

Knowledge Extraction (KnoX) in Deep Learning: Application to the Gardon de Mialet Flash Floods Modelling

Bob E. Saint Fleur, Guillaume Artigue, Anne Johannet, Séverin Pistre

▶ **To cite this version:**

Bob E. Saint Fleur, Guillaume Artigue, Anne Johannet, Séverin Pistre. Knowledge Extraction (KnoX) in Deep Learning: Application to the Gardon de Mialet Flash Floods Modelling. ITISE 2019 - International Conference on Time Series and Forecasting, Sep 2019, Granada, Spain. hal-02462501

HAL Id: hal-02462501

https://imt-mines-ales.hal.science/hal-02462501

Submitted on 31 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Knowledge Extraction (KnoX) in Deep Learning: Application to the *Gardon de Mialet* Flash Floods Modelling

Bob E. Saint Fleur [1,2], Guillaume Artigue [1], Anne Johannet [1], Severin Pistre [2]

[1] IMT Mines Alès, Laboratoire de Génie et de l'Environnement Industriel (LGEI), Alès, France
[2] Hydrosciences, Univ Montpellier, CNRS, IRD, 34090 Montpellier, France

Corresponding author Guillaume ARTIGUE (guillaume.artigue@mines-ales.fr)

Abstract. Flash floods frequently hit Southern France and cause heavy damages and fatalities. To better protect persons and goods, official flood forecasting services in France need accurate information and efficient models to optimize their decision and policy. Since heavy rainfalls that cause such floods are very heterogeneous, it becomes a serious challenge for forecasters. Such phenomena are typically nonlinear and more complex than classical floods events. That problem leads to consider complementary alternatives to enhance the management of such situations. For decades, artificial neural networks have been very efficient to model nonlinear phenomena, particularly rainfall-discharge relations in various types of basins. They are applied in this study with two main goals: first modelling flash floods on the *Gardon de Mialet* basin; second, extract internal information from the model by using the Knowledge eXtraction method to provide new ways to improve models. The first analysis shows that the kind of nonlinear predictor influences strongly the representation of information: e.g. the main influent variable (rainfall) is more important in the recurrent and static models than in the feed-forward one. For understanding flash floods genesis, recurrent and static models appear thus as better candidates, even if their results are better.

1 Introduction

In the Mediterranean regions, flash floods due to heavy rainfalls frequently occur and cause numerous fatalities and costly damages. During the last few years, the south of France has been particularly exposed to these catastrophic situations. In such cases, damages can reach more than one billion euros, and, in only one event, there can be more than 20 fatalities [1]. Facing these issues, authorities need reliable forecasts for early warning purposes. Unfortunately, both the short-term rainfall forecasts and the processes leading to the discharge response remain poorly known at the space and time scales required. It is thus difficult to provide forecasts using the traditional coupling between a meteorological model and a physically based hydrological model.

Artificial Neural Networks therefore appear as an alternative paradigm as they are able to provide forecasts of an output (discharge) without making any other hypothesis

on the system than the causality between rainfall and discharge. ANN have been applied in a wide variety of domains as they are essentially based on data and training [2]. They appear as particularly suitable for identifying the generating processes in hydrological time series because of their ability to model nonlinear dynamic systems [3,4]. However, due to their statistical origin, it is difficult to associate meaning to their internal parameters and they are rightly considered as black-box models. For this reason and to enhance the understanding of the behavior of the model, several works have been done to bring more transparency in the operating mode and introduced concepts of gray-box and transparent-box models [5,6]. In hydrology, several works have been conducted to make neural networks models more physically meaningful [6, 7, 8].

To be considered as gray-box (or transparent-box), ANN internal information or data must be accessible. In this paper, it will not be discussed *deep learning* itself, but an intermediate method to analyze the meaningful of internal information about neuronal models in hydrology operating on deep models. That method is termed ***Knowledge eXtraction*** (KnoX), it has been proposed by [7]. It was proved efficient on a fictitious basin, before being applied, by simulation, to estimate contributions and response times of various parts of a karst aquifer: the *Lez* aquifer (Southern France). It was later used by [8] for better apprehend the contributions of surface or underground processes in generation of floods on the Lavallette basin (Southern France).

Several studies were performed on the *Mialet* basin: first [4] showed that flash flood discharge can be forecasted by a multilayer perceptron with reasonable quality up to two-hours lead time; second, [9] showed that the initial value of the neural network parameters in flash floods forecasting has a major impact on the result. The purpose of this work is thus to better understand how the main variables influence the basin's outflow, regarding the model scheme used, in order to diminish the sensitivity of the model to the initialization of its parameters.

