E-proceedings of the 38th IAHR World Congress
September 1-6, 2019, Panama City, Panama
doi:10.3850/38WC092019-0274

CAN MACHINE LEARNING IMPROVE THE ACCURACY OF WATER LEVEL FORECASTS FOR INLAND NAVIGATION? CASE STUDY: RHINE RIVER BASIN, GERMANY

YUELING MA[1], ELENA MATTA[2], DENNIS MEIßNER[3], HANNO SCHELLENBERG[4] & REINHARD HINKELMANN[5]

[1,2,5] Technische Universität Berlin, Berlin, Germany
mayueling-2017@hotmail.com; elena.matta@wahyd.tu-berlin.de; reinhard.hinkelmann@wahyd.tu-berlin.de

[3] Federal Institute of Hydrology, Koblenz, Germany
meissner@bafg.de

[4] BearingPoint GmbH, München, Germany
hanno.schellenberg@bearingpoint.com

ABSTRACT

In a context where the German Federal Ministry of Transport and Digital Infrastructure (BMVI) expects a 23% growth of traffic in inland navigation in Germany by the year 2030, a higher efficiency and an optimized logistics in traffic management are required. The BMVI-funded project Digital Skipper Assistant (DSA) had the objective to develop a cloud-based application to support inland navigation, able to calculate best routes and Estimated Times of Arrival (ETA). Contributing to the DSA, this study investigates the competence of artificial neural networks (ANNs) to predict water levels up to 10 days ahead in some crucial gauges of the Rhine River Basin in Germany. A multiple-outputs model based on long short-term memory (LSTM) networks was implemented, adopting as inputs firstly the water level measurements at specific gauges. In a second work phase, the water level forecasts of the hydrological model chain of the German Federal Institute of Hydrology (BfG) were included as additional predictor into the model, highly improving the results (the coefficient of determination R^2 increased of about 20%). The LSTM model has been trained, validated and tested (respectively with 80%, 10% and 10% of the dataset) using the historical data and the BfG hindcasts from January 2008 until December 2015. The results of the model evaluation were very good (i.e. R^2 around 90% for 7-days prediction). Several tests have been run during the DSA field test (from July 2018 to December 2018) and the results were promising.

Keywords: Water level predictions; River Rhine, Germany; ANN, LSTM; hydrological modeling.

1 INTRODUCTION

The German Federal Ministry of Transport and Digital Infrastructure (BMVI, 2016) is expecting an increase of about 23% of traffic in inland navigation in Germany between 2010 and 2030. Due to few foreseen changes to the current infrastructures, a more efficient and optimized traffic management is demanded. In this context, the BMVI-funded project Digital Skipper Assistant (DSA) allowed the development of a cloud-based application, to support the inland navigation stakeholders, providing user-specific route guides and forecast information, for the first time in unique digital source. The DSA was a collaborative work between the management and technology consultancy firm BearingPoint GmbH (coordinator of the project) and BearingPoint Technology GmbH, together with the German Federal Ministry of Hydrology (BfG) and the Chair of Water Resources and Modeling of Hydrosystems at TU Berlin.

In this work, different machine learning (ML) techniques for water level forecasts in the Rhine River in Germany have been investigated. Artificial neural networks (ANNs) have gained considerable popularity in Hydroscience over the last two decades, and hydrologists have successfully implemented such approaches to predict water levels in various rivers over the world (e.g. the Anyangcheon Stream in Korea by Sung et al. (2017); the Reno River in Italy by Alvisi et al. (2006); the Tagliamento River in Italy by Campolo et al. (1999); the Mosel River in Germany by Stüber and Gemmar (1997); the Pisuena River in Spain by Crespo and Mora (1993)).

Compared to traditional hydrological models, ANNs are attractive for many water researchers, because of their low requirements of physical background (e.g. topography) and high toleration of input errors (Thirumalaiah and Deo, 1998). In addition, the possibility of extensive parallel processing and parsimonious data storage enable ANNs to execute often faster than numerical models (Tanty and Desmukh, 2015). Nevertheless, although the high performances of ANNs in Hydroscience have been widely proven, their shortages should not be ignored. The precondition of a good ANN application is both the quality and quantity of data available. It is usually hard to achieve, as there are often not such long hydrological records available. The lack of physical concepts and relations is another major disadvantage of this methodology, resulting in no standardized way of choosing a network structure. The design of an ANN is usually based on the previous experience and preference of the developer instead of the physical aspects of the problem, which is why ANNs are sometimes difficult to reproduce (Govindaraju, 2000).

Many works have shown that ANNs can be very successful for short-term water level predictions (less than 1 day ahead). However, the focus of this study is on the longer term, since it is more interesting for the stakeholders in inland navigation. Indeed, a survey reported by Meißner et al. (2018) demonstrated that many skippers would prefer an up to 7-days in advance forecasts to optimize navigation whereas longer-term predictions are also desired by them e.g. for a complete cycle of transport plan. Hydrological models usually take a long time (up to hours) to calculate long-term forecasting and may have many input noises included. Therefore, it is worth to explore the performance of ANNs in long-term forecasting, where still limited studies are available. In detail, the focus of this work – as part of the DSA project – was to investigate the capabilities of ANNs in predicting deterministic water levels up to 10 days ahead in some critical gauges of the German section of the Rhine River Basin. The word ‘critical’ is used to define those gauges, which are relevant for navigation in order to determine the maximum load during periods of low water levels.

