Data-driven modeling for river flood forecasting based on a piecewise linear ARX system identification

B. Hadid, E. Duviella, S. Lecoeuche

Version of Record: https://www.sciencedirect.com/science/article/pii/S0959152419304688
Manuscript_092dc7b25ad74a8c1014e2eff9d50060

Data-driven modeling for river flood forecasting based on a Piecewise Linear ARX system identification ⋆

Baya Hadid[a,*], Eric Duviella[a], Stéphane Lecoeuche[a]

[a] *IMT Lille Douai, University of Lille, Informatics and Automatics Research Unit, F-59000 Lille, France.*

Abstract

Most of the studies related to the rainfall-runoff modeling of rivers consist of data-driven models, given that the corresponding physical modeling approaches are based on a thorough geological knowledge of the river in addition to a time consuming simulation. Indeed, flood forecasting services have the difficult task of avoiding natural and human disasters and choose for that to use input-output or grey box models for their simplicity and easy calibration updates. However, these models are not evolving according to the variations of environmental conditions or need at least the evapotranspiration and the soil humidity measurements in addition to the rainfall quantity. This paper gives an alternative approach to the existing rainfall/runoff linear and nonlinear models by the utilization of a hybrid system consisting in a Piecewise Auto-Regressive eXogeneous (PWARX) structure identified using

⋆This work has been supported by ARMINES through the contract n° 1908V/1700661 and the IIW (Institution Intercommunale des Wateringues). Authors also thank the DREAL Hauts-De-France flood forecasting services (SPC) for providing access to data. Conflict of interest - none declared

*Corresponding author

Email addresses: baya.hadid@imt-lille-douai.fr (Baya Hadid), eric.duviella@imt-lille-douai.fr (Eric Duviella), stephane.lecoeuche@imt-lille-douai.fr (Stéphane Lecoeuche)

an approach that alternates between data assignment and parameter estimation. The usage of this special kind of nonlinear systems bears a potential to handle the nonlinearities and varying-time delays mainly induced by the soil water storage and evapotranspiration.
Keywords: Rainfall-Runoff model, Hybrid system, Data-driven model, Non-supervised clustering, Data assignment

1. Introduction

Nonlinearities are the biggest challenge faced when one seeks for a complex large-scale natural system model that has to be reproducible, especially when we have a limited knowledge of the intrinsic phenomena governing the system and when this system is virtually unique, for example in terms of geophysical properties. This is particularly true when it comes to model the rainfall-runoff relationship of a river for a precise flood forecast with having as an only knowledge the precipitation past measurements and forecast and the runoff measurements. Indeed, the generation process of a flood starting from precipitations in a natural catchment in order to deliver an early warning for extreme hydrological events is one of the biggest challenges faced by hydrologists. [1] assessed in its study at least 19 existing daily rainfall-runoff models, leading rapidly to a bigger number of models when it comes to take into account the seasonality and the climate change we are currently undergoing. The reason is that the stream level or discharge is not similarly impacted by the same amount of rainfall according to seasons due to the influence of the temperature on the evapotranpiration of the ground and the vegetative cover. Other factors affecting the runoff come into play such as

the soil permeability, the river slope, the drainage due to agriculture activity, factors that differ from a river to another.

The common goal of all the rainfall-runoff modeling approaches is to provide a flood forecast with a minimum lead-time (time between its announcement and arrival). The overwhelming majority of the proposed approaches are either physical/mathematical approaches, data-based alternately named black box approaches, and conceptual approaches. Mathematical/hydrological modeling is known to be time consuming due to the huge number of parameters to calibrate and it is thus critical to use for an online forecast. In addition to the computational burden, it requires a non negligible amount of detailed information on each concerned river such as a bathymetric survey.

On the contrary, data-driven, black box fully numerical modelling approaches establish models by using only input and output measurements. Many researchers have developed numerical runoff/rainfall models with varying degrees of success. A good survey on non parametric data driven modeling techniques is presented in [2]. Neural networks (ANNs), genetic programming, evolutionary polynomial regression, Support vector machines (SVM), M5 model trees, K-nearest neighbors, and multiple linear regression techniques are implemented and evaluated using daily stream flow data. The SVM approach was also explored in [3], where a short-term stream flow prediction was performed on hourly data. Other approaches based on the neural networks were developed by for instance, introducing an *a priori* knowledge on the evapotranspiration (see [4] and references therein). Neural Fuzzy Networks (NFS) and preciseley ANFIS (Adaptive Network-based Fuzzy Inference Systems) were also applied through the years ([5], wavelet neuro-

fuzzy models [6], and radial basis function artificial neural network extreme training on monthly data in [7]) but still suffer from a non flexible structure that depend on the expert's knowledge. Recent studies use adaptive fuzzy-techniques for a more flexible structure achievement, such as the Self-adaptive Fuzzy Inference Network (SaFIN) presented in [8]. However, and despite respectable results provided by Machine learning techniques, they still need a huge database to achieve a correct training and validation results.

In the family of black box parametric approaches, the linear models consisting in ARX and ARMAX models were firstly used due to their simplicity [9, 10]. These methods have been abandoned because of the non linearities due to, as examples, evapotranspiration phenomenon and saturation due to soil storage. Recently, a black box Linear Parameter Varying (LPV) model was investigated for the Rainfall-Runoff Relationship (RRR) in urban drainage networks [11] and rural catchment [12]. This kind of systems consider a lot of nonlinearities and depend on one or several external variables, called scheduling variables, and then could be linearized at different operating points resulting in a set of local Linear Time-Invariant (LTI) systems. The issue with this approach comes from how to choose the right scheduling variable which is not trivial and could be a non observable variable or unknown measurement. In [11], the scheduling variable was chosen as the output of a non parametric model and the model was identified using the least-squares algorithm when the scheduling parameter is taken as the output of the best linear model in [12] but the optimal identification Simplified Refined Instrument Variable (SRIV) algorithm was applied. Both LPV methods lead to acceptable results. In [13], the authors proposed an online recursive nonlinear

identification algorithm applied to Liane river (France) and was compared to a recursive least-squares linear model over a future horizon of 24 hours. Indeed, the online estimation allows to track the catchment intrinsic variations by updating at each time-step the model parameters. The study over a horizon is innovative comparing to the previous cited approaches and shows that even the Fit score [14] or the Nash coefficient [15] are good, the introduction of a prediction horizon deteriorates the estimation results and then increases the number of false alarms and missed alarms.

In [16], a conceptual daily lumped rainfall/runoff model called GR4J (from the french *"Génie Rural à 4 paramètres Journaliers"*) and its hourly version GR4H, is presented as an improvement of the GR3J [17, 18] and the performance was tested using five criteria. This conceptual model is based on a reservoir system and its initialization requires the knowledge of the soil saturation of the previous day. This approach has shown its reliability once it was properly calibrated using a large amount of data and was applied on several case studies [19, 20, 21].

Hydromax tool used currently as a part of the Belgium forecasting system and presented in [22] represents the nonlinearities by a Hammerstein structure using an *a priori* knowledge of the hydrological system to characterize the static function on the river basin maximum storage, the runoff coefficient and the percolation parameter. An evapotranspiration estimation is also required to calculate the effective rainfall which is the final input of a predictive ARX model. Even if this approach is effective for important flow forecasts, a larger prediction error is observed for low water. The authors propose a combination or a switch between a linear ARX prediction model

and nonlinear ARX (NARX).

In this paper, a rainfall-river level PieceWise Affine model (PWA) is proposed as a new solution to handle the maximum of nonlinearities. PWA systems falls into the category of Hybrid systems. The term "hybrid" is related to the interaction between continuous and discrete dynamics. The discrete state sequence could be a result of the interaction between logic devices and continuous processes or a physical phenomena that shows a discontinuous behavior. A hybrid system could also be the result of an approximation of a nonlinear system by a set of linear submodels. PWA model has the property of being universal due to its particular mapping and make this kind of systems particularly attractive for nonlinear dynamical system identification [23, 24].

The principle is to consider the model as a combination of different submodels. Each submodel is assimilated to a state or a mode and the switching between the states depends for example on the change in the operating conditions. However, it is not obvious to know *a priori* the number of modes and the state sequence, especially in case of natural large-scale systems. Indeed, it depends on the knowledge degree of our system and even so, there is no guarantee to achieve satisfactory prediction performances. The literature proposes multiple approaches of PWA estimation. These approaches can be based on Algebraic procedures[25, 26], Bayesian procedures [27], Bounded-error (set-membership) procedures [28], Clustering procedures [29, 30, 31], Optimization-based procedures [32, 33]Recursive procedures [34, 35, 36, 37]. Since the objective is to model the rainfall-runoff relationship without *a priori* knowledge on the physical system using real data provided by gauges,

it then becomes important to choose carefully the right method that makes possible a PWA approximation with a minimum knowledge on the structure especially on the number of modes. We prefer among them the unsupervised techniques and precisely, the PWA model identification algorithm proposed by [31]. This algorithm is an unsupervised clustering technique based on the theory of evidence introduced by [38] and explored later by [39] for clustering purposes on the basis of the evidence of nearby observation samples called the k-nearest neighbors according to the euclidean distance between data. It solves simultaneously the problem of data assignment to discrete modes, the parameter estimation of submodel dynamics and the number of possible regimes or modes, thanks to an evidential procedure. The unsupervised feature of this approach allows therefore to do not make any assumption on the total number of operating modes, only the linear model orders are fixed *a priori*.