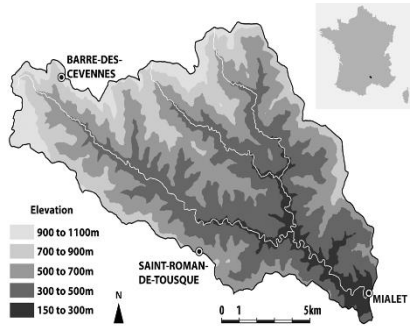
In the next sections, we will briefly present neural networks, their operating principles in hydrology, the deep multilayer perceptron used, as well as a reminder about the KnoX method and the models designed. The focus is set on a discussion about the behavior of the variable's weights according to the model type used, by applying the KnoX method to extract that information.

2 Materials and methods

2.1 Study area: location and general description

The *Gardon de Mialet* basin covers 220 sq.km in southern France. It is part of the *Cévennes* range which is known as a preferential location for the well-known meteorological phenomenon named *cevenols episodes* (Fig. 1). These episodes consist in short duration (less than 2 days) very heavy rainfall events. The elevation of Mialet basin ranges from 150 *m.a.s.l.* to 1170 *m.a.s.l.* and its mean slope is about 33 %. As for the most of basins of the *Cévennes*, these characteristics lead to limited infiltration or underground flow and thus to a high drainage density. Its response time is relatively short: between 2-4 hours [4]. The area is dominated by a metamorphic formation essentially with 95 % of mica-schist and gneiss, which lead to a poorly porous and impermeable

rocky sub-soil. The land use is almost homogeneous while covered by natural vegetation (chestnut trees, conifers, mixed forest and bush) for 92 %. The rest is shared between rocks and urban areas.



Typically, in Mediterranean regions, heavy rainfalls sometimes exceed 500 mm in only 24 h, to be compared to the 600 mm that fall on Paris annually. They are mainly produced by convective events, triggered either by relief, by a wind convergence, or by both. For example, in September 2002, the Gard (France) department has registered 687 mm of rainfall in 24h with 137 mm in only one hour at *Anduze* (a few km distant from *Mialet*).

Fig. 1. The study area (by Artigue, 2012)

2.2 Database

2.2.1. Presentation.

The database used in this study is essentially compounded with hourly observations from 1992 to 2002 and 5 minutes time step observations from 2002 to 2008 on three rain gauges and one hydrometric station at the outlet at *Mialet* (Fig. 1). From upstream to downstream, these stations are: BDC (*Barre des Cévennes*), SRDT (*Saint-Roman de Tousque*) and *Mialet* which coincide with the discharge station. They are all managed by the local Flood Forecasting Service (*SPC Grand Delta*). 58 events were extracted at 30 min time-step (based on rainfall events having at least 100 mm accumulation in 48 h on any of the rain gauges). Data description is synthetized in Tables 1 & 2.

Table 1. Data description

	Rainfall (mm)			Discharge	
	BDC	***SRDT***	***Mialet***	***(m^3s^{-1})***	***($m^3s^{-1}km^{-2}$)***
Maximum (30 min)	33.3	41.8	62.0	819.3	3.72
Median (30 min)	0.3	0.3	0.2	29.3	0.13
Moy	1.0	1.3	1.2	43.4	0.20
Min	0	0	0	2.13	0.010

Table 2. Test event description

Event	Date	Duration	Maximum of discharge (m^3s^{-1})	Mean discharge (m^3s^{-1})	Cumulative rainfall (mm)	Intensity ($mm.h^{-1}$)
13	Sept. 00	26 h	454,2	70	230	40

2.3 Artificial Neural network

2.3.1. General presentation.

A neural network is a combination of parametrized functions called neurons that calculate their parameters thanks to a database using a training process [10]. The most

popular model is the multilayer perceptron (MLP), which generally contains one or more hidden layers of nonlinear neurons and one output linear neuron. Each hidden neuron computes a non-linear function of a weighted sum of the input variables, then the output neuron computes the linear combination of the outputs of the hidden ones.

