8-1-2014

Computationally Efficient Lumped Floodplain Modelling Using Data-Driven Methods

Vincent Wolfs

Patrick Willems

11th International Conference on Hydroinformatics
HIC 2014, New York City, USA

COMPUTATIONALLY EFFICIENT LUMPED FLOODPLAIN MODELLING USING DATA-DRIVEN METHODS

VINCENT WOLFS, PATRICK WILLEMS
KU Leuven, Hydraulics Division, Kasteelpark Arenberg 40, BE-3001 Leuven, Belgium. E-mail: vincent.wolfs@bwk.kuleuven.be

ABSTRACT

Modelling floodplains adequately is crucial for numerous water management applications. Many of these require a large number of simulations or long term calculations, thereby necessitating the use of models with a limited calculation time. This paper investigates two different flexible data-driven methods for computationally efficient lumped floodplain modelling that predict the inundation level in the floodplain and the flow between the river and the floodplain for a given set of river stages. The flow can be calculated using a set of optimized discharge equations. Alternatively, a hybrid approach can be followed that combines physical principles with neural networks that rely on water levels to obtain flow estimates. The derived methodology was tested on a case study using simulation results from a detailed MIKE11 model that solves the full de Saint-Venant equations. The results show that both approaches deliver accurate results, although the neural networks outperform the fixed weir equations. At the same time, the simulations are more than 10.000 faster compared to the original full hydrodynamic model. The obtained floodplain models can be used for various applications that require very short computation times.

1 INTRODUCTION

Floodplains are a crucial element in river models. They cannot be neglected in most cases, since they can significantly alter the routing in the river itself by delaying and diffusing peak flows. Moreover, floodplains are often the direct subject of research, such as in flood management and damage analyses, water quality in wetlands or ecological investigations. Hence, floodplains should be incorporated in river models and the desired accuracy depends on the final applications in mind. Ideally, information from detailed digital elevation models (DEM) is combined with a full 3D solution of the de Saint-Venant equations. Such set-up is generally infeasible due to prolonged time to create such models and computational expense. Therefore, the problem is usually reduced to two dimensions in most detailed hydrodynamic models.

In the last decade, applications that require long or a large amount of simulations have gained importance, such as uncertainty analyses, real time control of structures using MPC and integrated catchment modelling. Two dimensional models that solve the full de Saint-Venant equations are still too expensive for these applications. Alternative faster river models are described extensively in literature, but the modelling of floodplains are often overlooked. In general, two different strategies are applied to model floodplains computationally more efficient. The first focuses on reducing the dimension of the problem. Willems et al. [1] presented a quasi-2D approach that strives for a trade-off between accuracy and calculation speed. Further simplification to solely 1D models often reduce accuracy significantly because such models neglect the lateral diffusion of the flood wave. The second approach involves the simplification of the used equations by omitting some terms to yield the diffusive or kinematic flow equations. Such simplified 2D schemes were tested using measurements, and analytical and physical models (see Hunter et al. [2] for a comprehensive overview). It was concluded that the use of simplified equations instead of the complete only had a small effect on the accuracy compared to the uncertainties inherent to describing the topology and roughness (e.g. Bates and De Roo [3]). An interesting approach that combines both strategies is the use of storage cells in a rigid raster in combination with simple flow relationships, such as Manning equations. Bates and De Roo [3] give an overview of different variants of this approach and propose a novel technique, named LISFLOOD-FP, although this method also has its limitations (e.g. Horrit and Bates [4]).

Most of the aforementioned approaches still suffer from too long simulation times, or miss flexibility due to a limited parameter space. This paper presents a data-driven hybrid floodplain modelling approach that can emulate the results of detailed full hydrodynamic models and measurements, but with a minimal calculation time. To reduce the computational expense as much as possible, all processes are lumped in space and only discrete and explicit relationships are used. The floodplain topology is also schematized as a series of cells. To enhance flexibility, artificial neural networks (ANNs) are employed to model the flow between these cells. Such networks have an adaptable structure and can be tailored to target values. The performance of the ANNs is compared to models that use a set of weir equations. The hybrid calculation scheme incorporates basic physical principles explicitly, such as dike levels and modules to mimic infiltration and evaporation. The presented approach fits within the conceptual river modelling framework and software proposed by Wolfs et al. [5], but are also compatible with other river routing models.

