Automated ARIMA-Type Model Selection

Comparisons of Data-Driven Flow Forecast Models

Ari Jóhannesson

DTU Environment
Department of Environmental Engineering

Outline

- Introduction
  - Backround
  - Aim of This Work
  - Research questions
- Theory
  - Current Models in Hydrology
  - Auto-regressive intergrated moving-average (ARIMA) type models
- Case Study
- Exploratory Data Analysis

Outline

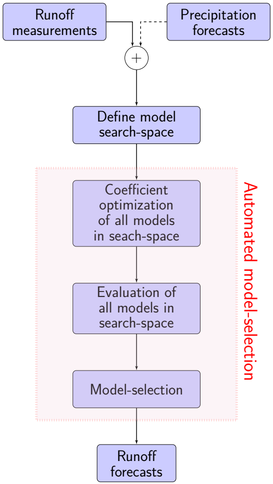
- Methods
  - Data Cleaning and Preparation
  - Objective Function Criteria
  - Coefficient Estimation
  - Meta-Optimization
  - Evaluation
- Results
  - Comparisons of Optimization Methods in Coefficient Estimation
  - Comparisons of Nelder-Mead optimized Hyper-Models
  - Selection of Best Performing Model for Operational Purposes
- Discussion
- Outlook

Background

- Modern sewage systems are centralized around WWTP
- Older cities often have combined sewage (CS)
  - Combination of sewage and surface water
  - In heavy rain events, CS can increase and lead to combined sewage overflow CSO
  - CSO is disposal of untreated hazardous wastewater into the environment
- Dynamic control of the combined sewage can reduce CSO

Aim of This Work

- Automated model selection of ARIMA type models
- Runoff measurements for two locations in Copenhagen, and 'perfect' precipitation forecasts
- Meta-optimization
- Two evaluation measures will be used for model selection



Research questions

Research Questions

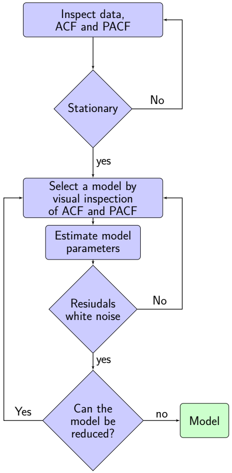
1. Can competent ARIMA type models be selected in an automated and efficient manner?
2. How should the parameter search-space be constrained such that parameter selection can be performed in a computationally efficient manner, while still selecting parameters that adequately capture the complex behavior of the system?
3. Do local and global-searches in the coefficient estimation produce significantly different models?
4. Do different objective function criteria (i.e., calibrating models to single/multi-step forecasts) generate substantially different models?
5. Will proposed error metrics result in analogous models.
6. Does precipitation as an external regressor improve forecasting?
7. How do ARIMA models compare to the current models in use?

Current Models in Hydrology

- Economical and environmental benefits of hydological forecasting
- Conceptual white-box models
  - Rely on physical structure of the system
  - Complex, computationally heavy, slow
  - Calibration has to be done manually
- Data-driven black-box models
  - Use statistics in a data-diven way
  - Have been gaining attention in recent years
  - f.x. SVM, ANN, ARIMA,

ARIMA (Auto-regressive intergrated moving-average)

- ARIMA models introduced in 1950s
- Revisited in 1970s by Box and Jenkins
  - Three step iterative analysis
    - Identification
    - Estimation
    - Verification
- Composed of three components: AR(p), I(d), and MA(q)



AR(p) term

- Classical regression of degree p
- Past observations multiplied with coefficients to form a prediction

$$y_{t+1} = \mu + \phi_1 y_t + \phi_2 y_{t-1} + \cdots + \phi_p y_{t+1-p} + \epsilon_{t+1} \tag{1}$$

$$\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i \tag{2}$$

$$\phi(B) y_{t+1} = \epsilon_{t+1} \tag{3}$$

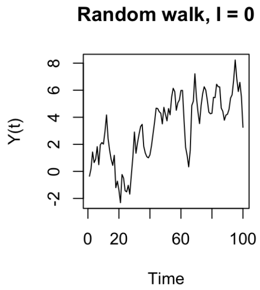
I term

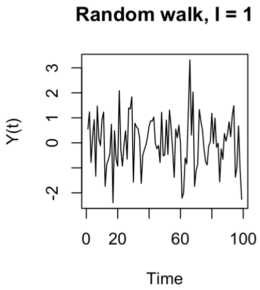
- Ensures stationarity
- Linear trend removed with first degree

$$\nabla y_{t+1} = y_{t+1} - y_t \tag{4}$$

- Quadratic trend removed with second degree

$$\phi(B)\nabla^2 y_{t+1} = \epsilon_{t+1} \tag{5}$$





MA(q) term

- Past errors multiplied by coefficients to predict future observations

$$y_{t+1} = \theta_1 \epsilon_t + \theta_2 \epsilon_{t-1} + \ldots + \theta_p \epsilon_{t+1-p} + \epsilon_{t+1} \tag{6}$$

$$\theta(B) = 1 + \sum_{i=1}^{q} \theta_i B^i \tag{7}$$

$$y_{t+1} = \theta(B) \epsilon_{t+1} \tag{8}$$

Assembly of ARIMA models

- Non-differentiated

$$y_{t+1} = \mu + \sum_{i=1}^{p} \phi_i y_{t+1-i} + \epsilon_{t+1} + \sum_{i=1}^{q} \theta_i \epsilon_{t+1-i} \tag{9}$$

- Differentiated of degree d

$$\phi(B)\nabla^d y_{t+1} = \alpha + \theta(B)\epsilon_{t+1} \tag{10}$$

External regressor

- Adding external regressors x, can be done
- ARIMAX$(p, d, q, \rho_{\text{lag}}, \rho_n)$

$$y_{t+1} = \underbrace{\mu + \sum_{i=1}^{p} \phi_i y_{t+1-i} + \epsilon_{t+1} + \sum_{i=1}^{q} \theta_i \epsilon_{t+1-i}}_{\text{ARIMA}} + \underbrace{\sum_{i=1}^{\rho_n} \lambda_i x_{t+1-i-\rho_{lag}}}_{\text{External regressor}} \tag{11}$$

Coefficient estimation

- Numerical optimization algorithms minimize/maximize objective function
- Sum-of-squares

$$S(\phi, \theta) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{12}$$

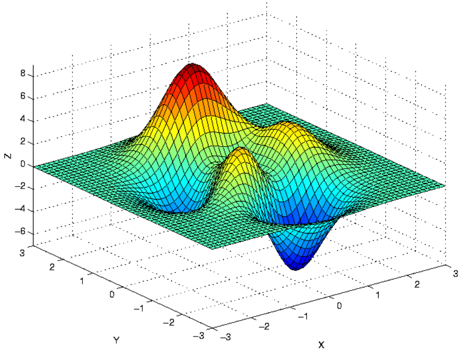
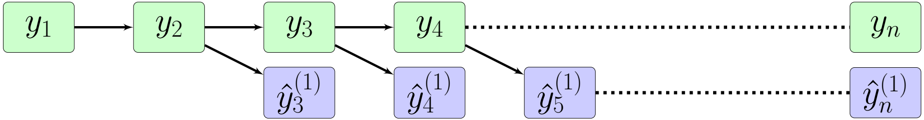