2 METHODOLOGY

2.1 Artificial neural networks

An ANN is a massively parallel-distributed computing paradigm consisting of many basic information-processing elements (namely neurons or nodes) interconnected together (Haykin, 2009). Motivated by biological neural systems, ANNs have wide learning competence from historical data with high toleration of input noises, capable to implement complex operations and input-output mapping. Furthermore, their plasticity permits the developing networks to be retrained for the adaptation to the new environments. Therefore, ANNs are able to solve many real-world issues such as hydrodynamics processes, which can be highly nonlinear and complicated. The most common architecture of an ANN is multilayer (e.g. Figure 1), constructed by an input layer accepting inputs from the surrounding, an output layer releasing the network’s responses to its inputs and one or more layers in between named hidden layers (Liong et al., 2000).

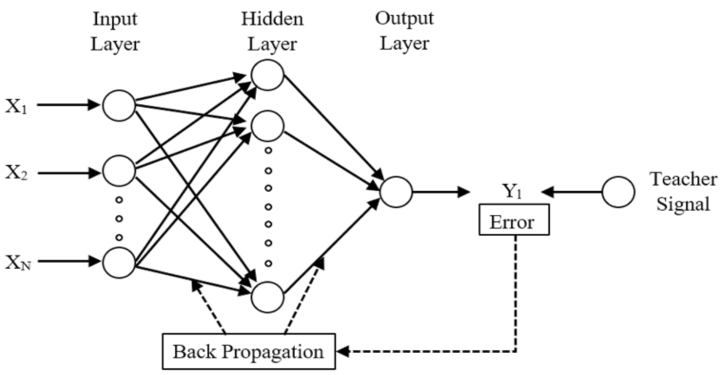

Figure 1. Topography architecture of a three-layer network which learns supervised (Liong et al., 2000).

ANNs can be categorized into feedforward (FF) networks and recurrent neural networks (RNNs) regarding the direction where information flows. FF networks are the most widely used neural networks, having a relatively elementary structure compared to others. The neurons only have the connections subsisting in adjacent layers in an orientation to the nodes in the subsequent layer, and each link is joined with a synaptic weight indicating the influence of the two neurons at both sides on the input-output predictions. Due to this architecture, the factors which determine the outcome of a neuron in a layer are solely the inputs produced by the previous layers and the corresponding weights, and the information processes from the input to the output side (Govindaraju, 2000).

In comparison with FF networks, RNNs consider not only the current inputs but also the information received in the past stored into the neural inputs, achieved by their workable memories as well as loops existing in their hidden layers. The majority of linkages in RNNs is bidirectional and exists between a pair of neurons in the same layer, so the information can also travel in an opposite direction (i.e. from outputs to inputs) in RNNs, which highly upgrades the networks’ performance on the temporal association (Karim and Rivera, 1992). Benefiting from this distinguished structure, RNNs usually show their superiority on the issues regarding the sequential information such as natural language processing (NLP) and long-term time series predictions. The comparative study conducted by Nagesh Kumar et al. (2004) states that RNNs surpass FF networks on both single-step ahead and multi-step ahead predictions with smaller architecture and less learning time. In addition, the research of Chang et al. (2014) pointed out the superiority of RNNs on detecting the lasting dependencies via their repeated outputs, which considerably diminish output instability. Given the proficiency of RNNs in long-term forecasting, they were chosen as the computing tool for this study.

Network training, validation and testing are indispensable processes in the application of an ANN, which guarantee the network to generate same or similar outcomes as the targets, and it is possible to only have network training and testing when the dataset is small. Training (or learning) is an operation to minimize the error from output layer by the adaptation of the parameters involved in the network and the equation to compute the error is called the loss function, commonly using mean squared error in time sequence problems. Generally, two types of training exist: supervised and unsupervised. Most problems in Hydroscience implement the former algorithm which requires an additional teacher signal to conduct the training process (as illustrated in Figure 1). Validation, in some papers also named crossing training, is a process to prevent the network from the phenomenon of overfitting. The training of a network will be stopped at a certain time when the error relevant to the training set continues to decrease but that for the validation set starts to increase, and it is assumed that the optimal parameters associated with the network can be obtained after such training. Finally, testing is a process to evaluate the quality of a trained network using a brand new dataset as input, which originates from the same source as the training set and has never seen by the network during training (Govindaraju, 2000; Liong et al., 2000).

2.2 Comparison with hydrological models

ANNs have many similarities as hydrological models, which are a simplified real-world system to simulate hydrological processes relying principally on cause-effect relationships based on mathematical equations (e.g. continuity equations) and /or empirical relationships (e.g. rainfall-runoff, storage volume-outflow) (Devia et al., 2015; Hinkelmann, 2005). An ANN can be considered as a black box model, category where some hydrological models also fall in, which are capable to establish a functional relation between historical inputs and outputs with little background knowledge. Network training, validation and testing are also analogous concepts as calibration and validation in hydrological models. However, ANNs can be more versatile because of their flexible structures and no explicit mathematical functions inside (Govindaraju, 2000). Furthermore, ANNs often have higher computation speed than hydrological models.

