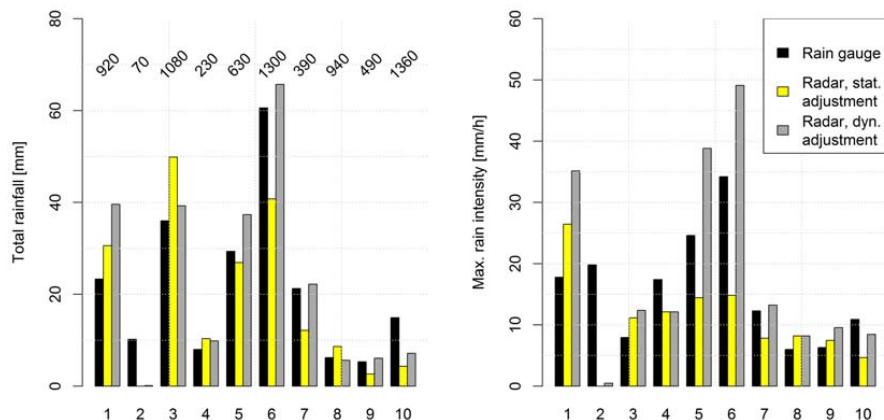
956

957 Figure 3. Simulation run identifiers depending on considered catchments, rainfall  
958 input and spatial resolution.

959

960

961



962

963

964 Figure 4. Rain event depths (left) and maximum intensity (right) derived for mean  
965 areal rainfall with different rainfall measurements in the Ballerup catchment. Left  
966 plot includes label of duration of rain event (in min).

967

968

969

970

971

972

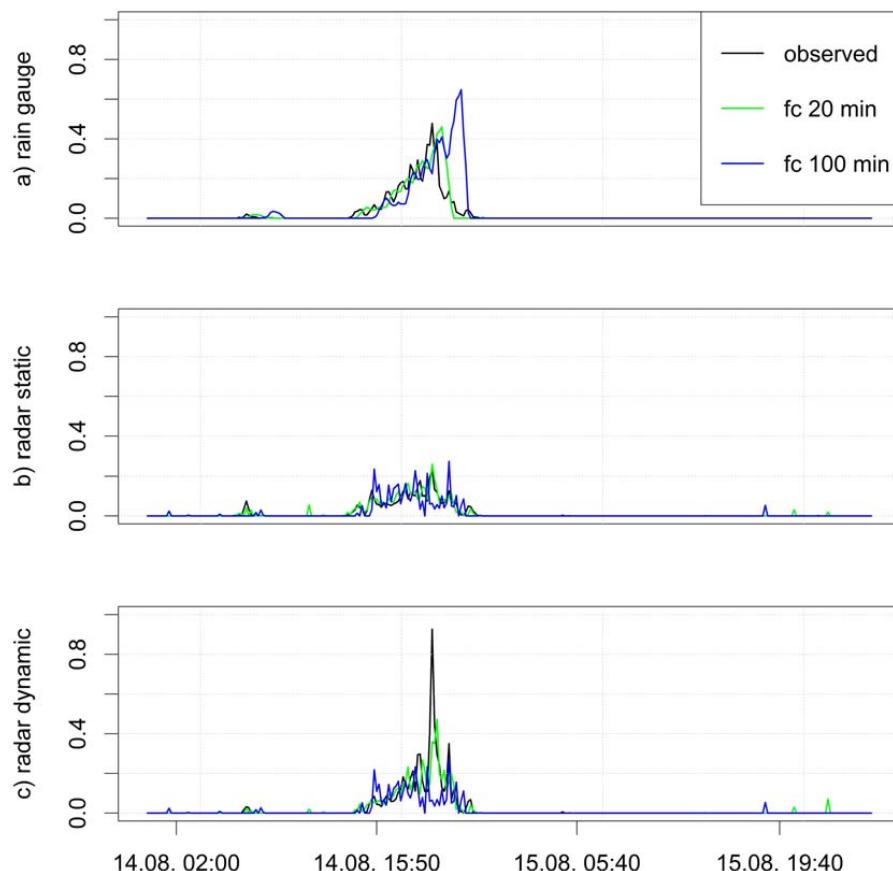
973

974

975

976

977



978

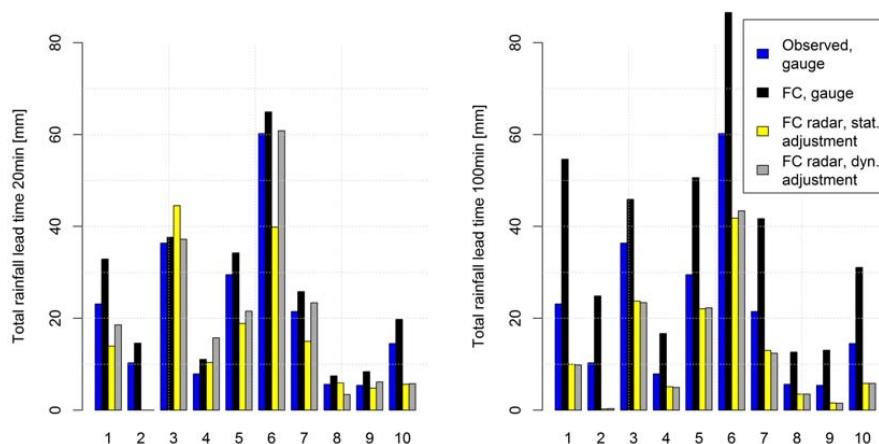
979

980 Figure 5. Sample rainfall event in the Ballerup catchment. Part a: rain gauge  
 981 observations (black, as in model B1a) and rainfall forecasts with lead times of 20 min  
 982 (green) and 100 min (blue), Part b: statically adjusted radar rainfall observations and  
 983 forecasts (B2a), Part c: dynamically adjusted radar rainfall observations and forecasts  
 984 (B3a).

985

986

987



988

989

990 Figure 6. Total forecasted (FC) rainfall amount for the Ballerup catchment for lead  
991 times of 20 (left) and 100min (right) for the considered rain events, together with  
992 rainfall amount observed by rain gauges.

993

994

995

996

997

998

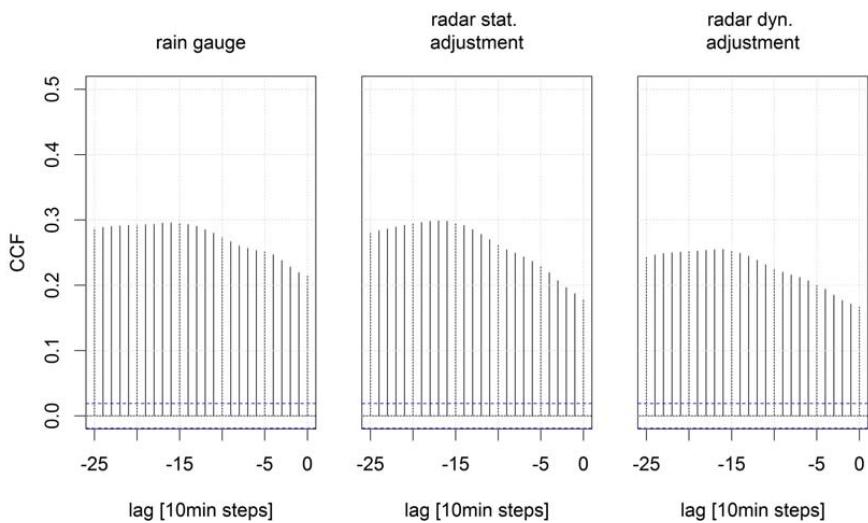
999

1000

1001

1002

1003



1004

lag [10min steps]

lag [10min steps]

lag [10min steps]

1005

1006 Figure 7. Cross correlation (CCF) between runoff and catchment averaged rainfall  
1007 observations in the Ballerup catchment. Rainfall observations are lagged in 10min  
1008 steps to the runoff observations.

1009

1010

1011

1012

1013

1014

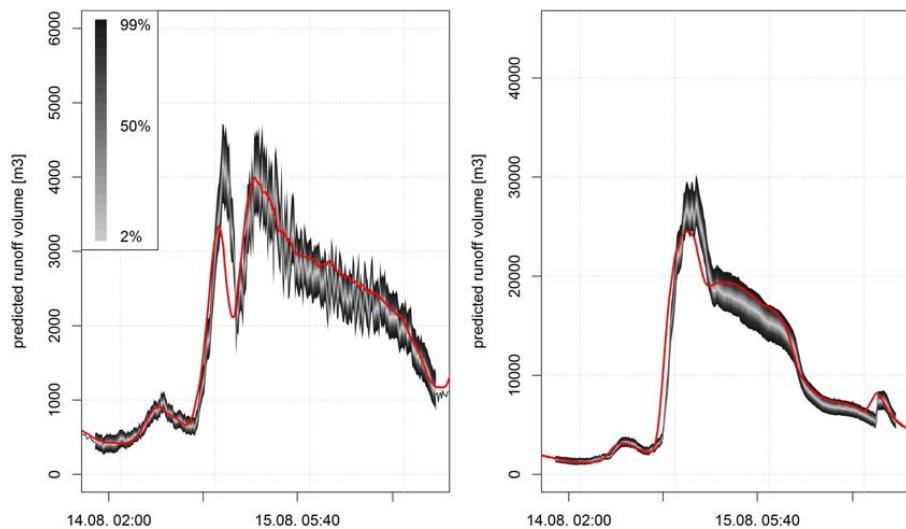
1015

1016

1017

1018

1019



1020

1021

1022 Figure 8. 10-step forecasts of runoff volume for event 6 in Ballerup (left, model B2a)  
1023 and Damhusåen (right, model D2a) catchments together with observation (red). The  
1024 shading corresponds to different prediction intervals with coverage rates from 2% to  
1025 98%.

1026

1027

1028

1029

1030

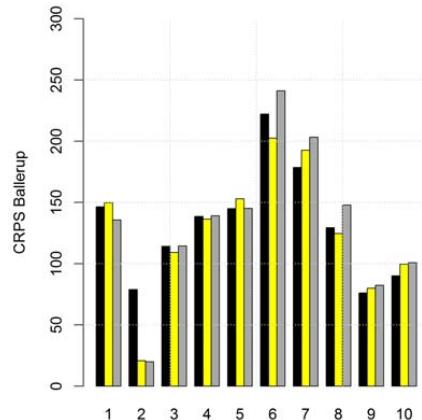
1031

1032

1033

1034

1035



1036

1037

1038 Figure 9. Quality of probabilistic runoff forecasts for 100 min horizon (10 step)

1039 expressed as CRPS for different rain events and inputs in Ballerup (left) and

1040 Damhusåen (right) catchments.

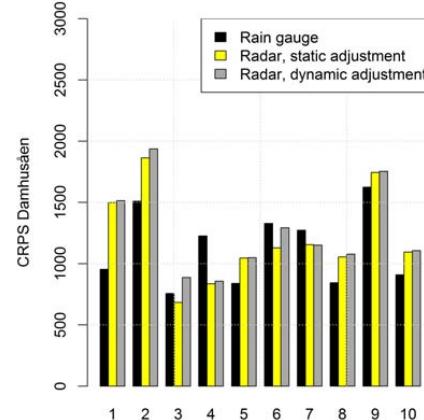
1041

1042

1043

1044

1045





PAPER C

# Stochastic rainfall-runoff forecasting: parameter estimation, multi-step prediction, and evaluation of overflow risk

---

**Authors:**

Roland Löwe, Peter Steen Mikkelsen, Henrik Madsen

**Published in:**

*Stochastic Environmental Research and Risk Assessment*, 28: 505–516, 2014.



# Stochastic rainfall-runoff forecasting: parameter estimation, multi-step prediction, and evaluation of overflow risk

Roland Löwe<sup>1</sup>, Peter Steen Mikkelsen<sup>2</sup>, Henrik Madsen<sup>1</sup>

## Abstract

Probabilistic runoff forecasts generated by stochastic greybox models can be notably useful for the improvement of the decision-making process in real-time control setups for urban drainage systems because the prediction risk relationships in these systems are often highly nonlinear. To date, research has primarily focused on one-step-ahead flow predictions for identifying, estimating, and evaluating greybox models. For control purposes, however, stochastic predictions are required for longer forecast horizons and for the prediction of runoff volumes, rather than flows. This article therefore analyzes the quality of multistep ahead forecasts of runoff volume and considers new estimation methods based on scoring rules for k-step-ahead predictions. The study shows that the score-based methods are, in principle, suitable for the estimation of model parameters and can therefore help the identification of models for cases with noisy in-sewer observations. For the prediction of the overflow risk, no improvement was demonstrated through the application of stochastic forecasts instead of point predictions, although this result is thought to be caused by the notably simplified setup used in this analysis. In conclusion, further research must focus on the development of model structures that allow the proper separation of dry and wet weather uncertainties and simulate runoff uncertainties depending on the rainfall input.

---

<sup>1</sup> Department of Applied Mathematics and Computer Science, Technical University of Denmark, Matematiktorvet, B322, DK-2800 Kgs. Lyngby, Denmark

<sup>2</sup> Department of Environmental Engineering, Technical University of Denmark, Miljøvej, B115, DK-2800 Kgs. Lyngby, Denmark

**Stochastic Environmental Research And Risk Assessment manuscript No.**  
(will be inserted by the editor)

---

## Stochastic rainfall-runoff forecasting: parameter estimation, multi-step prediction, and evaluation of overflow risk

Roland Löwe · Peter Steen Mikkelsen · Henrik  
Madsen

Received: 2013/05/03 / Accepted: date

**Abstract** Probabilistic runoff forecasts generated by stochastic greybox models can be notably useful for the improvement of the decision-making process in real-time control setups for urban drainage systems because the prediction risk relationships in these systems are often highly nonlinear. To date, research has primarily focused on one-step-ahead flow predictions for identifying, estimating, and evaluating greybox models. For control purposes, however, stochastic predictions are required for longer forecast horizons and for the prediction of runoff volumes, rather than flows. This article therefore analyzes the quality of multistep ahead forecasts of runoff volume and considers new estimation methods based on scoring rules for k-step-ahead predictions. The study shows that the score-based methods are, in principle, suitable for the estimation of model parameters and can therefore help the identification of models for cases with noisy in-sewer observations. For the prediction of the overflow risk, no improvement was demonstrated through the application of stochastic forecasts instead of point predictions, although this result is thought to be caused by the notably simplified setup used in this analysis. In conclusion, further research must focus on the development of model structures that allow the proper separation of dry and wet weather uncertainties and simulate runoff uncertainties depending on the rainfall input.

---

Roland Löwe · Henrik Madsen  
Technical University of Denmark (DTU), Dept. of Applied Mathematics and Computer Science, Matematiktorvet B322, DK-2800 Kgs. Lyngby, Denmark  
Tel.: +45-4525 3399  
Fax: +45-4588 2673  
E-mail: rolo@dtu.dk

Peter Steen Mikkelsen  
Technical University of Denmark (DTU), Dept. of Environmental Engineering, Miljøvej B113, DK-2800 Kgs. Lyngby, Denmark

---

**Keywords** stochastic greybox model · skill score · real-time control · urban drainage · multistep prediction · online forecasting

---

**1 Introduction**

2 Real-time control (RTC) often provides a method to efficiently operate sewer systems  
3 and reduce spills of sewage into lakes, rivers, and oceans (combined sewer overflows,  
4 CSOs). This reduces the need to build storage volumes in the sewer system, which  
5 makes the method economically attractive. A multitude of control systems are in op-  
6 eration today. The types of setup range from rule-based strategies that are determined  
7 offline (Fuchs and Beenken, 2005; Seggelke et al, 2012), to online optimizations  
8 of storage volumes (Pabst et al, 2011) and model predictive control (MPC) (Schütze  
9 et al, 2004; Puig et al, 2009).

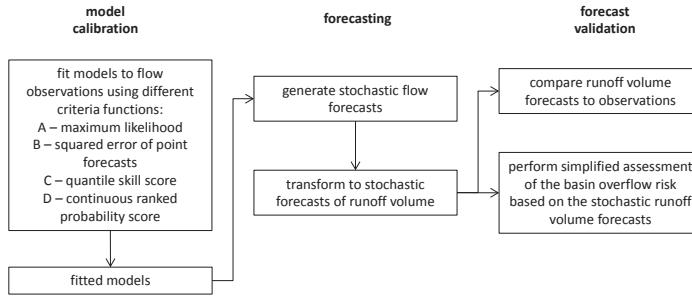
10 It is commonly expected that the combination of forecast information and global  
11 optimization as applied in MPC will yield the best control results. This is obscured  
12 by the complex side constraints that result from the operational requirements in the  
13 sewer system and by insufficient forecast quality. Recently, a new control setup was  
14 introduced in the Copenhagen area to minimize the total overflow risk from a num-  
15 ber of storage basins in the catchment through the dynamic adjustment of the basin  
16 outflows and the pumping capacities. The decisions in this algorithm for the global  
17 control of the system are based on forecasted runoff volumes for the catchment of  
18 each basin (Dynamic Overflow Risk Assessment, DORA) (Vezzaro and Grum, 2012;  
19 Grum et al, 2011).

20 Forecasts in such a setup need to be available at varying horizon lengths which  
21 makes models that provide multistep predictions attractive. Furthermore, the fore-  
22 cast uncertainties need to be considered in the decision-making process because  
23 prediction-risk relationships in urban drainage systems are typically nonlinear (Vez-  
24 zaro and Grum, 2012). However, no tools for the modeling of predictive uncertainties  
25 in an online setting are available. At present, the very simplified assumption that fore-  
26 cast uncertainties can be described by a Gamma distribution with shape parameters  
27 that depend on the predicted runoff volume is used.

28 Stochastic greybox models fulfill both requirements because these provide predic-  
29 tive uncertainties at varying horizons. For our purposes, stochastic greybox models  
30 are termed simplified models with physically interpretable parameters that provide a  
31 quantification of the model uncertainties. Several authors have demonstrated the gen-  
32 eral applicability of this class of models to urban drainage problems. Carstensen et al  
33 (1998) applied ARMAX models to simulate the inflow to a wastewater treatment  
34 plant. Bechmann et al (2000) simulated the first flush and later the pollutant loads  
35 (Bechmann et al, 1999) using stochastic differential equations. Breinholt et al (2011)  
36 investigated model setups for flow predictions based on linear reservoir cascades us-  
37 ing stochastic differential equations and took the initial steps required to quantify  
38 the predictive uncertainty. Furthermore, Thordarson et al (2012) investigated multi-  
39 step flow predictions for urban drainage systems and evaluated these using skill score  
40 criteria.

41 Previous works on stochastic forecasting of runoff in urban drainage systems have  
42 focused on the prediction of flows for one or several prediction horizons. However,  
43 the decision-making process in real-time control is typically based on the predicted  
44 runoff volume. The quality of the probabilistic multistep volume predictions obtained  
45 from the stochastic greybox models has not yet been evaluated. Furthermore, it is

- 46 unclear whether the currently used parameter estimation technique, which is based  
 47 on the maximization of the likelihood for one-step ahead predictions, also yields a  
 48 good model for multistep-ahead forecasts.



**Fig. 1** Flow scheme for comparing model estimation approaches and evaluation of multistep forecasts of runoff volume

49 Therefore, following the scheme shown in Figure 1, the stochastic multistep pre-  
 50 dictions of the runoff volume are generated using greybox models. New estimation  
 51 approaches for stochastic greybox models that focus on multistep predictions were  
 52 suggested, and the forecasts from the resulting models were compared.

53 A simplified assessment of the ability of the models to predict the overflow risk  
 54 was subsequently performed to evaluate the possible effects of the different forecasts  
 55 on real-time control.

56 **2 Methods**

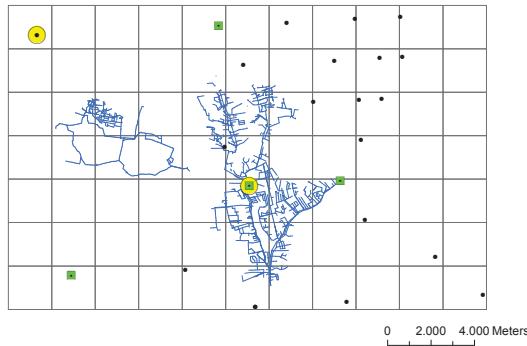
57 **2.1 Data and Catchments**

58 We consider two catchments in the Copenhagen area. The Ballerup catchment has  
 59 a total area of approximately 1,300 ha and is mainly laid out as a separate system,  
 60 although it does have a small combined section. The runoff in this area is also strongly  
 61 influenced by rainfall-dependent infiltration.

62 The Damhusåen catchment is located close to the Ballerup catchment but drains  
 63 to a different treatment plant. We consider the northern part of this catchment, which  
 64 has a total area of approximately 3,000 ha. The catchment is laid out as a combined  
 65 sewer system, and a multitude of CSOs are located in the area. Flow measurements  
 66 are available for both catchments at a 5-min resolution.

67 Numerous online rain gauge measurements are available from the Danish wastewater  
 68 committee's (SVK) network in the considered catchments (Jørgensen et al, 1998).  
 69 The gauges marked in Figure 2 were used as the input for the runoff forecasting models  
 70 for the two different catchments. These are the same gauges used in previous studies  
 71 on the Ballerup catchment (Breinholt et al, 2011; Thordarson et al, 2012) and for  
 72 radar rainfall calibration and real-time control in the Copenhagen area (Grum et al,  
 73 2011). These gauge measurements are also available with a temporal resolution of 5  
 74 minutes.

75 We have selected a 3-month measurement period from 25/06/2010 to 29/09/2010  
 76 for this study. The period contains several summer storms that can be considered  
 77 relevant for control applications in urban drainage systems. A modeling time step  
 78 of 10 min was adopted and corresponds to the temporal resolution used in previous  
 79 studies (Löwe et al, 2012a,b). The flow and rain gauge data were averaged to match  
 80 this time step.



**Fig. 2** Ballerup (left) and Damhusåen (right) catchments with online rain-gauge measurements in the area (black dots) and the gauges used as the input data for the Ballerup (circle) and the Damhusåen (rectangle) catchments

---

 81 2.2 Stochastic Greybox Models for Runoff Prediction

82 We predicted the runoff at the catchment outlets using stochastic greybox models,  
 83 which are briefly described in this section. The physical part of the models is based  
 84 on lumped reservoir approaches that transform the rainfall input into the flow output.  
 85 The principal model setup is described by Breinholt et al (2011). In this work, we  
 86 applied a simple two-reservoir cascade to both catchments. In a state space formula-  
 87 tion, we used two coupled Itô stochastic differential equations, which together form  
 88 the following so-called system or state equations

$$89 \quad d \begin{bmatrix} S_{1,t} \\ S_{2,t} \end{bmatrix} = \underbrace{\left[ A \cdot P_t + a_0 - \frac{1}{K} S_{1,t} \right] dt}_{\text{Drift term}} + \underbrace{\begin{bmatrix} \sigma_1 S_{1,t}^{\gamma_1} & 0 \\ 0 & \sigma_2 S_{2,t}^{\gamma_2} \end{bmatrix} d\omega_t}_{\text{Diffusion term}}$$

90

91 and the observation equation

$$92 \quad Y_k = \log(Q_k) = \log\left(\frac{1}{K} S_{2,k} + D_k\right) + e_k \quad (2)$$

93

94 where

$$95 \quad D_k = \sum_i^2 \left( s_i \sin \frac{i 2 \pi k}{24h} + c_i \cos \frac{i 2 \pi k}{24h} \right) \quad (3)$$

96

97  $S_1$  and  $S_2$  correspond to the states of the system, i.e. virtual storage fillings,  $A$  is the  
 98 sealed area in the catchment,  $a_0$  refers to the mean dry weather flow at the catch-  
 99 ment outlet, and  $K$  corresponds to the travel time constant. The rainfall input  $P_t$  was  
 100 determined as the mean area rainfall by averaging the rain-gauge measurements con-  
 101 sidered for every catchment. In the diffusion term, the variance of the state values was  
 102 scaled depending on the state value itself because the model predictions are more un-  
 103 certain in wet weather. The scaling was exponential to avoid extreme increases in the  
 104 variance in situations with high runoff.

105 In the observation equation,  $Q_k$  corresponds to the flow observation at time step  $k$   
 106 in discrete time, and  $D_k$  describes the variation of the dry weather flows as a harmonic  
 107 function with parameters  $s_1$ ,  $s_2$ ,  $c_1$ , and  $c_2$ . A log transform was used to avoid negative  
 108 flow predictions. Please refer to Breinholt et al (2011) for a detailed derivation and  
 109 description of the model structure.

110 The open source software CTSM (Kristensen and Madsen, 2003) was used for the  
 111 parameter estimation and the forecast generation. State-dependent diffusion terms,  
 112 such as those in equation (1), cannot be modeled in this setup (Vestergaard, 1998).  
 113 Therefore, a Lamperti transform was applied to the system equations (1), as described  
 114 by Breinholt et al (2011).

115 The multistep flow predictions were generated using the extended Kalman fil-  
 116 ter with updating. This setup provides a log-transformed flow prediction  $\hat{Y}_{i+k|i}$  with  
 117 variance  $R_{i+k|i}$  that is assumed to be normally distributed.

---

<sup>118</sup> 2.3 Parameter Estimation for Stochastic Runoff Prediction Models

<sup>119</sup> The purpose of the runoff prediction models considered is to describe the expected  
<sup>120</sup> runoff volume over a horizon of variable extent, which is defined by the control  
<sup>121</sup> scheme. When estimating the model parameters from historical data, we need to de-  
<sup>122</sup> sign the objective function such that the resulting model is actually optimal for the  
<sup>123</sup> generation of predictions for different horizons. Below, we introduce possible objec-  
<sup>124</sup> tive functions.

<sup>125</sup> All of the suggested objective functions focus on flow values rather than runoff  
<sup>126</sup> or even overflow volumes because the conversion from stochastic flow to runoff pre-  
<sup>127</sup> dictions is computationally demanding. In addition, the models should be estimated  
<sup>128</sup> to correctly describe the physical behavior of the system and thus reduce the risk  
<sup>129</sup> of overfitting (Weijs et al., 2010). The physical system behavior is captured when  
<sup>130</sup> focusing on flow values during the parameter estimation, whereas focusing on over-  
<sup>131</sup> flow volumes would likely introduce a partial loss of the information provided by the  
<sup>132</sup> measurements.

<sup>133</sup> Parameter estimation was performed automatically in all cases using a genetic al-  
<sup>134</sup> gorithm based on the concepts described by Whitley (1994), Spall (2003) and Hallam  
<sup>135</sup> (2010).

<sup>136</sup> 2.3.1 Maximum Likelihood Estimation (Model A)

<sup>137</sup> The most common approach for the estimation of parameters in stochastic greybox  
<sup>138</sup> models is to maximize the likelihood function for a given series of measurements  
<sup>139</sup> (Kristensen et al., 2004; Breinholdt et al., 2011). The computation of the likelihood  
<sup>140</sup> function is based on the computation of the one-step prediction errors or innovations  
<sup>141</sup> under the assumption that the one-step-ahead conditional densities are Gaussian:

$$\varepsilon_i = Y_i - \hat{Y}_{i|i-1} \quad (4)$$

<sup>144</sup> This approach may be difficult in the context of the estimation of models for multi-  
<sup>145</sup> step predictions in the urban runoff setting. The parameters found may not to be  
<sup>146</sup> optimal for multi-step predictions because these are based on the one-step prediction  
<sup>147</sup> errors. Furthermore, there is a clear risk of overfitting if the physical part of the model  
<sup>148</sup> fails to completely capture the system behavior. The one-step predictions are strongly  
<sup>149</sup> influenced by the updating of the states in the extended Kalman filter, and we may  
<sup>150</sup> identify parameters that are optimal for this updating but do not actually match the  
<sup>151</sup> physical system behavior, which would result in bad forecasts and simulations.

<sup>152</sup> One may argue that, in these situations, the modeler should attempt to improve  
<sup>153</sup> the physical part of the model and make it more suitable to the actual behavior of the  
<sup>154</sup> catchment. However, in practical applications, we will often face the situation that a  
<sup>155</sup> simple model will be sufficient for the forecasting purpose; moreover, the tailoring  
<sup>156</sup> of a model to each new catchment may be too time-consuming. This also indicates a  
<sup>157</sup> need for more robust estimation methods that focus on the forecasting purpose.

---

158 2.3.2 *Minimizing the Error of the Predicted Runoff Volumes (Model B)*

159 The fitting of forecast models in hydrology is typically performed by minimizing the  
160 forecast error variance (see, e.g., Nash and Sutcliffe (1970)). We suggest an objec-  
161 tive function based on the sum of the squared errors between the predicted and the  
162 observed runoff volumes over the prediction horizon:

$$\text{163} \quad S(\theta) = \sum_{i=1}^N \left( \sum_{j=1}^k Q_{i+j} - \sum_{j=1}^k \hat{Q}_{i+j|i}(\theta) \right)^2 \Delta t \quad (5)$$

164

165 At every time step  $i$  of length  $\Delta t$ , a  $k$ -step ahead flow prediction  $\hat{Q}$  is generated.  
166 The flow values are integrated to a runoff volume over the prediction horizon and  
167 compared to the observations  $Q$ . The minimization of the sum of these volume differ-  
168 ences for all  $N$  time steps gives an objective function for the estimation of the model  
169 parameter set  $\theta$ .

170 This objective function optimizes the model to give a good point forecast of the  
171 expected runoff volumes over the maximum prediction horizon of  $k$  steps (e.g.  $k=10$   
172 steps). Implicitly, we assume that we also obtain good predictions for shorter hori-  
173 zons.

174 2.3.3 *Estimation Based on the Interval Score (Model C)*

175 Minimizing the squared error of the predicted runoff volumes tunes the forecast mod-  
176 els to give good point predictions of the runoff volume for multistep prediction hori-  
177 zons. The quality of the forecast uncertainties is not evaluated in this criterion. How-  
178 ever, the forecast objective in the described setup is to obtain a probabilistic descrip-  
179 tion of the predicted runoffs. The predictive distribution should be as narrow (sharp)  
180 as possible and at the same time match the observations (be calibrated or reliable).

181 To account for the quality of the probabilistic predictions, we can modify the  
182 criterion developed in section 2.3.2. Assuming normality, we compute a  $(1 - \beta) \cdot$   
183  $100\% = 95\%$  prediction interval for forecast horizon  $j$  for the log-transformed flow  
184 values as

$$\begin{aligned} \text{185} \quad u\hat{Y}_{i+j|i} &= \hat{Y}_{i+j|i} + 1.96 \cdot \sigma_{\hat{Y}_{i+j|i}} \\ \text{186} \quad l\hat{Y}_{i+j|i} &= \hat{Y}_{i+j|i} - 1.96 \cdot \sigma_{\hat{Y}_{i+j|i}} \end{aligned} \quad (6)$$

187

188 where  $\sigma_{\hat{Y}_{i+j|i}}$  is the standard deviation of the  $j$ -step predictions.

189 The quality of this prediction interval can be evaluated using a number of meth-  
190 ods, e.g. the interval score described by Gneiting and Raftery (2007), which was  
191 applied to stochastic flow forecasts in urban drainage systems by Thordarson et al  
192 (2012). The score for the  $j$ -step prediction generated at time step  $i$  is

$$\begin{aligned} \text{193} \quad SC_{i,j}^\beta &= u\hat{Y}_{i+j|i} - l\hat{Y}_{i+j|i} + \frac{2}{\beta} (\hat{Y}_{i+j|i} - Y_{i+j}) \mathbb{1}\{Y_{i+j} < l\hat{Y}_{i+j|i}\} \\ \text{194} \quad &\quad + \frac{2}{\beta} (Y_{i+j} - u\hat{Y}_{i+j|i}) \mathbb{1}\{Y_{i+j} > u\hat{Y}_{i+j|i}\} \end{aligned} \quad (7)$$

195

<sup>196</sup> In equation 8, a reasonable scoring rule based on equation 7 is suggested and accounts  
<sup>197</sup> for several forecast horizons. More weight is placed on the flow forecasts for shorter  
<sup>198</sup> horizons, which have a stronger influence on forecasts of runoff volume because the  
<sup>199</sup> latter are generated as an integral over the flow forecasts for different horizons.

$$\text{SC}_i^\beta = \frac{1}{\sum_{j=1}^k (k-j+1)} \left( \sum_{j=1}^k (k-j+1) \cdot \text{SC}_{i,j} \right) \quad (8)$$

<sup>200</sup> By averaging over all  $N$  time steps, we obtain the objective function for parameter  
<sup>201</sup> estimation in model C.

$$S(\theta) = \frac{1}{N} \sum_{i=1}^N \text{SC}_i^\beta \quad (9)$$

#### <sup>206</sup> 2.3.4 Estimation Based on Continuous Ranked Probability Score (Model D)

<sup>207</sup> The interval score criterion described above was previously applied to flow forecasts  
<sup>208</sup> in urban drainage systems, but focuses on a 95% prediction interval, i.e., only the  
<sup>209</sup> tails of the predictive distribution. This may lead to a dislocation of the center of  
<sup>210</sup> the predicted flow distribution. The continuous ranked probability score (CRPS) is  
<sup>211</sup> a measure of the fit of the overall distribution; therefore, we introduce this score  
<sup>212</sup> here as the last objective function for parameter estimation in the stochastic runoff  
<sup>213</sup> forecasting models. Gneiting et al (2005) suggested the use of the CRPS in the fitting  
<sup>214</sup> of post-processing models for ensemble predictions and consider it robust toward  
<sup>215</sup> extreme events and outliers. A discussion of the score can be found in the manuscript  
<sup>216</sup> published by Gneiting and Raftery (2007).

<sup>217</sup> We obtained the CRPS for the  $j$ -step flow prediction generated at time step  $i$  as

$$\text{CRPS}_{i,j} = \int_{-\infty}^{\infty} (F(s) - \mathbb{1}\{s > Y_{i+j}\})^2 ds \quad (10)$$

<sup>218</sup> where  $F$  is the cumulative distribution function (CDF) for the (assumed normally  
<sup>219</sup> distributed) log-transformed flow prediction  $\hat{Y}_{i+j|i}$ , and  $Y_{i+j}$  is the corresponding  $(i+j)$ th value in the time series of the observations.  $\mathbb{1}$  denotes the Heaviside function  
<sup>220</sup> and takes the value 0 when  $s < Y_{i+j}$  and 1 otherwise. There exists a closed-form  
<sup>221</sup> solution for equation (10) if the predicted value is normally distributed. However, we  
<sup>222</sup> do not expect to be able to always rely on this assumption in practical situations and  
<sup>223</sup> therefore chose to evaluate the integral numerically.

<sup>224</sup> As in equation (8), we performed a weighting of the CRPS values derived for  
<sup>225</sup> different forecast horizons to obtain an average value for every time step. Ultimately,  
<sup>226</sup> we averaged the values for all of the considered time steps as in equation (9) to obtain  
<sup>227</sup> the value of the objective function. The optimal parameter set is found by minimizing  
<sup>228</sup> this value.

---

<sup>232</sup> 2.4 Generating Stochastic Forecasts of Runoff Volumes

<sup>233</sup> The applied greybox models provide flow forecasts for horizons 1 up to  $k$ . To derive  
<sup>234</sup> probabilistic forecasts of the runoff volume, we used a multivariate sampling ap-  
<sup>235</sup> proach. The correlations between the flow forecasts for different horizons are derived  
<sup>236</sup> from past forecast errors. The following steps were taken.

- <sup>237</sup> – Generate a 10-step forecast at time step  $i$  from the greybox models. We obtained  
<sup>238</sup> a vector of (assumed normal) log-transformed flow predictions  $\hat{Y}_i$  containing the  
<sup>239</sup> forecast values for horizons 1 through 10. The corresponding observations are  
<sup>240</sup> denoted  $Y_i$ .
- <sup>241</sup> – Find the error covariance contribution from this time step (Madsen, 2008):

$$\Sigma_i = (Y_i - \hat{Y}_i) \cdot (Y_i - \hat{Y}_i)^T \quad (11)$$

- <sup>242</sup> – Estimate the overall error covariance structure for time step  $i$  using exponential  
<sup>243</sup> smoothing. This allows for time variation of the considered correlations. We ap-  
<sup>244</sup> plied  $\lambda = 0.99$ .

$$\Sigma_i = \lambda \cdot \Sigma_{i-1} + (1 - \lambda) \cdot V_i \quad (12)$$

- <sup>245</sup> – Scale  $\Sigma_i$  to the predictive variances provided by the greybox model. We obtained  
<sup>246</sup> a covariance structure with variances according to those predicted by the model  
<sup>247</sup> and correlation values estimated from the forecast errors.
- <sup>248</sup> – Create 100,000 multivariate flow samples from the  $N(\hat{Y}_i, \Sigma_{S,i})$  distribution (using  
<sup>249</sup> the R-package MASS (Venables and Ripley, 2002)), each of which represents  
<sup>250</sup> a possible flow scenario for horizons 1 through 10. Integrate each sample into  
<sup>251</sup> runoff volumes and empirically derive the distribution of the runoff volumes.

<sup>252</sup> 2.5 Forecast Evaluation

<sup>253</sup> A set of measures was applied to evaluate the quality of the prediction intervals gen-  
<sup>254</sup> erated by the stochastic greybox models. These are described by Thordarson et al  
<sup>255</sup> (2012) and Jin et al (2010). All of the measures were applied not to flow predictions  
<sup>256</sup> as in Thordarson et al (2012) but to runoff volume predictions for different forecast  
<sup>257</sup> horizons.

- <sup>258</sup> – Reliability

$$REL = \frac{1}{N} \sum_{i=1}^N n_i^\beta \quad (13)$$

<sup>259</sup> where  $N$  is the number of observations,  $\beta$  is the significance level, and  $n_i^\beta$  is an  
<sup>260</sup> indicator variable with value 1 if an observation is not contained in the  $(1 - \beta)\%$   
<sup>261</sup> prediction interval and 0 otherwise. The measure corresponds to the percentage  
<sup>262</sup> of observations not contained in the  $(1 - \beta)\%$  prediction interval. A reliability  
<sup>263</sup> bias can be defined as

$$RB = \beta - REL \quad (14)$$

and becomes negative if the prediction bands fail to include more than  $\beta\%$  of the observations (it is otherwise positive). Ideally, the reliability bias should be 0. The bias is bounded depending on the significance level.

- Average relative interval length

$$ARIL = \frac{1}{N} \sum_{i=1}^N \frac{\hat{U}_{i+k|i} - \hat{L}_{i+k|i}}{V_{i+k|i}} \quad (15)$$

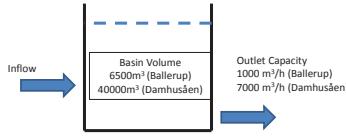
This refers to the average width of a volume prediction interval with lower bound  $\hat{L}_{i+k|i}$  and upper bound  $\hat{U}_{i+k|i}$  generated for a forecast horizon of  $k$  time steps relative to the observed value  $V_{i+k|i}$ . We consider 95% prediction intervals.

- $CRPS$  (10)

In general, a good stochastic forecast will be calibrated, i.e., generate reliabilities close to the significance level of the required prediction interval. Given a calibrated model, the spread of the prediction bounds should of course be as narrow as possible, which is indicated by low  $ARIL$  values. As an overall criterion, we aim to obtain the minimal  $CRPS$  for the forecasts of runoff volume.

## 2.6 Evaluating the Overflow Risk for Different Forecast Types

To evaluate the effect of the considered forecasting models on the RTC, a simplified assessment of the model ability to correctly predict the overflow cost according to equation (16) was used. We assumed a basin at the outlet of both catchments studied. The basin outlet capacity is fixed. The outlet capacity and volume were both chosen somewhat arbitrarily but such that a reasonable amount of overflow is obtained in the summer period considered. The selected values are shown in Figure 3.



**Fig. 3** Simplified model setup used for the evaluation of the predicted overflow cost for the different models and catchments

We considered a prediction horizon of 10 time steps or 100 min. Evaluating the basin mass balance with the selected characteristics and the measured time series of catchment outflows, we determined a series of 'predicted' overflow volumes over a 100-min horizon at every time step. Assuming a unit cost of overflow volume, this amount also corresponds to the true 'predicted' overflow cost  $C_{f,t}$ :

$$C_{f,t} = \int C(V_f) \cdot p(V_f)_t dV_f \quad (16)$$

301 where  $C(V_{f,t})$  corresponds to the cost value forecasted at time step t and  $p(V_f)_t$  is the  
302 forecasted probability that a runoff volume  $V_f$  occurs.

303 Forecasts of the runoff volumes were again derived from the probabilistic flow  
304 forecasts using the sampling approach described in section 2.5. Each sample forms a  
305 time series of flow predictions for the different horizons. We can evaluate the basin  
306 mass balance for this time series and compute the predicted overflow cost for each  
307 sample. Ideally, the predicted overflow cost derived from the stochastic models will  
308 match the reference derived from the observations at every time step.

---

309 **3 Results**

310 **3.1 Forecast Performance of Different Objective Functions**

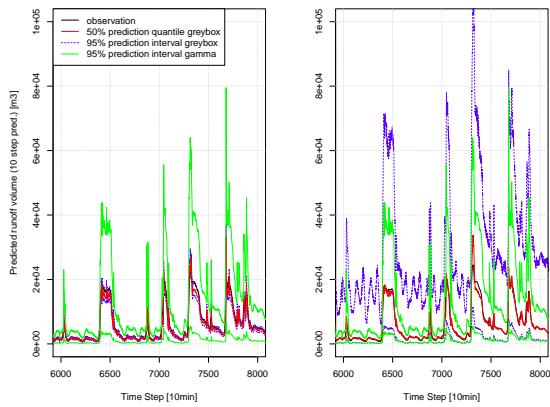
311 ***3.1.1 Runoff Predictions for a Number of Rain Events***

312 Figures 4 and 5 compare the predicted runoff volumes from the different models to  
313 the observed runoff volumes in the Damhusåen catchment. We also included predic-  
314 tion intervals that are based on the point prediction of model A, which describe the  
315 uncertainty of the runoff forecasts by a Gamma distribution, as detailed by Vezzaro  
316 and Grum (2012).

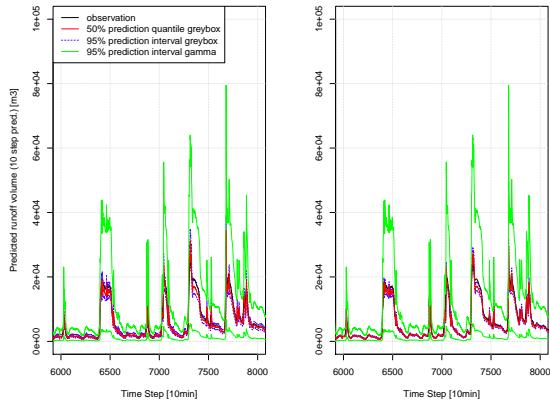
317 We found that model A satisfactorily captures the characteristics of the observed  
318 runoff curve. The prediction intervals, however, appear to be rather small. Model  
319 B provides very wide prediction intervals, whereas model C gives wider prediction  
320 intervals than model A. The forecasts from model D appear similar to those from  
321 model A, although the 50% quantile of the forecasts appears to match the observa-  
322 tions slightly better than model A. The prediction bounds from model D are narrower  
323 than those from model A.

324 With the exception of model B, all of the models appear to provide better esti-  
325 mates of forecast uncertainty than the Gamma distribution.

326 The estimated model parameters, which are shown in Table 1, exhibit the fol-  
327 lowing tendency: models estimated using multistep predictions produce more pro-  
328 nounced runoff peaks as a result of the smaller  $K$  values in the Ballerup catchment  
329 and the larger effective areas  $A$  in the Damhusåen catchment. Note that the models  
330 do not necessarily respect the mass balance due to the state updating. For all models,  
331 we obtained rather small observation uncertainties  $\sigma_e$  compared with the uncertainty  
332 of the model states ( $\sigma_1, \sigma_2$ ). This result demonstrates, that we consider the informa-  
333 tion content in the flow observations to be high and update the model to stay close  
334 to these observations. The different forecast uncertainties apparent in Figures 4 and 5  
335 are a result of the different state uncertainties  $\sigma_1$  and  $\sigma_2$ , which are shown in Table 1.



**Fig. 4** 95% prediction intervals for the 10-step runoff volume forecasts from model A (left) and model B (right). The 50% prediction quantile and the observation and the prediction intervals derived from the Gamma distribution are also shown (Damhusåen catchment).



**Fig. 5** 95% prediction intervals for the 10-step runoff volume forecasts from model C (left) and model D (right). The 50% prediction quantile and the observation and the prediction intervals derived from the Gamma distribution are also shown (Damhusåen catchment).

**Table 1** Parameter estimates for the two catchments obtained with different estimation approaches.

	$a_0$ [ $m^3/h$ ]	$K$ [h]	$A$ [ha]	$\sigma_1$	$\sigma_2$	$\sigma_e$
<b>Ballerup</b>						
A	393	7.52	206	1.61E+00	1.28E-02	6.88E-06
B	400	5.28	55	1.26E+00	5.54E-02	6.76E-07
C	372	3.03	74	4.35E-01	1.82E-02	4.35E-09
D	307	3.63	78	3.62E-01	1.09E-02	6.03E-06
<b>Damhusåen</b>						
A	841	1.95	94	1.16E+00	6.79E-03	1.24E-08
B	1678	2.49	270	7.70E+00	1.29E-01	1.07E-10
C	997	1.88	207	1.34E+00	7.34E-03	6.72E-10
D	933	2.51	122	7.64E-01	5.92E-03	7.24E-10

<sup>336</sup> *3.1.2 Evaluation of the Predictive Distributions*

<sup>337</sup> The first step in this analysis is to study the overall quality of the predictive distributions.  
<sup>338</sup> The *CRPS* was used to compare the forecasts and the observations. Note that,  
<sup>339</sup> other than in the criterion derived for the model estimation in section 2.3.4, we based  
<sup>340</sup> the analysis on the predicted runoff volumes for a given horizon.

<sup>341</sup> The estimation based on the volume prediction errors (Model B) clearly gives the  
<sup>342</sup> worst *CRPS* values. For the other models, we cannot easily identify the differences  
<sup>343</sup> based on this criterion. In both catchments, the volume forecasts generated by models  
<sup>344</sup> A, C, and D are very similar with respect to the *CRPS*.

<sup>345</sup> Table 3 shows the *ARIL* values for the 95% prediction intervals of the runoff  
<sup>346</sup> volumes for different forecast horizons. Model B yields very wide prediction intervals  
<sup>347</sup> because it considers only the point prediction in the model estimation. Large state  
<sup>348</sup> uncertainties facilitate the state updating and, if the quality of the observations is  
<sup>349</sup> acceptable, lead to better point predictions. The predicted uncertainties, however, are  
<sup>350</sup> too large.

<sup>351</sup> There is less difference between the forecasts generated by models A and C with  
<sup>352</sup> respect to *ARIL*. The forecasts generated by model D are clearly sharper than those  
<sup>353</sup> obtained with the other models. This tendency of the *CRPS*-based estimation was  
<sup>354</sup> also noted by Gneiting et al (2005).

<sup>355</sup> Although we assumed that the simple lumped reservoir model is much less suited  
<sup>356</sup> to the prediction of the runoff in the bigger and more complex Damhusåen catchment  
<sup>357</sup> than in the Ballerup catchment, we cannot identify a trend toward relatively larger  
<sup>358</sup> forecast uncertainties for this catchment.

<sup>359</sup> We continued this analysis by evaluating the distribution of the predicted runoff  
<sup>360</sup> volumes. Figure 6 shows the reliability biases *RB* of the runoff volume predictions

**Table 2** CRPS for volume predictions in  $m^3$  considering different prediction horizons (in time steps, step length = 10 min) and different estimation approaches for the two catchments.