2 METHODOLOGY

2.1 Model structure

The examined floodplain is divided into multiple cells, denoted as storage areas, which are connected with each other and the river via various paths. In contrary to conventional methods which combine a fine-scaled mesh with digital elevation models, the presented methodology lumps the modelled cells on a larger scale. In this set-up, a single storage area can cover multiple hectares. Elevations in the landscape define the boundaries of a storage area. The links that describe the flow between the different storage areas and a nearby river are schematized as well. The lumped character of this topology limits the number of calculation nodes, thereby reducing the computation time of future simulations. Small-scale effects cannot be studied in this approach, since water levels are only calculated at one location in a cell. Instead, an optimal balance is sought between the required model detail and computation time.

2.2 Calculation scheme

To calculate the flow between the different elements a hybrid method is proposed that combines elementary physical principles with two different data-driven methods, namely a fixed set of weir equations and neural networks. Both rely on up- and downstream water levels as discussed hereafter (§2.3). By taking all incoming and outgoing flows of a storage cell into account, the water balance is closed and the volume present in the storage area estimated. Next, the stage is related to the volume via a hypsometric curve. Finally, a module is implemented that models the emptying of the floodplain due to infiltration and evaporation of the water in a storage area after a flood has passed. In this study, a direct relationship between the stage and water loss was explored as is often done in hydrodynamic models that do not focus on the exact timing and strategy to empty the storage area. Of course, different and more complex modules can be implemented if required. Figure 1 summarizes the entire calculation scheme.

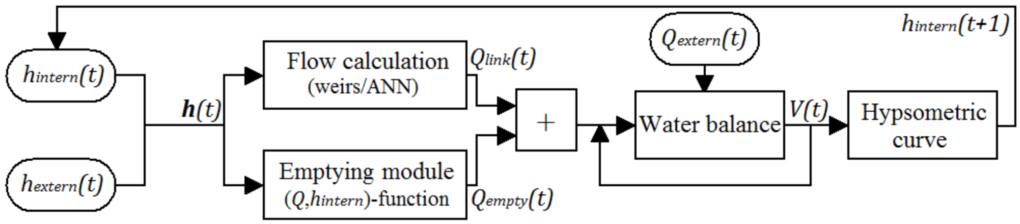


Figure 1: Calculation scheme of a floodplain cell with implemented discharge calculation over one link. Other incoming and outgoing flows are represented by the variable Q_{extern}.

2.3 Flow calculation

a. Fixed set equations

Simple equations, such as Manning-Strickler, have a limited validity and are often not flexible enough to capture the complex behavior of the flow between different volume cells. Instead, this study employs a set of equations that is deemed to describe the flow over a hydraulic structure. The equations from the hydrodynamic InfoWorks RS software (Innovyze [6]) that predict the flow over a broad crested weir were implemented with some minor corrections based on measurements. These equations are ideally suited to model the flow over in-line regular and irregular fixed weirs, as well as lateral flows over embankments. The equations (Eqs. 1 – 5) take the dike level explicitly into account, can make the distinction between free and drowned flow, and have different coefficients for positive and negative flow.