 Figure from https://medium.com/@tomdeore/deep-learning-in-gradient-decent-style-part-1-ed747e7cc2a3

Single-Step Forecasting

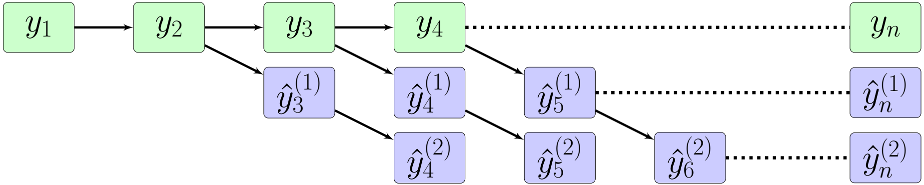
$$
\begin{aligned}
\hat{y}_{t+1} &= \mu + \phi_1 y_t + \phi_2 y_{t-1} + \epsilon_{t+1} + \theta_1 \epsilon_t + \theta_2 \epsilon_{t-1} \\
\hat{y}_{t+1} &= \mu + \phi_1 y_t + \phi_2 y_{t-1} + 0 + \theta_1 \epsilon_t + \theta_2 \epsilon_{t-1}
\end{aligned} \tag{13}
$$



Multi-Step Forecasting

$$\hat{y}^{(2)}_{t+2|t+1} = \mu + \phi_1 \hat{y}_{t+1} + \phi_2 y_t + \epsilon_{t+2} + 0 + \theta_2 \epsilon_t \tag{14}$$

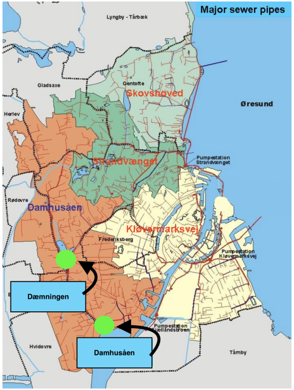
$$\begin{aligned} \hat{y}^{(3)}_{t+3|t+2 \text{ and } t+1} &= \mu + \phi_1 \hat{y}_{2,t+2} + \phi_2 \hat{y}_{1,t+1} + \epsilon_{t+3} + 0 + 0 \\ &= \mu + \phi_1 \hat{y}_{2,t+2} + \phi_2 \hat{y}_{1,t+1} + 0 + 0 + 0 \end{aligned} \tag{15}$$



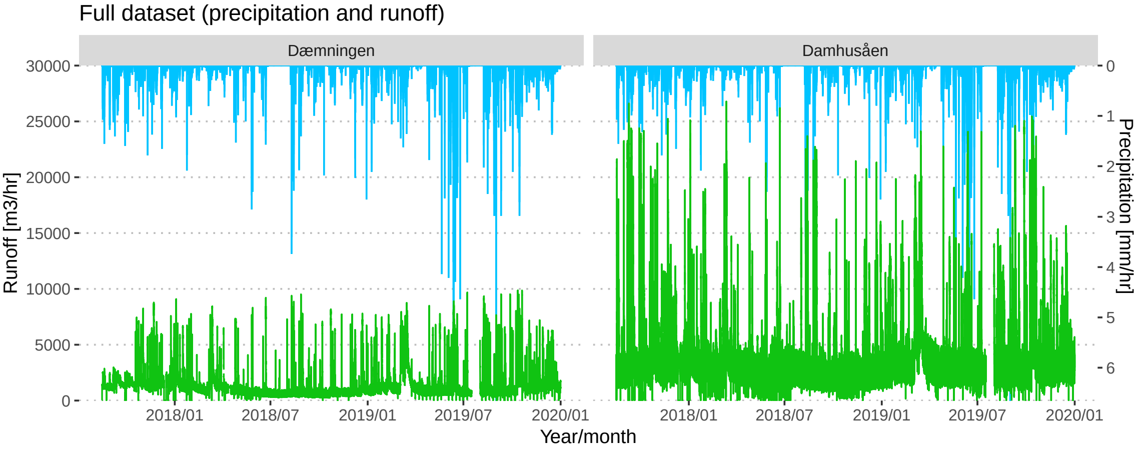
Catchments

- Two flow gauge data in Copenhagen
  - Damhusåen WWTP
  - Dæmningen
- 'Perfect' precipitation radar forecast

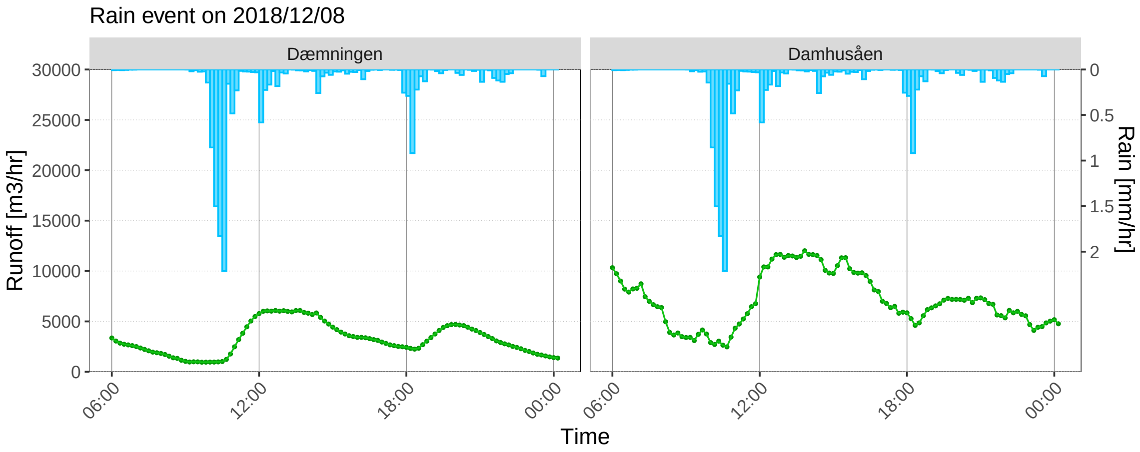
Type	Range
Training	August 16th 2017 - December 31st 2017
Validating	January1st 2018 - December 31st 2018