Horizon	Ballerup				Damhusåen			
	A	B	C	D	A	B	C	D
1	2	3	2	2	5	25	5	5
2	5	7	5	5	14	62	14	15
3	8	12	8	8	28	110	29	30
4	12	17	12	11	48	169	48	51
5	16	23	16	15	73	239	73	76
6	21	30	21	20	103	319	103	106
7	27	38	26	25	137	410	137	141
8	33	46	32	30	176	511	176	181
9	39	54	38	36	219	623	218	224
10	46	63	45	42	267	745	264	271
Mean	21	29	20	19	107	321	107	110

**Table 3** ARIL for 95% volume prediction intervals considering different prediction horizons (in time steps, step length = 10 min) and different estimation approaches for the two catchments.

Horizon	Ballerup				Damhusåen			
	A	B	C	D	A	B	C	D
1	0.21	0.67	0.22	0.19	0.10	1.65	0.12	0.08
2	0.22	0.75	0.24	0.19	0.14	2.15	0.16	0.10
3	0.24	0.85	0.27	0.20	0.18	2.73	0.20	0.12
4	0.27	0.95	0.30	0.22	0.22	3.38	0.25	0.14
5	0.30	1.04	0.33	0.23	0.26	4.13	0.29	0.16
6	0.33	1.14	0.35	0.25	0.29	4.97	0.34	0.19
7	0.36	1.23	0.38	0.26	0.33	5.93	0.39	0.21
8	0.39	1.31	0.40	0.27	0.37	7.00	0.43	0.23
9	0.42	1.40	0.43	0.29	0.40	8.19	0.48	0.25
10	0.45	1.48	0.45	0.30	0.44	9.50	0.52	0.27
Mean	0.32	1.08	0.34	0.24	0.27	4.96	0.32	0.18

361 considering different levels of significance  $\beta$  and prediction horizons. A significance  
 362 level of  $\beta = 0.05$  corresponds to a  $1 - 0.05 = 95\%$  prediction interval.

363 For model A, we observed small reliability biases for high coverage rates, i.e., at  
 364 the tails of the distribution. For smaller coverage rates, however, we overestimated the  
 365 forecast uncertainties, which led to positive reliability biases. This problem becomes  
 366 more pronounced for longer forecast horizons.

367 As indicated previously, model B clearly overestimates the forecast uncertainties  
 368 and yields strongly positive reliability bias values. Model C gives results that are  
 369 similar to those of model A, but generates smaller reliability biases at longer horizons.

370 Model D yields a slight underestimation of the forecast uncertainties for high  
 371 coverage rates. Compared with models A and C, however, the overestimation of the  
 372 uncertainties in the center of the distribution is also reduced. Similar to model C, we  
 373 observed smaller reliability bias values at longer horizons with model D compared  
 374 with model A.

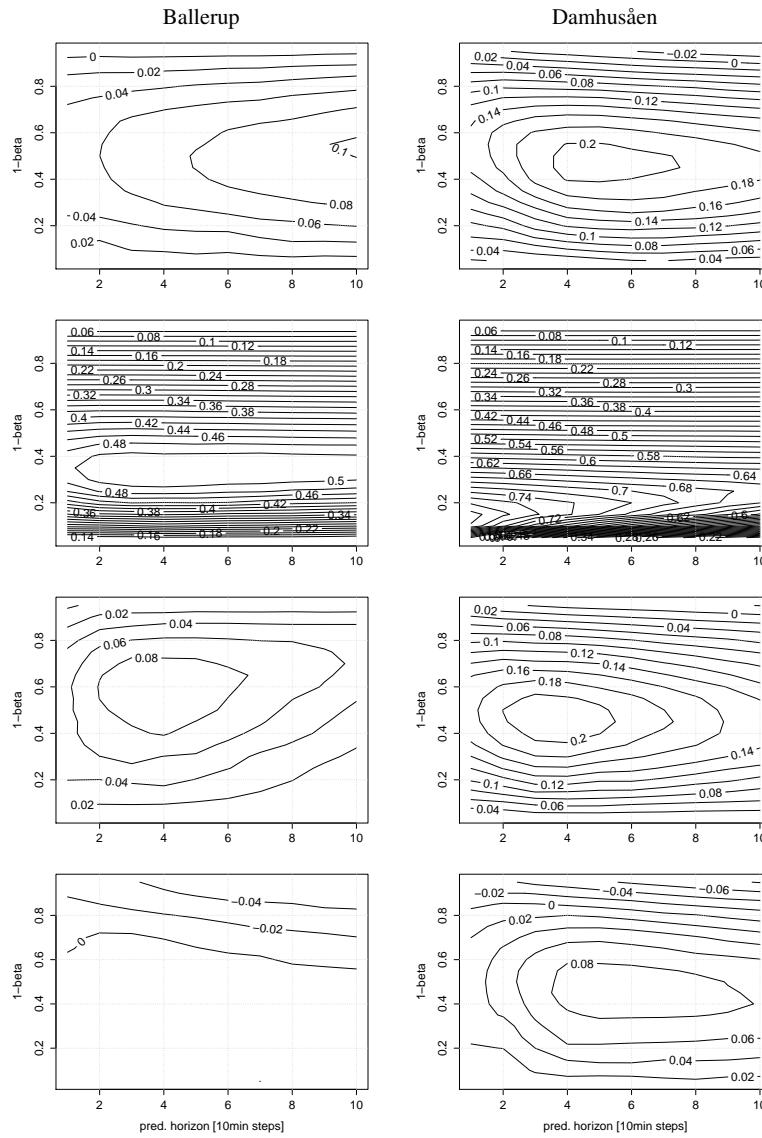
375 Models C and D account for multistep predictions in model estimation. In both  
 376 cases, this results in reduced reliability bias values at longer horizons compared to  
 377 model A. The parameter estimation in model C focuses on 95% prediction intervals.  
 378 This model consequently provides the best fit at the tails of the distribution.

379 For model D, a more balanced behavior can be observed with a reduced overes-  
 380 timation of the uncertainties at the center of the distribution but an underestimation  
 381 at the tails. The latter leads to slightly worse CRPS values of the forecasts of runoff  
 382 volume compared with model A.

383 In general, all of the models result in either an overestimation of the forecast  
 384 uncertainties at the center of the distribution or an underestimation at the tails. This  
 385 finding indicates that the normality assumption used in the multivariate sampling  
 386 approach may not hold.

387 There is a noticeable difference between the two catchments. Although we ob-  
388 tained somewhat reliable (or calibrated) forecasts for the Ballerup catchment, we  
389 tended to overestimate the uncertainties in the Damhusåen catchment because the  
390 applied model is less suitable to the description of the behavior of this system.

391 In both catchments, the forecast uncertainties during rain events are clearly un-  
392 derestimated by models A, C, and D (data not shown). This finding indicates that  
393 the applied model structure is not able to properly distinguish between dry and wet  
394 weather uncertainties.



**Fig. 6** Reliability bias for models A (top) through D (bottom) considering the prediction intervals for different levels of significance  $\beta$  and different forecast horizons for the Ballerup (left) and Damhusåen (right) catchments.

<sup>395</sup> 3.2 Predicted Overflow Cost for Fictive Basins

<sup>396</sup> Table 4 shows the overflow cost predicted using the simplified approach described in  
<sup>397</sup> section 2.6. The values for the true observations and the runoff predictions generated  
<sup>398</sup> by the different models over a horizon of 10 time steps are shown. The values shown  
<sup>399</sup> are integrated over the whole time period of 3 months. To compare the results with  
<sup>400</sup> the state-of-the-art method, we included two additional cases:

- <sup>401</sup> – Gamma - uses the point prediction from model A and derives the forecast uncer-  
<sup>402</sup> tainty for the runoff volumes from a Gamma distribution, the shape parameters of  
<sup>403</sup> which depend on the point value (Vezzaro and Grum, 2012)
- <sup>404</sup> – Model A Point - derives the predicted overflow volumes using the point forecast  
<sup>405</sup> of model A without considering the forecast uncertainties

<sup>406</sup> We found that model A produces values close to the true overflow volumes in  
<sup>407</sup> the Ballerup catchment and underestimates the true cost in the Damhusåen catch-  
<sup>408</sup> ment. When ignoring the forecast uncertainties provided by model A, we obtained  
<sup>409</sup> almost the same results (Model A Point), whereas the description of the forecast un-  
<sup>410</sup> certainties with a Gamma distribution results in a clear overestimation of the overflow  
<sup>411</sup> volumes. This finding indicates that a correct point forecast is essential for a good es-  
<sup>412</sup> timation of the overflow volumes in the simplified setup, whereas too small or no  
<sup>413</sup> estimates of forecast uncertainties hardly affect the estimation of overflow volumes.  
<sup>414</sup> In contrast, a too large estimate of the runoff forecast uncertainties, as obtained from  
<sup>415</sup> the Gamma distribution, will lead to an overestimation of the overflow risk.

<sup>416</sup> A similar result was obtained with model B, which generated reasonable re-  
<sup>417</sup> sults for the Ballerup catchment but strongly overestimated the overflow risk in the  
<sup>418</sup> Damhusåen catchment as a result of the very high estimates of the forecast uncer-  
<sup>419</sup> tainty.

<sup>420</sup> Models C and D exhibit a tendency to underestimate the overflow volumes. As in  
<sup>421</sup> model A, this underestimation is the result of the underestimation of the runoffs by  
<sup>422</sup> the point prediction, as observed in Figure 5.

<sup>423</sup> Considering the *RMSE* between the true overflow cost for a 10-step horizon and  
<sup>424</sup> the predicted overflow cost values derived from the different models, we obtained  
<sup>425</sup> a similar picture. A clear overestimation of the forecast uncertainties also results in  
<sup>426</sup> increased *RMSE* values for the overflow risk (model A  $\Gamma$  and model B), whereas  
<sup>427</sup> neglecting the forecast uncertainties hardly affects the estimated overflow cost values  
<sup>428</sup> (model A Point). Models C and D provide better point forecasts during the overflow  
<sup>429</sup> events in the Damhusåen catchment, which results in smaller *RMSE* values for the  
<sup>430</sup> overflow cost.

<sup>431</sup> In the authors' view, the most interesting outcome of this analysis is that no dif-  
<sup>432</sup> ference was found between the deterministic prediction of the overflow risk (model  
<sup>433</sup> A Point) and the use of forecast uncertainties (model A). Two possible reasons can be  
<sup>434</sup> suggested for this result. First, a linear relationship between the overflow volume and  
<sup>435</sup> the overflow cost was used in this simplified analysis. With a nonlinear relationship  
<sup>436</sup> that punishes (for example) the start of an overflow event, forecasts of the overflow  
<sup>437</sup> risk will profit more from a proper quantification of the uncertainty of the runoff  
<sup>438</sup> predictions.

**Table 4** Predicted overflow cost for 10-step horizon in  $m^3$  accumulated over all time steps.

	Ballerup	Damhusåen
True	3.1E+05	6.5E+05
A	3.2E+05	4.2E+05
B	3.8E+05	2.2E+06
C	2.5E+05	4.4E+05
D	2.6E+05	4.6E+05
$A\Gamma$	4.5E+05	8.2E+05
A Point	3.1E+05	4.1E+05

**Table 5** RMSE between the true overflow cost and the prediction in  $m^3$  from different models for a 10-step horizon.

	A	B	C	D	$A\Gamma$	A Point
Ballerup	37	43	45	40	62	37
Damhusåen	190	672	177	167	194	195

439 Second, the analysis performed here was static in the basin layout. Using the con-  
 440 sidered dataset, we obtained either only one event with a small overflow volume or  
 441 several events with rather large overflow volumes by choosing different basin dimen-  
 442 sions or outlet capacities. The effect of considering the forecast uncertainties is most  
 443 visible in those events where either the basin is close to being completely filled or  
 444 where only small overflow volumes are observed. In a predictive real-time control  
 445 system, the simulated basin outlet is varied in the optimization routine, which results  
 446 in strong variations in the simulated basin filling. A proper description of the forecast  
 447 uncertainties is more important in those cases.

---

**448 4 Conclusions**

449 We have evaluated the quality of probabilistic multistep runoff volume forecasts gen-  
450 erated by stochastic greybox models and compared the effect of different estimation  
451 methods on the forecast quality. Four methods were compared: a maximum like-  
452 lihood estimation based on one-step-ahead predictions (model A), a deterministic  
453 method that minimizes the error of the 10-step-ahead predictions (model B), and two  
454 methods that minimize the interval score for the 95% intervals of the multistep flow  
455 predictions (model C) or the continuous ranked probability score (*CRPS*, model D).

456 We concluded that, although it focuses on the whole prediction horizon, the de-  
457 terministic estimation method (model B) is unsuitable for estimating the stochastic  
458 models. The quality of the predictive uncertainty is not a criterion in the objective  
459 function for this method. In the cases considered here, this model results in too large  
460 estimates of the uncertainty for the states.

461 Models A, C, and D provided reasonable estimation results and multistep fore-  
462 casts of the runoff volume with similar skill values. Overfitting by model A was not  
463 observed as a result of the high quality of the considered flow observations. More  
464 noisy measurements will make the parameter estimation using one-step predictions  
465 more difficult and favor approaches focusing on multistep predictions.

466 However, the use of multistep predictions in parameter estimation (models C and  
467 D) clearly reduces the overestimation of the uncertainties at longer forecast horizons.  
468 Using the interval score for the parameter estimation (model C) results in forecasts  
469 that are suitable for the 95% prediction interval and overestimate the uncertainties in  
470 the center of the distribution. Applying the *CRPS* as the objective function (model D)  
471 allows the balance of this effect and gives forecasts that are more evenly calibrated  
472 over the whole distribution.

473 In the prediction of the overflow risk in a simplified setup, it was demonstrated  
474 that a significant overestimation of the runoff forecast uncertainties leads to a strong  
475 overestimation of the overflow risk. Consequently, models A, C, and D all outperform  
476 the reference model, which describes the forecast uncertainties with a simple gamma  
477 distribution.

478 In the applied setup, it is not possible to show that the risk of basin overflow  
479 can be predicted better through the dynamic modeling of the uncertainties of the  
480 runoff forecasts compared to the application of a simple point forecast. However, the  
481 analysis applied here is linear and static in the basin layout. It is expected that the  
482 forecast uncertainties will be relevant in a more realistic control setting that exhibits  
483 nonlinear relationships between the forecast values and the risk and where the basin  
484 outflows are dynamically modified as part of an optimization routine.

485 In addition, all of the models clearly underestimate the forecast uncertainties dur-  
486 ing rain events. This finding suggests that the model structure should be modified to  
487 allow a proper separation of the dry and the wet weather uncertainties.

488 We need to be aware that this study focuses strongly on the correct prediction  
489 of the overflow risk to improve the real-time control of sewer systems. The methods  
490 suggested for the prediction of these risks, however, are also applicable in other con-  
491 texts of the urban drainage system, such as the prediction of the critical operational  
492 states at a wastewater treatment plant, the risk of flooding induced by overloading of

<sup>493</sup> the sewer system, and the risk of microbial pollution as a result of sewer overflows  
<sup>494</sup> close to bathing areas.

495 **Acknowledgements** This research was financially supported by the Danish Council for Strategic Re-  
496 search, Program of the Commission on Sustainable Energy and Environment through the Storm- and  
497 Wastewater Informatics (SWI) Project. The catchment and flow data were kindly provided by Avedøre  
498 Wastewater Services and Copenhagen Energy. We thank Luca Vezzaro for ideas on the evaluation of the  
499 overflow risk, Julija Tastu and Pierre Pinson for their input on the generation of predictive distributions and  
500 the evaluation of probabilistic forecasts and Anne Katrina Duun-Henriksen for help with the preparation  
501 of this manuscript.

**References**

- 503 Bechmann H, Nielsen MK, Madsen H, Poulsen NK (1999) Grey-box modelling of  
504 pollutant loads from a sewer system. *Urban Water* 1(1):71–78
- 505 Bechmann H, Madsen H, Poulsen NK, Nielsen MK (2000) Grey box modeling of first  
506 flush and incoming wastewater at a wastewater treatment plant. *Environmetrics*  
507 11(1):1–12
- 508 Breinholt A, Thordarson FO, Møller JK, Grum M, Mikkelsen PS, Madsen H (2011)  
509 Greybox modelling of flow in sewer systems with state-dependent diffusion. *Envi-*  
510 *ronmetrics* 22(8):946–961
- 511 Carstensen J, Nielsen MK, Strandbæk H (1998) Prediction of hydraulic load for urban  
512 storm control of a municipal wwt plant. *Water science and technology* 37(12):363–  
513 370
- 514 Fuchs L, Beenken T (2005) Development and implementation of a real-time control  
515 strategy for the sewer system of the city of vienna. *Water Science and Technology*  
516 52(5):187–194
- 517 Gneiting T, Raftery AE (2007) Strictly proper scoring rules, prediction, and estima-  
518 tion. *Journal of the American Statistical Association* 102(477):359–378
- 519 Gneiting T, Raftery AE, III AHW, Goldman T (2005) Calibrated probabilistic fore-  
520 casting using ensemble model output statistics and minimum crps estimation.  
521 *Monthly Weather Review* 133(5):1098–1118
- 522 Grum M, Thornberg D, Christensen M, Shididi S, Thirsing C (2011) Full-scale real  
523 time control demonstration project in copenhagens largest urban drainage catch-  
524 ments. In: *Proceedings of the 12th International Conference on Urban Drainage*
- 525 Hallam JW (2010) Genetic algorithms with shrinking population size. *Computational  
526 statistics* 25(4):691
- 527 Jin X, Xu CY, Zhang Q, Singh V (2010) Parameter and modeling uncertainty simu-  
528 lated by glue and a formal bayesian method for a conceptual hydrological model.  
529 *Journal of Hydrology* 383(3–4):147–155
- 530 Jørgensen HK, Rosenørn S, Madsen H, Mikkelsen PS (1998) Quality control of rain  
531 data used for urban runoff systems. *Water science and technology* 37(11):113–120
- 532 Kristensen NR, Madsen H (2003) Continuous time stochastic modelling ctsm 2.3  
533 - mathematics guide. Tech. rep., Technical University of Denmark, URL <http://www2.imm.dtu.dk/~ctsm/MathGuide.pdf>
- 535 Kristensen NR, Madsen H, Jørgensen SB (2004) Parameter estimation in stochastic  
536 grey-box models. *Automatica* 40(2):225–237
- 537 Löwe R, Mikkelsen PS, Madsen H (2012a) Forecast generation for real-time control  
538 of urban drainage systems using greybox modelling and radar rainfall. In: *Proced-  
539 ings of the 10th International Conference on Hydroinformatics*
- 540 Löwe R, Mikkelsen PS, Rasmussen MR, Madsen H (2012b) State-space calibration  
541 of radar rainfall and stochastic flow forecasting for use in real-time control of urban  
542 drainage systems. In: *Proceedings of the 9th International Conference on Urban  
543 Drainage Modelling*
- 544 Madsen H (2008) Time series analysis. Chapman Hall/CRC, Boca Raton, FL
- 545 Nash JE, Sutcliffe J (1970) River flow forecasting through conceptual models part ia  
546 discussion of principles. *Journal of Hydrology* 10(3):282–290

- 547 Pabst M, Alex J, Beier M, Niclas C, Oigurek M, Peikert D, Schütze M (2011) Adesba  
548 - a new general global control system applied to the hildesheim sewage system. In:  
549 Proceedings of the 12th International Conference on Urban Drainage
- 550 Puig V, Cembrano GG, Romera J, Quevedo J, Aznar B, Ramón G, Cabot J (2009)  
551 Predictive optimal control of sewer networks using coral tool: application to riera  
552 blanca catchment in barcelona. Water Science and Technology 60(4):869–878
- 553 Schütze M, Campisano A, Colas H, Schilling W, Vanrolleghem PA (2004) Real time  
554 control of urban wastewater systemswhere do we stand today? Journal of hydrology  
555 299(3):335–348
- 556 Seggelke K, Löwe R, Beeneken T, Fuchs L (2012) Implementation of an integrated  
557 real-time control system of sewer system and wwtw in the city of wilhelmshaven.  
558 Urban Water Journal submitted
- 559 Spall JC (2003) Introduction to stochastic search and optimization / Estimation, simu-  
560 lation, and control. Wiley, Hoboken,NJ, wiley-Interscience series in discrete math-  
561 ematics
- 562 Thordarson FO, Breinholt A, Møller JK, Mikkelsen PS, Grum M, Madsen H (2012)  
563 Uncertainty assessment of flow predictions in sewer systems using grey box mod-  
564 els and skill score criterion. Stochastic Environmental Research and Risk Assess-  
565 ment 26(8):1151–1162
- 566 Venables WN, Ripley BD (2002) Modern Applied Statistics with S, 4th edn. Springer,  
567 New York, URL <http://www.stats.ox.ac.uk/pub/MASS4>
- 568 Vestergaard M (1998) Nonlinear filtering in stochastic volatility models
- 569 Vezzaro L, Grum M (2012) A generalized dynamic overflow risk assessment (dora)  
570 for urban drainage rtc. In: 9th International Conference on Urban Drainage Mod-  
571 elling
- 572 Weijs SV, Schoups G, van de Giesen N (2010) Why hydrological forecasts should be  
573 evaluated using information theory. Hydrology and Earth System Sciences Discus-  
574 sions 7:4657–4685
- 575 Whitley D (1994) A genetic algorithm tutorial. Statistics and computing 4(2):65–85



PAPER D

# Comparing two stochastic approaches to predict urban rainfall-runoff with explicit consideration of model bias

---

**Authors:**

Dario Del Giudice, Roland Löwe, Henrik Madsen, Peter Steen Mikkelsen, Jörg Rieckermann

**Manuscript in preparation for:**

*Water Resources Research*, 2014.



## Comparing two stochastic approaches to predict urban rainfall-runoff with explicit consideration of model bias

Dario Del Giudice<sup>1</sup>, Roland Löwe<sup>2</sup>, Henrik Madsen<sup>2</sup>, Peter Steen Mikkelsen<sup>3</sup>,  
Jörg Rieckermann<sup>1</sup>

### Abstract

The quantification of model simulation uncertainties resulting from, for example, structural deficiencies and uncertain inputs has received increasing attention in recent hydrological literature. Simple approaches making strong assumptions about independence and distribution are still widely applied but lead to insufficient parameter estimates and wrong quantifications of model uncertainty. We present and compare two approaches that explicitly account for the time-varying and autocorrelated character of model errors. These approaches originate from different fields. Namely, the Bayesian bias description was developed for environmental modelling and has a strong focus on the simulation of environmental systems. This approach describes errors in the model outputs using a time-varying, autocorrelated bias. The stochastic grey-box approach (or state noise description) originates from the control and automation field. It includes the uncertainty description in the model states instead of the model output and strongly focuses on on-line applications.

We compare both approaches theoretically and in a common urban case study. Both approaches are able to reliably quantify simulation and forecast uncertainty and both have their pros and cons. The bias approach focuses on simulations and demonstrates a stronger capacity to generate simulations on long horizons but is computationally demanding. The stochastic grey-box approach, on the other hand, focuses on on-line application and allows for a fast testing of different model structures and can be applied for forecasting runoffs

---

<sup>1</sup>Urban Water Management Department, Swiss Federal Institute of Aquatic Science and Technology (eawag), Überlandstrasse 133, CH-8600 Dübendorf, Switzerland

<sup>2</sup>Department of Applied Mathematics and Computer Science, Technical University of Denmark, Matematiktorvet, B322, DK-2800 Kgs. Lyngby, Denmark

<sup>3</sup>Department of Environmental Engineering, Technical University of Denmark, Miljøvej, B115, DK-2800 Kgs. Lyngby, Denmark

in an on-line setting.

# Comparing two stochastic approaches to predict urban rainfall-runoff with explicit consideration of model bias

Dario Del Giudice,<sup>1,2</sup> Roland Löwe,<sup>3</sup> Henrik Madsen,<sup>3</sup> Peter Steen Mikkelsen,<sup>4</sup>, and Jörg Rieckermann,<sup>1</sup>

**Abstract.** The quantification of model simulation uncertainties resulting from, for example, structural deficiencies and uncertain inputs has received increasing attention in recent hydrological literature. Simple approaches making strong assumptions about independence and distribution are still widely applied but lead to insufficient parameter estimates and wrong quantifications of model uncertainty. We present and compare two approaches that explicitly account for the time-varying and autocorrelated character of model errors. These approaches originate from different fields. Namely, the Bayesian bias description was developed for environmental modelling and has a strong focus on the simulation of environmental systems. This approach describes errors in the model outputs using a time-varying, autocorrelated bias. The stochastic grey-box approach (or state noise description) originates from the control and automation field. It includes the uncertainty description in the model states instead of the model output and strongly focuses on on-line applications.

We compare both approaches theoretically and in a common urban case study. Both approaches are able to reliably quantify simulation and forecast uncertainty and both have their pros and cons. The bias approach focuses on simulations and demonstrates a stronger capacity to generate simulations on long horizons but is computationally demanding. The stochastic grey-box approach, on the other hand, focuses on on-line application and allows for a fast testing of different model structures and can be applied for forecasting runoffs in an on-line setting.

## 1. Introduction

Hydrological and hydrodynamic models are applied in urban areas to assess problems related to the occurrence of sewer overflows, the possibility of flooding, and the dynamics of aquatic contaminants, among others. This information is used for optimal real time control, to effectively reduce pollution and inundations, and to support rational decision making in urban water management. Ideally, for decision support, the applied models would perfectly represent the behaviour of the existing systems and future management alternatives. Unfortunately, as models are mere abstractions of the relevant/dominant processes, it still holds that “all models are wrong, but some are useful”.

In environmental systems, we can typically observe this “wrongness” as a systematic mismatch between true system response, such as discharge, and the modelled output. In urban hydrology, recent investigations into the underlying reasons of this mismatches suggest that most discrepancies are caused by model structure deficits and input errors [Del Giudice et al., 2013]. For example, structural deficits

mainly correspond to the incorrect or oversimplified description of important mechanisms of the system [Reichert and Mieleitner, 2009] whereas, input errors can result from insufficient spatial coverage of pluviometers, systematic errors of radar-derived rain rates, errors in land-use data etc. [McMillan et al., 2011]. Other causes of modelling errors are due to errors in the output measurements, insufficient number of calibration data [Sikorska et al., 2013] and incomplete knowledge on the model parameters. To support rational decision making, we have to explicitly account for these sources of uncertainty to avoid i) misidentification of processes and parameters and ii) overconfident predictions [Reichert and Mieleitner, 2009].

Although several approaches have been suggested to account for the individual sources of uncertainty in hydrological applications, some do not take into account all error sources or make very strong assumptions (see below for details). Consequently, their performance can vary considerably and some of them (e.g., [Reichert and Mieleitner, 2009; Bulygina et al., 2009; Honti et al., 2013]) produce more reliable and statistically sound predictions than others. However, it is currently not clear which of the suggested techniques can be recommended in which field of application. Also, it is not clear what are the advantages of different strategies to account for model bias due to i) the incorrect or oversimplified description of important mechanisms of the system and ii) input errors.

In this study, we therefore discuss two approaches that account for model discrepancy and apply them to probabilistically predict the sewage flow in near real-time as well as over a period of 2 weeks on a common case study. Specifically, we compare an approach that uses a deterministic model together with a statistical description of bias to a method that suggest to rather use stochastic models, formulated as stochastic differential equations (SDE) with system noise.

<sup>1</sup>Urban Water Management Department, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland.

<sup>2</sup>Institute of Environmental Engineering, Swiss Federal Institute of Technology Zürich, Zürich, Switzerland.

<sup>3</sup>Department of Applied Mathematics and Computer Science, Technical University of Denmark, Lyngby, Denmark.

<sup>4</sup>Department of Environmental Engineering, Technical University of Denmark, Lyngby, Denmark.

To this aim we unify the two terminologies, which makes it possible to compare the two methodologies regarding their i) underlying concepts, ii) specific configurations for hydrological applications and, finally, iii) specific numerical implementations. Our goal is not to assess which approach is superior to the other, but rather to evaluate them on different criteria. This will support modellers to better choose how to account for systematic mismatches between true system response and modelled output for their particular problem.

To the best of our knowledge, we present the first detailed comparison of a Bayesian description of model output bias and a stochastic description of system noise in environmental modelling. Both approaches originate from different traditions and we present them in a common framework and with a unified terminology to make them comparable. Further innovations are the application of the bias description to a wastewater system and applying a statistical description of bias, based on a Gaussian process, to on-line predictions of rainfall-runoff.

The remainder of the paper is structured as follows. Section 2 gives a concise review of the different approaches of uncertainty analysis, classifies them according to their main characteristics and motivates why the statistical description of bias should be compared to the state noise description. Section 3 explains the two methods in detail and discusses their similarities and differences. In addition, we present the numerical experiments and suggest suitable performance statistics we form the basis for our comparison. In Section 4 we describe the case study catchment and the conceptual hydrological model as well as the monitoring data used for model calibration and prediction. The inferred parameters and predicted flows are presented in Section 5. In Section 6, we discuss the main results and provide recommendations for future studies. Finally, we summarize our main findings in Section 7.

## 2. Brief review of methods for uncertainty quantification in hydrology

As mentioned above, several approaches have been suggested to quantify uncertainty in hydrological systems. Although their goal is to provide reliable uncertainty estimates that account for the individual sources of uncertainty in hydrological applications, some do not take into account all error sources or make very strong assumptions about the errors. In table 1, we compared the different approaches according to their main characteristics: i) model formulation, ii) assumptions regarding modelling errors and iii) representation of the different error sources.

First, it seems intuitive to distinguish methodologies for probabilistic flow predictions based on how they formulate the model [Renard et al., 2010]. Traditionally, one tries to model the output of a system by using a deterministic model (a.k.a. simulator) and one or more probabilistic error terms (first row of Table 1). We name this approach “output error modelling”. Alternatively, recent studies have represented the dynamics of a hydrosystem as temporal evolution of internal model states that are not necessarily observable (second row of Table 1). The model outcome, which is a function of these states (possibly) perturbed by noise is then affected by an observation error term. We name this approach “internal error modelling”.

Complementary to how they formulate the model, methods for uncertainty analysis of runoff predictions can be distinguished regarding their characterization of modelling errors [Yang et al., 2008]. Unfortunately, the vast majority of studies in urban hydrology considers the residual errors to be independent and identically distributed (iid) (first column of Table 1). This assumption is very strong and usually not justifiable in real applications, because it generates too

narrow (i.e. unreliable) predictions and uninterpretable parameter estimates.

Conceptually superior are some statistical approaches, mainly applied in natural hydrology, that tried to explicitly describe the different sources of errors (third column of Table 1). These techniques are appealing because of their rigorous foundation and the expected ability to produce reliable predictive distributions and to reduce systematic modelling errors. Their drawbacks are linked to the high conceptual complexity and the computational inefficiency.

Alternatively, a statistical description of model bias (or model inadequacy or discrepancy) uses an additional auto-correlated error term alongside the observation error. This is on the intermediate range of complexity, in between the two approaches mentioned above. The model bias is usually described as an additive autoregressive error process in the output space [Reichert and Schuwirth, 2012] or as a noise term (or diffusion term) in the state space [Breinholt et al., 2012]. Both approaches to explicitly model bias, once externally and once internally, have recently been applied in the hydrological literature to produce short or long term predictions [Honti et al., 2013; Löwe et al., 2013]. They both have shown to produce more reliable predictions than approaches neglecting structural errors, without being as computationally intensive as the frameworks describing and propagating the error sources [Del Giudice et al., 2013; Breinholt et al., 2012]. Therefore, they seem to be ideally suited for our comparison.

## 3. Methods

### 3.1. Two approaches to explicitly account for bias in rainfall-runoff modelling

We here explain and compare the concepts behind the bias description (BD) and state noise description (SND), which both compute probabilistic sewer flow predictions in the presence of autocorrelated residual errors. Both approaches infer parameters and uncertainties through a formal likelihood evaluation. However, while the BD adds stochasticity at the system output level, the SND adds noise in the system equations *and* in the output space. While the second framework has the advantage of naturally avoiding unrealistic outputs, it also has the drawback that we need to make assumptions for the evaluation of the likelihood function.

In the following we first present the general concept of representing the systematic deviations externally or internally of the model equations and then suggest how these error processes can be parametrized. As the deterministic model is less general and, to some degree, specific to the case study we describe it in the Materials (Sect. 4).

#### 3.1.1. Output error modelling and bias description

When modelling the behaviour of a system (for example a sewer watershed) its observed output  $\mathbf{Y}_o$  can be approximated as the sum of a deterministic model outcome plus an error term. In the single output case, the deterministic model (or simulator)  $f_M$  transforms a vector of deterministic driving forces (e.g., a rainfall time series)  $\mathbf{x}$  into a vector of output  $\mathbf{y}_M$

$$\mathbf{y}_M = f_M(\mathbf{x}, \theta). \quad (1)$$

When inferring the model parameters and making predictions for future system outputs it is usual to consider an error term additional to  $f_M$ . This stochastic component can be normal and independently distributed as in most of the environmental modelling literature (e.g., [Kleidorfer et al., 2009; Freni and Mannina, 2012], just to cite some from the urban drainage literature) or autocorrelated in time (as, e.g.,

in Kuczera [1983]; Bates and Campbell [2001]; Frey et al. [2011]; Sikorska et al. [2012]; Evin et al. [2013]). Several studies have demonstrated that autoregressive error models can characterize predictive uncertainty much more reliably than simplistic ones [e.g., Yang et al., 2007a; Honti et al., 2013].

Recent statistical literature has suggested a more satisfactory form of error parametrization, which, while accounting for autocorrelated residual errors, also distinguishes observation noise [Craig et al., 2001; Kennedy and O'Hagan, 2001; Higdon et al., 2005; Bayarri et al., 2007]. By following the notation of Reichert and Schuwirth [2012], who transferred this approach to environmental modelling, we model the system output as:

$$\mathbf{Y}_o = \mathbf{y}_M(\mathbf{x}, \theta) + \mathbf{B}_M(\mathbf{x}, \psi) + \mathbf{E}(\psi), \quad (2)$$

where  $\mathbf{B}_M$  is a random process accounting for systematic deviation of model results from the true system output,  $\mathbf{E}$  represents the uncorrelated part of the observation errors, and  $(\theta, \psi)$  are the simulator and error model parameters, respectively. All the terms are implicitly dependent on time. Modelled and observed outputs could be additionally transformed by a function (e.g., a logarithm). This can be useful in hydrology where the error variance increases during peak discharge. This effect can however be also reproduced by an heteroskedastic error model. While  $\mathbf{B}_M$  accounts for self-dependence of residuals due to input errors, structural deficits, and/or measurement biases, the output transformation compensates for residual heteroskedasticity and non-Gaussianity [O'Hagan, 2006; Dietzel and Reichert, 2012]. Del Giudice et al. [2013] suggested different parametrizations for the bias term  $\mathbf{B}_M$ .

In this study, we achieved satisfactory results with an input-dependent bias description. This formulation assumes that the bias follows an Ornstein-Uhlenbeck process where the variance is scaled depending on the rainfall input (see Honti et al. [2013] for derivation). In other words  $\mathbf{B}_M$  is modelled as a continuous version of a first-order autoregressive process with normal independent noise whose variance grows with the rain rate,  $x$ , shifted in time by a lag  $d$ . The evolution of  $\mathbf{B}_M$  and  $\mathbf{E}$  for the scalar case are described by the stochastic differential equations (SDEs)

$$d\mathbf{B}_M(t) = -\frac{\mathbf{B}_M(t)}{\tau} dt + \sqrt{\frac{2}{\tau} (\sigma_{B_{ct}}^2 + (\kappa x(t-d))^2)} dW(t), \quad (3)$$

$$E(t) = \sigma_E dW(t), \quad (4)$$

where  $\kappa$  is a scaling factor,  $d$  denotes the response time of the system to rainfall,  $\tau$  is the correlation time,  $\sigma_{B_{ct}}$  is the asymptotic standard deviation of the random fluctuations around the equilibrium.  $dW(t)$  represents increments of a standard Wiener process which may be interpreted as a continuous random walk with independent Gaussian increments [Kloeden and Platen, 1999; Iacus, 2008].

### 3.1.2. Internal error modelling and state noise description (SND)

An alternative way to account for uncertainties when modelling the behaviour of a hydrosystem is via internal error modelling (a.k.a. stochastic state-space modelling or grey-box modelling) [Vrugt et al., 2013; Kristensen et al., 2004]. Instead of adding stochasticity only to the system output, this approach considers the internal evolution of the system as a stochastic process, which essentially requires stochastic models.

$$d\mathbf{S} = f_M(\mathbf{S}, \mathbf{x}, t, \theta) dt + \sigma(\mathbf{S}, \mathbf{x}, t, \psi) d\mathbf{W}(t). \quad (5)$$

This so-called state (or transition or system) equation describes the continuous evolution of some hidden (or latent) states  $\mathbf{S}$ .  $\mathbf{S}$  is a vector of random variables which are not necessarily observable (e.g. water level in an unobserved CSO tank, hydraulic heads in specific points of an aquifer, values of soil moisture in a catchment) but which can be estimated from the observable outputs.  $\sigma$  is called "diffusion", "state noise", "stochastic forcing" or "level disturbance" and represents model bias, while  $d\mathbf{W}(t)$  are infinitesimal increments from a multidimensional Wiener process.  $f_M(\cdot)$  is called "drift term" and corresponds to the functions constituting the deterministic (part of the) model  $M$ . Adding noise to the model equations leads to a stochastic model. This is an important distinction from the BD where randomness is only added to the simulator's output.

The dynamics of the measurable output  $\mathbf{Y}_o$  are related to the state equations via an observation equation:

$$\mathbf{Y}_o = h(\mathbf{S}, \mathbf{x}, \theta, \psi, t) + \mathbf{E}(\psi), \quad (6)$$

which is a potentially non-linear function of states  $\mathbf{S}$  and parameters  $(\theta, \psi)$ . The modelled observation process  $\mathbf{Y}_o$  is assumed to be subject to independent random normal observation errors  $\mathbf{E}$ . The modelled system response  $h(\mathbf{S}, \mathbf{x}, \theta, \psi, t)$  can be additionally transformed by a function to better comply with assumptions made during parameter inference and to account for increasing observation error with increasing output, a situation often observed in hydrology [Breinholt et al., 2012]. We apply a logarithmic transformation in this work as suggested by Breinholt et al. [2011].

In this investigation we parametrized the state noise as linearly increasing with the model state

$$\sigma(\mathbf{S}, \mathbf{x}, t, \psi) = (\sigma_s \circ \mathbf{S}) \mathbf{1} \quad (7)$$

where  $\circ$  is the entry-wise product between the vector of diffusion parameters  $\sigma_s$  and the vector of states  $\mathbf{S}$ , and  $\mathbf{1}$  is the identity matrix. This formulation produced satisfactory results in previous urban hydrological studies [Breinholt et al., 2011]. The assumption of state-dependent noise in the SND is another relevant distinction from the BD where the additive noise terms can depend on the input or output, but not on a (hidden) state variable.

Our experience shows that state-dependent diffusion can be numerically very challenging. A Lamperti transform is therefore commonly applied to obtain a set of transformed equations where the state dependence of the diffusion is shifted into the drift term (see Breinholt et al. [2011]; Iacus [2008]; Møller [2010]).

Furthermore, we assumed the observation noise to have a standard deviation proportional to the flow. We did that by applying a logarithmic transformation of the model output  $h(\mathbf{S}, \mathbf{x}, \theta, \psi, t)$  and back-transforming  $h(\mathbf{S}, \mathbf{x}, \theta, \psi, t) + \mathbf{E}(\psi)$  in the real space.

### 3.2. Inference and predictions

To describe how the BD and SND differ regarding parameter estimation and predictions, we first compare the approaches on a conceptual level and then analyse their numerical implementation.

#### 3.2.1. Parameter estimation

In a probabilistic framework the inverse problem of parameter estimation requires assumptions about the error distribution. These assumptions are formalized by a likelihood function  $\mathcal{L}_M(\mathbf{y}_o | \theta, \psi, \mathbf{x})$  that describes the conditional probability density of producing the observed output data given a certain model structure  $M$ , inputs  $\mathbf{x}$ , and parameters  $(\theta, \psi)$ .

**Parameter estimation in the BD approach:**

In the BD approach the likelihood function assumes that the data generating process has a normal distribution:

$$\begin{aligned} \mathcal{L}_M(\mathbf{y}_o \mid \boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{x}) = & \\ & \frac{(2\pi)^{-\frac{n}{2}}}{\sqrt{\det(\boldsymbol{\Sigma}(\boldsymbol{\psi}, \mathbf{x}))}} \exp\left(-\frac{1}{2} [\mathbf{y}_o - \mathbf{y}_M(\boldsymbol{\theta}, \mathbf{x})]^T \boldsymbol{\Sigma}(\boldsymbol{\psi}, \mathbf{x})^{-1} \right. \\ & \left. [\mathbf{y}_o - \mathbf{y}_M(\boldsymbol{\theta}, \mathbf{x})]\right) \end{aligned} \quad (8)$$

where  $n$  is the number of observations i.e. the length of the vector  $\mathbf{y}_o$  (e.g., a measured discharge time series at the outlet of a catchment) and  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{B_M} + \boldsymbol{\Sigma}_E$  is the total error covariance matrix accounting for the autocorrelated and heteroskedastic bias process and for iid observation errors.

In the BD the parameters are interpreted as uncertain variables described by a probability distribution. Due to identifiability problems with the model and the bias term, a Bayesian approach is necessary to identify model bias in the output equation (Eq.2). This requires specifying a meaningful prior for the hyper-parameters of the bias. For statistical inference, the likelihood function is combined with the prior information about parameters to infer their posterior distribution according to Bayes' law:

$$f_{post}(\boldsymbol{\theta}, \boldsymbol{\psi} \mid \mathbf{y}_o, \mathbf{x}) = \frac{f(\boldsymbol{\theta}, \boldsymbol{\psi}) \mathcal{L}_M(\mathbf{y}_o \mid \boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{x})}{\int f(\boldsymbol{\theta}', \boldsymbol{\psi}') \mathcal{L}_M(\mathbf{y}_o \mid \boldsymbol{\theta}', \boldsymbol{\psi}', \mathbf{x}) d\boldsymbol{\theta}' d\boldsymbol{\psi}'} \quad . \quad (9)$$

We refer to Reichert and Schuwirth [2012] and Del Giudice et al. [2013] for further details. As discussed there, having a larger part of the residuals with a purely Bayesian interpretation (the bias) significantly reduces the usefulness of frequentist distributional tests.

Numerically, the posterior parameter distribution  $f(\boldsymbol{\theta}, \boldsymbol{\psi} \mid \mathbf{y}_o, \mathbf{x})$  is approximated by a Markov chain Monte Carlo (MCMC) algorithm. To have a satisfactory approximation of the posterior a large MCMC sample is required.

#### Parameter estimation in the SND approach:

As in the SND approach the model is stochastic, the full likelihood function for an arbitrary number of time steps is a path integral (see e.g., Quinn and Abarbanel [2010]). Sampling from this infinite-dimensional integral over all possible realizations of the system is however conceptually and numerically challenging. Therefore, in the current hydrological research on state noise modelling (e.g., Breinholt et al. [2011]) the likelihood function is only defined as a product of one step ahead conditional likelihoods:

$$\begin{aligned} \mathcal{L}_M(\mathbf{y}_o \mid \boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{x}) &= \left( \prod_{i=2}^n p(\mathbf{y}_{o_i} \mid \mathbf{y}_{o_{i-1}}, \boldsymbol{\theta}, \boldsymbol{\psi}) \right) p(\mathbf{y}_{o_1} \mid \boldsymbol{\theta}, \boldsymbol{\psi}) \\ &= \frac{(2\pi)^{-\frac{n}{2}}}{\sqrt{\det(\boldsymbol{\Sigma}(\boldsymbol{\psi}, \mathbf{x}))}} \cdot \exp\left(\sum_{i=2}^n \left( -\frac{1}{2} [\mathbf{y}_{o_i} - \right. \right. \\ &\quad \left. E(\mathbf{y}_{o_i} \mid \mathbf{y}_{o_{i-1}}, \boldsymbol{\theta}, \boldsymbol{\psi})]^T \boldsymbol{\Sigma}(\mathbf{y}_{o_i} \mid \mathbf{y}_{o_{i-1}}, \boldsymbol{\theta}, \boldsymbol{\psi})^{-1} \right. \\ &\quad \left. \left. [\mathbf{y}_{o_i} - E(\mathbf{y}_{o_i} \mid \mathbf{y}_{o_{i-1}}, \boldsymbol{\theta}, \boldsymbol{\psi})]\right)\right) p(\mathbf{y}_{o_1} \mid \boldsymbol{\theta}, \boldsymbol{\psi}), \end{aligned} \quad (10)$$

where  $E(\mathbf{y}_{o_i} \mid \mathbf{y}_{o_{i-1}}, \boldsymbol{\theta}, \boldsymbol{\psi})$  is the mean and  $\boldsymbol{\Sigma}(\mathbf{y}_{o_i} \mid \mathbf{y}_{o_{i-1}}, \boldsymbol{\theta}, \boldsymbol{\psi})$  covariance of the one step prediction. This product of conditional densities assumes the normality of the observations at each time-step given the observations up to time  $i-1$ . In other words, the innovations (i.e. the one step ahead errors given the observations) are assumed to be normal. This implies the assumption that the states given all observations up to  $i-1$  are normal. To prove this assumption, however, one would need to compute from the full likelihood of the

problem which is currently not possible. As an approximation, we can analyse if simulations of the SDEs one time-step ahead can be assumed normal.

In the SND inference can be either performed on a frequentist [Breinholt et al., 2012] or a Bayesian basis [Melgaard, 1994; Sadegh et al., 1994]. In the first case the parameters are assumed to have a fixed true value, which is however unknown, whereas in the second case parameters are convenient unknown quantities with a probability distribution. In the second framework we can make use of existing knowledge about parameters by formulating prior distributions. For the sake of comparability with the BD, we here adopted a Bayesian calibration. Traditionally, in the SND, Bayesian updating has consisted in maximizing the posterior  $f(\boldsymbol{\theta}, \boldsymbol{\psi} \mid \mathbf{y}_o, \mathbf{x})$  rather than characterizing its full distribution [Melgaard, 1994; Sadegh et al., 1994; Waller and Pronzato, 1997].

Numerically the so-called maximum a posteriori (MAP) estimation is performed with an extended Kalman filter (EKF). The EKF provides a consistent first-order approximation to the estimate of a non-linear state at the observation time, as well as the errors of this estimate [Kao et al., 2004]. Using an EKF accounts for non-linearity of the model (by linearising it and the evolution operators) and is computationally less expensive than an ensemble Kalman filter [Breinholt et al., 2012] or particle filter. Further details on the prediction (or ahead projection), innovation, updating, and Kalman gain equations can be found in Kristensen et al. [2004] and Kao et al. [2004].

#### 3.2.2. Rainfall-runoff predictions conditioned on flow observations

In a study it can be useful to predict system output and/or states for points in time (and space) where data have been observed and employed for parameter inference. We can call this set of points calibration period (or layout). The advantage of predicting in the past (a.k.a. smoothing [Bulygina et al., 2009]) is that it provides the distribution of the true system output and/or states. Comparing predicted and observed outputs can provide information about the nature of model discrepancy.

**Predictions in the calibration period with BD:** We here condition the bias error processes on the observation and updated parameters and propagate the parametric uncertainty of the model and both error models via Monte Carlo simulations. During model calibration, we want to predict the observed system response in the calibration layout. This implies approximating the distributions of  $\mathbf{y}_M + \mathbf{B}_M + \mathbf{E}$  for every temporal point  $i$  of the calibration dataset, i.e. for  $i = 1, \dots, n$ .

To approximate this conditioned Gaussian process, one has to propagate a sufficiently-large sample of the converged Markov chains through the simulator  $f_M$  and draw realizations of  $\mathbf{y}_M + \mathbf{B}_M + \mathbf{E}$ .

More information on how to predict in the calibration period when describing model bias can be found in Reichert and Schuwirth [2012]; Honti et al. [2013]; Del Giudice et al. [2013].

