Assimilating a multitude of rainfall and runoff data using a stochastic state space modeling approach

Morten Grum, Poul Harremoës and Jens Jørgen Linde

PH-Consult, Ordruphøjvej 4, 2920 Charlottenlund, Denmark, mg@phc.dk

Abstract

Today both rainfall and runoff are measured in many different ways at the same time. Interpretation and use of this multitude of data together with our a priori engineering knowledge of the system and the way it works, requires methods for handling a multitude of data concerning a single process. In this paper presents an approach based on stochastic state space modeling to assimilate data rainfall runoff data consisting of several point rain gauges, microwave link line rain gauges, radar rainfall measurements and level and flow measurement in the resulting runoff stream. The system is described for both an urban sewer and a rural river catchment. A Kalman filter technique in which both observation and modeling errors are estimated from the available data (together with the interpretable model parameters) using a search routine and a maximum likelihood criterion. In evaluating the usefulness of the approach for planning and real time situations emphasis is placed on the engineers ability include his a priori knowledge in the model. Emphasis is also on the importance of the methodologies applicability to water and water quality engineering problems in general.

Introduction

Adequate characterization of the rainfall inputs is often critical to the success of urban hydrological and hydraulic engineering projects. Over the past decades a number of very different rainfall measurement techniques have been presented. The high spatial variability of rainfall has prompted an increasing interest in alternatives to the traditional point rainfall measurements. Radar measurements presented an interesting alternative but, in spite of the much better spatial coverage, radar has proven relatively to have poor depth coverage. New initiatives involve the use of microwave attenuation, which provides a measurement of the mean rainfall intensity between two points in the rain plane. Other approaches to rainfall measurement are continually being presented and will be presented in the future. A sub-catchment upstream of a flow-point or level measurement could also be considered to be a rain gauge assuming that one is able to give a fair description of the lumping and delay that takes place.

Characteristic of these approaches is that they provide very different types of information about the main driving force of urban runoff processes, namely rainfall. As depicted in Figure 1 point rainfall measurements have a poor spatial coverage but can have a relatively high depth precision. On the contrary, radar measurements generally have a low depth precision but a high special precision. Microwave attenuation provides an estimate of the mean rainfall intensity between two points and one may therefore expect it to lie somewhere in-between with respect to

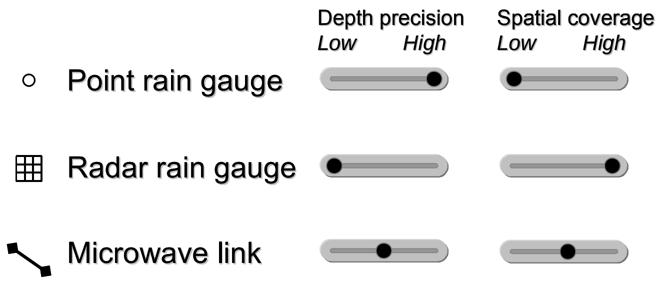
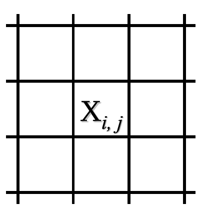


Figure 1. Rainfall measurements relative strengths. Point rainfall measurement with high depth precision, radar with good spatial coverage and microwave link measurements which may be somewhere inbetween.

both depth and spatial precision. Aspects of the use of microwave attenuation for rainfall measurement are discussed in Holt et al. (2000) and D'Amico (2002).

Used in isolation, each source of information has its drawbacks, falls short in some way or other and thus gives only part of the picture. This spread in the nature of the information suggests that there is a potential in using different measurement techniques simultaneously. In the presented approach this is achieved by constructing a model of the rainfall plane and the hydrological system. All measured data is considered to be different ways of looking at the same thing. That is, different observations of the current state of this combined rainfall and rainfall-runoff model. On the presence of any observations the state of the rainfall plane and the hydrological system is up-dated by weighting between what the model has predicted and the available observations. The weighting is done based on the uncertainties of the predictions and observations respectively.



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

where the coefficients $\alpha_{k,l}$ are given by the table

i | *j*	-1	0	+1
-1	0.05	0.05	0.05
0	0.05	0.60	0.05
+1	0.05	0.05	0.05

or $\alpha_{k,l} = f$ (wind speed, wind direction)

Figure 2. A very simple model which, one step ahead, predicts the rainfall intensity to be a weighting of its own current value with its immediate neighbors.