Time Series Vizualisation

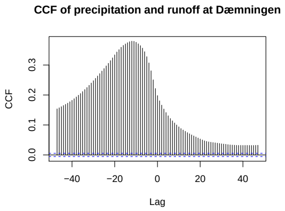


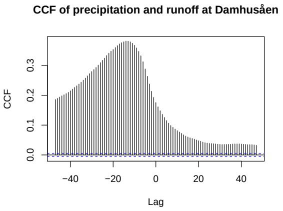
Time Series Vizualisation



Cross correlations

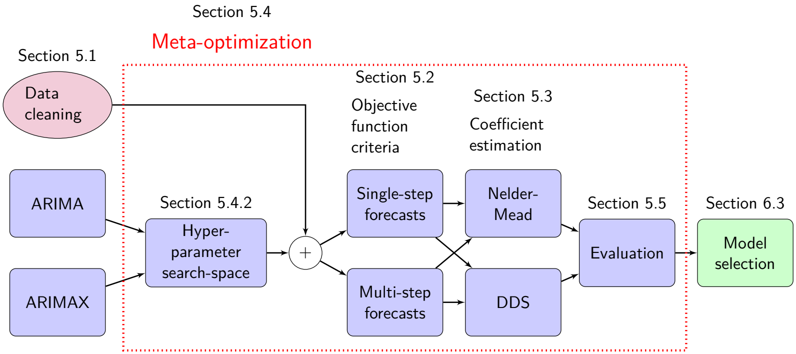
- Cross correlations between catchments
  - Highest for lag-3
- Cross correlations between runoff and precipitation
  - Highest for lag-12 for Dæmningen
  - Highest for lag-13-14 for Damhusåen





Methods

- **Parameters:** Constants that control the model structure i.e. p, d, q, ρ_{lag}, ρ_n. Model hyper-parameters are used to tune the parameters.
- **Coefficients:** Variables that are adjusted to minimize the objective function to generate adequate predictions. These predictions are generated by multiplying the coefficients with observations i.e. ϕ, θ, λ.

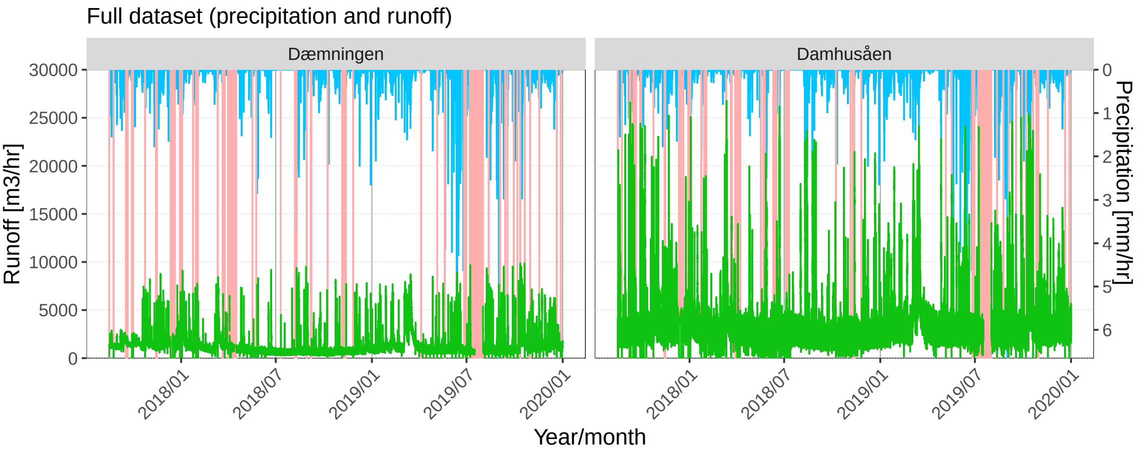


Data Cleaning and Preparation

- Normalization
- Different ranges of data sets
- Different frequency (temporal resolution) of data sets
- Daylight saving time shifts
- Flatlines in sensor data
- Missing data

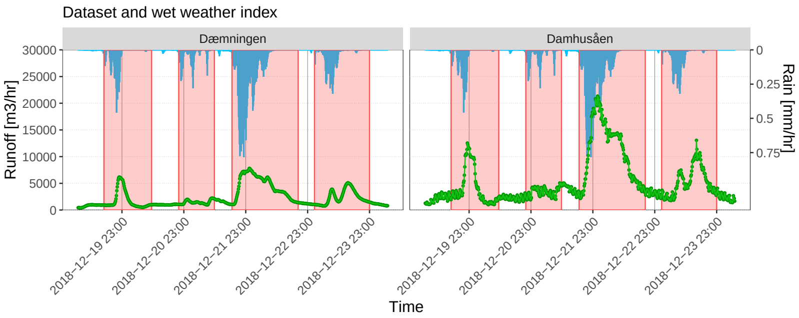
Flatlines and missing data

- Rolling window of size 5 (corresponding to 50 minutes)
- If all values are same, flatline flagged



Wet-weather index

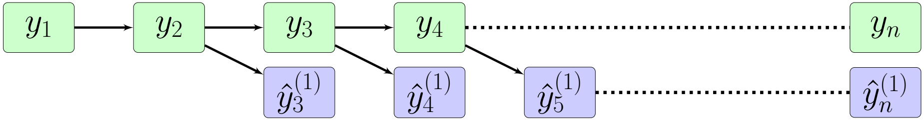
- Wet-weather flagged when precipitation $\geq 0.665 mm/hr$ (0.1 after normalization).
- Two rain events are merged and considered as one if:
  - The time difference between the start of two rain events is less than 4 hours.
  - The time difference between the stop and start of two rain events is less than 2 hour.
- 12 hour tail added to wet-weather event



Single-Step Objective Function Criterion

- Minimize the objective function of single-step predictions.

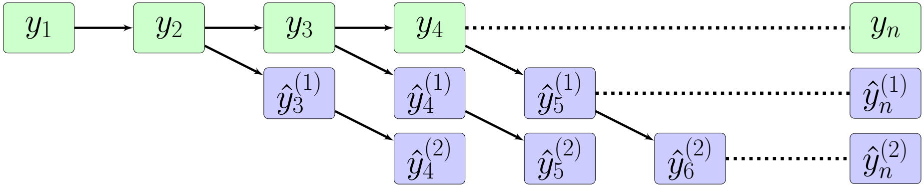
$$S(\phi, \theta) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{16}$$



Multi-Step Objective Function Criterion (perfomance estimation)

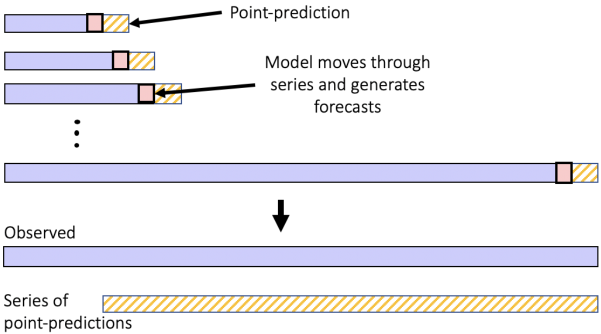
- Hydrological models need to forecast for various forecasting horizons
- Forcing models to perfom well for multi-step forcasting

$$SC_i = \frac{1}{\sum_{j=1}^{k}(k-j+1)} \left(\sum_{j=1}^{k} (k-j+1) \cdot SC_{i,j} \right) \tag{17}$$



Multi-Step Objective Function Criterion (efficient implementation)

- R packages such as STATS::ARIMA and STATS::PREDICT are primarily used for generating forecasts for a single starting point at the end of time series



