Distinguishing high and low flow domains in urban drainage systems 2 days ahead using numerical weather prediction ensembles

Courdent, Vianney Augustin Thomas; Grum, Morten; Mikkelsen, Peter Steen

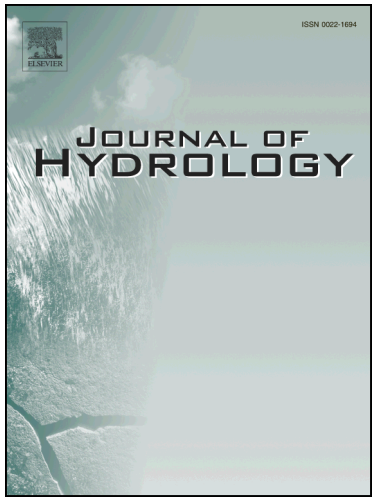
Published in:
Journal of Hydrology

Link to article, DOI:
10.1016/j.jhydrol.2016.08.015

Publication date:
2018

Document Version
Peer reviewed version

Citation (APA):
Courdent, V. A. T., Grum, M., & Mikkelsen, P. S. (2018). Distinguishing high and low flow domains in urban drainage systems 2 days ahead using numerical weather prediction ensembles. Journal of Hydrology, 556, 1013-1025. DOI: 10.1016/j.jhydrol.2016.08.015



Research papers

Distinguishing high and low flow domains in urban drainage systems 2 days ahead using numerical weather prediction ensembles

Vianney Courdent, Morten Grum, Peter Steen Mikkelsen

PII: S0022-1694(16)30506-6
DOI: http://dx.doi.org/10.1016/j.jhydrol.2016.08.015
Reference: HYDROL 21450

To appear in: *Journal of Hydrology*

Received Date: 15 March 2016
Revised Date: 10 July 2016
Accepted Date: 10 August 2016

Please cite this article as: Courdent, V., Grum, M., Steen Mikkelsen, P., Distinguishing high and low flow domains in urban drainage systems 2 days ahead using numerical weather prediction ensembles, *Journal of Hydrology* (2016), doi: http://dx.doi.org/10.1016/j.jhydrol.2016.08.015

Distinguishing high and low flow domains in urban drainage systems 2 days ahead using numerical weather prediction ensembles.

Vianney Courdent[a,b,*], Morten Grum[a], Peter Steen Mikkelsen[b]

* correspondence to vatc@env.dtu.dk. +45 60904238

[a] Krüger Veolia, Gladsaxavej 363, 2860, Søborg, Denmark, vic@kruger.dk, mg@kruger.dk

[b] Department of Environmental Engineering, Technical University of Denmark (DTU Environment), DTU Environment, Bygningstorvet, Bygning 115, 2800, Kgs. Lyngby, Denmark, vatc@env.dtu.dk, psmi@env.dtu.dk

ABSTRACT

Precipitation constitutes a major contribution to the flow in urban storm- and wastewater systems. Forecasts of the anticipated runoff flows, created from radar extrapolation and/or numerical weather predictions, can potentially be used to optimize operation in both wet and dry weather periods. However, flow forecasts are inevitably uncertain and their use will ultimately require a trade-off between the value of knowing what will happen in the future and the probability and consequence of being wrong.

In this study we examine how ensemble forecasts from the HIRLAM-DMI-S05 numerical weather prediction (NWP) model subject to three different ensemble post-processing approaches can be used to forecast flow exceedance in a combined sewer for a wide range of ratios between the probability of detection (POD) and the probability of false detection (POFD). We use a hydrological rainfall-runoff model to transform the forecasted rainfall into forecasted flow series and evaluate three different approaches to establishing the relative operating characteristics (ROC) diagram of the forecast, which is a plot of POD against POFD for each fraction of concordant ensemble members and can be used to select the weight of evidence that matches the desired trade-off between POD and POFD. In the first approach, the rainfall input to the model is calculated for each of 25 ensemble members as a weighted average of rainfall from the NWP cells over the catchment where the weights are proportional to the areal intersection between the catchment and the NWP cells. In the second approach, a total of 2825 flow ensembles are generated using rainfall input from the neighbouring NWP cells up to approximately 6 cells in all directions from the catchment. In the third approach, the first approach is extended spatially by successively increasing the area covered and for each spatial increase and each time step selecting only the cell with the highest intensity resulting in a total of 175 ensemble members. While the first and second approaches have the disadvantage of not covering the full range of the ROC diagram and being computationally heavy, respectively, the third approach leads to both a broad coverage of the ROC diagram range at a relatively low computational cost. A broad coverage of the ROC diagram offers a larger selection of prediction skill to choose from to best match to the prediction purpose.

The study distinguishes itself from earlier research in being the first application to urban hydrology, with fast runoff and small catchments that are highly sensitive to local extremes. Furthermore, no earlier reference has been found on the highly efficient third approach using only neighbouring cells

with the highest threat to expand the range of the ROC diagram. This study provides an efficient and robust approach to using ensemble rainfall forecasts affected by bias and misplacement errors for predicting flow threshold exceedance in urban drainage systems.

KEYWORDS

Urban hydrology; Numerical Weather Prediction (NWP); Ensemble Prediction System (EPS); Probability of Detection (POD); Probability of False Detection (POFD); Relative operating characteristic (ROC)

1 INTRODUCTION

Knowledge about the future generally allows for better decisions in the present, even when the knowledge about the future is uncertain. Weather forecasts are e.g. increasingly used as input to hydrological hazard warning systems (e.g.(Cloke and Pappenberger, 2009; Demeritt et al., 2013)). Within urban hydrology, rainfall forecasts are used as input to hydrologic and hydrodynamic models to predict combined sewer overflows (CSOs) and flooding, and they are used to trigger warnings and to optimize urban stormwater and wastewater systems in real time. Radar extrapolation is e.g. often used to forecast flow (Rouault et al. 2008; Thorndahl & Rasmussen 2013; Vezzaro & Grum 2014). However, radar forecast are limited by their prediction horizon of 1 to 3 hours and by the need to dynamically calibrate the radar images to ground observations from rain gauges (Thorndahl and Rasmussen, 2012). Real time control of urban drainage systems can potentially benefit from longer lead times, e.g. to improve actions towards reducing combined sewer overflow pollution for combined rain events where the optimal management of an event depends on antecedent event characteristics (Courdent et al., 2015), to prepare wastewater treatment plants (WWTPs) for wet-weather conditions (Thorndahl et al. (2013)), and to minimise the energy consumption of WWTPs based on the Smart-Grid control concept (Bjerg et al., 2015). The latter

utilises the storage volume in the drainage system upstream the WWTP to detain water during hours with high electricity prices and releases the water when the price is low, which is the case when e.g. wind or solar energy is available. Smart-grid control is, like e.g. initiation of maintenance and rehabilitation works, only feasible when the system is not under hydrologic stress, i.e. when the discharge is close to the dry weather flow or at least below a defined threshold. The emptying time of urban drainage systems in large cities can easily be in the order of 1-2 days, and therefore there is a need to look for other tools to supplement the radar extrapolation forecasts. Accurate and precise rainfall forecasts several days ahead would be ideal, but distinguishing between high-flow and low-flow domains in selected locations of an urban drainage system is actually sufficient information for some of the above mentioned purposes. This means that weather forecasts may be useful even when associated with a substantial uncertainty.

Numerical Weather Prediction (NWP) models can be used to increase the prediction horizon by generating Quantitative Precipitation Forecasts (QPF) and this potential has been studied in the context of urban hydrology by several authors. Thorndahl et al. (2013) compared NWP and radar extrapolation models for urban runoff forecasting and showed that the weather model performed better predicting the rainfall with lead times from 6 to 12 h than with shorter lead times, because of inadequate initial conditions. Liguori et al. (2012) merged radar extrapolation data and high-resolution NWP forecast data for urban runoff flow prediction purposes (6 hours lead time) and concluded that the overall performance of their rainfall forecasting system decreased with increasing rainfall intensities. NWPs have also been used for forecasting/prediction in other fields: from frost prediction used for optimising road salting and prediction of power production from wind and solar energy (Bacher et al., 2009; Giebel et al., 2005) to streamflow forecasting (Cuo et al. 2011; Shrestha et al. 2013), reservoir inflow prediction (Collischonn et al., 2007) and flood forecasting (Damrath et al., 2000).

The development of high-resolution limited area NWP models have led to improved and more realistic-appearing forecasts, describing mesoscale atmospheric processes including convective rain in an explicit and more detailed way (Sun et al., 2014). This does however not necessarily improve the forecast grid point accuracy, as inevitable errors in timing and position of rain cells are amplified with the increase in spatial resolution. (Walser et al., 2004) demonstrated that the uncertainties in Limited Area Model (LAM) precipitation forecasts increase rapidly with decreasing scale, and that traditional skills scores based on point-to-point comparison at fixed locations or grid points therefore are profoundly degraded by small errors in timing and spatial position (Mass et al., 2002). If the magnitude of a weather event is correctly forecasted but slightly displaced in space then the model will be penalized twice: once for missing the observations and once again for giving a false alarm; this is known as the ‘‘double penalty’’ (Michaelides, 2008). Hence the value of using a finer spatial resolution may be underestimated when using traditional verification procedures, and two overall categories of new performance evaluation methods have therefore been developed: filtering methods and displacement methods (Gilleland et al., 2009).

The fuzzy verification, also known as the neighbourhood approach developed by (Ebert, 2008), is an example of a filtering method. ‘Fuzzy’ verification techniques require that the forecasts are in approximate agreement with the observations, meaning that forecasts are close in space, time and intensity. These techniques typically measure the strength of the agreement as the closeness requirements are varied. Displacement methods such as the object-based approach developed by (Johnson and Wang, 2012) identify and compare features of weather events.

NWP models are highly nonlinear and chaotic and a tiny difference in initial states may thus be amplified into large differences in future states. Hence without notifying the uncertainty of NWP forecasts they may be considered incomplete, especially regarding precipitation which is one of the

most difficult variables to forecast on an urban scale due to its large variability in space, time and intensity (Du, 2007).

Forecast uncertainty is commonly described by meteorologist using Ensemble Prediction Systems (EPS) that are based on a set of deterministic forecasts. Multiple simulations are run with different initial conditions and/or numerical representations of the atmosphere, thereby addressing the two major sources of weather forecast uncertainty (Du, 2007; Gneiting and Raftery, 2005). In comparison, in Climatology, major sources of forecast uncertainty of Global Climate Model (GCM) are model structure, emission scenario and natural uncertainty (Woldemeskel et al., 2012), and GCM EPS are therefore generated based on those parameters.

The motivation of EPS is to use the ensemble spread to characterize the forecast uncertainty, but EPS are generally unable to incorporate all sources of uncertainty. They are generally under dispersive and exhibit systemic biases and therefore EPS do not produce Probabilistic Quantitative Precipitation Forecast (PQPF) directly. Ensemble post-processing methods (sometimes called pre-processing from a hydrological modelling point of view) are necessary to obtain reliable probabilistic forecast as explained in (WWRP/WGNE 2009), and an overview of ensemble post-processing techniques was given by (Wilks, 2011). The neighbourhood method (Theis et al., 2005) and the time-lagged approach (Mittermaier, 2007) were combined by (Ben Bouallègue et al., 2013). Ensemble Model Output Statistics (EMOS) were further developed by (Scheuerer, 2014) using a distribution to characterize the uncertainty and similar approaches have also been developed by (Bentzien and Friederichs, 2012; Kleiber et al., 2011; Sloughter et al., 2007).