2.3 Long short-term memory networks

Within the RNNs, the long short-term memory (LSTM) networks were chosen for this study. The networks were first proposed by Hochreiter and Schmidhuber (1997) and later optimized by Gers et al. (2000) with the introduction of a forget gate in the constant error carousel (CEC) (highlighted with a red circle in Figure 2) in order to prevent linear rising cell states in the repeating modules, i.e. the units to loop in the hidden layers. Four interacting layers are included in the repeating modules, ensuring constant backpropagated error signals via time and layers. Benefiting from the distinctive architecture, LSTM networks have the competence to capture relations over 1000 time steps, regardless of input noises (Hochreiter and Schmidhuber, 1997). This leads to the excellence of LSTM networks on the investigation concerning long-term dependencies. Figure 2 illustrates the structure of an optimized LSTM repeating module, where its working process is also labelled.

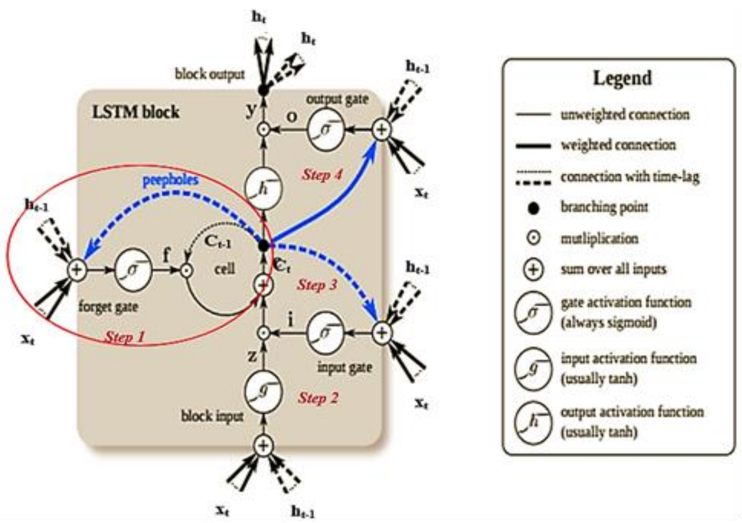

Figure 2. Repeating modules of a long short-term memory (LSTM) network (modified from Skymind, 2017)

(x: input; h: output; C: cell state; t-1 and t: the previous step and current step, e.g. h_{t-1}—the output from the previous step)

The computing processes of a LSTM repeating unit and their relevant mathematical equations are as follows (Olah, 2015):

- Step 1: Determine the information to forget from the cell state
  The forget gate estbalished by a sigmoid activation function examines h_{t-1} and x_t and derives a value in the range of 0 and 1 for each number in the cell state C_{t-1}, where 0 and 1 identify "fully forget the number" and "fully keep the number", respectively:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] - b_f) \quad [1]$$

  where, W_f and b_f are the weight vector and bias related to the forget gate, and the function $\sigma(\cdot)$ is the signoid function presented in Eq. [2].

$$\sigma(\text{net}) = \frac{1}{1 + e^{-\text{net}}} \quad [2]$$

- Step 2: Determine the new information to reserve in the cell state
  The input gate constructed by a sigmoid activation function chooses the values to update and the input activation function that produces a vector of new appropriate values $\widetilde{C_t}$.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] - b_i) \quad [3]$$

$$\widetilde{C_t} = \tanh(W_C \cdot [h_{t-1}, x_t] - b_C) \quad [4]$$

  where, W_i, b_i and W_C, b_C are the weight vectors and biases associated to the input gate and the input activation function respectively.
- Step 3: Update the cell state
  The old cell state C_{t-1} and f_t calculated in Step 1 are multiplied to abandon the unnecessary information, while the product of i_t and $\widetilde{C_t}$ is completed to scale the new appropriate values from Step 2 satisfying certain requirements. The new cell state C_t is the sum of these two results.

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C_t} \quad [5]$$

- Step 4: Determine the output information based on the cell state
  The output gate formed by a sigmoid activation function filters the output cell state, and then the output activation function tanh scales the cell state between -1 and 1, whose outcome is later multiplied by the result of the output gate. The multiplication is the final chosen output of the current cell:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] - b_o) \quad [6]$$

$$h_t = o_t * \tanh(C_t) \quad [7]$$

  where, W_o and b_o are the weight vector and bias relevant to the output gate.

3 STUDY AREA AND DATASETS

3.1 German section of the Rhine River Basin

The study area is the German section of the Rhine River Basin up to the gauge Emmerich (shown in Figure 3). The Rhine River is a transboundary river, connecting nine European countries, i.e. Austria, Belgium, France, Germany, Italy, Liechtenstein, Luxemburg, the Netherlands and Switzerland (Cioc, 2002). In Germany, the Rhine River is the most crucial waterway. Many German large cities are situated along the river, such as Cologne, Dusseldorf and the Ruhr area. It also takes a very important role in navigation and transportation, roughly 716 km of whose length are navigable for large ships (WSV, 2014). Approximately 80% of the water carriage in Germany happens on the Rhine River (toponline.org, n.d.). Oestrich and Kaub are two critical gauges among the river for navigation, because of their location and low water levels. Therefore, the gauge Oestrich was chosen as one of the gauges to study in this research.

Shipping activities in the Rhine River are very vulnerable to long-term droughts which result in frequent low water levels. The capability of the navigation and the vessel speed are bound to the dropped water stages, inducing the increase of the travelling time and fuel consumption (Meißner et al., 2018).