In the algorithm scheme, each data is assigned and the number of submodels is reduced at each iteration thanks to a criteria that combines the prediction error and the Euclidean distance between data samples, and then the parameters of each submodel are updated. The problem of the PWA model estimation is thus solved by clustering data and estimating a parameter vector using data contained in each class and an identification technique of a linear model such as the least squares algorithm, with ensuring a minimum number of submodels that gives the better performances. This algorithm is therefore an alternation between data assignment and parameter adaptation.

This paper is organized as follows: Section 2 introduces the PWA identification problem statement with considering a short-term and a long-term

prediction horizon, states the Evidential algorithm principle and details the Dempster-Shafer rules used for data reassignment. The problem of mode sequence forecasting with a lead time considering validation data is addressed in Section 3 through the prism of the region estimation based on machine learning techniques. The water level forecasting results of a river located in the north of France are reported in Section 4. Finally, Section 5 summarizes the key findings and foreshadows possible improvements in future research.

2. PWA with a prediction horizon identification problem statement

The basic principle of the proposed method is to put together data that are more likely to have been generated by the same and underlying linear submodel. For each group of data, we can hence estimate one parameter vector to represent the corresponding submodel. The method is based on a unsupervised clustering by alternating the regression data assignment to a submodel and the submodel estimation.

2.1. Problem description

Let us consider the following PieceWise AutoRegressive eXogenous (PWARX) model with a prediction horizon H

$$y\left(k\right) = f_{\sigma_k}^{(k)}\left(\varphi_H\left(k\right)\right) + e\left(k\right), \tag{1}$$

σ_k is the discrete state $\sigma_k \in \{1, \ldots, s\}$ and $\varphi_H(k)$ is the regression vector having the following structure

$$\varphi_H(k) = [\, y(k-H-1) \;\cdots\; y(k-H-n_a)\; u(k)\; u(k-1) \;\cdots\; u(k-H-n_b)\,]^\top, \tag{2}$$

where $u(k) \in \mathbb{R}^{n_u}$ and $y(k) \in \mathbb{R}^{n_y}$ are respectively the input and the measured output of the system at time $k \in \mathbb{Z}$, n_a and n_b are the orders of the discrete-time ARX model transfer function and H is the prediction horizon. $e(k)$ is a noise/error term regarded as a zero-mean Gaussian noise with a standard deviation σ^2. For the sake of exposition and according to the case study, only the *Single Input Single Output* (SISO) case is discussed in the following ($n_u = n_y = 1$). $\mathcal{F}$ is a piecewise linear map of the form

$$\mathcal{F}^{(k)}_{\sigma_k}(\varphi_H) = \begin{cases} \theta_1^\top \bar{\varphi}_H(k) & \text{if } \sigma_k = 1 \\ \;\vdots & \\ \theta_s^\top \bar{\varphi}_H(k) & \text{if } \sigma_k = s, \end{cases} \tag{3}$$

where θ_{σ_k} is the parameter vector defining the linear submodel $\mathcal{M}_{\sigma_k}$ and $\bar{\varphi}_H(k) = \left[\varphi_H(k)^\top\; 1\right]^\top$ is the extended regression vector.

σ_k is described by a polyhedral partition of the regression space and is then defined by

$$\sigma_k = i \quad \text{iff} \quad \varphi_H(k) \in \mathcal{R}_i, \quad i = 1, \ldots, s. \tag{4}$$

The regions $\{\mathcal{R}_i\}_{i=1}^s$ form a complete polyhedral partition of the regression space $\mathcal{R} \subset \mathbb{R}^n$ where $n = n_a + n_b + 1$. Each distinct region is described by

$$\mathcal{R}_i = \{\varphi_H(k) \in \mathbb{R}^n : \mathcal{H}_i \bar{\varphi} \leq \mathbf{0}\}. \tag{5}$$

$\mathcal{H}_i \in \mathbb{R}^{(n+1)}$ is a function defining the separating boundaries delimiting the region $\mathcal{R}_i$. Finally, the piecewise affine map $\mathcal{F}$ can be rewritten as follows

$$\mathcal{F}_{\sigma_k}^{(k)}\left(\varphi_H\right) = \begin{cases} \theta_1^\top \bar{\varphi}_H(k) & \text{if } \mathcal{H}_1 \bar{\varphi}_H \leq \mathbf{0} \\ \vdots & \\ \theta_s^\top \bar{\varphi}_H(k) & \text{if } \mathcal{H}_s \bar{\varphi}_H \leq \mathbf{0}. \end{cases} \tag{6}$$

The PWARX model identification consists in estimating the parameter vectors $\{\theta_i\}_{i=1}^s$, given a set of N observations $\{\varphi_H(k)^\top, y(k)\}_{k=1}^N$ generated by the PWARX system defined by (1) and the boundaries' parameters of the partitions $\{\mathcal{R}_i\}_{i=1}^s$ under the assumption that the discrete mode sequence $\{\sigma_k\}_{k=1}^N$ and the number of submodels s are unknown.

PWARX model estimation is classically used with a null prediction horizon $H = 0$ and the algorithm is a one-step ahead prediction model. Since the objective is to forecast the output at a future time $k+f$ at time k, noted $k+f|k$ with $f \geq 1$, the prediction at time $k+f$ with a null prediction horizon is given by

$$\hat{y}\left(k+f|k\right) = \hat{\theta}_i^\top \hat{\varphi}_0\left(k+f|k\right) \quad i = 1, \ldots, s, \tag{7}$$

$$\begin{aligned} \hat{\varphi}_0\left(k+f|k\right) = [\ \hat{y}(k+f-1|k)\ \cdots\ \hat{y}(k+f-n_a|k) \\ u(k+f)\ \cdots\ u(k+f-n_b)\ ]^\top, \end{aligned} \tag{8}$$

where $\hat{\varphi}_0$ is the regression vector containing the estimated output $\hat{y}$ instead of the real output y. We note here that $\hat{\varphi}_0$ contains the predicted outputs

previously and repeatedly estimated using the forecasted inputs $u(k+1)$ to $u(k+f-1)$ to simulate the model output using $\hat{y}(k) = \hat{\theta}_i^\top \hat{\varphi}_0(k)$. The repetition of this operation causes an error propagation. To achieve directly $\hat{y}\,(k+f|k)$, a prediction horizon equal to $H = f$ is used. The inconvenient is that the model is more complex due to the higher number of parameters to estimate. The predicted output is then defined by

$$\hat{y}\,(k+f|k) = \hat{\theta}_i^\top \varphi_H\,(k+f) \quad i = 1, \ldots, s, \tag{9}$$

$$\varphi_H\,(k+f) = [\; y(k-1) \;\cdots\; y(k-n_a)\; u(k+f) \;\cdots\; u(k-n_b)\;]^\top. \tag{10}$$

Increasing the prediction horizon leads then to an increasing of the regression vector dimension. However, the regression vector is commonly used as feature vectors to estimate the regions boundaries based on a supervised machine learning technique (see Section 3). It is yet recommended to keep the dimension of feature vectors as low as possible to avoid simulation time issues while ensuring that the clustering performances are not compromised due to a poor dimension. On the other hand, increasing the prediction horizon limits the effects of the error propagation. A high training precision remains however a necessity. This is possible thanks to a good value selection of number of nearest neighbors which is the main tuning parameter of the evidential algorithm defined in Section 2.2.

Based on this, we define first a new prediction horizon H' where $1 \leq H' < H$ in Eq. (2) selected depending on the system dynamic and more

specifically smaller than the response time of the system. The predicted output is then simulated as for the case $H = 0$, *i.e.* applying repeatedly the operation $\hat{y}(k + H') = \hat{\theta}_i^\top \hat{\varphi}_{H'}(k + H')$ with

$$\hat{\varphi}_{H'}\,(k + H') = [\ \hat{y}(k-1|k)\ \cdots\ \hat{y}(k-n_a|k)\ u(k+H')\ \cdots\ u(k+H'-n_b)\ ]^\top. \tag{11}$$

Note that at each new prediction, the real values of the output are assimilated by replacing the estimated output by its real value in the regression vector $\varphi_{H'}\,(k + H')$.

2.2. Algorithm principle and initialization

The approach is based on an unsupervised clustering combined with a linear regression to merge regression data belonging to the same submodel. In unsupervised clustering techniques, the class labels/indexes of regression data are not known *a priori*. They hence need a clustering rule or a similarity criterion to build the data partitioning.

The clustering based procedures consider that the local linearity characterizing the PWA map can be reconstructed using a resemblance between a submodel parameter vector and an estimated parameter vector generated from small local data sets. Each algorithm iteration consists of an alternation between regression data assignment to a class built using the unsupervised clustering and a submodel parameter estimation for this new updated class using identification approaches dedicated to linear ARX systems.

Since the number of submodels is unknown *a priori*, regression data are first partitioned into N clusters $\mathcal{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_N\}$, where $\mathcal{C}_i = \{X(i)\}$. We assign the initial parameter matrix $\Theta^{(0)} = \left[\hat{\theta}_1^{(0)}, \ldots, \hat{\theta}_N^{(0)}\right]$ to these N clusters.

The objective is to reduce the number of clusters at each iteration and to converge to s models by means of a decision rule that determine how data migrate towards the most representative clusters. Thanks to the adaptation procedure, the less representative clusters become empty and are eliminated resulting in a decreasing number of clusters. $\hat{\theta}_i$ is computed by considering the data of the cluster $\mathcal{C}_i$ and its "c" nearest neighbors and by using a *least-squares* (LS) on these $c+1$ data. Each data is achieved by concatenating the regression vector and the observed output $X(i) = \left[\varphi(i)^\top, y(i)\right]^\top$, $i = 1, \ldots, N$.