It is common meteorological practice to analyse forecasts based on thresholds of precipitation (e.g. exceedance of 1 mm/24 h). However, from a hydrological viewpoint these thresholds are rather arbitrary. To distinguishing high-flow and low-flow domains as mentioned above, a hydrologically

focused evaluation based on coupling of meteorological and hydrological forecast models in a cascade as suggested by (Pappenberger et al., 2008) will give more useful information. The prediction skills can then be assessed based on the modelled catchment discharge, which will however be affected both by the hydrological model uncertainty and the weather forecast uncertainty.

In this article we evaluate the feasibility of using forecast data from a NWP EPS in combination with different ensemble post-processing approaches and a simple hydrological model to distinguish high-flow and low-flow domains in an urban drainage system up to two days ahead. Particular focus is in quantifying the relative operating characteristic (ROC) diagram, which illustrates the probability of detection (POD) against the probability of false detection (POFD) of high-flow events for different fractions of concordant ensemble members. In section 2 we describe the coupled meteorological and hydrological model used, starting with the NWP EPS data, then the study case and the hydrological model developed for the purpose of this study. The performance evaluation methods based on contingency tables and ROC diagrams are explained in section 3 together with the NWP EPS post-processing approaches. Section 4 finally presents and discusses the results and section 5 contains the conclusions.

2. MATERIAL: NWP DATA AND HYDROLOGICAL MODEL

As underlined by (Shrestha et al., 2013) the evaluation of NWP model output for streamflow forecasting purposes should be done with a hydrological perspective, and we thus wish to do the same for urban drainage flow forecasting. Hence as suggested by (Pappenberger et al., 2008), we based the forecast evaluation on a coupled meteorological and hydrological model, using discharge predictions and discharge observations (rather than precipitation forecasts and observations). The methodology respects the importance of dominant hydrological processes and the non-linear error

transformation by the hydrological model. It should be noted that this approach has two main drawbacks: (i) the hydrological model is often calibrated based on rain gauge observation data which have a different error structure than that of the forecasts of the meteorological model, and (ii) the uncertainty of the hydrological model itself has to be acknowledged in this model cascade.

This section describes the Numerical Weather Prediction (NWP) model ensemble rainfall forecast data used. Then the urban catchment used for the study case is presented and the hydrological model is described. Finally outputs from the coupled model are presented.

2.1. The HIRLAM-DMI-S05 ensemble prediction system (EPS)

The international research program HIRLAM (HIgh Resolution Limited Area Model) is a research cooperation of European meteorological institutes, amongst those the Danish Meteorological Institute (DMI). The aim of this consortium is to develop and maintain up-to-date NWP systems for 1- and 2-day forecasts within a limited area. The HIRLAM model is a hydrostatic grid-point model for short-range weather forecasting at synoptic scale (5 - 15 km horizontal resolution).

This study uses the HIRLAM Ensemble Prediction System (EPS) run by DMI: DMI-HIRLAM-S05. This system covers the Scandinavian countries and north Europe, see Figure 1, with a horizontal resolution of 0.05° (approx. 5.6 km). To estimate boundary conditions this model is nested into the coarser (0.15° horizontal resolution) and larger DMI-HIRLAM-T15 model (Mahura et al., 2006), which is itself nested into the global ECMWF IFS model (http://www.ecmwf.int).

FIGURE 1 APPROX. HERE

The DMI-HIRLAM-S05 ensemble is a 2-dimensional EPS (i.e. considering both different initial conditions and different physical processes (Du, 2007)) comprising 25 members as described in Table 1. The different initial conditions are generated using the most recent atmospheric

measurements analysis which is then perturbed based on previous 6h and 12h old forecasts from the model.

Each of the five initial conditions is run with different model structures, using either the STRACO or the Kain-Fritsch/Rasch-Kristjansson schemes to model convection and condensation processes with or without the inclusion of stochastic physics resulting in a total of twenty ensemble members. The five remaining ensemble members are dedicated to study the impact of perturbing the roughness lengths for urban areas, which are parameters describing the surface fluxes (interaction between the land cover and the atmosphere); the roughness lengths is randomly chosen in the range 0.05-1.1 m (Feddersen, 2009). For further description of the processes and parameters mentioned above, see the HIRLAM technical documentation (Unden et al., 2002) and the HIRLAM website (http://www.hirlam.org/).

TABLE 1

Ensemble members	STRACO		KF/RK		STRACO
		Stoc. Phys.		Stoc. Phys.	Pert. Roughn.
Ini. cond. 1	1	6	11	16	21
Ini. cond. 2	2	7	12	17	22
Ini. cond. 3	3	8	13	18	23
Ini. cond. 4	4	9	14	19	24
Ini. cond. 5	5	10	15	20	25

It is important to underline that the perturbations of the initial conditions are not random and they do not sample the (unknown) observation uncertainty. Hence the ensemble members are not equally likely. Indeed, the perturbations tend to maximize the ensemble dispersion, and so the individual ensemble members will be, in general, less likely but closer to more extreme outcomes than the random approach. Nevertheless, experience has shown that ensemble forecasts tend to be

underdispersive in the sense that (too) many verifying observations fall outside the ensemble range (Feddersen, 2009).

The DMI-HIRLAM-S05 model is run every 6 hours and generates output with hourly time steps over a 54h forecast horizon. Ensemble data from June 2014 to January 2016 (20 months) were used in this study. Figure 2 gives an example of the spatial discrepancy among ensemble members at a given time step. Major dissimilarities are observed between the localisation as well as the intensity of rain of the individual ensemble members.

FIGURE 2 APPROX. HERE

2.2. Catchment, urban drainage system and hydro-meteorological observations

The Damhuså urban drainage catchment (Copenhagen, Denmark) was used for this study. The catchment covers a highly urbanised area with compact residential housing and some industrial activity and is equipped with a combined sewer system which conveys wastewater, rainfall runoff from paved surfaces and infiltration inflow especially in the winter months. Modern urban drainage systems are often equipped with various flow control infrastructures (e.g. detention basins and pumping stations implemented to improve the UDS management, e.g. (Vezzaro and Grum, 2014). The Damhuså catchment was chosen for the absence of flow control infrastructures affecting the catchment response, in order to simplify the modelling approach.

FIGURE 3 APPROX. HERE

Precipitation is measured using tipping bucket rain gauge stations placed at several locations in the catchment (blue circles on Figure 3). These rain gauges are part of the national Danish SVK rain gauge network which is operated by the Danish Meteorological Institute (DMI) and the Water Pollution Committee of the Danish Engineers Society (SVK - Spildevandskomiteen in Danish)

since 1979, and the recorded data has a 1 minute temporal resolution (Jørgensen et al., 1998). The rainfall events labelled "defect" by the SVK system were removed with the exception for the defect event affiliated to temperature below 3°. The catchment outlet, represented by the red hexagon on Figure 3, is a combined sewer pipe with a maximum capacity of 10,000 m^3/h. This outlet is monitored by the utility company HOFOR using an electromagnetic flow meter, and the flow data has a 2 minutes temporal resolution.

Monthly average potential evaporation data were calculated based on a historical (from 1989 to 2010) time series covering the Damhuså catchment (20x20 km grid cell), see (Scharling and Kern-hansen, 2012) for further description of the data.

2.3 Model description

The hydrological model is composed of 3 main conceptual parts describing wastewater flow from households, fast rainfall-runoff from impervious areas and slow runoff caused e.g. by infiltration-inflow and correlated to the potential evaporation (Figure 4).

FIGURE 4 APPROX. HERE

2.3.1. Household waste water flow variations

The wastewater pattern was assumed to be constant throughout the year and was modelled using a second order Fourier transform, which is common practice in this field e.g. (Carstensen et al., 1998; Langergraber et al., 2008; Talebizadeh et al., 2016), see equation (1) where a_0, a_1, b_1, a_2 and b_2 are the Fourier series, parameters and w the frequency of the pattern.

$$ww(t) = a_0 + a_1 \cos(w * t) + b_1 * \sin(w * t) + a_2 * \cos(2 * w * t) + b_2 * \sin(2 * w * t) \quad (1)$$

2.3.2. Fast and slow rainfall-runoff processes

The rainfall-runoff process was modelled using a lumped conceptual model based on the Nash linear reservoir cascade concept (Nash, 1957). Hence, the catchment response was represented by a set of reservoirs in series with a linear relationship between the reservoir outflow, Q(t), and the amount of water stored, S(t), as shown in equation (2).

$$S(t) = K * Q(t) \quad (2)$$

Assuming continuity, the Instantaneous Unit Hydrograph (IUH) is described by equation (3) with two hydrologic parameters: the number of linear reservoirs, n, and the storage coefficient, K.

$$IUH(t) = \frac{1}{K^n} * \frac{1}{\Gamma(n)} * t^{n-1} * \exp\left(-\frac{t}{K}\right) \quad (3)$$

where $\Gamma(n)$ is the gamma function defined by $\Gamma(n) = \int_0^{+inf} e^{-t} * t^{n-1} * dt$. If n is an integer the equation can be simplified to $\Gamma(n) = (n-1)!$

Different runoff patterns were observed during winter and summer months. During the summer period, the discharge reverts to the dry weather condition only few hours after the rainfall event (up to 10 hours), we call this process fast runoff. Whereas during winter period a rainfall event impacts the discharge during several days (up to 10 days), we called this process slow runoff.

The fast runoff is assumed to have a constant pattern through the year, whereas the slow runoff is assumed to be connected to a wetness index which is calculated based on potential evaporation and previous rain events (catchment's memory). The monthly average potential evaporation was calculated based on the historical (from 1989 to 2010) time series covering the Damhuså catchment (20x20 km grid cell), see (Scharling and Kern-hansen, 2012) for further description of the data.

2.3.3. Parameter estimation

The wastewater flow parameters were estimated first, using flow observations from summer periods without rainfall to avoid influence from the two other processes. Using fixed wastewater parameters, the fast runoff process parameters were then estimated based on rain and flow data for rain events during summer months to avoid influence from the slow runoff process, which was calibrated last using rainfall, flow and potential evaporation data for the all period (from November 2012 to November 2014). In all cases calibration was conducted using the Differential Evolution Adaptive Metropolis (DREAM) method (Laloy and Vrugt, 2012) considering the root mean square error as objective function.

2.3.4. Coupled meteorological and hydrological model output

Figure 5 shows example output from the coupled meteorological and hydrological model for a day in August 2015. The top figure shows the expected rainfall over the catchment (using the first NWP post-processing method (3.2.1)), and the bottom figure shows the output from the hydrological model based on this rainfall input. The flow ensembles appear to embrace the flow observations to a larger extent than rainfall ensembles embrace the averaged rain gauges measured rainfall. The hydrological model was initialised using measured rain gauge data prior running the NWP model.

FIGURE 5 APPROX. HERE

Table 2 displays the performance of the ensemble forecast from the hydrological model for different flow domains (<3,000 m^3/h, …, >9,500 m^3/h). The first section of Table 2 shows the performance of the ensemble spread, i.e. the percentage of observation that are either below, contained within or above the ensemble spread, as well as the average band wind (ABW), i.e. the average distance between the lower (5%) and upper (95%) prediction quantile. The second section of Table 2 shows measures of performance for the median of the ensemble. The mean absolute error from the EPS median (MAE_{median}) is the average of the absolute error between the EPS median and the

observation. The bias from the EPS median ($Bias_{median}$) is the average distance between the ensemble median and the observation.