$$\begin{cases} Q = 0 & \text{if } h_{up} < dike \;\&\; h_{down} < dike & (1) \\ Q = C_1 \cdot b \cdot (h_{up} - dike)^e & \text{if } h_{up} \geq h_{down} \;\&\; h_{up} \geq dike \;\&\; (h_{down} - dike)/(h_{up} - dike) \leq m & (2) \\ Q = C_1 \cdot b \cdot (h_{up} - dike) \cdot \sqrt{\dfrac{h_{up} - h_{down}}{1-m}} & \text{if } h_{up} \geq h_{down} \;\&\; h_{up} \geq dike \;\&\; (h_{down} - dike)/(h_{up} - dike) > m & (3) \\ Q = -C_2 \cdot b \cdot (h_{down} - dike)^e & \text{if } h_{up} < h_{down} \;\&\; h_{down} \geq dike \;\&\; (h_{up} - dike)/(h_{down} - dike) \leq m & (4) \\ Q = -C_2 \cdot b \cdot (h_{down} - dike) \cdot \sqrt{\dfrac{h_{down} - h_{up}}{1-m}} & \text{if } h_{up} < h_{down} \;\&\; h_{down} \geq dike \;\&\; (h_{up} - dike)/(h_{down} - dike) > m & (5) \end{cases}$$

where Q is the discharge [m³/s], h_{up} and h_{down} are the up- and downstream stages [m], *dike* is the dike level [m], m is the modular limit [-], C are constant coefficients, e corresponds to the shape of the weir and can range from 1.5 to 2.5, and b is the width [m]. Due to its fixed model structure, optimization algorithms are only able to alter the parameters of this set. The user can of course also use different equations, including controllable structures such as gated weirs.

b. Artificial neural networks

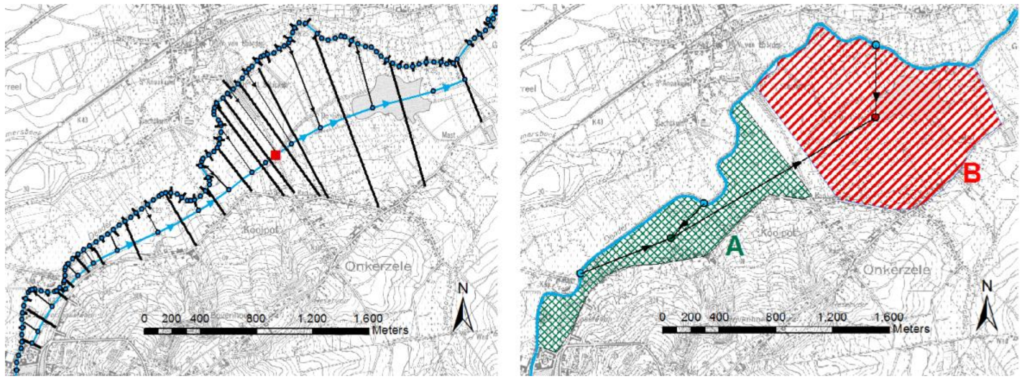
ANNs are employed increasingly in water engineering due to their flexible nature and learning capabilities. ANNs are characterized by an architecture that can adapt itself to the data presented during training. Hornik [7] stated that these networks are universal approximators, meaning that they can approximate any measurable function to any desired degree of accuracy. Despite their appealing characteristics and widespread use in applications, neural computing techniques are not a panacea. Their effectiveness and accuracy largely depends on a careful executed calibration procedure. The most critical challenge in the development of ANNs is generalization: a too simple network will fail to identify fully the underlying relationships, while an overcomplicated network can easily lead to overfitting and consequently poor predictions for unseen data. This study uses Feedforward Multilayer Perceptron networks and two different training procedures to obtain a model with optimal number of hidden nodes and good generalization capabilities.