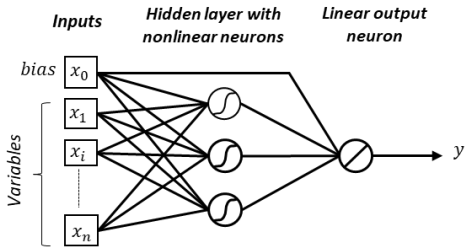


Fig. 2. Multilayer perceptron with a single hidden layer

The MLP is very popular due to its two main properties: universal approximation and parsimony. The first one states the capability to successfully approximate any differentiable function with an arbitrary level of accuracy [11]. The latter states how the multilayer perceptron needs fewer parameters to successfully fit a non-linear function, compared to others statistic model that linearly depend on their parameters [12]. The more general model of neuron calculates it output *y* as following:

$$y = f\left(\sum_{j=1}^{n} c_j \,.\, x_j\right) = f(x_1, \dots, x_n;\ c_1, \dots, c_n), \tag{1}$$

with x_j, the input variable *j*; c_j, the parameter linking the variable x_j to the neuron; $f(.)$, the activation function (usually a sigmoid). The dynamic properties of the identified process can be considered thanks to three kinds of models [13].

- ***Static model***

The static model is a digital filter with a finite impulse response. It calculates the following equation:

$$\hat{y}(k) = \varphi(\,\mathbf{x}(k), \dots, \mathbf{x}(k - n_r + 1), \mathbf{C}) \tag{2}$$

with $\hat{y}(k)$**,** the estimated output at the discrete time *k*; φ_{rn}, the non-linear function implemented by the model; **x** is the input vector**;** ***n*,** the sliding time-windows size defining the length of the necessary exogenous data**; C**, the vector of the parameters. This model is known for having more parameters than the following models.

- ***Recurrent model***

The recurrent model allows identification of dynamical processes (Infinite Impulse Response), it is implemented following the equation (3).

$$\hat{y}(k) = \varphi(\,\hat{\mathbf{y}}\,(k-1), \dots, \hat{\mathbf{y}}\,(k-r);\ \mathbf{x}(k), \mathbf{x}(k-1) \dots, \mathbf{x}(k - n_r + 1);\ \mathbf{C}) \tag{3}$$

With *r*, the order of the recurrent model; n_r, the depth of the sliding time-window used to consider the input variables. Ones must distinguish the recurrent variable (*y*) from the exogenous variables (**x**). This model can deliver forecasts for an undetermined forecasting horizon providing the availability of the exogenous variables.

- ***Feed-forward model***

In the feed-forward model, the recurrent input is substituted by the measurements of the process output at previous times step. This model is non recurrent; but it can identify dynamical processes. This model is the most used and generally provides the best results. Nevertheless, we have observed that it generally has difficulties to model the dynamics of the process (cited in Artigue et al 2012). It calculates:

$$\hat{y}\,(k) = \varphi(\mathbf{y}\,(k-1), \ldots, \mathbf{y}(k-r);\ \mathbf{x}(k), \mathbf{x}(k-1)\ \ldots, \mathbf{x}(k-n_r+1);\ \mathbf{C}) \qquad (4)$$

with $y(.)$, the observed value of the modelled variable at the discrete time *k*.
These three categories of models will be compared in this study.

2.3.2. Training

As data-driven models, neural networks design is based on a database. Training consists in calculating the set of parameters of the model in order to minimize the least square cost function on the training set [10]. Because the model is non linear, this minimization is iteratively calculated.

Nevertheless, as the goal of the model is to be able to generalize the trained behavior to any set of data never seen, the quality of the model must be validated on another set, independent from the training set that is called "test set". The bias-variance dilemma [14] shows an important limitation: the training error is not representative of the test error, and the difference increases with the complexity of the model (*i.e.* the number of free parameters of the model). The bias-variance dilemma may be avoided using regularization methods.

2.3.2. Regularization methods

Early stopping

Early stopping was presented by [15] as a regularization method. It consists in stopping the training before the full convergence. To this end a supplementary subset, called stop set, is defined those goal is to evaluate the ability of generalization of the model during the training. This subset is independent from the training set. Training is stopped when the error on the stop set begins to increase. The stop set is used to stop the training, the performances of this set are thus overestimated compared to those of the test set. Nevertheless, this set is usually (improperly) called "validation set" in the literature.

Cross validation

Proposed by [16], cross validation allows to select a model having the lower variance. To this end the training set is divided in *K* subset and the error is calculated on the remaining (*K*-1) subsets in the training set. After *K* trainings, the cross-validation score is calculated, for example by the mean of the previously obtained errors. Based on the cross validation score it is possible to select the model that has the lowest variance, minimizing by this way the bias on the training set and the variance on validation sets. This method allows to select input variables, the order (r), and the number of hidden neurons.

Ensemble model

Darras et al. [9] showed that, surprisingly, cross validation was not able to successfully select the best initialization of parameters. In order to diminish the sensitivity of the output to the parameter's initialization, they propose to create an ensemble model of $\mathcal{M}$ members [17] and to calculate the output of the ensemble, at each time step, by the median of the $\mathcal{M}$ members.

