Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com

Research papers

An ensemble neural network model for real-time prediction of urban floods

Simon Berkhahn [a,*], Lothar Fuchs [b], Insa Neuweiler [a]

[a] Institute of Fluid Mechanics and Environmental Physics in Civil Engineering, Leibniz Universität Hannover, Appelstrasse 9a, 30167 Hannover, Germany
[b] Institute for Technical and Scientific Hydrology (itwh GmbH), Engelbosteler Damm 22, 30167 Hannover, Germany

ARTICLE INFO

Keywords:
Artificial neural network
Ensemble neural network
Real-time forecast
Urban flooding

ABSTRACT

The real-time forecasting of urban flooding is a challenging task for the following two reasons: (1) urban flooding is often characterized by short lead times, (2) the uncertainty in precipitation forecasting is usually high. Standard physically based numerical models are often too slow for the use in real-time forecasting systems. Data driven models have small computational costs and fast computation times and may be useful to overcome this problem. The present study presents an artificial neural network based model for the prediction of maximum water levels during a flash flood event. The challenge of finding a suitable structure for the neural network was solved with a new growing algorithm. The model is successfully tested for spatially uniformly distributed synthetic rain events in two real but slightly modified urban catchments with different surface slopes. The computation time of the model in the order of seconds and the accuracy of the results are convincing, which suggest that the method may be useful for real-time forecasts.

1. Introduction

Urbanization may have a significant impact on the rainfall runoff relation (Ferguson and Suckling, 1990; Aronica et al., 2012). Related decrease of pervious surface area may increase urban floods. Mainly short term precipitation events with high intensity can lead to an exceedance of the capacity of drainage systems and flooding of the surface. With climate change, those events will occur more often in the future and the monetary and the non-monetary damage may increase significantly (Schreider et al., 2000; Dewan, 2013; Hölscher et al., 2014;36.). For the implementation of early warning systems, fast forecast models are indispensable (Henonin et al., 2013). Such models include the rain forecasting and the prediction of inundated areas.

In Henonin et al. (2013), several different approaches for water level prediction in real-time are discussed. The main outcome of that study is that accurate physically based numerical models are too slow for real time forecasting. Simplified physically based models (e.g. reduced in spatial dimension) are used to overcome the problem of computation time. Leandro et al. (2009) compared the coupling of a 1D sewer system with a 1D and a 2D surface flow model. This study shows that results of the 1D surface model were in good agreement with results of the 2D model. In that study, overland flow was restricted to the streets. Jahanbazi and Egger (2014) proposed to combine 1D and 2D surface flow models by defining regions prone to surface flooding. Only these regions were calculated with a 2D surface flow model. Thorndahl et al. (2016) have shown in a case study in Denmark that the computation time of the physically based model Mike Flood (DHI, 2012) can be minimized such that computation time is half the real-time. Also, René et al. (2018) used a physically based overland flow model for flood forecasting in the city of Castries, St. Lucia.

In both studies by Thorndahl et al. (2016) and René et al. (2018), the spatial extent of the prediction model was moderate. As pointed out in (Bermúdez et al., 2018), for model domains of larger spatial extent, the computation time of physically based numerical models can be limiting. Also, ensemble methods might be needed to capture the influence of uncertainties. For real-time forecasting of flooding, very fast prediction methods are needed.

In order to reduce computation time and enable real-time forecasting, the use of data driven models has become popular in the last years. Bruen and Yang (2006) combined a physically based model and an artificial neural network (ANN) modeland compared model results with observations. Chang et al. (2014) presented an ANN based model for floodwater prediction in storage ponds. Tayfur et al. (2007) used an ANN to predict real-time storm hydrographs for the upper Tiber river basin in Italy. Rjeily et al. (2017) used non-linear autoregressive exogenous (NARX) neural networks to calculate water depth at critical manholes in an urban catchment. Input of that model were both, rainfall in-

* Corresponding author.
Email address: berkhahn@hydromech.uni-hannover.de (S. Berkhahn)

https://doi.org/10.1016/j.jhydrol.2019.05.066
Received 19 February 2019; Received in revised form 17 May 2019; Accepted 20 May 2019
Available online xxx
0022-1694/

tensities and water depth at the previous time step. Results of that study have shown that the NARX model performed well on both minor and severe storm events. In above listed studies, flooded areas were predicted point-wise and 2D results for inundated area were not generated. A two-dimensional real-time forecast model for urban flooding is described in Bermúdez et al. (2018). The model described in that study is based on two different ANNs. One ANN uses the rainfall-runoff volume determined by a conceptual sewer model as input to detect flooding from the rain sewer system. The second ANN model is based on an ensemble approach and simulates the maximum flood volume. Results from both ANNs enable to choose a discrete flood depth map, which is selected from a database of physically based pre-simulated scenarios. Interpolation between pre-simulated scenarios is not possible with the approach described in that study, Regarding the variability of rain events, there is a need for real-time prediction of flood depth with interpolation abilities. An approach to interpolate between different rainfall events is given in Bermúdez et al. (2019:). In this study a hydrological model is used to generate discharge as input for support vector machine (SVM) models at about 25000 control points in an urban catchment. The output of the SVM models is the maximum flood depth and velocity. Jhong et al. (2017) presented a real-time forecast model for inundation maps during typhoons in Taiwan. The model is also based on SVM and uses a two stage approach. In the first stage inundation depths at reference points are predicted and in the second stage the spatial expansion based on geographic information is estimated. To the authors' best knowledge, a real-time forecast model based on ANN which provides 2D flood maps with interpolation skills has not been tested so far.

In this paper we present an approach to real-time predict 2D maps of maximum water level during a pluvial flood event in a city. Input of the approach is the precipitation forecast in full length. The approach described in the present study is based on an ANN and uses an ensemble training approach in order to overcome the problem of overfitting. The topology of the network was chosen using a network growing algorithm. The model is tested in two test areas of different sizes and different surface slopes with synthetic rain events. All tested rain events where full-length events with maximum duration of two hours. The database for training and testing of the ANN model is generated with the 2D hydrodynamic model HYSTEM-EXTRAN 2D [HE 2D], (itwh, 2017). The ANN model is analysed by creating an artificial flood event with the physically based model (HE 2D) and reproducing the maximum water levels during the flood event with the ANN. The general structure of the model is given in Fig. 1.

2. Methods

2.1. Artificial neural network for prediction of maximum water depth

Artificial neural network (ANN) roots in the McCulloch-Pitts-Neuron presented in McCulloch and Pitts (1943) and the perceptron algorithm by Rosenblatt (1958) who trained the perceptron for binary classification. Zhang et al. (1998) have given an overview on how the method can be extended for forecasting purposes in many ways. In order to predict maximum flood height in the present study, multilayer perceptrons (MLP) are used. A MLP consists of an input layer, an output layer, and several hidden layers in between. Each layer consist of a certain number of neurons, which process incoming information. Neurons are connected via synapses, which are multiplied with weight factors. We use a feed forward network (Bebis and Georgiopoulos, 1994), in which each layer k receives input from each neuron of the previous layer $k-1$. This structure directs the information through the network from the input to the output without inner circles. The network can be seen as a function $\vec{y} = f(\vec{x})$, where $\vec{y}$ stands for the vector of output variables and $\vec{x}$ for the vector of input variables. In the present study, $\vec{y}$ represents maximum water levels in the different grid cells of the model domain and $\vec{x}$ contains precipitation intensities in five minutes time steps. The output o_j of a neuron j is defined as:

$$o_j = \varphi\left(\sum_{i=1}^{n} w_{ij} o_i\right) \tag{1}$$

where φ is the activation function (in this study a hyperbolic tangent sigmoid function) and n the number of precursors. The output o_i of the previous neuron i is multiplied with the weight w_{ij}.

The number of neurons and layers is called topology and has two boundary conditions: (1) the number of input neurons is defined by the number of time steps in the rain event (for a rain event of 2 h duration: 24), (2) the number of output neurons, defined by the number of maximum water levels. In the present study, the number of hidden layers and number of neurons per hidden layer are determined by a growing algorithm (see Section 2.3).