The first training procedure is "early stopping". First, the calibration data is divided into three subsets. An ANN is trained using the Levenberg-Marquardt algorithm on the training data subset. The training is stopped when the error on the validation subset increases during a selected number of subsequent epochs. This procedure is repeated for several networks with different numbers of hidden neurons and the network with lowest error for the test set is retained. The early stopping technique is fast and easy to set-up, but the division of the data is subjective, has an impact on the results and is wasteful in its use of data. The second approach is Bayesian learning and is more sophisticated. Instead of solely minimizing the squared errors during training, the weights are also taken into account. By constraining the size of the network weights, the final response of the network should be smoother. Large weights can cause excessive variance of the output, which could lead to output values far beyond the calibration range. The Bayesian framework of MacKay [8] was used in this research to find the optimal configuration of the objective function parameters (see Foresee and Hagen [9] for a detailed description). Regardless of which training procedure is followed, one could also combine multiple trained networks. Such ANN ensembles can improve the generalization and have a smoother response. In this research, simple averaging (e.g. Lincoln and Skrzypek [10]) was employed for some networks. If the available data is too limited or uniform, the user could try to artificially create additional data. In this specific application, one could assume that the system reacts symmetrical, meaning that switching the up- and downstream water levels would only change the sign of the flow. Of course, this assumption is not valid under all circumstances and can have a significant impact on the outcome of the training algorithm.

After the ANN is created, its response is transformed into a three dimensional staggered grid: for several combinations of up- and downstream water levels, the outcome of the ANN is calculated (see Fig. 3a). This transformation allows for a straightforward implementation in different software environments, faster calculation times (since fairly sized lookup tables are computationally less expensive than calculating an ANN) and a clear and transparent visualization of the response. The latter is crucial, since each trained network should be checked for physical consistency.

3 STUDY AREA

The modelling approach was tested on a floodplain located next to the navigable river Dender in Belgium. The examined floodplain covers a surface of approximately 212 ha. Due to the lack of accurate and sufficient measurement data, simulation results from a detailed, full hydrodynamic MIKE11 model were used. The model was built using two dimensional spatial measurements of the river's and floodplains' topology. Simulation results were validated based on measurements of flow and stage and maps of recently flooded areas. Floodplains were incorporated in a quasi 2D setting to reduce calculation costs (Willems et al. [1]; Willems [11]): they are described as a network of fictitious river branches with detailed cross sections that are connected to the river via various spills. Figure 2a shows the topology of the examined floodplain in the detailed MIKE11 model. A road divides the floodplain into two parts, which is represented in the MIKE11 model as a weir.



a. b.

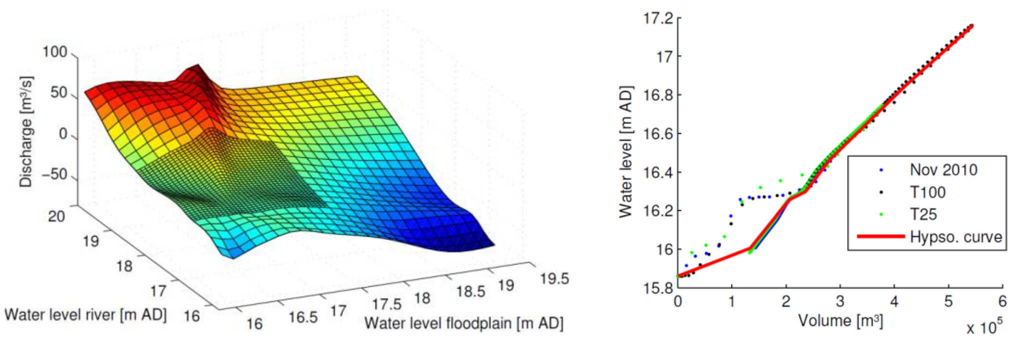
Figure 2: a. Topology of the examined floodplain of the detailed MIKE11 model. The cross-sections are indicated by the thin black lines, while the weir is shown by the red square. b. Conceptual schematization of the floodplain.

Two historical events ('Nov 2010': 01/11 - 30/11/2010; 'Jan 1995': 15/01 – 14/02/1995) and three synthetic rainfall events of one month were simulated ('T25', 'T100' and 'T250') with return periods of respectively 25, 100 and 250 years. The results were stored every 15 minutes.