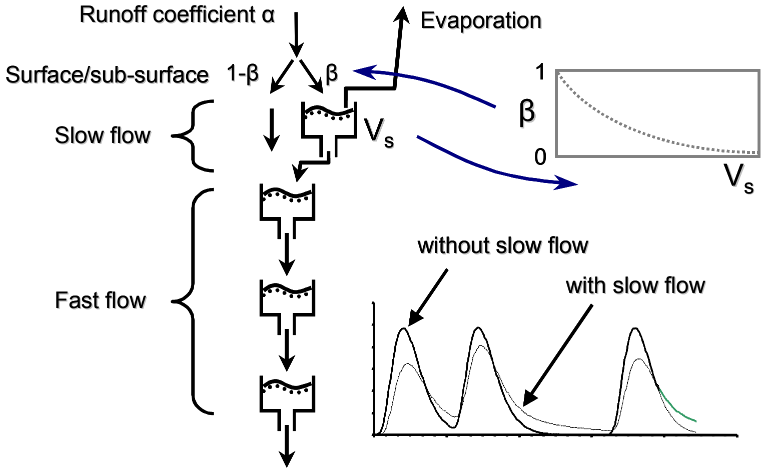


Figure 3. Schematic diagram of a simple sub-catchment model based on one slow and three fast flowing linear reservoirs.

Methodology

This section describes the methodology through the main steps and elements to be considered. The example used involves a very simple empirical rainfall model and a rainfall-runoff model consisting of linear reservoirs in series with a slow reservoir describing surface describing surface wetness and the slow runoff component.

Use a Model to Predict. A model of the rainfall plane can be constructed by dividing the rainfall plane into cells and saying (as in the equation below) that in the next time step the rainfall intensity in any given cell will be a weighting of the current intensity in that cell and the neighboring cells.

This somewhat naïve prediction would probably be greatly improved by incorporating the speed and direction of movement of the rain plane. These two parameters would then like the intensities have to become state variables of the rainfall model.

Similarly a simple rainfall-runoff model can be constructed by using linear reservoirs in series with a slow reservoir to describe surface wetness and a slow runoff component. The proportion of the water flowing into the slow reservoir is exponentially inversely proportional to the volume of water in the reservoir. At any given moment in time the state of the runoff system is defined by volume of water present in each of the reservoirs. This is illustrated in Figure 3.

In order to apply the Kalman filter one needs not only a prediction of the state variables themselves but also a prediction of their variances. The model therefore also serves to calculate the propagation of the states covariance matrix covariance. The model itself may be empirical or may to a smaller or larger extent be based on characterizations of the processes involved. The main criteria is that the model can be to following general discrete-time non-linear form:

$$\underline{X}_t = \underline{f}\left(\underline{X}_{t-1}, \underline{u}_{t-1}\right) + \underline{e}_{1,t}$$

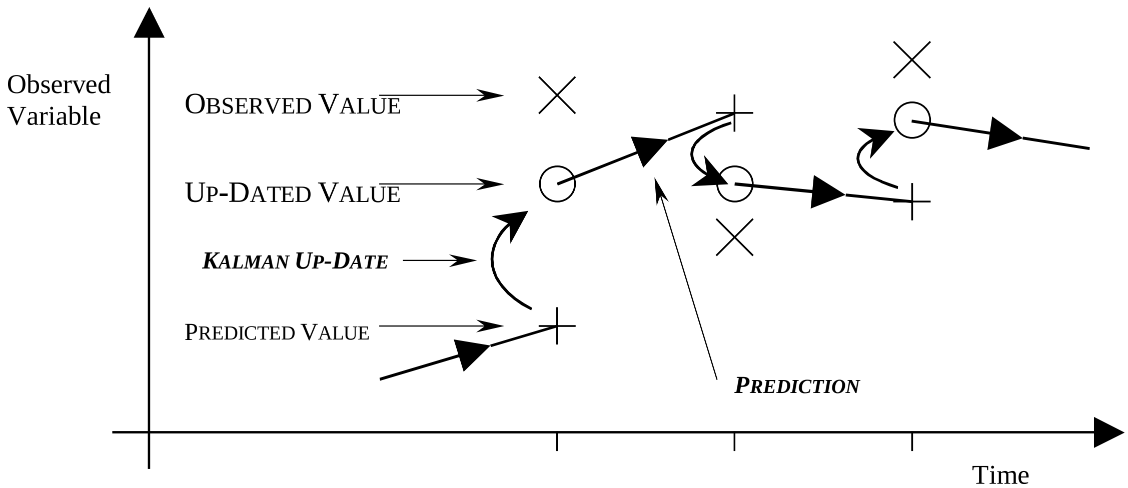


Figure 4. Sketch of the how the Kalman filter is used weight between the predicted and observed values. The illustration is for a one-dimensional case where the one state variable is observed directly.

where X is a vector of the state variables, $f(\ldots)$ is the set of functions transforming the state variables and known input variables at time t-1 to the state variables at time t. In some cases such as the rainfall and rainfall-runoff models of this paper there are no input variables. The model or system errors are assumed to be independent and normally distributed with a mean of zero.