#### Predictions in the calibration period with SND :

The extended Kalman filter implemented in the SND setting provides the possibility to create both, smoothed as well as filtered estimates of the model states and outputs. While the former conditions the model output on the whole dataset [Kristensen and Madsen, 2003], the latter does so only up to the current data point. To compare model results during the calibration period to those of the BD in commensurable way we therefore generated smoothed output.

#### 3.2.3. Predictions of future rainfall-runoff observations

In environmental modelling we often want to predict the system behaviour in points in time and/or locations where no measurements are available. After their calibration, models are typically used to predict the system output for future conditions. We define the probabilistic generation of outputs in “the future” (i.e. without directly conditioning on data), predicting in the extrapolation period. Output data available here can be used for conditional validation of model performances.

Some literature also distinguishes between “forecasting” and “simulation”. The term forecasting, “ex-post” forecasting or, more rigorously, “ex-hind-casting”, is used when we start predicting from the last observation. In contrast, “simulations” refers to unconditioned predictions for future periods were there is no influence of available observations [Beven and Young, 2013] (see Appendix).

#### Predictions with BD:

The posterior predictive distribution of rainfall-runoff at new points in time and/or space is computed via Monte Carlo Simulations. To this aim, we propagate a sample of the posterior through the simulator and the error model. In this process no conditioning on data takes places and therefore predictive uncertainty becomes larger than in the calibration period. Calibration data, however, can still have an influence on these “forecasts” if those are only few correlation lengths  $\tau$  away from the last observed output.

To approximate the distribution of  $\mathbf{y}_M + \mathbf{B}M + \mathbf{E}$  we first propagate a sample of  $\theta_{post}$  through  $f_M$  and so obtain realizations  $\mathbf{y}_M$ . Second, we sample for every time step  $t_j$  from the Gaussian processes with  $\psi_{post}$  describing the errors,  $\mathbf{B}$  and  $\mathbf{E}$ . While the observation error behaves like white noise, the bias process is Markovian. Finally, we add the simulated trajectories of the error process to those of the deterministic model. See Reichert and Schuwirth [2012]; Honti et al. [2013]; Del Giudice et al. [2013] for details.

#### Predictions with SND:

Stochastic output can be generated from stochastic grey-box models without conditioning on the observations by performing “scenario (or ensemble) simulations” [Platen and Bruti-Liberati, 2010] from Eq. 5.

To compute trajectories from the stochastic differential equations describing the state-space model we use discrete-time approximations. For each solution of Eq. 5 the predictions for  $\mathbf{Y}_o$  are derived by inserting the simulated path of the states into the observation equation (Eq. 6). This approach to make (unconditional) predictions with SDEs only assumes normality of the states and observations at time  $j$ , conditional on  $\mathbf{S}_{j-1}$ , provided that the time interval between  $j$  and  $j-1$  is small.

### 3.3. Numerical experiments

To compare the performances of the two approaches for probabilistic modelling we perform three numerical experiments. First, we analyse the parameter estimates we obtain after calibration. Second, we compare the quality of short-term predictions over 100min (20 time steps) and, third, long-term predictions over 14 days (5328 time steps). The results of these experiments are assessed quantitatively with 4 different performance metrics.

#### 3.3.1. Experiment 1: Parameter estimation

We first examine the results of the inference procedure which consist in solving Eq. 9. As discussed in Sect. 3.2.1 for the BD approach we obtain the full updated distribution of parameters via MCMC methods, whereas in the SND we identify only the maximum of this distribution.

While the updated parameters  $\theta$  of the deterministic model  $f_M$  can be easily compared, the parameters  $\psi$  are specific for each error description and therefore are not directly comparable.

#### 3.3.2. Experiment 2: Ex-post long-term forecasting

Off-line ex-post forecasting implies predicting the system response starting from the last available output measurement for a long period ahead. This kind of forecasting in simply involves the extension of the prediction horizon ahead with respect to the application described below. We may also consider this procedure as simulation of the stochastic model because the conditioning induced by the observations affects only the first few time steps.

#### 3.3.3. Experiment 3: Ex-post short-term forecasting

In contrast to predictions in the extrapolation period, the idea behind ex-post short-time forecasts, often in real-time is to sequentially assimilate output data as they become available in an on-line setting. For each assimilated measurement predictions are generated for the next few time steps only. This is similar to the simulation procedure, but with three differences. First, during the first future time steps, the predictions are conditioned on the available observations and the bias-corrected model outputs (for the BD) or states (for the SND) are not only corrected in the mean, but also have a reduced prediction uncertainty. Second, the starting point of the predictions continuously moves forward in time. Third, the on-line forecasting horizon is typically short term, i.e. only few time steps ahead. For the case study, eight forecasts at different starting points were performed (over a horizon of 20 time steps).

### 3.4. Performance criteria

To evaluate the performances of the BD and SND we use four performance metrics together with a visual inspection. To assess the quality of the underlying deterministic model, we consider the median of the probabilistic simulations. We use i) the Nash-Sutcliffe efficiency index (NS, optimally approaching 1 from below) and ii) the normalized (or relative) bias NB (NB, optimally approaching 0). Both statistics are commonly used in hydrology to assess the accuracy in fitting the peaks of the hydrographs and preserve water balance, respectively [Bulygina et al., 2009; Coutou et al., 2012; Bennett et al., 2013].

To assess the quality of ex-post forecasts, we focus on 95% prediction intervals. Specifically, we evaluate their iii) “reliability”, which measures the percentage of validation measurements falling into the 95% prediction intervals, iv) “sharpness”, which is the average width of the 95% prediction bands and v) the interval (skill) score  $S_{0.05}^{int}$  (optimally approaching 0 from above), which favours precise and reliable prediction intervals [Gneiting and Raftery, 2007].

$$S_{\alpha}^{int} = (u-l) + \frac{2}{\alpha}(l-y_{o_j})H\{l-y_{o_j}\} + \frac{2}{\alpha}(y_{o_j}-u)H\{y_{o_j}-u\} \quad (11)$$

where  $\alpha = 0.05$  corresponds to the confidence level,  $u$  and  $l$  to the 97.5 and 2.5 quantiles of the predictive distribution of  $\mathbf{Y}_o$  at the extrapolation time point  $j$  and  $y_{o_j}$  to the validation observation.  $H$  denotes the unit step function which takes the value of 1 if its argument is greater than 0 and 0 otherwise. We average  $S_{0.05}^{int}$  over all considered time steps.

### 4. Materials

To compare the stochastic approaches not only on a conceptual but also on an applied level, we investigate their ability to estimate the parameters of the rainfall-runoff model and predict the flow of a combined sewer system. In the following we briefly describe the analysed urban watershed, the deterministic model used, the available hydrological measurements, the chosen model priors and the specific computer implementation adopted in this study.

#### 4.1. Case study and data

As a case study we chose the sewer watershed located in the Ballerup area close to Copenhagen (Denmark) (Figure 1). The catchment has a total area of approximately 1300 ha and is mainly laid out as a separate system, although it does have a small combined section. The runoff in this area is also strongly influenced by rainfall-dependent infiltration. Several previous studies were undertaken using data from this catchment [Breinholt et al., 2011; Breinholt et al., 2012; Löwe et al., 2013]. Flow measurements are available with a temporal resolution of 5 minutes.

Tipping bucket rain gauge measurements are available from the Danish wastewater committee's (SVK) network in the considered catchment [Jørgensen et al., 1998]. Observations from the two pluviometers located near the catchment were averaged and used as input for the runoff forecasting models in this study. These measurements had a temporal resolution of 1 minute. In addition, flow measurements are available with a temporal resolution of 5 minutes.

A modelling time step of 10 min was adopted and the flow and rain gauge data were averaged to match this time step. Considering a time of concentration of approximately 60 minutes for the catchment, this temporal resolution is appropriate for capturing the hydrosystem dynamics. Schilling [1991] recommends a temporal resolution of rainfall measurements of at least 0.2 to 0.33 times the concentration time of an urban watershed.

#### 4.2. A simple rainfall runoff model

The sewer flow at the monitoring point is modelled as a superposition of wastewater flow and rainfall-runoff. While the wastewater hydrograph is modelled as a superposition of 4 harmonic functions, the rainfall-runoff is described by a cascade of two virtual reservoirs (Figure 2). The model dynamics are described by the following deterministic equations:

$$f_M(\mathbf{S}, \mathbf{x}, t, \theta) dt = d \begin{bmatrix} S_1(t) \\ S_2(t) \end{bmatrix} = \begin{bmatrix} A_{imp} \cdot x(t) + a_0 - \frac{1}{k} S_1(t) \\ \frac{1}{k} S_1(t) - \frac{1}{k} S_2(t) \end{bmatrix} dt \quad (12)$$

with output

$$y_M(\mathbf{x}, t, \theta) = \frac{1}{k} S_2(t) + d(t) \quad (13)$$

and dry weather wastewater flow

$$d(t, \theta) = \sum_i^2 (s_i \sin \frac{i 2\pi t}{24} + c_i \cos \frac{i 2\pi t}{24}) \quad (14)$$

$S_1$  and  $S_2$  correspond to the states of the system, i.e. the levels in the virtual storage tanks, and vary in function of time (in hours). The vector  $\theta$  of physical model parameters includes the impervious catchment area  $A_{imp}$ , the mean dry weather flow at the catchment outlet  $a_0$ , the mean travel time (or reservoir residence time)  $k$  and parameters  $s_1, s_2, c_1$ , and  $c_2$  which describe the dry weather variation of the catchment outflow as an harmonic function. The vector  $\mathbf{x}$  of model inputs includes the rainfall measurements averaged from the two pluviometers near the watershed.

This simplified model deliberately disregards infiltration and does not include existing stormwater and combined sewer overflows. As such, it is too simple to exactly reproduce the flow observations. However, as a so-called “grey-box model” it captures the major processes with components that have a physical meaning. As such, its major advantages is that its equations are suitable to be incorporated into the SND framework and it is computationally fast enough to be applied in a forecast setting with data

assimilation [Breinholt et al., 2011; Löwe et al., 2013]. Simple models have often proven useful in on-line applications [Coutou et al., 2012; Wolfs et al., 2013] and in applications requiring a large number of simulations such as modelling the integrated urban water system [Achleitner et al., 2007].

#### 4.3. Prior parameters

We selected prior distributions for the BD mainly based on expert knowledge based on the results of previous simulation studies in the same catchment. Prior knowledge about simulator parameters is described by log-normal distributions with coefficient of variation of 0.2. These estimates are based on experience from previous simulation studies in the same catchment. We chose an intermediate spread of the distributions to benefit from the Bayesian ability to make use of prior knowledge.

Prior distributions for the error parameters in BD are defined in the real space. As suggested in Reichert and Schuwirth [2012] and Del Giudice et al. [2013], we defined the prior of the bias in a way to avoid model inadequacy as much as possible. This is obtained by a probability density decreasing with increasing values of  $\sigma_{B_{et}}$  and  $\kappa$  (here truncated normal distributions). This helps to reduce the identifiability problem between the deterministic model and the bias. Regarding the correlation time of the bias,  $\tau$ , we chose a prior value close to 1/3 of the hydrograph recession time.

For the SND approach all parameters, besides  $A_{imp}$  and  $k$ , are defined in a logarithmic space. With respect to the standard deviation of the observation error  $\sigma_\epsilon$  we specified a prior as consistent as possible with the one of the bias description. Since the error models are different, these two terms are not exactly comparable.

Regarding the initial transformed model states it was not easy to define meaningful prior knowledge; we therefore assigned them non-informative priors for these parameters. Similarly, prior knowledge is not available for the diffusion parameters of the SND model and we also used non-informative priors for these.

The baseflow parameters of the model were not inferred from the data simultaneously with the other parameters due to numerical difficulties encountered in the SND estimation routine. Instead, we inferred these parameters beforehand with a least squares method on a dry weather period from 18/07/2010 until 28/07/2010 to  $a_0 = 281.5 \frac{m^3}{h}$ ,  $c_1 = -47.4$ ,  $c_2 = 21.3$ ,  $s_1 = -43.4$ ,  $s_2 = -84.2$  and kept them fixed during parameter estimation for both techniques.

The prior distributions of model and error model parameters in the BD framework are described in Tab. 2.

#### 4.4. Computer implementation

The conceptual hydrological model and the BD routine for uncertainty analysis were implemented in R [R Core Team, 2013]. During inference (Eq. 9) we first obtained an appropriate jump distribution and chain starting point by sequentially using the stochastic techniques described by Haario et al. [2001] and Vihola [2012], and then sampled from the target distribution by using a Metropolis-Hastings algorithm [Hastings, 1970]. Finally, to approximate the predictive distribution of  $\mathbf{Y}_o$  we propagated 2000 posterior parameter sets through the simulator and the error model.

The SND routine was implemented in the open source software CTSW [Juhl et al., 2013] which is available as a package for R. Posterior maximization was performed using the PORT algorithm through the R function nlmminb [Gay, 1990]. For predictions with the SDEs we apply an Euler-Maruyama scheme (see e.g. Kloeden and Platen [1999]; Marcus [2008]) to generate 1000 realizations of the process  $\mathbf{S}$ .

## 5. Results

By applying the BD and SND approaches to stochastically predict the flow at the outlet of a combined sewer system we find that: i) both methodologies provide forecast coverage which is very close to the nominal 95% coverage of the validation data in the next 3.3 hours; ii) reproducing the observations in longer term forecasts horizons is to some extent more challenging, which is demonstrated by a data coverage slightly lower than expected. In general the two stochastic methodologies produce similar results in terms of reliability of the uncertainty intervals. The BD however produces average model predictions which fit the validation data better.

### 5.1. Experiment 1: Parameter estimation

The conceptual hydrological model described in Sect. 4.2 was calibrated using the data presented in Fig. 6 and on the left side of Fig. 8. Due to the different ways of considering errors in the two modelling approaches, the stochastic process parameters  $\psi$  cannot be compared directly.

The parameters inferred for the different modelling approaches are shown in Fig. 7. As previously mentioned, the BD approach provides a full distribution of parameters, while only the mode of the posterior distribution is considered in the SND. Parametric uncertainty is therefore neglected in the current implementation of the SND.

In the bias description the inference produced well-behaved updated marginals. The computation time on a recent personal desktop computer for the full posterior characterization was in the order of 1 day. The only distribution with a complex shape is  $\delta$  which represents the time steps after which the rainfall influences runoff uncertainty. For the SND, the optimization problem was solved in the order of few minutes.

The initial model states  $S_1$  and  $S_2$  stay close to the prior in the BD approach while substantially different posterior values are obtained in the SND approach. For the effective area parameter  $A$ , we observe bigger values of approximately 38 ha for the BD approach while the SND estimates an optimum of 31 ha. For the time constant  $k$  a slightly higher value is obtained for the BD approach (2.45 h) in comparison to the SND approach (2.43 h). In both approaches the inferred observation noise was significantly smaller than the bias or diffusion term.

### 5.2. Verifying Assumptions for Parameter Inference

In the SND both, the extended Kalman filtering procedure and the evaluation of the likelihood function, rely on the assumption that the one-step ahead conditional densities are normal. Since the SDE's are driven by a Wiener process with Gaussian increments, this assumption is usually reasonable for short sample times. We can verify this assumption by performing model simulations one-step ahead and analysing if the resulting sample can be assumed normally distributed.

At time step  $i - 1$  we sample 2000 times from the multivariate normal distribution [Venables and Ripley, 2002]  $N(\mathbf{Z}_{i-1|i-1}, \mathbf{P}_{i-1|i-1})$  for the updated model states which is provided by the extended Kalman filter. Each sample of states is used as the starting point for a simulation of the state equations (5) up to time step  $k$  as described in section 3.2.3. We can then derive the corresponding model outputs  $\hat{\mathbf{y}}_{k|i-1}$  using the observation equation (6) and analyse if these are normally distributed.

Figure 3 shows histograms and normal QQ-plots of the predicted model output  $\hat{\mathbf{y}}_{k|i-1}$  for situations of low and high flow in the sewer system. Similar plots can be generated for all time points in the calibration dataset. The model output forecasted one time step ahead can certainly be assumed normal.

A second assumption during parameter inference of the SND approach is that the innovations  $\epsilon_k = \mathbf{y}_k - \hat{\mathbf{y}}_{k|i-1}$  are

normally distributed and iid. This will be the case if the one-step ahead predictions  $\hat{\mathbf{y}}_{k|i-1}$  are normal and if we can demonstrate that the innovations can be considered white noise. As is clear from figure 4, the latter is not the case. We can detect significant autocorrelation and periodicities in the innovations generated during the calibration period.

However, this is not the case if parameter inference for the SND approach is performed without prior information. We do in this case obtain a clearly better model fit for the calibration data as is illustrated in figure 5. The reason is that by considering prior information for our model parameters, we assume to have knowledge about the system which cannot be inferred from the (limited) set of observations. An optimal model fit to the observations is not expected in this case. On the other hand, estimating the same model with the same prior information on a dataset including all possible system states would certainly yield uncorrelated innovations.

### 5.3. Experiment 2: Ex-post long-term forecasting

As observed for short term forecasts, also long term predictions for the two approaches were similar. Credible intervals for SDN predictions were slightly wider than for the BD and therefore covered the validation data better. Higher data coverage was associated a lower (i.e. better) average interval score  $S_{0.05}^{int} = 800$  than for the BD where  $S_{0.05}^{int} = 1162$ .

Our best knowledge of the truth in the future (i.e. the median of the underlying deterministic model) was nonetheless closer to reality when describing output bias than when modelling the state disturbance. The model calibrated with the BD obtained a higher  $NS$  index of 0.723 while the SND resulted in a model with  $NS = 0.526$ . Both error descriptions, nevertheless, predicted wet weather flows which were systematically lower than what was measured.

### 5.4. Experiment 3: Ex-post short-term forecasting

As shown in Figure 9 the percentage of data points falling into the 95% credible interval of the short term predictions was very close to the nominal coverage. This means that the prediction can be considered reliable, although the underlying simulator appears to systematically deviate from reality. This is particularly interesting during the flood event on the right side of Figure 9 where the previously calibrated model heavily underestimated the receding section of the hydrograph but the probabilistic predictions still encompassed most of the validation data.

As indicated by the interval scores  $S_{0.05}^{int}$ , which penalizes too wide and unreliable uncertainty bands, during dry weather the BD and SND produced similarly performing predictions. During storm events however, interval scores were worse (i.e. higher) for the SND in particular in the decreasing limb of the flood hydrograph. By observing the flow predictions this can be associated to the lower resolution of the forecasts generated by the stochastic grey-box modelling approach in this period. The computational time for prediction generation was in the order of few minutes for both techniques.

## 6. Discussion

### 6.1. Results interpretation

As shown in the case study application, both methodologies were able to provide both short term and long term reliable predictions. This is remarkable for two reasons. First,

the underlying lumped reservoir model was a simplified representation of reality and therefore unable to consider all mechanisms occurring in the catchment (e.g. spatially varying soil characteristics and water content or infiltration). Second, the validation conditions were consistently different from the calibration circumstances. This non-stationarity is evident in peak discharges higher than observed before and important infiltration-inflow phenomena previously unobserved. This two considerations imply that the methods are relatively resistant to non-stationarity of inputs and boundary conditions and structural errors of the model. Furthermore, conditioning on data could provide for both methodologies short term forecasts reliable and precise in all flow conditions, even when the previously calibrated simulator was heavily deviating from the measurements (Fig. 9). As demonstrated in previous studies least squares methods (first column of Tab. 1) would have produced much more overconfident (and therefore less reliable) predictions than the BD [Honti et al., 2013; Del Giudice et al., 2013] and the SND [Breinholz et al., 2012].

The selected error representations besides differing in the location of the systematic error term, in the output space for the BD and in the state space for the SND, also differ in their parametrization. While the bias is a Gaussian process whose variance increases with precipitation intensity, the state noise has variance raising with the state value. This explains the much higher uncertainties observed with the SND in the decreasing limb of the hydrograph than with the BD. Other parametrizations of the error terms are however theoretically possible for both approaches. In the BD one could for instance implement an error variance depending on the output via data transformation [Del Giudice et al., 2013]. Similarly, the SND could accommodate an input-dependence of the diffusion term which however has not been investigated yet.

The sub-optimal simulator performances, besides being ascribable to the oversimplified model structure can also be connected to the error description. As shown in Del Giudice et al. [2013] and Bayarri et al. [2007] the bias description produces model performances which are slightly inferior to least square estimations. This can be explained by the fact that the inference with the BD do not force the simulator to reproduce the observations with biased (or over-tuned) parameters.

The fair deterministic model fit obtained with the SND may be related to the state-dependent scaling of the diffusion term as well as assumptions made during parameter inference. However, simulations of the SDEs one time step ahead give no indication of a violation of the normality assumption. Similarly, the model fitted using a frequentist approach without prior information yields perfectly uncorrelated residuals.

## 6.2. Comparison of strengths and weaknesses

Combining the theoretical considerations discussed in the Method Section and the experience gained in the application of the two approaches on a common case study, we here discuss the advantages and inconveniences of the BD and the SND. We first elaborate on the conceptual level and then consider practical aspects.

By describing the bias in the model equations the SND is theoretically sounder since it is in principle not restrained to the assumption that states and output are normally distributed. Furthermore, since the bias is described in the state equation and not at the output, no negative unrealistic flow can be generated. This is not always assured when modelling the errors in the output equations (as in the BD).

In this framework, instead, output transformation might be required to ensure plausible predictions (e.g. in Frey et al. [2011] and Sikorska et al. [2012]).

Although in principle more rigorous, making inference in the SND framework is difficult, because the likelihood is a path-integral (see e.g. Quinn and Abarbanel [2010]). This implies that in current environmental applications of the SND we have to use the extended Kalman filter which assumes normality of the predicted outflow given the outflow at the previous time-step and of the states. These assumptions need to be verified and can lead to difficulties in correctly identifying model parameters in particular if the model structure is not suitable for the data.

Other than in the BD, the updating of model states in the SDN (using the ensemble Kalman filter) violates the water balance (see e.g., Salamon and Feyen [2010]). In on-line applications, however, a correct reproduction of the water balance is of minor importance as the focus of the models is typically to provide accurate and precise runoff forecasts.

Regarding the application of the BD several parametrizations for hydrological modelling are possible [Del Giudice et al., 2013]. This, while being an advantage in terms of flexibility and adaptability to very different case studies, has also the drawback of requiring some initial trial and error to define the most appropriate description of the bias process. The SND has theoretically the same pros and cons, although only a linear dependence of the bias on the state has been explored in hydrology. Libraries of different model structures can be implemented and easily tested for a given dataset.

Both techniques can in principle distinguish between parametric uncertainty, bias uncertainty and observational uncertainty. In its current implementation the SND neglects parametric uncertainty and estimates very low observation noise. This does not seem critical in urban drainage applications where model discrepancies are the predominant error type [Del Giudice et al., 2013].

A comparison of the computer implementation reveals that current calibration with the SND is approximately two order of magnitude faster than the BD. This is very probably associated to the simpler posterior surface to explore in the SND and especially to neglecting the shape of the posterior distribution and therefore parametric uncertainty. This difference in the computational effort for solving the inference problem does not affect real-time forecasts which in both cases are very rapid (in the order of seconds to minutes). Notwithstanding its relative higher computational expensiveness, the BD is still much easier than other statistical techniques aiming at identifying the causes of errors and not just describing its effects (e.g., Bulygina et al. [2009]; Reichert and Mieleitner [2009]; Renard et al. [2011])

A final practical comparison level of the two approaches regards the possibility to accommodate existing computer codes. Current hydrodynamic models such as SWMM or MIKE-SHE can be easily integrated in the BD framework, since only the outputs of these simulators have to be known and not the model equations (see Honti et al. [2013]; Del Giudice et al. [2013]). In the SDN the use of external models is not foreseen for current applications where the focus is on exploiting very simplified model structures. The application of the approach to complex existing software would be challenging since it requires an implementation of the model as stochastic differential equations (SDEs).

## 6.3. Outlook

The Bayesian description of model discrepancies and the description of state noise within a grey-box model framework, both appeared promising to produce reliable hydrological predictions. Here we describe areas of future research

that we think would be interesting to explore.

For the SND a useful development would be to develop an inference procedure with less restrictive assumptions. This could be done by applying Hamiltonian Monte Carlo methods or Particle Filters to solve the infinite-dimensional integral constituting the likelihood of this approach. At the same time the computational time for inference with a fully defined likelihood would probably rise considerably. This is the subject of current research.

A satisfactory implementation of input-dependence of the diffusion term would also be of interest for hydrological applications.

Future research for the BD should be directed in understanding the reasons for systematic deviations, namely input errors and structural deficits. While we are currently investigating this area by analysing the bias of different models (Del Giudice et al, in prep.) an even more rigorous alternative would be to describe and propagate input uncertainty. This approach, however, will imply a sensible computational burden for the analyses and will therefore require the use of statistical emulators.

## 7. Conclusions

The objective of this study was to discuss and evaluate in a commensurable way two probabilistic techniques for uncertainty estimation in hydrology. The first approach was the bias description (BD), representing systematic model discrepancies in the output space. The second was the state noise description (SND), considering model inadequacies in the system equations. To our best knowledge these two promising approaches coming from different fields (Bayesian statistics the first and control theory the second) had never been analysed in a comparable way. To clarify their potentials and limitations for the benefits of the hydrological community we compared the two approaches using a unified terminology, a similar notation and the same case study. Based on theoretical considerations and the results of an urban hydrological application we can conclude that:

1. Both approaches describe model bias in a way suitable for hydrological modelling. Indeed, they both can produce reliable forecasts in the short term, which is useful e.g. for real time control of sewer networks, and in the long term. This, although using a very simple rainfall-runoff model and analysing a sewer system with non-stationary behaviour.
2. Both techniques can make use of previously available information contained in output observations not only for model calibration but also for conditioning. This second effect allows one to generate more accurate and precise short term forecasts.
3. Each approach, besides the mentioned strengths presents some limitations. Both can only “cleverly” describe the effects of model inadequacies but cannot individuate and reduce their main reasons, namely model structural deficits and input errors. The bias description can be intermediate demanding on a computational level since it requires thousand of Monte Carlo Markov chain simulations. On the other hand, the state noise descriptions, in its current implementation, makes some strong assumptions on the states distribution which can lead to suboptimal deterministic model performances.
4. Given the clear benefits of the BD and the SND over the least squares methods currently used in urban hydrology, we recommend the use of one of them as minimal technique for obtaining reliable probabilistic discharge predictions in the short and long term.

## Appendix A: Terminology

Different terminologies are used in the field. We therefore here give a brief summary of the keywords used in this article. Most of these definitions are given in more detailed form by *Beven and Young* [2013]. We denote as

1. **Forecast** - model results generated for a future time step  $t + k$  given inputs and observations of model outputs until the starting point  $t$ . Future inputs and output observations are considered unknown. In shorthand, we write  $t + k|t$ .

2. **Ex-post forecasting** - model results generated in a forecast setting but assuming the future (rainfall) inputs known.

3. **Simulation** - model results generated given some defined inputs but without conditioning on any observations of model outputs

4. **Filtering** - model results for time step  $t$  where both, inputs and observations, are known for this time step and the model is updated to the observations. In shorthand, we write  $t|t$  and we refer to characterizing the system state at the current time [Bulygina et al., 2009].

The term **prediction** is used used ambiguously in the literature. Time series applications often use it in the sense of *forecasting*, however, without making assumptions on the future input [Madsen, 2008]. In hydrology, the term has been used in the sense of generating model outputs in general [Reichert and Schuwerth, 2012] or the generation of forecasts where future inputs are assumed known [Renard et al., 2010]. Finally, we denote as **innovations** the one-step ahead ex-post forecast errors [Chatfield, 2003].

**Acknowledgments.** The authors are grateful to Carlo Albert for the interesting discussions about inference in state space modelling. This work was partially supported by a grant from Swiss National Science Foundation and by the DSF. Drawing by Frits Ahlefeldt.

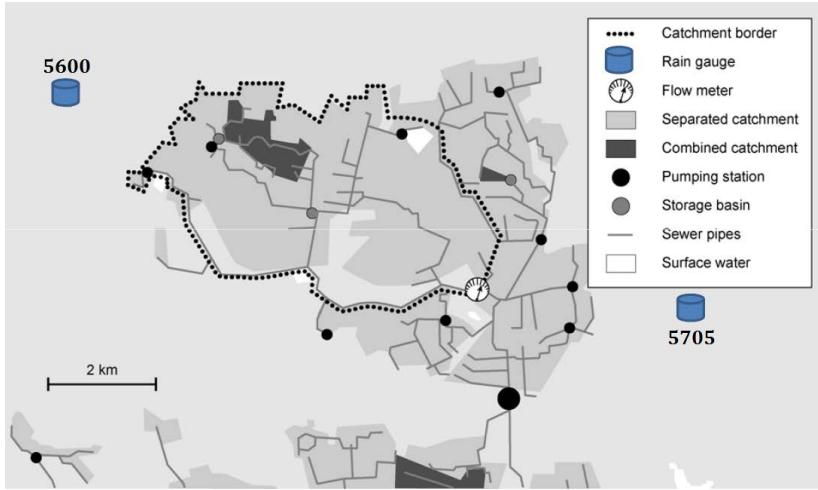
## References

- Achleitner, S., Möderl, M. and Rauch, W.: CITY DRAIN – An open source approach for simulation of integrated urban drainage systems, *Environmental Modelling and Software*, 22, 1184–1195, doi:10.1016/j.envsoft.2006.06.013.
- Bates, B. and Campbell, E.: A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling, *Water Resources Research*, 37, 937–947, 2001.
- Bayarri, M., Berger, J., Paulo, R., Sacks, J., Cafeo, J., Cavendish, J., Lin, C., and Tu, J.: A framework for validation of computer models, *Technometrics*, 49, 138–154, 2007.
- Bennett, N. D., B. F. W. Croke, G. Guariso, J. H. A. Guillaume, S. H. Hamilton, A. J. Jakeman, S. Marsili-Libelli, L. T. H. Newham, J. P. Norton, C. Perrin, and et al. (2013), Characterising performance of environmental models, *Environmental Modelling and Software*, 40, 1–20, doi:10.1016/j.envsoft.2012.09.011.
- Beven, K., and P. Young (2013), A guide to good practice in modeling semantics for authors and referees, *Water Resources Research*, 49(8), 5092–5098, doi:10.1002/wrcr.20393
- Breinholt, A., F. O. Thordarson, J. K. Møller, M. Grum, P. S. Mikkelsen, and H. Madsen (2011), Grey-box modelling of flow in sewer systems with state-dependent diffusion, *Environmetrics*, 22(8), 946–961, doi:10.1002/env.1135.
- Breinholt, A., Møller, J., Madsen, H., and Mikkelsen, P.: A formal statistical approach to representing uncertainty in rainfall-runoff modelling with focus on residual analysis and probabilistic output evaluation-distinguishing simulation and prediction, *J. Hydrol.*, 472–473, 36–52, doi:10.1016/j.jhydrol.2012.09.014, 2012.
- Bulygina, Nataliya and Gupta, Hoshin (2009), Estimating the uncertain mathematical structure of a water balance model via bayesian data assimilation, *Water Resources Research*, 45(12).
- Chatfield, C. (2003), *The Analysis of Time Series: An Introduction*, sixth edit ed., 352 pp., Chapman and Hall/CRC.
- Coutu, S., Del Giudice, D., Rossi, L., and Barry, D.: Parsimonious hydrological modeling of urban sewer and river catchments, *Journal of Hydrology*, 464 - 465, 477 – 484, doi:10.1016/j.jhydrol.2012.07.039, 2012.
- Craig, P., Goldstein, M., Rougier, J., and Seheult, A.: Bayesian forecasting for complex systems using computer simulators, *Journal of the American Statistical Association*, 96, 717–729, 2001.
- Del Giudice, D., Honti, M., Scheidegger, A., Albert, C., Reichert, P., and Rieckermann, J.: Improving uncertainty estimation in urban hydrological modeling by statistically describing bias, *Hydrology and Earth System Sciences*, 17, 4209–4225, doi:10.5194/hess-17-4209-2013, 2013.
- Dietzel, A. and Reichert, P.: Calibration of computationally demanding and structurally uncertain models with an application to a lake water quality model, *Environmental Modelling and Software*, 38, 129–146, 2012.
- Evin, G., Kavetski, D., M., T., and G., K.: Pitfalls and improvements in the joint inference of heteroscedasticity and autocorrelation in hydrological model calibration, *Water Resources Research*, pp. n/a–n/a, doi:10.1002/wrcr.20284, 2013.
- Freni, G. and Mannina, G.: Uncertainty estimation of a complex water quality model: The influence of Box–Cox transformation on Bayesian approaches and comparison with a non-Bayesian method, *Physics and Chemistry of the Earth, Parts A/B/C*, 42, 31–41, 2012.
- Frey, M. P., Stamm, C., Schneider, M. K., and Reichert, P.: Using discharge data to reduce structural deficits in a hydrological model with a Bayesian inference approach and the implications for the prediction of critical source areas, *Water Resources Research*, 47, doi:10.1029/2010WR009993, 2011.
- Gay, M. (1990), Usage Summary for Selected Optimization Routines, *Computing Science Technical Report No. 153*, AT&T Bell Laboratories, Murray Hill, NJ, United States, url. 1990.
- Goodwin, G. C., and R. L. Payne (1977), *Dynamic system identification : Experiment design and data analysis*, 291 pp., Academic Press.
- Gneiting, T., and A. E. Raftery (2007), Strictly proper scoring rules, prediction, and estimation, *Journal of the American Statistical Association*, 102(477), 359–378, doi:10.1198/016214506000001437.
- Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G., and Ye, M.: Towards a comprehensive assessment of model structural adequacy, *Water Resources Research*, 48, n/a–n/a, doi:10.1029/2011WR011044, 2012.
- Haario, H., Saksman, E., and Tamminen, J.: An Adaptive Metropolis Algorithm, *Bernoulli*, 7, pp. 223–242, 2001.
- Hastings, W.: Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, 57, 97–109, 1970.
- Hemri, S., F. Fundel, and M. Zappa (2013), Simultaneous calibration of ensemble river flow predictions over an entire range of lead times, *Water Resources Research*, 49, 6744–6755, doi:10.1002/wrcr.20542.
- Higdon, D., Kennedy, M., Cavendish, J. C., Cafeo, J. A., and Ryne, R. D.: Combining Field Data and Computer Simulations for Calibration and Prediction, *SIAM J. Sci. Comput.*, 26, 448–466, doi:10.1137/S1064827503426693, 2005.
- Honti, M., Stamm, C., and Reichert, P.: Integrated uncertainty assessment of discharge predictions with a statistical error model, *Water Resources Research*, doi:doi:10.1002/wrcr.20374, 2013.
- Iacus, S. M.: Simulation and inference for stochastic differential equations: with R examples, Springer, 2008.
- Jørgensen, H. K., S. Rosenørn, H. Madsen, and P. S. Mikkelsen (1998), Quality control of rain data used for urban runoff systems, *Water science and technology*, 37(11), 113–120.
- Juhl, R., N. R. Kristensen, P. Bacher, J. K. Kloppenborg, and H. Madsen (2013), CTSM: Continuous Time Stochastic Modeling, Version 0.6.7., Dept. Applied Mathematics and Computer Science, Technical University of Denmark, Kgs. Lyngby, Denmark, url. 2013.
- Kao, Jim and Flicker, Dawn and Henninger, Rudy and Frey, Sarah and Ghil, Michael and Ide, Kayo (2004), Data assimilation with an extended Kalman filter for impact-produced shock-wave dynamics, *Journal of Computational Physics*, 196, 705–723, doi:10.1016/j.jcp.2003.11.028, 2004.

- Kennedy, M. and O'Hagan, A.: Bayesian calibration of computer models, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63, 425–464, 2001.
- Kleidorfer, M., Deletic, A., Fletcher, T. D., and Rauch, W.: Impact of input data uncertainties on urban stormwater model parameters, *WATER SCIENCE AND TECHNOLOGY*, 60, 1545–1554, doi:10.2166/wst.2009.493, 2009.
- Kloeden, P. E., and E. Platen (1999), *Numerical Solution of Stochastic Differential Equations*, vol. 3rd, Springer.
- Kristensen, N. R., and H. Madsen (2003), *Continuous Time Stochastic Modelling CTSM 2.3 - Mathematics Guide*.
- Kristensen, N. R., H. Madsen, and S. B. Jørgensen (2004), Parameter estimation in stochastic grey-box models, *Automatica*, 40(2), 225–237, doi:10.1016/j.automatica.2003.10.001.
- Krzysztofowicz, R. (2001), The case for probabilistic forecasting in hydrology, *Journal of Hydrology*, 249(1–4), 2–9, doi:10.1016/S0022-1694(01)00420-6
- Kuczera, G.: Improved parameter inference in catchment models: 1. Evaluating parameter uncertainty, *Water Resources Research*, 19, 1151–1162, doi:10.1029/WR019i005p01151, 1983.
- Kuczera, G., D. Kavetski, S. Franks, and M. Thyre (2006), Towards a bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters, *Journal of Hydrology*, 331(1), 161–177, doi:10.1016/j.jhydrol.2006.05.010
- Lin, Z. and Beck, M.: On the identification of model structure in hydrological and environmental systems, *Water resources research*, 43, W02402, 2007.
- Löwe, R., Mikkelsen, P., and Madsen, H.: Stochastic rainfall-runoff forecasting: parameter estimation, multi-step prediction, and evaluation of overflow risk, *Stochastic Environmental Research and Risk Assessment*, pp. 1–12, doi:10.1007/s00477-013-0768-0, 2013.
- Adaptive state updating in real-time river flow forecasting — A combined filtering and error forecasting procedure, *Journal of Hydrology*, (1–4), 302–312.
- Madsen, H. (2008), *Time series analysis*, 380 pp., Chapman and Hall/CRC.
- McMillan, H., Jackson, B., Clark, M., Kavetski, D., and Woods, R.: Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models, *Journal of Hydrology*, 400, 83–94, 2011.
- Melgaard, H. (1994), Identification of physical models, Phd, Technical University of Denmark.
- Moradkhani, Hamid and DeChant, Caleb M. and Sorooshian, Soroosh (2012), Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-markov chain monte carlo method, *Water Resources Research*, 48(12), n/a–n/a, doi:10.1029/2012WR012144.
- Møller, J. K. (2010), Stochastic state space modelling of nonlinear systems, Ph.D. thesis, Technical University of Denmark (DTU).
- O'Hagan, A.: Bayesian analysis of computer code outputs: A tutorial, *Reliability Engineering & System Safety*, 91, 1290–1300, 2006.
- Platen, E., and Bruti-Liberati, N. (2010), *Numerical Solution of Stochastic Differential Equations with Jumps in Finance*, vol. 64, Springer.
- Quinn, J. C. and Abarbanel, H. D.: State and parameter estimation using Monte Carlo evaluation of path integrals, *Quarterly Journal of the Royal Meteorological Society*, 136, 1855–1867, doi:10.1002/qj.690, 2010.
- R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/>, 2013.
- Reichert, P.: Conceptual and practical aspects of quantifying uncertainty in environmental modelling and decision support, 2012 International Congress on Environmental Modelling and Software, International Environmental Modelling and Software Society (iEMSs), [http://www.iemss.org/sites/iemss2012/proceedings/D1\\_2\\_0666\\_Reichert.pdf](http://www.iemss.org/sites/iemss2012/proceedings/D1_2_0666_Reichert.pdf), 2012.
- Reichert, P. and Mieleitner, J.: Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters, *Water Resour. Res.*, 45, W10402, doi:10.1029/2009WR007814, 2009.
- Reichert, P. and Schuwirth, N.: Linking statistical bias description to multiobjective model calibration, *Water Resources Research*, 48, W09543, doi:doi:10.1029/2011WR011391, 2012.
- Renard, Benjamin and Kavetski, Dmitri and Kuczera, George and Thyer, Mark and Franks, Stewart W. , Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resour. Res.*, 46(5), W05521, doi:10.1029/2009WR008328, 2010.
- Renard, B., Kavetski, D., Leblois, E., Thyer, M., Kuczera, G., and Franks, S. W.: Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation, *Water Resour. Res.*, 47, W11516, doi:10.1029/2011WR010643, 2011.
- Sadegh, P., H. Melgaard, H. Madsen, and J. Holst (1994), Optimal experiment design for identification of grey-box models, in *Proc. 1994 Am. Control Conf. - ACC '94*, vol. 1, pp. 132–137, IEEE, doi:10.1109/ACC.1994.751709.
- Salamon, P. and Feyen, L. (2010), Disentangling uncertainties in distributed hydrological modeling using multiplicative error models and sequential data assimilation, *Water Resources Research*, 46(12), W12,501.
- Schilling, W. (1991), Rainfall data for urban hydrology: what do we need?, *Atmos. Res.*, 27(1), 5–21.
- Sikorska, A., Scheidegger, A., Banasik, K., and Rieckermann, J.: Bayesian uncertainty assessment of flood predictions in ungauged urban basins for conceptual rainfall-runoff models, *Hydrology and Earth System Sciences*, 16, 1221, doi:10.5194/hess-16-1221-2012, 2012.
- Sikorska, A. E., Scheidegger, A., Banasik, K., and Rieckermann, J.: Considering rating curve uncertainty in water level predictions, *Hydrology and Earth System Sciences*, 17, 4415–4427, doi:10.5194/hess-17-4415-2013, 2013.
- Schoups, G. and Vrugt, J. A.: A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors, *Water Resour. Res.*, 46, W10531, doi:10.1029/2007WR006720, 2010.
- Venables, W. N., and B. D. Ripley (2002), *Modern Applied Statistics with S*, vol. Fourth, Springer, New York.
- Vezzaro, L., and M. Grum (2013), A generalized dynamic overflow risk assessment (dora) for urban drainage real time control, *Journal of Hydrology*, submitted.
- Vihola, M.: Robust adaptive Metropolis algorithm with coerced acceptance rate, *Statistics and Computing*, 22, 997–1008, 2012.
- Vrugt, J. A., C. G. H. Diks, H. V. Gupta, W. Bouten, and J. M. Verstraten (2005), Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, *Water Resources Research*, 41(1), W01017, doi:10.1029/2004WR003059.
- Vrugt, J. A., ter Braak, C. J., Diks, C. G., and Schoups, G. (2013): Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: Theory, concepts and applications, *Advances in Water Resources*, 51, 457 – 478, doi: <http://dx.doi.org/10.1016/j.advwatres.2012.04.002>, 2013.
- Walter, E., and L. Pronzato (1997), *Identification of parametric models from experimental data*, 413 pp., Springer.
- Wilkinson, RichardD. and Vrettas, Michail and Cornford, Dan and Oakley, JeremyE. , *Quantifying Simulator Discrepancy in Discrete-Time Dynamical Simulators*, *Journal of Agricultural, Biological, and Environmental Statistics*, doi:10.1007/s13253-011-0077-3, 2011.
- Wolfs, V., Villazon, M., and Willemse, P.: Development of a semi-automated model identification and calibration tool for conceptual modelling of sewer systems, *Water Science and Technology*, 68, 167–175, doi:10.2166/wst.2013.237.
- Yang, J., Reichert, P., and Abbaspour, K.: Bayesian uncertainty analysis in distributed hydrologic modeling: A case study in the Thur River basin (Switzerland), *Water Resources Research*, 43, W10401, doi:10.1029/2006WR005497, 2007a.
- Yang, J., P. Reichert, K. C. Abbaspour, and H. Yang (2007), Hydrological modelling of the Chaohe Basin in China: Statistical model formulation and Bayesian inference, *Journal of Hydrology*, 340(3–4), 167 – 182, doi:10.1016/j.jhydrol.2007.04.006, 2007b.
- Yang, J., Reichert, P., Abbaspour, K. C., Xia, J., and Yang, H.: Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China, *Journal of Hydrology*, 358, 1–23, 2008.

Corresponding author: Dario Del Giudice, Urban Water Management Department (SWW), Swiss Federal Institute of Aquatic Science and Technology (EAWAG), Dübendorf, Switzerland. (Dario.DelGiudice@eawag.ch)





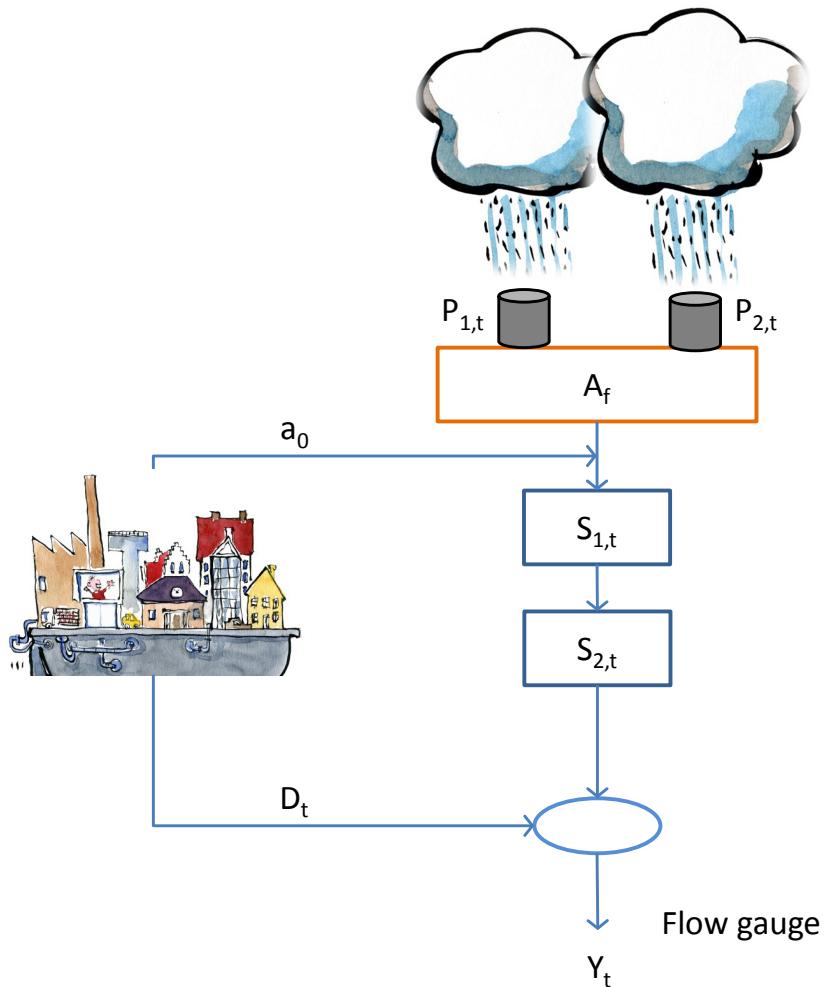
**Figure 1.** The considered Ballerup catchment together with locations of the rain gauges used for deriving the mean areal rainfall and the location of the flow meter [Breinholt et al., 2012]

**Table 1.** Probabilistic approaches for runoff predictions

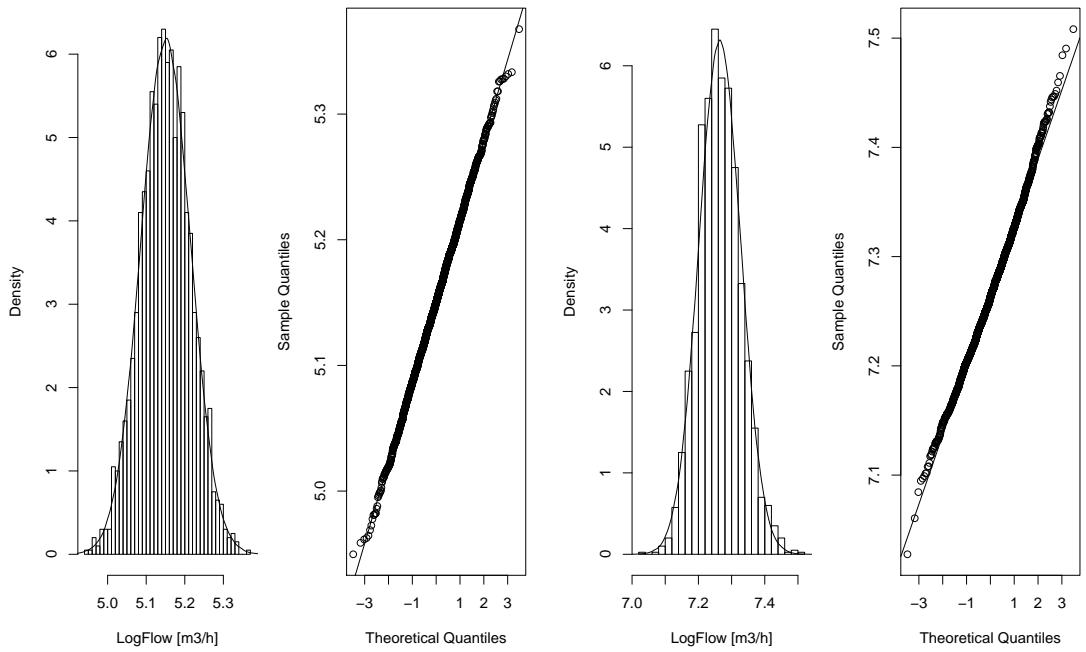
	Errors iid	Systematic deviation described	Error sources represented
<b>Output error modeling</b> (deterministic model + stochastic errors)	Freni and Mannina [2012] Kleidorfer et al. [2009]	Del Giudice et al. [2013] Reichert and Schuwirth [2012] Wilkinson et al. [2011]	Kuczera et al. [2006] Renard et al. [2011]
<b>Internal error modeling</b> (stochastic model + stochastic errors)	Moradkhani et al. [2012] Vrugt et al. [2013]	Breinholt et al. [2012] Vrugt et al. [2005] Madsen and Skotner [2005]	Reichert and Mieleitner [2009] Bulygina et al. [2009] Lin and Beck [2007] Salamon and Feyen [2010]

**Table 2.** Conceptual model and error model calibration parameters ( $\theta, \psi$ ). The notation for prior distributions is: LN( $\mu, \sigma$ ): lognormal, TN( $\mu, \sigma, a_1, a_2$ ): truncated normal, Exp( $\lambda^{-1}$ ): exponential. The symbol meaning is:  $\mu$ : expected value,  $\sigma$ : standard deviation,  $a_1$ : lower limit,  $a_2$ : upper limit,  $\lambda$ : rate.