2.2. Metric for prediction evaluation

To evaluate the results of the prediction model, the database is divided into two sets, the training set and the test set. The test set is ex-

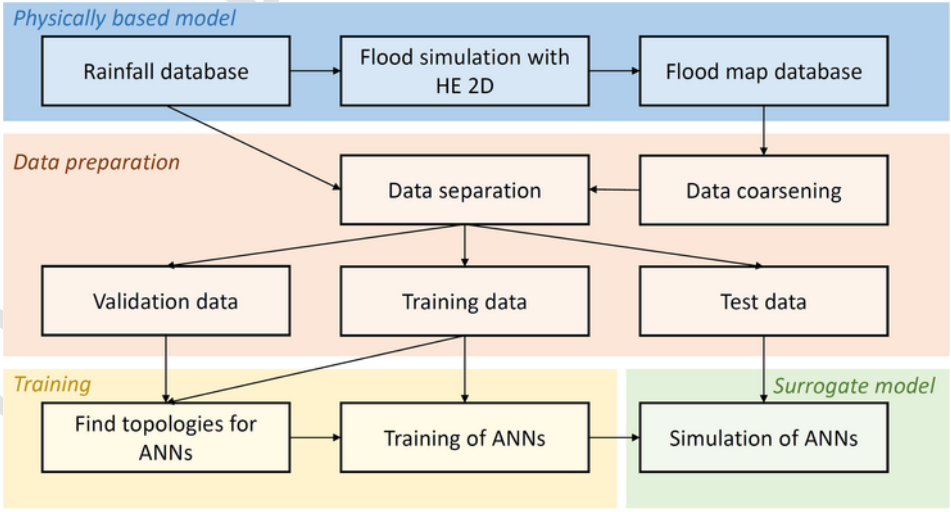

Fig. 1. General structure of the forecast model in a flowchart scheme (illustration according to Jhong et al. (2017)).

clusively used to evaluate the model. For the evaluation of the network, the root mean squared error (RMSE) and the Nash-Sutcliffe efficiency coefficient (NSE) of the predicted and measured water levels are used. In the present study the RMSE is defined as:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(wl_{pred,i} - wl_{\exp,i}\right)^2}{n}} \tag{2}$$

where i indicates the location of the water level and n the maximum number of cells that the urban area is divided into, where we restrict the analysis to cells for that is flooding predicted by the physically based model. The wl_{pred} is the predicted and the wl_{exp} the expected water level, which is in this paper taken from a physically based urban flow model. The RMSE has the unit of a length. To evaluate the performance of the prediction model for all rain events in the test set, the RMSE is evaluated for each test set sample and averaged over the domain. This performance value will be denoted as $\overline{\mathrm{RMSE}}$. The RMSE as given in Eq. (2) gives averaged values over the whole catchment. The spatial distribution of the model performance was evaluated by using the dimensionless NSE. The NSE for one location is defined as:

$$\mathrm{NSE} = 1 - \frac{\sum_{j=1}^{m}\left(wl_{pred,j} - wl_{\exp,j}\right)^2}{\sum_{j=1}^{m}\left(wl_{\exp,j} - \overline{wl}_{\exp}\right)^2} \tag{3}$$

where j indicates rain events and m is the maximum number of rain events. The mean of all maximum water levels from the pre-calculated scenarios for the specific location is written as $\overline{wl}_{\exp}$. By evaluating Eq. (3) for each location, a map of NSE can be shown. A location with a NSE smaller than zero is better represented by the mean of all pre-calculated water levels than by the forecast model.

2.3. Training algorithm

The training of an ANN can be done with different methods (Kröse et al., 1993). In this study, supervised learning was used for the training. Supervised learning is based on samples of input and output data. The ANN can be optimised by comparing the simulated output with the expected output followed by a subsequent modification of the ANN. In the present study, the ANN modification was realized by changing the topology and changing the weights. The ANN is adapted iteratively to minimize the difference between expectation and simulation output data. Three different training algorithms for the adaptation of the weights were compared, namely the backpropagation method (Rumelhart et al., 1986), a Levenberg-Marquardt algorithm (Moré, 1978), and a resilient backpropagation algorithm Rprop (Riedmiller and Braun, 1993). The method was tested with flood prediction runs on a sub-area of one of the test scenarios (scenario 1), which are introduced in Section 3.2. In order to compare the different methods, the training was not done for best prediction results but as a standardized test, here called test-training. As the topology for the ANN is not known during the test-training, an ANN with 2 hidden layers and 20 neurons per hidden layer was chosen. The test-training was carried out until an absolute error of $2{\cdot}10^{-5}$ was reached for the normalized output values. The number of iteration steps and the computation time for the test-training are shown in Table 1. All simulations were done on the same desktop PC (CPU: i7 3.6 GHz, RAM: 32 GB). Rprop turned out to be the best algorithm for the given problem with respect to the shortest computation time.

Table 1
Test-training with three different algorithms.

Algorithm	Iterations	Computation time
Backpropagation	120183	124 min
Levenberg–Marquardt	14	815 s
Rprop	103	2 s

Besides the adaptation of the weights, a suitable topology has to be found. Different approaches for topology adaptation can be found in the literature, although it should be noted that there is not an established best practice how to determine an appropriate topology for an ANN. Often, the topology is chosen by trial and error, e.g. in Tayfur et al. (2007). Reed (1993) gives a survey of pruning algorithms. These algorithms aim at finding a suitable topology by iteratively pruning a network which is more complex than necessary. A different group of approaches are so called growing algorithms. As described in Bishop (1995), the network topology is chosen by starting with a small network and adding neurons and layers during the training. In the present study a newly developed simplified growing algorithm is used to find a suitable topology. The approach is shown as pseudo code in Algorithm 1. In general, the number of neurons in a layer is increased as long as this improves the prediction error. Starting with one layer, new layers are added and the maximum number of neurons is determined again, as long as the prediction error is improved. The topology of the network is chosen as the resulting number of layers with the respective number of neurons obtained.

The presented growing algorithm uses four parameters to control the topology search, which has to be chosen by the user. (1) The n_{stop} value is the maximum number of times a neuron is added to a layer without improving the result. It is introduced to avoid the stopping of adding neurons due to oscillations. (2) The $threshold_N$ (ranges from 0 to 1) is the reduction of the prediction error that is required to add a neuron to the layer. It is used to guarantee a benefit from adding a neuron to a layer. (3) The $threshold_L$ (ranges from 0 to 1) is the reduction of the prediction error needed for adding a new layer. It is used to guarantee a benefit from adding a new layer to the topology. (4) The search distance s_d (natural number), which was implemented to start with a number of neurons in a new layer that is not too far away from the maximum number of the previous layer and therefore reducing the computation time.

The hidden size vector $\vec{h}$ defines the structure of the hidden layers. For example $\vec{h} = (8, 5)$ leads to an ANN with 8 neurons in the first hidden layer and 5 in the second. The $threshold_N$ was chosen to 0.95 and the $threshold_L$ to 0.90 in the present study. The n_{stop} value was set to 5. The search distance s_d was set to 5. As *error* in Algorithm 1 the RMSE is used as described in Eq. (2).

Algorithm 1 Growing algorithm for finding a suitable topology for an ANN

initialize hidden size vector $\vec{h}$ where h_i is number of neurons in layer i
$i = 0$
$h_i = 1$
training of ANN with $\vec{h}$
determine *error* and save as $error_{new}$
initialize $+\infty \gg error_{old} \gg error_{new}$
initialize counter $n = 0$
$addlayer = true$
while $addlayer$ **do**
  $error_{temp} = error_{old}$
  $\vec{h}_{temp} = \vec{h}$
  while $error_{new}/error_{temp} \leq threshold_N$ **or** $n < n_{stop}$ **do**
    if $error_{new} \geq error_{temp}$ **then**
      $n = n + 1$
    else
      $error_{temp} = error_{new}$
      $n = 0$
      $\vec{h} = \vec{h}_{temp}$
    end if
    $h_{temp,i} = h_{temp,i} + 1$
    training of ANN with $\vec{h}_{temp}$
    determine *error* and save as $error_{new}$
  end while
  if $error_{new}/error_{old} \leq threshold_L$ **then**
    $i = i + 1$

```
        h⃗_best = h⃗
        h_i = max(h_{i−1} − s_d, 0)
        error_old = error_new
    else
        addlayer = false
    end if
end while
return h⃗_best
```

The oscillation of ANN simulations between the training data due to high parameterization is called overfitting. To avoid overfitting of the network, different strategies can be found in the literature. One method is to use a sub-set of the training data for validating the training progress. That method was used in the present study by evaluating the error with respect to a validation set during the topology growing. In addition, an ensemble-approach (Hansen and Salamon, 1990) was tested in combination with the growing algorithm.