4 APPLICATION AND RESULTS

The events 'Nov 2010', 'T100' and 'T25' were selected as calibration events, 'Jan 1995' and 'T250' for validation. By choosing a more extreme event for validation, the extrapolation capabilities of the obtained models could be tested. The floodplain topology includes two large storage areas and four flow links through which these cells are filled and emptied (see Fig. 2b). This set-up was chosen after an analysis of the simulation results of the MIKE11 model. For each link that connects a storage area with the river, the optimal water level location in the river was selected via trial and error: the model element of this specific link is optimized for each available water level time series in the river as described below. The location in the river that yielded the lowest R^2 value for the calibration events was retained. In each storage area, the water level location was selected that was flooded first. During the calibration of the model structure of a flow link, the dike level is estimated based on the selected water levels.

Two different models were created: one that uses solely weir equations to describe the flow through the links, and another one using only neural networks. The weir equations were optimized via an interior-point method. The ANNs were initially calibrated using the early stopping method with just 5 hidden nodes. An important element is to retain only data couples of the calibration events for training with one or both water levels exceeding the dike level. This ensures that the calibration can focus on yielding accurate non-zero flow predictions, while the initiation of the flow is explicitly determined in the calculation scheme by the dike level. This set-up also constrains the predictions, precluding flow estimations when the up- and downstream stages are below the dike level. The retained data was randomly divided into 70% training, 15% validation and 15% testing sets and rescaled to [-1,1]. For each link, 5 networks were calibrated, each with random initial weights. Since the provided training data is rather uniform in some cases, the trained networks can generate significantly different results for data outside the calibration range. Visual inspection can be required to select the network that has the most logical response. For instance, when the stage in the river is much larger than that in the floodplain, the predicted flow should be directed from the river towards the floodplain and vice versa. Such simple checks could also be incorporated in the final calculation scheme to constrain the outcome of the ANNs for data far beyond the calibration range, although this could limit the flexibility. If trained properly with diverse data, such improvements are unnecessary. This straightforward procedure failed to deliver accurate results for one of the four flow links. Therefore, ANNs were calibrated using the Bayesian learning technique (see Fig. 3a). Additional data was created by assuming symmetry (see §2.3.b). A network with 15 hidden nodes yielded accurate predictions. The number of effective parameters (69) was close to the actual number of parameters (81), but training larger networks produced similar results with the same number of effective parameters. Next, hypsometric curves were calibrated via piecewise linear relationships based on volume and stage couples (see Fig. 3b). Finally, the emptying of each storage area is modelled by assuming a relationship between the stage in the storage area and the simulated flow. This relationship will be used when the up- and downstream stages of a link are below the dike but there is still water present in the storage area.



a. b.

Figure 3: a. Grid with variable resolution that emulates the outcome of the trained ANN obtained after Bayesian learning. b. Calibrated piecewise linear hypsometric curve for storage area ‘A’ for the three calibration events.

After calibration of all individual elements, a “full simulation” can be performed, meaning that predictions rely on previously calculated values. Figure 4 shows the water levels for storage area ‘A’ (upper figure) and ‘B’ (lower figure) for simulations with a time step of 15 minutes. Table 1 summarizes all RMSE, MAE and R^2 values. To keep the results concise, only stage predictions are shown. Since these rely directly on the predicted discharge, they give a clear representation of the overall results. The performance criteria (see Table 1) and plot (Fig. 4) show that both models manage to predict the water levels in general accurately with R^2 values close to unity, although the models that use ANNs to calculate the flow consistently outperform those with weir equations. The latter clearly have difficulties in modelling the falling branch of the water level series, while the ANNs can model these adequately. The maximal stages are predicted mostly with minor deviations, although the weir equations fail to predict flow through some links for the smaller events (e.g. storage area ‘B’, ‘Nov 2010’). Notable is that the ‘T250’ event is predicted very well by both models, especially since this event lies outside of the calibration range.