TABLE 2

	Observed flow [m^3/h]					
	< 3000	3000-4000	4000-5000	5000-7500	7500-9500	> 9500
Below EPS spread	59 %	31%	18%	9%	5%	4%
Contained within the EPS spread	31%	47%	58%	55%	43%	48%
Above EPS spread	10 %	22%	24%	35%	52%	48%
ABW [m^3/h]	1,022	1,883	2,766	3,193	3,115	3,565
$\boldsymbol{MAE_{median}}$ **[m^3/h]**	661	1,129	1,517	2,104	3,382	2,954
$\boldsymbol{Bias_{median}}$ **[m^3/h]**	594	315	4	- 1,180	- 3,052	- 2,792

ABW: Average band width; MAE: Mean absolute error

The ensemble spread is generated by routing the rainfall through the runoff model whereas the wastewater from households is based on a deterministic model (2.3.1) without contributing to the spread. Hence for the low flow domain (< 3000 m^3/h), which is mainly driven by the wastewater process, the ensemble spread shows poor quality and has lower ABW than the other flow domains. The large ABW of the EPS and MAE of the median quantile underline the substantial prediction uncertainty and support flow domains prediction rather than quantitative prediction. A shift in the bias is observed between low and high flow thresholds, low flows are over estimated while high flows are under estimated. This can be related to the rainfall prediction skill of NWP, which tend to over predict small rain events and miss high intensity rain events (Feddersen, 2009). The large percentage of the observations above the ensemble spread and the large negative bias for the high flows underlines the need for approaches to limit the missing/underestimation of events.

3. METHODOLOGY

This section is separated into two parts. The first part describes the prediction performance evaluation methods. The second part describes NWP post-processing strategies developed to enhance flow domain prediction skills.

3.1. Forecast performance evaluation methods

Weather forecast evaluations are most often conducted on variables and methods of meteorological relevance. To match our model purpose, distinguishing urban flow domains, we have chosen to use a hydrologically focused evaluation of meteorological forecasts based on a coupled (meteorological & hydrological) model as suggested in (Pappenberger et al., 2008). Thus, the prediction skill is assessed directly on the basis of the catchment outlet discharge. Using this approach, the hydrological model structure uncertainty is added to the overall prediction uncertainty.

3.1.1. Contingency table

The distinction between 2 flow domains based on a given flow threshold generates binary outputs. Hence each time step only has 2 outcomes: either an event occurs if the flow is above the threshold or no-event occurs if the flow is below the threshold. The occurrence/non-occurrence of an event are represented using the binary notation Obs and $\overline{Obs}$. Similarly the prediction/non-prediction of an event are noted Nwp and $\overline{Nwp}$. Comparing the predictions and the observations leads to 2 x 2 possible outcomes (contingencies) as shown in Table 3. The "Hits" represents the number of high-flow events that were correctly forecasted, i.e. both the forecast and the observation exceeded the flow threshold ($Obs \cap Nwp$). The "Misses" represents the number of high-flow events that were not forecasted ($Obs \cap \overline{Nwp}$). The "False Alarms" represents the number of occurrences that were forecasted to be high but which turned out not to be high ($\overline{Obs} \cap Nwp$) and the "Correct Negatives"

represents the number of occurrences where high flow was neither forecasted nor observed ($\overline{Obs} \cap \overline{Nwp}$). Those contingency table outputs are used to define prediction skill scores in Table 4.

TABLE 3

Forecast exceeding a given threshold	**Observation exceeding a given threshold**	
	Yes	No
Yes	Hits (H)	False Alarms (FA)
No	Misses (M)	Correct Negatives (CN)

The probability of detection (POD) measures the fraction of observed events that were correctly forecasted and the probability of false detection (POFD) gives the fraction of observed non-events that were forecasted to be events: these two key quantities are used to generate relative operating characteristic (ROC) curves as described below (3.2.1). The POD represents a major objective: to correctly predict high flow periods, whereas the POFD considers the false alarms and represents the cost of being too conservative when aiming to avoid missing high flow periods. An overview of criteria and verification methods can be found in meteorological literature (WWRP/WGNE, 2009) and (Wilks, 2011).

TABLE 4

Score	**Formula**	**Detailed Formula**	**Range**	**Perfect**
POD	H/(H + M)	$POD = \frac{\sum_{j=1}^{nForecast}\sum_{k=1}^{nTimeStep}(Obs_{j,k} \cap Nwp_{j,k})}{\sum_{j=1}^{nForecast}\sum_{k=1}^{nTimeStep}(Obs_{j,k})}$	[0,1]	1
POFD	FA/(FA + CN)	$POFD = \frac{\sum_{j=1}^{nForecast}\sum_{k=1}^{nTimeStep}(\overline{Obs_{j,k}} \cap \overline{Nwp_{j,k}})}{\sum_{j=1}^{nForecast}\sum_{k=1}^{nTimeStep}(\overline{Obs_{j,k}})}$	[0,1]	0

Contingency tables, as well as the prediction skill scores introduced in Table 3, can be computed for a single/fraction of ensemble member(s) and a selected forecast horizon (thus allowing to distinguish the forecast skill for different forecast horizons), or a for single/fraction of ensemble

member(s) and a range of forecast horizons as a whole (thus evaluating the average forecast skill across forecast horizons).

3.1.2. Relative operating characteristic (ROC)

The relative operating characteristic (ROC) is the graph of the probability of detection (POD) against the probability of false detection (POFD) that is mapped out as the decision threshold varies, Figure 6. In our case the decision threshold corresponds to the fraction of ensemble members (f_{EM}) that predict an event and which is used to trigger a decision. In order to avoid confusion with the flow threshold, this decision threshold is called thereafter "weight of evidence". The ROC diagram illustrates the ability of the forecast to discriminate between the occurrence and the non-occurrence of an event. A perfect forecast lies at the point (POFD, POD) = (0, 1), in the top left corner point as shown by the red line in Figure 6. The point (0, 0) corresponds to never forecasting an occurrence, while the point (1, 1) corresponds to constantly (and wrongly) forecasting occurrences.

The ROC diagram of an EPS can be plotted by generating 2x2 contingency table and values for POD and POFD through each weight of evidence possible (f_{EM}, which ranges from 1/N to N/N with N the size of the EPS). The point on the furthest left corresponds to the prediction skill for a decision requiring all ensemble members to predict an event (high weight of evidence ($f_{EM} = {}^{N}/_{N} = 1$), while for the point on the furthest right only a single positive ensemble member is required to trigger a decision (low weight of evidence ($f_{EM} = {}^{1}/_{N}$). The selection of a given weight of evidence converts the EPS to a single deterministic prediction as shown by eq. (4).

$$Nwp_{j,k}\,(f_{EM}) = \left(\frac{\sum_{EM_i=1}^{N} Nwp_{EM_i,j\,k}}{N} \geq f_{EM}\right) \qquad (4)$$

The expressions of POD and POFD for a given weight of evidence are shown by eq. (5) and (6) which combine formula of table 3 and the eq. (4).

$$POD(f_{EM}) = \frac{\sum_{j=1}^{nForecast}\sum_{k=1}^{nTimeStep}\left(Obs_{j,k} \cap Nwp_{j,k}(f_{EM})\right)}{\sum_{j=1}^{nForecast}\sum_{k=1}^{nTimeStep}(Obs_{j,k})} \quad (5)$$

$$POFD(f_{EM}) = \frac{\sum_{j=1}^{nForecast}\sum_{k=1}^{nTimeStep}(\overline{Obs_{j,k}} \cap \overline{Nwp_{j,k}(f_{EM})})}{\sum_{j=1}^{nForecast}\sum_{k=1}^{nTimeStep}(\overline{Obs_{j,k}})} \quad (6)$$

The score skill of a ROC diagram is calculated based on the area under the curve (ROCA). The ROCA ranges from 0 to 1 where a score of 1 corresponds to a perfect forecast and a score of 0.5 corresponds to a random forecast (i.e. no skill).

FIGURE 6 APPROX. HERE

3.2 NWP EPS post-processing

We investigated three NWP post-processing approaches. The first approach, which is used as a reference point for the two others, considers the intersection of the catchment with the NWP model grid cells directly above the catchment. The two other approaches explore how inclusion of forecasts from neighbouring grid cells can improve the prediction characteristics by compensating for inherent spatial misplacement of the rainfall predictions. In these approaches we included neighbouring grid cells up to a radius of six cells from the centre of the catchment. Those two methods are more "conservative" than the first one, as they attempt to minimize the occurrence of missed events.

3.2.1. Weighted areal overlap method

In the first approach, only the rainfall predictions from the grid cells overlapping the catchment were considered (Figure 7). The predictions from these grid points were lumped into one rainfall intensity input to the hydrological model based on their percentage of overlap with the hydrological

catchment. The percentages of areal overlap of each grid points were 5% (1966 and 2080), 10% (1965), 15% (2024), 30% (2022 and 2081) and 90% (2023). Based on these percentages of areal overlap, each grid cell was associated with a weight to calculate the mean rainfall intensity value predicted over the catchment. Similarly, (Johnson et al., 1999) used mean areal estimate from radar data as input for the 8 hydrological basins implemented in their hydrological model.

FIGURE 7 APPROX. HERE

3.2.2. Neighbourhood inclusion methods

The aim of this second approach is to investigate to what extent inclusion of rainfall predictions from neighbouring grid cells is able to compensate for the spatial errors of the NWP model and, thus, improve the ROC. Our expectation that this will improve the ROC is motivated by the observation that location errors dominate precipitation forecast errors when compared to errors in forecasted volume and pattern (Ebert and McBride, 2000). Simulated convective storm are often a few hours off, or a few tens of kilometres away from the observed one (Bernardet et al., 2000).

This approach was first developed by (Theis et al., 2005) as a pragmatic, low-budget post-processing procedure that derives probabilistic precipitation forecasts from deterministic NWP model output by creating a "pseudo ensemble". The concept was further developed by (Schaffer et al., 2011) and (Ben Bouallègue et al., 2013) to expand EPS generating a ''super-ensemble''. This concept has also been used for forecast verification purposes, e.g. (Ebert, 2008), (Gilleland et al., 2009) and (Ebert, 2009).

The surrounding areas used for the neighbourhood ensemble expansion method were defined based on a circle centred in the catchment. Figure 1 displays the scope of the neighbourhood, from the smallest vicinity in dark grey to the largest vicinity (radius of 6 grid cells) in light grey. The radii and corresponding number of ensemble members are shown in Table 5 below. Note that, whereas

we in the first approach (3.2.1) created a weighted average of the overlapping cells and thus left the ensemble size unchanged, we here pooled all the ensemble members within the considered radius and simulated a flow forecast for each ensemble member individually.

FIGURE 8 APPROX. HERE

TABLE 5

Inclusion radius	1	2	3	4	5	6
EPS size	125	325	625	1225	1825	2825

3.2.3 Maximal Threat EPS

To limit the ensemble size, considering that a hydrological simulation needs to be run for each ensemble member, a different post-processing strategy was defined based on the worst case scenario within the area defined by the inclusion radius. For each ensemble member and at each time step the highest rainfall intensity with the considered area was selected. The aim was here not to produce QPF but to assess the maximal threat within a given surrounding in order to avoid miss predicted event. This approach is increasingly conservative with the increase of the encompassed area's radius.