2.4. Ensemble of ANNs

The ensemble-approach uses a positive number n_{ens} of ANNs with identical topology trained with different initial weights. In this way, it captures the uncertainty due to the random initialization. The result of the maximum water level wl_{ens}, generated with the ANN ensemble is the ensemble mean, which reads

$$wl_{ens} = \frac{1}{n_{ens}} \sum_{i=0}^{n_{ens}} wl_i \tag{4}$$

where wl_{ens} is the maximum water level predicted by the ensemble for one location and wl_i the maximum water level predicted by one of the realizations for the same location. To maintain reasonable results, all simulated water levels which are negative are set to zero. The optimal ensemble size depends on the catchment and has to be found by testing. Studies on the convergence are given in Section 4 for two different catchments.

2.5. Physically based numerical model

The purpose of the ANN is to provide a fast simplified model as substitute for a slow physically based numerical model. Therefore, we use flood simulations generated with a physically based numerical model as virtual truth. The model results are used for training of the ANN as well as to assess the ANN as a method for flood prediction. The physically based numerical model is a hydrodynamic coupled surface and sewer network model. The governing equations for the surface flow are the diffusive wave approximation of the shallow water equations. The flow in the sewer network is described by the 1D Saint Venant equations. A Preissmann slot (Preissmann, 1961) is used for the pipes under filled up conditions. The surface and the sewer system model are linked by manholes and street inlets. A description and discussion on the model is given in Jahanbazi and Egger (2014).

2.6. Resolution of the ANN model

Physically based numerical models for urban flood prediction need a high spatial resolution of the surface. This resolution is not necessary for data driven models like the ANN model in the present study, as the relation between neighbouring cells via conservation laws is not captured. In order to reduce the number of outputs, the irregular triangular mesh used in the physically based model (HE 2D) is reduced to a regular rectangular grid with the cell size of $dx \times dy$. For every triangular grid cell i of the HE 2D model that overlaps with the rectangular grid cell j an area related weight $a_{i,j}$ is defined as:

$$a_{i,j} = \frac{A_{i,j}}{dx \times dy} \tag{5}$$

where $A_{i,j}$ is the sub-area of the triangle i which lies inside the rectangle j. The maximum water levels which are used as output in the samples are calculated as:

$$wl_j = \sum_{i=1}^{n} a_{i,j} \times wl_i \tag{6}$$

where wl_j is the maximum water level of rectangular cell j and wl_i is the maximum water level of triangle cell i. All weights $a_{i,j}$ for a rectangular cell j are set to zero if j is filled less than a chosen threshold. This strategy can be formulated as:

$$a_{i,j} = \begin{cases} 0 & \text{for } \sum_{k=1}^{n} a_{k,j} < t \\ a_{i,j} & \text{for } \sum_{k=1}^{n} a_{k,j} \geqslant t \end{cases} \tag{7}$$

The k represents the counter for the triangles lying in the rectangle j. The threshold t can be used to represent building edges in a more proper way compared to the grid without a threshold. Depending on the HE 2D mesh, the number of calculation cells has been reduced by up to 90 percent.

Note that by averaging the maximum water level as described in Eq. (6), we use a coarse grained water level that yields a good representation of the flood volume, but not necessary for other measures of flooding, such as maximum water level or flooded area. Alternatively, one could use the maximum value instead of an average in order to get a good representation of the highest water level in the area. Such a measure would then lead to an overestimation of the flood volume. Also, a coarse grained water level could be formulated that gives a good representation of the flooded area, which would then again lead to a poor representation of the flood volume.

2.7. Zero-cells

Sajikumar and Thandaveswara (1999) described that ANN should be used for interpolation, but not for extrapolation purposes. Only cells which have the potential to be flooded, are of interest in an early warning system. To forecast flooding in a cell which is never flooded in the training dataset, the water level in that cell would have to be extrapolated, which is not a useful approach. Thus cells which have a maximum water level of zero in every sample of the training set are neglected. This constraint reduces the amount neurons in the output layer.

2.8. Grid of the ANN model

If one ANN is used to simulate an entire catchment, the output vector has the length of the number of output cells. Considering a medium urban catchment this would easily lead to a vector length of over one hundred thousand. Because each output neuron is connected to each hidden neuron in the previous layer, a huge number of weights has to be adapted during the training phase. In order to reduce the memory requirement and the computation time, the calculation area is divided into individual subnets. The subnets are here generated by dividing the domain into rectangular cells with a regular grid. The basis for the division was the maximum number of cells that can be handled by the PC that was used. It should be noted that other ways to divide the domain would be possible, such as a division into further sub-catchments, which could consist of few streets only. This is discussed later in Section 5. For the cases analysed in this paper, we do not consider the division of the domain into rectangles as critical, as no unphysical results were obtained. Every subnet has its own ANN and thereby a much smaller output vector. This means, instead of using one ANN model,

several smaller ANN model for different areas are set up. The subnets are independent, which implies that the correlation between cells with a short distance that belong to different subnets gets lost. For bigger catchments a strategy for sub-catchments delineation and coupling presented in Wolfs and Willems (2017) could be useful. However, the philosophy of a data driven model is that every correlation is contained in the data and is therefore captured. This means the interconnection of the subnets is included in the data.

3. Test cases

The developed ANN model for maximum water level prediction was tested with two synthetic test cases. The first test case is a model of a part of a city that has a flat area. The second is a model in a part of a city with steep slopes. Both urban catchments were modelled with HE 2D to generate test datasets and to assess predictions with the ANN model. From the results of the hydrodynamic model, only the maximum water level was considered, although more information would in principle be available.

Both test cases were derived from real catchments and were slightly modified to maintain anonymity. Because of missing measurements and the rarity of the considered flood events, it is not possible to calibrate and validate the physically based model. Nevertheless the plausibility was checked for the real catchment which is related to test case 1. For this purpose, pictures from the press and reports of the fire brigade were evaluated for one flood event which occurred during the summer of 2017. This comparison showed good results in the sense that areas that were documented as have been flooded were also flooded in the model. The modelling process with the physically based model is, however, not part of the present study and will thus not be discussed further. Uncertainties of the physically based model are not of interest in the present paper.

The mean cell size for test area 1 and 2 is 3.1 m^2 respectively 4 m^2 for the physically based model (more information on the triangular grid is given in Section 3.2). To find a proper value for the grid size and the threshold t (see Eq. (7)) the discretized areas of different grid sizes where investigated. Differences of the discretized area are due to the different resolution of buildings and streets. This investigation was done for test area 1. The rectangular grid size was set to $dx = dy$. To find the grid size three cases where tested, (1) $dx = 3$ m, (2) $dx = 6$ m and (3) $dx = 9$ m. The threshold value t was varied from zero to one in 0.25 steps. The results for the $dx = 6$ m grids are shown in Table 2. For all grid sizes the threshold $t = 0.5$ gives the smallest deviation in size of discretized area compared to the original grid. For the grid size with $dx = 3$ m the resulting resolution of the regular grid was much higher than the underlying triangular grid in some areas. Therefore the $dx = 3$ m grids were not considered further. With a $dx = 9$ m regular grid, complex structures of a urban catchment can barely be represented. The grid with $dx = 6$ m and $t = 0.5$ is a good compromise between computation time and complexity. This grid reduces the number of cells about 90 percent and has still a relatively small difference of the discretized area. To maintain comparability, the same values for *dx* and *t* are used for test area 2.