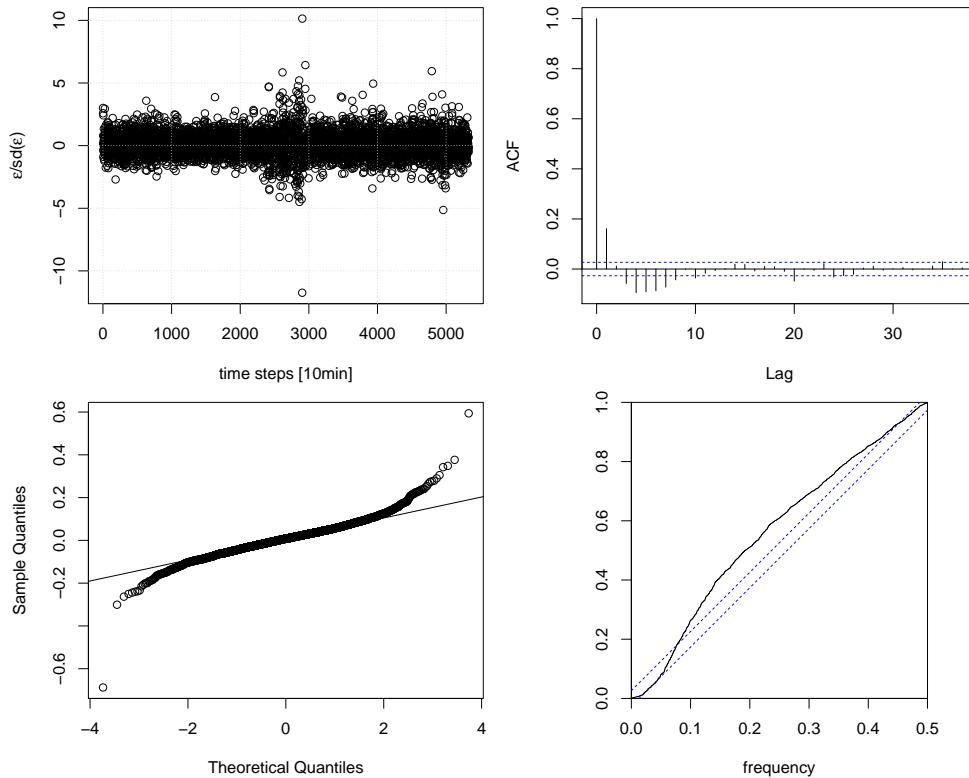
Name	Description	Units	Prior (for BD)	Prior (for SND)
$A_{imp}$	...	ha	LN(4.31,0.86)	LN(4.31,0.86)
$S_{1,0}$		$m^3$	LN(200,40)	LN(7,1000)
$S_{2,0}$		$m^3$	LN(200,40)	LN(7,1000)
$K$		h	LN(4.5,0.9)	TN(4.5,0.9,0,1000)
corrlen	BD: Correlation Length of $\mathbf{B}$ ( $\tau$ )	10min	LN(10,3)	-
sd.Eps_Q	Standard Deviation of $\mathbf{E}$ ( $\sigma_E$ )	$m^3/h$	LN(20,2)	LN(-2.55,0.255)
sd.B_Q	BD: Standard Deviation of $\mathbf{B}$ ( $\sigma_{B_{ct}}$ )	$m^3/h$	TN(0,40,0,10 <sup>8</sup> )	-
ks_Q	BD: Proportionality Constant between precipitation and uncertainty increase ( $\kappa$ )	$m^2$	TN(0,57965,0,10 <sup>8</sup> )	-
Delta	BD: Lag (in timesteps) between precipitation and uncertainty increase ( $\delta$ )	[ $\cdot$ ]	Exp(6)	-
$\sigma_1$	SND: diffusion scaling for state $Z_1$	[ $\cdot$ ]	-	LN(-10,1000)
$\sigma_2$	SND: diffusion scaling for state $Z_2$	[ $\cdot$ ]	-	LN(-10,1000)



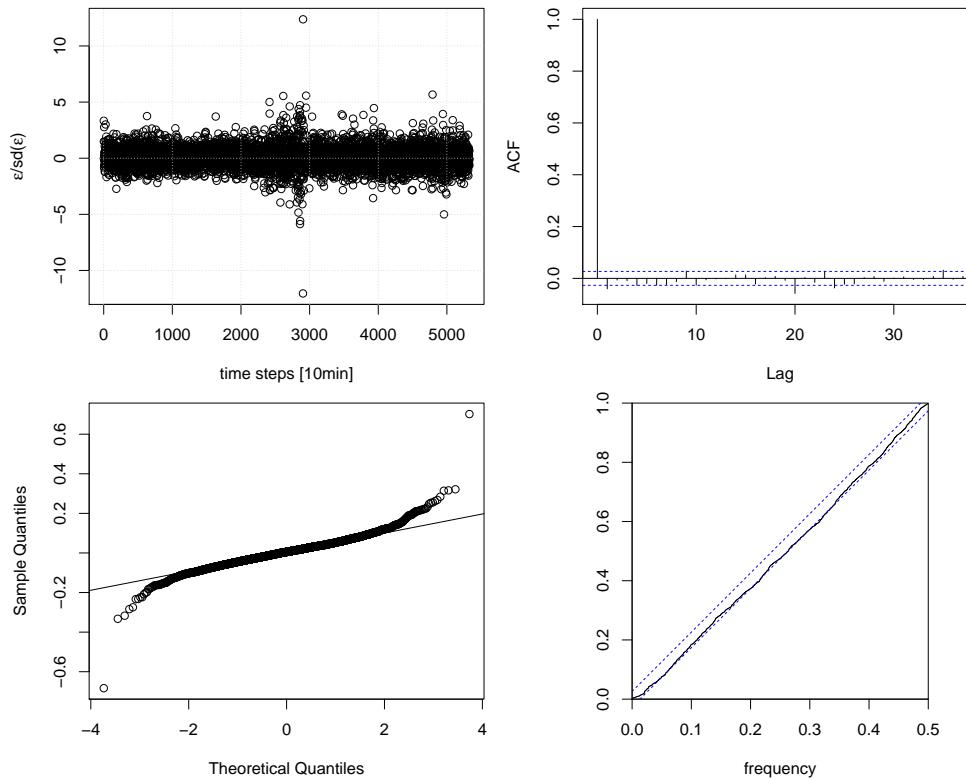
**Figure 2.** The linear reservoir cascade model considered for rainfall runoff modelling



**Figure 3.** Histogram and normal QQ-Plot for one-step ahead predictions generated by simulating the SND model. Left - low flow situation, right - high flow situation.



**Figure 4.** Analysis of innovations  $\epsilon$  during calibration period, top left: innovations standardized by standard deviation, top right: sample autocorrelation, bottom left: normal QQ-plot, bottom right: cumulated periodogram



**Figure 5.** Analysis of innovations  $\epsilon$  during calibration period when performing parameter inference without prior information, top left: innovations standardized by standard deviation, top right: sample autocorrelation, bottom left: normal QQ-plot, bottom right: cumulated periodogram

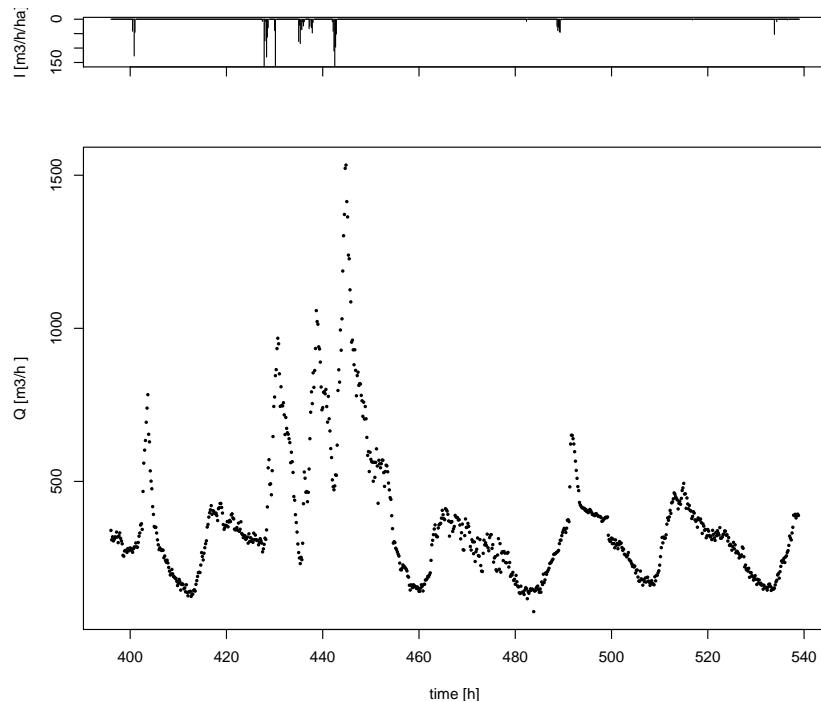
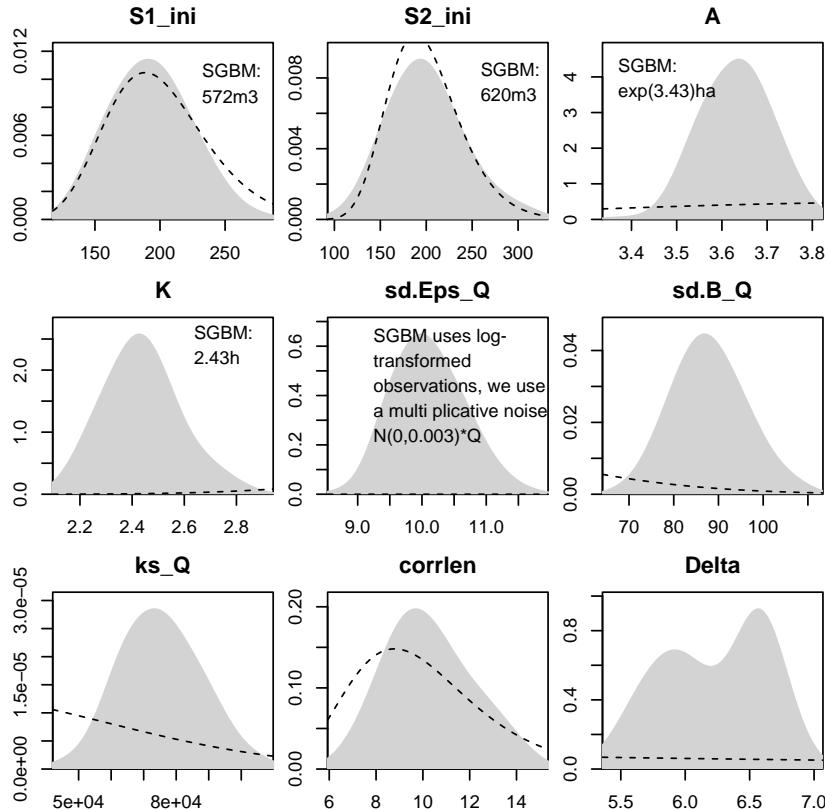
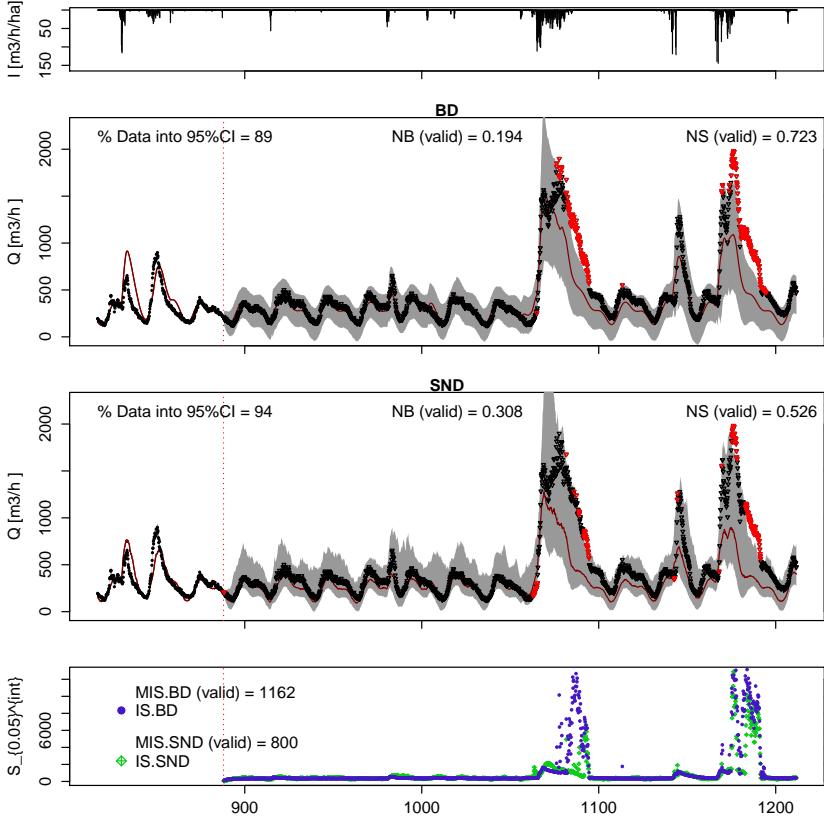


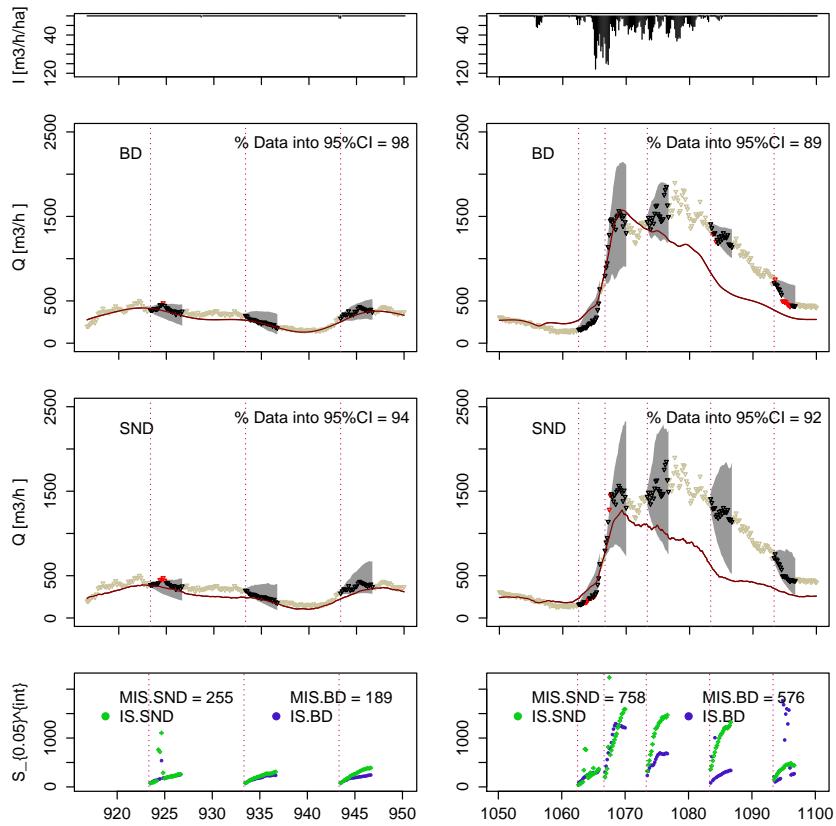
Figure 6. Observed flow ( $Q$ ) and rain intensity ( $I$ ) during the first event used for parameter inference



**Figure 7.** Prior distributions (blue, dashed) and posterior margins (grey) from parameter inference in the BD framework together with corresponding parameter estimates from the SDN framework (mode of the posterior distribution)



**Figure 8.** Simulation results for the validation period. From top to bottom: Rain intensity  $I$ ; simulations using the BD approach (95% prediction intervals (grey), observations (dots), median of simulations (red line)); simulations using the SND approach; interval skill scores  $IS$  for the simulated period together with mean value  $MIS$ .



**Figure 9.** On-line forecast results for 2 sample periods. From top to bottom: Rain intensity  $I$ ; 20-step flow forecasts for the BD approach (95% prediction intervals (grey), observations (dots), median for pure simulation (no updating to data, red line)); 20-step flow forecasts for the SDN approach; interval skill score  $IS$  for the different forecast horizons together with mean value  $MIS$ .



PAPER E

# Investigating the use of probabilistic forecasts for RTC of urban drainage systems - A Layout for Probabilistic Online Forecasting of Sewer Flows

---

**Authors:**

Roland Löwe, Luca Vezzaro, Peter Steen Mikkelsen, Morten Grum, Henrik Madsen

**in preparation for:**

*Environmental Modelling and Software*, 2014.



# Investigating the use of probabilistic forecasts for RTC of urban drainage systems - A Layout for Probabilistic Online Forecasting of Sewer Flows

Roland Löwe<sup>1</sup>, Luca Vezzaro<sup>2,3</sup>, Peter Steen Mikkelsen<sup>2</sup>, Morten Grum<sup>3</sup>,  
Henrik Madsen<sup>1</sup>

## Abstract

We present a framework for generating probabilistic on-line runoff forecasts that can be used in the context of real-time control of urban drainage systems. The forecasts are generated using linear reservoir cascades in a stochastic grey-box model framework with state-dependent description of forecast uncertainties. The grey-box approach provides a flexible tool for dynamic modelling of forecast uncertainties and at the same time permits us to use statistical techniques such as state updating and automated parameter estimation.

This article focuses on the on-line implementation of the forecasting layout, in particular probabilistic forecast generation using radar rainfall measurements. Forecast performance is evaluated in six urban catchments during eight sample events and high skill is demonstrated for forecasts of runoff volume on a horizon of 120 minutes. A benchmark forecast approach which applies a maximum a posteriori method for recalibration of the model parameters at every time step is consistently outperformed by the stochastic grey-box models. Making model parameters time varying using the extended Kalman filter did, on the other hand, in most cases not prove successful for the considered periods of one to few days.

We conclude that the proposed framework is suitable for on-line runoff forecasting in urban catchments and can be applied even if only indirect measurements for catchment outflow are available. The stochastic models act as software sensors, extracting relevant information from the noisy data. In practice, to improve forecast reliability, the very simple model structure should be improved for each

---

<sup>1</sup>Department of Applied Mathematics and Computer Science, Technical University of Denmark, Matematiktorvet, B322, DK-2800 Kgs. Lyngby, Denmark

<sup>2</sup>Department of Environmental Engineering, Technical University of Denmark, Miljøvej, B115, DK-2800 Kgs. Lyngby, Denmark

<sup>3</sup>Krüger AS, Gladsaxevej 363, DK-2860 Søborg, Denmark

considered catchment and forecast uncertainties should be modelled depending on the rainfall forecasts.

# **Investigating the use of probabilistic forecasts for RTC of urban drainage systems - A Layout for Probabilistic On-line Forecasting of Sewer Flows**

4

5 Roland Löwe<sup>a,\*</sup>, Luca Vezzaro<sup>b</sup>, Peter Steen Mikkelsen<sup>b</sup>, Morten Grum<sup>c</sup>, Henrik Madsen<sup>a</sup>

6

<sup>7</sup> \*correspondence to:rolo@dtu.dk, +45-45253399

<sup>8</sup> <sup>a</sup>Department of Applied Mathematics and Computer Science, Technical University of Denmark (DTU)

9 Compute), Matematiktorvet B322, Kgs. Lyngby, 2800, Denmark, [rolo@dtu.dk](mailto:rolo@dtu.dk), [hmad@dtu.dk](mailto:hmad@dtu.dk)

<sup>b</sup>Department of Environmental Engineering, Technical University of Denmark (DTU Environment),

11 Bygningstorvet B115, Kgs. Lyngby, 2800, Denmark, [luve@env.dtu.dk](mailto:luve@env.dtu.dk), [psmi@env.dtu.dk](mailto:psmi@env.dtu.dk)

12 <sup>c</sup>Krüger A/S, Veolia Water Solutions and Technologies, Gladsaxevej 363, Søborg, 2860, Denmark,

13 [mg@kruger.dk](mailto:mg@kruger.dk)

14

15 *ABSTRACT*

16 We present a framework for generating probabilistic on-line runoff forecasts that can be used  
17 in the context of real-time control of urban drainage systems. The forecasts are generated  
18 using linear reservoir cascades in a stochastic grey-box model framework with state-  
19 dependent description of forecast uncertainties. The grey-box approach provides a flexible  
20 tool for dynamic modelling of forecast uncertainties and at the same time permits us to use  
21 statistical techniques such as state updating and automated parameter estimation.

22 This article focuses on the on-line implementation of the forecasting layout, in particular  
23 probabilistic forecast generation using radar rainfall measurements. Forecast performance is  
24 evaluated in six urban catchments during eight sample events and high skill is demonstrated  
25 for forecasts of runoff volume on a horizon of 120 minutes. A benchmark forecast approach

26 which applies a maximum a posteriori method for recalibration of the model parameters at  
27 every time step is consistently outperformed by the stochastic grey-box models. Making  
28 model parameters time varying using the extended Kalman filter did, on the other hand, in  
29 most cases not prove successful for the considered periods of one to few days.

30 We conclude that the proposed framework is suitable for on-line runoff forecasting in urban  
31 catchments and can be applied even if only indirect measurements for catchment outflow are  
32 available. The stochastic models act as software sensors, extracting relevant information from  
33 the noisy data.

34 In practice, to improve forecast reliability, the very simple model structure should be  
35 improved for each considered catchment and forecast uncertainties should be modelled  
36 depending on the rainfall forecasts.

37

38

39 ***KEYWORDS***

40 grey-box model, probabilistic forecasting, real-time control, urban drainage, radar rainfall,  
41 storm water management

42

43 Received

44 2014/02/28

45

46

47

48

49 **1 INTRODUCTION**

50 This study investigates the implementation of probabilistic forecast models for runoffs from  
51 urban catchments in an on-line setting. This development is motivated by recent  
52 developments in real-time control of urban drainage systems. Traditionally, implementation  
53 of real-time control (RTC) in sewer systems has been based on rules derived from previous  
54 experience with a given sewer system (Schütze et al., 2002; Seggelke et al., 2013), whereas  
55 more recent approaches rely on an on-line optimisation of the system based on current states  
56 (Pabst et al., 2011) and in some cases also runoff forecasts (Pleau et al., 2005; Puig et al.,  
57 2009; Vezzaro and Grum, 2014a).

58

59 Traditionally, urban hydrologists simulate the drainage system with complex distributed  
60 models such as MOUSE, SWMM or Infoworks and apply simplified (conceptual) models if  
61 computational requirements do not permit the use of distributed models. In an on-line context,  
62 the computational burden is clearly a limiting factor for the applied forecasting models.  
63 Forecasts over several hours need to be generated within time steps of several minutes and a  
64 multitude of model runs may be required as part of a numerical optimisation.

65

66 Moreover, following the discussions in Nash and Sutcliffe (1970) and in Harremoës and  
67 Madsen (1999), we argue that models applied for forecasting should be simple but stochastic.  
68 Significant uncertainty of runoff forecasts results from rainfall forecasts that may be biased  
69 and delayed in time (Achleitner et al., 2009; Borup et al., 2013; Thorndahl and Rasmussen,  
70 2013b). The model that physically describes the system best is then not necessarily the one  
71 that minimizes the forecast error and the model parameters should be automatically tuned for  
72 the forecasting purpose given a certain type of rainfall input. Also, techniques such as state  
73 updating and statistical parameter validation are easily implemented on simplified models but

74 difficult to handle for complex models. Consequently, we focus on conceptual models in our  
75 forecasting approach.

76

77 Urban catchments are characterised by fast runoff reactions. Timescales typically range from  
78 the order of 10 minutes for basin filling to 1 day for the emptying of basins. The short  
79 timescales make it appealing to include short-term runoff forecasts in the range of a few hours  
80 into the control system. At the same time, the rapid changes in the flow regime pose a major  
81 challenge to the forecasting system. This may be one of the reasons that actual operational  
82 forecasting systems for urban drainage systems are rare in the literature. Pleau et al. (2005)  
83 and Ocampo-Martínez and Puig (2010) both describe the application of conceptual models for  
84 forecasting in the context of RTC of sewer systems, whereas Cassar and Verworn (1999)  
85 describe the use of distributed models in an on-line setting. Neither author provides an  
86 evaluation of forecast performance.

87

88 Thorndahl et al. (2013a) evaluate flow forecasts generated using radar rainfall forecasts and  
89 auto-calibrated conceptual models in a large urban catchment (total area 7700ha) and find  
90 reasonable uncertainties for forecasts of runoff volume up to 2 hours ahead. This is in line  
91 with results obtained by Löwe et al. (2014b) using rain gauge input for two urban catchments  
92 (total area 1300 and 3000ha) and lead times of 100 minutes. Achleitner et al. (2009) consider  
93 flow forecasts generated using conceptual models and radar rainfall forecasts. The authors  
94 conclude that forecast uncertainty becomes unacceptable for horizons larger than 90 minutes.  
95 Liguori et al. (2012) generate flow forecasts from a distributed model using combined radar  
96 and weather model rainfall forecasts. The authors find a tendency for the forecasted flows to  
97 underestimate simulated flows but do not provide a general evaluation on forecast quality.  
98 Thorndahl and Rasmussen (2013b) evaluate the quality of runoff volume forecasts generated

99 using radar rainfall forecasts and an auto-calibrated distributed sewer model. For 50 events in  
100 an 80ha urban catchment, the authors find acceptable forecast uncertainties for lead times of  
101 30 and 60 minutes, whereas lead times of 120 minutes yield almost no forecast value, even  
102 when considering only the integrated runoff volume over the whole forecast horizon.

103

104 We do not account for rain forecast uncertainty in this work but assume the future rainfall  
105 known and focus on model structure uncertainties and the implementation of probabilistic  
106 forecast models. However, the study by Thorndahl and Rasmussen (2013b) on a very small  
107 catchment suggests that in an on-line setting, acceptable rainfall forecasts are available for  
108 horizons of up to 60 minutes.

109

110 The value of forecast uncertainty is increasingly recognized in hydrology (Krzysztofowicz,  
111 2001). Accounting for model uncertainty allows us to obtain better parameter estimates  
112 (Reichert and Schuwirth, 2012) and can be expected to yield better results in real-time control  
113 (Schütze et al., 2004). A real-time control framework for urban drainage systems that allows  
114 us to account for forecast uncertainties is described by Vezzaro and Grum (2014a). The value  
115 of accounting for forecast uncertainties in real-time control of urban drainage systems is, to  
116 our knowledge for the first time, demonstrated in the companion paper by Vezzaro et al.  
117 (2014b).

118

119 Drawing on the issues presented above, we believe an on-line runoff forecasting framework  
120 should be based on simplified stochastic models and should provide a quantification of  
121 forecast uncertainties. We have developed such a framework and implemented it in the  
122 probabilistic real-time control setup described by Vezzaro and Grum (2012). We apply  
123 stochastic grey-box models for runoff forecasting. The applicability of these models in the

124 context of urban drainage has in principle been demonstrated by Carstensen et al. (1998) and  
125 their suitability for modelling uncertainties by Breinholt et al. (2011). However, stochastic  
126 grey-box models have not been implemented in an on-line context in urban hydrology before.

127

128 In this article we summarize the methodologies for and experiences in implementing the  
129 stochastic models with on-line (and erroneous) data from a control server, updating model  
130 parameters and generating probabilistic forecasts. As a case study, we apply the models to six  
131 subcatchments in the Lynetten catchment in Copenhagen and evaluate the quality of the  
132 generated forecasts on a 120min horizon.

133

134 As a benchmark, we consider auto-calibrated conceptual forecast models that are currently  
135 implemented operationally for real-time control in the catchment.

136

137 In the companion paper by Vezzaro et al. (2014b) we subsequently perform a simulation  
138 study to evaluate the effect of the new stochastic runoff forecasting approach on real-time  
139 control. This work demonstrates that a correct quantification of the uncertainty of forecasts of  
140 runoff volume can benefit real-time control. However, it is important that forecast uncertainty  
141 is quantified somewhat reliably. In particular, largely overestimating forecast uncertainty will  
142 have a negative impact on the efficiency of the control scheme.

143

## 144 **2 MATERIAL AND METHODS**

### 145 **2.1 CATCHMENT**

146 We consider the Lynetten catchment as a case study for the implementation of the stochastic  
147 models. This catchment covers the central part of Copenhagen (Denmark) and it has an area  
148 of approximately 76 km<sup>2</sup>. The catchment discharges to the Lynetten wastewater treatment

149 plant (WWTP). A global real-time control algorithm was recently implemented in the  
150 catchment with the aim of minimizing combined sewer overflows (CSO) into the harbour  
151 (Vezzaro and Grum, 2014a). We refer to the companion paper by Vezzaro et al. (2014b) for a  
152 description of the catchment. In the present study we consider the six subcatchments specified  
153 in Table 1.

154

155 Table 1. Considered control points with reduced area of associated subcatchment, storage  
156 volume and available runoff measurements

Subcatchment	Reduced Catchment Area [ha]	Runoff measurement
East Amager (EAM)	228	level in basin, basin outflow
Colosseum (COL)	211	level in basin, basin outflow
Kloevermarken (KLO)	777	outflow from pumping station
Lersoeledning (LER)	733	level in storage pipe, level in downstream sewer
Strandvaenget (STB)	92	outflow from pumping station
West Amager (WAM)	97	level in basin, basin outflow

157

## 158 **2.2 DATA**

159 The presented study is based on on-line data available in the global control setup since 2012.  
160 As the control framework was only implemented in the course of 2011, rainfall and in-sewer  
161 measurements were not reliable for numerous periods and a manual selection of valid datasets  
162 was performed to demonstrate the applicability of the setup. A time step of 2 minutes was  
163 adopted according to the forecasting and control time step currently applied in the on-line  
164 system. The maximal forecast horizon considered is 120 minutes or 60 time steps.

165 **2.2.1 Rainfall Measurements**

166 Radar rainfall measurements were applied to generate flow forecasts. The measurements are  
167 derived from a C-band Doppler radar located at Stevns approximately 50 km south of the  
168 catchment. The radar is operated by the Danish Meteorological Institute (DMI). Quantitative  
169 precipitation estimates are derived from the radar data using the Marshall and Palmer  
170 relationship. The precipitation estimates are mean-field bias adjusted according to six rain  
171 gauges in the catchment based on daily accumulations (Thorndahl et al., 2013a).

172

173 Radar measurements are available with a temporal resolution of 10 minutes. The radar  
174 measurements on a 2km by 2km grid are averaged to a mean areal rainfall for each  
175 subcatchment in the control setup according to the portion of a pixel that is contained in the  
176 subcatchment. For this study, we had access to the processed radar measurements.

177

178 Rainfall forecasts from the radar data are used in the control setup but are not available as  
179 historical data for the period before autumn 2013. Runoff forecasts were therefore in the  
180 following generated under the assumption that the future rainfall is known and that it  
181 corresponds to the future rainfall measurements. This does not hinder the principal  
182 applicability of the presented approach but will, of course, affect the uncertainty of runoff  
183 forecasts in practice.

184

185 **2.2.2 In Sewer Measurements**

186 Runoff measurements for the six subcatchments are available with a 2-minute resolution  
187 according to the control time step. The objective of the runoff forecasting models is to predict  
188 inflow to the control points. Corresponding observations are required for calibration and

189 updating of the models. In reality, however, the required quantities are in most cases not  
190 observed directly. Table 1 shows the measured values available in each subcatchment.

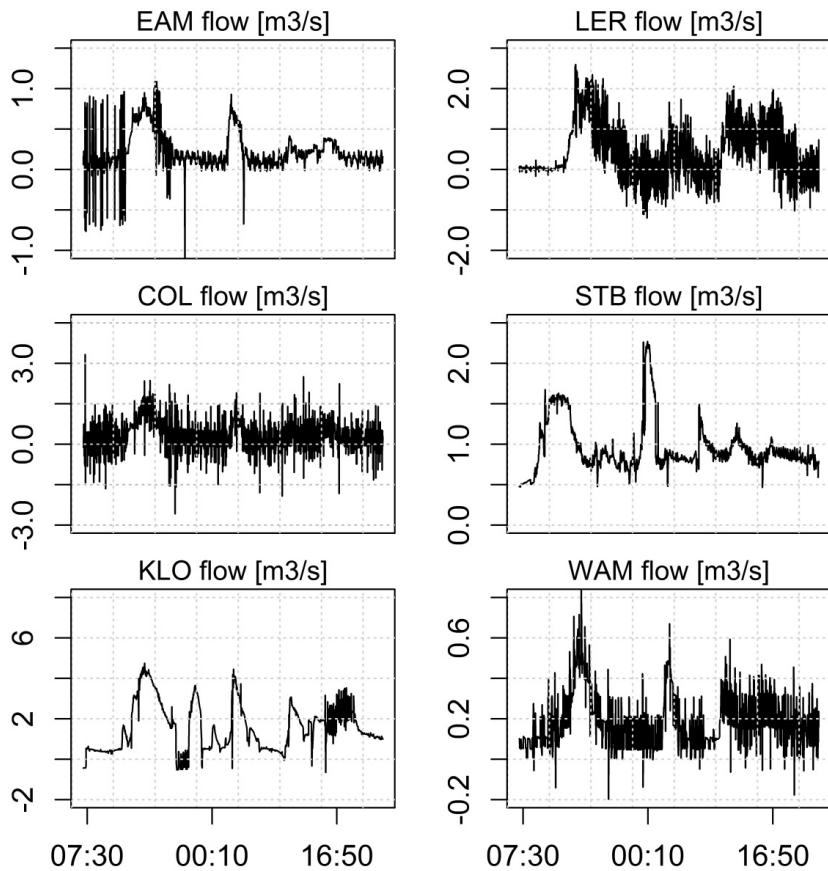
191

192 In the EAM, COL, LER and WAM catchments, level measurements in the storage basin and  
193 outflow measurements from the basin are available. We derived the basin inflow from the  
194 basin water balance. We obtained flow observations as shown in Figure 1.

195

196 The observations are rather noisy as a result of the high temporal resolution. In the LER  
197 subcatchment negative flow observations can often be observed at the end of an event, also  
198 after smoothening. This effect occurs because not all basin outflows are observed during the  
199 emptying phase. This problem also affects the observations in the STB subcatchment as  
200 inflows from this subcatchment are derived as the measured flow from this pumping station  
201 minus inflows from the LER catchment.

202



203

204

205 Figure 1. Flow observations from the six subcatchments for calibration event No. 1

206 **2.2.3 Calibration and Validation Periods**

207 We consider eight rain events in 2012 and 2013 where no more than 10 % of the radar  
 208 observations and the in-sewer measurements of all considered subcatchments are missing. All  
 209 events led to significant runoff and storage filling in the control points. However, none of the  
 210 events was extreme. This kind of rain event is most relevant for real-time control of sewer  
 211 systems because potential for distributing storage and avoiding CSO is available.

212

213 Table 2 gives an overview of the event characteristics derived from the average rainfall over  
214 the KLO subcatchment. We consider a rain event to start if a rain intensity of more than  
215 0.6 mm/h is observed by the radar. The event ends if rain intensities below 0.6 mm/h have  
216 been observed for more than 15 hours.

217

218 Events 1 to 4 are used for model fitting whereas events 5 to 8 are considered only for the  
219 evaluation of forecast performance. According to Arnbjerg-Nielsen et al. (2005), a rain event  
220 with return period of 0.5 years has a maximal intensity of 15.2 mm/h for a duration of 30  
221 minutes. However, these statistics are derived for rain gauge observations, whereas we are  
222 here concerned with radar observations averaged over a catchment with a reduced area of  
223 780 ha.

224

225 Table 2. Characteristics of the considered rain events No. 1 to 8 derived from a spatial  
226 average over the KLO subcatchment

	1	2	3	4	5	6	7	8
Month of Year	Jan	Apr	Jul	Oct	Jan	Nov	Jun	Jul
Total Rainfall	19.7	20.1	9.6	9.0	15.8	8.7	4.1	7.5
[mm]								
Maximal Intensity 10min [mm/h]	5.3	5.4	9.3	2.8	10.9	6.0	3.6	7.3
Maximal Intensity 30min [mm/h]	4.2	3.7	6.8	2.4	9.7	5.3	2.3	6.0
Duration	1690	2540	270	560	650	220	290	680
[min]								

227

228 **2.3 PROBABILISTIC RUNOFF FORECASTING**229 **2.3.1 Model Layout**

230 A simple cascade of three linear reservoirs is applied for runoff forecasting in each  
 231 subcatchment. The same layout is currently applied operationally for deterministic runoff  
 232 forecasting in the considered catchment (Section 2.5). We do not consider more elaborated  
 233 model structures as the purpose of this article is to give a proof of concept.

234

235 The model is implemented as a stochastic grey-box model in a state-space layout as described  
 236 by Breinholt et al. (2011) and shown in state (or system) equations (1) and observation  
 237 equation (2). The model is implemented in the open source software CTSM-R (Juhl et al.,  
 238 2013).

239

$$d \begin{bmatrix} S_{1,t} \\ S_{2,t} \\ S_{3,t} \end{bmatrix} = \begin{bmatrix} A \cdot P + a_0 - \frac{1}{K} S_{1,t} \\ \frac{1}{K} S_{1,t} - \frac{1}{K} S_{2,t} \\ \frac{1}{K} S_{2,t} - \frac{1}{K} S_{3,t} \end{bmatrix} dt + \begin{bmatrix} \sigma_1 S_{1,t} \\ \sigma_2 S_{2,t} \\ \sigma_3 S_{3,t} \end{bmatrix} d\omega_t \quad (1)$$

$$Q_k = Y_k = \frac{1}{K} S_{3,k} + D_k + e_k \quad (2)$$

240

241  $S_1$ ,  $S_2$  and  $S_3$  correspond to the storage states,  $A$  to the impervious catchment area,  $P$  to the  
 242 rain intensity,  $a_0$  to the mean dry weather flow and  $K$  to the travel time constant. The  
 243 uncertainty of model predictions is described in the so-called diffusion term by a vector  
 244 Wiener process  $d\omega_t$ . Considering a time step  $\Delta t$ , an increment  $\Delta\omega_t$  of this process is normally  
 245 distributed with mean 0 and covariance  $diag(\Delta t^2, \Delta t^2, \Delta t^2)$ . The parameters  $\sigma_i$  scale the standard  
 246 deviation of the diffusion process which here increases linearly with the state values  $S_i$ .

247

248 The observation equation (2) relates time-continuous model predictions and flow observations  
249  $Y_k$  at discrete time steps  $k$ . It includes a trigonometric function  $D$  to describe the variation of  
250 dry-weather flows and the observation error  $e_k$  with standard deviation  $\sigma_e$ .

251

252 A Lamperti transformation (Iacus, 2008; Øksendal, 1998) is applied to the state equations (1)  
253 to remove the dependency of the noise description on the state (Breinholt et al., 2011). The  
254 transformation further guarantees positive flow predictions in the layout suggested here. The  
255 Lamperti transformation results in the following new set of equations:

256

$$d \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ Z_{3,t} \end{bmatrix} = \begin{bmatrix} (A \cdot P + a_0) \cdot e^{-Z_{1,t}} - \frac{1}{K} - \frac{\sigma_1^2}{2} \\ (\frac{1}{K} \cdot S_{1,t}) \cdot e^{-Z_{2,t}} - \frac{1}{K} - \frac{\sigma_2^2}{2} \\ (\frac{1}{K} \cdot S_{2,t}) \cdot e^{-Z_{3,t}} - \frac{1}{K} - \frac{\sigma_3^2}{2} \end{bmatrix} dt + \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \end{bmatrix} d\omega_t \quad (3)$$

257 where

258

$$S_{i,t} = \exp(Z_{i,t}) \quad (4)$$

259 The uncertainty defined by the state equations (1) lumps input, structural and parameter  
260 uncertainty into one term. We consider this approach reasonable for on-line forecasting, as the  
261 forecast uncertainty results as the sum of these three components. We can estimate the  
262 uncertainty scaling parameters  $\sigma_i$  as part of the parameter calibration procedure (Kristensen et  
263 al., 2004).

264

265 In four of the six considered catchments, only measurements of the level in the storage basin  
266 and the outflow from the storage basin are available, rather than direct measurements of the  
267 basin inflow (EAM, COL, LER, WAM). The state equations (1) can in this case be extended  
268 with an additional state representing the storage basin. Using the level measurements as  
269 observations in equation (2), the model can then account for the situation where combined  
270 sewer overflow occurs from the basin. The basin outflow is in this case considered as a model  
271 input.

272

273 We have tested the setup with an additional state in the four concerned catchments but could  
274 not obtain improved forecast skill in any of the catchments. Therefore, in the following we  
275 maintain the simpler layout shown in equation (1). In practice, this setup also has the  
276 advantage that short periods of missing values for both level and outflow measurements can  
277 easily be handled on-line by the extended Kalman filter (Section 2.3.2). Using the basin  
278 outflow as a model input, on the contrary, requires routines that fill gaps in the outflow  
279 measurements.

280

### 281 **2.3.2 State Updating**

282 CTSM-R applies the extended Kalman Filter (Jazwinski, 1970; Kristensen et al., 2004) for  
283 updating the state equations (3) when new flow observations are available. The degree of  
284 updating depends on the standard deviations of the model states and observations which are  
285 scaled by parameters  $\sigma_i$  and  $\sigma_e$ . These are part of the parameter estimation routine.

286

287 Missing flow observations are handled in the extended Kalman filter by setting the variance  
288 of the missing observations (equation (2)) to a large value (see Kristensen et al. (2004)). As a  
289 result no updating occurs and the variance of the model states increases.

290

291 **2.3.3 Parameter Estimation**

292 All model parameters, including the variance scaling of the diffusion terms, are determined in  
293 an automatic calibration routine. The objective function is derived to minimize the multistep  
294 forecast error. At each time step, a 60-step-ahead forecast is generated and the continuous  
295 ranked probability score (*CRPS*, Gneiting et al., 2005) of the flow forecasts for every horizon  
296 is minimized as described by Löwe et al. (2014a).

297

298 The “dynamically dimensioned search algorithm” (Tolson and Shoemaker, 2007) is applied  
299 for automatic parameter estimation and offers the possibility to handle unbehavioural  
300 parameter sets. Numerical model failures can occur in stochastic grey-box models for  
301 unsuitable parameter combinations because variance matrices may not be positive semi-  
302 definite, for example. We consider 1500 objective function evaluations and, after finishing,  
303 restart the optimization algorithm from the identified optimum to ensure that a good  
304 parameter set is identified.

305

306 During parameter estimation, gaps in the flow observations in equation (2) are filled by linear  
307 interpolation if the gap is shorter than 10 time steps or 20 minutes. Longer gaps are excluded  
308 from the estimation. Missing values in the rainfall input are set to 0.

309

310 **2.3.4 On-line Forecast Generation**

311 Real-time control requires forecasts of runoff volume for a given horizon. These are derived  
312 as integral over the flow forecasts provided by the model for all forecast time steps up to the  
313 considered horizon. When creating probabilistic forecasts of runoff volume, we need to

314 consider the multivariate distribution of flow forecasts for all lead times and thus describe the  
315 correlation between different forecast horizons.

316

317 Further, when solving the system of stochastic differential equations (1), the forecast  
318 uncertainty cannot generally be described parametrically, for example because state-  
319 dependent noise descriptions as defined in (1) will lead to skewed distributions.

320

321 We adopt a scenario modelling approach for the generation of probabilistic multistep runoff  
322 forecasts in an on-line setting. This approach is computationally feasible and accounts for the  
323 aforementioned issues when generating runoff forecasts. We neglect observation uncertainty  
324 during forecast generation because for RTC we are interested only in the future state of the  
325 system. Probabilistic runoff forecasts are derived through the steps shown below.

326

327 a) At time step  $t$ , perform Kalman filtering with new observation  $Q_t$  and obtain updated  
328 states  $Z_{i,t|t}$  with covariance matrix  $\Sigma_{t|t}$  ( $i=1,2,3$ ).

329 b) Initialize scenarios by creating 1000 multivariate samples from the multivariate  
330 normal distribution  $N(Z_{i,t|t}, \Sigma_{t|t})$  (Gentle, 2009).

331 c) Simulate the transformed state equations (3)  $k=60$  time steps or 120 min ahead and  
332 create 1000 scenarios of state predictions using an Euler-Maruyama scheme (Kloeden  
333 and Platen, 1999) with a simulation time step  $\Delta t=1.2\text{s}$ .

334 d) Convert state prediction from a scenario into flow using (2) but neglecting the  
335 observation error  $e$ . For each scenario determine expected runoff volume for every  
336 forecast horizon  $k=1,\dots,60$ .