4. RESULTS AND DISCUSSION

This section presents the prediction skill of the coupled meteorological and hydrological model to distinguish between flow domains, in terms of POD, POFD and ROC diagrams. In addition to the NWP post-processing method presented in the previous section, we investigated how the prediction skill depends on various parameters such as the forecast lead time, the fraction of ensemble members (EMs) predicting an event, and the selected high flow threshold.

4.1. Variation of POD and POFD with forecast lead time

The predictions were assessed based on the Probability of Detection (POD) and the Probability of False Detection (POFD) as descripted in Table 4. These two parameters are often in competition as a conservative approach improves the detection but generates more false alarms.

FIGURE 9 APPROX. HERE

The general tendency is for the prediction skills to degrade (POD decreases and POFD increases) with increasing lead time, Figure 9. However, in the case of the conservative strategies, maximal threat EPS within a radius of 3 (dotted line on the right Figures) and 5 grid cells (dot-dashed line on the right Figures), there is no significant decrease of the POD with lead time (according to the Mann Kendall trend test). The improved detection performance comes at the expense of large increase of their POFD. In contrast, the POFD does not show a significant increase with lead time for the weighted areal overlap method with a number of EMs of 16 (dot-dashed line on the left Figures). The choice of NWP post-processing method, the chosen inclusion radius, and the fraction of EMs predicting high-flow all impact the variation of the prediction skill with lead time. Strategies can be selected to avoid the degradation of one criterion at the expense of the other.

It is noteworthy that initial prediction skills are alike for all prediction strategies (Figure 9). This similarity originates from the hydrological model initialisation using measured rain gauge data and the response time of the hydrological catchment. Despite being driven by measurement data those first time steps do not achieve perfect prediction. Indeed rain gauge measurements using a zero forecast horizon has an almost perfect POFD of 0.1% but scores only 80.1% for its POD, as shown on the ROC diagrams of Figure 10 by the purple asterisk. The missed high flow events result from measurement errors of the rain gauges used (Jørgensen et al., 1998) and from processes that are not

included in the hydrological model (e.g. snow melt, higher groundwater table and infiltration). Indeed, those are mainly winter processes and Table 6 shows lower prediction skill during the winter period.

TABLE 6

	Full period	**Winter**	**Spring**	**Summer**	**Autumn**
POD	*0.80*	*0.70*	*0.83*	*0.89*	*0.86*
POFD	*0.010*	*0.009*	*0.003*	*0.013*	*0.013*

4.2. Using ensembles to compute the ROC diagram

The prediction consistency among ensemble members (EMs) determines the weight of evidence of a prediction, i.e. a larger number of EMs predicting an event leads to a stronger evidence for an event to occur. Hence low weights of evidence lead to more conservative predictions. Table 7 shows an extract of the contingency table and skill scores for different fractions of ensemble members that predict an event, calculated as an average across the 48 hours forecast horizon and using the weighted areal overlap post-processing method and a flow threshold of 4,000 m^3/h.

TABLE 7

	Hit	False Alarm	Miss	Correct Negative	POD	POFD
Average of individual EMs	4.8 %	4.8 %	3.8 %	86.6 %	58.3 %	4.5 %
Forecast at least by 1 Ensemble Member (EM_1)	7.1 %	16.3 %	1.6 %	75.0 %	81.9 %	17.9 %
EM_2	6.7 %	12.6 %	1.9 %	78.8 %	77.3 %	13.8 %
…	…	…	…	…	…	…
EM_{10}	5.3 %	4.9 %	3.3 %	86.5 %	61.5 %	5.4 %
…	…	…	…	…	…	…
EM_{20}	3.6 %	1.5 %	5.0 %	90.0 %	42.3 %	0.17 %
…	**...**	…	…	…	…	…

EM_{25}	2.3 %	0.4 %	6.3 %	91.0 %	27.3 %	0.005 %

The range of the available prediction skills are displayed on the ROC diagram, Figure 10a. Each triangle represents a fraction of EMs predicting a high flow event. The point furthest to the left corresponds to a fraction of 25/25, i.e. all EMs are predicting a high flow. Correspondingly, the triangle furthest to the right corresponds to a fraction of 1/25, i.e. only one of the EMs is predicting a high flow event. The skill scores of individual EMs, blue circles in Figure 10a, are gathered in the same area with an average POD of 58.3% and an average POFD of 4.5% (Table 6). This result underlines the benefit of EPS compared to individual EMs. Indeed, using ensembles gives the opportunity to select the weight of evidence the (i.e. fraction of EM predicting an event) that most appropriately balances POD against POFD for the prediction purpose at hand.

As displayed in Table 6 and on Figure 10a, the approach considering the areal intersection between the catchment and the NWP grid cells as input to the hydrological model (weighted overlap method) leads to a maximal POD of 81.9%. Hence a significant proportion of high flow events are missed by the predictions using this approach, which suggests that more conservative approaches are employed to complete the ROC diagram.

FIGURE 10 APPROX. HERE

4.3. Neighbourhood Ensemble Expansion

The nearest neighbourhood approach was performed for different inclusion areas, Figure 10b shows the ROC curve for the largest inclusion area tested, a radius of 6 grid cells (approx. 33.6 km), which generates an ensemble with 2825 members. This method provides a large and dense range of prediction skills, with a maximal POD of 94.3% that is significantly higher than when using rain

gauge measurement as input: 80.1% (purple star). This result suggests that the precaution on the NWP input partly compensates for hydrological model errors.

4.4. Maximal Threat EPS

In this approach, for each of the 25 EMs, the highest threat within the given vicinity is used. The ROC diagram, Figure 10c, was created by merging predictions strategies using increasingly larger vicinity areas. This method generated a large range of prediction skills. Figure 10c shows that the variation of the prediction skills generated with different radii are complementary and gave a similar but scarcer ROC diagram than the neighbourhood ensemble method. The most conservative prediction has a POD of 97.0% for a POFD of 42.3%. Despite this extremely high rate of false alarms such prediction can still prove to be useful, e.g. for energy optimisation and maintenance purposes which require high POD to avoid jeopardizing the system. Furthermore as mentioned in section 4.1 part of the prediction error comes from the hydrological model and could be improved.

4.5. Influence of the discharge threshold

The methods developed in this article were evaluated using three different flow thresholds (Figure 11):

FIGURE 11 APPROX. HERE

• 2,500 m^3/h which is close to the highest discharge during dry periods.

• 4,000 m^3/h which is used as distinction between low and high flow domains.

• 9,500 m^3/h which is close to the pipe maximal capacity (approx. 10,000 m^3/h).

The high threshold can be used for CSO prediction, whereas the medium and low threshold can be used for maintenance planning or energy consumption optimisation.

The ROC diagram in Figure 11 is generated by joining 2 prediction strategies, the weighed areal overlap (circles) and the maximal threat method with a radius of 6 grid cells (triangles). The skills shown in grey colour correspond to Figure 10 a and c (4,000 m^3/h threshold). The prediction of exceeding the low flow threshold (in blue) shows lower performance than the two other thresholds. Its higher POFD may be explained by the tendency of NWPs to over-predict small rain events, whereas the lower POD may be explained by the simplicity of the hydrological model, which represents the tail of the hydrograph poorly during the winter period. Medium (in grey) and high (in black) thresholds demonstrate similar prediction skills but the prediction of the high threshold fails to reach high POD with a maximum of 89.3%. This limitation may be explained by the difficulties of the NWP model to accurately predict high intensity events.

5. CONCLUSION

In this paper the prediction of flow domains for an urban catchment using EPS NWP data was assessed. In order to be purpose oriented the predictions were evaluated using a hydrological perspective focusing on catchment discharge rather than the rainfall intensity. Hence a hydrological rainfall-runoff model was developed and coupled with output from the DMI-HIRLAM-S05 NWP ensemble prediction system (EPS).

This coupled model was run first using the weighted overlap method where rainfall was post-processed considering the areal overlap between the hydrological catchment and the NWP grid cells. Forecasting using a single deterministic NWP output provides a specific prediction skill, i.e. a single combination of POD and POFD. However, ensembles provide a large range of prediction skill more or less conservative depending on the fraction of ensemble members used to trigger an event. Hence an optimal weight of evidence can be selected to match the prediction purpose.

The ensemble approach showed limitations, failing to reach high probabilities of detection (POD), which may be explained by the spatial uncertainty of rainfall prediction. Indeed, urban scale applications require weather models with fine resolution and this accentuates the problematic of rainfall forecast misplacement. In order to cope with this limitation, two EPS post-processing methods were implemented to incorporate the predictions from surrounding grid cells. The (i) neighbourhood ensemble expansion method increases the EPS dimension by adding predictions from the surrounding grid cells, whereas (ii) the maximum threat EPS method selected the worst case scenario within the inclusion area. Both approaches improved the scope of the prediction skills available and increased the maximal value POD available. Thereby NWP post-processing is necessary to improve the scope of conservative approach, avoiding missed events by accounting for misplaced rainfall prediction.

The three approaches were tested for different flow thresholds: low to medium flow thresholds can e.g. provide information for maintenance planning or energy consumption optimisation, whereas a high flow threshold can be used for predicting e.g. CSO events. Low threshold prediction showed inferior prediction skill due to limitations of the used hydrological model. Indeed, using a coupled meteorological and hydrological model lead to an overall prediction skill is influenced by both models performance. Hence the predictions skills might be improved using a more advanced hydrological model with e.g. data assimilation and snowmelt processes incorporated.

Additional NWP post-processing methodologies could be assessed and applied to improve the prediction even more: The neighbourhood method could be extended to the temporal dimension to account for mistiming. Consistency between consecutive and overlapping NWP model runs could be included as well.

Acknowledgments

This research was financially supported by the industrial PhD programme of the Innovation Fund Denmark. The catchment and flow data were kindly provided by Copenhagen Utility Company (HOFOR). We would like to thank the Danish Meteorological Institute (DMI), especially Henrik Feddersen, for providing the NWP data from their DMI-HIRLAM-S05 model.

Supplementary material

Animation of Figure 2 through forecast lead time: "Example of precipitation ensemble forecast.avi"

References

Bacher, P., Madsen, H., Nielsen, H.A., 2009. Online short-term solar power forecasting. Sol. Energy 83, 1772–1783. doi:10.1016/j.solener.2009.05.016

Ben Bouallègue, Z., Theis, S.E., Gebhardt, C., 2013. Enhancing COSMO-DE ensemble forecasts by inexpensive techniques. Meteorol. Zeitschrift 22, 49–59. doi:10.1127/0941-2948/2013/0374

Bentzien, S., Friederichs, P., 2012. Generating and Calibrating Probabilistic Quantitative Precipitation Forecasts from the High-Resolution NWP Model COSMO-DE. Weather Forecast. 27, 988–1002. doi:10.1175/WAF-D-11-00101.1

Bernardet, L.R., Grasso, L.D., Nachamkin, J.E., Finley, C. a., Cotton, W.R., 2000. Simulating convective events using a high-resolution mesoscale model. J. Geophys. Res. 105, 14963. doi:10.1029/2000JD900100

Bjerg, J.E., Grum, M., Courdent, V., Halvgaard, R., Vezzaro, L., Mikkelsen, 2015. Coupling of Weather Forecasts and Smart Grid-Control of Wastewater inlet to Kolding WWTP (

Denmark), in: 10th International Urban Drainage Modelling Conference. Mont Sainte-Anne, Québec, Canada, pp. 47–59.