Table 2
Grid size analyses for test area 1.

dx [m]	threshold *t*	Number of cells	Discretized area [m^2]	Difference to HE2D grid [m^2]
HE2D		1193222	3652336	0
6	0.00	118206	4255416	603080
6	0.25	109333	3935988	283652
6	0.50	102119	3676284	23948
6	0.75	93880	3379680	−272656
6	1.00	40995	1475820	−2176516

3.1. Rain events

Synthetic rain events were used for testing the model. The rain events are spatially uniformly distributed over the test domain. For the distribution in time, according to DWA-A 118 (2006) Euler type I and II events were used (see Fig. 2). The rain events were mirrored to expand the sample size. Rain events with the duration of one and two hours were used. The rain events have eight different return periods from two up to one hundred years. The total number of rain event samples is 64. These samples are divided into 56 samples for the training set, 4 samples for validation during the topology search with the growing algorithm and 4 for the test set. This division was made manually to obtain a good representation of strong and weak rain events in all sets.

3.2. Test areas

The test area 1 (see Fig. 3) is about 600 ha of size and has a mean slope of less than one percent. The sewer system consists of 1224 pipe sections and 1143 manholes. The pipe system has a total length of about 53 km. The study area is divided into residential, commercial and industrial areas. The proportion of paved area is slightly less than 50 percent. The HE 2D surface model consists of 1 193 222 triangular cells. The mean cell size is about 3.1 m^2 with sizes in the range from 0.108 m^2 to 17.7 m^2. The buildings are represented by 2186 polygons. The reduced rectangular grid for the ANN model consists of 9623 cells after pruning zero-cells according to Section 2.7. The resulting grid of the catchment was divided into 23 subnets.

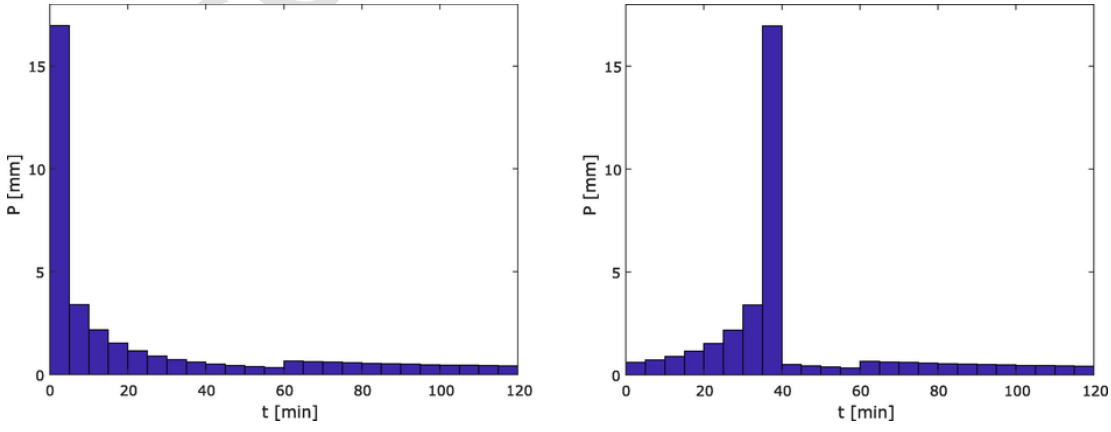

Fig. 2. Examples for Euler type I (left) and Euler type II (right) events. P is the precipitation intensity.

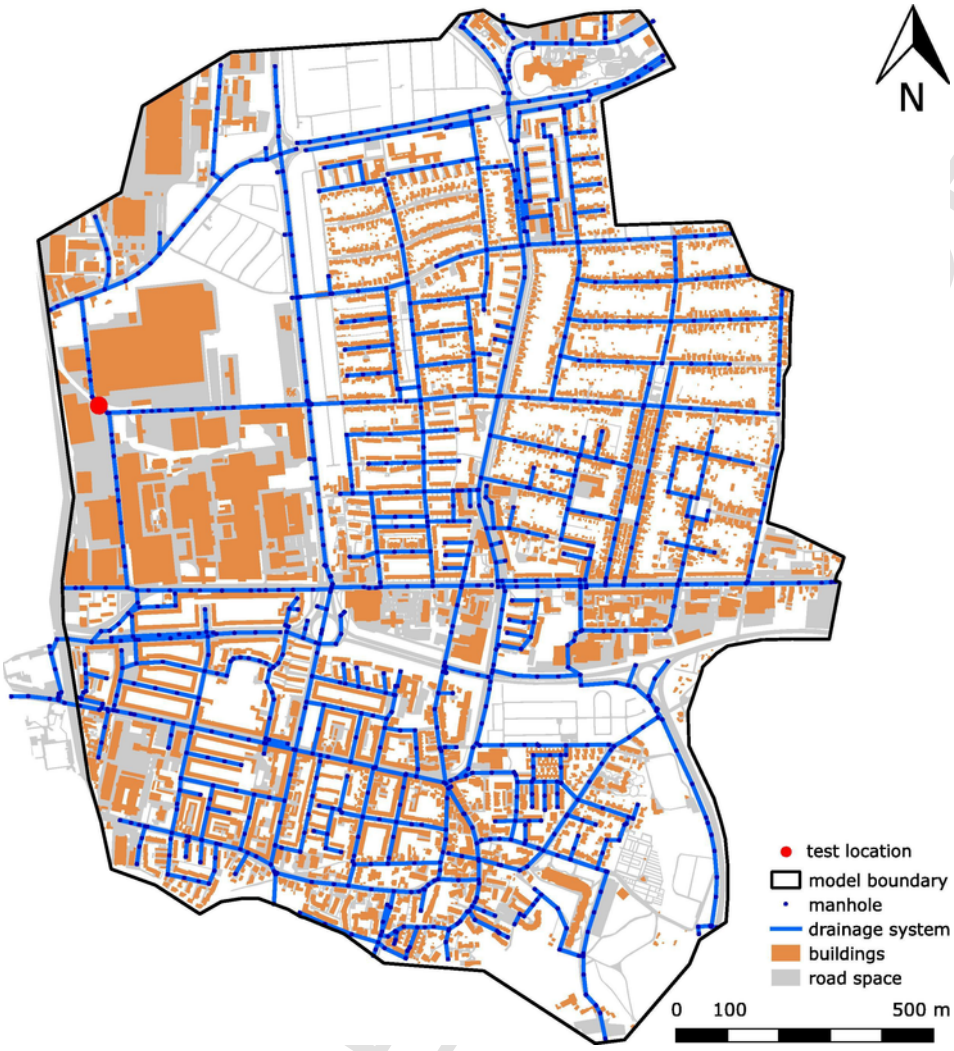

Fig. 3. Schematic representation of test area 1, including the elements of the drainage system, road space and buildings.

To test the model in an area with steep slopes, test area 2 (see Fig. 4) was used. The 86 ha area has a mean slope of 10 percent. The proportion of paved area is about 25 percent. The sewer system has a length of about 10 km. It consists of 299 pipe sections and 291 manholes. The HE 2D surface model consists of 174226 triangular cells. The mean size is about 4 m^2 with sizes in the range from 0.001 m^2 to 8.6 m^2. The buildings are represented by 965 polygons. The reduced rectangular grid consist of 7796 cells. The grid for the catchment was divided into 9 subnets.

4. Results

Results for the different test cases are shown in Table 4 for test case 1 and in Table 6 for test case 2. Resulting topology of the respective ANN obtained through the growing algorithm are shown in Table 3 for test case 1 and in Table 5 for test case 2.

4.1. Test area 1

For the 23 subnets of test area 1, the topology search with the growing algorithm was conducted for ensemble sizes of 1, 5 and 10. The resulting numbers of hidden layers and neurons per layer are shown in Table 3. The number of weights is a measure for the complexity of an ANN. The grid of ANN with ensemble size $n_{ens} = 10$ shows the biggest range of number of weights, but also the smallest average number of weights.

The computation time for whole test area 1 with an ensemble size of 10 for one rain event is in the range from 2 to 3 s. The computation time with the physically based numerical model was in a range from 10 to 55 min. The results for the events from the test dataset are summarized in Table 4. The $\overline{\text{RMSE}}$ decreases with the ensemble size from 0.35 cm to 0.26 cm. The highest RMSE of the simulation with an ensemble size equal to 10 is 0.43 cm for a Euler type 2 event with return period of 50 years and duration of 1 h. Table 4 illustrates that with increasing ensemble size the number of cells with a NSE > 0.9 is increasing. The number of cells with NSE values lower than zero, which can be seen as outliers is decreasing with increasing ensemble size.