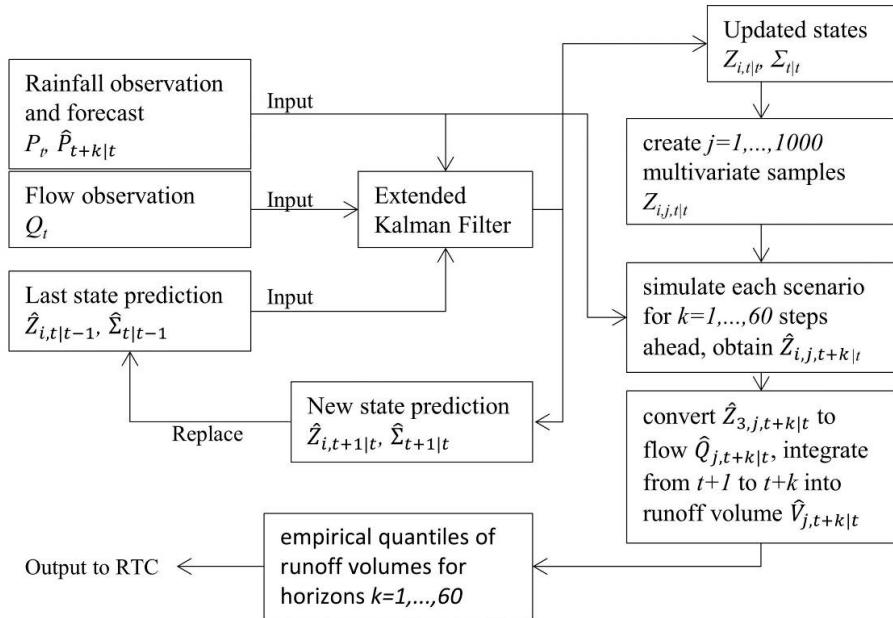
337 e) Evaluate distribution of runoff volumes from 1000 scenarios for every forecast  
338 horizon. Determine quantiles from 1 to 99 % in 2 % resolution.

339

340 This scheme is summarized in Figure 2. In a few cases, the extended Kalman filtering step  
 341 fails, for example due to extreme changes in the observations. In this case, in the next time  
 342 step the model is reinitialized from the stored states of the previous time step.

343

344 The model simulations in step c) are performed on the Lamperti transformed state equations  
 345 (3). This simplifies the solution of the stochastic differential equations as the diffusion process  
 346 then has a time-constant variance (see Kloeden and Platen (1999)). The physical part of the  
 347 stochastic differential equation, however, becomes more complex and more likely to be stiff.  
 348 We have in this work obtained stable solutions using a simple Euler-Maruyama scheme for  
 349 simulating the stochastic differential equations. For more complex model structures, implicit  
 350 solvers (Higham et al. (2002)) or predictor-corrector schemes (Bruti-Liberati and Platen  
 351 (2008)) may be more appropriate.



352

353 Figure 2. On-line operation of probabilistic runoff forecasting models as shown in (3), 60 time  
 354 steps correspond to a forecast horizon of 120 min

355

356 **2.3.5 Time Varying Parameters**

357 A simple conceptual model obviously cannot perfectly describe the runoff reaction of a  
358 catchment. An approach to reduce such structural model errors is to use the Kalman filtering  
359 setup to allow model parameters to vary in time by including them in (3) as model states  
360 instead of constants. Such approaches were suggested for conceptual models for example by  
361 Da Ros and Borga (1997) and Rajaram and Georgakakos (1989), but discouraged by Todini  
362 (2006) on the grounds of providing unstable solutions.

363

364 Equation (5) shows an according model layout (after Lamperti transformation), in this case  
365 including the dry-weather flow  $a_0$  as model state. This case is relevant in particular for the  
366 STB catchment where dry weather flows vary strongly.

367

368 In (5)  $I$  is an index which is 1 in the updating step of the model and 0 when generating  
369 scenario forecasts. In the forecasting step, it is desirable to use the last-known estimate of  $a_0$   
370 which is based on the current flow observation. Allowing  $a_0$  to vary also when generating  
371 forecasts would increase the uncertainty of the forecasts, but not provide further knowledge  
372 about the parameter.

373

$$d \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ Z_{3,t} \\ a_0 \end{bmatrix} = \begin{bmatrix} (A \cdot P + a_0) \cdot e^{-Z_{1,t}} - \frac{1}{K} - \frac{\sigma_1^2}{2} \\ (\frac{1}{K} \cdot S_{1,t}) \cdot e^{-Z_{2,t}} - \frac{1}{K} - \frac{\sigma_2^2}{2} \\ (\frac{1}{K} \cdot S_{2,t}) \cdot e^{-Z_{3,t}} - \frac{1}{K} - \frac{\sigma_3^2}{2} \\ 0 \end{bmatrix} dt + \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \sigma_3 & \\ & & & \sigma_4 \cdot I \end{bmatrix} d\omega_t \quad (5)$$

374

375 **2.4 FORECAST VERIFICATION**

376 We here introduce the tools used for evaluating the runoff forecast quality. We focus on  
377 forecasts of runoff volume with a lead time of 60 time steps or 120 minutes. This is the  
378 longest forecast horizon considered in the global real-time control layout suggested by  
379 Vezzaro and Grum (2014a) and consequently the case which poses the biggest challenge for  
380 the simple forecasting models in this setup.

381

382 We consider point and probabilistic forecast quality in the evaluation. A forecast skill is  
383 commonly defined as (Wilks, 2011)

384

$$SS = \frac{A - A_{ref}}{A_{perf} - A_{ref}} \quad (6)$$

385 Where  $A$  is a measure of forecast accuracy,  $A_{ref}$  the accuracy obtained by a reference forecast  
386 and  $A_{perf}$  the accuracy obtained for a perfect forecast.

387

388 For evaluating the point forecast skill, we consider the 50 % quantile of the probabilistic  
389 forecasts of runoff volume. Considering  $A$  as the squared error of runoff volume forecasts and  
390  $A_{ref}$  as the squared error of a forecast obtained by averaging all flow observations during an  
391 event, we obtain the Nash-Sutcliffe Efficiency  $NSE$  (Nash and Sutcliffe, 1970). This score  
392 function, however, is problematic in that it assumes a very weak reference and may therefore  
393 lead the user to falsely conclude on forecast skill of the model (see also Smith et al. (2012)).

394

395 We find a stronger reference  $A_{ref}$  as the squared error of a runoff volume forecast derived from  
396 the last known flow observation. Considering a horizon of 60 time steps and flow  
397 observations  $Q_t$  that are available for  $N$  observations in intervals  $\Delta t=2$  min, this results in

$$A_{ref} = \sum_{t=1}^N \left( \sum_{i=1}^{60} Q_t \cdot \Delta t - \sum_{i=1}^{60} Q_{t+i} \cdot \Delta t \right)^2 \quad (7)$$

400 Such a reference will, however, perform badly for very noisy flow observations (Figure 1).  
 401 We therefore replace the last known flow observation  $Q_t$  in equation (7) with the  
 402 corresponding exponentially smoothed flow value (Brown and Meyer, 1961). We adjust the  
 403 smoothing parameter for each of the six subcatchments, such that the squared error described  
 404 by equation (7) is minimized during the four rain events used for model calibration. We  
 405 denote the resulting score function as smoothed persistence index  $PI$  (see Bennett et al. (2013)  
 406 for the persistence index).

408 In a probabilistic sense, it is desirable that the runoff forecasts are reliable. This means that for  
 409 a given coverage, rate  $\alpha \%$ ,  $\alpha \%$  of the observations should be included in the corresponding  
 410 prediction interval. This property of the probabilistic forecasts can be assessed by comparing  
 411 predicted and observed coverage rates in reliability diagrams (Murphy and Winkler, 1977).  
 412 Such diagrams are easier to understand and simplify the communication with practitioners  
 413 and are therefore here preferred over the probability integral transform used by, for example,  
 414 Hemri et al. (2013) and Renard et al. (2010).

416 Finally, given a reliable probabilistic forecast, it is desirable it has a high resolution, i.e. is as  
 417 sharp as possible. A common measure is the sharpness or average width of a  $\alpha \%$  prediction  
 418 interval. Jin et al. (2010) normalize this measure with the observation to obtain the average  
 419 interval width  $ARIL$ , which we here apply for the 90 % prediction interval as a measure of  
 420 forecast uncertainty.

421

$$ARIL = \frac{1}{N} \sum_{t=1}^N \frac{\hat{V}_{95\%,t+60|t} - \hat{V}_{5\%,t+60|t}}{V_{t+60|t}} \quad (8)$$

422 In (8)  $\hat{V}_{95\%,t+60|t}$  and  $\hat{V}_{5\%,t+60|t}$  correspond to the 5 % and 95 % quantiles of the probabilistic  
423 runoff volume forecasts generated at time step t for a lead time of 60 time steps.  $V_{t+60|t}$   
424 corresponds to the associated observation which is derived by integrating the flow  
425 observations  $Q_{t+1}$  to  $Q_{t+60}$ .

426

427 **2.5 FORECAST BENCHMARK**

428 For runoff forecasting, the real-time control setup in the considered Lynetten catchment  
429 (Vezzaro et al., 2014a) currently applies deterministic conceptual models that, equivalently to  
430 equation (1) employ a series of three linear reservoirs. The catchment area parameter  $A$ , the  
431 time constant  $k$  and the mean dry weather flow  $a_0$  of these models are autocalibrated at every  
432 control time step (2 minutes) against the last 12 hours of observations from the corresponding  
433 catchment as described by Lund et al. (2014).

434

435 The autocalibration applies a maximum a posteriori (MAP) approach that considers prior  
436 normal distributions for the three parameters. The mean of the distribution corresponds to the  
437 optimal parameter identified in the previous time step, while the variance needs to be  
438 specified by the user (Lund et al. (2014)).

439

440 For the four validation events, runoff forecasts over a 120 minute horizon were available for  
441 these auto-calibrated models. We included them in the evaluation of point forecast skill as a  
442 benchmark (*BE*) for the new stochastic forecast models.

443

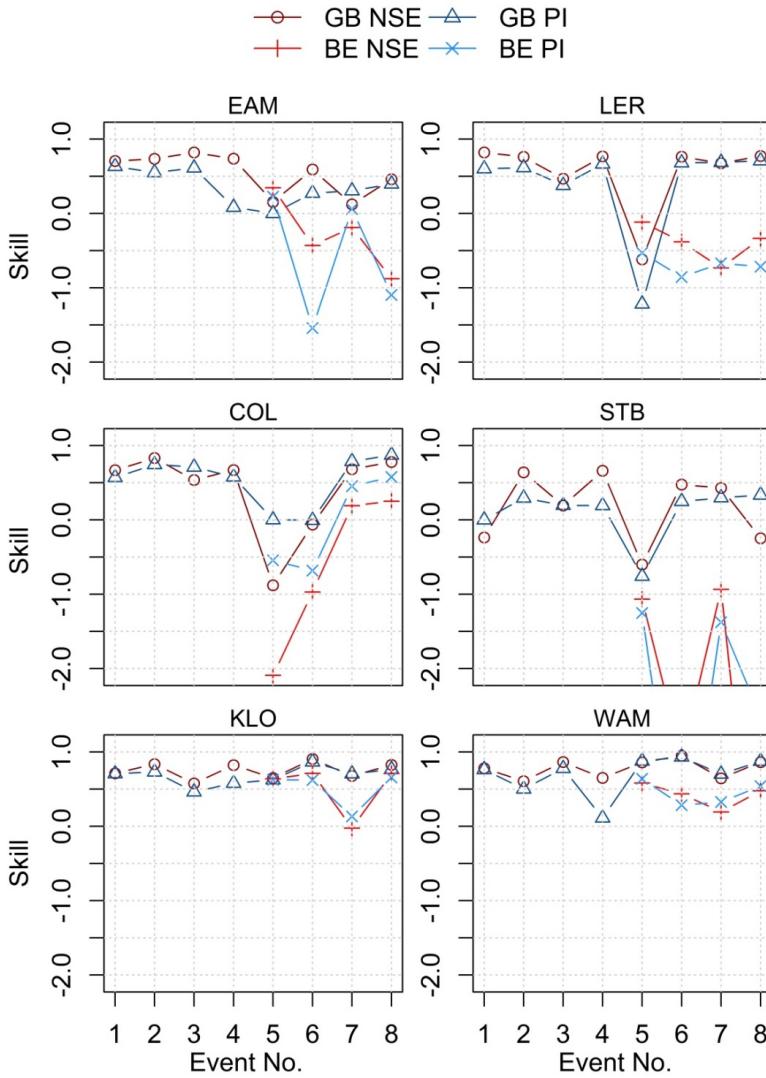
444 **3 RESULTS**

445 We analyse point and probabilistic forecast skill of forecasts of runoff volume on a horizon of  
446 2 hours assuming a perfect rainfall forecast derived from the C-band radar measurements. In  
447 the analysis we exclude periods where the observed runoff volume over a 2 hour horizon is  
448 negative to avoid an influence from false flow measurements, in particular in the LER  
449 catchment (see Section 2.2.2).

450

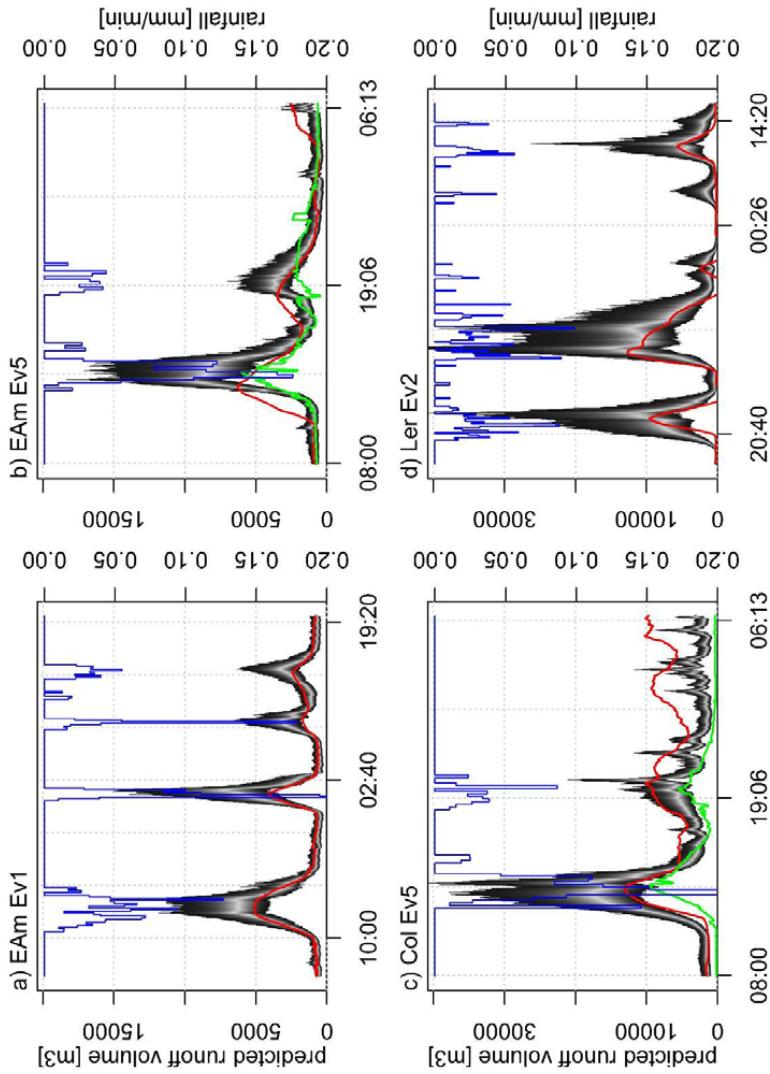
451 **3.1 POINT FORECAST SKILL**

452



453

454 Figure 3. Point forecast skill for 120-minute forecasts of runoff volume during events 1 to 8 in  
 455 the six subcatchments using grey-box (GB) forecast models with time constant parameters  
 456 and the benchmark approach (BE). Results for the grey-box models are based on 50%  
 457 quantiles.



458

459 Figure 4. Predicted distribution of runoff volumes generated by stochastic grey-box models at  
 460 a time step  $t$  for lead time  $t+60$  together with point prediction from benchmark BE for  $t+60$   
 461 (green), observed runoff volume for  $t+60$  (red) and rain intensity measured at  $t$  (blue); a)  
 462 Subcatchment EAM, Event 1, b) EAM, Event 5, c) COL, Event 5, d) LER, Event 2

463

464 Figure 3 shows the average point forecast skill for 120-minute forecasts of runoff volume  
 465 during each event. Figure 4 shows the predicted runoff volumes for a 120-minute horizon

466 together with the rain intensity. Note, that in Figure 4 the runoff volume shown at time  $t$  is an  
467 integral of the flow observations or predictions from  $t+1$  to  $t+60$  while the depicted rainfall is  
468 a measurement at time  $t$  and thus lagged to the runoff values.

469

470 We observe that the stochastic grey-box models (GB) generally provided runoff forecasts  
471 with high skill and during the validation events consistently outperformed the autocalibrated  
472 benchmark models (BE) which are currently implemented in practice.

473

474 Effects from data errors were identified in several cases. During event 5, the radar  
475 measurements suggest high rain intensities in the northeastern catchments (EAM, COL, LER,  
476 STB) which are not reflected in the runoff measurements to the same degree. As a result, the  
477 forecast models overestimated the observed runoff in all 4 catchments.

478

479 In the COL catchment, flow measurements are corrupted during the end of events 5 and 6.  
480 The flow measurements oscillate strongly and on average suggest an increased runoff which  
481 is not supported by the rainfall measurements. The latter can be noted from the red line in the  
482 second half of the rain event depicted in Figure 4c. Excluding these periods from the  
483 evaluation led to much higher forecast skill values for this catchment during events 5 and 6  
484 ( $NSE=0.50$  and  $0.65$ ,  $PI=0.72$  and  $0.62$ ).

485

486 A general drop in forecast performance was observed in the EAM catchment, when  
487 comparing the validation events 5-8 to the calibration events 1-4. The catchment tends to  
488 react faster during the validation events which can be observed from the rapidly increasing  
489 runoff in Figure 4c, which was not captured by the forecast model. This behaviour was not

490 properly represented in the calibration dataset, where the catchment reacts slower despite high  
491 rain intensities (Figure 4a).

492

493 For the other catchments, there was no general trend for a difference in forecast performance  
494 during calibration and validation periods. However, generally low forecast skills were  
495 observed in the STB catchment due to an insufficient representation of dry weather flows and  
496 slow runoffs in the model (c.f. Section 3.3).

497

498 In Figure 4 and Figure 8 we observe that the autocalibrated benchmark forecasts show strong  
499 variations because the model parameters vary strongly. Such issues are also discussed by  
500 Lund et al. (2014). Although a problem in the operational implementation of the benchmark  
501 models is suggested by the systematic underestimation of dry weather flows in the COL  
502 catchment, the graphical analysis (Figure 4 and Figure 8) and the lower score values in all  
503 catchments (Figure 3) indicate that the benchmark approach is less robust than the stochastic  
504 grey-box models.

505

### 506 **3.2 PROBABILISTIC FORECAST QUALITY**

507 Figure 5 compares predicted and observed coverage rates for-120 minute forecasts of runoff  
508 volume. Figure 6 shows the average interval length for the corresponding 90 % prediction  
509 intervals. In the COL catchment, for this analysis we have excluded the periods of faulty flow  
510 measurements during events 5 and 6 (c.f. Section 3.1).

511

512 In terms of reliability of the probabilistic forecasts, again we did not observe a generally  
513 different behaviour between the calibration and validation period. An exception, however, is  
514 the EAM catchment, where the low point forecast quality during the validation period also  
515 reflects on the reliability of the runoff forecasts (Figure 5).

516

517 We did, however, note a general tendency of the runoff forecasts to be unreliable. A 90 %  
518 prediction interval typically includes approximately 70 % of the runoff observations. One  
519 issue to be considered here is the parameter estimation approach as discussed in Section 4.  
520 However, we can also note that forecast uncertainties were typically underestimated in the  
521 beginning of a rain event (Figure 4b, c and d). This suggests that the applied, linearly state  
522 dependent structure for modelling forecast uncertainties is too simple.

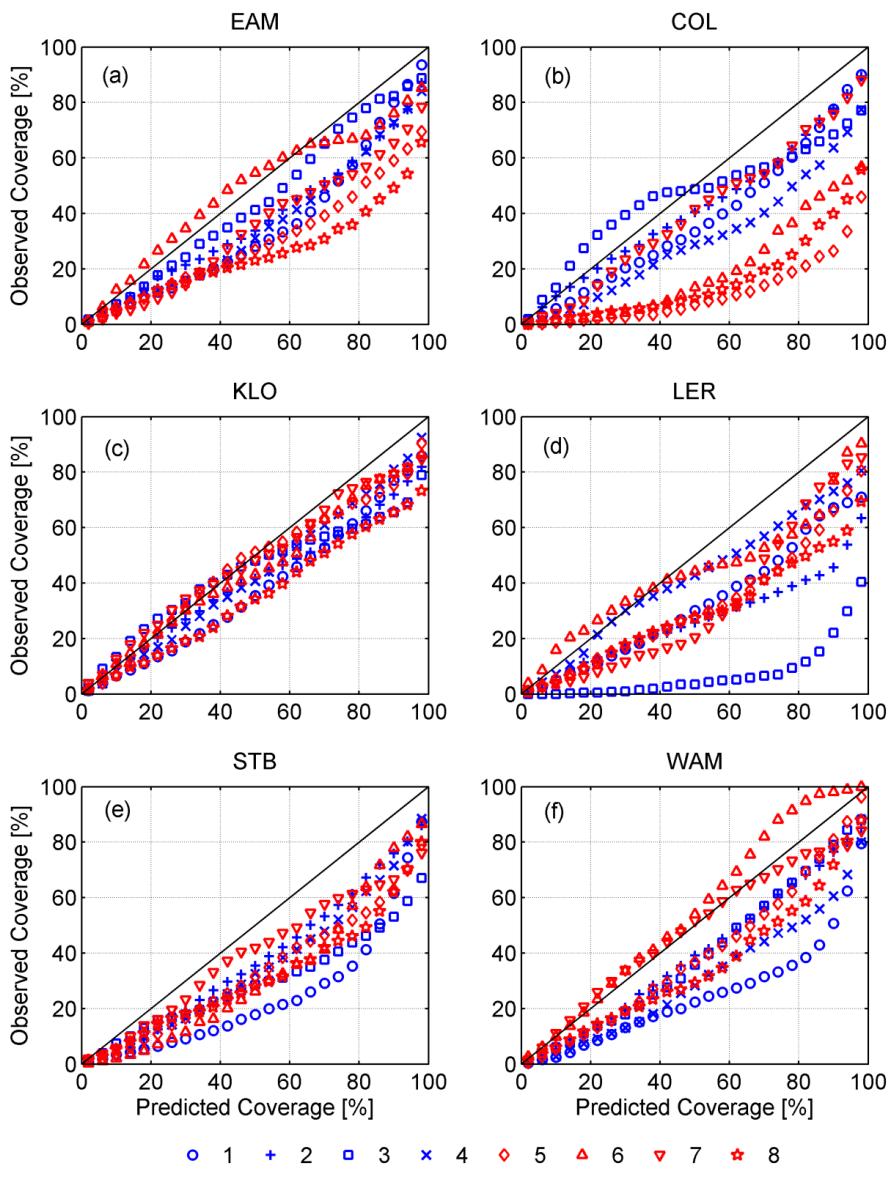
523

524 The 90 % runoff volume prediction interval for lead times of 120 minutes typically has a  
525 width of approximately 50 to 100 % of the corresponding observations. In the KLO  
526 subcatchment very small dry weather runoff volume observations close to 0 led to increased  
527 ARIL values during event 2.

528

529 In the LER subcatchment, forecast uncertainties were much higher than in the other  
530 subcatchments. This was caused by the unreliable flow measurements during low flows (c.f.  
531 Section 2.2.2). The simulated model states  $S$  in these periods are very small. Applying a  
532 linearly state dependent uncertainty description as shown in equation (1) therefore requires  
533 large scaling parameters  $\sigma_i$  to account for such errors during dry weather. As a result we  
534 obtained large forecast uncertainties during rain periods. This behaviour can be observed in  
535 Figure 4d. Again, the applied model structure for describing forecast uncertainties proved to  
536 be too simple.

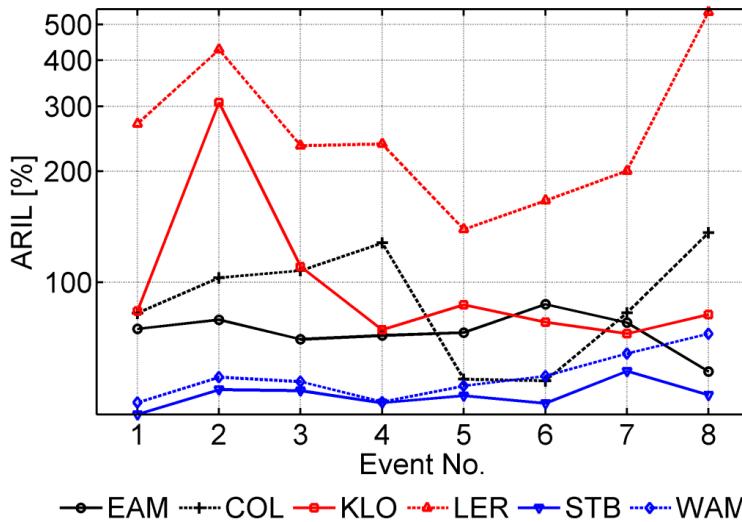
537



538

539 Figure 5. Observed vs. predicted coverage rates for 120-minute forecasts of runoff volume  
 540 during events 1 to 8 in the six subcatchments using forecast models with time constant  
 541 parameters. Calibration events are marked blue, while validation events are marked red.

542



543

544

545 Figure 6. *ARIL* for 90 % intervals of 120-minute forecasts of runoff volume during events 1 to  
 546 8 in the six subcatchments using forecast models with time constant parameters.

547

### 548 3.3 FORECAST QUALITY TIME VARYING PARAMETERS

549 We evaluated the effect of including time-varying model parameters in the setup of the  
 550 stochastic forecast models. For this purpose, we included the physical model parameters  $A$ ,  $k$   
 551 and  $a_0$  as states in the model structure as demonstrated in equation (5) and estimated the  
 552 model parameters (including time constant variance parameters for the new parameter states)  
 553 as described in Section 2.3.3.

554

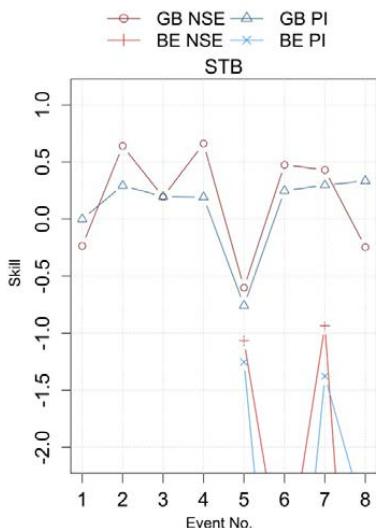
555 The resulting variance estimates for the new parameter states, however, were generally very  
 556 small for the  $A$  and  $k$  parameters and the parameter states in an on-line setting hardly deviated  
 557 from their starting values. Also a time-variable  $a_0$  parameter did not lead to, in all but one  
 558 case, improved runoff forecasts skills as compared to the values shown in Figure 3.

559

560 An exception from this behaviour is the dry weather flow parameter  $a_0$  in the STB catchment.  
561 The deficiencies of the model in terms of capturing the catchment's behaviour during dry  
562 weather and slow runoff situations can partly be compensated for by this parameter. We also  
563 obtained improved parameter estimates for the area  $A$  and time constant  $k$  parameters and as a  
564 result improved forecast skill values as depicted in Figure 7.

565

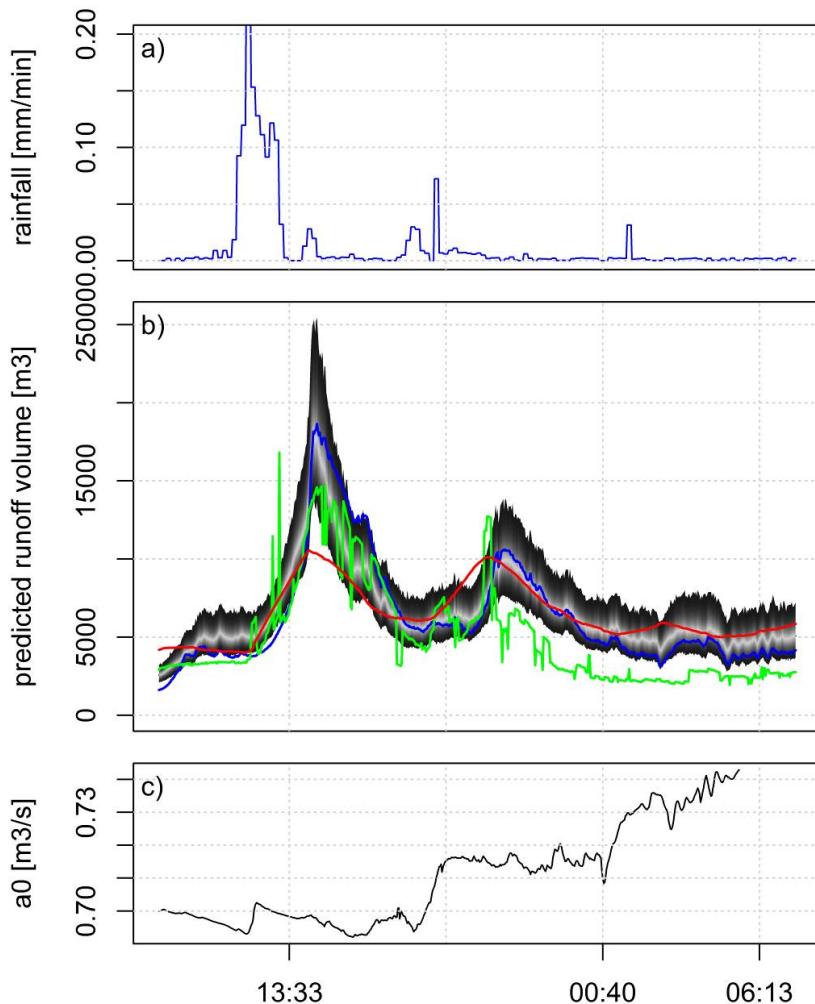
566 Figure 8 compares the runoff forecasts generated by a model with a constant and a model with  
567 time-varying  $a_0$  parameter during event 5 in the STB catchment. In the end of the event the  
568 parameter state starts to increase as a result of the underestimation of slow runoffs by the  
569 model. The increase is limited due to the short simulation time in this case. The biggest range  
570 of variation observed for this parameter during the considered 8 events was  $0.2\text{m}^3/\text{s}$  during  
571 event 4.



572

573 Figure 7. Point forecast skill for 120-minute forecasts of runoff volume during events 1 to 8 in  
574 the STB subcatchment using a forecast model with time-adaptive dry weather flow  $a_0$ .  
575 Results for the grey-box model are based on 50% quantiles.

576



13:33 00:40 06:13

579 Figure 8. Event 5 in the STB subcatchment, a) observed rain intensity,  
 580 b) 120-min runoff  
 581 volume forecast for time adaptive  $a_0$  (grey, 1-99% quantile), runoff volume forecast using  
 582 constant  $a_0$  (blue, 50% quantile), benchmark (green) and observation (red), c) time variable  
 583 estimates of dry weather flow  $a_0$

585 **4 DISCUSSION**

586 **4.1 FORECAST RESULTS IN AN ON-LINE SETTING**

587 We could generally obtain suitable forecasts of runoff volume with high skill over a horizon  
588 of 120 minutes (or 60 time steps) with the stochastic grey-box approach applied in this work  
589 although only a simplistic conceptual model structure is considered. As with any model,  
590 however, the quality of the considered measurements can have severe effects on the forecast's  
591 skill.

592

593 The derivation of basin inflow observations from level and basin outflow measurements led to  
594 noisy flow observations (Figure 2), but was not problematic in general. On the other hand,  
595 positively (as in the COL catchment in the end of events 5 and 6) or negatively (as in the LER  
596 catchment in the end of most rain events) biased observations did have a negative impact on  
597 the forecast models.

598

599 In an on-line setting, the positively biased observations during validation events in the COL  
600 catchment did lead to an updating of the forecast model states to higher values and as a result  
601 to wrong forecasts of runoff volume. The negative flow observations in the end of most rain  
602 events in the LER catchment (Figure 5d) pose a problem to parameter estimation in  
603 combination with the considered structure of the uncertainty model (Section 4.2).

604

605 Both faulty rainfall inputs and in-sewer observations need to be identified in a pre-processing  
606 of the data (Campisano et al., 2013). A correctly identified stochastic grey-box model can be  
607 used in this context to compare probabilistic simulations of the system to the observed flows.  
608 In addition, the results obtained for the EAM catchment in Section 3.1 make it clear that we

609 need to ensure that the rain events used for calibration of the models comprise the whole  
610 range of observed behaviour in a catchment.

611

612 The validity of the analysis of runoff forecast skill presented here is limited in that we did not  
613 have access to actual radar rainfall forecasts. Runoff forecast uncertainties therefore solely  
614 arise from deficiencies in model structure and errors in rainfall and flow observations. We  
615 refer to Thorndahl et al. (2013a) who were able to demonstrate runoff forecast skill using  
616 radar rainfall forecasts in a catchment with a total area of 80 ha up to forecast horizons of  
617 60 min, and Löwe et al. (2014b) who demonstrate good runoff forecast skill up to a horizon of  
618 100 min using rain gauge input in larger catchments with total areas of 1300 and 3000 ha. The  
619 size of the catchments considered in this work ranges in between these references.

620

621 Finally, comparing the score values obtained using the *NSE* and *PI* indices (Figure 4), the  
622 latter tends to be more critical and indicates a lower forecast skill of the rainfall runoff model.  
623 This is not the case where flow observations are noisy to a degree that cannot be removed by  
624 smoothing, for example during event 5 in the COL catchment or event 8 in the STB  
625 catchment (Figure 4).

626

#### 627 **4.2 STOCHASTIC MODEL STRUCTURE AND PARAMETER ESTIMATION**

628 Forecast uncertainties are generally underestimated (Figure 6). This tendency has been noted  
629 previously when estimating model parameters using the *CRPS* (Löwe et al., 2014a) and  
630 motivates the investigation of different score functions for parameter estimation.

631

632 We can, however, also identify structural deficiencies in the model. We here consider a model  
633 structure where forecast uncertainty increase linearly with the predicted model states

634 (equation (1)). This implies that a strong under- or overestimation of flows by the model will  
635 also lead to a false quantification of forecast uncertainty, which is particularly evident in the  
636 beginning of rain events (Figure 5b, c, d).

637

638 In addition, if relatively large errors occur in low flow situations, this will lead to large scaling  
639 parameters  $\sigma$  for the linearly state dependent uncertainty description. For high flows we then  
640 overestimate forecast uncertainties. This problem can be observed for the LER catchment in  
641 Figure 5d.

642

643 Consequently, future work must focus on the development of model structures that allow for a  
644 proper separation of the different uncertainties in dry and wet weather situations and that can  
645 describe forecast uncertainty, for example with a dependence on the rainfall input rather than  
646 the model states alone.

647

#### 648 **4.3 TIME VARIATION OF MODEL PARAMETERS**

649 We did not obtain improved runoff forecasts by making the area ( $A$ ) and time delay ( $k$ )  
650 parameters of our model time-variable by including these parameters in the state vector of the  
651 model. The corresponding variances were estimated as extremely small values and the  
652 parameters became time-invariant as a result.

653

654 The reason for this is that model errors that are observed on the considered time scale of one  
655 to few days are not related to actual variations in these parameters but caused by reasons such  
656 as errors in the radar rainfall measurements or problems in the model structure. These causes  
657 may vary in the course of an event.

658

659 An adjustment of our model parameters according to such errors corresponds to an overfitting  
660 of the model and will quite likely result in a reduced forecast quality over longer horizons.  
661 This is especially problematic for the  $A$  and  $k$  parameters because they are multiplicative in  
662 the model structure (1) and strongly affect the forecast results for all runoff processes. We can  
663 also see that not representing the whole range of catchment reactions in the calibration  
664 dataset, led to clearly reduced forecast skill in the EAM catchment.

665

666 The low forecast skill of the autocalibrated benchmark model (Figure 4) supports this  
667 argumentation. The selected variances in the prior parameter distributions allow for too strong  
668 variation of the parameters and thus too fast “forgetting” of previous observations. At the  
669 same time, allowing for less flexibility in the model parameters would reduce the model’s  
670 ability to adapt to new observations. This task is elegantly solved by the extended Kalman  
671 filter in the stochastic grey-box models.

672

673 We cannot exclude that it may be possible to track actual parameter variations over long time  
674 scales using the extended Kalman filter (or autocalibration approaches). Such variations may  
675 for example be caused by seasonally different reactions of the catchment. This, however,  
676 needs to be investigated using long, continuous time series of at least one year which were not  
677 available for this work.

678

679 For the time-variable dry weather flow parameter  $a_0$  we were able to obtain increased runoff  
680 forecast skill in the STB catchment (Figure 8). The parameter is additive in the model and  
681 variations hardly affect the runoff forecasts obtained during rain events. The time variable  
682 parameter acts similarly to a bias correction, affecting only the slow runoff processes.

683

684 **5 CONCLUSIONS AND OUTLOOK**

685 We have presented a framework for the generation of probabilistic on-line runoff forecasts in  
686 an urban context for forecast horizons of up to 120 minutes. We apply conceptual forecast  
687 models that were cast into a stochastic grey-box model layout. The forecast models were  
688 implemented to generate inflow forecasts for six dynamically controlled structures in the  
689 Lynetten catchment in Copenhagen, Denmark and the forecast performance was evaluated on  
690 a 120-minute forecast horizon. We draw the following conclusions and recommendations  
691 from this implementation:

692

- 693 1. We can generate runoff forecasts with high skill using the very simple model structure in  
694 combination with data assimilation and automated parameter estimation techniques. In  
695 practice, the forecast skill will be limited by the quality of rainfall forecasts, which were  
696 assumed to be known in our case. Future investigations will investigate the relation  
697 between catchment size, forecast horizon and quality of the rainfall forecast that is  
698 required to obtain an acceptable runoff forecast.
- 699 2. We can operate the runoff forecasting models without an explicit flow measurement in the  
700 inflow of the considered control points, relying solely on level measurements in the basins  
701 and outflow measurements from the basins. We do, however, need to ensure that the  
702 considered measurements capture all volume changes and outflows from the structure.
- 703 3. Noisy observations can be handled by the forecast models, while biased rainfall and flow  
704 measurements will corrupt the runoff forecasts and need to be identified in the pre-  
705 processing of the data. A correctly identified stochastic rainfall-runoff model can assist in  
706 the identification of corrupted data.

707 4. A time-variability of model parameters by including them in the extended Kalman filter  
708 needs to be handled with care. Modifications of area and time delay parameters of the  
709 model over short time periods are likely to induce over-fitting and thus reduced forecast  
710 skill. We were not able to obtain improved forecast skill by making these parameters time-  
711 variable in any of the six considered catchments.

712

713 It is possible that a time-variability of these parameters will lead to improved runoff  
714 forecasts if longer time scales (months) are considered for parameter variation. We were  
715 limited to event durations of a few days in this work.

716

717 Although we obtained improved runoff forecast skill in one of the catchments using a  
718 time-variable dry weather flow parameter, we suggest that improvements of the model  
719 structure are investigated before resolving to time-varying model parameters. The very  
720 simple model structure considered here leaves sufficient room to consider the spatial  
721 rainfall variability in an improved way, for example.

722 5. The operational deterministic forecast model layout used today applies a recalibration of  
723 model parameters every 2 minutes using a MAP approach. This layout was consistently  
724 outperformed by the stochastic grey-box models. Similarly to the previous points, we see  
725 the main reason for this result in too strong parameter variations in the auto-calibrated  
726 benchmark approach. Allowing only for smaller variations of the parameters in this  
727 approach, however, would comprise the model's ability to adapt to current flow  
728 observations. Adjusting the model states is a more robust method to update the model to  
729 current observations than adjusting the model parameters.

730 6. The uncertainty description in the stochastic grey-box approach currently builds on a  
731 linear dependence on the model states. This structure is problematic because it does not  
732 allow for a proper separation of the different uncertainties occurring in dry weather  
733 periods and during rain events and because wrong forecasts will also lead to false  
734 uncertainty estimates. Future work will focus on new model structures that describe  
735 forecast uncertainty depending on the rainfall input.

736 7. In a practical sense, the stochastic grey-box models make it easy to generate libraries of  
737 models of varying structural complexity. The user can automatically calibrate different  
738 model structures to observations from a catchment and select the required complexity.  
739 This process is simple and makes the method attractive, for example in comparison with  
740 the considered benchmark models where users need to understand the methodology and  
741 specify parameter distributions for the MAP estimation.

742 8. Finally, the companion paper by Vezzaro et al. (2014b) demonstrates that the  
743 consideration of runoff forecast uncertainties has significant effect on the efficiency of  
744 real-time control schemes. This work applies the stochastic grey-box models in an  
745 operational real-time control layout.

746

## 747 6 ACKNOWLEDGEMENTS

748

749 This research is financially supported by the Danish Council for Strategic Research,  
750 programme of the Commission on Sustainable Energy and Environment through the Storm-  
751 and Wastewater Informatics (SWI) project. The catchment and flow data were kindly  
752 provided by Copenhagen Utility Company (HOFOR). We thank the Danish Weather Service  
753 (DMI) for providing data from the C-Band radar at Stevns.

754

755 **7 REFERENCES**

756

757 Achleitner, S., Fach, S., Einfalt, T., Rauch, W., 2009. Nowcasting of rainfall and of combined  
758 sewage flow in urban drainage systems. Water Sci Technol 59, 1145-1151.

759

760 Arnbjerg-Nielsen, K., Madsen, H., Mikkelsen, P.S., 2006. Regional variation af ekstremregn i  
761 Danmark - ny bearbejdning (1979-2005). IDA Spildevandskomiteen Skrift 28, Copenhagen,  
762 Denmark.

763

764 Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J.,  
765 Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B.,  
766 Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., 2013. Characterising performance of  
767 environmental models. Environ. Model. Softw. 40, 1–20, doi 10.1016/j.envsoft.2012.09.011.

768

769 Borup, M., Grum, M., Mikkelsen, P.S., 2013. Comparing the impact of time displaced and  
770 biased precipitation estimates for online updated urban runoff models. Water Sci Technol 68,  
771 109-116.

772

773 Breinholt, A., Thordarson, F.O., Møller, J.K., Grum, M., Mikkelsen, P.S., Madsen, H., 2011.  
774 Grey-box modelling of flow in sewer systems with state-dependent diffusion. Environmetrics  
775 22, 946-961.

776

777 Brown, R.G., Meyer, R.F., 1961. The Fundamental Theorem of Exponential Smoothing. Oper  
778 Res 9, 673-685.

779

780 Bruti-Liberati, N., Platen, E., 2008. Strong predictor-corrector Euler methods for stochastic  
781 differential equations. *Stochastics Dyn.* 8, 561–581, doi 10.1142/S0219493708002457.

782

783 Campisano, A., Cabot Ple, J., Muschalla, D., Pleau, M., Vanrolleghem, P.A., 2013. Potential  
784 and limitations of modern equipment for real time control of urban wastewater systems.  
785 *Urban Water J.* 10, 300–311, doi 10.1080/1573062X.2013.763996.

786

787 Carstensen, J., Nielsen, M.K., Strandbæk, H., 1998. Prediction of hydraulic load for urban  
788 storm control of a municipal WWT plant. *Water Sci Technol* 37, 363-370.

789

790 Cassar, A., Verworn, H.R., 1999. Modifications of rainfall runoff and decision finding models  
791 for on-line simulation in real time control. *Water Sci Technol* 39, 201-207.

792

793 Da Ros, D., Borga, M., 1997. Adaptive use of a conceptual model for real time flood  
794 forecasting. *Nordic Hydrol* 28, 169-188.

795

796 Gentle, J.E., 2009. Computational Statistics, Springer, New York.

797

798 Gneiting, T., Raftery, A.E., Westveld III, A.H., Goldman, T., 2005. Calibrated probabilistic  
799 forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon*  
800 *Weather Rev* 133, 1098-1118.

801

802 Gneiting, T., 2007. Strictly proper scoring rules, prediction, and estimation. *J Am Stat Assoc*  
803 102, 359-378.

804

805 Harremoës, P., Madsen, H., 1999. Fiction and reality in the modelling world–Balance  
806 between simplicity and complexity, calibration and identifiability, verification and  
807 falsification. Water Sci Technol 39, 1-8.

808

809 Hemri, S., Fundel, F., Zappa, M., 2013. Simultaneous calibration of ensemble river flow  
810 predictions over an entire range of lead times. Water Resour. Res. 49, 6744–6755, doi  
811 10.1002/wrcr.20542.

812

813 Higham, D.J., Mao, X., Stuart, A.M., 2002. Strong convergence of Euler-type methods for  
814 nonlinear stochastic differential equations. SIAM J. Numer. Anal. 40, 1041–1063.

815

816 Iacus, S.M., 2008. Simulation and inference for stochastic differential equations: with R  
817 examples. Springer, Milan, Italy.

818

819 Jazwinski, A., 1970. Stochastic Processes and Filtering Theory. Academic press, New York,  
820 United States.

821

822 Jin, X., Xu, C.Y., Zhang, Q., Singh, V.P., 2010. Parameter and modeling uncertainty  
823 simulated by GLUE and a formal Bayesian method for a conceptual hydrological model. J  
824 Hydrol 383, 147-155.

825

826 Juhl, R., Kristensen, N.R., Bacher, P., Møller, J.K., Madsen, H., 2013. CTSM-R User Guide.  
827 Technical University of Denmark Lyngby, Denmark, [www.ctsm.info](http://www.ctsm.info).