Carstensen, J., Marinus K. Nielsen, Strandbæk, H., 1998. Prediction of hydraulic load for urban storm control of a municipal WWT plant. Wat. Sci. Tech. 37, 363–370. doi:10.1016/S0273-1223(98)00355-2

Cloke, H.L., Pappenberger, F., 2009. Ensemble flood forecasting: A review. J. Hydrol. 375, 613–626. doi:10.1016/j.jhydrol.2009.06.005

Collischonn, W., Morelli Tucci, C.E., Clarke, R.T., Chou, S.C., Guilhon, L.G., Cataldi, M., Allasia, D., 2007. Medium-range reservoir inflow predictions based on quantitative precipitation forecasts. J. Hydrol. 344, 112–122. doi:10.1016/j.jhydrol.2007.06.025

Courdent, V., Vezzaro, L., Mikkelsen, P.S., Mollerup, A.L., Grum, M., 2015. Using ensemble weather forecast in a risk based real time optimization of urban drainage systems. La Houille Blanche 101–107. doi:10.1051/lhb/20150025

Damrath, U., Doms, G., Frühwald, D., Heise, E., Richter, B., Steppeler, J., 2000. Operational quantitative precipitation forecasting at the German Weather Service. J. Hydrol. 239, 260–285. doi:10.1016/S0022-1694(00)00353-X

Demeritt, D., Nobert, S., Cloke, H.L., Pappenberger, F., 2013. The European Flood Alert System and the communication, perception, and use of ensemble predictions for operational flood risk management. Hydrol. Process. 27, 147–157. doi:10.1002/hyp.9419

Du, J., 2007. Uncertainty and Ensemble Forecast. NOAA's National Weather Service Office of Science and Technology.

Ebert, E., McBride, J.., 2000. Verification of precipitation in weather systems: determination of systematic errors. J. Hydrol. 239, 179–202. doi:10.1016/S0022-1694(00)00343-7

Ebert, E.E., 2009. Neighborhood Verification: A Strategy for Rewarding Close Forecasts. Weather Forecast. 24, 1498–1510. doi:10.1175/2009WAF2222251.1

Ebert, E.E., 2008. Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. Meteorol. Appl. 15, 51–64. doi:10.1002/met

Feddersen, H., 2009. A Short-Range Limited Area Ensemble Prediction System. Copenhagen.

Giebel, G., Badger, J., Landberg, L., Nielsen, H.A., Nielsen, T.S., Madsen, H., Sattler, K., Feddersen, H., Vedel, H., Tøfting, J., Kruse, L., Voulund, L., 2005. Wind power prediction using ensembles, Report Risø. Roskilde.

Gilleland, E., Ahijevych, D., Brown, B.G., Casati, B., Ebert, E.E., 2009. Intercomparison of Spatial Forecast Verification Methods. Weather Forecast. 24, 1416–1430. doi:10.1175/2009WAF2222269.1

Gneiting, T., Raftery, A.E., 2005. Weather Forecasting with Ensemble Methods. Science. 310, 248–249. doi:10.1126/science.1115255

Johnson, A., Wang, X., 2012. Verification and Calibration of Neighborhood and Object-Based Probabilistic Precipitation Forecasts from a Multimodel Convection-Allowing Ensemble. Mon. Weather Rev. 140, 3054–3077. doi:10.1175/MWR-D-11-00356.1

Johnson, D., Smith, M., Koren, V., Finnerty, B., 1999. Comparing mean areal precipitation estimates from NEXRAD and rain gauge networks. J. Hydrol. Eng. 4, 117–124. doi:10.1061/(ASCE)1084-0699(1999)4:2(117)

Jørgensen, H.K., Rosenorn, S., Madsen, H., Mikkelsen, P.S., 1998. Quality control of rain data used for urban runoff systems, in: Water Science and Technology. pp. 113–120. doi:10.1016/S0273-1223(98)00323-0

Kleiber, W., Raftery, A.E., Gneiting, T., 2011. Geostatistical Model Averaging for Locally Calibrated Probabilistic Quantitative Precipitation Forecasting. J. Am. Stat. Assoc. 106, 1291–1303. doi:10.1198/jasa.2011.ap10433

Langergraber, G., Alex, J., Weissenbacher, N., Woerner, D., Ahnert, M., Frehmann, T., Halft, N., Hobus, L., Plattes, M., Spering, V., Winkler, S., 2008. Generation of diurnal variation for influent data for dynamic simulation. Water Sci. Technol. 57, 1483–1486. doi:10.2166/wst.2008.228

Liguori, S., Rico-Ramirez, M. a., Schellart, A.N. a., Saul, A.J., 2012. Using probabilistic radar rainfall nowcasts and NWP forecasts for flow prediction in urban catchments. Atmos. Res. 103, 80–95. doi:10.1016/j.atmosres.2011.05.004

Mahura, A., Sattler, K., Petersen, C., Amstrup, B., Baklanov, A., 2006. Technical Report 05-12 DMI-HIRLAM Modelling with High Resolution Setup and Simulations for Areas of Denmark.

Mass, C.F., Ovens, D., Westrick, K., Colle, B. a., 2002. Does increasing horizontal resolution produce more skillful forecasts? The results of two years of real-time numerical weather prediction over the Pacific Northwest. Bull. Am. Meteorol. Soc. 83, 407–430+341. doi:10.1175/1520-0477(2002)083<0407:DIHRPM>2.3.CO;2

Michaelides, S., 2008. Chapter 16: Verification of quantitative precipitation forecasts. In: Precipitation: Advances in Measurement, Estimation and Prediction. Berlin: Springer-Verlag, pp. 419-452.

Mittermaier, M.P., 2007. Improving short-range high-resolution model precipitation forecast skill using time-lagged ensembles. Q. J. R. … 133, 1487–1500. doi:10.1002/qj

Nash, S.E., 1957. The Form of the Instantaneous Unit Hydrograph. IASH Publ. 114–121.

Pappenberger, F., Scipal, K., Buizza, R., 2008. Hydrological aspect of meteorological verification. Atmos. Sci. Lett. 9, 43–52. doi 10.1002/asl.171

Rouault, P., Schroeder, K., Pawlowsky-Reusing, E., Reimer, E., 2008. Consideration of online rainfall measurement and nowcasting for RTC of the combined sewage system. Water Sci. Technol. 57, 1799–1804. doi:10.2166/wst.2008.289

Schaffer, C.J., Gallus, W. a., Segal, M., 2011. Improving Probabilistic Ensemble Forecasts of Convection through the Application of QPF–POP Relationships. Weather Forecast. 26, 319–336. doi:10.1175/2011WAF2222447.1

Scharling, M., Kern-hansen, C., 2012. Climate Grid Denmark - Dateset for use in research and education, DMI Technical Report. Copenhagen.

Scheuerer, M., 2014. Probabilistic quantitative precipitation forecasting using Ensemble Model Output Statistics. Q. J. R. Meteorol. Soc. 140, 1086–1096. doi:10.1002/qj.2183

Shrestha, D.L., Robertson, D.E., Wang, Q.J., Pagano, T.C., Hapuarachchi, H. a P., 2013. Evaluation of numerical weather prediction model precipitation forecasts for short-term

streamflow forecasting purpose. Hydrol. Earth Syst. Sci. 17, 1913–1931. doi:10.5194/hess-17-1913-2013

Sloughter, J.M.L., Raftery, A.E., Gneiting, T., Fraley, C., 2007. Probabilistic Quantitative Precipitation Forecasting Using Bayesian Model Averaging. Mon. Weather Rev. 135, 3209–3220. doi:10.1175/MWR3441.1

Sun, J., Xue, M., Wilson, J.W., Zawadzki, I., Ballard, S.P., Onvlee-Hooimeyer, J., Joe, P., Barker, D.M., Li, P.W., Golding, B., Xu, M., Pinto, J., 2014. Use of nwp for nowcasting convective precipitation: Recent progress and challenges. Bull. Am. Meteorol. Soc. 95, 409–426. doi:10.1175/BAMS-D-11-00263.1

Talebizadeh, M., Belia, E., Vanrolleghem, P.A., 2016. Influent generator for probabilistic modeling of nutrient removal wastewater treatment plants. Environ. Model. Softw. 77, 32–49. doi:10.1016/j.envsoft.2015.11.005

Theis, S.E., Hense, a, Damrath, U., 2005. Probabilistic precipitation forecasts from a deterministic model: a pragmatic approach. Meteorol. Appl. 12, 257. doi:10.1017/S1350482705001763

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 modelbased flow prediction in urban drainage systems. Water Sci. Technol. 68, 472–478. doi:10.2166/wst.2013.274

Thorndahl, S., Rasmussen, M.R., 2013. Short-term forecasting of urban storm water runoff in real-time using extrapolated radar rainfall data. J. Hydroinformatics 15, 897–912. doi:10.2166/hydro.2013.161

Thorndahl, S., Rasmussen, M.R., 2012. Marine X-band weather radar data calibration. Atmos. Res. 103, 33–44. doi:10.1016/j.atmosres.2011.04.023

Unden, P., Rontu, L., Jarvinen, H., Lynch, P., Calvo, J., Cats, G., Cuxart, J., Eerola, K., Fortelius, C., Garcia-moya, J.A., Jones, C., Lenderlink, G., Mcdonald, A., Mcgrath, R., Navascues, B., 2002. HIRLAM-5 Scientific Documentation. Norrkoping.

Vezzaro, L., Grum, M., 2014. A generalised Dynamic Overflow Risk Assessment (DORA) for Real Time Control of urban drainage systems. J. Hydrol. 515, 292–303. doi:10.1016/j.jhydrol.2014.05.019

Walser, A., Lüthi, D., Schär, C., 2004. Predictability of Precipitation in a Cloud-Resolving Model. Mon. Weather Rev. 132, 560–577. doi:10.1175/1520-0493(2004)132<0560:POPIAC>2.0.CO;2

Wilks, D. S., 2011. Chapter 7.6 - Ensemble Forecasting. In: I. Geophysics, ed. Statistical Methods in Atmospheric Science. s.l.:Elsevier, pp. 267-292.

Woldemeskel, F.M., Sharma, A., Sivakumar, B., Mehrotra, R., 2012. An error estimation method for precipitation and temperature projections for future climates. J. Geophys. Res. Atmos. 117, 1–13. doi:10.1029/2012JD018062

WWRP/WGNE, 2009. Recommendations for the Verification and Intercomparison of QPFs and PQPFs from Operational NWP Models, World Meteorological Organization, WMO/TD - No. 1485

Caption Figures

Figure 1: Geographical domains of the DMI-HIRLAM T15 and S05 models (Feddersen, 2009).

Figure 2: Spatial distribution and intensity of forecasted total precipitation [mm/h] in each of the 25 ensemble members of the HIRLAM-S05 NWP EPS generated the 31st-August 2014 at midnight with a lead time of 18 hours. The area covered is a 80 km square with 0.05˚ grid cells (North Zealand). An animation is available in the supplementary material.

Figure 3: The Damhuså urban drainage catchment, Copenhagen, Denmark (contributing area, green area on the map, 67 km2).

Figure 4: Overall scheme of the conceptual hydrological model.

Figure 5: Predicted rainfall intensity (top) and flow (bottom) through the 48 hours forecast horizon. The blue plumes show different quantiles and the red dots show measurements.

Figure 6: Relation operating characteristic (ROC) diagram with examples of random prediction as well as predictions with some, better and perfect skill.