Multi-Step Objective Function Criterion (efficient implementation)

$$\overbrace{\begin{pmatrix} 1 & y_r & y_{r-1} & \cdots & y_{r+1-p} & \epsilon_r & \epsilon_{r-1} & \cdots & \epsilon_{r+1-q} \\ 1 & y_{r+1} & y_r & \cdots & y_{r+2-p} & \epsilon_{r+1} & \epsilon_r & \cdots & \epsilon_{r+2-q} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & y_{n-1} & y_{n-2} & \cdots & y_{n-p} & \epsilon_{n-1} & \epsilon_{n-2} & \cdots & \epsilon_{n-q} \end{pmatrix}}^{\Psi} \delta = \overbrace{\begin{pmatrix} \hat{y}^{(1)}_{r+1} \\ \hat{y}^{(1)}_{r+2} \\ \vdots \\ \hat{y}^{(1)}_{n} \end{pmatrix}}^{\hat{y}^{(1)}} \qquad (18)$$

Multi-Step Objective Function Criterion (efficient implementation)

$$
\begin{array}{c}
\begin{array}{ccccccccc}
 & \overbrace{\qquad}^{\hat{y}_*^{(1)}} & & & & \overbrace{\qquad\qquad}^{\hat{\epsilon}_*^{(1)}} & & &
\end{array}
\\
\begin{pmatrix}
1 & \hat{y}_{r+1}^{(1)} & y_r & \dots & y_{r+2-p} & \hat{\epsilon}_{r+1}^{(1)} = 0 & \epsilon_r & \dots & \epsilon_{r+2-q} \\
1 & \hat{y}_{r+2}^{(1)} & y_{r+1} & \dots & y_{r+3-p} & \hat{\epsilon}_{r+2}^{(1)} = 0 & \epsilon_{r+1} & \dots & \epsilon_{r+3-q} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & \hat{y}_{n-1}^{(1)} & y_{n-2} & \dots & y_{n-p} & \hat{\epsilon}_{n-1}^{(1)} = 0 & \epsilon_{n-2} & \dots & \epsilon_{n-q}
\end{pmatrix}
\end{array}
\delta =
\begin{array}{c}
\overbrace{\qquad}^{\hat{y}^{(2)}} \\
\begin{pmatrix}
\hat{y}_{r+2}^{(2)} \\
\hat{y}_{r+3}^{(2)} \\
\vdots \\
\hat{y}_{n}^{(2)}
\end{pmatrix}
\end{array}
\qquad (19)
$$

Multi-Step Objective Function Criterion (efficient implementation)

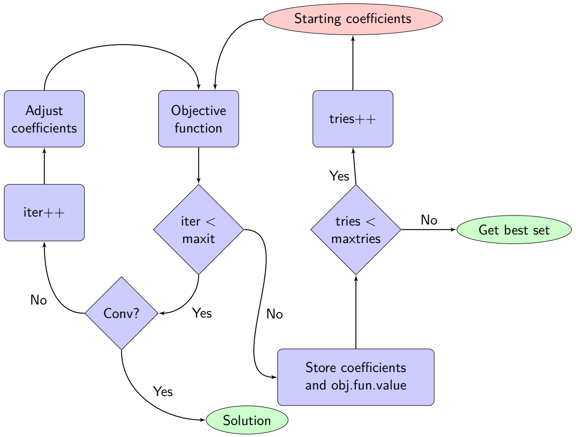
$$
\begin{array}{cc}
\begin{array}{ccccccccc}
 & \overbrace{\qquad}^{\hat{y}_*^{(2)}} & \overbrace{\qquad}^{\hat{y}_*^{(1)}} & & & \overbrace{\quad}^{\hat{\epsilon}_*^{(2)}} & \overbrace{\quad}^{\hat{\epsilon}_*^{(1)}} & &
\end{array} & \overbrace{\qquad}^{\hat{y}_*^{(3)}} \\
\begin{pmatrix}
1 & \hat{y}_{r+2}^{(2)} & \hat{y}_{r+1}^{(2)} & \dots & y_{r+3-p} & 0 & 0 & \dots & \epsilon_{r+3-q} \\
1 & \hat{y}_{r+3}^{(2)} & \hat{y}_{r+2}^{(2)} & \dots & y_{r+4-p} & 0 & 0 & \dots & \epsilon_{r+4-q} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & \hat{y}_{n-1}^{(2)} & \hat{y}_{n-2}^{(2)} & \dots & y_{n-p} & 0 & 0 & \dots & \epsilon_{n-q}
\end{pmatrix} \delta = &
\begin{pmatrix}
\hat{y}_{r+3}^{(2)} \\
\hat{y}_{r+3}^{(2)} \\
\vdots \\
\hat{y}_{n}^{(2)}
\end{pmatrix}
\end{array}
\tag{20}
$$

Multi-Step Objective Function Criterion (efficient implementation)

$$\Omega = \begin{pmatrix} \hat{y}_{r+1}^{(1)} & \hat{y}_{r+2}^{(1)} & \hat{y}_{r+3}^{(1)} & \hat{y}_{r+4}^{(1)} & \hat{y}_{r+5}^{(1)} & \hat{y}_{r+6}^{(1)} & \hat{y}_{r+7}^{(1)} & \cdots & \hat{y}_{n}^{(1)} \\ & \hat{y}_{r+2}^{(2)} & \hat{y}_{r+3}^{(2)} & \hat{y}_{r+4}^{(2)} & \hat{y}_{r+5}^{(2)} & \hat{y}_{r+6}^{(2)} & \hat{y}_{r+7}^{(2)} & \cdots & \hat{y}_{n}^{(2)} \\ & & \hat{y}_{r+3}^{(3)} & \hat{y}_{r+4}^{(3)} & \hat{y}_{r+5}^{(3)} & \hat{y}_{r+6}^{(3)} & \hat{y}_{r+7}^{(3)} & \cdots & \hat{y}_{n}^{(3)} \\ & & & \hat{y}_{r+4}^{(4)} & \hat{y}_{r+5}^{(4)} & \hat{y}_{r+6}^{(4)} & \hat{y}_{r+7}^{(4)} & \cdots & \hat{y}_{n}^{(4)} \\ & & & \ddots & \ddots & \ddots & \ddots & \cdots & \hat{y}_{n}^{(k)} \end{pmatrix} \tag{21}$$

Coefficient Estimation

- Nelder-Mead local-search
  - $1,000$ iterations
  - If no convergence, run again (max 5 times)
- DDS global-search
  - $2,500$ objective function evaluations



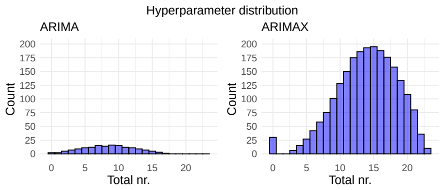
Hyper-Model-Types

	Type	Objective function criteria	Coefficient optimization
1	ARIMA	Single-step	Nelder-Mead
2	ARIMA	Single-step	DDS
3	ARIMA	Multi-step	Nelder-Mead
4	ARIMA	Multi-step	DDS
5	ARIMAX	Single-step	Nelder-Mead
6	ARIMAX	Single-step	DDS
7	ARIMAX	Multi-step	Nelder-Mead
8	ARIMAX	Multi-step	DDS

Selection of Hyper-Model Parameters

- Hyper-model parameter search-space must be kept limited
- Calibrations are parallelized

Parameter	ARIMA	ARIMAX
p	$0, 1, 2, 3, 4, 5, 6, 7, 8$	$0, 1, 2, 3, 4, 5, 6, 7, 8$
d	0, 1	0, 1
q	$0, 1, 2, 3, 4, 5, 6, 7, 8$	$0, 1, 2, 3, 4, 5, 6, 7, 8$
ρ_{nr}	0	$2, 4, 6, 8, 10$
ρ_{lag}	0	$5, 10, 15$
Nr. models	162	2.430



Evaluation

Hydrological models will be evaluated with two distinct error measures:

- Persistence Index (PI) skill-score.

$$\mathsf{PI} = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2}{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_{i-1}\right)^2} \tag{22}$$

- Accuracy in predicting a simplified version of ATS activation.