We briefly recall that a LS estimation based on an observed data sequence $\{u(k), y(k)\}_{k=1}^{N_i}$ of the parameter vector defining the ARX system is given by

$$\hat{\theta}_i = \left(\sum_{k=1}^{N_i} \bar{\varphi}_H(k) \bar{\varphi}_H^\top(k) \right)^{-1} \sum_{k=1}^{N_i} \bar{\varphi}_H(k) y(k). \tag{12}$$

The tuning parameter c value is a compromise between the noise level and the risk of achieving a number of mixed submodels generated from mixed local data belonging to different submodels. If the noise level is high, its effect is filtered by using a large value of c. In the case of rainfall/runoff observations provided by gauges, the effect of noise is weak if not non-existent. A small c is then recommended to avoid the problem of mixed submodels. In addition, a small c value results in a higher number of submodels leading inevitably to higher performances. Numerical examples show that beyond a certain value, the number of submodels converges to the real s beyond a certain value [31].

2.3. Data reassignement based on Dempster-Shafer theory

Dempster-Shafer theory, also called theory of Belief functions, is used to model information uncertainty [38]. This theory is thus based on the belief functions that allows reasoning on uncertain facts, not relying on probabilistic quantification, but in a more general way than the Bayesian model. The *Transferable Belief Model* (TBM) represents the quantified beliefs of an agent and was subject to varying interpretations [40]. The Dempster-Shafer theory is composed of a first level called *credal* where the beliefs are quantified and merged and a second level called *pignistic* which comes from the Latin *pignus* which means literally *a bet* where decisions are made by a transformation of the belief functions to probability functions.

Let $\mathcal{C}$ a set of propositions called a frame of discernment and A a subset of $\mathcal{C}$, $A \subseteq \mathcal{C}$, $m(A)$ is a part of the belief in the membership of the agent X to the subset A. $m : \mathcal{C} \rightarrow [0, 1]$ is defined as follows

$$m(\emptyset) = 0\,, \tag{13a}$$

$$\sum_{A \subseteq \mathcal{C}} m(A) = 1. \tag{13b}$$

The $m(A)$ values are called the *Basic Belief Masses* (BBM) and the m function is the *Basic Belief Assignment* (BBA). As an example, a total lack of information concerning the potential membership of the agent X to a subset of $\mathcal{C}$ is represented by $m(A) = 0$ for any strict subset A of $\mathcal{C}$.

If we transfer this interpretation to the case of the nearest neighbors for the data assignment, we define $\Gamma_c(X(i))$ as a set of the c nearest neighbors

of $X(i)$. Each neighbor $X(j) \in \Gamma_c(X(i))$, $j = 1, \ldots, c$ can be considered as a piece of evidence that influences one's belief in the cluster $\mathcal{C}$ membership of the data $X(i)$, and a mass of belief m_j^i is associated to their relationship.

In other words, if $X(j)$ is a member of the class $\mathcal{C}_P$, $P \in \{1, \ldots, \bar{s}\}$ where $\bar{s}$ is the estimated number of submodels, then a part of the mass of belief is allocated to the action of assigning the data point $X(i)$ to the cluster $\mathcal{C}_P$ given by $m_j^i(\{\mathcal{C}_P\})$ and the remaining part, to the set $\mathcal{C}$ as described as follows

$$m_j^i(A) = \begin{cases} \alpha_0 \phi_j^i & \text{if} \quad A = \mathcal{C}_P \\ 1 - \alpha_0 \phi_j^i & \text{if} \quad A = \mathcal{C} \\ 0 & \text{if} \quad A \in 2^{\mathcal{C}} \setminus \{\{\mathcal{C}_P\}, \mathcal{C}\} \end{cases}, \tag{14}$$

where

$$\phi_j^i = \exp\left(-\gamma_P \|X(i) - X(j)\|^2 - \beta_P \left(y(i) - \theta_P^\top \bar{\varphi}_H(i)\right)^2\right). \tag{15}$$

What is notable in ϕ_j^i expression, is that the exponent is composed of two terms. The first one expresses the Euclidean distance between $X(i)$ and the neighbor $X(j)$ represented by $d_j^i = \sqrt{(X(i) - X(j))^\top (X(i) - X(j))}$, and the second term represents the prediction error related to the estimation of the parameter vector θ_P of the submodel P, associated to the cluster $\mathcal{C}_P$. This formulation aims at minimizing simultaneously the euclidean distance and the output error (both shown in Figure 1.

α_0 is a positive parameter such as $0 << \alpha_0 < 1$ that prevents from allocating the entire mass of belief to the class $\mathcal{C}_P$. Indeed, even if there is zero distance between $X(i)$ and $X(j)$ and the prediction error is also towards zero, it remains an uncertainty concerning the belonging to the same class.

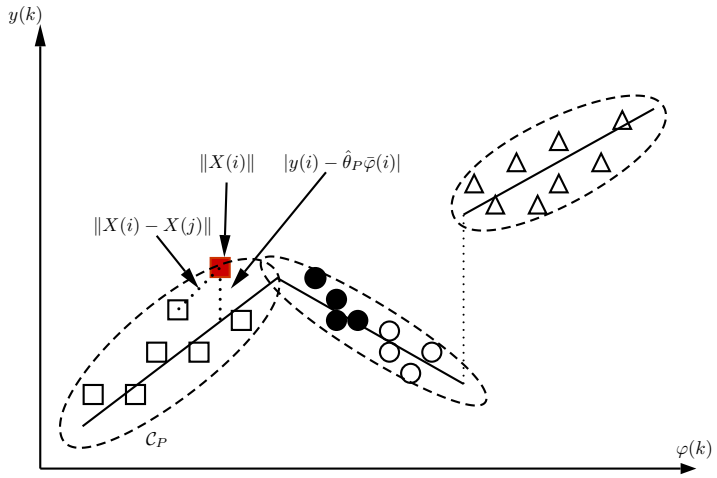

Figure 1: Example of three affine submodels in the augmented regression space

The mass of belief of $X(j)$ on cluster $\mathcal{C}_p$ *i.e.* $m_j^i(\{\mathcal{C}_P\})$ is therefore function of the mean distance d_P between two data belonging to the same class $\mathcal{C}_P$ and its proximity with the linear model expressed by the average error e_P between the measured output and the submodel output as described in the following

$$\beta_P = \frac{1}{d_P^2}, \tag{16a}$$

$$\gamma_P = \frac{1}{e_P^2}. \tag{16b}$$

Simple heuristics for the choice of α_0 and the initial value of γ_P is presented in Section 4.2. One can note that if $X(i)$ is far from $X(j)$ and in the same time the error $\left(y(i) - \theta_P^\top \bar{\varphi}(i)\right)^2$ is large, the class $\mathcal{C}_P$ of $X(j)$ will be considered as providing very little information regarding the class of $X(i)$.

For more detailed and comprehensive explanation on how the decision rules are built based on the combination between the belief functions and

Dempster's rule, the reader is invited to refer to [39, 31, 41].

Credal level:

After tuning the c BBAs $m_1^i, m_2^i, \ldots, m_c^i$, the combination of all the BBAs using Dempster's combination rule is performed using the orthogonal sum $\oplus$ and yields to a single mass

$$m^i = m_1^i \oplus \ldots \oplus m_c^i, \tag{17}$$

which can be calculated using a second rule called the conjunctive rule of combination. In case of $A = \mathcal{C}$, it is defined by

$$m^i(\mathcal{C}) = \sum_{A_1, A_2, \ldots, A_c | \bigcap_j A_j = \mathcal{C}} \prod_{j=1}^{c} m_i^j(A_j). \tag{18}$$

It is obvious that: $\left(\bigcap_j A_j = \mathcal{C}\right) \Rightarrow A_j = \mathcal{C}$. According to Eq. (14), $m^i(\mathcal{C})$ becomes

$$m^i(\mathcal{C}) = \prod_{j=1}^{c} (1 - \alpha_0 \phi_j^i). \tag{19}$$

The same reasoning applies to the belief function $m^i(\{\mathcal{C}_P\})$ which is given by

$$m^i(\{\mathcal{C}_P\}) = (1 - \prod_{j/X(j) \in \mathcal{C}_P} (1 - \alpha_0 \phi_i^j)) \prod_{\substack{q=1 \\ q \neq P}}^{\bar{s}} \prod_{j/X(j) \in \mathcal{C}_P} (1 - \alpha_0 \phi_i^j). \tag{20}$$

Pignistic level:

Let $\mathcal{P}$ a set of probability distributions derived from the BBAs. the unknown pignistic probability measure $\mathcal{P}$ satisfies

$$Bel(A) < \mathcal{P}(A) < Pl(A), \tag{21}$$

where $Bel(A)$ is the belief function also called the credibility of A and $Pl(A)$ is the plausibility of A defined respectively by

$$Bel^i\left(\{\mathcal{C}_P\}\right) = m^i\left(\{\mathcal{C}_P\}\right), \tag{22a}$$

$$Pl^i\left(\{\mathcal{C}_P\}\right) = m^i\left(\{\mathcal{C}_P\}\right) + m^i\left(\mathcal{C}\right). \tag{22b}$$

$\mathcal{P}^i$ is then expressed by

$$\mathcal{P}^i\left(\{\mathcal{C}_P\}\right) = m^i\left(\{\mathcal{C}_P\}\right) + \frac{m^i(\mathcal{C})}{\bar{s}}. \tag{23}$$

The decision is made by assigning the data $X(i)$ to the cluster $\mathcal{C}_q$ with maximum of pignistic probability. Then the decision rule is given by:

$$\mathcal{C}_q = \mathcal{C}_q \cup \{X(i)\} \quad \text{such that} \quad q = \max_{P,P=1,\ldots,\bar{s}} \left(\mathcal{P}^i\left(\{\mathcal{C}_P\}\right)\right). \tag{24}$$

An iteration procedure allows the convergence of the clusters and their parameters using a stop criterion based on a comparison between the old and the new parameter vectors. Indeed, an adaptation of the parameter vector $\hat{\theta}_P$ is performed at each new iteration it after a new data reassignment. The clustering is stabilized if

$$\left\|\Theta^{it+1} - \Theta^{it}\right\| \leq \epsilon, \tag{25}$$

where ϵ is fixed by the user.