2.3.3 Design of the model

In this study, regularization methods are applied by: (*i*) dividing the dataset in three subsets (training, stop and test sets), (*ii*) using cross correlation to select the architecture of the model in the following succession: inputs (n_r) except for rain gauges, order (r), number of hidden neurons (h), and (*iii*) using 20 members in the ensemble.

Three kinds of sliding window widths are tried based on the rainfall-runoff cross-correlogram.

2.3.4. Performance criteria

Several criteria are used to assess the performance of a model. The determination coefficient R^2 [18]; the *Synchronous percentage of the peak discharge (S_{PPD})* and the *Peak delay* as two peak assessment criteria [4]. They are briefly detailed below:

- **R^2 criterion**

$$R^2 = 1 - \frac{\sum_{k=1}^{n}(y_k - \hat{y}_k)^2}{\sum_{k=1}^{n}(y_k - \overline{y_o})^2}, \quad (5)$$

with the same notations as previously.

The nearest than 1 the Nash-Sutcliff efficiency is, the best the results are. Nevertheless, this criterion can reach good values even if the model proposes bad forecasts.

- **Peak analysis**

The quality of the flood prediction is analyzed regarding the quality of the peak using two criteria defined by [4].

Synchronous percentage of the peak discharge: SPPD

The synchronous percentage of the peak discharge: SPPD [4] is a relevant criterion to assess flash flood modeling performance of a model on the peak discharge. It shows the simulation quality at the peak discharge through the ratio between the observed and simulated discharges at the observed peak discharge moment (k_o^{max}).

$$S_{PPD} = 100\frac{\hat{y}^{k_o^{max}}}{y^{k_o^{max}}}, \quad (6)$$

Peak delay (P_D)

The peak delay [4] indicates the duration between the maximum of simulated peak and measured peak. When the estimated peak is in advance, the peak delay is negative.

$$P_D = k_s^{max} - k_o^{max}, \quad (7)$$

with $k_{\cdot}^{max}$ the instant of the peak of discharge (simulated or observed).

2.5. Extracting information: KnoX method

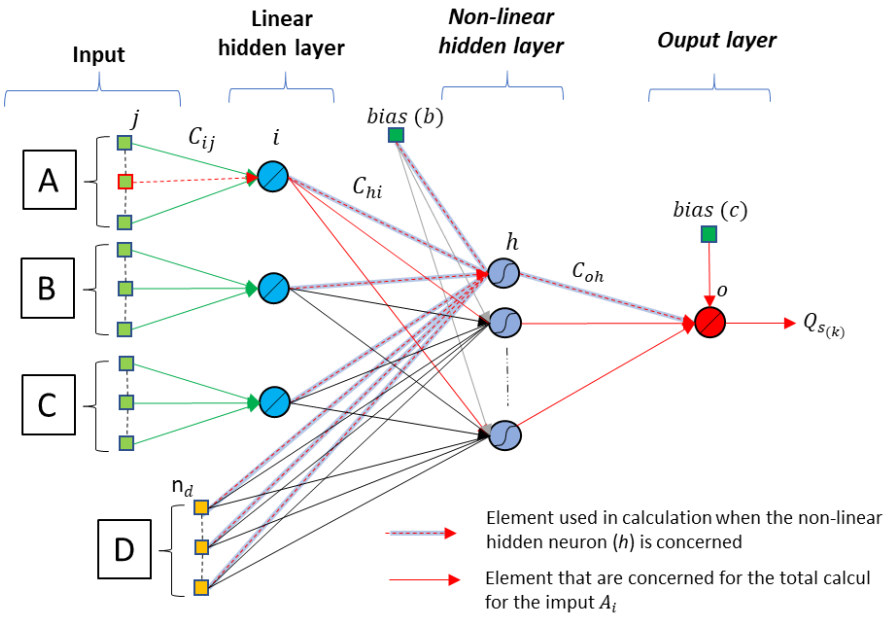


Fig. 3. Application of the KnoX method on the deep multilayer perceptron

$$P_{A(j)} = \frac{{}^{M}|C_{ij}|}{\sum_{i=1}^{n_A} {}^{M}|C_{ij}|} \sum_{h=1}^{H} \left(\frac{{}^{M}|C_{hi}|}{\sum_{i=1}^{n_i} {}^{M}|C_{hi}| + \sum_{d=1}^{n_d} {}^{M}|C_{hd}| + b_h} \right) \left(\frac{{}^{M}|C_{oh}|}{\sum_{h=1}^{H} {}^{M}|C_{oh}| + c_o} \right), \quad (8)$$

and: $P_A = \sum_{j=1}^{n_A} (P_{A(j)})$ (9)