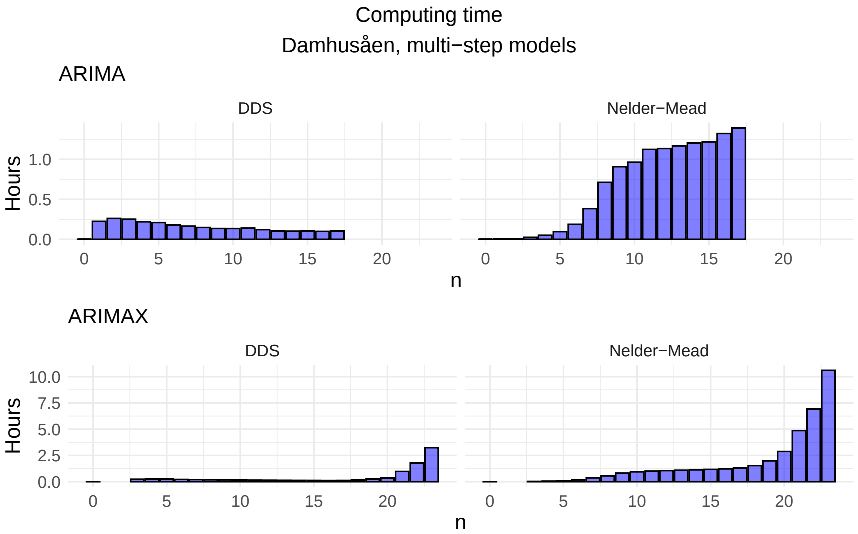
$$\mathsf{Accuracy}_c = \frac{\mathrm{TP}_c}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{23}$$

$$\mathsf{Accuracy}_{c+e} = \frac{\mathrm{TP}_c + \mathrm{TP}_e}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{24}$$

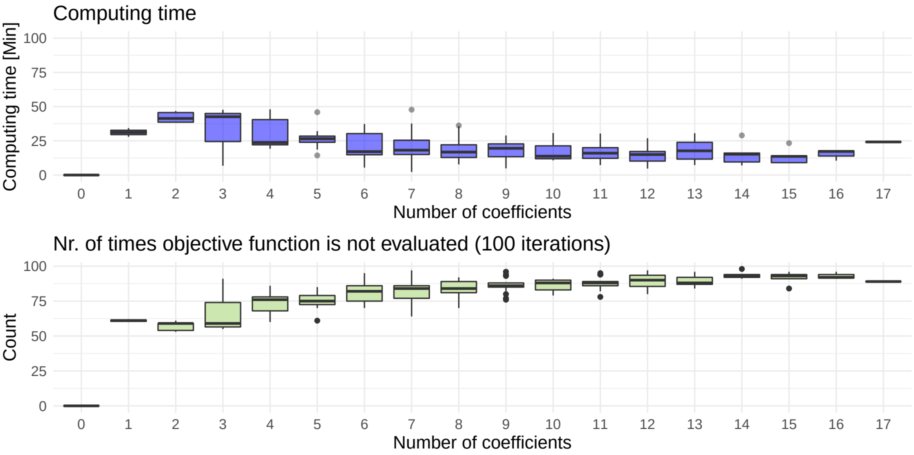
Performance of ATS Activation

- Correct (TP_c): If WET activation was predicted within a 75 min interval (from 60 min before measured flow exceeded the threshold to 15 min after)
- Early (TP_e): If WET activation was predicted earlier than 60 min before the measured flow exceeded the threshold
- False alarm (FP): WET event forecasted, but measured runoff did not exceed the threshold.
- Missed (FN): If WET activation was predicted more than 15 min after the measured flow exceeded the threshold or if no WET was predicted (False Negative)
- Down: no data are available for the specific time interval where the WET event took place (i.e., it is not possible to evaluate the performance of the forecast)

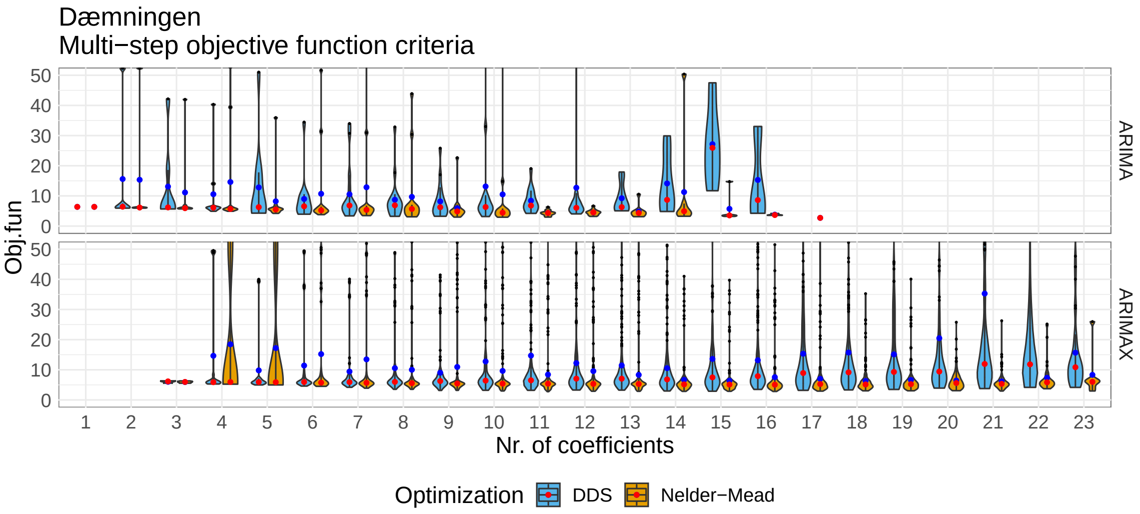
Comparisons of Optimization Methods in Coefficient Estimation