3.2 Datasets

The datasets investigated in this study are the daily water level measurements and predictions at different gauges among the Rhine River from December 2007 to January 2016 for network setup and from July 2018 to December 2018 for the field test. All data were provided by the Federal Institute of Hydrology (BfG, Koblenz), also DSA project partner. The dataset includes the forecasts up to 10-days ahead at the gauges Oestrich, Kaub, Koblenz, Cologne, Dusseldorf, Ruhrort, Emmerich (daily and hourly). They were computed by the BfG hydrological model chain, consisting of a hydrological model, fed by external weather forecasts, for flow predictions and a hydrodynamic model for water level forecasts (Meißner et al., 2018).

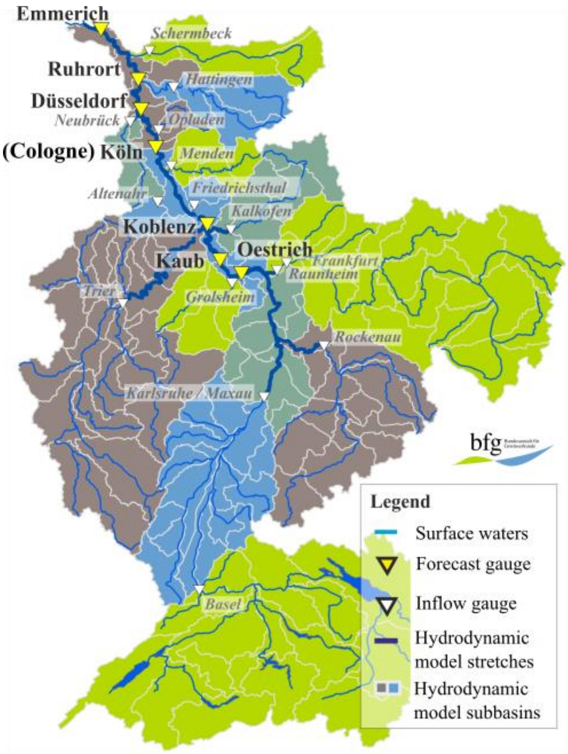

Figure 3. Study area, the German section of the Rhine River Basin from Basel up to Emmerich

4 NETWORK CONFIGURATION

4.1 Basic idea of model setup

In a river system, a strong relationship exists between the upstream and the downstream, and thus the information at a downstream target is determined by its previous measurements and those at its influencing upstream gauges. For this reason and such physical linkage, this study focuses exclusively on water levels.

4.2 Work flow

The work flow of the model setup in this study is displayed in Figure 4.

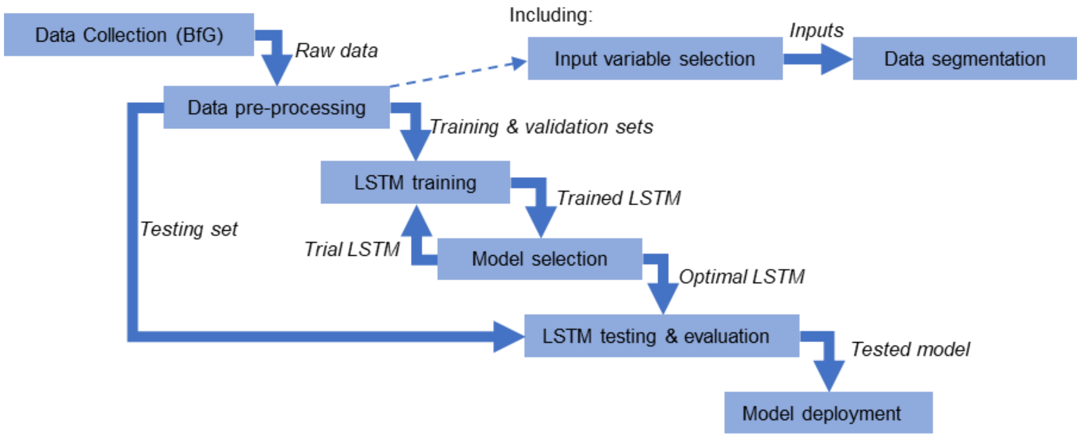

Figure 4. Procedure of model setup

The raw data were provided by BfG as mentioned in the last section. Firstly, these data were cleaned, removing missing values, and transformed into the format of Figure 5, which includes teaching signals to supervise the training process. Then the data were separated into a training set (80% of the whole data), a validation set (10% of them) and a testing set (10% of them). The former two sets were used for model training, and the best-trained model with certain parameters was assumed to be obtained when the error for training set kept dropping but that for validation set began to rise. Next, after several sensitivity tests (i.e. varying parameters such as the number of hidden layers and neurons, scaling method and batch size, seeking the improvement of the results), the optimal LSTM network was obtained. Finally, its performances were evaluated using the testing set and further deployed to predict the water levels not only at Oestrich, but also at other gauges along the river.