Assimilate Observations using the Extended Kalman Filter. When encountering observations whilst modeling in time, one may do one of three things: Firstly, one may choose to have complete confidence in the model and assume that any deviation between the modeled and observed is a result of measurement or observation error. In this case one would note the deviation (which one minimizes on calibration) and proceed modeling from the modeled value. Secondly, one could choose to say "seeing is believing", rely on the observation and have no confidence in the model. In this case one would adjust or update the model state variables to fit with the observed values. The third option is to say "my model is good but there is probably some truth in those observations" and as a consequence choose to update the model states based on a weighting between "what we see" and "what we model". In this case one would calculate the uncertainty of the state variables, which would give a basis for updating the states to the most probable values (given also the uncertainty of the observations). The third option of weighing between "what we see" and "what we model" is illustrated in the Figure 4.

Note that also unobserved state variables will be updated. This happens in the Kalman filter due to their interdependency and their common variances, expressed as a covariance matrix. Very often state variables will not be observed directly. A water level measurement may be an indirect way of observing a water volume state variable. In such cases it is generally not a good idea to measly to convert all the observed values using some direct relationship to values for the corresponding state variable. This has to do with the underlying assumptions concerning the error distributions and is a discussion that falls outside the scope of this paper. A better solution is to incorporate the relationship between the state variable, for example a water volume, and the observed variable, for example a measured water level, into the observation equation.