Finally, all the previous steps are summarized in Algorithm 1. β_0 is initialized using Eq. (16a), $\gamma_0 = \nu\beta_0$ where ν is a positive coefficient arbitrarily chosen by the user. This parameter allows to adjust the weights in each of the two terms of the exponent in Eq. (15) corresponding to the euclidean distance and the submodel error to give each more or less importance depending on the case study. The parameters γ_P and β_P are adapted after each iteration in order to take into account the evolution of the clusters. The number of the nearest neighbors c is the main tuning parameter. After a finite number of iterations, the procedure converges relatively fast to a reasonable number of submodels. $\bar{s}$ is the number of non empty clusters but, in the case of noisy data, the cardinality of some clusters falling under a given number are removed. The problem of noisy data does not generally arise in case of measurements provided by hydrometric stations. However, some outliers could be encountered due to defective or blocked sensors.

3. Model validation for event forecasts

The computation of the river flow forecasts over a prediction horizon H involves necessarily the knowledge of the discrete state with a time-step ahead equal to this prediction horizon. In fact, the issue of the estimation of the discrete state on validation data was rarely addressed in literature. The literature survey is even rarer when it comes to real case studies, especially

Algorithm 1

1: **Initialization**

- Set c, α_0, β_0, γ_0, $\bar{s} = N$, $it = 0$ and ϵ (*e.g.* $\epsilon = 10^{-4}$).
- Set $\mathcal{C}_i = \{X(i)\}, i = 1, \ldots, \bar{s}$.
- Estimate $\Theta^{(0)} = \left[\hat{\theta}_1^{(0)}, \ldots, \hat{\theta}_{\bar{s}}^{(0)}\right]$.

2: **Data reassignment**

for $i = 1, .., N$

- For all $X(j) \in \Gamma_c(X(i))$, $j = 1, \ldots, c$, calculate ϕ_i^j, Eq. (15).
- Calculate $m_j^i(\{\mathcal{C}_P\})$ and $m_j^i(\{\mathcal{C}\})$, Eq. (14).
- Combine all the BBA functions using Dempster's rule, Eq. (19) and (20).
- Calculate all the $\mathcal{P}^i(\{\mathcal{C}_P\})$ using Eq. (23).
- Decide on the assignment of $X(i)$ based on Eq. (24).

end for.

- $\bar{s}$ = number of non empty clusters.
- Adaptation of the parameters:
  - Adapt $\Theta^{(it)}$ using LS algorithm on the data of each non empty cluster.
  - Adapt the parameters β_P and γ_P, Eq. (16a) and (16b).

3: **Convergence test**

If $\left\|\Theta^{(it+1)} - \Theta^{(it)}\right\| \leq \epsilon$

$s = \bar{s}$.

else

Set $it = it + 1$ and return to step 2.

end.

4: **Return** Θ^{it} and $\hat{\sigma}(k)$.

when the number of states is relatively important, thus making the task of state prediction even more delicate. Also, the case of discrete state prediction

over a prediction horizon was not investigated to the best of the author's knowledge.

The output of the data classification step provides the discrete state estimation $\hat{\sigma}(k)$ and the estimates of the submodel parameters Θ. It is now possible to estimate the shapes of the polyhedral region boundaries. These regions are in reality hyperplanes that form a complete polyhedral partition $\{\mathcal{R}_i\}_{i=1}^{s}$ of the regression space. The problem can then be equivalent to the separation to s sets using linear classifiers [24]. However, the same author highlights two problems encountered when a linear classification is used. First, it is not possible to exactly estimate the regions starting from a finite set of data which induces small errors in shaping, leading to the second problem consisting in a misclassification of the regression vector located near the discontinuities and larger prediction error can then be observed. A posterior re-attribution to the correct submodel could be envisaged during the validation. For a more complete discussion about class separation methods applied for SARX and PWARX systems, the reader is referred to Section 4.2 of [24].

The *Multicategory Support Vector Machine* (SVM) algorithm [42, 43] is used to predict the shapes of partitions. This technique is commonly used in literature related to PWARX identification approaches in order to estimate the polyhedral regions. In [44], a SVM classifier using a one-versus-the-rest approach was applied.

4. Application to flood forecasting

4.1. Case study

Our study focuses on the Liane coastal river of 37 km long. It is situated in the north of France (see Fig. 2) and flows in the *La Manche* sea (the English Channel as called by the British). The catchment drained by this river covers a total area of 244 km^2 and flows out a significant annual amount of rainfall higher than 480 mm, which is much greater then the national average. The Liane river has an average annual flow of 1.81 m^3/s [45] corresponding to a stream level of 47 mm but knows an increasingly number of severe flood events due to the drainage network practiced in agriculture and a limited exchange between the ground and surface water, exceeding frequently the 20 m^3/s. We can cite for example the flood of November 2012 where its level achieved 4,37 m.

A hydrometric station is situated in Wirwignes (red marker in Figure 2) and is equipped with a telemetered rain gauge since 2003 and a limnimeter since 1970. The rain gauge gives an information on the quantity of rain falling on the catchment and the limnimeter gives the measurement of the river upstream level. The data are collected with a basic time-step of 1 hour. It is important to note that most of the studies related to flood forecasting deal with the prediction of the streamflow instead of the water level. However, in practice, the streamflow is not directly measured but is inferred using a periodically revised relation between the measured stream level and the streamflow. This relation is determined by a simultaneous measurements of these two physical quantities over the natural range of flows from the lowest to the highest values corresponding to the floods.

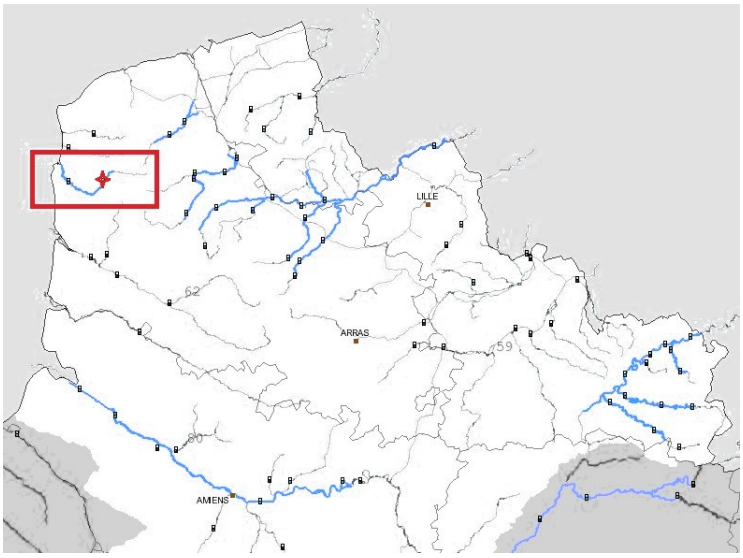

Figure 2: Localisation of Liane river and the hydrometric station (red marker) in the north of France.

Figures 1.a and 1.b shows respectively the hourly rainfall and the hourly water level measurements, from January 2010 to June 2018 *i.e.* over a period of eight years and a half. Two high water level periods are noticed: during winter and at beginning of spring. The low level period is during summer and autumn. The Liane is also characterized by a short hydrological response time and a flood period less than 24 hours. The objective of the study is to forecast the river level with the prediction horizon of at least 24 hours, using the rainfall forecast provided by weather services and the observed stream levels.

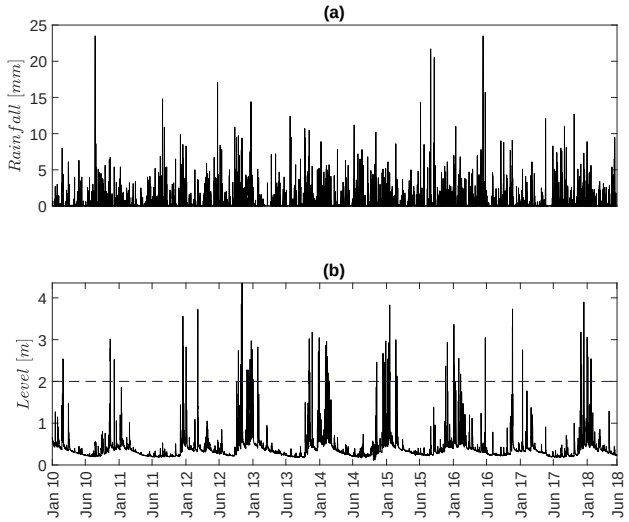

Figure 3: Model input and output from January 2010 to June 2018. (a) Rainfall in [mm] (b) Liane level [m]. The blue dashed line refers to the green threshold.

4.2. Prediction results

The Liane river underwent from January 1^{st}, 2010 to June 13^{th}, 2018, 55 significant floods where the river level exceeds the threshold of 2 m which corresponds to a return period of one year (the green threshold, see Fig. 3) which still relatively a low threshold considering that in practice, it is taken equivalent to a return period of two years (yellow threshold). Since we are interested in the flood prediction, only the flood periods corresponding to heavy rainfall seasons are manually selected for the study. Each period duration is between 3000 and 5000 hours. Hence, only 31072 samples on 74045 are selected. About 70% of the 55 most significant events are used for the training (which corresponds to 40 events and 20715 samples) and the remaining events are used for validation (15 events and 10357 samples). Figure 4 represents level samples used for estimation (black line) and for

validation (red line).

