# State-space adjustment of radar rainfall and stochastic flow forecasting for use in real-time control of urban drainage systems

Roland Löwe[1] , Peter Steen Mikkelsen[2], Michael R. Rasmussen[3], Henrik Madsen[4]

[1] Department of Informatics and Mathematical Modelling, Technical University of Denmark (DTU), Denmark, rolo@imm.dtu.dk
[2] Department of Environmental Engineering, Technical University of Denmark (DTU), Denmark, psmi@env.dtu.dk
[3] Department of Civil Engineering, Aalborg University, Denmark, mr@civil.aau.dk
[4] Department of Informatics and Mathematical Modelling, Technical University of Denmark (DTU), Denmark, hm@imm.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 good 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 further 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 red 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. From 25/08/2010 until 14/09/2010 several extensive gaps can be observed in the radar data also during some smaller rain events.

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.

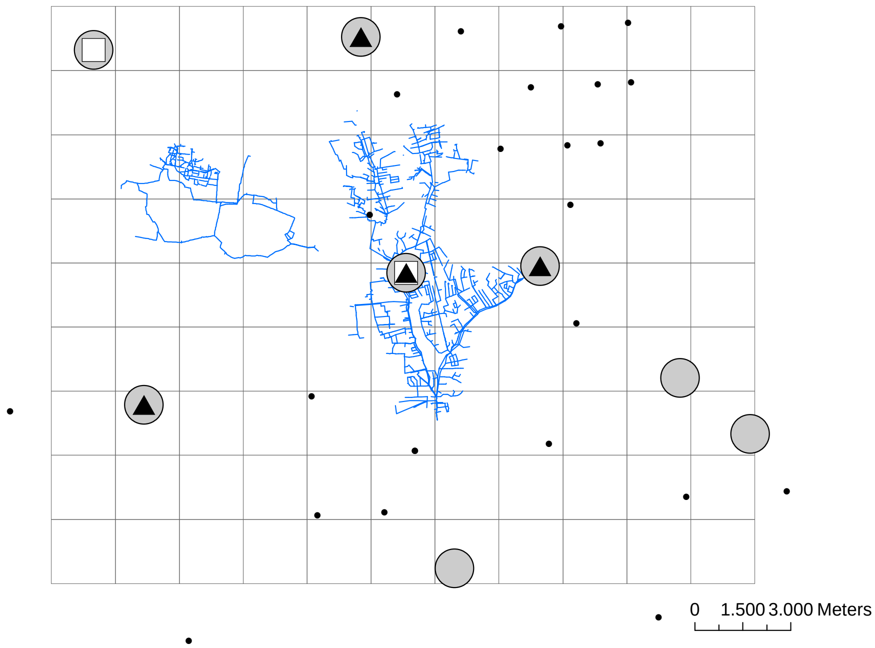


Figure 1. Considered area with C-band radar pixels, Ballerup (left) and Damhusåen (right) catchments, rain gauges in the area (small dots), gauges used for radar adjustment (grey circles), gauges used as input for reference simulations with gauges for Ballerup (white rectangles) and Damhusåen (black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

## 2.2 Radar Adjustment

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

- Create a model to predict rainfall at the next time step for every pixel (system or state equation)
- Relate model prediction and observed rainfall values from different sources in a set of observation equations
- Determine the adjusted rainfall values by weighting between model and observation uncertainties using a Kalman filter

In the state equations as well as in the variance matrices 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.

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

$X_{i,j,t}$ 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. We define

$$\alpha_{k,l} = \begin{matrix} a & k=l=0 \\ \frac{1-a}{8} & k\neq 0, l\neq 0 \end{matrix} \qquad (2)$$

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

$$X_t = A\cdot X_{t-1} + e_t \qquad (3)$$

where $X$ is a vector containing 99 rainfall state values corresponding to 9x11 pixels, $A$ is a weighting matrix according to (1) and $e$ is a 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):

$$Y_t = C\cdot X_t + s_t \qquad (4)$$

The observation vector $Y_t$ contains 99 non-adjusted measurements from all radar pixels and 8 rain gauge measurements. The matrix $C$ relates states and observations (see Grum et al. (2002)) and $s$ is a 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.

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

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.

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

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.

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:

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

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

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, $P$ 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 Lamberti 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 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 $\Delta t$=10min) starting from time step $k$ and integrate them to a predicted runoff volume (Eq. 9).