828

- 829 Kloeden, P.E., Platen, E., 1999. Numerical Solution of Stochastic Differential Equations. third  
830 ed., Springer, Berlin-Heidelberg-New York.
- 831
- 832 Kristensen, N.R., Madsen, H., Jørgensen, S.B., 2004. Parameter estimation in stochastic grey-  
833 box models. *Automatica* 40, 225–237, doi 10.1016/j.automatica.2003.10.001.
- 834
- 835 Krzysztofowicz, R., 2001. The case for probabilistic forecasting in hydrology. *J Hydrol* 249,  
836 2-9.
- 837
- 838 Liguori, S., Rico-Ramirez, M.A., Schellart, A.N.A., Saul, A.J., 2012. Using probabilistic  
839 radar rainfall nowcasts and NWP forecasts for flow prediction in urban catchments. *Atmos*  
840 *Res* 103, 80-95.
- 841
- 842 Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I—A  
843 discussion of principles. *J Hydrol* 10, 282-290.
- 844
- 845 Löwe, R., Mikkelsen, P.S., Madsen, H., 2014a. Stochastic rainfall-runoff forecasting:  
846 parameter estimation, multi-step prediction, and evaluation of overflow risk. *Stoch. Environ.*  
847 *Res. Risk Assess.* 28, 505-516, doi 10.1007/s00477-013-0768-0.
- 848
- 849 Löwe, R., Thorndahl, S., Mikkelsen, P.S., Rasmussen, M.R., Madsen, H., 2014b. Probabilistic  
850 online runoff forecasting for urban catchments using inputs from rain gauges as well as  
851 statically and dynamically adjusted weather radar. *J Hydrol*, submitted.
- 852

- 853 Lund, N.S. V, Pedersen, J.W., Borup, M., Mikkelsen, P.S., Grum, M., 2014. Auto-Calibration  
854 For Updating Of Linear Reservoir Models Used In Flow Forecasting Of Urban Runoff.  
855 Hydrol. Process., in preparation.
- 856
- 857 Murphy, A.H., Winkler, R.L., 1977. Reliability of Subjective Probability Forecasts of  
858 Precipitation and Temperature. J Roy Stat Soc C-App 26, 41-47.
- 859
- 860 Ocampo-Martínez, C.A., Puig, V., 2010. Piece-wise linear functions-based model predictive  
861 control of large-scale sewage systems. IET Control Theory Appl 4, 1581-1593.
- 862
- 863 Øksendal, B., 1998. Stochastic differential equations : An introduction with applications, 6th  
864 ed., Springer, Berlin-Heidelberg-New York..
- 865
- 866 Pabst,M., Alex,J., Beier,M., Niclas,C., Ogurek,M., Peikert,D., Schütze,M., 2011. ADESBA -  
867 A new general global control system applied to the Hildesheim sewage system, 12th  
868 International Conference on Urban Drainage, 11-16 September, Porto Alegre, Brazil.
- 869
- 870 Pleau, M., Colas, H., Lavallée, P., Pelletier, G., Bonin, R., 2005. Global optimal real-time  
871 control of the Quebec urban drainage system. Environ Modell Softw 20, 401-413.
- 872
- 873 Puig, V., Cembrano, G.G., Romera, J., Quevedo, J., Aznar, B., Ramón, G., Cabot, J., 2009.  
874 Predictive optimal control of sewer networks using CORAL tool: application to Riera Blanca  
875 catchment in Barcelona. Water Sci Technol 60, 869-878.
- 876

- 877 Rajaram, H., Georgakakos, K.P., 1989. Recursive parameter estimation of hydrologic models.  
878 Water Resour Res 25, 281-294.
- 879
- 880 Reichert, P., Schuwirth, N., 2012. Linking statistical bias description to multiobjective model  
881 calibration. Water Resour Res 48, W09543.
- 882
- 883 Renard, B., Kavetski, D., Kuczera, G., Thyre, M., Franks, S., 2010. Understanding predictive  
884 uncertainty in hydrologic modeling: The challenge of identifying input and structural errors.  
885 Water Resour. Res. 46, W05521, doi 10.1029/2009WR008328.
- 886
- 887 Seggelke, K., Löwe, R., Beenken, T., Fuchs, L., 2013. Implementation of an integrated real-  
888 time control system of sewer system and waste water treatment plant in the city of  
889 Wilhelmshaven. Urban Water J. 10, 330–341, doi 10.1080/1573062X.2013.820331.
- 890
- 891 Schütze, M., Butler, D., Beck, M.B., 2002. Modelling, Simulation and Control of Urban  
892 Wastewater Systems. Springer Verlag.
- 893
- 894 Schütze, M., Campisano, A., Colas, H., Schilling, W., Vanrolleghem, P.A., 2004. Real time  
895 control of urban wastewater systems—where do we stand today? J Hydrol 299, 335-348.
- 896
- 897 Smith, M., Koren, V., Zhang, Z., Zhang, Y., Reed, S., Cui, Z., Moreda, F., Cosgrove, B. a.,  
898 Mizukami, N., Anderson, E., 2012. Results of the DMIP 2 Oklahoma experiments. J. Hydrol.  
899 418-419, 17–48, doi 10.1016/j.jhydrol.2011.08.056.
- 900

- 901 Thorndahl, S., Poulsen, T.S., Bøvith, T., Borup, M., Ahm, M., Nielsen, J.E., Grum, M.,  
902 Rasmussen, M.R., Gill, R., Mikkelsen, P.S., 2013a. Comparison of short term rainfall  
903 forecasts for model based flow prediction in urban drainage systems. Water Sci Technol 68,  
904 472-478.
- 905
- 906 Thorndahl, S., Rasmussen, M.R., 2013b. Short-term forecasting of urban storm water runoff  
907 in real-time using extrapolated radar rainfall data. J Hydroinf, In Press, doi  
908 10.2166/hydro.2013.161.
- 909
- 910 Todini, E., 2006. Rainfall-Runoff Models for Real-Time Forecasting, Encyclopedia of  
911 Hydrological Sciences, John Wiley & Sons, New York, United States.
- 912
- 913 Tolson, B. A., Shoemaker, C. A., 2007. Dynamically dimensioned search algorithm for  
914 computationally efficient watershed model calibration, Water Resour Res 43, W01413.
- 915
- 916 Vezzaro, L., Grum, M., 2014a. A generalized Dynamic Overflow Risk Assessment (DORA)  
917 for urban drainage Real Time Control. J. Hydrol., submitted.
- 918
- 919 Vezzaro, L., Löwe, R., Grum, M., Madsen, H., Mikkelsen, P.S., 2014b. Investigating the use  
920 of probabilistic forecasts for RTC of urban drainage systems – full scale testing in the city of  
921 Copenhagen, Denmark. Environ. Model. Softw., submitted.
- 922
- 923 Wilks, D.S., 2011. Statistical Methods in the Atmospheric Sciences, 2nd ed., Academic press,  
924 Oxford, United Kingdom.



PAPER F

# Investigating the use of probabilistic forecasts for RTC of urban drainage systems – Full scale testing in the city of Copenhagen, Denmark

---

**Authors:**

Luca Vezzaro, Roland Löwe, Henrik Madsen, Morten Grum, Peter Steen Mikkelsen

**in preparation for:**

*Environmental Modelling and Software*, 2014.



# Investigating the use of probabilistic forecasts for RTC of urban drainage systems – Full scale testing in the city of Copenhagen, Denmark

Luca Vezzaro<sup>1,2</sup>, Roland Löwe<sup>3</sup>, Peter Steen Mikkelsen<sup>1</sup>, Morten Grum<sup>2</sup>,  
Henrik Madsen<sup>3</sup>

## Abstract

Global Real Time Control (RTC) is increasingly becoming an efficient tool to improve the performance of urban drainage systems. This article assesses the potential of integrating the most recent developments in (i) radar-based runoff nowcasting, (ii) stochastic rainfall-runoff modelling, and (iii) uncertainty-based control schemes for RTC of storm- and wastewater systems. These tools are applied in the Lynetten urban catchment (76 km<sup>2</sup>, located in Copenhagen, Denmark) to reduce the volume of Combined Sewer Overflows (CSO) discharged into the natural waters. Stochastic runoff forecasts are generated for each subcatchment based on radar measurements, providing a 2-hour estimation of the expected runoff volumes. These so-called grey-box models also provide a dynamic estimation of the prediction uncertainty. The forecasted runoff volumes, along with their level of uncertainty, are then used by the global RTC scheme to reduce the CSO risk in the catchment. This study investigated the advantages of using stochastic runoff predictions against deterministic forecasts (currently applied in the Lynetten catchment) by simulating 13 rain events with a simplified hydrological model of the catchment, which was previously calibrated against a detailed hydraulic model. The comparison focused on the CSO reduction obtained by the control scheme, expressed both in terms of absolute and yearly CSO volumes. Based on the simulation results, RTC employing stochastic runoff forecasts achieved a 70% reduction in yearly CSO volume compared to the current local RTC, while the RTC employing deterministic forecasts showed worse performance

---

<sup>1</sup>Department of Environmental Engineering, Technical University of Denmark, Miljøvej, B115, DK-2800 Kgs. Lyngby, Denmark

<sup>2</sup>Krüger AS, Gladsaxevej 363, DK-2860 Søborg, Denmark

<sup>3</sup>Department of Applied Mathematics and Computer Science, Technical University of Denmark, Matematiktorvet, B322, DK-2800 Kgs. Lyngby, Denmark

(CSO reduction around 30%). This study thus demonstrates how the integration of stochastic runoff forecasts into global RTC strategies can contribute to reduce the impact of urban drainage system on the natural waters.

**Title: Investigating the use of probabilistic forecasts for RTC of urban drainage systems - full scale testing in the city of Copenhagen, Denmark**

**Authors:** Luca Vezzaro<sup>1,2</sup>, Roland Löwe<sup>3</sup>, Henrik Madsen<sup>3</sup>, Morten Grum<sup>2</sup>, Peter Steen Mikkelsen<sup>1</sup>

<sup>1</sup>Technical University of Denmark, Department of Environmental Engineering (DTU Environment), Miljøevej, Building 113, 2800 Kgs. Lyngby, Denmark

<sup>2</sup>Kruger A/S, Veolia Water Solutions&Technology, Gladsaxevej 363, 2860 Gladsaxe, Denmark

<sup>3</sup>Technical University of Denmark, Department of Applied Mathematics and Computer Science (DTU Compute), Richard Petersens Plads, Building 305, 2800 Kgs. Lyngby, Denmark

**Corresponding author:** Luca Vezzaro, e-mail: [luve@env.dtu.dk](mailto:luve@env.dtu.dk), tel. (+45) 45251579

## **ABSTRACT**

Global Real Time Control (RTC) is increasingly becoming an efficient tool to improve the performance of urban drainage systems. This article assesses the potential of integrating the most recent developments in (i) radar-based runoff nowcasting, (ii) stochastic rainfall-runoff modelling, and (iii) uncertainty-based control schemes for RTC of storm- and wastewater systems. These tools are applied in the Lynetten urban catchment ( $76 \text{ km}^2$ , located in Copenhagen, Denmark) to reduce the volume of Combined Sewer Overflows (CSO) discharged into the natural waters. Stochastic runoff forecasts are generated for each subcatchment based on radar measurements, providing a 2-hour estimation of the expected runoff volumes. These so-called grey-box models also provide a dynamic estimation of the prediction uncertainty. The forecasted runoff volumes, along with their level of uncertainty, are then used by the global RTC scheme to reduce the CSO risk in the catchment. This study investigated the advantages of using stochastic runoff predictions against deterministic forecasts (currently applied in the Lynetten catchment) by simulating 13 rain events with a simplified hydrological model of the catchment, which was previously calibrated against a detailed hydraulic

model. The comparison focused on the CSO reduction obtained by the control scheme, expressed both in terms of absolute and yearly CSO volumes. Based on the simulation results, RTC employing stochastic runoff forecasts achieved a 70% reduction in yearly CSO volume compared to the current local RTC, while the RTC employing deterministic forecasts showed worse performance (CSO reduction around 30%). This study thus demonstrates how the integration of stochastic runoff forecasts into global RTC strategies can contribute to reduce the impact of urban drainage system on the natural waters.

## KEYWORDS

Global Real Time Control, runoff forecast, urban drainage system, uncertainty, greybox model

### 1. Introduction

Global real time control (RTC) of urban drainage networks is increasingly seen as an efficient tool to respond to the increasing demands that are defined for these systems (see the discussion on integrated systems in Rauch et al.,2005). Global RTC can help urban water managers (i) to fulfil stricter environmental regulations and to improve the environmental status of natural waters (see the examples in Langeveld et al.,2013; Vanrolleghem et al.,2005) in an economically feasible manner, (ii) to meet the increasing demands from population (in terms of both reduction of risk of pluvial flooding and access to natural water for recreational purpose), (iii) to reduce operational costs (e.g. by reducing energy consumption or by automatic flushing and aeration), and (iv) to cope with unknown future challenges for the system due to for example changes in precipitation patterns or in the catchment land usage. The benefits of RTC compared to other structural solutions have been discussed in Beenenken et al. (2013) and demonstrated by both simulation studies (e.g. Dirckx et al.,2011) and by several full scale applications (e.g. Fradet et al.,2011; Nielsen et al.,2010; Ocampo-Martinez and Puig,2010; Pleau et al.,2005; Seggelke et al.,2013).

With respect to the reduction of Combined Sewer Overflows (CSO), the major benefit of global RTC derives from the ability of better exploiting the available storage in the drainage network. Clearly, information on the future evolution of the system (i.e. the expected runoff expected in the short future)

can contribute to a better optimization of the system. Recent development in radar-based runoff nowcasting (Kramer et al.,2005; Thorndahl and Rasmussen,2013; Thorndahl et al.,2013), along with an increasing number of on-line measurement and the development of data assimilation techniques are paving the way towards more advanced model predictive control strategies.

Given the usually short reaction time of urban drainage networks (around or below hours), nowcasting of runoff volumes relies on simplified (and fast) conceptual models (based on, for example, time-area method, linear reservoir cascade or time series methods). Examples applying such models for real-time control can be found in Pleau et al. (2005) and Puig et al. (2009)

However, as discussed by Deletic et al. (2012) and Schilling and Fuchs (1986), several sources of uncertainty can affect the results of urban drainage models (including model structure uncertainty, parameter uncertainty, etc.). Input uncertainty is clearly among the major factor affecting the on-line performance of these models (Schilling and Fuchs (1986), Schilling (1991)). Radar-based rainfall nowcasts provide an updated input that can be used to estimate runoff volumes in the near future (up to 2 hours – see also Thorndahl and Rasmussen (2013) and Borup et al. (2013)), but, as also pointed out by Achleitner et al. (2009), their uncertainty can significantly affect the quality of the runoff forecasts (and thus impact the performance of the control strategy).

Given the uncertainty of runoff forecasts resulting from simplified model structures and, in particular, uncertain rainfall inputs, forecast models should apply statistical techniques to exploit information from on-line measurements about the current state of the system to update the model. In addition, we argue that a dynamic quantification of forecast uncertainty is required to allow for a valid decision making in real time control.

An on-line correction of the forecast models based on current measurements can for example take the form of automated recursive parameter estimation based on a moving data window (Lund et al., submitted), time varying model parameters through Kalman filtering (Pleau et al.,2005) or model state updating through Kalman filtering (Carstensen et al.,1996). These approaches provide improved runoff forecasts as compared to their non-updated counterparts, yet they do not provide a calibrated estimate of forecast uncertainty.

Techniques that are applied for quantifying uncertainties in (urban) hydrology comprise for example Bayesian techniques (Del Giudice et al.,2013; Sun and Bertrand-Krajewski,2013) or GLUE (Thorndahl et al.,2008). Both approaches require Monte Carlo simulations to obtain an estimate of forecast uncertainty. Although steps towards making such techniques more applicable in an on-line context are being investigated (Moradkhani et al.,2012; Vrugt et al.,2005) they are still to computationally demanding to be applied in an on-line setup.

An alternative approaches for forecasting runoffs that combines the above requirements of exploiting information from on-line measurements and providing an estimate of forecast uncertainty is based on stochastic grey-box models. Here uncertainty is explicitly addressed by including a stochastic term in the model structure (Breinholt et al.,2011). Greybox models provide a dynamic estimation of the model output uncertainty, information that can directly be used by the control strategy. Further, through an extended Kalman filter setup, the models can adapt to new runoff observations in an online setting. This state updating usually ensures that forecasts are generated from a correct starting point but also provides flexibility for handling erroneous or missing observations. By using a lumped model structure, which is computationally fast and thus suitable for RTC applications, it is possible to control the drainage system based on the dynamically estimated level of uncertainty.

Various control strategies which explicitly account for uncertainty in the model predictions have recently been presented in literature. For example, Raso et al. (2012), described a stochastic optimization approach based on multiple model simulations, while an RTC approach using ensemble rainfall forecasts is presented in Raso et al. (in press). Vezzaro and Grum (2012) introduced the Dynamic Overflow Risk Assessement (DORA) approach. DORA utilizes the estimated uncertainty in the runoff prediction to calculate and minimize the overflow risk across the entire drainage network.

This article aims at demonstrating the potential for using stochastic runoff forecasts within global RTC schemes of urban drainage networks. The study focused on the Lynetten catchment (located in Copenhagen, Denmark), which is controlled by the DORA global strategy. Currently, short-term runoff forecasts are generated by using an auto-updated deterministic rainfall-runoff model. The comparison was performed by running a simplified hydrological model of the Lynetten catchment

with five different scenarios (baseline, deterministic/stochastic, with/without uncertainty in forecasted runoff volumes). These scenarios allowed identifying the effect of using deterministic or stochastic runoff forecast in combination with DORA, as well as showing the importance of including uncertainty within global control schemes of urban drainage systems.

## 2. Material and methods

### 2.1. Catchment description

The Lynetten catchment covers the central area of Copenhagen (Denmark) and it has an area of approximately 76 km<sup>2</sup>. The catchment discharges to the Lynetten wastewater treatment plant (WWTP), which applies the Aerated Tank Settling approach (Nielsen et al., 2000; 1996) to increase of the plant capacity during rain events. By integrating the WWTP and catchment control schemes into the same infrastructure (in the specific case, a STAR® control tool - Thomsen and Ônnerth (2009))) it is thus possible to increase the WWTP capacity based on the forecasted inflow to the plant, and to optimize the upstream storage capacity by including the actual capacity of the biological treatment unit at the WWTP into DORA (see section 2.2). A detailed description of the study area can be found in Breinholt and Sharma (2010).

The integrated control of the Lynetten catchment considers eight overflow structures (connected to basins and pumping stations – see Figure 1), discharging to recipients with different sensitivity to CSO. The total storage capacity is about 151,400 m<sup>3</sup>. Rainfall falling on the catchment is quantified by radar measurements, which are time-dynamically adjusted to data provided by the rain gauges belonging to the network of the Danish Water Pollution Committee, operated by the Danish Meteorological Institute (Jørgensen et al.(1998)). Further details on the rainfall forecasting methods can be found in Thorndahl et al. (2013) and Thorndahl and Rasmussen (2013). While rain gauge measurements have been available for several decades, radar data of reasonable quality have become available only in the period 2012-2013.

Measurements available at the control points are commonly outflows from basins and water levels, i.e. there are no direct measurements of inflows to the basins and/or of overflows. CSO volumes are

commonly estimated by the water utility by using a detailed hydrodynamic model (implemented in the MIKE URBAN software, DHI (2008)), which utilizes radar measurements as input.

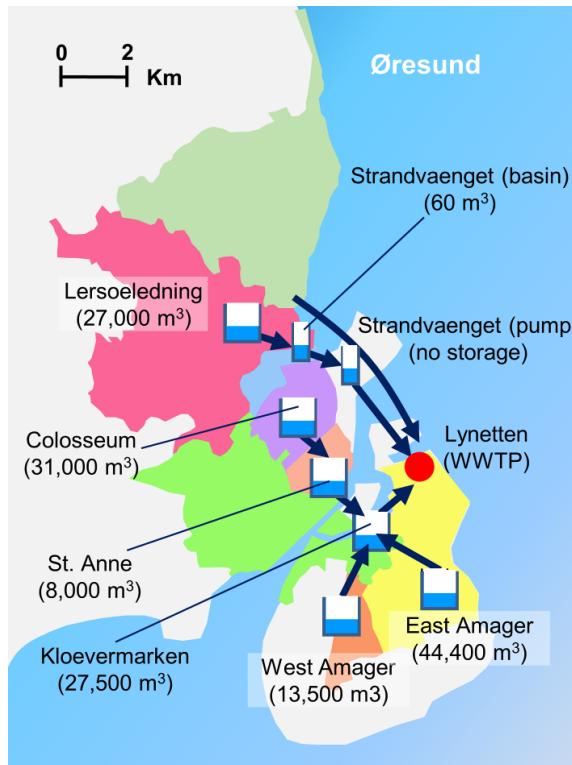


Figure 1. Scheme of the Lynetten catchment. Storage volumes and (in brackets) are listed for each controlled node of the network.

Table 1. Main characteristics of the points controlled by DORA.

Subcatchment	Impermeable area [ha]	Storage available for RTC [m <sup>3</sup> ]	Max outflow [m <sup>3</sup> /s]	CSO cost [€m <sup>3</sup> ]
Lersoeledning (LER)	733	27,000	1.1	25
Strandvaenget – Basin (STB)	92	60	3.9	25
Strandvaenget – Pump (STP)	-	0	2.4	5
Colosseum (COL)	211	30,914	0.9	5
St. Anne (STA)	77	7,987	1.3	5
East Amager (EAM)	228	44,425	2.1	25
West Amager (WAM)	97	13,460	1.0	5

Kloevermarken (KLO)	777	27,500	7.5	5
Lynetten WWTP (LYN)	564	76	5 (6.4 <sup>a</sup> )	1
Total	2,279	151,432		

<sup>a</sup> during wet weather operation of the WWTP

## 2.2. Global control strategy (DORA)

The Dynamic Overflow Risk Assessment (DORA) strategy (Vezzaro and Grum (2012)) aims at reducing the overflow risk across the entire catchment. The overflow risk for each controlled point is calculated by multiplying the CSO cost (which reflects the sensitivity of the receiving waters) by the forecasted probability of overflow volume (obtained by subtracting the available storage volume from the forecasted distribution of runoff volume). The CSO cost is calculated by using a global function which considers:

- The current water storage in the system (provided by measurements);
- The expected runoff volumes (estimated by using radar-based nowcasts as input to conceptual rainfall-runoff models);
- The uncertainty of the runoff predictions (see Section 2.3 for further details);
- The sensitivity of the receiving water body (expressed through the CSO cost) which was assumed to be linearly proportional to the overflow volume and constant in time. More sensitive points (e.g. bathing areas) have higher CSO prices than less sensitive points (e.g. inlet to WWTP - see Table 1 for the list of used CSO prices). We refer to Vezzaro and Grum (2012) for details.

At each time step (in this study set to 2 minutes, i.e. each time a new set of measurements from the catchment is available in the STAR® database), DORA executes the following loop (Figure 2):

- Step 1: The available storage volume for each basin is calculated by using on-line measurements.
- Step 2: Runoff forecasts (and the associated uncertainty) are used to calculate the probability of overflow for each basin.
- Step 3: a genetic algorithm is used to identify the optimal set of flows between all the basins in

the catchment which minimized the total CSO cost. When CSO risk is low (e.g. after the end of a rain event, with no new rainfall within the forecast horizon), DORA identifies the flows that ensure a faster emptying of the system.

- Step 4: optimal setpoints for each basin outflow are sent to the actuators in the system.

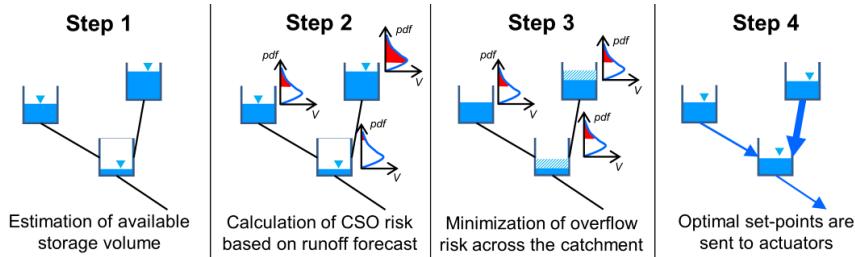


Figure 2. Schematic representation of basic steps in DORA (Probability of overflow is shown in red).

### 2.3. Runoff forecast models

The estimation of the expected runoff at each control point (step 2 in DORA) can be performed by using simple rainfall-runoff models, which are computationally fast and that can easily assimilate information on the current state of the system. Two different approaches are available in the Lynetten catchment: autocalibrated deterministic models (currently used) and stochastic models (presented in Löwe et al., submitted). In the operational setting, a three-tank in series model is in both cases used to generate runoff predictions based on radar-based rainfall nowcasts (Thorndahl and Rasmussen, 2013) which are provided every 10 minutes as input. A detailed description of the two methods is given in the following sections.

#### 2.3.1. Deterministic model with recursive parameter estimation

The runoff for each of the eight subcatchments (Table 1) is currently estimated by using a rainfall-runoff model which employs three linear reservoirs. The three parameters of the model (fraction of reduced area, linear reservoir constant, and average dry weather flow) are automatically estimated every 2 minutes by using a Maximum A Posteriori approach in combination with flow measurements from the catchment (Lund et al., submitted). The models are implemented in Wateraspects<sup>TM</sup> (Grum et

al.,2004) and hydrographs for the following 2 hours are forecasted. From the numerical integration of these hydrographs, DORA estimates the expected runoff volume at each controlled point.

The uncertainty of the estimated runoff volume is assumed to follow a gamma distribution:

$$p(V_{F,i}) = V_{F,i}^{k-1} \cdot \frac{e^{-\frac{V_{F,i}}{\theta}}}{\Gamma(k) \cdot \theta^k} \quad (1)$$

where  $\Theta$  and  $k$  are the two parameters of the gamma distribution. These parameters can be determined by using a maximum likelihood principle, such that the forecasted runoff for the  $i$ -th control point  $V_{F,i}$  [ $\text{m}^3$ ] corresponds to the mean of the distribution (Vezzaro and Grum, 2012; Löwe et al., submitted). Based on these assumptions, it is possible to analytically estimate the overflow risk for each basin.

### 2.3.2. Stochastic model

Similar to the deterministic case, we apply conceptual hydrological rainfall-runoff models for generating stochastic runoff predictions with radar nowcasts as input. Breinholt et al. (2011) give a detailed description of the model structure. The open-source software CTSM (Kristensen and Madsen,2003) is applied for the modelling process and we obtain a state space model layout with the following system equations:

$$d \begin{bmatrix} S_1 \\ S_2 \\ S_3 \end{bmatrix} = \begin{bmatrix} A \cdot P_t + a_0 - \frac{1}{K} S_1 \\ \frac{1}{K} S_1 - \frac{1}{K} S_2 \\ \frac{1}{K} S_2 - \frac{1}{K} S_3 \end{bmatrix} dt + \begin{bmatrix} \sigma_1 \cdot S_1 \\ \sigma_2 \cdot S_2 \\ \sigma_3 \cdot S_3 \end{bmatrix} d\omega_t \quad (2)$$

In eq. 2 ( $S_1$ ,  $S_2$ ,  $S_3$ ) correspond to the reservoir states in the lumped model, ( $A$ ) to the effective area, ( $P_t$ ) to the rainfall input, ( $a_0$ ) to the mean dry weather flow and ( $K$ ) to the time lag constant. In addition to the physical model part, the system equations also include a stochastic term consisting of a random process ( $d\omega_t$ ) with state dependent variance ( $\sigma \cdot S$ ). This term is used to model uncertainties resulting from uncertain (rainfall) inputs and an incomplete description of reality by the model.

The so-called observation equation relates flow predictions from the model to basin inflow observations ( $Q_i$ ) at time step ( $i$ ). We describe variations in dry weather flow by a harmonic function ( $D_i$ ) and consider the flow observations subject to a random normal error ( $e_i$ ).

$$Q_i = \frac{1}{K} S_{3,i} + D_i + e_i \quad (3)$$

Depending on the variance of states and observations, this model layout allows for an adjustment of the states at every time step to match the observations. This state updating is performed through an extended Kalman filtering routine. It is also possible to include the model parameters ( $A$ ) and ( $K$ ) as additional states corresponding to e.g. purely random variables in the system equations 3. The parameters are then estimated as part of the state updating and the model can adapt to different behaviour of the system e.g. in summer and winter or for different rainfall characteristics. We intend to implement this kind of behaviour in the next version of forecasting models and hence do not perform a constant re-calibration as for the deterministic models.

Parameters for the stochastic flow forecasting models used in this work are derived by minimizing the difference between the distributions of multistep flow predictions and the empirical distribution of the corresponding flow observations using the continuous ranked probability score (CRPS) as criterion (Gneiting and Raftery, 2007; Löwe et al., 2014). Four rainfall events from all seasons in 2012 were used for estimation of the model parameters.

Probabilistic predictions of runoff volumes are obtained through an ensemble approach. 1000 possible realisations of the stochastic flow forecasts up to the maximal considered horizon of 60 time steps or 2 hours are generated from the grey-box models. For each realisation, the predicted runoff volume up to a given horizon is determined and subsequently a distribution of runoff volumes is derived from the different realisations. (Löwe et al., submitted)

## 2.4. Comparison setting

### 2.4.1. Catchment model

The effect of using a deterministic or a stochastic model to provide runoff forecast for global control of the Lynetten catchment was evaluated through simulations with a conceptual model of the Lynetten catchment. Following the procedure presented by Borsanyi et al. (2008), the conceptual model (coded in Wateraspects™) was calibrated against the detailed MIKE-URBAN model of the catchment.

The calibration was performed with respect to the cumulative runoff volume discharged from the

subcatchment as well as to the runoff flow. Model performances on runoff flows were evaluated by using the Nash-Sutcliffe Efficiency (NSE - Nash and Sutcliffe,1970) over three rain events (one for calibration and two for validation).

The runoff generation and transport processes in the Wateraspects™ software (Grum et al.,2004) are those typically described in similar conceptual hydrological models (such as CITYDRAIN® (Achleitner et al.,2007) or SIMBA (Erbe and Schuetze,2005)): a time-area method is used to generate runoff and a simple time delay is used for routing in pipes. Local controls existing in the catchment (e.g. pumping based on filling degree in basins) are implemented in the model, and they can be overridden by the DORA setpoints when global control strategies are simulated.

#### **2.4.2. Scenario comparison**

To illustrate the effect of different methods to estimate runoff volumes, five different scenarios were simulated:

- *Baseline scenario*: the Lynetten model is run without any global control (i.e. only local controls within the catchment were simulated).
- *Scenario A (deterministic)*: the deterministic method is used to estimate runoff forecasts as described in Section 2.3.1. Given the poor performance of the deterministic forecast model (Löwe et al., submitted), the estimation of the gamma distribution parameters was considered as unfeasible. Default parameters used in operational conditions of DORA were applied in this study, i.e. the simulated set-up of DORA corresponds to the one currently used in the Lynetten catchment;
- *Scenario B (deterministic without uncertainty)*: same as scenario A, but runoff forecasts are assumed to have a negligible uncertainty (i.e. a normal distribution with standard deviation of  $10^{-4}$  was used to calculate the CSO risk);
- *Scenario C (stochastic)*: the grey-box method is used to estimate runoff forecasts and their uncertainty bounds as described in Section 2.3.2;
- *Scenario D (stochastic without uncertainty)*: the median of the flows forecasted by the grey-box model is used to estimate runoff volumes. Similarly to Scenario B, this forecast is

assumed to have a negligible uncertainty (i.e. normally distributed with standard deviation  $10^{-4}$ ).

Control results in the five scenarios were evaluated by using the simulation set-up illustrated in Figure 3: i.e. inputs, control strategy and controlled systems were identical. Simulations are run by using a STAR® emulator, i.e. software settings and configurations are identical to those applied for on-line operations. Measured data were used to update the forecast models while the control decisions were evaluated in the conceptual WaterAspects model of the whole catchment.

Scenario B and D were included in the analysis to illustrate the benefit of including uncertainty in the control scheme. The comparison focused on (i) total number of CSO, (ii) overflow volume discharged during each event, (iii) total CSO volume and estimated yearly CSO discharge.

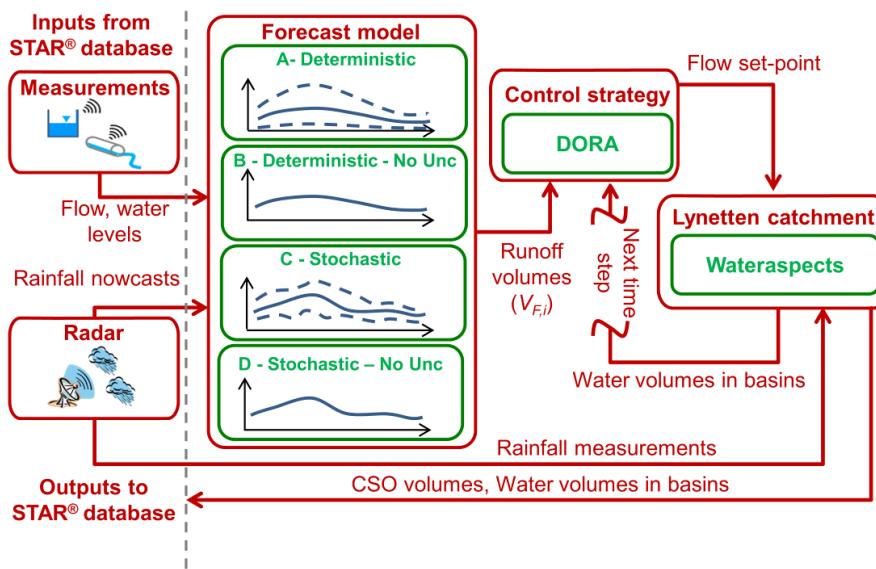


Figure 3. Scheme of information flow for each time step of simulation.

#### 2.4.3. Simulated events

Rain events were selected based on (i) the total rainfall height, to ensure CSO events in the system, (ii) the availability and quality of input radar data, and (iii) the availability of measurements from the catchment. Due to the bad quality of the measurements available for the St. Anne basin, this control

point was not considered in the control scheme (i.e. DORA used the setpoints from the local control point).

If two or more overflow events occurred within 8 hours (i.e. in the same order of magnitude of the emptying time of the system), these events were merged into a single events. Based on these criteria, a total of 13 events were identified in the period January 2012 – July 2013.

To evaluate the frequency of the simulated events (and thus their statistical representativity), historical rain data from the period January 2000-July 2013 (for a total of 921 events) were used to simulate overflow events in the simplified conceptual model of the Lynetten catchment and to extrapolate statistics on CSO events. A three parameter generalized Pareto distribution was fitted to the simulated total CSO volumes (see Figure 2).

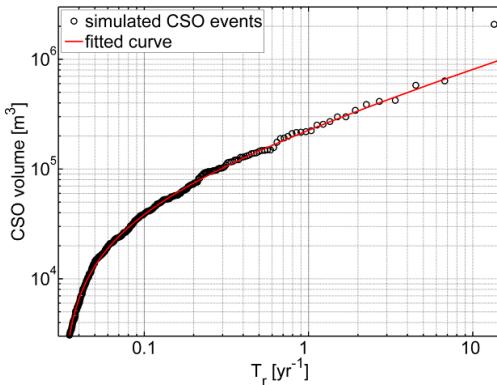


Figure 4. Simulated total CSO volume for the Lynetten catchment for the period January 2000- June 2013. The red line shows the fitted three-parameter Generalized Pareto distribution.

The return period of the events included in the scenario comparison was estimated by using the fitted distribution. An expected annual variation of CSO volume was then derived and used to compare the overflow volumes in scenarios A to D to the baseline scenario. A similar approach is commonly applied in flood risk management for calculating the Expected Annual Damage and comparing different solutions (see Zhou et al. (2012) for an example).

As a first step, we defined a *CSO density curve* for scenario X ( $f_{CSO,X}(t)$ ), which was obtained by

multiplying the CSO volume by the event frequency. The expected relative variation in yearly CSO volume for scenario  $X$  ( $\Delta CSO_X$ ) was then calculated as:

$$\Delta CSO_X = 1 - \frac{\int_0^{t=1\text{yr}} f_{CSO,X}(t) dt}{\int_0^{t=1\text{yr}} f_{CSO,baseline}(t) dt} \quad (4)$$

Where  $f_{CSO,baseline}(t)$  [ $\text{m}^3/\text{yr}$ ] is the CSO density curve for the baseline scenario. The relative variation in yearly CSO volume can be compared graphically for different scenarios as shown in Figure 5

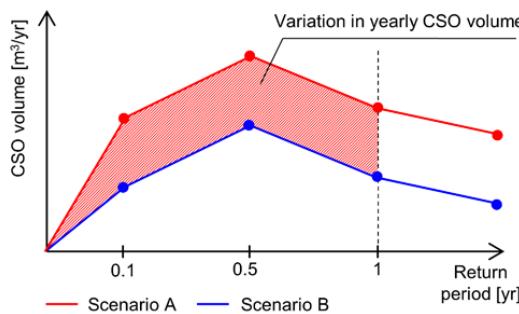


Figure 5. Schematic representation of use of CSO volume density curve (adapted from Zhou et al., 2012).

### 3. Results and discussion

#### 3.1. Performance of simplified hydrological model

Figure 6 shows the comparison between the simulated flow by the detailed hydraulic model (black line) and the simple hydrological model (red line). The figure shows how the time-area method managed to achieve similar results to the detailed hydraulic model with respect to fast runoff and peak flows.

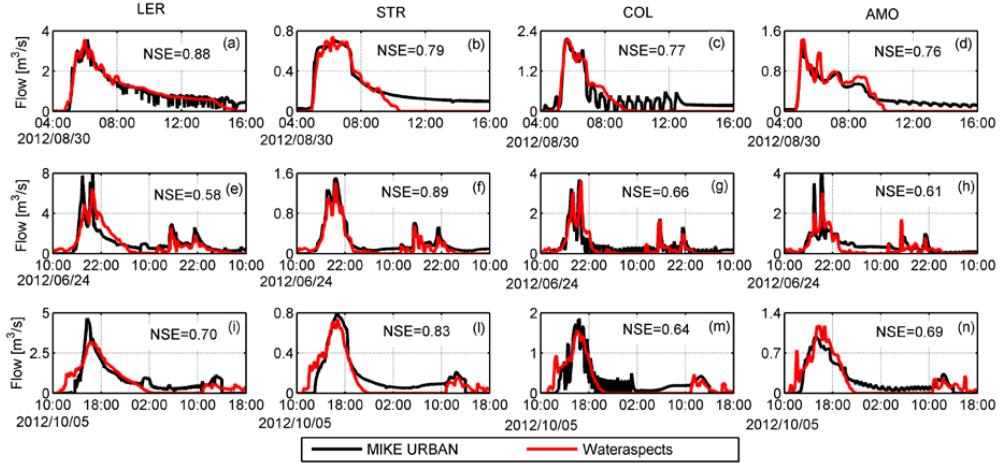


Figure 6. Examples of runoff generated by the fully hydrodynamic model MIKE URBAN (black line) and by the hydrological model WaterAspects (red line) for some sub-catchments (LER, STR, COL, AMO) for calibration (a-d) and validation (e-n) events.

The NSE for the calibration events ranged between 0.41 and 0.90, while the NSE ranged between 0.46 and 0.89 for the validation events (excluding the LYN catchment, which had negative NSE, and the STA catchment, where the ranged between 0.21 and 0.33). The performance for the validation event at the LYN subcatchment were extremely poor due to (i) instability in the MIKE-URBAN model (e.g. leading to negative flows) and (ii) the fact that the flow from that catchment is entirely driven by an upstream pumping station, whose behaviour is not possible to represent with a time-area method. A neglected pumping station also caused the low NSE values at STA; the average flow, however, was almost overlapping and the results were therefore assumed to be satisfactory. The simple hydrological model also encountered difficulties in fully simulating the slow runoff (as can be seen in some examples in Figure 6b,c,l,n). When looking at the water balance (Figure 7), the WaterAspects model tended to underestimate the runoff volume compared to the MIKE-URBAN model (8% less on average), mainly corresponding to the slow runoff.

Given that (i) CSO events are usually not affected by slow runoff, and (ii) WaterAspects succeeded in representing the initial phase of the hydrograph and the peak flows, the conceptual model was regarded as suitable to evaluate the effect of different control strategies of the catchment.

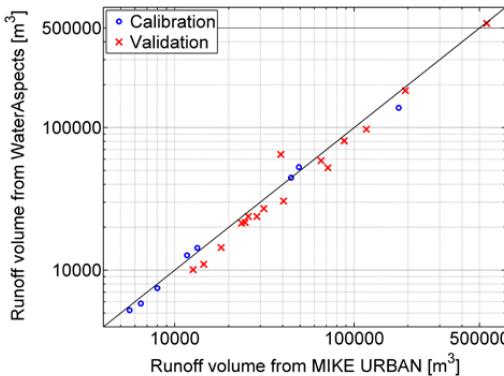


Figure 7. QQ plot of simulated runoff volumes from the eight subcatchments.

### 3.2. Baseline scenario

The estimated return period of the 13 compared events ranged between 0.03 and 1.3 years, i.e. the simulated events were frequent and with a limited magnitude of overflow volume discharged into the receiving waters. As a consequence, CSO events were simulate only at a limited number of discharge points (Figure 8), with the greatest number of overflow taking place at Lynetten and Lersøledning.

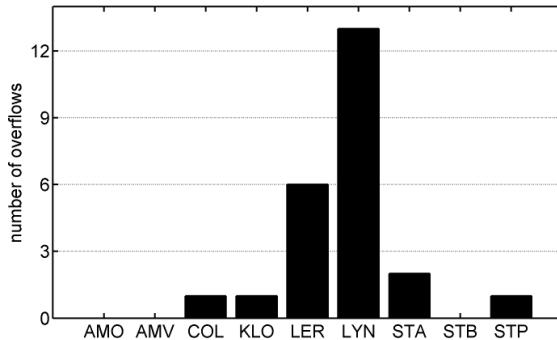


Figure 8. Number of simulated CSO events at the different CSO structures during the simulations of the baseline scenario.

### 3.3. Single events comparison

Figure 9 shows some examples of the runoff predictions, along with their uncertainty bounds, for selected events (events that were included in the calibration and validation of the stochastic models – see Löwe et al., submitted). When looking at the median value (point forecast), the deterministic

(scenario A) and the stochastic (scenario C) approaches produced similar forecasts. When the simulated flow considerably deviated from the measured flow (as for STR), both forecast approaches showed similar bias due to the updating of the models to current measurements.

The major difference between Scenario A and C can be seen when looking at the width of the predicted uncertainty bounds. The uncertainty bounds for Scenario A were generally wider than those estimated for Scenario C (as for the examples shown in Figure 9).

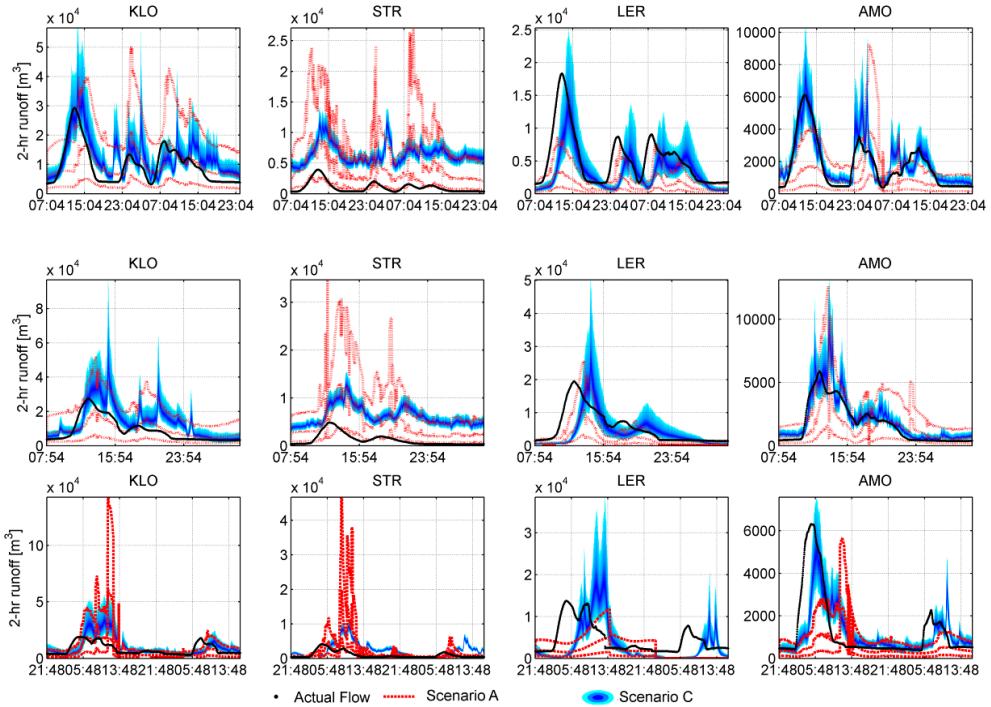


Figure 9. Simulated runoff forecasts for the deterministic model (A - dotted lines) and the stochastic approach (C - blue area) against the runoff simulated by the catchment model (WaterAspects) for selected subcatchments (Lersøledning, Strandvænget, Colosseum, Amager East - black dots). Top: event: start at 2012/01/21. Middle: event start at 2012-01-19. Bottom: Top: event. 2013-05-07.

The narrower bounds estimated in scenario C (stochastic forecast models) resulted in lower overflow risk compared to scenario A (deterministic forecast models). For example, uncertainty bounds at

Strandvænget (STB) were generally greater for the deterministic approach, resulting in a greater amount of water stored in the system by DORA in order to decrease the CSO risk at STR (Figure 10). This resulted in overflow events which were avoided when the forecast uncertainty was lower. When uncertainty in predictions is neglected (scenarios B and D), the forecasted overflow risk is close to zero unless basins are completely filled. This resembles the final phase of rain events, when the forecasted runoff volumes are low and DORA optimizes the emptying of the system. Therefore, the results obtained for Scenario B and D resembled those of a traditional RTC approach, where priority is given to emptying detention basins. This can be noticed in Figure 10 for the upstream basins (AMO, COL, AMV), which were emptied earlier compared to scenarios A and C. Similarly, the KLO basin, receiving more water from the upstream part of the system, tended to empty at a lower rate during scenarios B and D than during scenarios A and C.

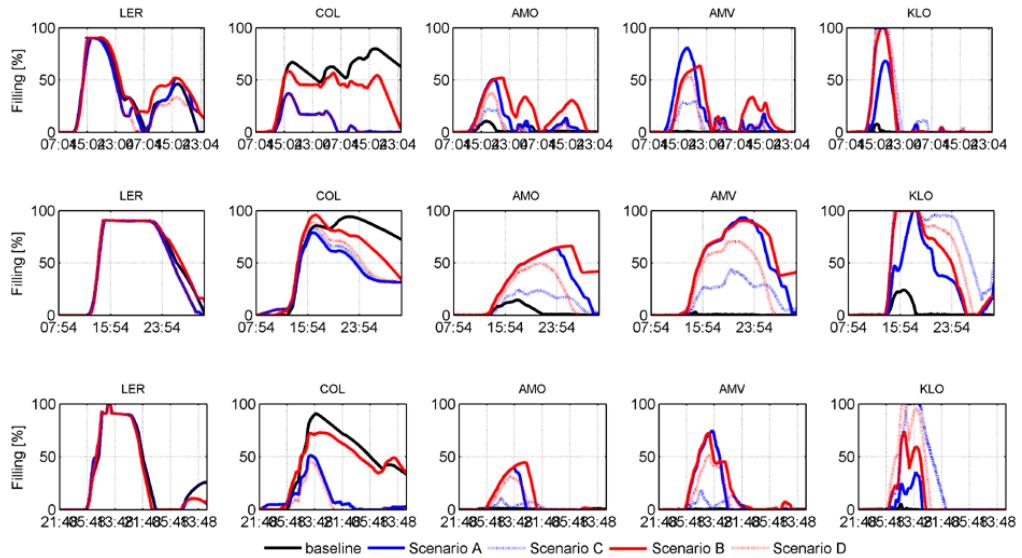


Figure 10. Comparison of simulated degree of filling in controlled detention basins for event starting at 2012/01/21 (top), 2012-01-19 (middle), and 2013-05-07 (bottom).

### 3.4. Scenario comparison

Figure 11 shows a comparison between the CSO volumes for the different controlled scenarios against the corresponding CSO volume for the baseline scenario. Using this graph, it is possible to evaluate the relative variation in the CSO volume for several rain events. The main effect of DORA can be seen at the WWTP inlet (Lynetten, i.e. the control point which is considered least sensitive in the control strategy) and at Lersøledning (LER) (a highly sensitive discharge point). A large number of CSO events was avoided at Lynetten, while discharge at Lersøledning was increased (the volume is more than doubled for scenarios A-B, and increased by 12% and 80% for scenarios C and D, respectively). A similar effect of DORA in the Lynetten catchment (i.e. decrease at LYN and increase at LER) was simulated by Vezzaro et al. (2013), who used “perfect forecasts” (i.e. the exact runoff flows were used to calculate overflow risks). DORA also generated new overflow events compared to the baseline scenario in other points of the system, mainly at Kløvermarken and Strandvænget (Figure 12). Although the creation of new overflow events through the catchment might appear as a negative result, this was a direct result of the integrated approach, as the total amount of CSO volume discharged by all the CSO structures was reduced.

Scenarios B and D, where uncertainty in runoff volume forecasts was neglected, both resulted in similar overflow volumes. As mentioned in the previous sections, the median point forecasts used in those scenarios were similar, thus resulting in similar optimal set-points and control performance. The importance of including forecast uncertainty in the control scheme is stressed by the greater CSO volume for scenario D than scenario C. One of the structures which greatly benefitted from the inclusion of uncertainty is Kløvermarken: as shown in the examples in Figure 10, scenarios B and D tended to store more water in this structure due to neglecting the CSO risk. Conversely, scenarios A and C stored a lower amount of water in order to minimize the CSO risk: this resulted in a lower CSO volume over all the events.