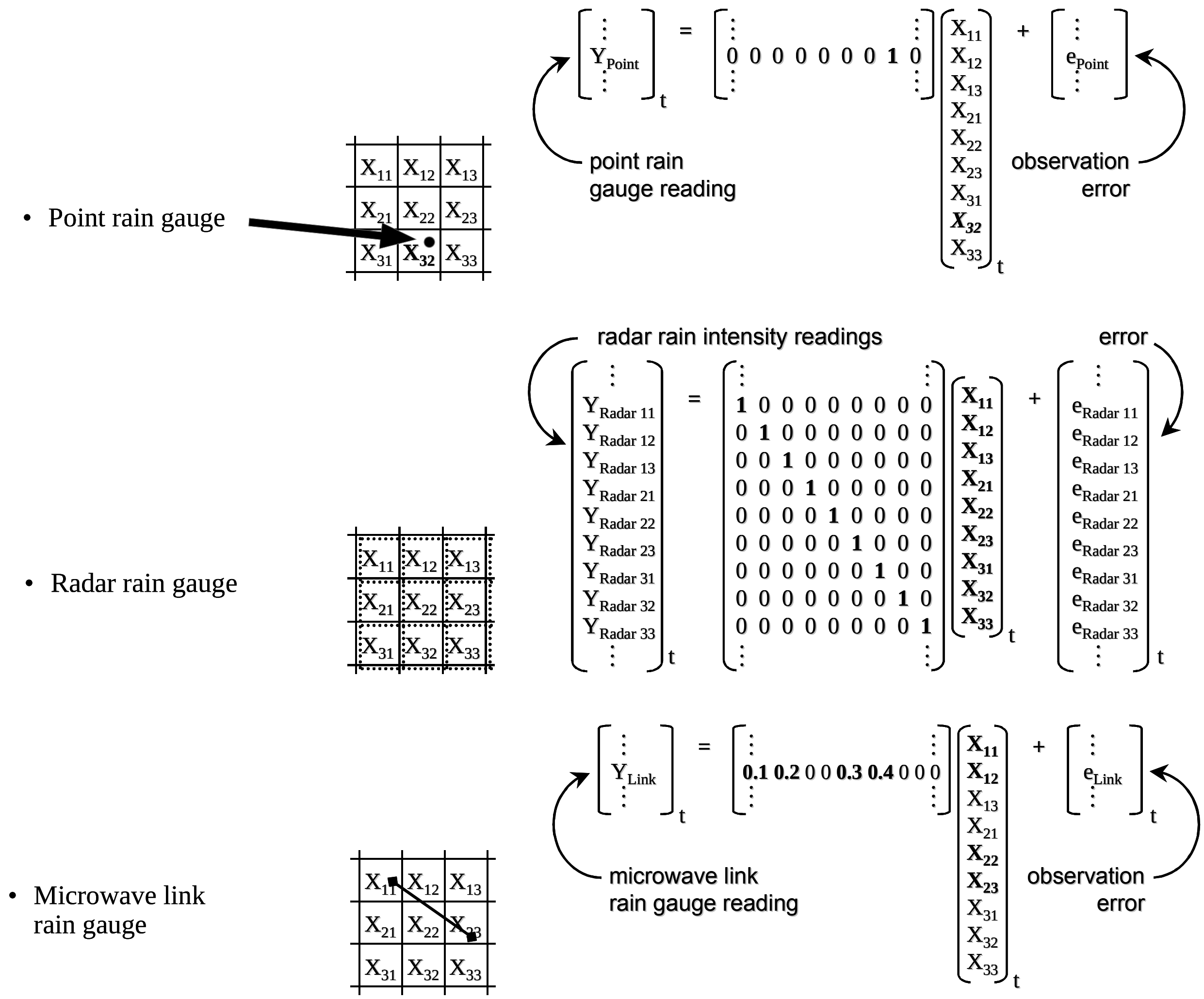


Figure 5. Illustration of how the various ways of looking at the same nine rainfall intensity state variables could be expressed in the observation equations.

The observation equation defines how one is observing the state of the system and has the following general form:

$$\underline{Y}_t = \underline{h}\left(\underline{X}_t\right) + \underline{e}_{2,t}$$

where Y is a vector of the observations, h(…) is the set of function defining how the state variables, X, are observed and the e_2 is a vector of the observation

With rainfall intensities of each cell in a rain plane as state variables, the observation equations for point rain gauges, radar and a microwave link measurement would all be linear. How the observation equations for these three types of rainfall measurements come about is illustrated in Figure 5. The point rain gauge is placed in cell (3,2) and its observation equation defines that it

- Flow

$$\underbrace{\begin{bmatrix}\vdots\\ Y_{Flow}\\ \vdots\end{bmatrix}_t}_{\text{flow reading or simulation}} = \begin{bmatrix}\vdots & & & & & & & & \vdots\\ \cdots & 0 & 0 & \frac{\mathbf{1}}{\mathbf{T_1}} & 0 & \frac{\mathbf{1}}{\mathbf{T_2}} & 0 & 0 & \frac{\mathbf{1}}{\mathbf{T_3}}\\ \vdots & & & & & & & & \vdots\end{bmatrix}\begin{bmatrix}\cdots\\ V_{11}\\ V_{12}\\ \mathbf{V_{13}}\\ V_{21}\\ \mathbf{V_{22}}\\ V_{31}\\ V_{32}\\ \mathbf{V_{33}}\end{bmatrix}_t + \underbrace{\begin{bmatrix}\vdots\\ e_{Flow}\\ \vdots\end{bmatrix}_t}_{\text{observation error}}$$

Figure 6. Illustration of the observation equation for a flow measurement at a point receiving flow from three sub-catchments (see text).

corresponds directly to intensity (3,2). Nevertheless, when updating, its value will have an influence on many of the other state variables too. This happens due to the interrelationship and co-variation between the state variables.

The observation equation for the flow out of the sub-catchments depends on the volume of water in the final reservoir. This relationship is illustrated in Figure 6 where the flow from three sub-catchments is measured at a single point.

Estimating the Variances. The success of this approach depends highly on ones knowledge of how much the state variable's uncertainty increases with time and the values of the observation variances. Normally, one will have very little process related basis for coming up with good estimates of these variables. Even the magnitude of observation errors often have little to do with for example the measurement errors related to sample analysis. Observation error encompasses much more, including in-homogeneity and sampling error. With in adequate estimates values for these model and observation error variances one is likely to get poor updates and in the worst case an instable modeling system.

However, based on a given set of parameter and variance values one can calculate the likelihood of actually obtaining the observed values. One can then choose the best set of parameters and variables to be that has the highest likelihood. Using an off-line optimization procedure on this maximum likelihood criterion, it is possible to estimate the model parameters and the required error variances. This approach is described in detail be Madsen et al. (1998).