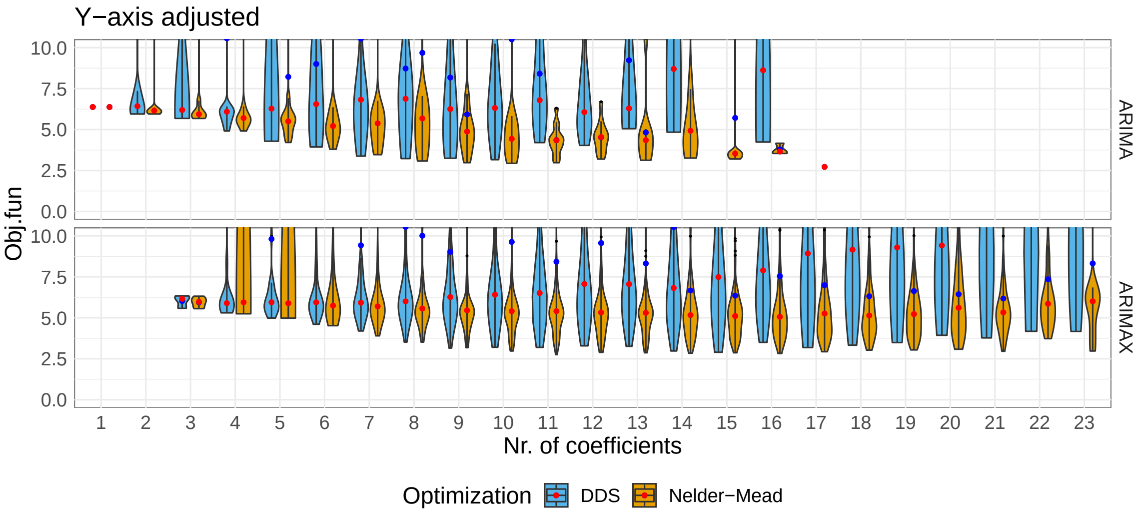
Comparisons of Optimization Methods in Coefficient Estimation



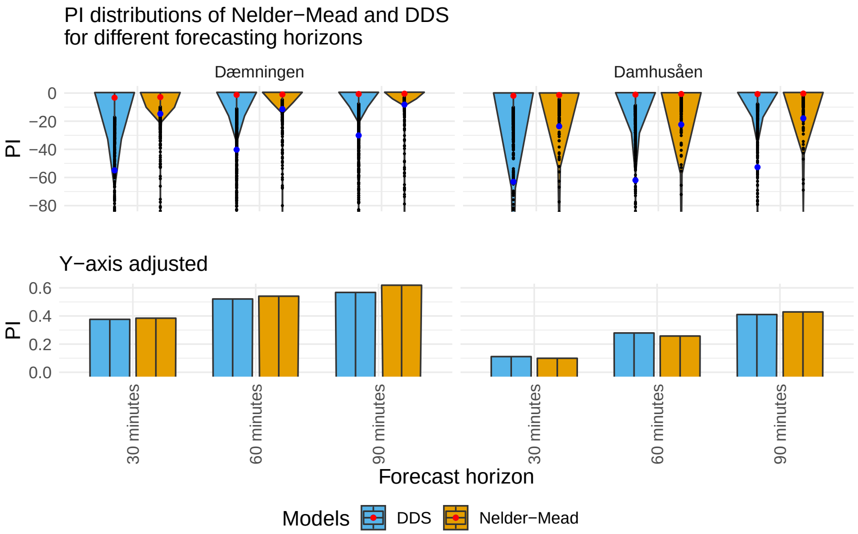
Minimized Objective Function



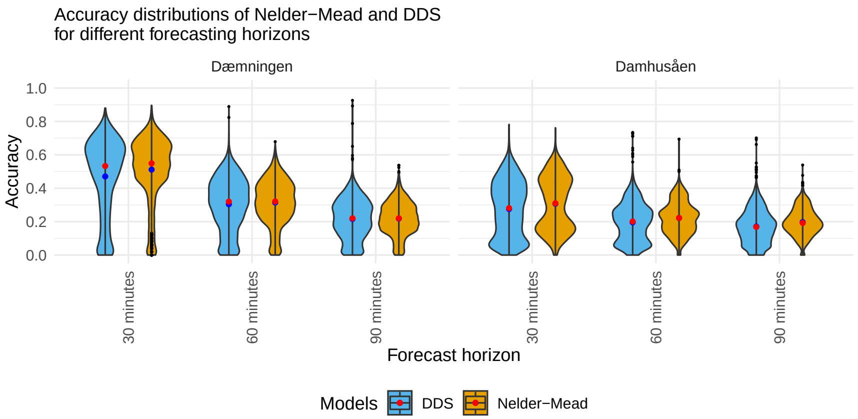
Minimized Objective Function



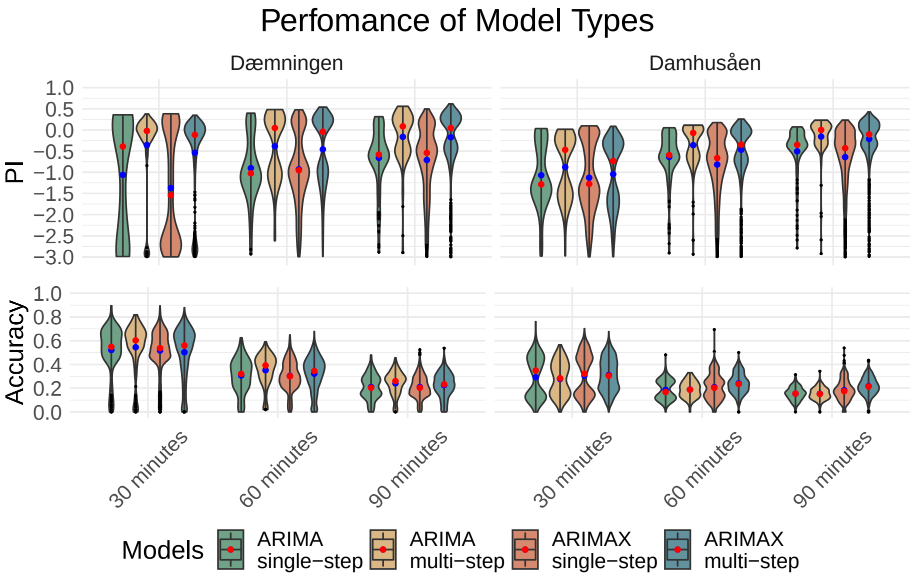
Persistence Index



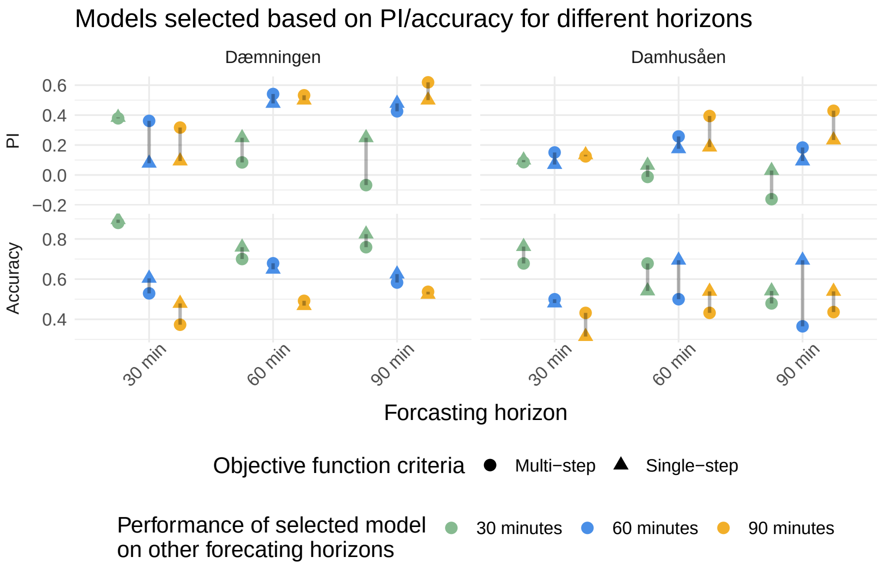
Accuracy of ATS activation



Perfomance of Model Types



Best models compared



On Evalaution Measures (sorting based on PI)