The ensemble approach was tested at the test location in test area 1 (see Fig. 3). In total 99 rain events were generated by interpolating between the Euler type II rain events. For this purpose, the precipitation rate was interpolated in the range from a return period of two years up to one hundred years. The model was set up with an ensemble size of 10 for those events. To show the correlation between the precipitation and the flooding, the maximum water level is plotted over the cumulative rainfall P_{sum}. The result is shown in Fig. 5. The data points of the training set (red circles) are well represented by the simulation. The maximum water level for the rain event with a return period of 50 years with a rainfall volume of about 29 mm (green cross) is overes-

Fig. 4. Schematic representation of test area 2, including the elements of the drainage system, road space and buildings.

Table 3
Resulting topologies from growing algorithm with different ensemble sizes for test area 1. Abbreviation are: ensemble size (n_{ens}), number of hidden layers (n_{hl}), number of neurons per layer (n_n), number of weights ($n_{weights}$).

n_{ens}	n_{hl}	$\bar{n}_{hl}$	n_n	$\bar{n}_n$	$n_{weights}$	$\bar{n}_{weights}$
1	1–4	1.96	7–28	10.84	273–16775	4815
5	1–5	2.44	5–20	11.57	273–23019	5723
10	1–4	2.00	7–18	9.34	273–27627	4402

timated by the model. This rain event is part of the test dataset. Also the maximum water level for the validation data point with a return period of 20 years with a rainfall volume of almost 26 mm (black square) is overestimated. Fig. 5 shows that the ensemble mean (blue graph) smooths out the oscillations of the single realizations (gray graphs) that occur due to the strong parametrization of the model.

Table 4
Comparison of the results from the ANN model against the HE 2D model for the test data in area 1.

n_{ens}	computation time [s]	$\overline{\text{RMSE}}$ [cm]	$\text{cells}_{nse < 0}$	$\text{cells}_{nse > 0.9}$
1	1.4	0.35	408	4387 (46%)
5	2.0	0.29	285	4794 (50%)
10	2.4	0.26	49	6693 (70%)

The comparison of flood depth maps in Fig. 6 shows good agreement of the physically based numerical model and the ANN based model. Shown is a representative detail of the urban catchment with a

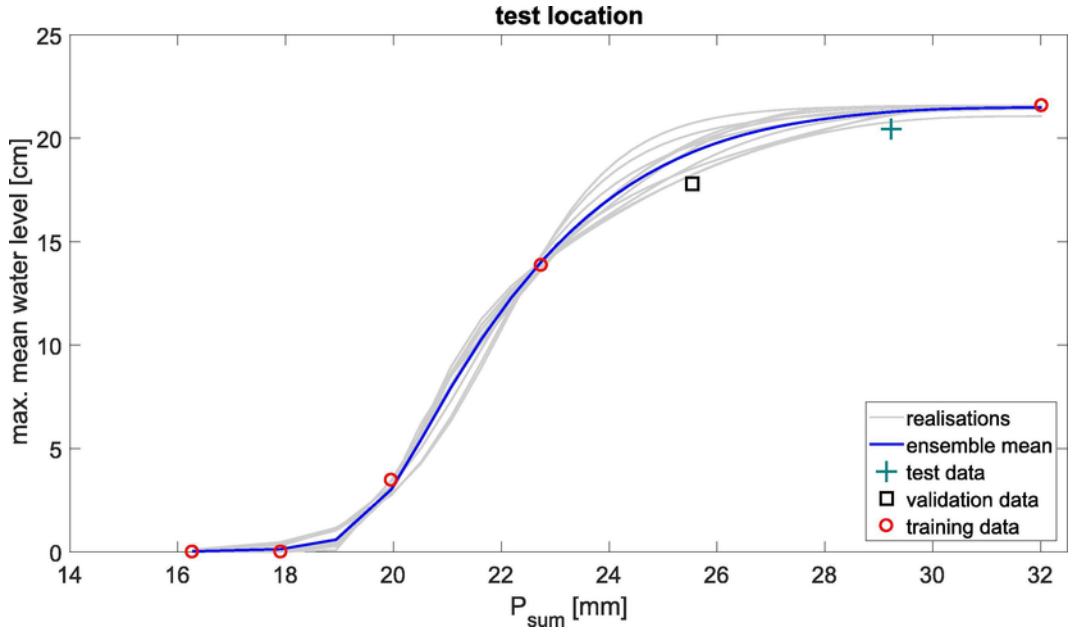

Fig. 5. Results with ensemble size 10 for the test location in test area 1 for Euler type II events with duration of 1 h. Training data is marked with red circles, validation data with a black square and test data with a green cross. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

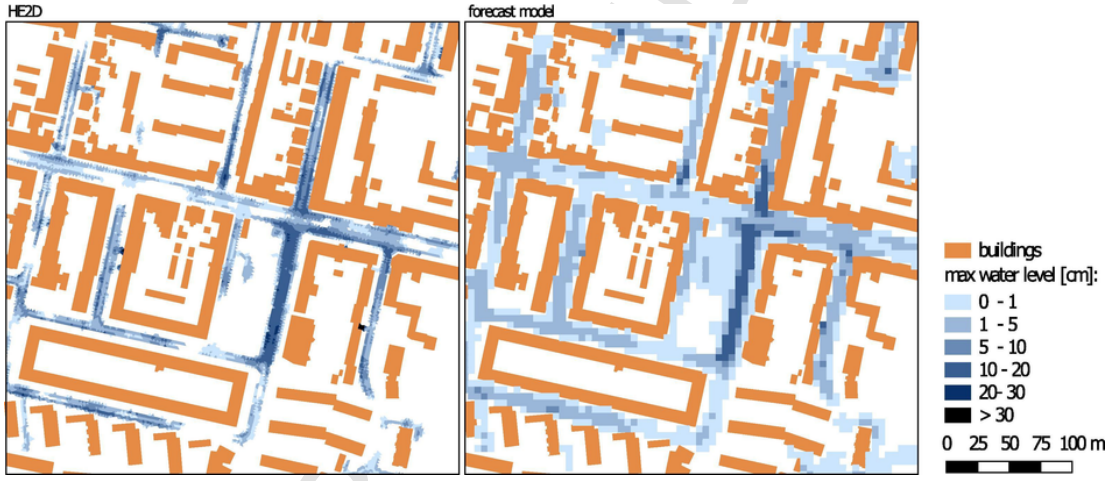

Fig. 6. Maximum water level maps for a detail of test area 1 for an Euler type II event with duration of 1 h and a return period of 50 years. HE 2D model (left) and ANN based forecast model with $n_{ens} = 10$ (right).

size of about 0.11 km^2. At some places the ANN based model overestimates the maximum water level. Due to the reduction of spatial resolution, the water levels are more spread, but the general shape of inundation is represented by the forecast model. The inundated area for the event shown in Fig. 6 is 96233 m^2 for the physically based model and 138204 m^2 for the ANN model. The mismatch is mainly due to the coarsening of the grid and the definition used here for the coarse grained water level (Eq. (6)). To illustrate this, the inundated area with the water level defined as in Eq. (6) was calculated for the coarsened model, resulting in 130104 m^2. Therefore, 33871 m^2 of the mismatch can be attributed to the coarsening. As mentioned before, the definition of the coarse grained water level was chosen such that a good representation of the flood volume is achieved. The flood volume is 5372 m^3 for the physically based model and 5463 m^3 for the ANN model.

The map of NSE in Fig. 7 shows the NSE for all events in the test dataset for the detail of test area 1. In general the map shows NSE values higher than 0.9. However there are locations which have NSE values less than 0.5. Most of these locations have low maximum water levels in a range from 0 to 1 cm. The number of cells with negative NSE decreases with increasing ensemble size. For ensemble size 1 the number of these cells is 408 out of 9623 and for ensemble size 10 the number is 49.

4.2. Test area 2

For the 9 subnets of test area 2, the topology search with the growing algorithm was conducted for ensemble sizes of 1, 5 and 10. The resulting numbers of hidden layers and neurons per layer are shown in Table 5. The grid of ANN with ensemble size $n_{ens} = 1$ shows the biggest range of number of weights, and also the highest average number of weights. The grid with ensemble size $n_{ens} = 5$ shows the smallest average number of weights.