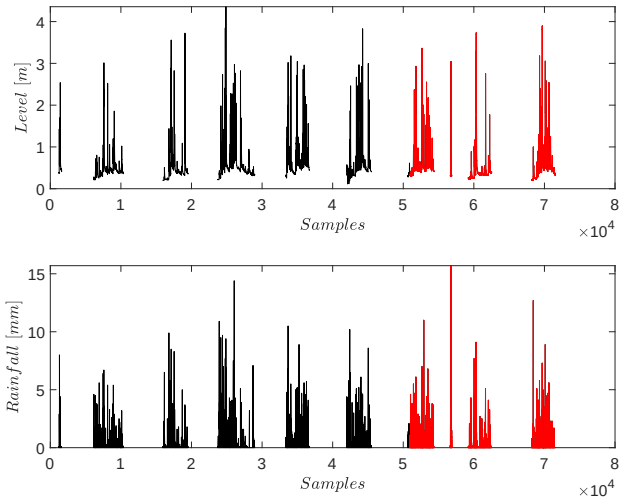

Figure 4: Training data (black line) together with checking data (red line)

Different criteria are used to globally appreciate the performances of each model. The FIT criterion introduced by Ljung [14] defined by (26), the Mean Square Error defined by (27) and the Nash-Sutcliff coefficient [15] defined by (28) give an information about the global fitting of the predicted output in the observed data, considering that a prediction is performed each f points (of the f next water level samples).

$$FIT = 100 \times (1 - \frac{\|y - \hat{y}\|_2}{\|y - \overline{y}\|_2}), \tag{26}$$

$$MSE = \frac{1}{N} \sum_{k=1}^{N} (y(k) - \hat{y}(k))^2, \tag{27}$$

$$NASH = 1 - \frac{\frac{1}{N}\sum_{k=1}^{N}\left[\hat{y}(k) - y(k)\right]^2}{\mathrm{var}(y(k))}, \tag{28}$$

where $\overline{y}$ is the mean of the observed output, N is the total number of samples and var(.) is the signal variance. Other indicators are used specifically to the flood events in hydrological modeling. The most important one is the *Critical Success Index* CSI [46] which represents the ratio between the number of "Correct alarms" CA which are the number of the well-estimated crests with respect to the total number of alarms *i.e.* the significant events. CSI is defined by (29).

$$CSI = \frac{CA}{CA + MA + FA} \times 100. \tag{29}$$

If the observed event is not simulated or underestimated with more than 10% of the crest maximum, a Missed Alarm occurs and MA is the number of Missed Alarms. Conversely, a False Alarm occurs if a non observed event is simulated or overestimated with more than 10% and FA is the number of False Alarms. The MA and FA indicators are then defined as follows for each significant event

If $|y(k) - \hat{y}(k|k-f)| > 0.1 \times |y(k)|$:

$$\begin{cases} MA: & \text{if } y(k) > \hat{y}(k|k-f) \\ FA: & \text{if } y(k) < \hat{y}(k|k-f) \end{cases}. \tag{30}$$

We thus simply define a CA for events fulfilling the following condition : $|y(k) - \hat{y}(k|k-f)| \leq 0.1 \times |y(k)|$.

4.2.1. Short-term forecasting

The selection of the short-term prediction horizon depends usually on the sample time and the response time of a unit hydrograph. A typical choice of the natural response time is one fifth to one third of a unit hydrograph peak time in order to consider the slightest dynamic changes [22]. However, since SVM technique is used to predict the discrete sequence mode, it is recommended to shorten the prediction horizon to ensure a better validation. Indeed, the dimension of the features used for training which consists in the regression vectors depends on the horizon value. Higher is H, higher is the dimension of features and more restrictive is the data recognition. A classical solution commonly used in machine learning algorithms consists in reducing the features dimension by replacing the regression vector by some statistical metrics such as the mean, the variance and the kurtosis values. The problem of this procedure is the risk of loss of information related to the exact location of the local data provided initially by the regression vectors leading to a confusion between submodels resulting in different data located in the same region with same parameter vectors.

The peak time of the Liane river is about twelve hours which implies a response time between two and four hours. Finally, the prediction horizon is reduced to one hour for the previously stated reasons.

Model orders for the tested methods were chosen according to the performances structures built by increasing n_a and n_b from 1 to 20. Due to a high signal to noise ratio, it is allowed to choose the tuning parameter c small relatively to the number of training data samples to avoid a mixed submodel stemming from an important number of mixed local data as mentioned in

the previous section.

The choice of the parameter c is difficult to make *a priori*, especially when the method is sensitive to this parameter. A systematic search is necessary to obtain optimal results with a minimum number of modes.

To see the influence of the parameter c on the estimation of the number of submodels and for a selection that will approach the optimum as near as possible, the PWARX identification algorithm is applied to the same training set for different c values. The FIT score is calculated after the evaluation of the PWARX model on the validation set and for a long-term forecasting chosen arbitrarily equal to $f = 24$.

Selection of the hyperparameters:

Fig. 5 shows the resulting number of modes and FIT scores on validation data for $f = 24$ with c ranging from 50 to 500. It can be noticed that, non surprisingly, the smaller the c value, the larger is the number of estimated modes and FIT score. However, FIT scores for $c = 150$ and $c = 200$ are substantially equal to each other whereas the resulting number of modes are respectively $\bar{s} = 85$ and $\bar{s} = 60$. The choice was naturally directed towards $c = 200$.

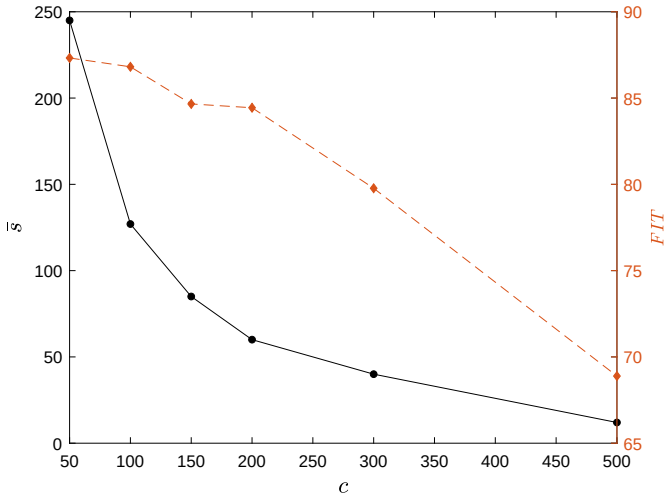

Figure 5: Number of submodels $\bar{s}$ estimated by the Evidential algorithm and the FIT resulting values on validation data set for $f = 24$ as a function of c nearest neighbors.

γ is usually chosen between 0.1 and 5 and β is taken 10 to 40 times greater then γ. The ratio between these two weighting parameters is in fact rather more important then the values themselves given that the goal is to favour the prediction error in this case study rather then the euclidean distance in Eq. (15). These hyperparameter c gives the user the possibility to achieve a high accuracy depending on the application but at the expense of an important number of resulting modes. It is although possible to achieve equivalent accuracy values with a smaller number of modes by testing one of the parameters β or γ.

The best PWA structure in terms of performances criteria using estimation data is for $n_a = 1$, $n_b = 1$, $c = 200$, $\gamma = 0.5$, $\beta = 20$, leading to the training scores $FIT = 93.7\%$, $Nash = 0.996$ and 60 classes. Fig. 6 shows the simulated output, the prediction error and the estimated discrete state

493 resulting from the training step.

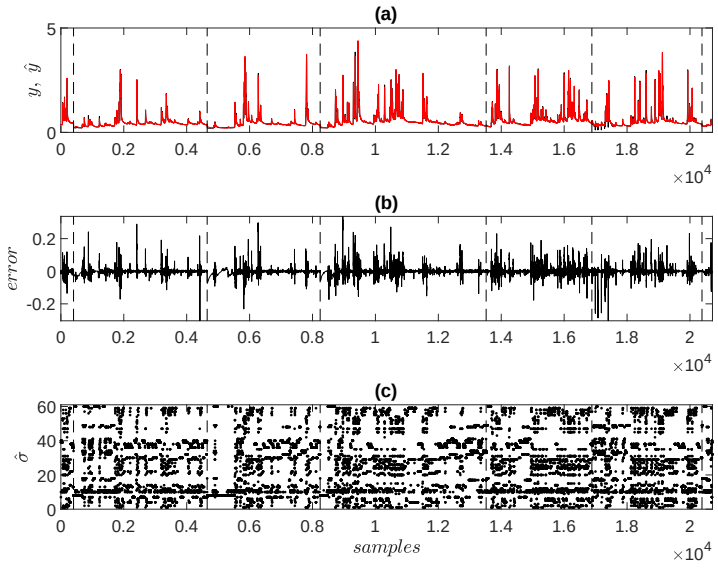

Figure 6: (a) Observed level used to train the PWARX model (black solid line) together with PWA model output (red solid line) (b) Prediction error (c) Predicted discrete state, for $H = 1$ and $f = 1$

494 The Evidential algorithm gives the number of modes and the discrete
495 state sequence. The latter information is considered as a labelling of data
496 and is used to train a Support Vector Machine (SVM) classifier [42] in order
497 to predict the discrete state sequence corresponding to the validation data.
498 Although the fact that the model is piecewise linear, the kernel function
499 used for training is the RBF (Radial Basis Function) instead of a linear
500 function to minimize a wrong classification of the regression vectors close
501 to linear boundaries (see Section 3). The normalized regression vectors are
502 used as inputs of the SVM classifier. The feature vectors are normalized
503 by subtracting the mean in the numerator and dividing by the standard

deviation in order to make the values of each feature have zero-mean and unit-variance [47]. Feature normalization is a classical preprocessing step in machine learning that is required when features have different ranges and it has shown in our case study better performance results.