Historical measurements							BfG predictions	Teaching signals			
B(t-14d)	**…**	**O(t-14d)**	**…**	**B(t-1d)**	**…**	**O(t-1d)**	**10-days forecasts from BfG**	**O(t)**	**O(t+1d)**	**…**	**O(t+10d)**
…	…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…	…

Figure 5. Formats of the adapted data for the implemented LSTM network in various scenarios (B-Basel; O-Oestrich; d-day(s))

The statistical methods used for model evaluation are the root mean square error (RMSE), the coefficient of determination (R^2) and the percentage of the forecast with the absolute error below 10 cm and 20 cm (i.e. $P_{E<10}$ and $P_{E<20}$). RMSE and R^2 are common approaches to evaluate the accuracy of a dataset. $P_{E<10}$ and $P_{E<20}$ are the skipper-relevant information to evaluate the accuracy of the predictions. The formulae of these four approaches are reported in Eq. [8] to Eq. [13].

$$\mathrm{RMSE} = \sqrt{\frac{\sum(y_{exp} - y_{pred})^2}{N}} \qquad [8]$$

$$R^2 = 1 - \frac{\sum_{i=0}^{N-1}(y_{exp} - y_{pred})^2}{\sum_{i=0}^{N-1}(y_{exp} - \overline{y_{exp}})^2} \qquad [9]$$

$$\overline{y_{exp}} = \frac{1}{N}\sum_{i=0}^{N-1} y_{exp} \qquad [10]$$

where, y_{exp} is the expected value, and y_{pred} is the predicted value, and N is the number of the compared pairs.

$$\text{Absolute error} = |y_{exp} - y_{pred}| \qquad [11]$$

$$P_{E<10} = \frac{N_{\text{Absolute error}<10}}{N} \times 100\% \qquad [12]$$

$$P_{E<20} = \frac{N_{\text{Absolute error}<20}}{N} \times 100\% \qquad [13]$$

where, $N_{\text{Absolute error}<10}$ is the number of the forecasts with the error below 10 cm, and $N_{\text{Absolute error}<20}$ is similarly defined.

5 RESULTS AND DISCUSSION

5.1 Architecture of the optimal LSTM network

The gauges Oestrich and Cologne were selected as target gauges in this study. Based on the observed and simulated data collected at Oestrich between 2007 and 2016, the network was optimized and its structure is listed in Table 1. Due to the direct physical relationship between upstream and downstream in a river system, a similar strategy is applicable to forecast water levels at downstream gauges. Therefore, the architecture chosen for the better performing network for Oestrich was also adopted for the predictions at Cologne.

In this study, the input gauges for water level predictions at Oestrich are Basel, Maxau, Rockenau, Frankfurt, Oestrich, and those to forecast water levels at Cologne are the input gauges for Oestrich plus two extra gauges Kaub and Cologne. The locations of all the input gauges are labelled in Figure 3. For both target gauges, the inputs were collected starting from the gauge Basel, which is situated upstream to Maxau (Figure 3). In the Rhine River, the waterway from Basel to Maxau is highly controlled by weirs, while the waterway between Maxau and Emmerich is free-flowing. Although the flow is highly regulated near Basel, its water levels are still directly associated with those at its downstream gauges such as Maxau and Oestrich. With the measurements at Basel involved, the network can receive additional forecasting time from the travelling time of water from Basel to Oestrich (almost 1 day), which can improve the performance of the network. Furthermore, the inputs of the network also introduced the water level forecasts from the BfG hydrological chain, in order to include some physical knowledge, e.g. precipitation, temperature etc. (Meißner et al., 2018).

Table 1. Information regarding the structure of the better performing LSTM network

Input	Previous 2-week water levels collected from input gauges with a frequency of 1 day + 10-day predictions produced from the BfG hydrological chain
Output	Daily water level prediction at Oestrich or Cologne (1 day ahead, 2 days ahead, …, up to 10 days ahead)
Time period	From December 19th, 2007 to January 10th, 2016
Dataset distribution	Training set (80%), validation set (10%), testing set (10%)
Structure	A visible layer with 2D input tensors (determined by input dataset) + a hidden layer with 64 LSTM neurons + an output layer with 10 outputs
Activation function	Hyperbolic tangent activation function (tanh)
Loss function	Mean Square Error (MSE)
Number of epochs	Depending on the condition to cease training
Batch size	1024

5.2 Model evaluations

The accuracy of the developed network was calculated using the results from the testing process, whose inputs were water levels approximately between 2015 and 2016, and calculating RMSE, R^2, $P_{E<10}$ and $P_{E<20}$ as statistical parameters.

The performances of the following three models have been compared: 1) the LSTM network with only daily measurements as inputs; 2) the LSTM network with the historical measurements and the simulations from the BfG hydrological model chain as inputs (indicated as 'the LSTM network with the integrated inputs'); 3) the BfG hydrological model chain.

Considering the long-term water level predictions (up to 10 days) at Oestrich, the LSTM network with the integrated inputs delivered better results, mainly reflected in two aspects. Firstly, it outperforms in regards to the overall accuracy (i.e. RMSE and R^2) of its forecasts with a leading time from 2 to 10 days. Taking the 10-days in advance forecast as an example (Figure 6), R^2 from the combined model is 77.4%, above 10% higher than for the hydrological model chain (R^2 = 63.5%) and more than 15% higher than that from the LSTM network having only the daily measurements as inputs (R^2 = 51.7%). Secondly, the distribution of the forecasting errors (i.e. $P_{E<10}$ and $P_{E<20}$) from the 2nd to the 7th day is lower compared with ones of the hydrological model chain. In other words, considering e.g. $P_{E<10}$ (Figure 7), the LSTM network with the integrated inputs has the ability to produce a higher number of predictions with a difference smaller than 10 cm between measurements and precipitations.