The computation time for the whole test area 2 for one rain event is about one second for all 3 investigated ensemble sizes. The computation time with the physically based numerical model was in a range from 8 to 49 min. The results for the events from the test dataset are summarized in Table 6. The $\overline{RMSE}$ decreases from ensemble size $n_{ens} = 1$ to $n_{ens} = 5$ from 0.19 cm to 0.12 cm. The $\overline{RMSE}$ for $n_{ens} = 10$

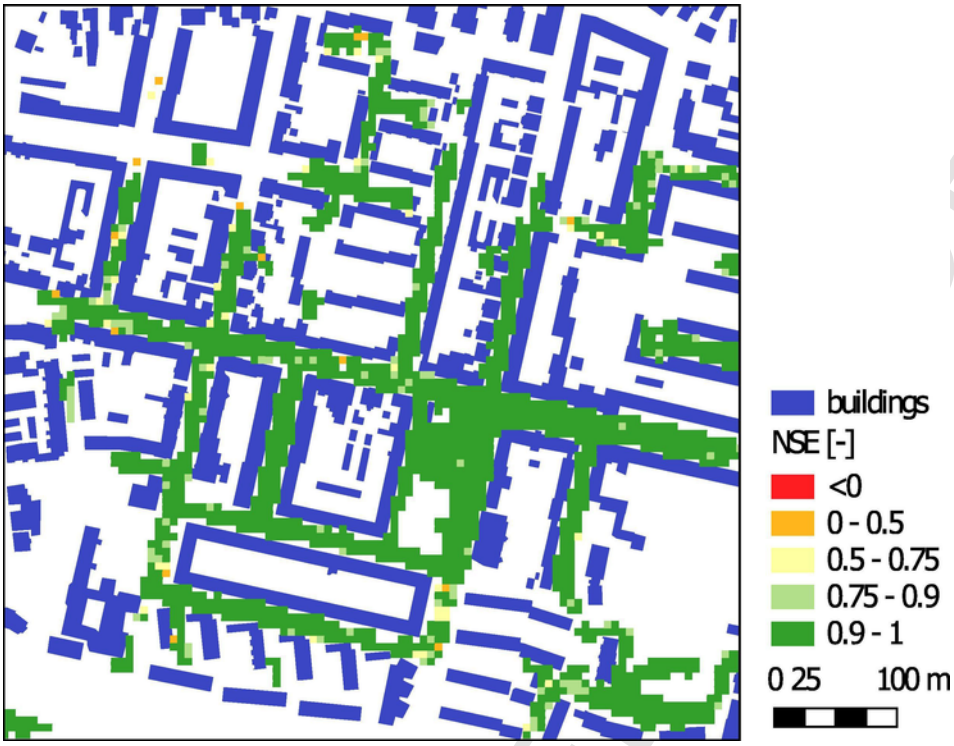

Fig. 7. NSE map for a detail of test area 1 for all events in the test dataset simulated with the ANN model with $n_{ens} = 10$.

Table 5
Resulting topologies from growing algorithm with different ensemble sizes for test area 2. Abbreviation are: ensemble size (n_{ens}), number of hidden layers (n_{hl}), number of neurons per layer (n_n), number of weights ($n_{weights}$).

n_{ens}	n_{hl}	$\bar{n}_{hl}$	n_n	$\bar{n}_n$	$n_{weights}$	$\bar{n}_{weights}$
1	1–3	1.78	7–26	12.56	1908–31221	12513
5	1–4	2.11	7–21	12.53	1910–18105	10927
10	1–3	1.56	8–26	13.93	2483–22904	11153

Table 6
Comparison of the results from the ANN model against the HE 2D model for the test data in area 2.

n_{ens}	Computation time [s]	$\overline{\text{RMSE}}$ [cm]	$\text{cells}_{nse < 0}$	$\text{cells}_{nse > 0.9}$
1	0.7	0.19	3	7388 (95%)
5	0.9	0.12	0	7572 (97%)
10	1.1	0.13	1	7568 (97%)

with 0.13 cm is not further improved. The highest RMSE of the simulation with an ensemble size equal to 5 is 0.14 cm for a Euler type 1 event with return period of 10 years and duration of 2 h. Table 4 illustrates that with increasing ensemble size from $n_{ens} = 1$ to $n_{ens} = 5$ the number of cells with a NSE > 0.9 is increasing. Further increasing to ensemble size $n_{ens} = 10$ gives only marginal benefit. The number of cells with NSE values lower than zero is almost zero for all tested ensemble sizes. The comparison of flood depth maps in Fig. 8 shows good agreement of the physically based numerical model and the ANN based model. (Fig. 9).

5. Discussion

When looking at the general performance of the ANN to predict maximum water levels during a flood event, simulation results of the ANN based forecast model are generally in good agreement with simulations of the physically based numerical model. In particular, no unreasonable results, such as water towers in single cells or large differences between neighbouring cells were generated with the ANN. Many studies stated, that ANN models cannot [e.g. Minns and Hall, 1996], or only barely [e.g. Tayfur et al., 2007] be applied for extrapolation of data. This means that ANN based forecast models can only be used in the range of training data. Thus, a classification of the rainfall input might be beneficial for the presented flood forecast model as a criterion for the applicability of the model.

The inundated area is overestimated by the forecast model by more than 40 percent. The inundated area of the rectangular grid compared to the triangular grid is overestimated by about 35 percent. The ANN model overestimates the inundated area of the coarser grid by about 6 percent. Thus the overestimation is mainly due to the coarser grid compared to the physically based model. With the formulation in Eqs. (5) and (6) the water levels are averaged over the area. With this approach the inundated area is overestimated because water levels are spread to the whole area of the coarser grid cell. In contrast to that the predicted flood volume is in good agreement with the physically based model.

The forecast model gives maximum water levels without temporal distribution. A temporal distribution of water levels is a challenge for further developments of the model, but the model can be seen as a first step for early warning purposes. Also, flood durations could be predicted. First tests showed that the network structure would have to be chosen very differently from the ones obtained for the maximum flood height, so that this was here not considered. First tests to use the same ANN structure that was applied here to predict maximum water height in order to predict the time of the maximum peak showed that a direct transfer of the model is not possible. It would, in any case, be most useful to have a more complete information about the time characteristics of a flood event. In future studies, the model should be expanded to cover information about the duration of a flood. A recursive neural network to predict the full temporal distribution of the flood could be tested for this purpose. Instead of using only precipitation data as input one could additionally use an initial state of water levels as input to