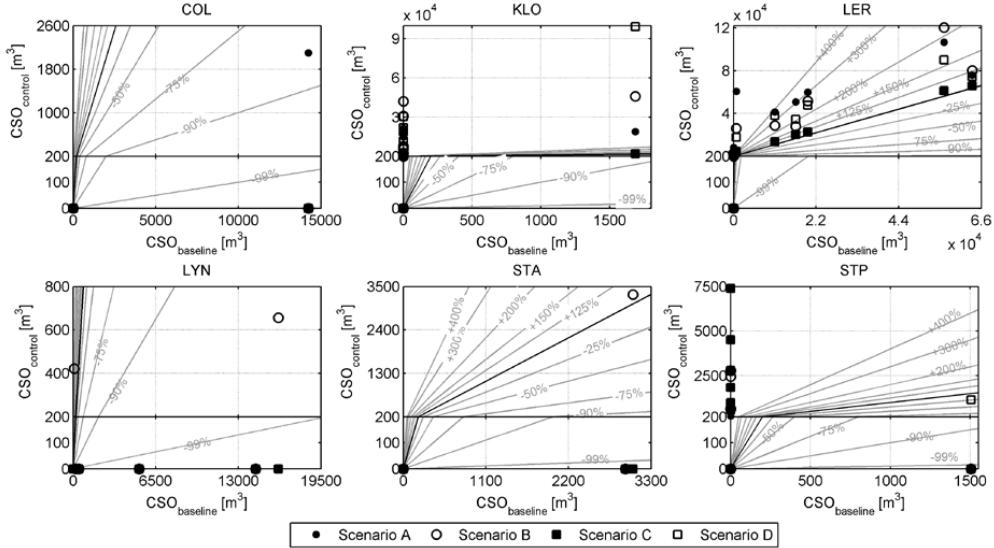


Figure 11. Comparison between CSO volumes with DORA (y-axis – not linear scale) against CSO volumes for local control (x-axis) for points where overflow occurred in baseline scenario. Values below the diagonal line corresponds to reduction in CSO volumes, while values above the diagonal corresponds to an increase in CSO compared to the baseline scenario.

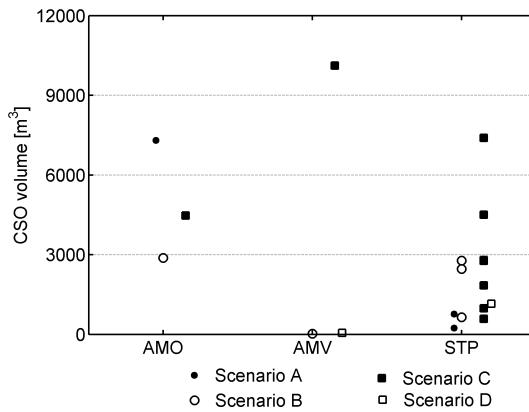


Figure 12. Simulated CSO volumes for points where no CSO occurred during baseline scenario.

When the results for the 13 events where deterministic forecasts were available are summed together (Figure 13), a clear pattern can be identified:

- As mentioned earlier, DORA significantly reduced the number of overflows (from 13 events

to below 5) at the downstream point of LYN (inlet to the WWTP) by storing a greater amount of water in the upstream parts of the system.

- The total CSO volume discharged by the Lynetten catchment is reduced for all the simulated scenarios. DORA in combination with the deterministic forecast models in scenario A obtained a reduction of 32.3% in the CSO volume. When stochastic forecast models were used in scenario C, the CSO reduction reached 58.7%.
- The CSO volume reduction obtained in scenario A (around 32%) was lower when compared to scenario C (which obtained a reduction of almost 70% in the total CSO volume when forecast uncertainty was taken into account– see Table 2). As discussed in Section 3.2, narrower uncertainty bounds allowed a better exploitation of the available storage, and thus resulted in better overall performance of the drainage network. It is thus important to provide a dynamic description of forecast uncertainty and use robust forecasting models.
- Scenarios B and D obtained CSO reductions which were about 20% lower than the corresponding scenarios with uncertainty (A and C). These results further stressed the paramount importance of including forecast uncertainty when information on the future evolution of rainfall is included in the control scheme.

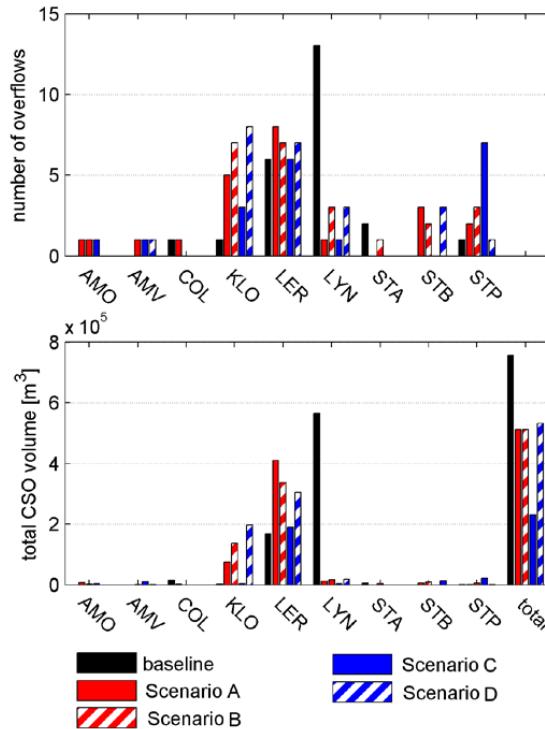


Figure 13. Top: Number of simulated overflow events for the different CSO structures. Bottom: total CSO volume for the different scenarios.

Table 2. Total CSO volumes estimated for the XX simulated events and percentage reduction, along with extrapolated yearly reduction (based on eq. 4).

Scenario	Total CSO volume [m³]	Variation from baseline [%]	
		Simulated events	Extrapolated yearly variation
Baseline	$7.56 \cdot 10^5$		0
Scenario A	$5.12 \cdot 10^5$	-32.3	-33.5
Scenario B	$5.11 \cdot 10^5$	-32.4	-42.6
Scenario C	$2.32 \cdot 10^5$	-69.4	-75.9
Scenario D	$5.33 \cdot 10^5$	-29.6	-41.8

When the performance of the different scenarios is analysed with respect to the magnitude of the simulated CSO events (expressed as return period - Figure 14), Scenario C showed more consistent

performance irrespective of the characteristics of the rain event: the blue solid line, in fact, is always below all the lines representing all the other scenarios. Conversely, scenario A showed quite significant variations in performance for different events, with some cases where the global control resulted in a greater CSO volume compared to the baseline scenario (i.e. where the CSO density curve for scenario A – red solid line - is above the curve for the baseline scenario). The similarities in the performance of scenarios B and D are shown also in Figure 14, as the two dotted lines are almost overlapping, suggesting that the similar CSO reductions were obtained for the simulated events.

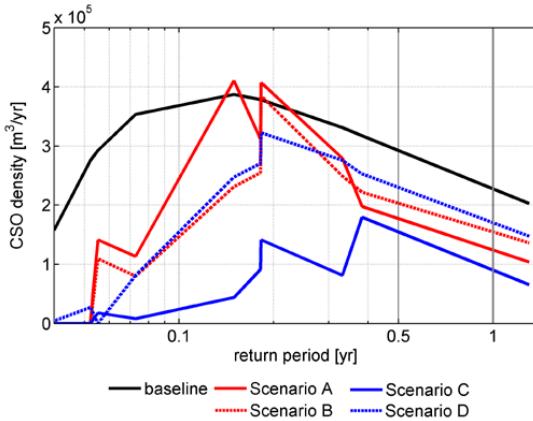


Figure 14. CSO density curve as function of return period. The curves for scenarios using stochastic forecasts (C-D) are generally below the baseline curve and the deterministic scenarios (A-B)

#### 4. Conclusions

The potential for the implementation of stochastic runoff forecasting models into a strategy for RTC of urban drainage systems has been assessed. Based on the simulation results, RTC based on the stochastic forecasts led to lower CSO volumes and to a better exploitation of the available storage capacity.

The analysis of single overflow events confirmed that the DORA control strategy was able to adapt the optimal flows according to the estimated uncertainty in the runoff volume forecasts: in fact, forecasts with similar median values but different level of uncertainty resulted in different set-points sent to the

actuators. The importance of including uncertainty of runoff forecasts clearly emerged from the analysis of both single events and event series. When overflow risk was overlooked, the DORA control strategy prioritized emptying of the system, resulting in lower storage in the systems and increased CSO volumes discharged into the environment.

The more realistic estimate of forecast uncertainty provided by the stochastic forecast models resulted in a better utilization of the available storage volume in the systems, and thus in a reduction of overflow volumes as compared to the scenario where deterministic forecast models were applied. The benefit of the stochastic approach was clearly shown when analysing the DORA performance over 13 events. While the reduction in total CSO volume obtained by the deterministic approach was around 32%, the stochastic approach obtained a reduction of almost 70%.

This study illustrates how the integration of stochastic runoff forecasts into global RTC strategies can provide a significant contribution to the improvement of the performance of urban drainage network and to reduce their impact on the natural waters.

## **5. Acknowledgements**

The results presented in this study are obtained under the framework of the SWI project (Storm- and Wastewater Informatics), a strategic Danish Research Project financed by the Danish Agency for Science Technology and Innovation under the Programme commission on sustainable energy and environment. Data are kindly provided by HOFOR under the framework of the METSAM project (MiljøEffektiv Teknologi til SAMstyring af afløb og renseanlæg), financed by the Danish Environmental Agency. The authors thank Tina Kunnerup Hestbæk (Krüger A/S) for providing the results of the detailed hydrodynamic model. Luca Vezzaro is an industrial postdoc financed by the Danish National Advanced Technology Foundation under the project “MOPSUS - Model predictive control of urban drainage systems under uncertainty”.

## **6. References**

Achleitner, S., Möderl, M., Rauch, W., 2007. CITY DRAIN © - An open source approach for

simulation of integrated urban drainage systems. Environmental Modelling and Software 22, 1184-1195.

Achleitner, S., Fach, S., Einfalt, T., Rauch, W., 2009. Nowcasting of rainfall and of combined sewage flow in urban drainage systems. Water Sci.Techol. 59, 1145-1151.

Beenken, T., Erbe, V., Messmer, A., Reder, C., Rohlfing, R., Scheer, M., Schuetze, M., Schumacher, B., Weilandt, M., Weyand, M., 2013. Real time control (RTC) of urban drainage systems GÇô A discussion of the additional efforts compared to conventionally operated systems. Urban Water Journal 10, 293-299.

Borsanyi, P., Benedetti, L., Dirckx, G., De Keyser, W., Muschalla, D., Solvi, A.M., Vandenberghe, V., Weyand, M., Vanrolleghem, P.A., 2008. Modelling real-time control options on virtual sewer systems. Journal of Environmental Engineering and Science 7, 395-410.

Borup, M., Grum, M., Mikkelsen, P.S., 2013. Comparing the impact of time displaced and biased precipitation estimates for online updated urban runoff models. Water Sci.Techol. 68,

Breinholt, A., Sharma, A.K., 2010 Case Area Baseline Report - Copenhagen Energy and Lynette Fælleskabet.

Breinholt, A., Thordarson, F.Ö., Møller, J.K., Grum, M., Mikkelsen, P.S., Madsen, H., 2011. Grey-box modelling of flow in sewer systems with state-dependent diffusion. Environmetrics 22, 946-961.

Carstensen, J., Nielsen, M.K., Harremoes, P., 1996. Predictive control of sewer systems by means of grey-box models. Water Sci.Techol. 34, 189-194.

Del Giudice, D., Honti, M., Scheidegger, A., Albert, C., Reichert, P., Rieckermann, J., 2013. Improving uncertainty estimation in urban hydrological modeling by statistically describing bias. Hydrology and Earth System Sciences 17, 4209-4225.

Deletic, A., Dotto, C.B.S., McCarthy, D.T., Kleidorfer, M., Freni, G., Mannina, G., Uhl, M., Henrichs, M., Fletcher, T.D., Rauch, W., Bertrand-Krajewski, J.L., Tait, S., 2012. Assessing uncertainties in urban drainage models. Physics and Chemistry of the Earth, Parts A/B/C 42-44, 3-10.

DHI, 2008 MIKE URBAN release 2008.

Dirckx, G., Schtze, M., Kroll, S., Thoeye, C., De Gueldre, G., Van De Steene, B., 2011. Cost-efficiency of RTC for CSO impact mitigation. Urban Water Journal 8.

Erbe, V., Schuetze, M., 2005. An integrated modelling concept for immission-based management of sewer system, wastewater treatment plant and river. Water Sci.Technol. 52, 95-103.

Fradet, O., Pleau, M., Marcoux, C., 2011. Reducing CSOs and giving the river back to the public: innovative combined sewer overflow control and riverbanks restoration of the St Charles River in Quebec City. Water Sci.Technol. 63, 331-338.

Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102, 359-378.

Grum, M., Longin, E., Linde, J.J., 2004 A Flexible and Extensible Open Source Tool for Urban Drainage Modelling: [www.WaterAspects.org](http://www.WaterAspects.org). 6th Urban Drainage Conference

Jørgensen, H.K., Rosenørn, S., Madsen, H., Mikkelsen, P.S., 1998. Quality control of rain data used

for urban runoff systems. Water Sci.Technol. 37, 113-120.

Krämer, S., Grum, M., Verworn, H.R., Redder, A., 2005. Runoff modelling using radar data and flow measurements in a stochastic state space approach. Water Sci.Technol. 52, 1-8.

Kristensen, N.R., Madsen, H., 2003 Continuous Time Stochastic Modeling - CTSM 2.3 - Mathematics Guide. Technical University of Denmark, Technical Report, DTU, Lyngby, Denmark ed.

Langeveld, J.G., Benedetti, L., de Klein, J.J.M., Nopens, I., Amerlinck, Y., van Nieuwenhuijzen, A., Flamelin, T., van Zanten, O., Weijers, S., 2013. Impact-based integrated real-time control for improvement of the Dommel River water quality. Urban Water Journal 10, 312-329.

Löwe, R., Mikkelsen, P.S., Madsen, H., 2014. Stochastic rainfall-runoff forecasting: parameter estimation, multi-step prediction, and evaluation of overflow risk. Stoch Environ Res Risk Assess 28, 505-516.

Löwe, R., Vezzaro, L., Mikkelsen, P.S., Grum, M., Madsen, H. (submitted) Investigating the use of probabilistic forecasts for RTC of urban drainage systems - A Layout for Probabilistic Online Forecasting of Sewer Flows. Submitted to Environmental Modelling & Software

Lund, N.S. V, Pedersen, J.W., Borup, M., Mikkelsen, P.S., Grum, M., 2014. Auto-Calibration For Updating Of Linear Reservoir Models Used In Flow Forecasting Of Urban Runoff. Hydrol. Processes. in preparation.

Moradkhani, H., DeChant, C.M., Sorooshian, S., 2012. Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method. Water Resour.Res. 48.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I GÇö A discussion of principles. *Journal of Hydrology* 10, 282-290.

Nielsen, M.K., Bechmann, H., Henze, M., 2000. Modelling and test of aeration tank settling (ATS). *Water Sci.Technol.* 41, 179-184.

Nielsen, M.K., Carstensen, J., Harremoes, P., 1996. Combined control of sewer and treatment plant during rainstorm. *Water Sci.Technol.* 34, 181-187.

Nielsen, N.H., Ravn, C., Mølbye, N., 2010 Implementation and design of a RTC strategy in the sewage system in Kolding, Denmark.

Ocampo-Martinez, C., Puig, V., 2010. Piece-wise linear functions-based model predictive control of large-scale sewage systems. *Iet Control Theory and Applications* 4, 1581-1593.

Pleau, M., Colas, H., Lavalle, P., Pelletier, G.v., Bonin, R., 2005. Global optimal real-time control of the Quebec urban drainage system. *Environmental Modelling and Software* 20, 401-413.

Puig, V., Cembrano, G., Romera, J., Quevedo, J., Aznar, B., Ramon, G., Cabot, J., 2009. Predictive optimal control of sewer networks using CORAL tool: application to Riera Blanca catchment in Barcelona. *Water Sci.Technol.* 60, 869-878.

Raso, L., van Overloop, P.J., Schwanenberg, D., Optimal control of water systems under uncertainty using stochastic approximation algorithm. In Proceedings of 10th International Conference on Hydroinformatics, HIC2012, Hamburg, Germany

Rauch, W., Seggelke, K., Brown, R., Krebs, P., 2005. Integrated Approaches in Urban Storm Drainage: Where Do We Stand? Environmental Management 35, 396-409.

Schilling, W., Fuchs, L., 1986. Errors in Stormwater Modeling - Quantitative Assessment. J.Hydraul.Eng. 112, 111-123.

Schilling, W., 1991. Rainfall data for urban hydrology: what do we need? Atmospheric Research 27, 5-21.

Seggelke, K., Löwe, R., Beeneken, T., Fuchs, L., 2013. Implementation of an integrated real-time control system of sewer system and waste water treatment plant in the city of Wilhelmshaven. Urban Water Journal 10, 330-341.

Sun, S., Bertrand-Krajewski, J.-L., 2013. Separately accounting for uncertainties in rainfall and runoff: Calibration of event-based conceptual hydrological models in small urban catchments using Bayesian method. Water Resour.Res. 49.

Thomsen, H.R., Önnnerth, T.B., 2009 Results and benefits from practical application of ICA on more than 50 wastewater systems over a period of 15 years. Keynote paper, 10th IWA Instrumentation, Control and Automation conference, Cairns, Queensland, Australia, 14-17 June 2009

Thorndahl, S., Beven, K.J., Jensen, J.B., Schaarup-Jensen, K., 2008. Event based uncertainty assessment in urban drainage modelling, applying the GLUE methodology. Journal of Hydrology 357, 421-437.

Thorndahl, S., Rasmussen, M.R., 2013. Short-term forecasting of urban storm water runoff in real-time using extrapolated radar rainfall data. Journal of Hydroinformatics 15, 897-912.

Thorndahl, S., Poulsen, T.S., Bøvith, T., Borup, M., Ahm, M., Nielsen, J.E., Grum, M., Rasmussen, M.R., Gill, R., Mikkelsen, P.S., 2013. Comparison of short-term rainfall forecasts for model-based flow prediction in urban drainage systems. *Water Sci Technol* 68, 472-478.

Vanrolleghem, P.A., Benedetti, L., Meirlaen, J., 2005. Modelling and real-time control of the integrated urban wastewater system. *Environmental Modelling & Software* 20, 427-442.

Vezzaro, L., Grum, M., 2012 A generalized Dynamic Overflow Risk Assessment (DORA) for urban drainage RTC. *Urban Drainage Modelling : Proceedings of the Ninth International Conference on Urban Drainage Modelling*, Belgrade, Serbia, 4-6 September 2012

Vezzaro, L., Löwe, R., Madsen, H., Grum, M., Mikkelsen, P.S., 2013 Investigating the use of stochastic forecast for RTC of urban drainage systems, *Proceedings of NOVATECH 2013*, Lyon, France, 2013.

Vrugt, J.A., Diks, C.G.H., Gupta, H.V., Bouten, W., Verstraten, J.M., 2005. Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. *Water Resour.Res.* 41.

Zhou, Q., Mikkelsen, P.S., Halsnæs, K., Arnbjerg-Nielsen, K., 2012. Framework for economic pluvial flood risk assessment considering climate change effects and adaptation benefits. *Journal of Hydrology* 414-415, 539-549.

PAPER G

# Forecasting Operational Runoff Forecast Uncertainties - State, Rainfall and Error Dependencies

---

**Authors:**

Roland Löwe, Rune Juhl, Peter Steen Mikkelsen, Henrik Madsen

**in preparation for:**

*Hydrology and Earth System Science Discussions*, 2014.



# Forecasting Operational Runoff Forecast Uncertainties - State, Rainfall and Error Dependencies

Roland Löwe<sup>1</sup>, Rune Juhl<sup>1</sup>, Peter Steen Mikkelsen<sup>2</sup>, Henrik Madsen<sup>1</sup>

## Abstract

This work focuses on structures for describing rainfall-runoff forecast uncertainties in stochastic grey-box models. Such models are applied for runoff forecasting in on-line applications such as real-time control. We consider four urban catchments with reduced areas between 43 and 900ha and three different model structures for describing forecast uncertainty. These structures describe forecast uncertainty depending on i)the forecasted model state, ii) a smoothed version of the rainfall input and iii) a smoothed version of the previously observed forecast errors.

The state dependent model structure is currently applied operationally and may be considered a benchmark. The results in this work suggest that probabilistic runoff forecasts would benefit from scaling the uncertainty term in the stochastic models depending on the smoothed rainfall input rather than the model state. This is supported by a correlation analysis between runoff forecast errors and an assessment of the skill of the probabilistic runoff forecasts generated in the different catchments on a horizon of 100min.

The presented article is a working paper which will undergo significant revisions in accordance to the limitations described in the discussions part. In particular, models estimated using the MAP approach will not be considered in the future as they do not lead to significantly different forecast performance. In addition, issues related to the violation of assumptions made for parameter inference will be addressed.

---

<sup>1</sup> Department of Applied Mathematics and Computer Science, Technical University of Denmark, Matematiktorvet, B322, DK-2800 Kgs. Lyngby, Denmark

<sup>2</sup> Department of Environmental Engineering, Technical University of Denmark, Miljøvej, B115, DK-2800 Kgs. Lyngby, Denmark

# Forecasting Operational Runoff Forecast Uncertainties - State, Rainfall and Error Dependencies

Roland Löwe<sup>1</sup>, Rune Juhl<sup>1</sup>, Peter Steen Mikkelsen<sup>2</sup>, and Henrik Madsen<sup>1</sup>

<sup>1</sup>Department of Applied Mathematics and Computer Science, Technical University of Denmark (DTU), Kgs. Lyngby, Denmark.

<sup>2</sup>Department of Environmental Engineering, Technical University of Denmark (DTU), Kgs. Lyngby, Denmark.

*Correspondence to:* Roland Löwe  
(rolo@dtu.dk)

**Abstract.** This work focuses on structures for describing rainfall-runoff forecast uncertainties in stochastic grey-box models. Such models are applied for runoff forecasting in online applications such as real-time control. We consider four urban catchments with reduced areas between 43 and 900ha and three different model structures for describing forecast uncertainty. These structures describe forecast uncertainty depending on i) the forecasted model state, ii) a smoothed version of the rainfall input and iii) a smoothed version of the previously observed forecast errors.

The state dependent model structure is currently applied operationally and may be considered a benchmark. The results in this work suggest that probabilistic runoff forecasts would benefit from scaling the uncertainty term in the stochastic models depending on the smoothed rainfall input rather than the model state. This is supported by a correlation analysis between runoff forecast errors and an assessment of the skill of the probabilistic runoff forecasts generated in the different catchments on a horizon of 100min.

The presented article is a working paper which will undergo significant revisions in accordance to the limitations described in the discussions part. In particular, models estimated using the MAP approach will not be considered in the future as they do not lead to significantly different forecast performance. In addition, issues related to the violation of assumptions made for parameter inference will be addressed.

## 1 Introduction

Recent real-time control schemes for urban catchments rely on on-line optimisation of the drainage network and the wastewater treatment plant (Pleau et al., 2005; Puig et al.,

2009; Vezzaro and Grum, 2014). Different control and optimisation algorithms are applied, but usually a forecast of the future states of the drainage system is required for an efficient control operation. This work concerns itself with the generation of on-line runoff forecasts for such operational control settings and in particular with the quantification of forecast uncertainties on forecast horizons of up to two hours.

In operational settings, forecast models need to be computationally efficient. A multitude of simulation runs may be required in an optimisation algorithm at every control time step and multiple ensemble simulations may be required for the consideration of forecast uncertainties (for example Raso et al. (2013)). The application of distributed models becomes infeasible in such contexts where exhaustive simulations are required (see also Achleitner et al. (2007); Wolfs et al. (2013)), although they are widely applied in urban hydrology.

In this work we therefore focus on the application of conceptual model structures for on-line forecasting. Such model structures also simplify the application of statistical techniques for model identification (Harremoës and Madsen, 1999; Kristensen et al., 2004b). Information from on-line measurements can easily be implemented in such models through on-line updating (Breinholt et al., 2011; Carstensen et al., 1998; Löwe et al., 2014c; Vrugt et al., 2005). Forecast errors resulting from errors or rainfall measurements and deficiencies in the model structure can thus be reduced.

Objective functions for real-time control typically depend non-linearly on the forecasted runoff value. Decision making is therefore expected to benefit from a quantification of forecast uncertainty. Vezzaro et al. (2014) and Löwe et al. (2014a) demonstrate such effects in urban catchments. On-

line runoff forecasting models should therefore provide a probabilistic forecast rather than a point value.

Stochastic grey-box models fulfill both requirements of being suitable for on-line purposes and providing a quantification of forecast uncertainties. As stochastic grey-box models we denote models based on stochastic differential equations (SDE) that are implemented in a state-space framework. The physical part of these models is a conceptual (or lumped) representation of the physical processes. The model parameters are, however, physically interpretable. Previous applications of such models for the identification and simulation of processes in urban hydrology have been documented by Bechmann et al. (1999), Bechmann et al. (2000), Breinholz et al. (2011), Carstensen et al. (1996), Carstensen et al. (1998) and Thordarson et al. (2012).

Breinholz et al. (2011) suggested a stochastic grey-box model structure for simulating and forecasting runoff from urban catchments where the predicted uncertainty depends directly on the predicted state value. Such a model structure is easy to implement but has drawbacks. First, it lumps forecast uncertainties for dry weather and rain periods into the same parameter. However, these uncertainties are clearly different and the lumping leads to misspecified parameters, often with too small forecast uncertainty during rain periods and too large forecast uncertainty during dry weather periods (Löwe et al., 2014b).

Second, the forecast uncertainty is not necessarily related to the forecast model states. In particular the runoff in the beginning of rain events has proven difficult to predict. If the forecast model underestimates the true runoff, the forecast uncertainty will also be underestimated if the uncertainty description involves a direct state dependency (Löwe et al., 2014b,c). On the contrary, runoff forecast uncertainties are typically small in the end of the rain event as we only need to describe the decreasing branch of the runoff curve. Flows in the sewer system, however, are still high, and a direct state dependency of the forecasted uncertainty will lead to an overestimation of forecast uncertainty.

The problems with a state-dependent description of forecast uncertainty motivate investigations into different model structures for describing forecast uncertainty. An obvious candidate is to model runoff forecast uncertainty depending on the measured and/or forecasted rainfall. If considering off-line simulation settings, we can find several such approaches in the literature. Rainfall multipliers are used in Bayesian approaches for a time-dynamic uncertainty description that varies from event to event, for example (Kuczera et al., 2006; Renard et al., 2010; Sun and Bertrand-Krajewski, 2013). Such approaches, however, require a huge number of model parameters and are not necessarily suitable for on-line purposes. Reichert and Schuwirth (2012) suggest a Bayesian approach for quantifying simulation where model outputs are corrected by a bias with both time-varying mean and variance. Del Giudice et al. (2013) suggest to vary the variance of this bias process depending on the rainfall input. In a direct

comparison, this approach appears to provide a better quantification of forecast uncertainties than the state dependency (Del Giudice et al., 2014). Also the bias approach is in its current implementation not suitable for on-line purposes due to the large number of model runs required, although more efficient implementations may be possible (Del Giudice et al., 2014).

A disadvantage of conditioning runoff forecast uncertainty on the rainfall input is that, other than in the state-dependent approach, the forecast uncertainty is not zero if the model states approach zero. This property, however, is very desirable to avoid negative flow predictions. In this work we investigate, how state and input dependent uncertainty description can be combined in a stochastic grey-box approach which is suitable for on-line purposes.

We commence the investigation by analysing the relation between runoff forecast errors and various potential explanatory variables. Such an analysis provides insight into the dependencies of forecast errors and we derive possible model structures for describing forecast uncertainty. Subsequently, we test the derived model structures by generating probabilistic forecasts in a setting corresponding to the on-line application. Probabilistic forecasts are generated through scenario simulations of the SDE's. An Euler-Maruyama scheme (Kloeden and Platen, 1999) has previously been applied for such simulations (Thordarson et al., 2012). However, such a scheme is problematic for stiff equations and consequently may not be applicable for on-line purposes. We therefore investigate a simple alternative and make suggestions for the required simulation time step and the number of ensemble members.

We conclude the article with an evaluation of the probabilistic forecasts.

## 2 Materials

### 2.1 Catchments

We consider four catchments in the Copenhagen area. In all catchments the majority of the runoff observed results from areas that are drained as combined sewer systems.

The Ballerup (BAL) and upper Damhusåen (DAM) catchments are neighboring catchments located in the western part of the Copenhagen. The reduced area is 43 and 900ha, respectively. These catchments were documented in a number of previous works (Breinholz et al., 2011, 2012, 2013; Löwe et al., 2014a,b; Thordarson et al., 2012). In the Ballerup catchment, a significant part of the runoff can result from the slow runoff component. The Damhusåen catchment is a larger catchment containing a number of combined sewer overflow points.

The Kløvermarken (KLO) and Lersøledning catchments (LER) are located next to each other in the western part of Copenhagen and were documented by Löwe et al. (2014c)

and Vezzaro et al. (2014). The KLO catchment is the catchment of the largest pumping station in Copenhagen and has a reduced area of 780ha. Lersøledning is a storage pipe with a volume of approximately 17,000m<sup>3</sup> and an upstream reduced catchment area of 730ha.

## 2.2 Considered Datasets

In the BAL and DAM catchments, we have rainfall measurements and in-sewer flow measurements available in 10min resolution. The rainfall measurements were derived by averaging the rain gauge measurements of 2 (BAL) and 4 (DAM) rain gauges located around the catchments as described in Löwe et al. (2014b). The considered data period comprises 40 days from 5th of July, 2010 until 14th of August, 2010. The period includes two smaller and two large rain events and a long dry weather period. The same dataset was used in Del Giudice et al. (2014). A 10min time step is considered for the simulation of these catchments, equivalently to the works by Del Giudice et al. (2014) and Löwe et al. (2014b).

The KLO and LER catchments are part of a global real-time control scheme implemented in the Copenhagen area (Grum et al., 2011). Rainfall and runoff data are available directly from the control server with a temporal resolution of 2 minutes, corresponding to the control time step. Rainfall observations are available mean areal rainfall derived from C-band radar measurements. The temporal resolution of the radar measurements is 10min and the rain intensity is kept constant for the five 2min intervals in between two radar observations.

Flow measurements in the KLO catchment correspond to the measured outflow from the pumping station with the inflow from upstream catchments (pumping stations) subtracted. In the LER catchment, flow measurements are derived from level measurements in and outflow measurements from the storage pipe using the water balance for the storage pipe. This process is described in Löwe et al. (2014c) and Vezzaro et al. (2014) and the same dataset was used for model calibration in this work. We consider four rain events from January, April, July and October 2012. The total length of the considered dataset is 130h or 3900 time steps.

Figure 1 shows the rainfall and runoff measurements for the BAL and KLO catchments. The dataset contains pronounced rain events that are relevant for on-line runoff forecasting. The rain events observed in the DAM and LER catchments are similar due to the proximity of the areas.

All datasets have flaws. In the BAL catchment the runoff can be strongly influenced by slow runoffs which has proven problematic for very simple on-line runoff forecasting models. Similarly, the rather big DAM catchment contains numerous structures which make runoff forecasting with simple model structures a challenge. In addition, in both catchments, the observations of rainfall through rain gauges (which in the case of the BAL catchment are located outside the catchment) can lead to significant bias in the estimated areal rain-

fall (Borup et al., 2013). In the KLO and LER catchments, the derivation of catchment outflow from water balances at the pumping station or the storage pipe leads to significant uncertainties in the flow measurements (see also Löwe et al. (2014c)).

The idea for this work is to reuse datasets that were applied in previous works and to use observations of a quality that is realistic with respect to the data that are available in an on-line setting. We need to keep this in mind when interpreting the model results. Still we will be able to demonstrate clear differences between the considered modelling approaches.

Finally, we do not consider actual rainfall forecast but assume the future rainfall known from the measurements when generating runoff forecasts. This is mainly due to the lack of historical data for the KLO and LER catchments. Consequently, the analysis below is limited to the modelling of runoff forecast uncertainties that result from structural model deficiencies and uncertainties in the rainfall observations.

## 3 Methods

### 3.1 Stochastic Greybox Models for On-line Runoff Forecasting

#### 3.1.1 Stochastic Rainfall Runoff Model

We apply a stochastic rainfall runoff model in a state space layout in the following form:

$$d \begin{bmatrix} S_{1,t} \\ S_{2,t} \end{bmatrix} = \underbrace{\begin{bmatrix} A \cdot P_t + a_0 - \frac{1}{K} S_{1,t} \\ \frac{1}{K} S_{1,t} - \frac{1}{K} S_{2,t} \end{bmatrix}}_{\text{Drift term}} dt + \underbrace{\begin{bmatrix} g(\sigma, t, u_t, S_{1,t}) & 0 \\ 0 & (\sigma, t, P_t, S_{2,t}) \end{bmatrix}}_{\text{Diffusion term}} d\omega_t \quad (1)$$

$$y_k = \frac{1}{K} S_{2,k} + d_k + e_k \quad (2)$$

In this setup Equation 1 is called the **state (or system) equation**. It consists of a system of coupled Itô stochastic differential equations. The physical part of these models is called the drift term and is based on lumped reservoir approaches that transform the rainfall input into the flow output. Breinholt et al. (2011) describe the principal model setup.  $S_1$  and  $S_2$  correspond to the states of the system, i.e. virtual storage fillings,  $A$  is the sealed area in the catchment,  $a_0$  refers to the mean dry weather flow at the catchment outlet, and  $K$  corresponds to the travel time constant. The rainfall input  $P_t$  is determined as the mean area rainfall by averaging the rainfall measurements considered for every catchment.

The drift term is combined with a stochastic part which we call diffusion term. The diffusion term is driven by a Wiener process  $\omega_t$ . Increments  $d\omega_t$  of this process are independent

and normally distributed with expectation zero and a variance corresponding to the considered time increment  $dt$ . In principle, it is possible to consider a multivariate process for all system states. However, as shown in Equation 1 we limit ourselves to the diagonal case. The variance of the stochastic process can be scaled time dynamically by a generic function  $g$  which can depend on time  $t$ , parameters  $\sigma$ , model inputs  $u_t$  or the system states  $S_{i,t}$ , for example. The diffusion term lumps structural and input uncertainty.

Equation 2 is called **observation equation** and relates the continuous system states to discrete flow observations  $y_k$  at time steps  $k$ . The variation of the dry weather flows  $d_k$  is described as a harmonic function with parameters  $s_1$ ,  $s_2$ ,  $c_1$ , and  $c_2$  and is shown in Equation 3. The observations are subject to a independent and normally distributed observation error  $e_k$  with expectation zero and variance  $\sigma_e^2$ .

An extended Kalman filter is applied to update the model states to current observations at every time step. For missing observations we increase the variance of the observation error  $\sigma_e^2$  to a large value (100) and consequently exclude the concerned value from updating and the evaluation of the likelihood function during parameter estimation (Section 3.1.4).

$$d_k = \sum_i^2 (s_i \sin \frac{i2\pi k}{24h} + c_i \cos \frac{i2\pi k}{24h}) \quad (3)$$

The physical model structure applied here is very simple and neglects infiltration, spatial distribution of rainfalls in the catchment and overflows. Nevertheless, it has demonstrated to provide reasonable forecasts on a two hour horizon in several catchments (Breinholt et al., 2011; Löwe et al., 2014b,c) and we deem it sufficient for this work where we mainly focus on the identification of functions  $g$  for scaling the variance in the diffusion term.

We refer the reader to Breinholt et al. (2011) for further details on stochastic grey-box model structures for urban runoff and to Kristensen et al. (2004a) for general information on the stochastic grey-box modelling setup.

### 3.1.2 Forecast Uncertainty Description in Stochastic Greybox Models

Breinholt et al. (2011) suggested a direct dependency of the variance scaling in the diffusion term of the  $i$ -th model state on the state value itself:

$$g_i = \sigma_{i,1} \cdot S_i. \quad (4)$$

We consider this approach a benchmark and denote it as **Model 1**. As described in section 1, this model structure can be disadvantageous because it does not distinguish between uncertainties for dry weather and rain periods and because wrong model forecasts will also lead to wrong uncertainty forecasts.

Our hypothesis is that we can obtain an improved description of forecast uncertainty by making the function  $g$  dependent on some external input such as the rainfall observations

instead. However, a positive feature of the state dependency is that it allows us to consider physical limitations for the state values. For example, in a linear reservoir cascade the model states cannot be less than zero.

A new model structure for forecast uncertainties should therefore be a combination of both approaches and we suggest the following structure for the diffusion term of state  $i$ :

$$g_i = (\sigma_{i,1} + \sigma_{i,2} \cdot F) \frac{S_i^2}{1 + S_i^2}. \quad (5)$$

In this structure,  $\sigma_{i,1}$  is a parameter that must be identified during model calibration and that describes a constant level of forecast uncertainty corresponding to the uncertainty observed during dry weather periods.  $F$  is an external forcing that describes the dynamic variation of forecast uncertainty. It is scaled by the parameter  $\sigma_{i,2}$ . The fraction ensures that forecast uncertainty approaches zero as the model state  $S_i$  approaches zero but remains unaffected by the state otherwise.

We investigate two variants for the external uncertainty forcing  $F$ . In the first variant, we use the rainfall input as external forcing but apply a smoothing to obtain a behaviour more similar to that of the runoff forecast errors. In addition, we lag the smoothed rainfall observations in time because runoff occurs lagged in time to the rainfall observations. We obtain the forcing  $F_k$  from the  $k$ -th time step as

$$F_k = P_{k-l}^S = \lambda \cdot P_{k-l-1}^S + (1 - \lambda) \cdot P_{k-l} \quad (6)$$

where  $P^S$  are the smoothed rainfall observations,  $l$  is a lag parameter (in no. of time steps),  $\lambda$  is a smoothing parameter and  $P_{k-l}$  is the rainfall measurement at time step  $k-l$ . We denote this approach as **Model 2**.

In the second variant, we try to scale the runoff forecast uncertainty by a smoothed version of the previously observed forecast errors. The approach is related to the generalized autoregressive conditional heteroskedasticity (GARCH) models that are applied for modelling error variances in econometrics (Bollerslev, 1986). We denote this approach as **Model 3** and obtain the following external forcing  $F_k$  for time step  $k$ :

$$F_k = \lambda \cdot F_{k-1} + (1 - \lambda) \cdot \epsilon_k. \quad (7)$$

In Equation 7  $\lambda$  is a smoothing parameter and  $\epsilon_k$  are the innovations (or one-step ahead prediction errors) obtained from extended Kalman filtering. Considering a flow observation  $y_k$  and the one-step ahead prediction  $\hat{y}_{k|k-1}$ , we find  $\epsilon_k = y_k - \hat{y}_{k|k-1}$ .

For both model 2 and model 3 the external forcing is kept constant when generating on-line forecasts, i.e.  $F_k$  is used to scale the forecast uncertainty for a runoff forecast from time step  $k$  up to time step  $k+j$ .

### 3.1.3 Lamperti Transformation

A state dependency in the diffusion term as in Equations 359  
4 and 5 requires higher order filters to obtain a stable results from the Kalman filtering (Vestergaard, 1998). A Lamperti transformation is therefore commonly applied to such state equations. This transformation results in a state equation which equivalently describes the stochastic process but 360  
has a time-constant variance in the diffusion term and a drift 395  
term which is modified accordingly. Such transformations are also favourable in the numerical simulation of SDE's because the transformed equation system is not subject to bounds of the state space which are imposed by the state dependence (such as states being required to be strictly greater than zero, (Møller, 2010)).

The Lamperti transformation is based on Itô's lemma (øksendal, 1998). Given an Itô-stochastic-differential-equation as in Equation 1, a transformed state 370  
 $Z_{i,t} = \phi(t, S_{i,t})$  can again be expressed as an Itô-process with

$$dZ_{i,t} = \left( \frac{\partial \phi}{\partial t} + f_i(\cdot) \frac{\partial \phi}{\partial S_{i,t}} + \frac{g_i^2(\cdot)}{2} \frac{\partial^2 \phi}{\partial S_{i,t}^2} \right) dt + g_i(\cdot) \frac{\partial \phi}{\partial S_{i,t}} d\omega_i. \quad (8)$$
405

In Equation 8  $f_i(\cdot)$  corresponds to the drift term of the original state equation. To obtain a transformed process with constant diffusion, the task is thus to define a transformation such 410  
that

$$\frac{1}{g_i(\cdot)} = \frac{\partial \phi}{\partial S_{i,t}} \quad (9)$$

Strictly speaking, we should consider a multivariate transformation as Equation 1 describes a system of state equations. However, we do not consider correlation between the states in the diffusion term and can therefore apply the transformation in Equation 9 separately to each state  $S_{i,t}$ . 415

The Lamperti transformation for the linear state dependency in Model 1 is described in detail in Breinholt et al. (2011) and Møller (2010). For the state dependency imposed by the fraction in Models 2 and 3, we apply Equation 9 to find 385  
the following relation between original state  $S_{i,t}$  and transformed state  $Z_{i,t}$ :

$$S_{i,t} = \frac{(\sigma_{i,1} + \sigma_{i,2}F) \cdot Z_i}{2} + \sqrt{\frac{((\sigma_{i,1} + \sigma_{i,2}F) \cdot Z_i)^2}{4} + 1}. \quad (10)$$
420

The new state equation for the transformed state  $Z_{i,t}$  becomes:

$$dZ_{i,t} = \left( 0 + f_i(\cdot) \frac{1 + S_{i,t}^2}{(\sigma_{i,1} + \sigma_{i,2}F) \cdot S_{i,t}^2} - \frac{(\sigma_{i,1} + \sigma_{i,2}F) \cdot S_{i,t}}{(1 + S_{i,t}^2)^2} \right) dt + 425 \\ + 1 \cdot d\omega_i \quad (11)$$
430

Equation 10 is substituted into Equation 11 to obtain the final transformed state equation. We note, that some rather strong non-linearities were introduced into the drift term as a result of the transformation.

### 3.1.4 Parameter Estimation

Parameter inference in stochastic grey-box models is typically performed in a frequentist setting and based on one-step ahead forecast errors (innovations) from extended Kalman filtering. The approach is described in detail by Kristensen et al. (2004a). The likelihood for a set of (in the general case  $l$ -dimensional) flow observations  $\mathbf{Y}_n$  for  $n$  time steps is expressed as a product of one-step ahead conditional densities:

$$L(\theta, \mathbf{Y}_n) = \left( \prod_{k=1}^n p(\mathbf{y}_k | \mathbf{y}_{k-1}, \theta) \right) p(\mathbf{y}_0 | \theta). \quad (12)$$

As the diffusion term is driven by a Wiener process with normal increments, it is usually reasonable to assume normal distribution for the one-step ahead conditional densities at least for short prediction horizons. We obtain these densities from extended Kalman filtering as one-step ahead predictions  $\hat{\mathbf{y}}_{k|k-1}$  with covariance matrix  $\mathbf{R}_{k|k-1}$ . If the model structure is suitable for the observations  $\mathbf{y}_n$ , the associated innovations  $\epsilon_k = \mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1}$  are normally distributed with mean zero and covariance matrix  $\mathbf{R}_{k|k-1}$ . The likelihood function can then be rewritten as

$$L(\theta, \mathbf{Y}_n) = \left( \prod_{k=1}^n \frac{\exp\left(-\frac{1}{2}\epsilon_k^T \mathbf{R}_{k|k-1}^{-1} \epsilon_k\right)}{\sqrt{\det(\mathbf{R}_{k|k-1})} (\sqrt{2\pi})^l} \right) p(\mathbf{y}_0 | \theta) \quad (13)$$

We do in this set assume that the stochastic system described by the state equations is characterized by a normal distribution on short horizons, such that the assumptions made during Kalman filtering and parameter inference are fulfilled. This assumption should be verified as demonstrated in Del Giudice et al. (2014). Further, the innovations will only have an expected value of zero if the model structure is suitable for the data. This assumption should be checked using residual analysis procedures on the innovations. An advantage of this estimation approach is that no assumption is made about the correlation of observations for different time steps.

Prior information about model parameters can be considered in this setting by performing maximum a posteriori estimation (MAP) (Goodwin and Payne, 1977; Walter and Pronzato, 1997). This approach is usually applied if the data do not provide sufficient information to identify the system (for example in Melgaard (1994) and Sadegh et al. (1994)).

Assuming that the priors for the  $p$ -dimensional parameter vector  $(\theta)$  are normally distributed with mean  $\mu_\theta$ , and covariance  $\Sigma_\theta$ , the posterior probability density function can

be rewritten as

$$\begin{aligned} \mathbf{L}(\theta, \psi, \mathbf{Y}_n) = & \left( \prod_{k=1}^N \frac{\exp\left(-\frac{1}{2}\epsilon_k^T \mathbf{R}_{k|k-1}^{-1} \epsilon_k\right)}{\sqrt{\det(\mathbf{R}_{k|k-1})}(\sqrt{2\pi})^l} \right) p(\mathbf{y}_0|\theta) \\ & \frac{\exp\left(-\frac{1}{2}\epsilon_\theta^T \boldsymbol{\Sigma}_\theta^{-1} \epsilon_\theta\right)}{\sqrt{\det(\boldsymbol{\Sigma}_\theta)}(\sqrt{2\pi})^p} \end{aligned} \quad (14)$$

where  $\epsilon_\theta = \theta - \mu_\theta$  (Kristensen et al., 2004a).

We infer the following model parameters from the datasets: initial (transformed) model states  $Z_{1,0}$ ,  $Z_{2,0}$ , reduced catchment area  $A$ , mean dry weather flow  $a_0$ , reservoir time constant  $k$ , state uncertainty scaling  $\sigma_{1,1}$ ,  $\sigma_{2,1}$  (all models) and  $\sigma_{1,2}$ ,  $\sigma_{2,2}$  (models 2 and 3), measurement error variance  $\sigma_e^2$ , smoothing parameter  $\lambda$  (models 2 and 3) and time lag  $l$  for the external forcing (model 2).

We consider the dry weather variation  $d_k$  in Equation 3 a deterministic bias in the model output and estimate the corresponding parameters  $c_1$ ,  $c_2$ ,  $s_1$  and  $s_2$  separately during a dry weather period using least squares.

### 3.1.5 On-line Forecast Generation Through Simulation of Stochastic Differential Equations

We consider forecasts of runoff volume over horizons of 100 minutes in this work. In the BAL and DAM catchments (data resolution 10 minutes), this corresponds to a forecast horizon of 10 time steps. In the KLO and LER catchments (data resolution 2 minutes), a forecast horizon of 50 time steps is considered. We generate probabilistic forecasts of runoff by creating scenario (or ensemble) simulations of the model.

The starting point for the simulations is a sample from the multivariate normal distribution (Venables and Ripley, 2002) of the updated model states obtained from extended Kalman filtering (Kristensen et al., 2004a). We consider  $N$  ensemble members (see section 4.1) and for each ensemble member we simulate the transformed state equations up to the considered forecast horizon. From the observation equation 2 we obtain a flow value at every forecast time step (in 10 minutes resolution for the BAL and DAM catchments and 2 minutes resolution for the KLO and LER catchments). Integrating these flow values, we can for each scenario derive the forecasted runoff volume at the considered horizon of 100 minutes. Considering the distribution of the  $N$  ensemble members, we obtain a probabilistic forecast of runoff volume. This procedure is documented in Löwe et al. (2014c).

The simulation of the stochastic differential equations is simplified by the Lamperti transformation which yields a constant diffusion term. In previous works in urban hydrology (Löwe et al., 2014c; Thordarson et al., 2012) and other fields (Møller et al., 2012), an explicit Euler-Maruyama scheme is commonly applied (Kloeden and Platen, 1999). This scheme is fast and thus very suitable for on-line purposes. However, it cannot be applied for stiff equation systems. For the strongly non-linear transformed state equations

in Equation 11, the Euler-Maruyama scheme is likely to yield unstable solutions.