In order to assess the effectiveness of the classification, some measures based on the Confusion Matrix are calculated: the Accuracy rate, the Recall rate and the Precision rate. We briefly recall that a Confusion Matrix [48, 49] is a concept coming from machine learning. It contains information about actual and predicted classifications and is then used to estimate the overall classification accuracy. A confusion matrix is a two-dimensional matrix, one is indexed by the actual class of an object, the other is indexed by the class that the classifier predicts.

Table 1 presents the basic form of confusion matrix for a multi-class classification task, with the classes $\mathcal{C}_1, \ldots, \mathcal{C}_{\bar{s}}$. N_{ij} represents the number of misclassed data *i.e.* belonging to class $\mathcal{C}_i$ but classified as class $\mathcal{C}_j$. The confusion matrix is computed using the matlab routine *"confusionmat"*.

		Predicted				
		$\mathcal{C}_1$	$\cdots$	$\mathcal{C}_j$	$\cdots$	$\mathcal{C}_{\bar{s}}$
Actual	$\mathcal{C}_1$	N_{11}	$\cdots$	N_{1j}	$\cdots$	$N_{1\bar{s}}$
	$\vdots$	$\vdots$		$\vdots$		$\vdots$
	$\mathcal{C}_i$	N_{11}	$\cdots$	N_{ij}	$\cdots$	$N_{i\bar{s}}$
	$\vdots$	$\vdots$		$\vdots$		$\vdots$
	$\mathcal{C}_{\bar{s}}$	$N_{\bar{s}1}$	$\cdots$	$N_{\bar{s}j}$	$\cdots$	$N_{\bar{s}\bar{s}}$

Table 1: Confusion matrix

Accuracy is the proportion of the total number of the correct predictions and is then defined as the ratio between the trace and the sum of the confusion

matrix as follows

$$Accuracy = \frac{\sum_{i=1}^{\bar{s}} N_{ii}}{\sum_{i=1}^{\bar{s}} \sum_{j=1}^{\bar{s}} N_{ij}}. \tag{31}$$

Recall is a measure of the ability of a prediction model to select instances of a certain class from a data set. The overall Recall rate is the average value of each individual recall rate and is provided by the following formula

$$Recall = \frac{1}{\bar{s}} \sum_{i=1}^{\bar{s}} \frac{N_{ii}}{\sum_{k=1}^{\bar{s}} N_{ik}}. \tag{32}$$

Precision is a measure of the accuracy provided that a specific class has been predicted. The overall Precision rate is the average value of each individual class precision rate and is provided by

$$Precision = \frac{1}{\bar{s}} \sum_{i=1}^{\bar{s}} \frac{N_{ii}}{\sum_{k=1}^{\bar{s}} N_{ki}}. \tag{33}$$

The resulting SVM Accuracy indicator is 91.78%, the Recall indicator is 88.48% and the Precision indicator is 88.81%, showing a good recovery of the estimated state sequence.

Table 2 compares the PWA performance indicators with one-step ahead prediction horizon ($H = 1$,$f = 1$) with the LPV approach on the checking data *i.e.* the 15 remaining flood events. The LPV approach is chosen for comparison to PWARX modeling instead of the Nonlinear ARX (NARX) which is the other data-driven nonlinear approach because of the difficulty

encountered to opt for the most effective static nonlinearities of both inputs and outputs. Several static functions were tested (linear, Radial Basis Function with two tuning parameters, logarithmic, sigmoid, etc) and none of them was able to give fair results. Bad performances would necessarily be attributed to a bad selection of the static functions.

The index in LPV_1 and PWA_1 is relative to $f = 1$. LPV model structure is chosen according to the performance indicators for $n_a = 1, \ldots, 10$ and $n_b = 1, \ldots, 10$. Concerning the LPV model, four different possibilities for the scheduling variable are tested: past rainfall samples $(u(k-1), u(k-2), \ldots, u(k-24))$, past water level samples $(y(k-1), \ldots, y(k-5))$, linear model output [50], identified using for example Least Squares algorithm or an Output Error (OE) algorithm based on a nonlinear minimization of the prediction error based on Levenberg-Marquardt algorithm [51, 52], or the Refined Instrumental Variable identification of a Box-Jenkins Transfer Function model [53]. The last possibility is a combination between the average of past rainfall inputs and past water level $([mean(u(k-1), \ldots, u(k-i))\ y(k-1)]$ where $i = 1, \ldots, 24)$. A polynomial dependency function to the scheduling variable with a polynomial degree $r = 1, \ldots, 5$ is tested [10]. The OE approach based on Levenberg-Marquardt algorithm is extended to the discrete-time LPV case and used for LPV model identification. PWA model outputs are presented in Fig. 7.

	$\%\ FIT$	$Nash$	MSE	Max	CA	MA	FA	$\%\ CSI$
PWA_1	90.9	0.9917	9.6×10^{-4}	0.99	15	0	0	100
LPV_1	94.7	0.997	3.3×10^{-4}	0.47	15	0	1	93.75

Table 2: Performance results of the tested models on validation set, for $H = 1$, $f = 1$

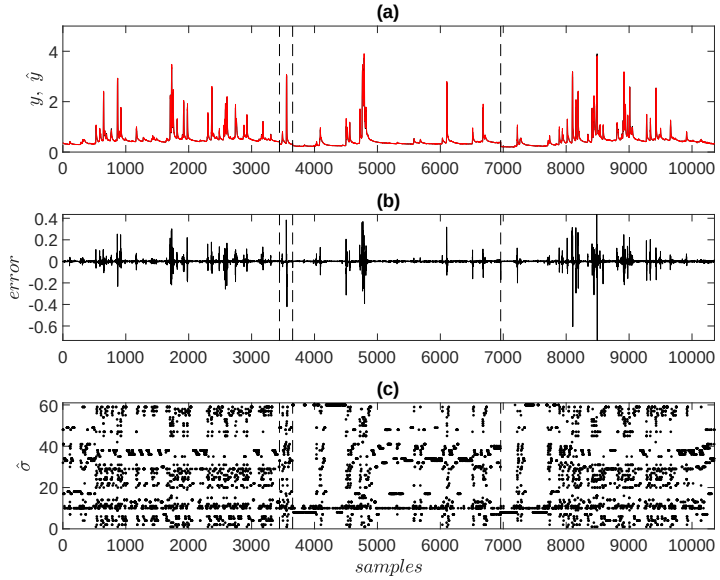

Figure 7: (a) Actual level (black solid line) together with PWA model output (red solid line) (b) Prediction error (c) Predicted discrete state, with $H = 1$ and $f = 1$

According to Table 2, the performance indicators are in favor of the LPV model for a very short-term forecast $f = 1$, albeit with one FA. The good performances are most observed for lower water levels. One can note that the PWA performance indicators did not significantly decrease in validation. This is primarily due to a good selection of training data and tuning parameters, and obviously, to the small prediction horizon value, thus limiting the loss of accuracy in the SVM prediction of the discrete state and the error accumulation due to the multiple iterations.

4.2.2. Long-term forecasting :

In case of prediction horizons larger than the natural response time, a good forecast of the precipitation is required and must be provided by the

user using weather forecasts services. In the following, an assumption on the exact knowledge of the future rainfall is made. Naturally, a discussion on a pessimistic and optimistic forecast must be conducted to quantify to what extent the model is sensitive to its inputs. A comparison between the LPV model and the PWARX model for 6h, 12h, 24h and 48h prediction horizons is driven and the performance indicators are drawn in Table 3 and it shows the clear superiority of the PWA performance indicators regardless to the lead time.

In addition, the percentage of forecasted warnings regarding the green threshold overrun with a temporal prediction accuracy smaller than 1 hour is 88% for a prediction horizon $H = 24$, making this approach one of the best ARX approaches in terms of prediction lag performances. The peaks are also globally forecasted in time where the peak prediction lag percentage smaller then 1 hour is 85%. Note that the CSI is also a good indicator of the peak prediction temporal accuracy since it is impacted in case of an important lag between real and forecasted peak. Fig. 8 shows the water level and the prediction error when considering 24h ahead PWA forecasts.

		$\%\ FIT$	$Nash$	MSE	Max	CA	MA	FA	$\%\ CSI$
LPV	$H = 6$	67.5	0.894	0.0123	1.74	10	5	1	62.5
	$H = 12$	57.03	0.815	0.021	1.96	3	11	1	20
	$H = 24$	50	0.75	0.029	1.88	2	12	1	13.3
	$H = 48$	46.4	0.713	0.033	1.64	1	14	0	6.7
PWA	$H = 6$	85.5	0.9790	0.0025	1.42	15	0	1	93.75
	$H = 12$	84.7	0.9767	0.0027	1.43	13	1	1	92.86
	$H = 24$	84.4	0.9758	0.0028	1.43	13	2	1	86.67
	$H = 48$	84.25	0.9752	0.0029	1.43	12	2	2	85

Table 3: Performance results of the LPV model and the PWARX model for river level forecasts and for lead times 6h, 12h, 24h and 48h.