Discussion

Assumptions. As stated earlier the error terms are assumed to be normally distributed with zero mean. In the case of rainfall intensities that rise from zero to high levels, back to zero and never become negative, this assumption will not at all be satisfied. One solution to this problem is to have transformed rainfall intensities as the state variables. A logarithmic transformation is not likely to be the right solution, as this would unrealistically amplify the state error term for the high intensities. Based on the general range and distribution of rainfall intensities and based on trial and error it is necessary to gain experience on what would be suitable for use with Kalman filtering. The same transformation should be applied probably be applied to the observation equation.

Optimization. As mentioned briefly above the model parameters including the variances of the error terms can be estimated on-line by using an optimization routine in combination with the maximum likelihood criteria. In practice there are a number of considerations to make that have not been discussed in this paper. These include aspects such as handling of missing observations on one or more of the signals, quantitative incorporation of *a priori* knowledge into the parameter estimation, robust estimators to minimize the influence of outliers, suitable parameter transformations, parameter statistics, critical parameter interchangeability and the suitability of different search routines.

Implementation. A number of research project have for many years presented case specific solutions that suggest that there is a great potential for the presented methodology for data assimilation within the field the water and water quality engineering (Grum, 2001, Carstensen, 1997, Qian, 1997, Carstensen, 1996, Jacobsen, 1996, Chiu, 1978, Beck, 1976). Many of these studies have combined physical, biological and/or chemical knowledge of the processes taking place with stochastic state space modeling. In the present study emphasis if on choosing solutions and software implementations specific to water engineering rather than to the present case study. The aim is to create obtain a modeling tool suitable for assimilating a multitude of real time water and water quality data using the described stochastic state space modeling approach. With a modulated software architecture the aim is to have an open and flexible tool where case specific needs can be incorporated at low cost.

Once the rainfall plane has been estimated, it can be used as input to more complex models, such as hydrodynamic models, which are unsuitable for stochastic state space modeling due to the large number of state variables.

Conclusion

This paper has presented a methodology for the assimilation of a multitude of rainfall and runoff data using a stochastic state space modeling approach. The approach employs known methods of combining Kalman filter with a maximum likelihood criteria and thereby updating the state of a system model whenever any observations were available.

Future project research and development will focus on:

- Gaining practical experience with the presented approach using observed rainfall-runoff data from both urban and rural catchments.
- Isolating generally applicable aspects and solutions from the case specific and using this knowledge in optimal tool design.
- Analyzing the relative importance of each of the different rainfall measurement sources and finding optimal combinations of the different types of rain gauges.
- To study the use of the approach for planning purposes and for real time tasks such as early warning and real time control.

References

Beck, M. B. and Young, P. C. (1976). Systematic identification of DO-BOD model structure, *Journal of Environmental Engineering Division*, ASCE, 102, 909-927, 1776.

Carstensen, J., Harremoës, P. and Strube, R. (1996). Software sensors based on the grey-box modelling approach. *Wat Sci. Tech.*, 33(1), pp 117-126

D'Amico, M., Pinotti, M. and Capsoni, C (2002). Performance of the Microwave radio links evaluated through radar simulations. In preparation/submitted.

Grum, M. (2001). Urban Runoff Pollution: Modelling and Uncertainty in Return Period Analysis. PhD. Thesis, Wageningen University.

Holt, A R, Goddard, J W F, Upton, G J G, Willis, M J, Rahimi, A R, Baxter, P D and Collier, C G (2000). RADIOCOMMUNICATION - Measurement of rainfall by dual-wavelength microwave attenuation. *Journal title Electronics Letters*, Vol.36 Issue.25, pp. 2099-2100.

Jacobsen, J. L., Madsen, H., Harremoes, P. (1997) A stochastic model for two-station hydraulics exhibiting transient impact. *Wat. Sci. Tech.*, 36(5), pp. 19-26.

Madsen, H. and Holst, J. (1998). Modelling Non-Linear and Non-Stationary Time Series. Department of Mathematical Modelling, Technical University of Denmark, Building 321/305, DK-2800 Kgs. Lyngby, Denmark. www.imm.dtu.dk

Qian, Song S. (1997). Water quality model structure identification using dynamic linear modeling: river cam case study revisited. *Wat. Sci. Tech.*, 36(5), pp. 27-35.

Aknowledgements

The presented work is carried out under the EU sponsored MANTISSA project. Further details and progress can be found at: http://prswww.essex.ac.uk/mantissa/.