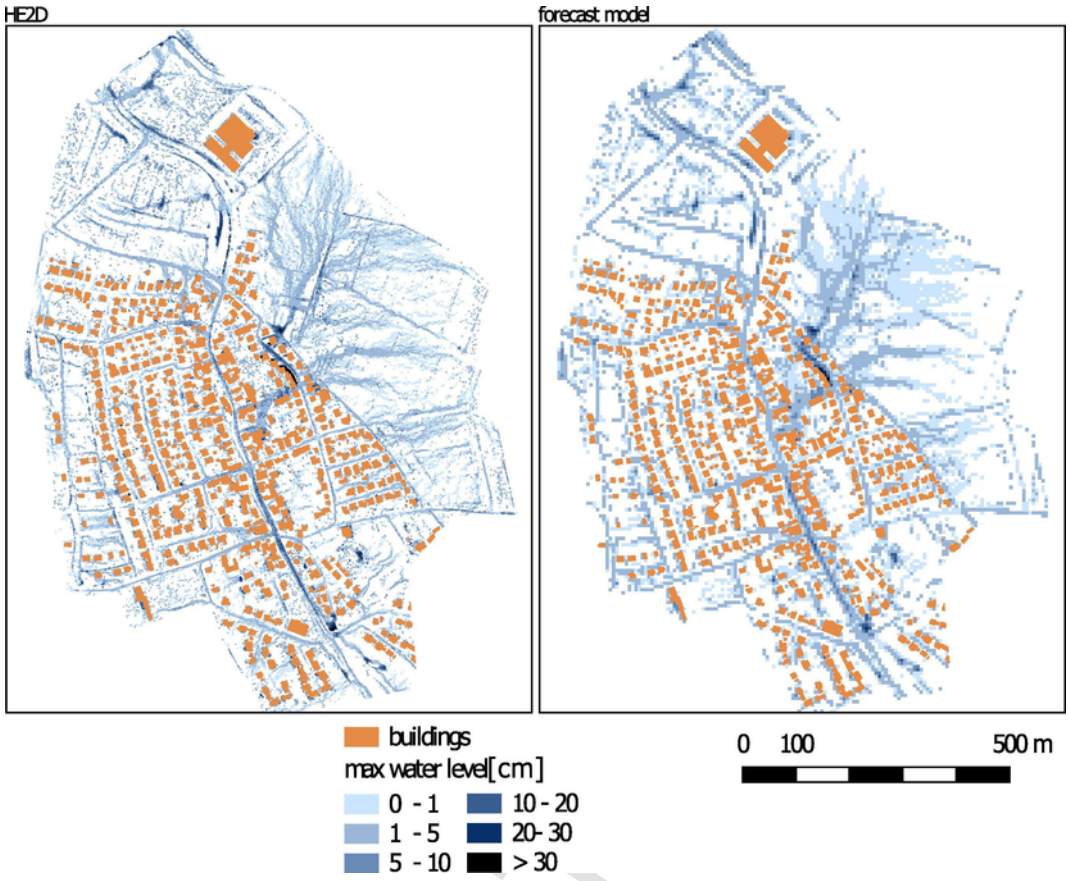

Fig. 8. Maximum water level maps for the test area 2 for an Euler type II event with duration of 1 h and a return period of 50 years. HE 2D model (left) and ANN based forecast model with $n_{ens} = 5$ (right).

calculate a future state of water levels as output. For single locations this approach was already used with success in Rjeily et al. (2017).

A further limitation of the presented model is that it was only tested in urban catchments without real-time controlled sewer systems. The failure of pumping stations or defects in the pipe system and exchange fluxes with the subsurface are also not captured by the ANN based model. Changes in the urban catchment have to be implemented in the physically based numerical model to rebuild the database and the ANN model requires new training. The presented forecast model cannot replace physically based simulations. Nevertheless, despite the necessary and not yet realized further developments listed above, the presented model can be seen as a step towards 2D real-time flood prediction for pluvial floods.

The ensemble approach for the ANN was successfully tested as a measure to overcome the overfitting problem in the strongly parameterized model. With an ensemble size of 10 respectively 5, oscillations are smoothed out and for both test areas the RMSE was at least 25 percent lower compared to the simulation with ensemble size 1. Ensemble sizes bigger than 10 were not investigated due to the immense computation times for the training process. The ensemble size 10 was chosen as a reasonable compromise between accuracy and computation time. Further improvements could be achieved by investigation on composition of the ensemble. Combining ANN with different topologies to an ensemble could be a next step.

Evaluation of NSE maps have shown good performance in general. However, some locations have small NSE values even lower than zero. Those locations are mostly in areas with maximum water levels in a range from 0 to 1 cm. Such levels cannot really be considered as flood heights. Focusing on high water levels for flood prediction by setting a threshold of a minimum water level that is taken into account in a prediction, this uncertainty can be seen as less important. It could in general be questioned how relevant or useful predictions of water levels in the range of 1 cm are. Considering the uncertainties of the prediction with the ANN (and with any model), this height is certainly in the range of uncertainty. The predictions should thus not be read in the sense that the model could really predict a flood height in the range of one centimeter above surface. We kept here also the small water levels in order to compare results of the ANN to results of the physically based model. It should, however, be noted that such results would probably not be used directly in a forecast. Instead, areas would, for example, be classified into different categories of flood hazard. Areas with flood heights in the range of 1 cm would be in a catogory of zero or maybe one.

The transfer of the model from a flat area to an area with steep slopes was successfully tested. The performance of the ANN model was better in this second test case. This could be due to the more pronounced flow paths caused by steeper slopes. Further investigation could be done for bigger catchments.

Simulation results generated with HE 2D were used as virtual truths. The structure of the ANN model allows to replace or expand the data base with results from other models. It would be good to include observation data into the training and testing datasets. For this purpose, it would be useful to install monitoring systems at critical points in an urban area and to make them available for research.

The present study uses synthetic rain events with spatial uniform distribution for model testing. For those rain events the model shows good performance. For more realistic test scenarios nature rain events with spatially distribution should be investigated. A large variety of rainfall durations can be challenging for the ANN and might require sub-networks for different categories. A further step is to use spatial

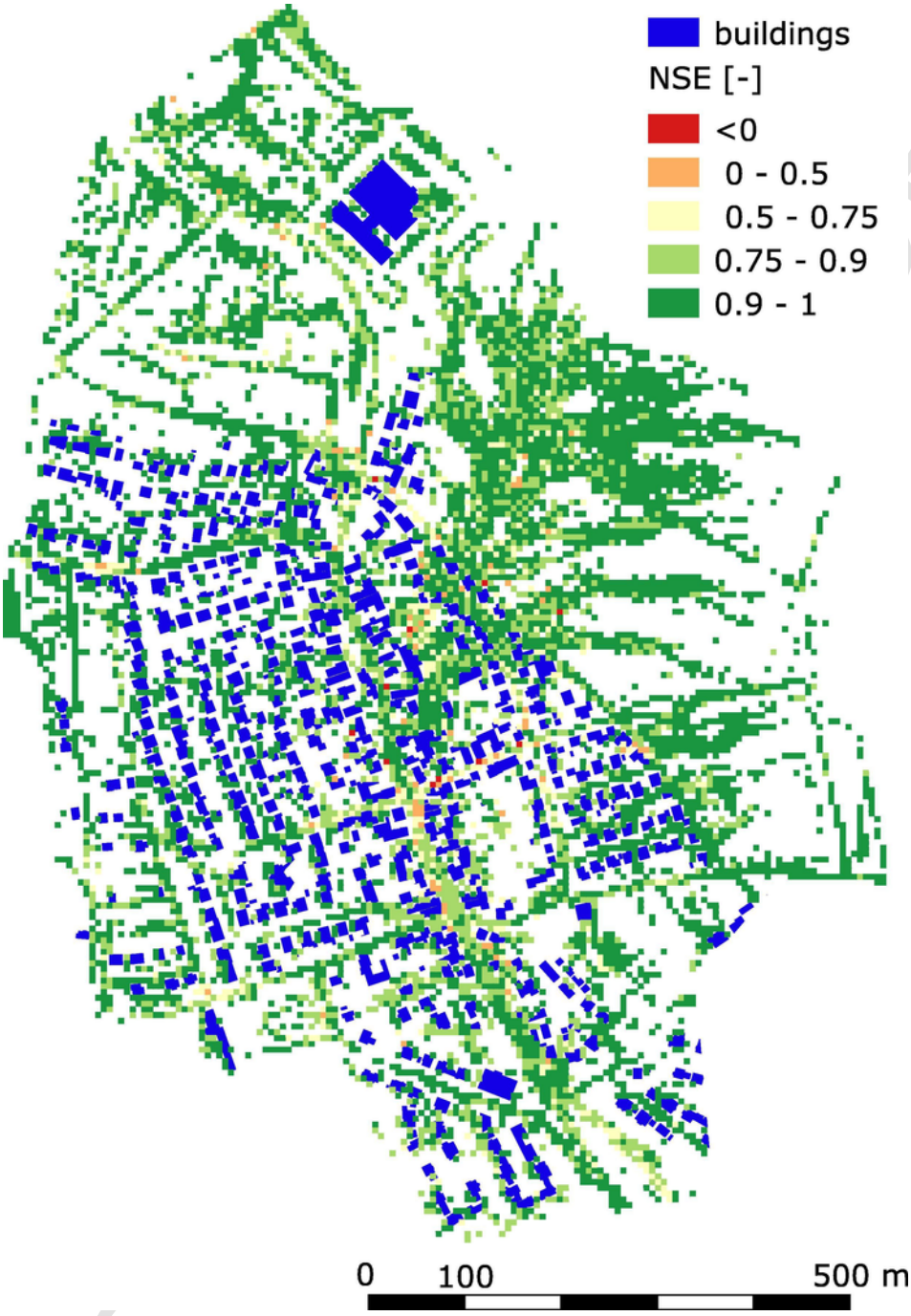

Fig. 9. NSE map for test area 2 for all events in the test dataset simulated with the ANN model with $n_{ens} = 5$.

distributed rain events from radar data as input. By using the subnets with different inputs the model is already prepared for testing spatially distributed rain events as input.