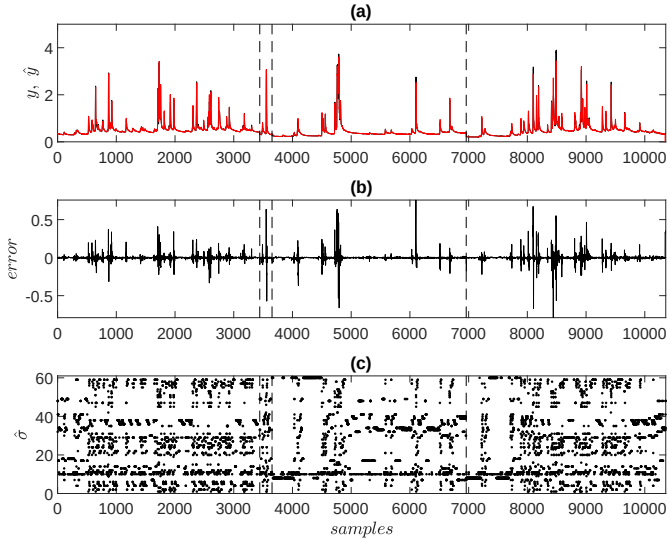

Figure 8: (a) Observed level (solid black line) together with PWA model output (solid red line) (b) Prediction error, with $H = 1$ and $f = 24$

Figures 9, 10 and 11 show long-term forecasts of the three most important flood events during the validation period, respectively during December, 13^{th}, 2017, November 19^{th}, 2016 and January 3^{rd}, 2016, with the smaller lead time $H = 6$ and the larger lead time $H = 48$ in order to assess the error propagation due to the multiple iterations. One can note that the PWARX output deviation does not significantly increase with the prediction horizon. This is due to the high accuracy of the estimated model during the training thus becoming the main condition for PWARX precise forecasts in the case of an exact knowledge of the rainfall forecasts. Errors in water level forecasting in long term will hence be attributed to the errors in the rainfall forecasts which not surprisingly are higher in longer term.

We also note a worse deviation for multiple hydrographs *i.e.* several

successive flood peaks like the phenomenon observed during November 19^{th}, 2016. It is most likely due to the fact that the data set used for training the model does not probably include the exact similar event, which is the common shortcoming of the black box approaches.

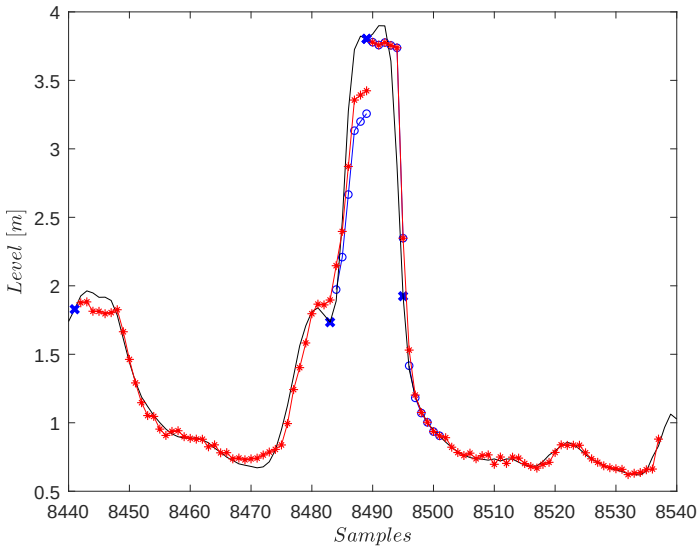

Figure 9: Observed river level (black line) together with PWA short-term forecasts $f = 6\ h$ (blue line) and long-term forecasts $f = 48\ h$ (red line) of the flood of the December, 13^{th}, 2017. The blue crosses indicate the time when forecasting is done.

5. Conclusion

This paper deals with the stream level prediction of a river using a hybrid model over a short-term and long-term prediction horizon in order to prevent damages by a short-term forecast of a flood. The choice of a nonlinear model is justified by numerous factors among them the soil saturation caused notably by intense rainy events during the previous days or the evapotranspiration of the soil and the canopy water interception. The choice of

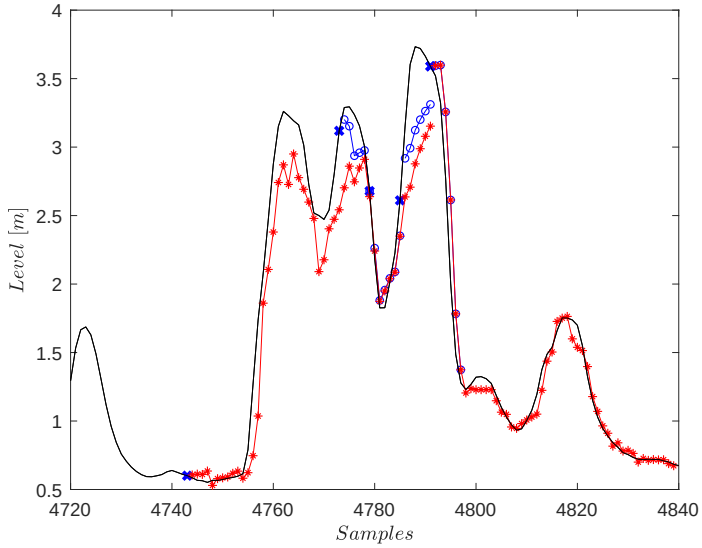

Figure 10: Observed river level (black line) together with PWA short-term forecasts $f = 6\ h$ (blue line) and long-term forecasts $f = 48\ h$ (red line) of the flood of November 19^{th}, 2016. The blue crosses indicate the time when forecasting is done.

a piecewise linear structure as a hybrid system is justified by the fact that a river acts differently for the same amount of precipitation and this can be assimilated to a switch between multiple models that might be linear according to the presented results, allowing the usage of an easy implementable algorithm with a low processing time according to the number of treated data samples. In addition, the utilization of an unsupervised clustering technique for the PWARX model identification facilitates the task of achieving a number of modes with no *a priori* knowledge and non systematic search. This is possible thanks to the algorithm based on Dempster-Shafer adapted for PWARX modeling. The main difficulty is to fix optimally the value of the main tuning parameter consisting in the number of the nearest neighbors. If this is properly done, one can achieve high performances both in a short-

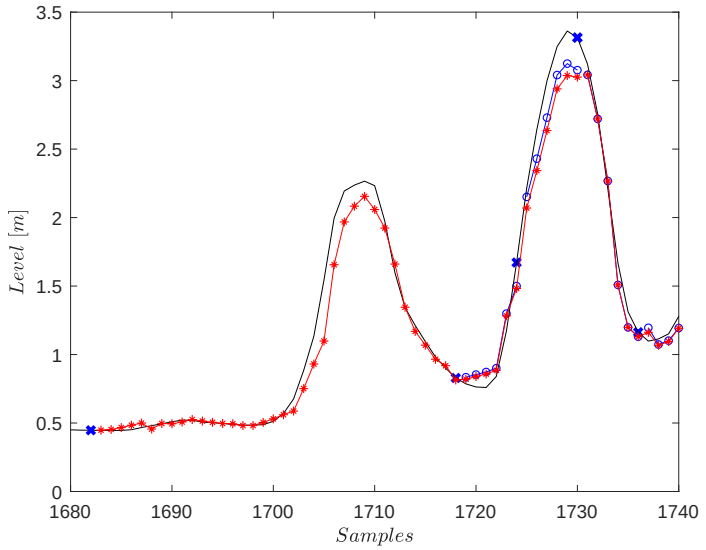

Figure 11: Observed river level (black line) together with PWA short-term forecasts $f = 6\ h$ (blue line) and long-term forecasts $f = 48\ h$ (red line) of the flood of January 3^{rd}, 2016. The blue crosses indicate the time when forecasting is done.

term and a long-term flood forecasting in terms of peak value, the instant when the threshold is crossed and the instant of the peak occurrence. It is however important to mention that the rainfall measurement issued by the rain gauges does not necessarily translate the real precipitation spilled on the catchment. This information has to be cross-checked with the weather radar information in order to reduce the uncertainty on any rainfall-runoff model input. This paper presented the first step of an exploratory work on the usage of PWARX systems in the flood forecast field. Future works will concern the robustness of these models to an uncertainty on the rainfall which has to be replaced by its forecast, a sensitivity analysis on the parameters, the model calibration and validation on a larger data base and finally, the evaluation of the model use in real-life settings.

References

[1] C. Perrin, V. Andréassian, Does a large number of parameters enhance model performance ? comparative assessment of common catchment model structures on 429 catchments, Journal of Hydrology 242 (2001) 275–301.

[2] A. Elshorbagy, G. Corzo, S. Srinivasulu, D. P. Solomatine, Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - part 2: Application, Hydrology and Earth System Sciences 14 (2010) 1943 – 1961.

[3] T. Asefa, M. Kemblowski, M. McKee, A. Khalil, Multi-time scale stream flow predictions: The support vector machines approach, Journal of Hydrology 318 (2006) 7 – 16.

[4] L. K. A. Siou, A. Johannet, S. Pistre, V. Borrell, Flash floods forecasting in a karstic basin using neural networks: the case of the lez basin (south of france), International Symposium on Karst – ISKA, Malaga. In Advances in research in karst media. Andreo et al eds Springer (2010).

[5] P. Nayak, K. Sudheer, D. Rangan, K. Ramasastri, A neuro-fuzzy computing technique for modeling hydrological time series, Journal of Hydrology 291 (2004) 52 – 66. doi:https://doi.org/10.1016/j.jhydrol.2003.12.010.

[6] H. Badrzadeh, R. Sarukkalige, A. W. Jayawardena, Intermittent stream flow forecasting and modelling with hybrid wavelet neuro-fuzzy model, Hydrology Research 49 (2017) 27–40. doi:10.2166/nh.2017.163.