Figure 7: Damhuså catchment overlap with DMI-HIRLAM-S05 grid.

Figure 8: Illustration of the grid cells used for the smallest expansion (dark grey) and the largest expansion (light grey) in the neighbourhood ensemble expansion approach.

Figure 9: Variation of prediction skill with forecast lead time for a flow threshold of 4,000 m^3/h. Left: POD and POFD for different fractions of ensemble members (iEM) predicting an event using the weighted aerial overlap method. Right: POD and POFD for the weighted aerial overlap method and the Maximal Threat EPS method with different radii based on the same fraction of ensemble members (4EM).

Figure 10: ROC diagram for the different NWP post-processing strategies.

Figure 11: ROC diagram, constructed with different NWP post-processing, for different flow thresholds.

Caption Tables

Table 1: Configuration of DMI-EPS ensemble members 1-25 (Feddersen 2009).

Table 2: Performance scores for the flow prediction ensemble spread and median quantile.

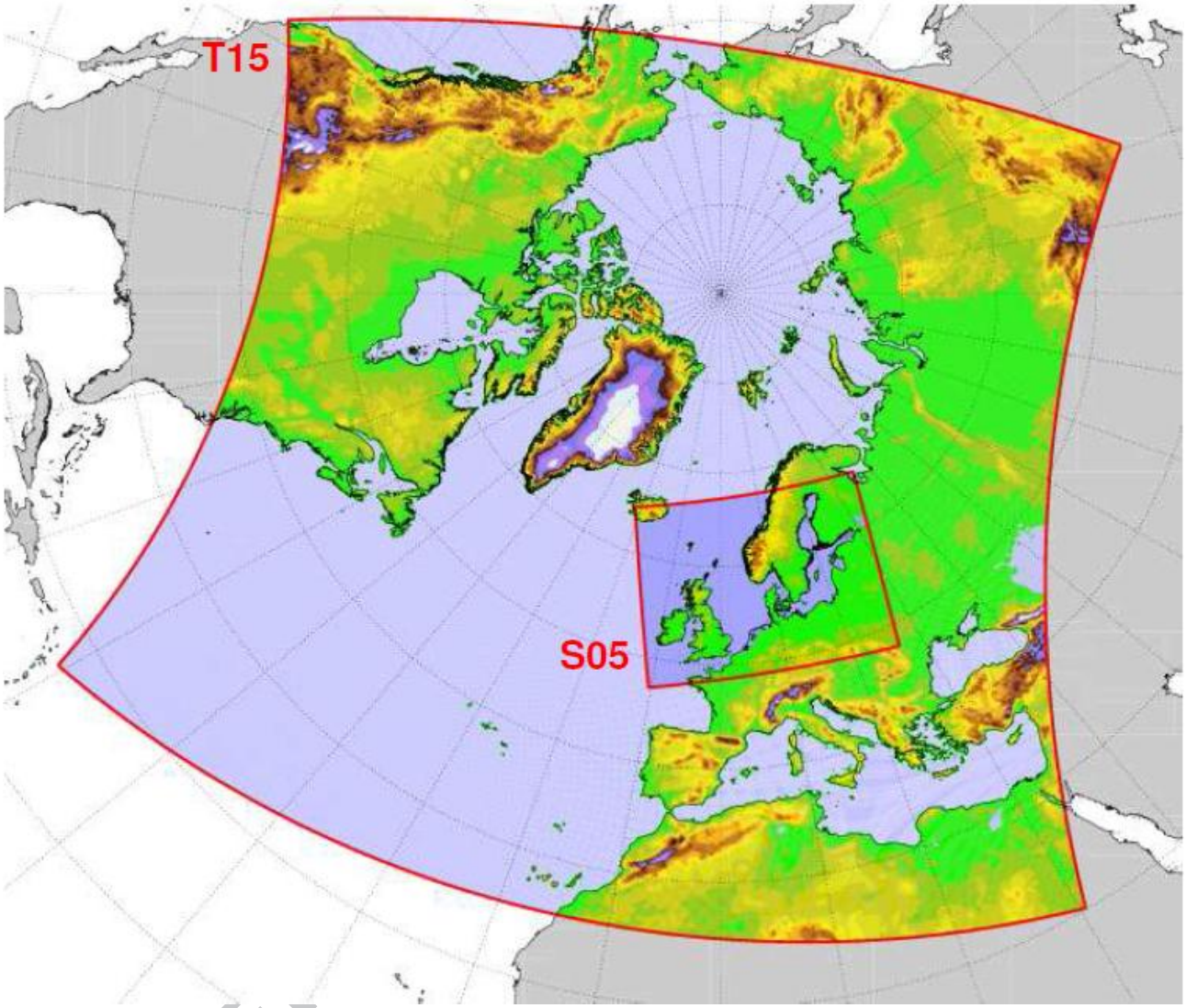
Table 3: Contingency table of binary events for categorical verification scores.

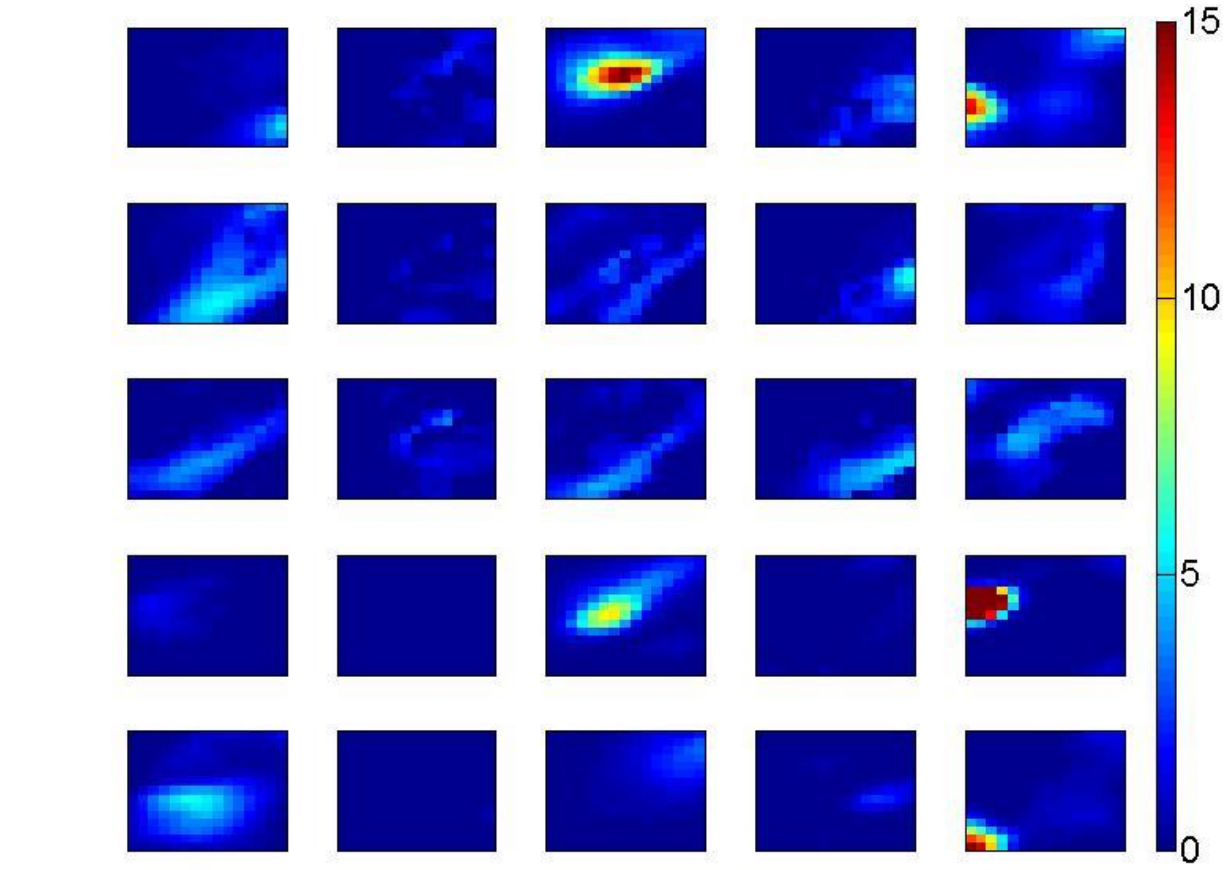
Table 4: POD and POFD verification skill scores for deterministic forecasts (*nForecast* represents the number forecast generations and *nTimeStep* represents the number of time steps of one forecast).

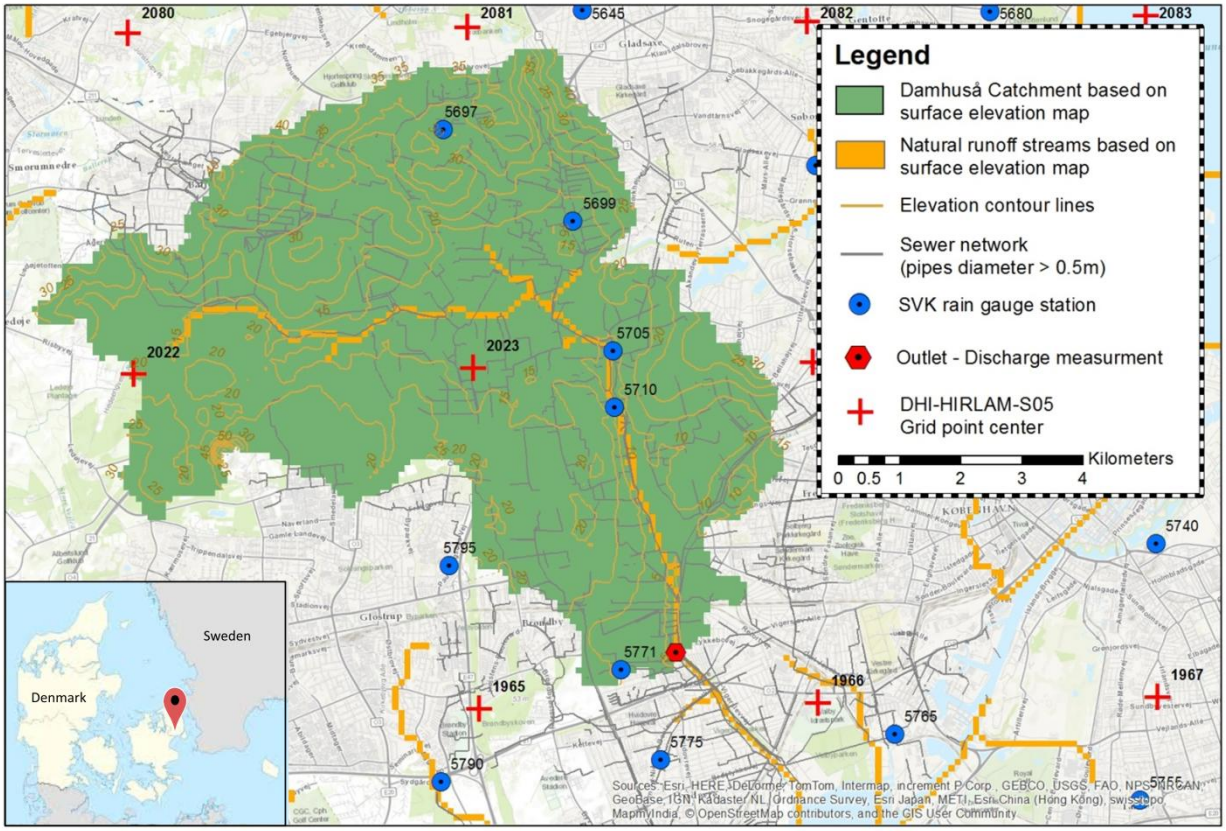
Table 5: Variation of EPS size with neighbourhood radius in grid cell (the spatial resolution is 0.05˚ i.e. approx. 5.6 km).