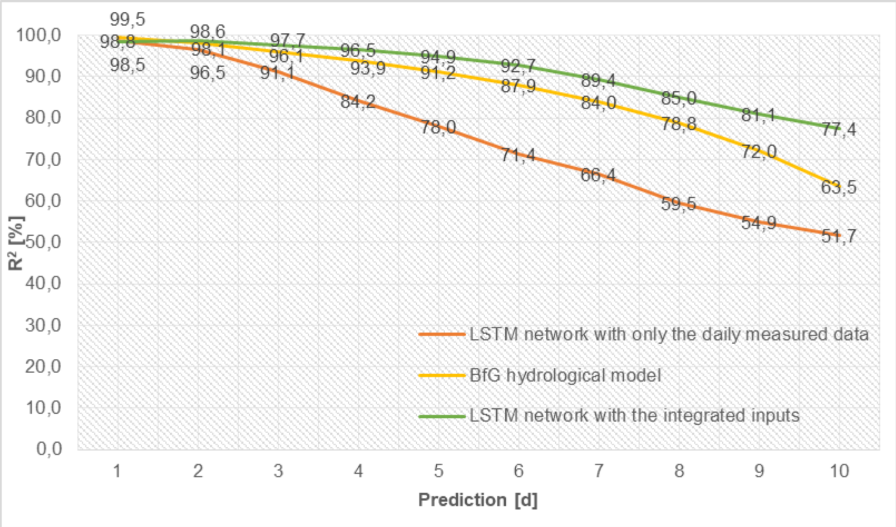

Figure 6. Coefficient of determination R^2 comparing the different models (Target: Oestrich)

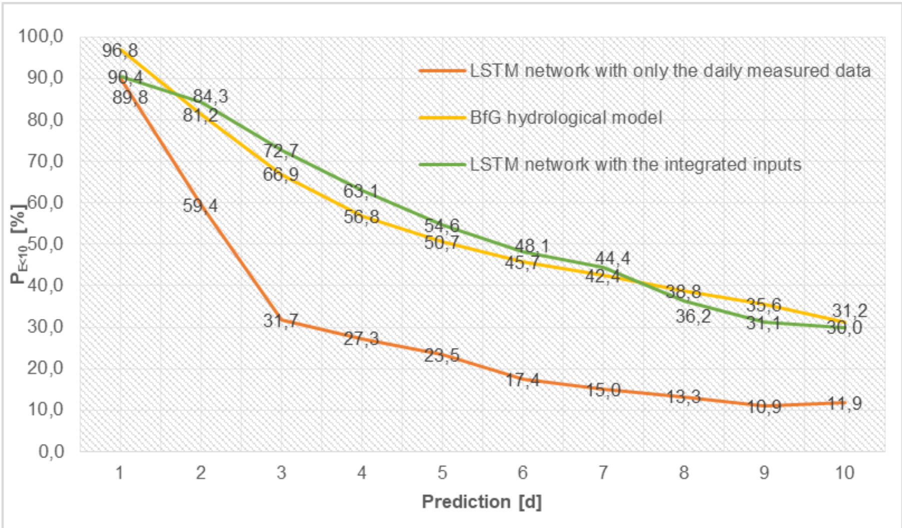

Figure 7. Percentage of the absolute error $P_{E<10}$ comparing the different models (Target: Oestrich)

The structure of the better performing LSTM network with the integrated inputs designed for Oestrich also yielded good results for the water level forecasts at Cologne (e.g. the comparison of R^2 of the different models shown in Figure 8). This demonstrates that the developed model has the capability to deliver reliable predictions at any gauges along the Rhine River with a properly prepared input dataset, which is an advantage of this multi-outputs model. Nevertheless, the performance of the model in Cologne showed a lower accuracy. With regard to R^2, the accuracy for both 1-day ahead and 2-days ahead predictions at Cologne are lower than the ones of the hydrological model chain. This occurs because the LSTM architecture was optimized for the target Oestrich; therefore, a lower accuracy will be encountered when predicting water levels for another gauges along the Rhine, compared to the one in Oestrich (Figure 6).

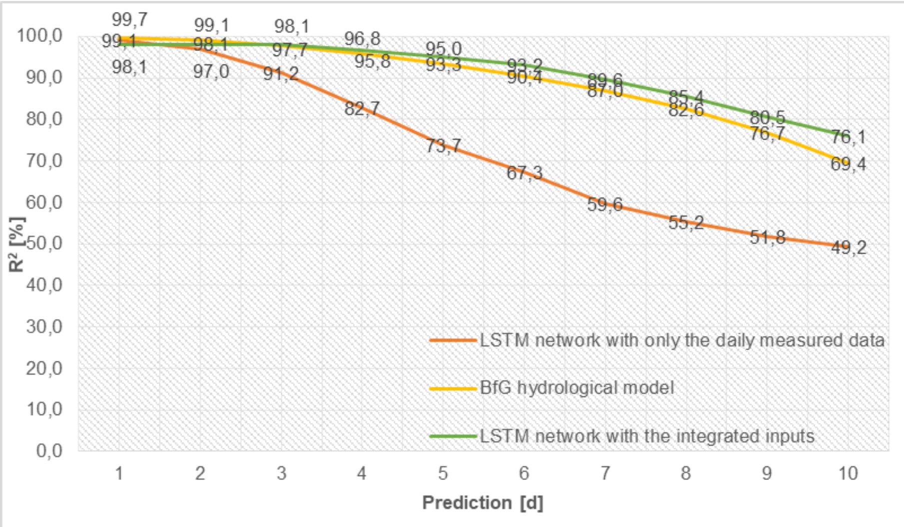

Figure 8. Coefficient of determination R^2 comparing the different models (Target: Cologne)