[7] A. B. Dariane, S. Azimi, Streamflow forecasting by combining neural networks and fuzzy models using advanced methods of input variable selection, Journal of Hydroinformatics 20 (2017) 520–532. doi:10.2166/hydro.2017.076.

[8] T. K. Chang, A. Talei, C. Quek, V. R. Pauwels, Rainfall-runoff modelling using a self-reliant fuzzy inference network with flexible structure, Journal of Hydrology 564 (2018) 1179 – 1193. doi:https://doi.org/10.1016/j.jhydrol.2018.07.074.

[9] P. C. Young, Time series methods and recursive estimation in hydrological systems analysis, In River flow modelling and forecasting, D. A. Kraijenhoff, J. R. Moll (eds). D. Reidel : Dordrecht (1986) 129 – 180.

[10] R. Tóth, F. Felici, P. S. C. Heuberger, P. M. J. V. den Hof, Discrete time lpv i/o and state space representations, differences of behavior and pitfalls of interpolation, Proc. of the European Control Conf., Kos, Greece (2007) 5418 – 5425.

[11] F. Previdi, M. Lovera, Identification of parametrically-varying models for the rainfall-runoff relationship in urban drainage networks, IFAC Proceedings Volumes 42 (2009) 1768 – 1773. 15th IFAC Symposium on System Identification.

[12] V. Laurain, M. Gilson, S. Payraudeau, C. Grégoire, H. Garnier, Identification de modèles LPV : Application à la modélisation pluie/débit d'un bassin versant viticole, Sixième Conférence Internationale Francophone d'Automatique, CIFA (2010).

[13] E. Duviella, L. Bako, Predictive black-box modeling approaches for flow forecasting of the liane river., SYSID'12, Bruxelles, Belgium, (2012).

[14] L. Ljung, System identification : theory for the user (2nd Edition), Prentice Hall PTR, NJ, USA, Upper Saddle River, 1999.

[15] J. E. Nash, J. V. Sutcliffe, River flow forecasting through conceptual models part I: a discussion of principles, Journal of Hydrology 10 (1970) 282–290.

[16] C. Perrin, C. Michel, V. Andréassian, Improvement of a parsimonious model for streamflow simulation, Journal of Hydrology 279 (2003) 275 – 289.

[17] Edijatno, C. Michel, Un modèle pluie-débit journalier à trois paramètres, La Houille Blanche 2 (1989) 113 – 121.

[18] Edijatno, N. D. O. Nascimento, X. Yang, Z. Makhlouf, C. Michel, GR3J: a daily watershed model with three free parameters, Hydrological Sciences Journal 44 (1999) 263–277.

[19] F. Bourgin, M. Ramos, G. Thirel, V. Andréassian, Investigating the interactions between data assimilation and post-processing in hydrological ensemble forecasting, Journal of Hydrology 519 (2014) 2775 – 2784. doi:https://doi.org/10.1016/j.jhydrol.2014.07.054.

[20] A. Ficchi, An adaptive hydrological model for multiple time-steps : diagnostics and improvements based on fluxes consistency, Ph.D. thesis, Université Pierre et Marie Curie - Paris VI, 2017.

[21] H. Dakhlaoui, D. Ruelland, Y. Tramblay, Z. Bargaoui, Evaluating the robustness of conceptual rainfall-runoff models under climate variability in northern tunisia, Journal of Hydrology 550 (2017) 201 – 217.

[22] G. Bastin, L. Moens, P. Dierick, Online river flow forecasting with hydromax : successes and challenges after twelve years of experience, In proceedings of the 15th IFAC Symposium on System Identification, Saint-Malo, France, July 6-8 (2009).

[23] J. Sjoberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P.-Y. Glorennec, H. Hjalmarsson, A. Juditsky, Nonlinear black-box modeling in system identification: a unified overview, Automatica 31 (1995) 1691 – 1724. Trends in System Identification.

[24] S. Paoletti, A. L. Juloski, G. Ferrari-Trecate, R. Vidal, Identification of hybrid systems a tutorial, European Journal of Control 13 (2007) 242 – 260.

[25] R. Vidal, S. Soatto, Yi Ma, S. Sastry, An algebraic geometric approach to the identification of a class of linear hybrid systems, in: 42nd IEEE International Conference on Decision and Control (IEEE Cat. No.03CH37475), volume 1, 2003, pp. 167–172 Vol.1.

[26] R. Vidal, Identification of PWARX hybrid models with unknown and possibly different orders, in: American Control Conference, volume 1, 2004, pp. 547–552 vol.1.

[27] A. L. Juloski, S. Weiland, W. P. M. H. Heemels, A bayesian approach

to identification of hybrid systems, IEEE Transactions on Automatic Control 50 (2005) 1520–1533.

[28] A. Bemporad, A. Garulli, S. Paoletti, A. Vicino, A bounded-error approach to piecewise affine system identification, IEEE Transactions on Automatic Control 50 (2005) 1567–1580. doi:10.1109/TAC.2005.856667.

[29] G. Ferrari-Trecate, M. Muselli, D. Liberati, M. Morari, A clustering technique for the identification of piecewise affine systems, Automatica 39(2) (2003) 205–2017.

[30] F. Lauer, G. Bloch, Piecewise smooth system identification in reproducing kernel hilbert space, in: 53rd IEEE Annual Conference on Decision and Control, 2014, pp. 6498–6503.

[31] K. Boukharouba, L. Bako, S. Lecoeuche, Identification of piecewise affine systems based on Dempster-Shafer theory, in: IFAC Symposium on System Identification, Saint-Malo, France, 2009, pp. 1662–1667.

[32] F. Lauer, G. Bloch, R. Vidal, A continuous optimization framework for hybrid system identification, Automatica 47 (2011) 608 – 613. doi:https://doi.org/10.1016/j.automatica.2011.01.020.

[33] L. Bako, Subspace clustering through parametric representation and sparse optimization, IEEE Signal Processing Letters 21 (2014) 356–360.

[34] S. Kersting, M. Buss, Recursive estimation in piecewise affine systems using parameter identifiers and concurrent learning, International Journal of Control 0 (2017) 1–18.

[35] L. Bako, K. Boukharouba, E. Duviella, S. Lecoeuche, A recursive identification algorithm for switched linear/affine models, Nonlinear Analysis: Hybrid Systems 5 (2011) 242 – 253.

[36] V. Breschi, A. Bemporad, D. Piga, Identification of hybrid and linear parameter varying models via recursive piecewise affine regression and discrimination, in: 2016 European Control Conference (ECC), Aalborg, Denmark, 2016, pp. 2632 – 2637.

[37] V. Breschi, D. Piga, A. Bemporad, Piecewise affine regression via recursive multiple least squares and multicategory discrimination, Automatica 73 (2016) 155 – 162.

[38] G. Shafer, A mathematical theory of evidence, Princeton University Press, Princeton, N.J., 1976.

[39] T. Denoeux, A k-nearest neighbor classification rule based on dempster-shafer theory, IEEE Transactions on Systems, Man, and Cybernetics 25 (1995) 804–813. doi:10.1109/21.376493.

[40] P. Smets, R. Kennes, The transferable belief model, Artificial Intelligence 66 (1994) 191–234.

[41] B. Hadid, S. Lecoeuche, Data assignment and parameter adaptation for switched LPV system estimation, in: 20th IFAC World Congress, Toulouse, France, 2017, pp. 4564 – 4569.

[42] V. Vapnik, The nature of statistical learning theory, Springer-Verlag New York Inc, 1995.

[43] E. J. Bredensteiner, K. P. Bennett, Multicategory classification by support vector machines, Computational Optimization and Applications 12 (1999) 53–79.

[44] H. Ohlsson, L. Ljung, Identification of switched linear regression models using sum-of-norms regularization, Automatica 49 (2013) 1045 – 1050.

[45] http://www.hydro.eaufrance.fr (2019).

[46] D. Norbiato, M. Borga, S. D. Esposti, E. Gaume, S. Anquetin, Flash flood warning based on rainfall thresholds and soil moisture conditions: An assessment for gauged and ungauged basins, Journal of Hydrology 362 (2008) 274 – 290.

[47] X. Kong, X. Liu, R. Shi, K. Y. Lee, Wind speed prediction using reduced support vector machines with feature selection, Neurocomputing 169 (2015) 449 – 456. doi:https://doi.org/10.1016/j.neucom.2014.09.090, learning for Visual Semantic Understanding in Big Data ESANN 2014 Industrial Data Processing and Analysis.

[48] A. Hay, The derivation of global estimates from a confusion matrix, International Journal of Remote Sensing 9 (1988) 1395–1398. doi:10.1080/01431168808954945.

[49] C. Sammut, G. I. Webb, Encyclopedia of Machine Learning, Springer New York, 2011.

[50] P. Young, Top-down and data-based mechanistic modelling of rainfall-flow dynamics at the catchment scale, Hydrological Processes 17 (2003) 2195 – 2217.

[51] K. Levenberg, A method for the solution of certain non-linear problems in least squares, Quarterly of Applied Mathematics 2 (1944) 164 – 168. doi:https://doi.org/10.1090/qam/10666.

[52] D. W. Marquardt, An algorithm for least-squares estimation of non-linear parameters, Journal of the Society for Industrial and Applied Mathematics 11 (1963) 431–441. URL: http://www.jstor.org/stable/2098941.

[53] P. Young, A. Jakeman, Refined instrumental variable methods of recursive time-series analysis part i. single input, single output systems, International Journal of Control 29 (1979) 1–30. doi:https://doi.org/10.1080/00207177908922676.