To avoid overloading the ANN with too many output neurons, the domain was here divided into subnets. This was done by dividing the domain into rectangles. In the cases considered in this study, this did not lead to unphysical results. However, as noted in Section 2.8, there are other methods to divide the domain. For larger areas or for more complex flooding patterns, a division into more physically based subnets might be beneficial. A strategy for the coupling can, for example, be found in Wolfs and Willems (2017), where the division would be based on a delineation of smaller sub-catchments. In this way one could aim at minimizing the interrelation of the response of the single subnets. The tested areas in the present study are already small sub-catchments.

A problem of physically based numerical models like HE 2D for fast flood predictions is the computation time. The computation time for the presented forecast model (in the range of seconds) is short enough for early warning system. Even the simulation of an ensemble of for Example 20 rain inputs to capture the uncertainty in rain forecast can be done in less than one minute. As a comparison the prediction of 20 rain events with HE 2D would take about 10 h.

6. Conclusions

An ANN based ensemble approach for real-time water level prediction in urban areas was implemented and tested in two different urban catchments. The model is based on feedforward neural network and trained with precipitation rates as input and 2D distributed maximum water levels as output. A growing algorithm was used to find suitable topologies for the ANN. The maximum water levels for the training were generated with a detailed 1D-2D dual drainage model. The model was tested with test events and compared with the physically based numerical model. Using an ensemble of ANNs was beneficial. The developed model achieves computation times and accuracies that can be considered as sufficient for real-time forecasts. The tests show that ANNs may be useful for flood forecast systems. It is also shown that extensions or alternatives, such as recursive networks or sub-networks should be studied to improve predictions quality. Nevertheless, the presented model can be seen as a step towards 2D real-time flood prediction for pluvial floods.

Declaration of Competing Interest

None.

Acknowledgements

We thank Aaron Peche, Robert Sämann and Tim Berthold for discussion and technical support. Furthermore, we thank the anonymous reviewers and Slobodan Djordjevic for challenging comments which helped to improve the present study. The research is being conducted with financial support of the state of Lower Saxony and the BMBF funded research project EVUS (EVUS – Real-Time Prediction of Pluvial Floods and Induced Water Contamination in Urban Areas) [BMBF, 03G0846A].

References

Aronica, G., Franza, F., Bates, P., Neal, J., 2012. Probabilistic evaluation of flood hazard in urban areas using monte carlo simulation. Hydrol. Process. 26 (26), 3962–3972.

Bebis, G., Georgiopoulos, M., 1994. Feed-forward neural networks. IEEE Potentials 13 (4), 27–31.

Bermúdez, M., Cea, L., Puertas, J., 2019:. A rapid flood inundation model for hazard mapping based on least squares support vector machine regression. J. Flood Risk Manage.e12522.

Bermúdez, M., Ntegeka, V., Wolfs, V., Willems, P., 2018. Development and comparison of two fast surrogate models for urban pluvial flood simulations. Water Resour. Manage 1–15.

Bishop, C.M., et al., 1995. Neural Networks for Pattern Recognition. Oxford University Press.

Bruen, M., Yang, J., 2006. Combined hydraulic and black-box models for flood forecasting in urban drainage systems. J. Hydrol. Eng. 11 (6), 589–596.

Chang, F.J., Chen, P.A., Lu, Y.R., Huang, E., Chang, K.Y., 2014. Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control. J. Hydrol. 517, 836–846.

Dewan, A., 2013. Floods in a Megacity: Geospatial Techniques in Assessing Hazards, Risk and Vulnerability. Springer Science & Business Media.

DHI, 2012. Mike flood 1d-2d modelling user manual.

DWA-A 118, 2006. Hydraulic dimensioning and verification of drain and sewer systems.

Ferguson, B.K., Suckling, P.W., 1990. Changing rainfall-runoff relationships in the urbanizing peachtree creek watershed, atlanta, georgia. JAWRA J. Am. Water Resour. Assoc. 26 (2), 313–322.

Hansen, L.K., Salamon, P., 1990. Neural network ensembles. IEEE Trans. Pattern Anal. Mach. Intell. 12 (10), 993–1001.

Henonin, J., Russo, B., Mark, O., Gourbesville, P., 2013. Real-time urban flood forecasting and modelling–a state of the art. J. Hydroinf. 15 (3), 717–736.

Hölscher, J., Petry, U., Anhalt, M., Meyer, S., Meon, G., Kreye, P., Haberlandt, U., Fangmann, A., Berndt, C., Wallner, M., et al., 2014;36.. Globaler Klimawandel: Wasserwirtschaftliche Folgenabschätzung für das Binnenland: Niedrigwasser. Oberirdische Gewässer

itwh, 2017. HYSTEM-EXTRAN 2D Modellbeschreibung (HYSTEM-EXTRAN 2D model description). Institut für technisch-wissenschaftliche Hydrologie GmbH, Hannover 2017

Jahanbazi, M., Egger, U., 2014. Application and comparison of two different dual drainage models to assess urban flooding. Urban Water J. 11 (7), 584–595.

Jhong, B.C., Wang, J.H., Lin, G.F., 2017. An integrated two-stage support vector machine approach to forecast inundation maps during typhoons. J. Hydrol. 547, 236–252.

Kröse, B., Krose, B., van der Smagt, P., Smagt, P., 1993. An introduction to neural networks. Citeseer

Leandro, J., Chen, A.S., Djordjević, S., Savić, D.A., 2009. Comparison of 1d/1d and 1d/2d coupled (sewer/surface) hydraulic models for urban flood simulation. J. Hydraul. Eng. 135 (6), 495–504.

McCulloch, W.S., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. 5 (4), 115–133.

Minns, A., Hall, M., 1996. Artificial neural networks as rainfall-runoff models. Hydrol. Sci. J. 41 (3), 399–417.

Moré, J.J., 1978. The levenberg-marquardt algorithm: implementation and theory. In: Numerical analysis. Springer, pp. 105–116.

Preissmann, A., 1961. Propagation des intumescences dans lescanaux et rivieres. 1st Cong French Assoc for Computation, Grenoble.

Reed, R., 1993. Pruning algorithms-a survey. IEEE Trans. Neural Networks 4 (5), 740–747.

René, J.R., Djordjević, S., Butler, D., Mark, O., Henonin, J., Eisum, N., Madsen, H., 2018. A real-time pluvial flood forecasting system for castries, St. Lucia. J. Flood Risk Manage. 11, S269–S283.

Riedmiller, M., Braun, H., 1993. A direct adaptive method for faster backpropagation learning: the rprop algorithm. In: Neural Networks. IEEE International Conference on IEEE, pp. 586–591.

Rjeily, Y.A., Abbas, O., Sadek, M., Shahrour, I., Chehade, F.H., 2017. Flood forecasting within urban drainage systems using NARX neural network. Water Sci. Technol. 76 (9), 2401–2412.

Rosenblatt, F., 1958. The perceptron: a probabilistic model for information storage and organization in the brain. Psychol. Rev. 65 (6), 386.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323 (6088), 533.

Sajikumar, N., Thandaveswara, B., 1999. A non-linear rainfall-runoff model using an artificial neural network. J. Hydrol. 216 (1–2), 32–55.

Schreider, S.Y., Smith, D., Jakeman, A., 2000. Climate change impacts on urban flooding. Clim. Change 47 (1), 91–115.

Tayfur, G., Moramarco, T., Singh, V.P., 2007. Predicting and forecasting flow discharge at sites receiving significant lateral inflow. Hydrol. Process.: Int. J. 21 (14), 1848–1859.

Thorndahl, S., Nielsen, J.E., Jensen, D.G., 2016. Urban pluvial flood prediction: a case study evaluating radar rainfall nowcasts and numerical weather prediction models as model inputs. Water Sci. Technol. 74 (11), 2599–2610.

Wolfs, V., Willems, P., 2017. Modular conceptual modelling approach and software for sewer hydraulic computations. Water Resour. Manage. 31 (1), 283–298.

Zhang, G., Patuwo, B.E., Hu, M.Y., 1998. Forecasting with artificial neural networks: the state of the art. Int. J. Forecasting 14 (1), 35–62.