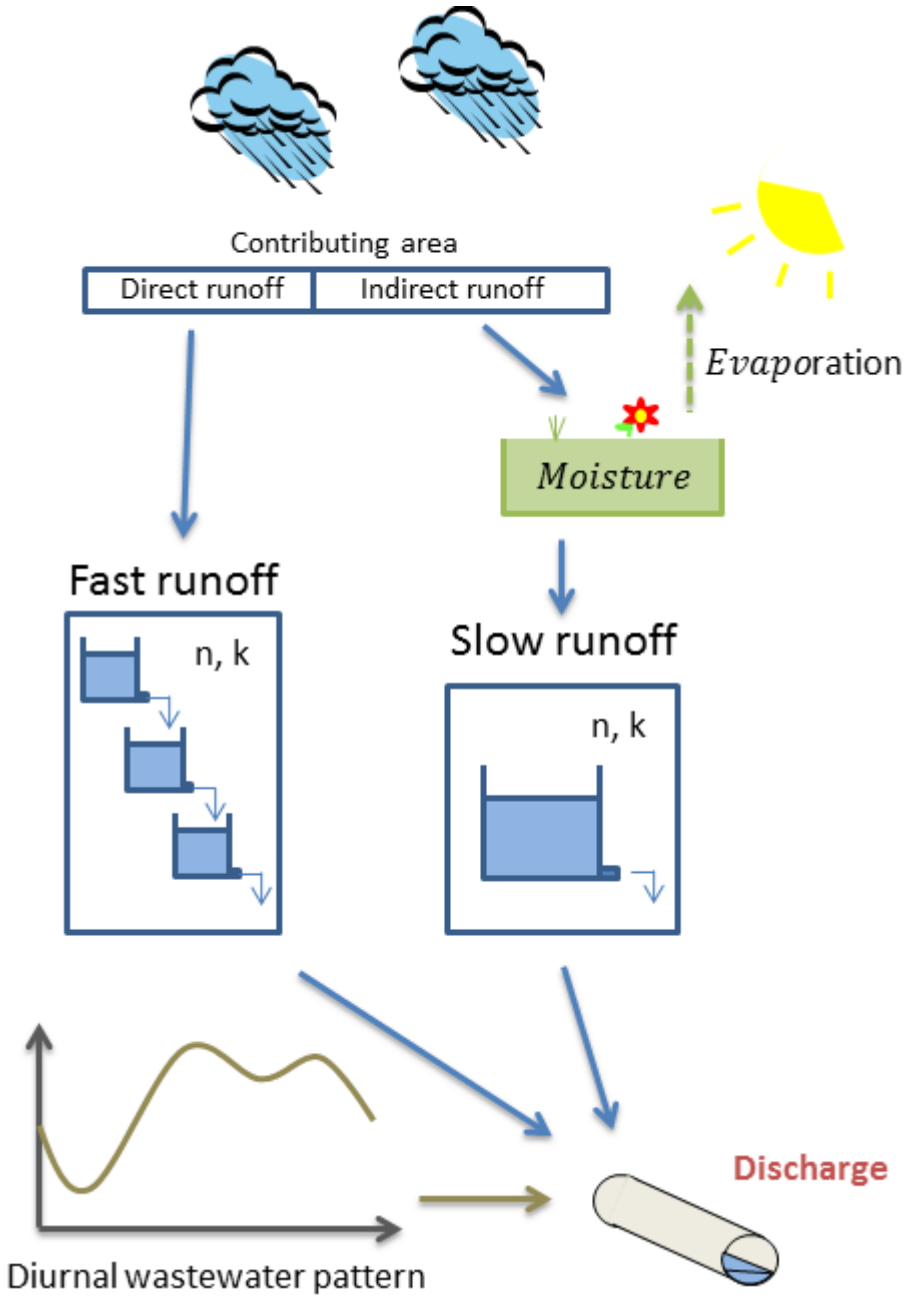
Table 6: Skills of the flow domain distinction including the entire forecast horizon for the different seasons using rain gauge measurements.

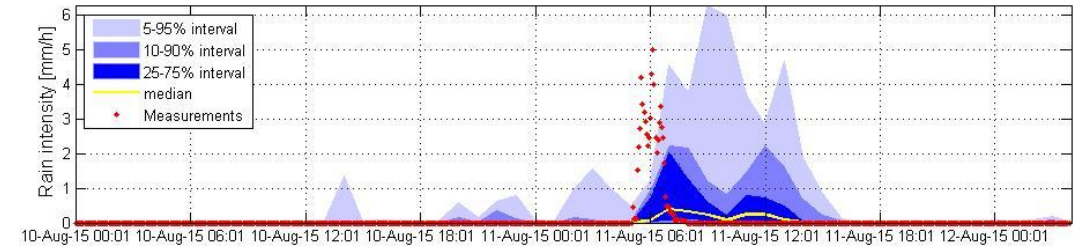
Table 7: Contingency table and skill scores for different fractions of ensemble members predicting an event using the weighted aerial overlap method and a flow threshold of 4,000 m^3/h. All time steps within the forecast horizon are included