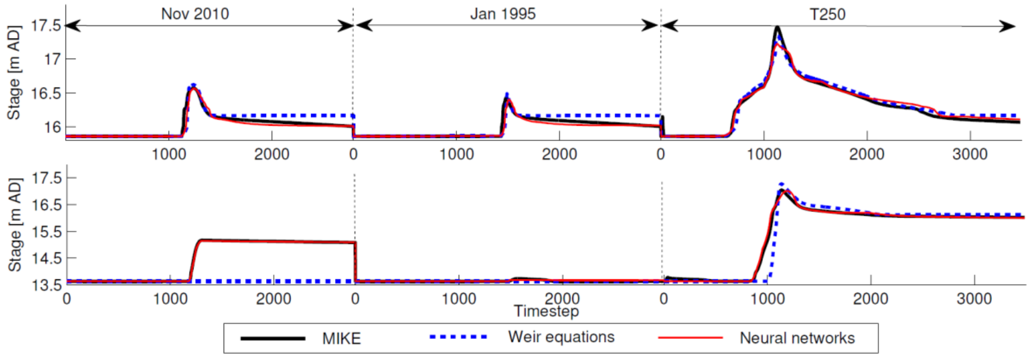


Figure 4: Water level predictions for storage areas 'A' (up) and 'B' (lower) for a calibration ('Nov 2010') and the validation events ('Jan 1995' and 'T250').

Table 1: Performance criteria for the water level predictions in storage areas 'A' and 'B' for the calibration ('Nov 2010', 'T100' and 'T25') and validation events ('Jan 1995' and 'T250'). The value before the slash symbol (/) represents the model with weir equations, the other with ANN (i.e. weir equations/ANN).

		Nov 2010	T100	T25	Jan 1995	T250
A	RMSE (m)	0.083 / 0.045	0.079 / 0.045	0.099 / 0.049	0.079 / 0.043	0.061 / 0.051
	MAE (m)	0.055 / 0.024	0.058 / 0.026	0.071 / 0.026	0.051 / 0.022	0.047 / 0.034
	R^2	0.746 / 0.926	0.927 / 0.977	0.701 / 0.925	0.622 / 0.890	0.970 / 0.979
B	RMSE (m)	1.112 / 0.013	0.215 / 0.021	0.914 / 0.019	0.027 / 0.028	0.315 / 0.071
	MAE (m)	0.837 / 0.007	0.098 / 0.012	0.756 / 0.010	0.010 / 0.020	0.149 / 0.040
	R^2	0.000 / 1.000	0.964 / 1.000	0.336 / 1.000	0.000 / 0.000	0.920 / 0.996

5 COMPUTATIONAL EFFICIENCY AND APPLICATIONS

The entire modelling approach, from the extraction of data from the simulation results to the final build-up of the model in C programming language was implemented in the modelling framework presented by Wolfs et al. [5]. To improve the computational efficiency, the grids that emulate the neural networks can have a variable resolution (see e.g. Fig. 3a): erratic spaces can be incorporated with more detail, while more smooth areas are implemented with a larger grid size. The calculation time amounts $3.7 \cdot 10^{-4}$ seconds for simulating a one-month event (averaged over 200 runs) for the model with weir equations and $4.1 \cdot 10^{-4}$ seconds for the ANN on an average laptop with i3 processor using a single core. This is more than 10.000 times faster than a comparable MIKE11 model that solves the complete de Saint-Venant equations.

These models with very short calculation time are ideally suited for applications that require a large amount of or long term simulations. The presented modelling approach was already used in models for uncertainty flood mapping (Wolfs et al. [12]), real time control of hydraulic structures to prevent flooding (Van den Zegel et al. [13]; Chiang and Willems [14]) and impact analyses (De Vleeschauwer et al. [15]).