Dæmningen													
Model					Persistence Index			Accuracy					
	ofc	order	ρ_{nr}	ρ_{lag}	30 min	60 min	90 min	correct 30 min	+early 30 min	correct 60 min	+early 60 min	correct 90 min	+early 90 min
Skill score of best performing models based on PI for 90 minute forecasting horizon													
1	multi-step	0, 1, 8	8	5	-0.07	0.43	0.62	0.84	0.84	0.55	0.55	0.47	0.47
2	multi-step	1, 1, 8	4	5	-0.01	0.38	0.59	0.65	0.65	0.46	0.46	0.29	0.29
3	multi-step	0, 1, 8	6	5	0.04	0.53	0.58	0.76	0.76	0.56	0.56	0.39	0.39
4	multi-step	1, 1, 8	8	5	-0.05	0.39	0.58	0.62	0.62	0.38	0.38	0.33	0.33
5	multi-step	1, 1, 8	6	5	0.03	0.51	0.57	0.63	0.63	0.47	0.47	0.30	0.30
6	multi-step	4, 1, 7	10	5	-0.05	0.39	0.57	0.63	0.63	0.39	0.39	0.33	0.33
7	multi-step	3, 1, 8	4	10	-0.07	0.38	0.57	0.71	0.71	0.47	0.47	0.35	0.35
8	multi-step	3, 1, 7	4	10	-0.03	0.46	0.57	0.67	0.67	0.41	0.41	0.37	0.37
9	multi-step	2, 1, 7	4	5	0.06	0.47	0.56	0.66	0.66	0.39	0.39	0.32	0.32
10	multi-step	1, 1, 7	4	10	0.03	0.45	0.56	0.60	0.60	0.36	0.36	0.33	0.33

On Evaluation Measures (sorting based on accuracy)

Dæmningen													
Model					Persistence Index			Accuracy					
	ofc	order	ρ_{nr}	ρ_{lag}	30 min	60 min	90 min	correct 30 min	+early 30 min	correct 60 min	+early 60 min	correct 90 min	+early 90 min
Skill score of best performing models based on accuracy for 90 minute forecasting horizon													
1	multi-step	1, 0, 6	10	5	-3.94	-1.13	-0.22	0.76	0.77	0.58	0.60	0.54	0.55
2	multi-step	7, 0, 7	4	5	-3.68	-1.07	-0.33	0.62	0.63	0.59	0.60	0.54	0.55
3	single-step	3, 0, 6	8	10	-172.94	-51.49	-25.62	0.82	0.88	0.62	0.70	0.52	0.60
4	single-step	4, 1, 6	10	5	-3.67	-2.06	-1.94	0.89	0.90	0.62	0.65	0.50	0.53
5	multi-step	5, 0, 5	8	5	-4.06	-0.90	-0.31	0.70	0.70	0.68	0.69	0.49	0.50
6	single-step	6, 0, 6	6	15	-2.57	-0.57	-0.28	0.75	0.75	0.65	0.66	0.49	0.49
7	single-step	2, 0, 2	4	5	-3.12	-1.14	-0.31	0.74	0.75	0.53	0.55	0.48	0.50
8	single-step	6, 1, 8	6	5	-0.44	-0.03	0.22	0.79	0.79	0.62	0.62	0.48	0.48
9	multi-step	6, 0, 8	2	5	-3.53	-0.89	-0.25	0.63	0.64	0.54	0.54	0.48	0.49
10	single-step	4, 1, 7			-2.02	-0.70	-0.47	0.90	0.90	0.60	0.63	0.48	0.50

On Evalaution Measures

- Selecting based on:
    - PI skill-score overfitts for longer forecasting horizons
    - accuracy gives terrible PI skill-score
- Average PI compromises for different forecasting horizons

Model Selection (sorting based on average-PI)

Skill score of best performing Nelder-Mead optimized models based on average-PI														
Model					Persistence Index			Accuracy						
	ofc	order	ρ_{nr}	ρ_{lag}	30 min	60 min	90 min	correct 30 min	+early 30 min	correct 60 min	+early 60 min	correct 90 min	+early 90 min
Dæmningen													
1	single-step	1, 1, 7	4	10	0.34	0.47	0.45	0.75	0.75	0.38	0.38	0.33	0.33
2	single-step	1, 1, 7	8	15	0.25	0.48	0.50	0.75	0.75	0.58	0.58	0.43	0.43
3	single-step	1, 1, 8	2	10	0.35	0.47	0.40	0.66	0.66	0.40	0.40	0.31	0.31
4	single-step	0, 1, 8	6	15	0.35	0.42	0.42	0.70	0.70	0.45	0.45	0.37	0.37
5	single-step	1, 1, 7	6	5	0.34	0.42	0.42	0.73	0.73	0.43	0.43	0.27	0.27
6	single-step	0, 1, 8	8	5	0.35	0.42	0.40	0.69	0.69	0.33	0.33	0.28	0.28
7	single-step	2, 1, 6	8	5	0.35	0.46	0.36	0.58	0.58	0.39	0.39	0.23	0.23
8	single-step	1, 1, 8	4	10	0.36	0.40	0.40	0.62	0.62	0.31	0.31	0.29	0.29
9	single-step	1, 1, 6	2	5	0.38	0.41	0.37	0.63	0.63	0.39	0.39	0.23	0.23
10	multi-step	0, 1, 6	2	5	0.15	0.52	0.49	0.73	0.73	0.56	0.57	0.36	0.37