5.3 DSA field test

The DSA field test started in July 2018 and ended in December 2018. The project partner BearingPoint developed a demonstrator and made available to the skippers for 6-months testing, feedbacks and improvements. During this period, numerous tests have been conducted in parallel testing the LSTM network with the integrated inputs, in order to prove its performance on real-time water level predictions, evaluating their differences from the water level measurements. Due to the 2018 European heat wave, Summer 2018 was very dry and characterized by extreme temperatures for Germany, as well as a significant lack of rainfall, leading to the lowest historical water level in the gauge Kaub (downstream to Oestrich, Figure 3) and Cologne in the month of October. Despite the exceptional dataset compared to the past values, the LSTM network with the integrated inputs produced promising results during the DSA field test, where the better predictions incurred at the beginning of summer and the worse ones were in October. E.g. considering the forecasting period from 12th to 21st July 2018 (reported in Table 2), the absolute differences between the predictions and the measurements at Oestrich and Cologne were within 22 cm. Clearly, the hydrological model chain of the BfG delivered better results, because continuously updated and linked with the newest information (e.g. weather). The LSTM model with the integrated inputs can be further improved by e.g. including Summer 2018 in its training set.

Table 2. Evaluation of the DSA field test from 12th to 21st July 2018

Target: Oestrich										
All values in [cm]	12 July	13 July	14 July	15 July	16 July	17 July	18 July	19 July	20 July	21 July
Measurements	140	133	125	116	116	119	116	117	117	114
LSTM+BfG model	141	137	132	129	123	122	121	122	125	129
Meas. – LSTM+BfG	-1	-4	-7	-13	-7	-3	-5	-5	-8	-15
BfG model	133	124	120	118	115	113	111	110	108	106
Meas. - BfG	7	9	5	-2	1	7	6	7	9	8
Target: Cologne										
All values in [cm]	12 July	13 July	14 July	15 July	16 July	17 July	18 July	19 July	20 July	21 July
Measurements	200	188	183	168	160	158	159	155	157	153
LSTM+BfG model	196	192	184	178	172	169	165	166	169	175
Meas. – LSTM+BfG	4	-4	-1	-10	-12	-11	-6	-11	-12	-22
BfG model	193	183	170	163	159	155	153	151	148	99
Meas. - BfG	7	5	13	5	1	3	6	4	9	22

6 CONCLUSION AND OUTLOOK

The current German waterways require an improvement concerning inland navigation logistics and management, in order to mitigate the significant increase in traffic expected by the year 2030 (BMVI, 2016). In this context, contributing to the BMVI-funded project Digital Skipper Assistant (DSA), this study had the motivation to explore the competence of ANNs in forecasting water levels up to 10-days ahead at some strategic gauges among the German section of the Rhine River. The basic idea is the determination of a target gauge based on its previous information (in time) and on some relevant upstream gauges, mainly focusing on water levels. A multi-outputs LSTM network was developed in this study, demonstrating the promising performance of LSTM networks in identifying long-term dependencies. The hydrological data and predictions were provided by the Federal Institute of Hydrology in Germany (BfG), also DSA project partner.

In this study, the focus was on two important gauges along the Rhine River Basin: Oestrich and Cologne. The better performing LSTM architecture was established for Oestrich and then applied to Cologne as well, considering the direct physical relationship between upstream and downstream in a river system. The results showed an impressive improvement, when not only the water level measurements were used as inputs, but also the predictions of the BfG. As the network was optimized to produce the best results at Oestrich, its outputs were better performing than the ones of Cologne. Nevertheless, the developed model still delivered reliable forecasts at Cologne, having the highest R^2 for the water level predictions with a leading time from 3 to 10 days. In addition, several tests were conducted on real-time forecasts during the DSA field test from July to December 2018. Influenced by the extremely dry and hot weather conditions in Europe, the water levels in Summer 2018 were significantly different compared to the historical ones used for the model training. However, the LSTM network with the integrated inputs gave promising results; although, a certain improvement would occur when including Summer 2018 in the training set of the model.

To conclude, this work shows that the implemented LSTM network integrating as additional inputs the forecasts of the BfG hydrological model is able to predict water levels not only at Oestrich, but also at other gauges along the Rhine River system. Furthermore, the developed network is also applicable for real-time predictions. In further research, the accuracy of the network can still be increased, e.g. by updating the training set or including further information as inputs, such as precipitation data.

ACKNOWLEDGEMENTS

The research reported herein contributed to the project Digital Skipper Assistant (DSA), which was one of the mFUND projects funded by the Federal Ministry of Transport and Digital Infrastructure (BMVI). The authors would like to thank the staff in BearingPoint GmbH, BearingPoint Technologie GmbH and the Federal Institute of Hydrology (BfG) for precious support in this research and the fruitful collaboration.