A simple alternative is the split-step backward Euler method (SSBE) (Higham et al., 2002), where the drift term of the SDE is solved implicitly:

$$Z_{i,t+h}^* = Z_{i,t} + h \cdot f_i(Z_{i,t+h}^*) , \quad (15a)$$

$$Z_{i,t+h} = Z_{i,t+h}^* + g_i(Z_{i,t+h}^*) \cdot \Delta \omega_{i,t} . \quad (15b)$$

In Equation 15  $h$  is the time increment considered for simulating the SDE,  $f_i$  is the drift term of the  $i$ -th transformed state equation and  $g_i$  the corresponding diffusion term. In our case,  $g_i = 1$ . The random numbers  $\Delta \omega_{i,t}$  are normally distributed with variance  $h$  and are not correlated for different states  $Z_i$  and time steps  $t$ . We generate the random disturbances as a Latin Hypercube sample (Stein, 1987) with  $N$  members at every time increment  $t+h$ . For the solution of the deterministic problem in Equation 15a we do not limit ourselves to the backward Euler method, but apply BDF methods up to order 5 through the DLSODE package instead (Hindmarsh, 1983; Radhakrishnan and Hindmarsh, 1993).

The time increment  $h$  in Equation 15 determines at what intervals noise is added to the deterministic model solution. It is not obvious what increment  $h$  and how many ensemble members  $N$  are required to obtain a reliable representation of the stochastic process. We therefore tested different combinations of  $h$  and  $N$  for Models 1 and 2 in the Ballerup catchment and chose  $N = 5000$  and  $h = 60$ s (see Section 4.1).

### 3.1.6 Prior Estimation of Measurement Errors

In previous works (Breinholt et al., 2011; Löwe et al., 2014a,b) it was observed that the identification of the measurement error variance can be difficult when applying stochastic grey-box models for runoff forecasting in urban catchments. An approach to avoid this behaviour can be to consider prior information for this variance during parameter estimation as described in Section 3.1.4 and we consider model variants accounting for prior information on observation error variance.

Dette et al. (1998) discuss estimators for observation error variance in nonparametric regression. We here consider the estimator proposed by Gasser et al. (1986) for  $n$  observations  $Y_i$ .

$$\hat{\sigma}_e^2 = \frac{2}{3(n-2)} \sum_{i=3}^n \left( \frac{1}{2} Y_{i-2} - Y_{i-1} + \frac{1}{2} Y_i \right)^2 \quad (16)$$

Generally, we should not assume the observation error variance constant as suggested by Vrugt et al. (2005). For different flow levels, different error sources may be relevant in the flow measurements. This is in particular true for the catchments where flow observations are derived not by direct velocity measurements in the pipes, but through indirect measurements such as stage curves or mass balance considerations.

However, in this work we only want to investigate if we can generally improve the model fit using prior information and therefore content ourselves with the estimator above. We apply lognormal prior distributions with a mean according to the above estimator and a standard deviation of 0.5.

### 3.2 Summary of Model Variants

In the following, we consider six different stochastic grey-box models for forecasting runoff in four different urban catchments. The models distinguish themselves in the structure of the diffusion term according to Section 3.1.2 and whether we consider prior information for the measurement error variance during parameter calibration or not. We name the approaches as follows:

- Models without prior information: Model1 (linear state dependence of diffusion), Model 2 (dependence of diffusion on external rainfall input forcing), Model 3 (dependence of diffusion on external GARCH forcing)
- Models with prior information: Model1\_P, Model2\_P, Model3\_P

Note, that prior information is exclusively used for the measurement error variance. We assume that no information is available for the other model parameters and consider non-informative prior distributions during the MAP estimation.

### 3.3 Forecast Verification

We evaluate forecasts of runoff volume on a forecast horizon of 100 minutes and compare them to the corresponding observations. We denote the runoff volume observed over a horizon from  $t$  to  $t + 100\text{min}$  as  $v_t$ . The (probabilistic) forecast of runoff volume on the same horizon is denoted as  $V_t$ .

**Reliability** of the probabilistic forecasts is analysed through the probability integral transform (PIT, (Gneiting et al., 2007)). If the forecasted distribution of runoff volumes is reliable, the observations correspond to realizations from this distribution. The values of the forecasted cumulated distribution functions  $F(v_t) = p(V_t \leq v_t)$  obtained for the observations  $v_t$  should then be uniformly distributed on the interval  $[0,1]$ . This can be checked in a predictive QQ-plot (Renard et al., 2010; Thyer et al., 2009).

The QQ-plots are summarized through two quantities denoted  $\alpha$  and  $\xi$  that were suggested by Renard et al. (2010). The index  $\alpha$  reflects the overall reliability of the forecasted distributions and is related to the area between the empirical QQ-line and 1:1 line. The index  $\xi$  is the complement of the fraction of observations which are not included in the predicted range. Both indices vary between zero (worst) and one (the perfect forecast) (Renard et al., 2010).

For the assessment of **resolution** (or sharpness) of the predictive distribution of runoff volumes we apply the coefficient of variation  $CV$  for the predictive distributions aver-

aged of all  $n$  time steps (Evin et al., 2014):

$$CV = \frac{1}{n} \sum_{t=1}^n \frac{sdev(Y_t)}{mean(Y_t)} \quad (17)$$

Both moments are estimated from the forecasted ensemble of runoff volumes. Given a reliable probabilistic forecast, we prefer to obtain small coefficients of variation. A measure based on quantiles (such as the quartile coefficient of dispersion) may be preferable for assessing forecast resolution because the forecasted distributions need not be symmetric. However, previous works (Evin et al., 2014; Renard et al., 2010) have applied the  $CV$  and we use it here for better comparability.

### 3.4 Numerical Implementation

The stochastic grey-box models are implemented in *CTSM-R* (Juhl et al., 2013) which is available as a package for the statistical software package *R* (R-Core-Team, 2013). We use this package for model implementation, automated symbolic differentiation and extended Kalman filtering.

For the generation of on-line runoff forecasts we generate multivariate normal samples from the updated model states using the R-package *MASS* (Venables and Ripley, 2002). The generation of random numbers for simulation of the SDE's is performed by generating a Latin Hypercube sample at every time increment using the R-package *LHS* (Carnell, 2012). The simulation of the SDE's itself is performed through calls from *R* to *Fortran* routines to gain computational speed. *DLSODE* from the ODE solver package *ODEPACK* (Hindmarsh, 1983; Radhakrishnan and Hindmarsh, 1993) is applied for solving the drift term of the SDE's.

For numerical optimization during parameter estimation we have applied the *PORT* algorithm (Gay, 1990) through *R* function *nlminb*. For model 3 this was combined with the *DDS* algorithm (Tolson and Shoemaker, 2007) because the time lag of the external forcing is an integer variable.

## 4 Results

### 4.1 Simulation of the Stochastic Differential Equations

To analyse what increment  $h$  and what number of ensemble members  $N$  should be used when simulation the state equations for on-line forecast generation, we perform a simulation experiment. We consider models 1 and 2 with ML parameter estimates in the Ballerup catchment. We omit model 3 as it is based on the same uncertainty description as model 2, using only a different external forcing  $F$ .

We consider a period of 400 time steps in  $\Delta t=10\text{min}$  resolution (66h), starting in a dry weather period and ending just after the flow peak of the biggest rain event in the dataset available for the catchment. The minimal flow during this period is  $0.03\text{m}^3/\text{s}$ , the maximal flow is  $0.53\text{m}^3/\text{s}$ .

Starting from a multivariate sample of the updated model states, we simulate the flow using the stochastic grey-box model 400 time steps forward in time as described in Section 3.1.5, considering varying numbers of ensemble members  $N$  and time increments  $h$ . We perform five ensemble simulations for each combination of  $N$  and  $h$  and, considering any time point within the considered period of 400 time steps, evaluate:

- the maximal variation of the ensemble standard deviation between the 5 simulation runs,
- the maximal variation of the ensemble median between the 5 simulation runs.

The results are shown in Figures 2 and 3. In both cases it is evident that a too small number of ensemble members leads to strong variations in both the median and the standard deviation of the probabilistic simulation. However, there is no clear trend regarding the time increment  $h$ . The variations are bigger for model 1, where we observe variations of up to 40% in terms of ensemble standard deviation and up to 10% in terms of ensemble median between the 5 simulation runs when considering too few ensemble members.

Even with  $N=10,000$  ensemble members and a time increment of  $h=5\text{s}$  variations between the standard deviations from the 5 simulation runs can still be observed (Figures 3 and 2). Considering very large numbers of ensembles is problematic for on-line purposes and we here define a maximal variation of  $0.01\text{m}^3$  for the ensemble standard deviation as an acceptable level of variation. We observe such variations during the peak of the rain event with a median flow of approximately  $0.4\text{m}^3\text{s}^{-1}$  (Figures 3 and 2). When choosing  $N=5,000$  and  $h=60\text{s}$  we fulfill this criterion.

## 4.2 Verification of Assumptions for Parameter Calibration

During parameter estimation we assume that simulations of the stochastic system described by Equations 1 and 2 can be assumed normal on short horizons of one time step (Kristensen et al., 2004a). We verify this assumption by generating an ensemble of 50,000 simulations for a rain period as described in Del Giudice et al. (2014).

Generally, normality can be assumed for models 2 and 3 in all catchments. The assumption is not fulfilled in all cases for model 1. This is induced by the Lamperti transformation for this model type which introduces an exponential backtransformation of the model states (Breinholst et al., 2011), while the backtransformation is approximately linear for models 2 and 3 (Equation 10). For the BAL and DAM catchments, the uncertainty of the states is rather small and the simulations can be assumed approximately normal (Figure 4a). However, for the KLO and LER catchments this assumption does not hold. A logarithmic transformation should be applied to the observation equation in these cases, but it cannot be applied to the negative flow observations for these catchments.

From Figures 4b and 4c we see, that the assumption of independent innovations is approximately fulfilled in THE BAL catchment only, while strong autocorrelations are observed for all models in the other catchments. These are the result of the far too simple model structure for these catchments.

## 4.3 Correlation between Forecast Errors and Input Variables

As an initial analysis into the dependencies between the runoff forecast errors we observe in an on-line setting and possible explanatory variables, we perform a correlation analysis. We use the median of the probabilistic runoff volume forecasts generated for model 1 in Löwe et al. (2014b) and Löwe et al. (2014c) and, for a given horizon, find a time series of forecast residuals by subtracting the observations and taking the absolute value. For forecast horizons of 10, 50 and 100 minutes we analyse how well this time series correlates with

- the forecast runoff volume for the considered horizon, estimated by the median of the probabilistic forecast,
- a smoothed time series of one-step ahead prediction errors (innovations) according to Equation 7, where the smoothing parameter is determined such that the correlation is maximized on average over the three considered horizons,
- the observed rainfall at the starting of the runoff forecast ( $P_t$  for a forecast from  $P_t$  to  $P_{t+k}$ ),
- a smoothed version of the observed rainfall up to the starting of the runoff forecast according to Equation 6, where the smoothing parameter is again determined such that the correlation is maximized.

The results of this analysis are shown in Figure 5. The results vary in between the different catchments. The forecast median is well correlated with the forecast error in the BAL catchment but yields only little information about forecast uncertainty in the DAM catchment, for example. The smoothed innovations, as expected, generally explain forecast uncertainty well on the short horizon but mostly yield little value on the longer horizons. An exception is the LER catchment ( $\lambda=0.87$ ).

On the long horizons, a smoothed version of the rainfall observations performs consistently well in explaining runoff forecast uncertainty and generally yields better results than directly using the rainfall observation.

The performance of the different approaches for modelling runoff forecast uncertainty considered in this article should depend on the catchment characteristics. However, the consistently high correlation values obtained for the smoothed rainfall observations suggest that model 2 may yield good forecast results.

#### 4.4 Probabilistic Forecast Performance on a 100min horizon

We consider first the models calibrated without prior information. Figures 6 and 8 suggest that model 2 yields the best calibrated forecasts for all catchments but the LER catchment. The predictive cumulative probabilities obtained for the observations more closely resemble a uniform distribution which results in a QQ-line that is more close to the diagonal in Figure 6 and smaller  $\alpha$ -values in Figure 8.

In the left part of Figure 9 it can be seen how the forecast uncertainty scales differently in model 2 as compared to model 1 in the DAM catchment. We can observe a slightly earlier increase of forecast uncertainty in the beginning of the rain event and the decrease of forecast uncertainty following the forecast value around time point 1174 is avoided. The new uncertainty structure proposed in model 2 thus appears to behave in a more desirable way than the direct state dependence in model 1.

Model 3 yields extremely small forecast uncertainties in the BAL and the DAM catchments and similar forecast uncertainties to model 2 in the KLO and LER catchment. The probabilistic forecasts obtained for this model tend to have lower  $\alpha$ -values than for the other two models and can consequently be considered less reliable. In addition, forecasts generated using this model include the least observations within their predictive range. Such behaviour is observed in all catchments, although the model achieved the best likelihood values during parameter estimation in the DAM and LER catchments (Table 1).

The  $\alpha$  values are generally smaller in the KLO and LER catchments. In addition, the  $\xi$ -values are smaller which indicates that a higher portion of the observations is not included in the predictive range. This problem is particularly pronounced for models 2 and 3. Model 1 includes a higher percentage of observations in the predicted range, however, at the cost of extremely wide predictive bands which can be seen in the right part of Figure 9 and lead to  $CV$  values between 3 and 5 (Figure 8).

Finally, we cannot identify a clear trend for a different forecast performance of the models calibrated using prior information for the measurement noise through a MAP approach. The results are generally very similar to the results obtained without prior information. Exceptions are model 2 in the DAM catchment, where a slightly worse model fit is obtained, and model 3 in the BAL catchment, where the forecast median seems to better match the observations which results in higher  $\alpha$  values but also even smaller forecast uncertainty than in the case without prior information.

## 5 Discussion

Considering first the simulation results over a long horizon that were obtained for models 1 and 2 in the BAL catch-

ment (Figures 2 and 3), it seems that model 2 yields more stable simulation results already for smaller numbers of ensembles. For both model structures the median of the simulations converges more quickly than the standard deviation of the ensembles. The stronger variations in the standard deviation of the simulations generated by model 1 may well be related to the Lamperti transformation applied for this model structure which applies an exponential backtransform of the states.

To our surprise, the simulation results were hardly affected by the considered time increment  $h$ . We have performed this experiment in the Ballerup catchment, where models 1 and 2 yield very good forecast results. It would be interesting to repeat the experiment in a different catchment where the forecast uncertainty is higher. Furthermore, it would be interesting to compare the simulation results for the Euler-type scheme that was applied here to those of higher order schemes and the time increment required for obtaining stable solutions in such schemes.

In terms of forecast performance, we see that model 2 consistently provides good results for all catchments and mostly outperforms model 1. This suggests that scaling the diffusion term of the stochastic grey-box models depending on the rainfall input is an attractive alternative to the linear state dependence used so far. For model 3, we have not been able to identify a model with strong forecast performance in any of the considered catchments, although the model yields the best likelihood values in the DAM and LER catchments.

An extremely big dry weather flow parameter  $a_0$  of  $6.5\text{m}^3/\text{s}$  is estimated for model 1 in the KLO catchment. A very large forecast uncertainty is estimated for this model and the (Lamperti transformed) drift term is strongly affected by this diffusion scaling. Simulations with the mean dry weather flow fixed to the (correct) value of approximately  $0.65\text{m}^3/\text{s}$  have yielded a strong underestimation of dry weather flows.

In the DAM, KLO and LER catchments the assumption of independent residuals is clearly not fulfilled (Figure 4). The reason is that the structure of the rainfall runoff model is too simple to describe the catchment behaviour. For these catchments, more complex model structures should be considered that account, for example, for the spatial variability of rainfall (see Löwe et al. (2014b)). However, for the KLO and the LER catchment we're bound by the available rainfall observations which were provided as an areal mean over these larger catchments.

Furthermore, the assumption of normality of the one-step ahead simulation is clearly not fulfilled for model 1 in the KLO and LER catchments. A log-transformation should be applied to the observation equation in these cases, but is hindered by negative flow observations. We should consider increasing the simulation time step for these models to 10 minutes, for instance, to smoothen short term variations in the flow data. An increased simulation time step would also have the advantage that the 100min forecast horizon becomes shorter in terms of numbers of time steps for these cases. The

rather long forecast horizon of 50 time steps may well be the reason of the generally inferior forecast performance in the KLO and the LER catchment, because the model parameters are estimated based on one-step ahead predictions.

880

Despite the above limitations, the forecasts generated by models 1 and 2 do yield value in terms of describing the future runoff volume. Further steps will focus on removing the above limitations and also on considering parameter estimation procedures that focus on multi-step forecasts (Löwe et al., 2014a).

885

## 6 Conclusions

Considering four urban drainage catchments with reduced areas between 43 and 900ha, we have evaluated the ability

to generate forecasts and to quantify forecast uncertainty of stochastic grey-box models with three different approaches for modelling uncertainty. The first approach (model 1) scales forecast uncertainty with a direct linear dependence on the model states, the second approach (model 2) applies a scaling depending on a smoothed version of the previous rainfall measurements, while the third approach applies a scaling depending on a smoothed version of the previously observed forecast errors.

The uncertainty analysis in this work is limited to the quantification of structural deficiencies in the model and uncertainty in the rainfall measurements. We do not consider the uncertainty of rainfall forecasts.

900

At this stage, we can identify the following conclusions:

1. The uncertainty scaling depending on the smoothed rainfall observations appears to be most robust and yields good forecasts in all cases. This is also backed by a correlation analysis between forecast errors and external variables, where the smoothed rainfall input yields the best results.
2. We can identify significant violations of the assumptions made for parameter inference. In the larger catchments, more appropriate physical model structures should be considered to obtain innovations that approximately fulfil the assumption of being independent. For model 1, a logarithmic transformation of the observation equation should be applied, such that simulations on short horizons can be assumed normal. However, this requires investigations into smoothing the data or increasing the simulation time step to remove the frequent negative flow observations in some catchments.
3. Imposing a prior on the measurement error does not yield a significant improvement in forecast performance over the models estimated without prior information.
4. The likelihood values obtained during parameter estimation do not necessarily coincide with the forecasts performance of the models. In particular, model 3 does

905

865

870

875

in some cases score very good likelihood values but generates forecasts where the uncertainty is strongly underestimated. Forecast models should be estimated with consideration to the application which is multi-step predictions in this case.

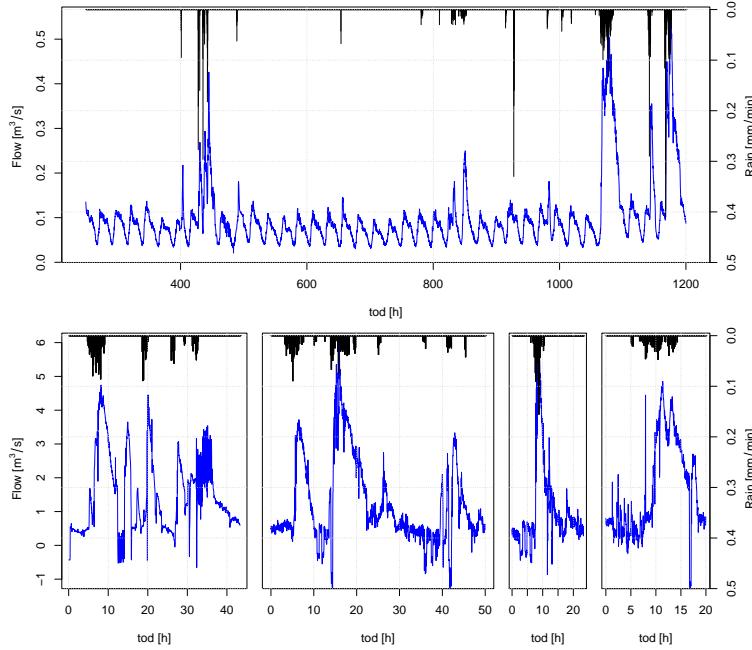
5. We can simulate the stochastic grey-box models using a combination of Euler scheme and an implicit solution of the drift part. The simulation result depends little on the considered time increment, but strongly on the considered number of scenarios.

## References

- Achleitner, S., Möderl, M., and Rauch, W.: CITY DRAIN © – An open source approach for simulation of integrated urban drainage systems, *Environ. Model. Softw.*, 22, 1184–1195, doi:10.1016/j.envsoft.2006.06.013, <http://linkinghub.elsevier.com/retrieve/pii/S136481520600171X>, 2007.
- Bechmann, H., Nielsen, M. K., Madsen, H., and Poulsen, N. K.: Grey-box modelling of pollutant loads from a sewer system, *Urban Water*, 1, 71–78, 1999.
- Bechmann, H., Madsen, H., Poulsen, N. K., and Nielsen, M. K.: Grey box modeling of first flush and incoming wastewater at a wastewater treatment plant, *Environmetrics*, 11, 1–12, 2000.
- Bollerslev, T.: Generalized autoregressive conditional heteroskedasticity, *J. Econom.*, 31, 307–327, doi:10.1016/0304-4076(86)90063-1, <http://www.sciencedirect.com/science/article/pii/0304407686900631>, 1986.
- Borup, M., Grum, M., and Mikkelsen, P. S.: Comparing the impact of time displaced and biased precipitation estimates for online updated urban runoff models, *Water Sci. Technol.*, 68, 109–116, doi:10.2166/wst.2013.221, 2013.
- Breinholt, A., Thordarson, F. O., Møller, J. K., Grum, M., Mikkelsen, P. S., and Madsen, H.: Grey-box modelling of flow in sewer systems with state-dependent diffusion, *Environmetrics*, 22, 946–961, doi:10.1002/env.1135, 2011.
- Breinholt, A., Møller, J. K., Madsen, H., and Mikkelsen, P. S.: A formal statistical approach to representing uncertainty in rainfall-runoff modelling with focus on residual analysis and probabilistic output evaluation – Distinguishing simulation and prediction, *J. Hydrol.*, 472–473, 36–52, doi:10.1016/j.jhydrol.2012.09.014, <http://www.sciencedirect.com.globalproxy.cvt.dk/science/article/pii/S0022169412007640>, 2012.
- Breinholt, A., Grum, M., Madsen, H., Örn Thordarson, F., and Mikkelsen, P. S.: Informal uncertainty analysis (GLUE) of continuous flow simulation in a hybrid sewer system with infiltration inflow – consistency of containment ratios in calibration and validation?, *Hydrol. Earth Syst. Sci.*, 17, 4159–4176, doi:10.5194/hess-17-4159-2013, <http://www.hydrol-earth-syst-sci.net/17/4159/2013/>, 2013.
- Carnell, R.: lhs: Latin Hypercube Samples, <http://cran.r-project.org/package=lhs>, 2012.
- Carstensen, J., Nielsen, M. K., and Harremoës, P.: Predictive Control of Sewer Systems by Means of Grey-box Models, *Water Sci Technol*, 34, 189–194, 1996.

- 930 Carstensen, J., Nielsen, M. K., and Strandbæk, H.: Prediction of hydraulic load for urban storm control of a municipal WWT plant, *Water Sci. Technol.*, 37, 363–370, 1998.
- 935 Del Giudice, D., Honti, M., Scheidegger, A., Albert, C., Reichert, P., and Rieckermann, J.: Improving uncertainty estimation in urban hydrological modeling by statistically describing bias, *Hydrol. Earth Syst. Sci.*, 17, 4209–4225, doi: 10.5194/hess-17-4209-2013, www.hydrol-earth-syst-sci.net/17/4209/2013/, 2013.
- 940 Del Giudice, D., Löwe, R., Madsen, H., Mikkelsen, P. S., and Rieckermann, J.: A critical appraisal of two stochastic approaches to predict urban rainfall-runoff with explicit consideration of model bias, *Water Resour. Res.*, submitted, 2014.
- Dette, H., Munk, A., and Wagner, T.: Estimating the variance in nonparametric regression - what is a reasonable choice?, *J. R. Stat. Soc. Ser. B-Methodological*, 60, 751–764, doi:10.1111/1467-9868.00152, 1998.
- Evin, G., Thyer, M., Kavetski, D., Mcinerney, D., and Kuczera, G.: Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity, *Water Resour. Res.*, accepted, 2014.
- Gasser, T., Sroka, L., and Jennen-Steinmetz, C.: Residual variance and residual pattern in nonlinear regression, *Biometrika*, 73, 625–633, 1986.
- 945 Gay, D. M.: Usage Summary for Selected Optimization Routines, Tech. rep., AT&T Bell Laboratories, Murray Hill, NJ, United States, <http://netlib.bell-labs.com/cm/cs/cstr/153.pdf>, 1990.
- Gneiting, T., Balabdaoui, F., and Raftery, A. E.: Probabilistic forecasts, calibration and sharpness, *J. R. Stat. Soc. Ser. B-Methodological*, 69, 243–268, doi:10.1111/j.1467-9868.2007.00587.x, 2007.
- 950 Goodwin, G. C. and Payne, R. L.: Dynamic system identification : Experiment design and data analysis, Academic Press, 1977.
- Grum, M., Thornberg, D., Christensen, M. L., Shididi, S. A., and Thirsing, C.: Full-Scale Real Time Control Demonstration Project in Copenhagen's Largest Urban Drainage Catchments, in: Proc. 12th Int. Conf. Urban Drain., Porto Alegre, Brazil, 2011.
- 955 Harremoës, P. and Madsen, H.: Fiction and reality in the modelling world—Balance between simplicity and complexity, calibration and identifiability, verification and falsification, *Water Sci. Technol.*, 39, 1–8, 1999.
- Higham, D. J., Mao, X., and Stuart, A. M.: Strong convergence of Euler-type methods for nonlinear stochastic differential equations, *SIAM J. Numer. Anal.*, 40, 1041–1063, 2002.
- 960 Hindmarsh, A. C.: ODEPACK, A Systematized Collection of ODE Solvers, *IMACS Trans. Sci. Comput.*, 1, 55–64, 1983.
- Juhl, R., Kristensen, N. R., Bacher, P., Möller, J. K., and Madsen, H.: CTSM-R User Guide, CTSM.info, 2013.
- Kloeden, P. E. and Platen, E.: Numerical Solution of Stochastic Differential Equations, vol. 3rd, Springer, Berlin-Heidelberg-New York, 1999.
- 965 Kristensen, N. R., Madsen, H., and Jørgensen, S. B.: Parameter estimation in stochastic grey-box models, *Automatica*, 40, 225–237, doi:10.1016/j.automatica.2003.10.001, 2004a.
- Kristensen, N. R., Madsen, H., and Jørgensen, S. B.: A method for systematic improvement of stochastic grey-box models, *Comput. Chem. Eng.*, 28, 1431–1449, doi:10.1016/j.compchemeng.2003.10.003, http://linkinghub.elsevier.com/retrieve/pii/S0098135403002758, 2004b.
- 970 Kuczera, G., Kavetski, D., Franks, S., and Thyer, M.: Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters, *J. Hydrol.*, 331, 161–177, doi:10.1016/j.jhydrol.2006.05.010, 2006.
- Löwe, R., Mikkelsen, P. S., and Madsen, H.: Stochastic rainfall-runoff forecasting: parameter estimation, multi-step prediction, and evaluation of overflow risk, *Stoch. Environ. Res. Risk Assess.*, 28, 505–516, doi:10.1007/s00477-013-0768-0, 2014a.
- 975 Löwe, R., Thorndahl, S., Mikkelsen, P. S., Rasmussen, M. R., and Madsen, H.: Probabilistic online runoff forecasting for urban catchments using inputs from rain gauges as well as statically and dynamically adjusted weather radar, *J. Hydrol.*, submitted, 2014b.
- Löwe, R., Vezzaro, L., Mikkelsen, P. S., Grum, M., and Madsen, H.: Investigating the use of stochastic forecasts for RTC of urban drainage systems - A Layout for Probabilistic Online Forecasting of Sewer Flows, *Environ. Model. Softw.*, submitted, 2014c.
- Melgaard, H.: Identification of physical models, Imm-phd-1994-4, Technical University of Denmark, [http://orbit.dtu.dk/fedora/objects/orbit:78530/datastreams/file\\\_\\\_2833556/content](http://orbit.dtu.dk/fedora/objects/orbit:78530/datastreams/file\_\_2833556/content), 1994.
- Möller, J. K.: Stochastic State Space Modelling of Nonlinear systems, Imm-phd-2010-246, Technical University of Denmark (DTU), 2010.
- Möller, J. K., Bergmann, K. R., Christiansen, L. E., and Madsen, H.: Development of a restricted state space stochastic differential equation model for bacterial growth in rich media, *J. Theor. Biol.*, 305, 78–87, doi:10.1016/j.jtbi.2012.04.015, 2012.
- 980 Øksendal, B.: Stochastic differential equations : An introduction with applications, Springer, 6th edn., 1998.
- Pleau, M., Colas, H., Lavalle, P., Pelletier, G., and Bonin, R.: Global optimal real-time control of the Quebec urban drainage system, *Environ. Model. Softw.*, 20, 401–413, doi:10.1016/j.envsoft.2004.02.009, 2005.
- Puig, V., Cembrano, G., Romera, J., Quevedo, J., Aznar, B., Ramón, G., and Cabot, J.: Predictive optimal control of sewer networks using CORAL tool: application to Riera Blanca catchment in Barcelona, *Water Sci. Technol.*, 60, 869–878, doi:10.2166/wst.2009.424, 2009.
- R-Core-Team: R: A Language and Environment for Statistical Computing, [www.r-project.org](http://www.r-project.org), 2013.
- 985 Radhakrishnan, K. and Hindmarsh, A. C.: Description and Use of LSODE, the Livermore Solver for Ordinary Differential Equations, Tech. rep., Lawrence Livermore National Laboratory, 1993.
- Raso, L., van de Giesen, N., Stive, P., Schwanenberg, D., and van Overloop, P. J.: Tree structure generation from ensemble forecasts for real time control, *Hydrol. Process.*, 27, 75–82, doi:10.1002/hyp.9473, <http://doi.wiley.com/10.1002/hyp.9473>, 2013.
- Reichert, P. and Schuwirth, N.: Linking statistical bias description to multiobjective model calibration, *Water Resour. Res.*, 48, W09543, doi:10.1029/2011WR011391, 2012.
- Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S.: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resour. Res.*, 46, W05521, doi:10.1029/2009WR008328, 2010.

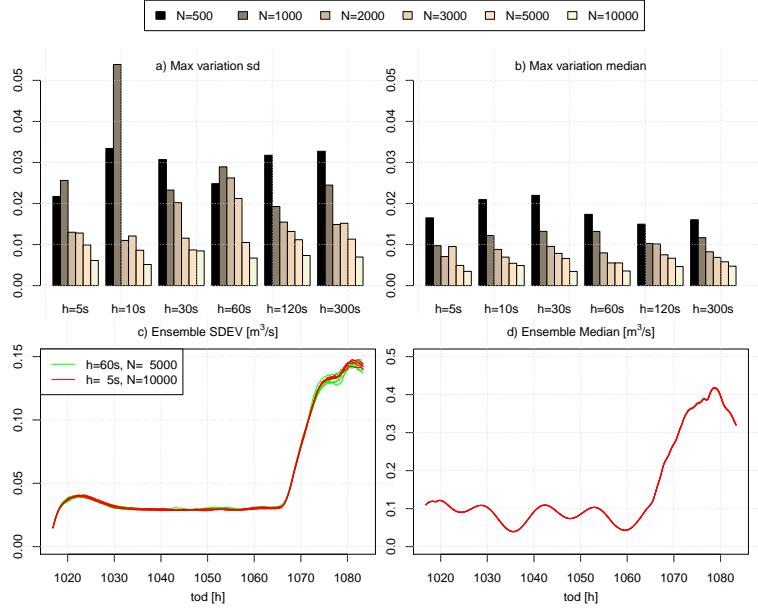
- Sadegh, P., Melgaard, H., Madsen, H., and Holst, J.: Optimal experiment design for identification of grey-box models, in: Proc. 1994 Am. Control Conf. - ACC '94, vol. 1, pp. 132–137, IEEE, doi:10.1109/ACC.1994.751709, <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=751709>, 1994.
- Stein, M.: Large Sample Properties of Simulations Using Latin Hypercube Sampling, *Technometrics*, 29, 143–151, doi:10.1080/00401706.1987.10488205, 1987.
- Sun, S. and Bertrand-Krajewski, J.-L.: Separately accounting for uncertainties in rainfall and runoff: Calibration of event-based conceptual hydrological models in small urban catchments using Bayesian method, *Water Resour. Res.*, 49, 5381–5394, doi:10.1002/wrcr.20444, 2013.
- Thordarson, F. O., Breinholt, A., Møller, J. K., Mikkelsen, P. S., Grum, M., and Madsen, H.: Uncertainty assessment of flow predictions in sewer systems using grey box models and skill score criterion, *Stoch. Environ. Res. Risk Assess.*, 26, 1151–1162, doi:10.1007/s00477-012-0563-3, 2012.
- Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S. W., and Srikanthan, S.: Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis, *Water Resour. Res.*, 45, doi:10.1029/2008WR006825, 2009.
- Tolson, B. A. and Shoemaker, C. A.: Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, *Water Resour. Res.*, 43, W01413, doi:10.1029/2005WR004723, 2007.
- Venables, W. N. and Ripley, B. D.: Modern Applied Statistics with S, vol. Fourth, Springer, New York, <http://www.stats.ox.ac.uk/pub/MASS4>, 2002.
- Vestergaard, M.: Nonlinear filtering in stochastic volatility models, Master, Technical University of Denmark (DTU), Lyngby, Denmark, 1998.
- Vezzaro, L. and Grum, M.: A generalized Dynamic Overflow Risk Assessment (DORA) for urban drainage Real Time Control, *J. Hydrol.*, submitted, 2014.
- Vezzaro, L., Löwe, R., Madsen, H., Grum, M., and Mikkelsen, P. S.: Investigating the use of stochastic forecast for RTC of urban drainage systems - full scale testing in the city of Copenhagen, Denmark, *Environ. Model. Software*2, Submitted, 2014.
- Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W., and Verstraten, J. M.: Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, *Water Resour. Res.*, 41, W01017, doi:10.1029/2004WR003059, 2005.
- Walter, E. and Pronzato, L.: Identification of parametric models from experimental data, Springer, 1997.
- Wolfs, V., Villazon, M. F., and Willems, P.: Development of a semi-automated model identification and calibration tool for conceptual modelling of sewer systems, *Water Sci. Technol.*, 68, 167–175, doi:10.2166/wst.2013.237, <http://www.ncbi.nlm.nih.gov/pubmed/23823553>, 2013.



**Fig. 1.** Flow measurements (blue) and rainfall measurements (black) for the considered data period in the BAL catchment (top) and the KLO catchment (bottom).

**Table 1.** Log-likelihood values from parameter estimation for the different models and catchment

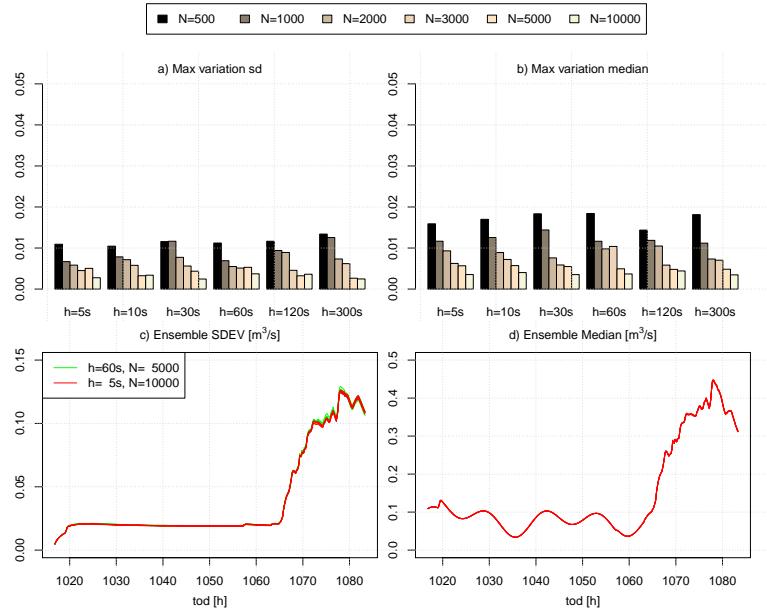
Catchment	Model1	Model2	Model3	Model1_P	Model2_P	Model3_P
BAL	-22105	-21929	-21360	-21882	-21700	-21235
DAM	-19464	-18236	-19596	-19162	-17905	-14339
KLO	1564	3590	3986	2054	4155	4050
LER	4867	5171	4248	5353	5618	4620



**Fig. 2.** Simulations of Model 1 in the Ballerup catchment with different increments  $h$  and different number of ensemble members  $N$ . a) maximal variation ensemble standard deviation during 5 simulation runs, b) maximal variation ensemble median during 5 simulation runs, c) maximal variation of ensemble standard deviation from any out of 5 simulation runs to any out of 5 reference runs with  $h = 5\text{s}$  and  $N = 10000$ , d) maximal variation of ensemble median from any out of 5 simulation runs to any out of 5 reference runs with  $h = 5\text{s}$  and  $N = 10000$ . Units are  $\text{m}^3/\text{s}$ .

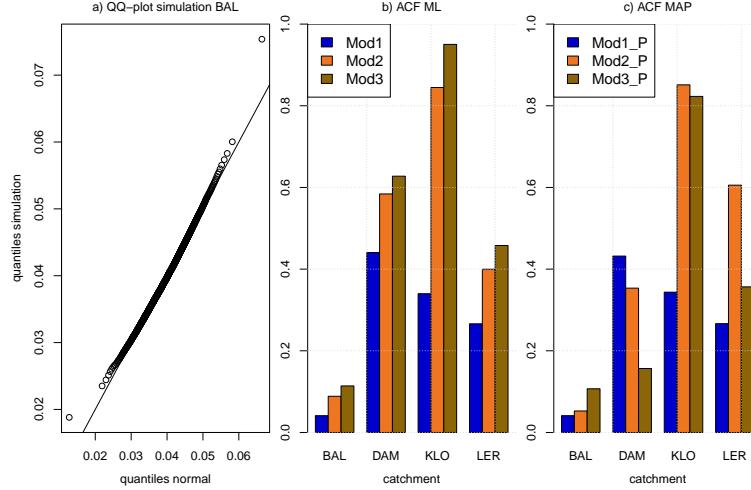
**Table 2.** Parameter estimates from Maximum Likelihood (ML) estimation for the four different catchments ("C") and models ("M") 1 to 3. The initial model states  $Z_{1,0}$  and  $Z_{2,0}$  are omitted.

C	M	A ha	$a_0$ $\text{m}^3/\text{s}$	$k$ s	$\ln(\sigma_{1,1})$	$\ln(\sigma_{1,2})$	$\ln(\sigma_{2,1})$	$\ln(\sigma_{2,2})$	$\ln(\sigma_e)$	$\lambda$	$l$
BAL	1	59	0.09	9904	-5.2		-6.3		-12.7		
	2	55	0.07	29958	-1.2	1.3	1.5	6.7	-13.2	0.98	3
	3	372	0.12	18971	1.7	-1.4	-4.8	6.3	-26.8	0.73	
DAM	1	17	0.25	4361	-5.0		-7.3		-17.1		
	2	61	0.23	7141	1.3	8.5	0.5	7.2	-40.0	0.97	5
	3	1	0.41	24288	-3.9	2.1	-4.4	6.9	-27.2	0.42	
KLO	1	2644	6.50	3772	-2.7		-4.7		-2.9		
	2	585	0.60	3750	2.2	7.4	-39.9	-37.6	-1.4	0.98	0
	3	787	1.06	18899	1.6	3.2	3.7	-5.4	-2.3	0.92	
LER	1	3946	0.00	11761	-3.1		-4.7		-1.8		
	2	219	0.00	5918	-4.9	0.2	1.9	4.9	-1.4	0.94	15
	3	230	0.10	22234	-3.5	-1.8	2.9	-5.0	-2.9	0.88	

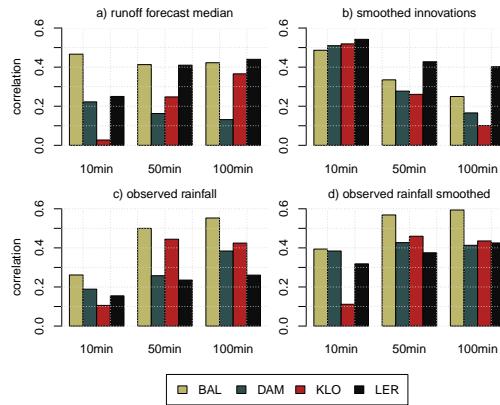


**Fig. 3.** Simulations of Model 2 in the Ballerup catchment with different increments  $h$  and different number of ensemble members  $N$ . a) maximal variation ensemble standard deviation during 5 simulation runs, b) maximal variation ensemble median during 5 simulation runs, c) maximal variation of ensemble standard deviation from any out of 5 simulation runs to any out of 5 reference runs with  $h = 5s$  and  $N = 10000$ , d) maximal variation of ensemble median from any out of 5 simulation runs to any out of 5 reference runs with  $h = 5s$  and  $N = 10000$ . Units are  $m^3/s$

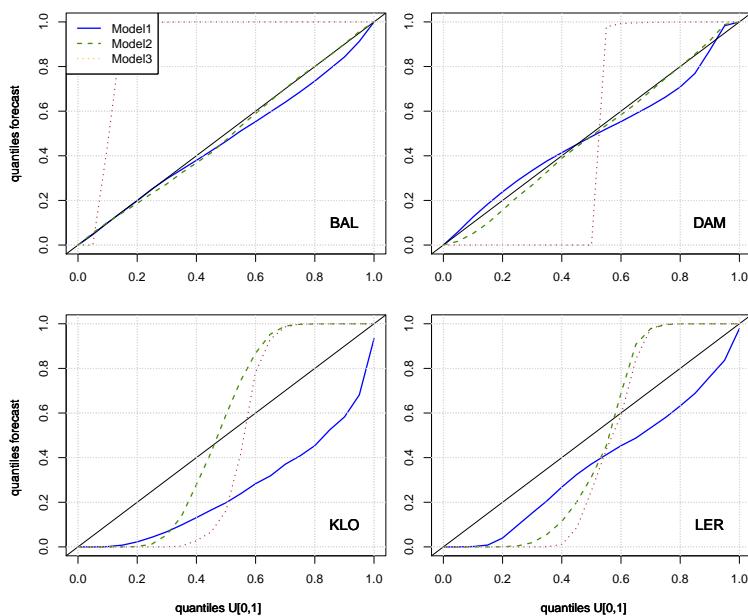
**Table 3.** Parameter estimates from Maximum a Posteriori (MAP) estimation for the four different catchments ("C") and models ("M") 1 to 3. The initial model states  $Z_{1,0}$  and  $Z_{2,0}$  are omitted. Row "P" shows the prior mean of the log-normal prior distribution with stdv=0.5 for the corresponding catchment



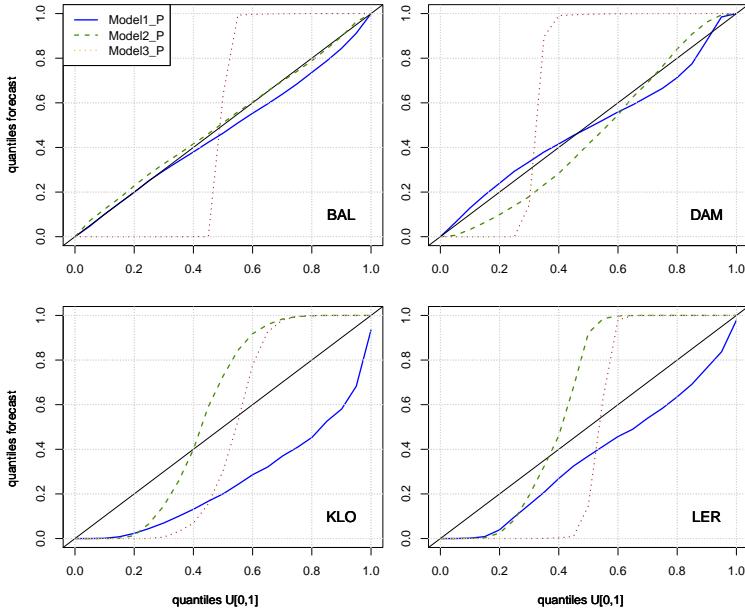
**Fig. 4.** Validation of assumptions for parameter inference: a) - QQ-plot for one-step ahead simulations of model 1 in the BAL catchment against a normal distribution, b) - maximal observed autocorrelation coefficient for the innovations obtained with ML parameter calibration, c) - maximal observed autocorrelation coefficient for the innovations obtained with MAP parameter calibration.



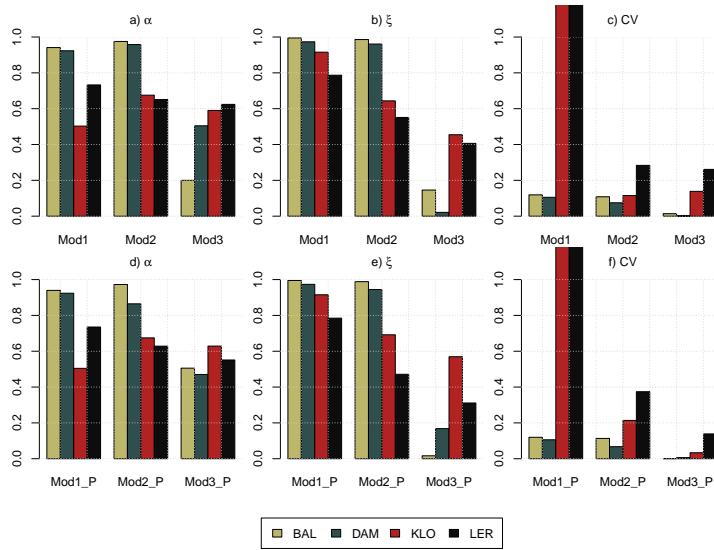
**Fig. 5.** Correlations between runoff forecast errors and different external variables on forecast horizons of 10, 50 and 100 minutes.



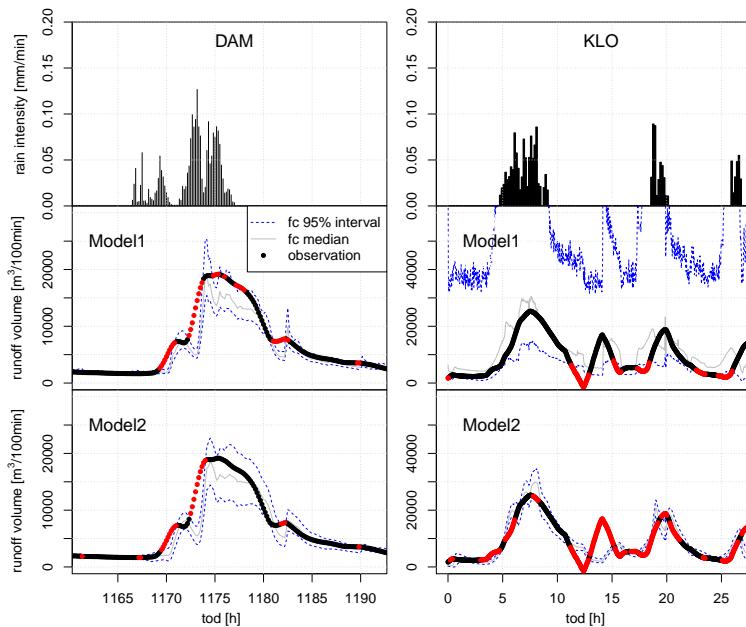
**Fig. 6.** Predictive QQ-Plot comparing predicted cumulated probabilities for the runoff volume observations against the  $U[0,1]$  distribution on a forecast horizon of 100min (models estimated using ML).



**Fig. 7.** Predictive QQ-Plot comparing predicted cumulated probabilities for the runoff volume observations against the  $U[0,1]$  distribution on a forecast horizon of 100min (models estimated using MAP).



**Fig. 8.** Summary of the predictive distribution on a 100min horizon as described in Section 3.3 for the models calibrated without (top) and with (bottom) prior information for the measurement error variance.



**Fig. 9.** 100min forecast of runoff volume from model 1 and model 2 for example events in the DAM and KLO catchments together with corresponding observations and rainfall measurements. Observations not included in the 95% prediction interval are marked red.