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

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 to derive upper and lower prediction bounds for the expected runoff volume over the whole prediction horizon.

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

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

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 $\sigma_{Q_{k+i}}$ the standard deviation of the flow prediction $i$ steps into the future starting from time step $k$.

Comparing the above stochastic volume forecast to the observed runoff volume, we find the optimal model parameters by minimizing the skill score ($Sk$) 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.

### 2.4 Forecast Evaluation

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

- Reliability (*Rel*) – percentage of observations not contained in a 95 % prediction 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.

## 3 RESULTS

Table 1 shows the parameters derived for the radar adjustment models 1-4. The small weighting factor $a$ 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) decreases the ratio between the variances of radar and 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 making identifiability of the models difficult. 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 | $1.23\cdot10^{-4}$ | $8.12\cdot10^{-4}$ | $9.12\cdot10^{-4}$ | - | - | - |
| Model 2 | 0.20 | $1.58\cdot10^{-4}$ | $4.90\cdot10^{-4}$ | $3.39\cdot10^{-1}$ | 0.22 | - | - |
| Model 3 | 0.20 | $1.51\cdot10^{-4}$ | $8.12\cdot10^{-4}$ | $3.80\cdot10^{-1}$ | - | 0.614 | 0.384 |
| Model 4 | 0.20 | $1.23\cdot10^{-4}$ | $8.12\cdot10^{-4}$ | $9.12\cdot10^{-4}$ | - | - | - |

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

| Model Input | Ballerup catchment | | | Damhusåen catchment | | |
|---|---|---|---|---|---|---|
| | *Rel* | *ARIL* | *Sk* | *Rel* | *ARIL* | *Sk* |
| 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 on 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 gauges.

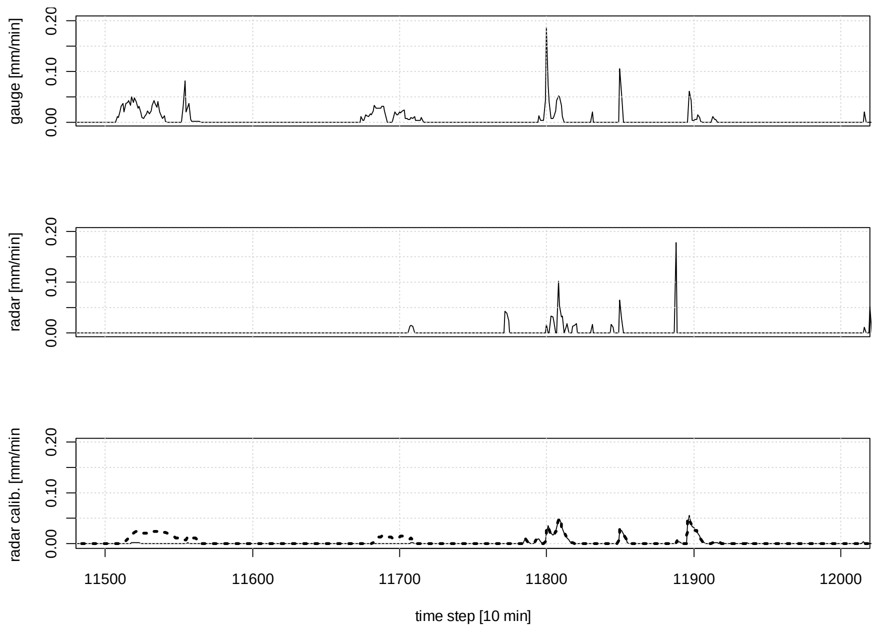


Figure 2. Mean area rain intensities for the Ballerup catchment from gauge measurements (top), non-adjusted radar measurements (centre) and adjusted radar measurements (bottom) without (Model 1, full thin line) and with (Model 4, dotted bold) error marker

## 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 distinguishing between dry and wet weather situations in the physical and stochastic model parts and 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 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 Apluim T. (2009). Adjustment of state space models in view of area rainfall estimation. *Environmetrics*, 22, 530-540.

Grum M., Harremöes 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., 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.

Thordarson F.Ø., Breinholt A., Møller J.K. ,Mikkelsen P.S., Grum M. and Madsen H. (in press). Evaluation of probabilistic flow predictions in sewer systems using grey box models and a skill score criterion. *Stochastic Environmental Research and Risk Assessment*, accepted.

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.