6 CONCLUSIONS

This research presents a lumped and computationally very efficient modelling approach for floodplains. Flows are predicted via a hybrid method that relies on fixed weir equations and/or artificial neural networks (ANNs). The presented methodology is tested on simulation results of a detailed MIKE11 model of the Dender River in Belgium. Performance criteria show that the stages can be emulated accurately, while the calculation time is more than 10.000 times shorter than a similar MIKE11 model that solves the full de Saint-Venant equations. The models involving ANNs outperform those with weir equations due to their enhanced flexibility and learning ability. Special attention is paid to the generalization capability and physical soundness of the model by applying a systematic model structure identification and calibration procedure.

This research fits within a larger project to set up a framework for conceptual modelling of river and sewer systems. These models can be used for numerous applications that require very fast or a large amount of simulations.

7 ACKNOWLEDGEMENTS

This research was supported by the Agency for Innovation by Science and Technology in Flanders (IWT). The authors would like to thank DHI for the MIKE11 software license.

8 REFERENCES

[1] Willems P., Vaes G., Popa C., Timbe L. and Berlamont J. Quasi 2D river flood modelling. River Flow 2002. In: International Conference on Fluvial Hydraulics. Louvain-La-Neuve, Belgium, 4-6 September 2002, pp. 1253-1259, (2002).

[2] Hunter N.M., Bates P.D., Horritt M.S., and Wilson M.D. Simple spatially-distributed models for predicting flood inundation: A review. Geomorphology 90, 208-225, (2007).

[3] Bates P.D. and De Roo A.P.J. A simple raster-based model for flood inundation simulation. Journal of Hydrology 236 (1-2), pp. 54-77, (2000).

[4] Horritt M.S. and Bates, P.D. Evaluation of 1-D and 2-D numerical models for predicting river flood inundation. Journal of Hydrology 268, pp. 87-99, (2002).

[5] Wolfs V., Meert P. and Willems P. Semi-automatic conceptual model identification and calibration tool for River hydraulic computations. International Conference on Fluvial Hydraulics, River Flow 2014, Sept. 3-5, 2014, Lausanne. (*submitted*)

[6] Innovyze. InfoWorks-RS help documentation [v 13.5], Innovyze, Oxfordshire, UK, (2014).

[7] Hornik K. Approximation Capabilities of Multilayer Feedforward Networks. Neural Networks 4(2), pp. 251-257. (1991)

[8] MacKay D.J.C. Bayesian Interpolation. Neural Computation 4, 415-447 (1992).

[9] Foresee F.D. and Hagan, M.T. Gauss-Newton approximation to Bayesian learning. Proceedings of the 1997 International Conference on Neural Networks. (1997).

[10] Lincoln W. and Skrzypek J. Synergy of clustering multiple back propagation networks. Advances in Neural Information Processing Systems 2, 650-657 (1990).

[11] Willems P. River modelling (in Dutch). Acco Leuven & Den Haag, 268 p., ISBN 978-90-334-9296-9, D/2013/0543/132, NUR 955. (2013)

[12] Wolfs V., Van Steenbergen N. and Willems P. Flood probability mapping by means of conceptual modeling. In Muñoz, R. (Ed.). River Flow 2012: Vol. 2. International Conference on Fluvial Hydraulics. Costa Rica, 5-7 Sept. 2012, pp. 1081-1085. (2012)

[13] Van den Zegel B., Vermuyten E., Wolfs V., Meert P. and Willems, P. Real-time control of floods along the Demer River, Belgium, by means of MPC in combination with GA and a fast conceptual model. 11th International Conference on Hydroinformatics, HIC 2014.

[14] Chiang P.-K. and Willems, P. Model predictive control combined with Genetic Algorithms for a river system. 11th International Conference on Hydroinformatics, HIC 2014.

[15] De Vleeschauwer K., Weustenraad J., Wolfs V. and Willems P. Comparison of the up-versus downstream storage on a sewer-river system: a quantitative case study. Proceedings of the 35th IAHR World Congress, Chengdu, China, 8-13 Sept. 2013 (art. nr. A10783).