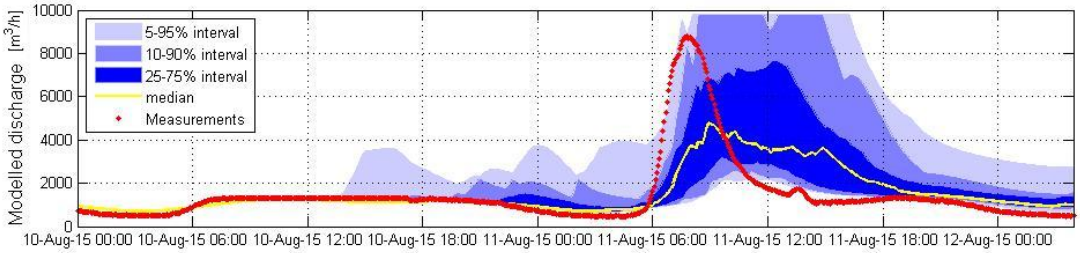


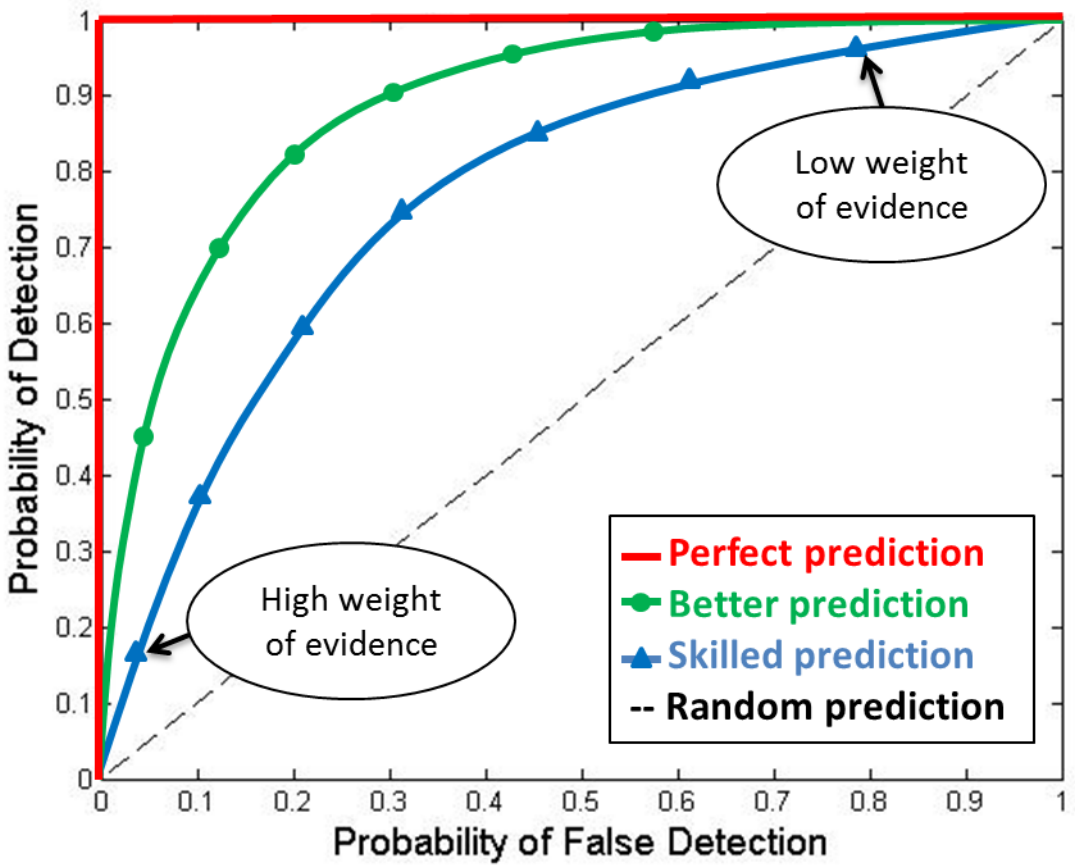


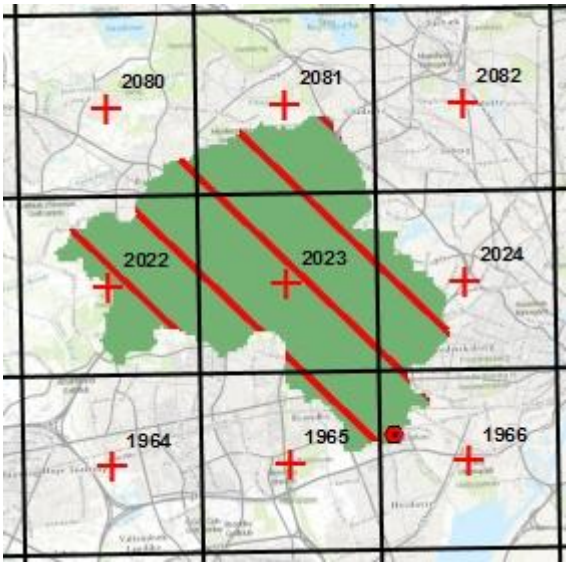


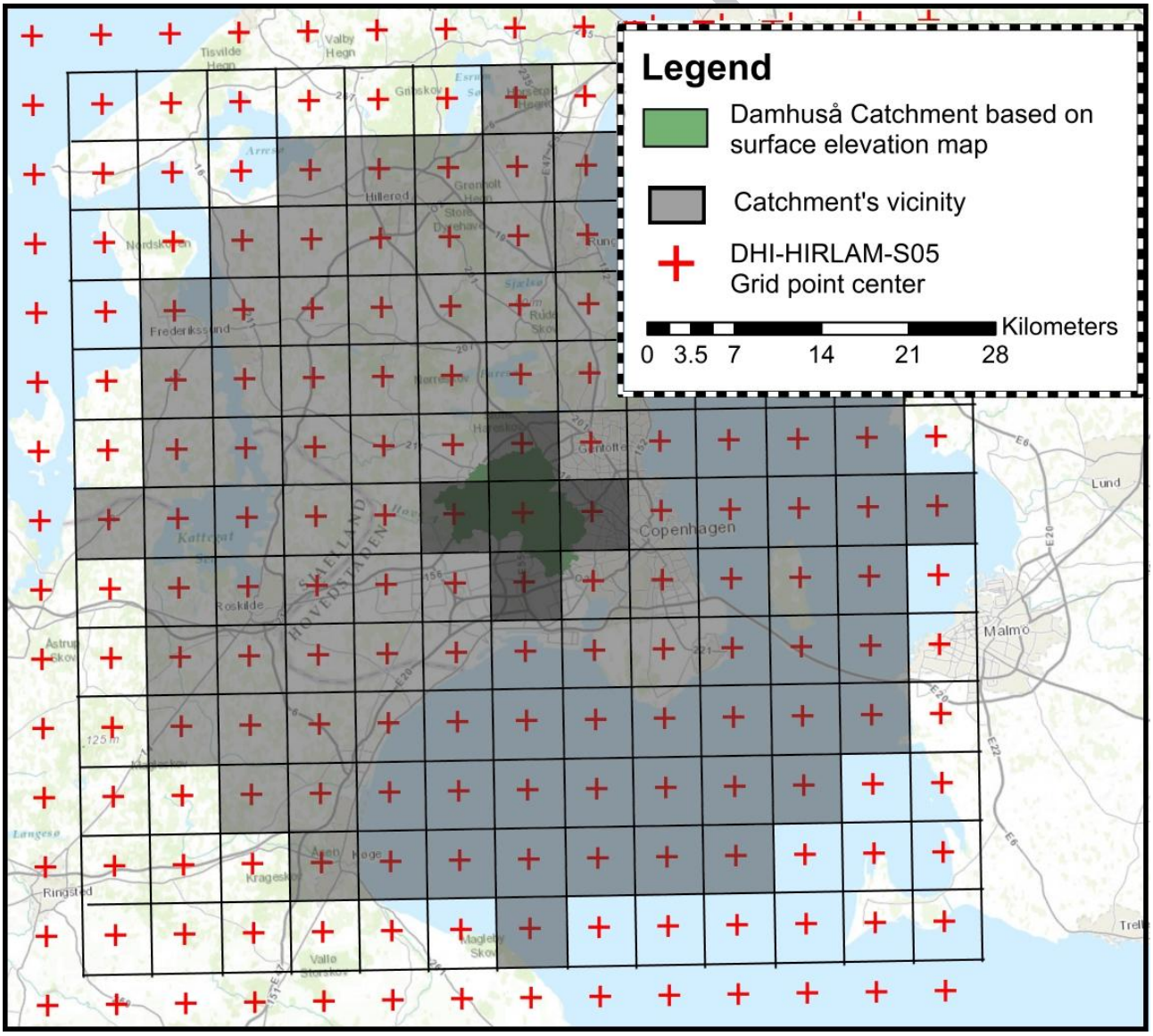


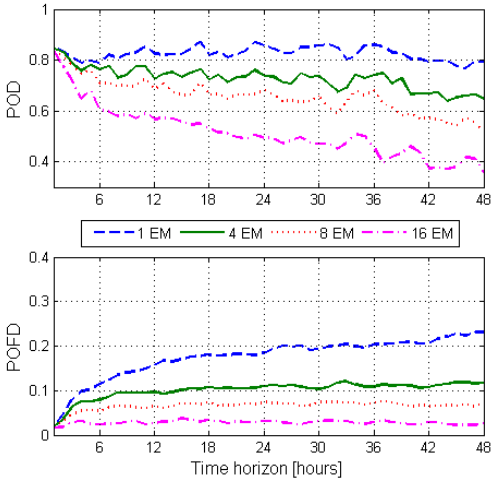

Rain intensity [mm/h]
5-95% interval
10-90% interval
25-75% interval
median
Measurements
10-Aug-15 00:01
10-Aug-15 06:01
10-Aug-15 12:01
10-Aug-15 18:01
11-Aug-15 00:01
11-Aug-15 06:01
11-Aug-15 12:01
11-Aug-15 18:01
12-Aug-15 00:01

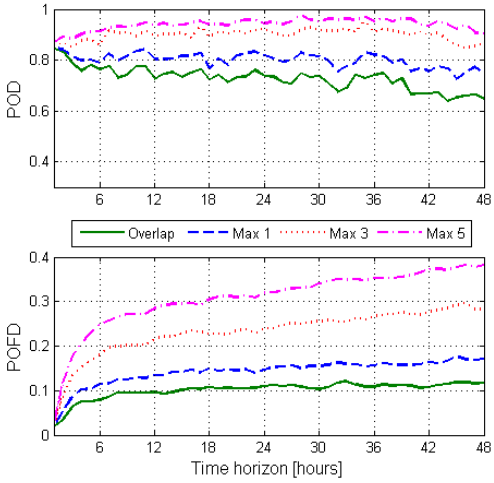

Modelled discharge [m3/h]
5-95% interval
10-90% interval
25-75% interval
median
Measurements
10-Aug-15 00:00
10-Aug-15 06:00
10-Aug-15 12:00
10-Aug-15 18:00
11-Aug-15 00:00
11-Aug-15 06:00
11-Aug-15 12:00
11-Aug-15 18:00
12-Aug-15 00:00

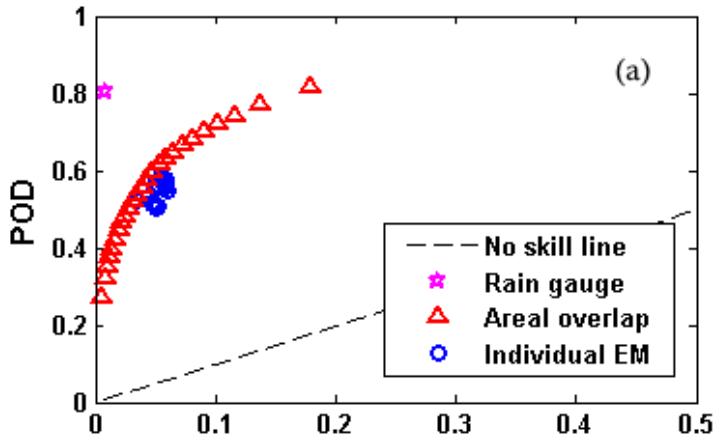

Probability of Detection
Probability of False Detection
Low weight of evidence
High weight of evidence
Perfect prediction
Better prediction
Skilled prediction
-- Random prediction

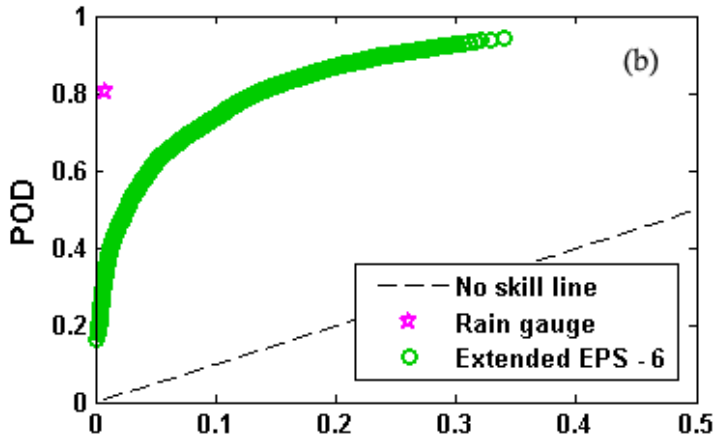


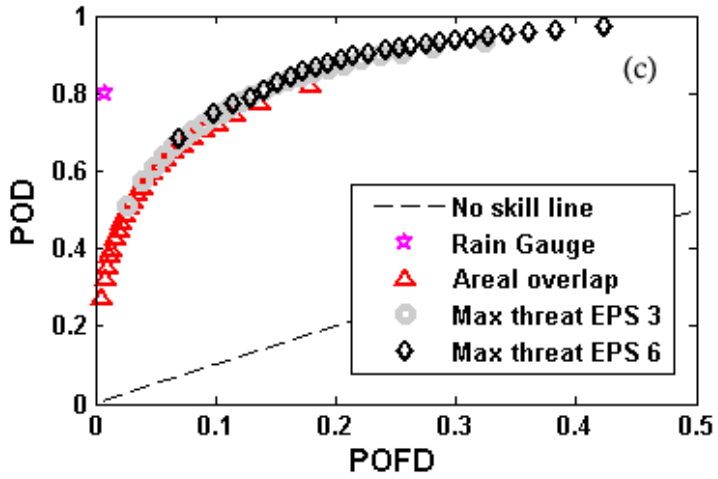


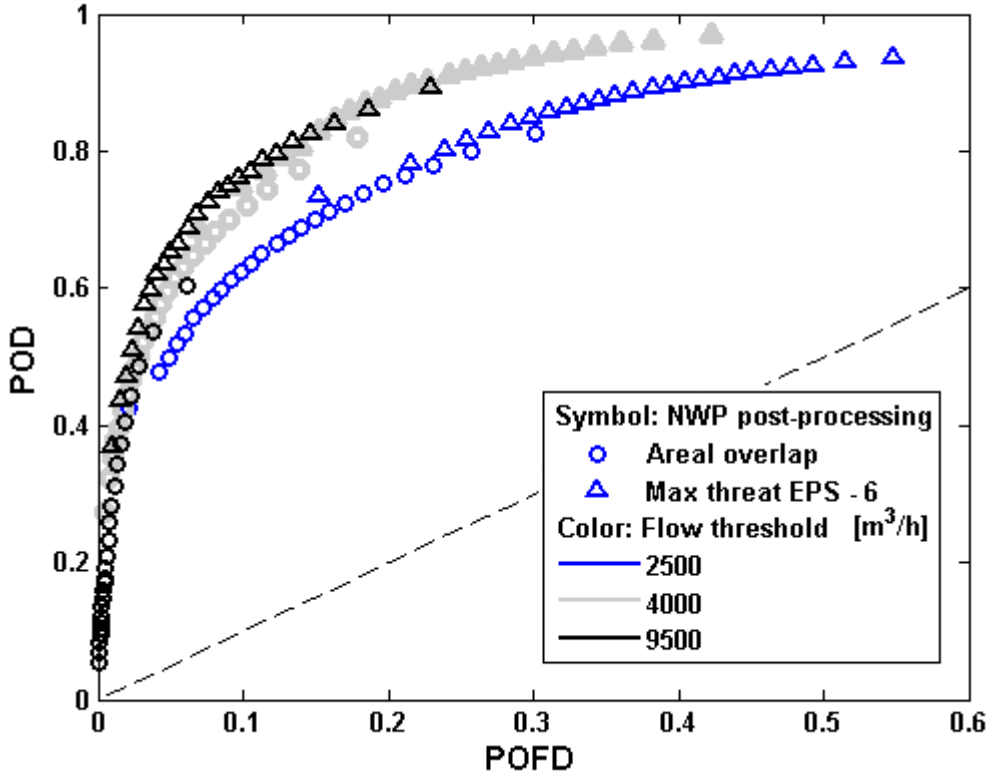












Highlight

- Weather models provide uncertain but valuable urban scale discharge predictions
- Flow threshold exceedance can be predicted with some confidence two days ahead
- Ensemble predictions allow computing the relative operating characteristic (ROC)
- Spatial misplacement errors may be compensated through ensemble post-processing
- The optimal prediction skill can be selected based on the prediction purpose