REFERENCES

Alvisi, S., Mascellani, G., Franchini, M. and Bárdossy, A. (2006). Water level forecasting through fuzzy logic and artificial neural network approaches. *Hydrology and Earth System Sciences*, 10(1), 1-17.

BMVI, German Federal Ministry of Transport and Digital Infrastructure (2016) *Bundesverkehrswegeplan 2030 (Federal Transport Infrastructure Plan, BVWP 2030)*; URL: https://www.bmvi.de/SharedDocs/DE/Anlage /VerkehrUndMobilitaet/BVWP/bvwp-2030-gesamtplan.pdf?__blob=publicationFile [Accessed 26 July 2018].

Campolo, M., Andreussi, P. and Soldati, A. (1999). River flood forecasting with a neural network model. *Water Resources Research*, 35(4), 1191-1197.

Chang, F., Chen, P., Lu, Y., Huang, E. and Chang, K. (2014). Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control. *Journal of Hydrology*, 517, 836-846.

Cioc, M. (2002). *The Rhine*. Seattle: University of Washington Press.

Crespo, J., Mora, E. (1993). Drought estimation with neural networks. *Advances in Engineering Software*, 18(3), 167-170.

Devia, G., Ganasri, B. and Dwarakish, G. (2015). A Review on Hydrological Models. *Aquatic Procedia*, 4, 1001-1007.

Gers, F., Schmidhuber, J. and Cummins, F. (2000). Learning to Forget: Continual Prediction with LSTM. *Neural Computation*, 12(10), 2451-2471.

Govindaraju, R. (2000). Artificial Neural Networks in Hydrology. I: Preliminary Concepts. *Journal of Hydrologic Engineering*, 5(2), 115-123.

Haykin, S. (2009). *Neural networks and learning machines*. 3rd ed. New York: Prentice Hall, 1-2.

Hinkelmann, R. (2005). *Efficient Numerical Methods and Information-Processing Techniques for Modeling Hydro- and Environmental Systems*. Springer-Verlag Berlin and Heidelberg & Co. KG.

Hochreiter, S. and Schmidhuber, J. (1997). LONG SHORT-TERM MEMORY. *Neural Computation*, 9(8), 1735-1780.

Karim, M. and Rivera, S. (1992). Comparison of feed-forward and recurrent neural networks for bioprocess state estimation. *Computers & Chemical Engineering*, 16, 369-377.

Liong, S., Lim, W. and Paudyal, G. (2000). River Stage Forecasting in Bangladesh: Neural Network Approach. *Journal of Computing in Civil Engineering*, 14(1), 1-8.

Ma, Y., Matta, E., Meißner, D., Richter, J., Schellenberg, H., Hinkelmann, R. (2018). Forecasting water levels for the German waterways of the Rhine River using artificial neural networks and physically based models. *In TERENO International Conference October 2018; Berlin, Germany*, 2018.

Meißner, D., Schellenberg, H., Tretter, G., Matta, E. (2018). Linking operational hydrological forecasts with a route guidance system to support inland navigation. *In EGU General Assembly April 2018; Vienna, Austria*, 2018.

Nagesh Kumar, D., Srinivasa Raju, K. and Sathish, T. (2004). River Flow Forecasting using Recurrent Neural Networks. *Water Resources Management*, 18(2), 143-161.
Olah, C. (2015). *Understanding LSTM Networks -- colah's blog*. [online] Colah.github.io. Available at: http://colah.github.io/posts/2015-08-Understanding-LSTMs/ [Accessed 16 Jun. 2018].
Skymind (2017). *A Beginner's Guide to Recurrent Networks and LSTMs - Deeplearning4j: Open-source, Distributed Deep Learning for the JVM*. [online] Deeplearning4j.org. Available at: https://deeplearning4j.org /lstm.html#feedforward [Accessed 18 Jun. 2018].
Stüber M., Gemmar P. (1997). An approach for data analysis and forecasting with neuro fuzzy systems — Demonstrated on flood events at river Mosel. In: Reusch B. (eds) Computational Intelligence Theory and Applications. Fuzzy Days 1997. *Lecture Notes in Computer Science*, vol 1226. Springer, Berlin, Heidelberg
Sung, J., Lee, J., Chung, I. and Heo, J. (2017). Hourly Water Level Forecasting at Tributary Affected by Main River Condition. *Water*, 9(9), 644 pp
Tanty, Rakesh, S. Desmukh, Tanweer. (2015). Application of Artificial Neural Network in Hydrology- A Review. *International Journal of Engineering and Technical Research*. V4. 10.17577/IJERTV4IS060247.
Thirumalaiah, K., Deo, M. (1998). River Stage Forecasting Using Artificial Neural Networks. *Journal of Hydrologic Engineering*, 3(1), 26-32.
toponline.org. (n.d.). *Germany's Transportation Systems*. [online] Available at: http://www.toponline.org/books /kits/welcome%20to%20germany/WTGpdf/Handout %204.pdf [Accessed 29 Apr. 2018].
WSV, Wsd-west.wsv.de. (2014). *Geografische Daten (Geographic data)*. [online] Available at: http://www.wsdwest.wsv.de/wasserstrassen/verkehrsweg_rhein/geografische_daten/index.html [Accessed 29 Apr. 2018].