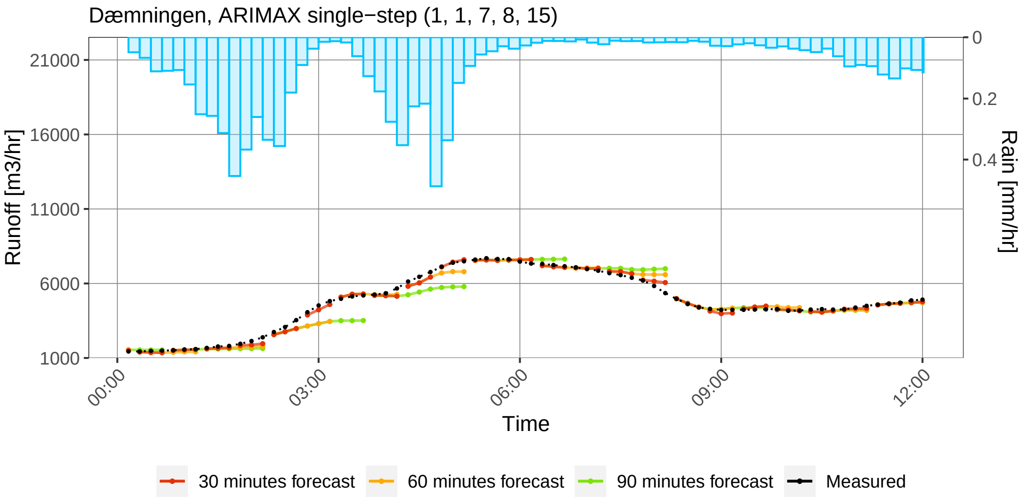
Model Selection (sorting based on average-PI)

Skill score of best performing Nelder-Mead optimized models based on average-PI														
Model					Persistence Index			Accuracy						
	ofc	order	ρ_{nr}	ρ_{lag}	30 min	60 min	90 min	correct 30 min	+early 30 min	correct 60 min	+early 60 min	correct 90 min	+early 90 min
Damhusåen														
1	multi-step	7, 1, 5	2	5	-0.01	0.26	0.39	0.39	0.40	0.31	0.32	0.22	0.23
2	multi-step	4, 1, 6	2	5	-0.05	0.21	0.42	0.45	0.48	0.34	0.37	0.27	0.30
3	multi-step	4, 1, 3	4	5	0.06	0.18	0.33	0.64	0.66	0.47	0.48	0.35	0.39
4	multi-step	4, 1, 7	2	5	-0.09	0.23	0.41	0.50	0.51	0.37	0.38	0.33	0.35
5	multi-step	5, 1, 3	4	5	0.06	0.21	0.27	0.51	0.52	0.38	0.39	0.28	0.32
6	multi-step	1, 1, 8	2	5	-0.08	0.24	0.39	0.49	0.51	0.34	0.36	0.29	0.32
7	multi-step	4, 1, 3	2	5	0.06	0.19	0.28	0.53	0.56	0.39	0.40	0.29	0.33
8	multi-step	4, 1, 2	6	5	0.03	0.18	0.31	0.51	0.52	0.39	0.41	0.34	0.36
9	multi-step	3, 1, 3	4	5	0.05	0.17	0.28	0.57	0.57	0.41	0.42	0.32	0.35
10	multi-step	6, 1, 2	2	5	0.06	0.17	0.28	0.57	0.59	0.39	0.40	0.33	0.34

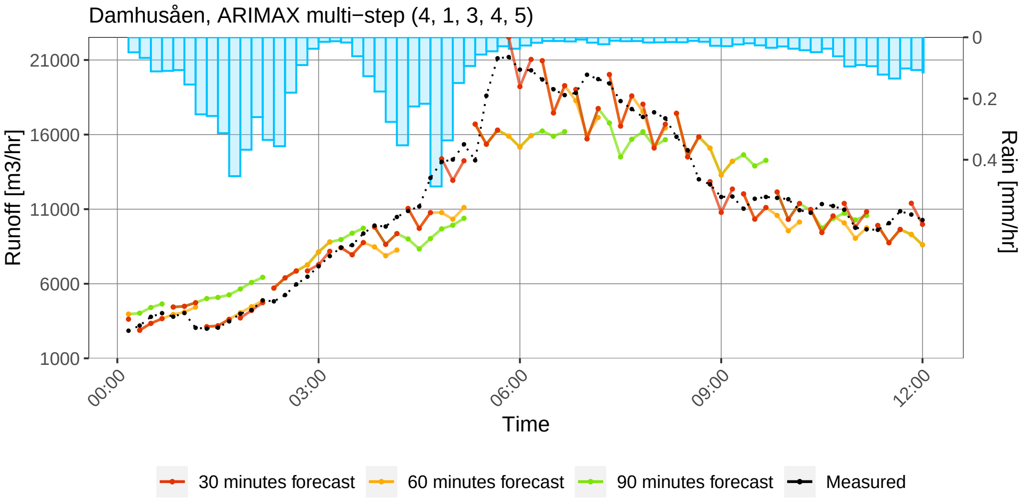
Model Selection

Dæmningen	Damhusaåen
Single-step objective function criteria	Multi-step objective function criteria
d = 1, short AR, and long MA terms	d = 1, medium AR/MA terms
External regressors on lags further back	External regressors on lags which are closer
Models selected	
Model 2: $(1, 1, 7, 8, 15)$	Model 3: $(4, 1, 3, 4, 5)$

Real-World-Forecast of Selected Model



Real-World-Forecast of Selected Model



Research Questions Revisited

R.Q.1 and 2

1. Can competent ARIMA type models be selected in an automated and efficient manner?
   - ARIMA Models selected with reasonable PI skill-score and accuracy in predicting ATS activations
   - For efficiency, calibrations must be done in parallel
2. How should the parameter search-space be constrained such that parameter selection can be performed in a computationally efficient manner, while still selecting parameters that adequately capture the complex behavior of the system?
   - Using Nelder-Mead requires limited search space
   - DDS search can be carried out with bigger search-space

Research Questions Revisited

R.Q.3 and 4

3. Do local and global-searches in the coefficient estimation procedure significantly different models?
    - Local-search: Many good models for high computation cost
    - Global-search: Some good models for low computation cost
    - Local-search achieves wee bit better models
    - increasing the maximum iterations of DDS could produce better models
4. Do different objective function criteria (i.e., calibrating models to single/multi-step forecasts) generate substantially different models?
    - Single-step models more robust
    - Multi-step models tend to overfit, especially on longer forecasting horizons
    - Objective function criteria preference seems to depend on catchment.

Research Questions Revisited

R.Q.5 and 6

5. Will proposed error metrics result in analogous models.
    - Selecting based on accuracy gives poor PI skill-score.
    - Selecting based on PI gives fair accuracy but overfit for longer forecasting horizons.
6. Does precipitation as an external regressor improve forecasting?
    - Precipitation seem to increase perfomance.
    - Makes sense due to correlation.

Research Questions Revisited

R.Q.7

7 How do ARIMA models compare to the current models in use?

- Performance of ARIMA type models promising but comparison biased.
- Current models take uncertainty into account and use actual radar rainfall forecast.

Forecasting horizon	Accuracy	Grey-box	ARIMA
30 min	Correct	0.14,	0.64
30 min	Correct + early	0.39	0.66
60 min	Correct	0.09	0.47,
60 min	Correct + early	0.39	0.48

Outlook

- Increasing iterations of global-search.
- Use 'non-perfect' precipitation forecasts as external regressors.
- How often models need to be re-calibrated, and how much data is needed.
- Account for uncertainty.
- Automated data-cleaning and corrections of measurements