The KnoX method [8, 19] allows to calculate a simplified contribution of each input to the model output. This method is described for the general deep model (2 hidden layers) shown in Fig. 3. The principle of the method is that a contribution of an individual input variable can be quantified after training, by the product of the parameters linking this input to the output. The considered parameters are (*i*) "normalized" by the sum of the parameters linked to the same targeted neuron, and (*ii*) regularized by calculating the median of absolute values of their values for 20 different random initializations. This regularized value is noted as ${}^{M}|C_{ij}|$ for the parameter C_{ij} linking the neuron (or input) *j* to the neuron *i*.

Regarding the model shown in Fig. 3, the contribution (P_A) of the input *A* (group of several delayed inputs) is the sum of the contributions of each individual delayed input of the group A. The equation calculating the contribution for just one element of the

input A is provided in eq. (8). It is not possible to explain more comprehensively the method in the short present paper, so we suggest to the reader to refer to [8].

3 Results

Starting from previous works of [4] we chose the following exogenous variables: (*i*) *Barre des Cevennes* rain gauge, *Saint-Roman de Tousque* rain rauge and *Mialet* rain gauge, each one with a sliding window length $\{k, \dots k\text{-}n_r\text{+}1\}$, (*ii*) the sum of the mean rain (over the three gauges) fallen from the beginning of the event. Of course, a bias input is used; several values were tried in order to evaluate the sensitivity of the KnoX method to its value.

3.1. Window widths selection thanks to correlation analysis

Widths of the rainfall windows applied to the model are selected thanks to cross correlation. Initially proposed by [20] Jenkins and Watts (1968), [1] generalizes the application of cross correlation in hydrology. The used equation in this study is presented in eq. (9).

$$C_{xy}(k) = \frac{Cov\,(x_i, y_{i+k})}{\sigma_x\,\sigma_y} = \frac{\frac{1}{n}\sum_{i=1}^{n-k}(x_i - \bar{x})(y_{i+k} - \bar{y})}{\sigma_x\,\sigma_y} \quad , \tag{9}$$

With $k = 0, 1, \dots$;. where m is the truncation which is recommended to be $m=n/3$ (Mangin 1984). [20]) indicated that 2 hydrological variables can be considered as statistically independent if their cross-correlation is superior to 0.2. We thus select three possible lengths for the sliding windows of rain gauges inputs: (*i*) the number of time step between C_{xy} =0 and C_{xy} =0.2, that defines the memory effect; (*ii*) the window between C_{xy} =0.2 (positive slope) and C_{xy} =0.2 (negative slope) and (*iii*) all the m positive values of C_{xy}. Based on [20] the correlations between gauges as well as response times are indicated in Table 3.

Table 3. Correlation analysis of the data

Rain gauge		Mialet (h)	SRDT (h)	*BDC* (h)
Average response-time		2	3	4.5
Response-time range		1 – 3.5	2.5 – 4.5	4 – 5.5
Rainfall-discharge average cross-correlation		0.40	0.455	0.44
Rain gauge cross-correlation	Mialet	--	0.58	0.45
	SRDT	--	--	0.61

3.2. Model selection

A partial cross-validation score was operated on a subset of 17 most intense events in the database [3]. The number of hidden neurons was increased from 1 to 10. The best

model was chosen according to the highest cross-validation score S_v estimated as following:

$$S_v = \frac{1}{K}\sqrt{\sum_{i=1}^{K}|E_i|^2}, \quad (9)$$

Where E_i is the validation error of the subset *i* used in partial cross validation.

The output values are the result of the median of the outputs of an ensemble of 20 members differing only by their initialization before training.

Three bias values are considered (0.01; 0.1; 1), three depths of sliding windows (see section 3.1) and three kinds of models (see section 2.3), 27 different models have been designed following the procedure indicated in section 2.3.3. The best one in each kind of models has been chosen, regarding the test event, in order to have efficient models to analyze. Architectures presented in Table 4 were thus selected.

Table 4. Selected models

Input variables	**Static**	**Recurrent**	***Feed-Forward***
Rain-gauge window width (n_r) (BDC/SRDT/Mialet)	32/32/23	27/28/20	32/32/23
Rain cumul window width	3	3	3
Order (r)	x	3	3
Number of hidden nonlinear neurons	2	10	5
Bias value	1	0.01	1

3.3 Results

Obtained test set hydrographs are shown in the Fig 4 and their performances described in Table 5. It appears in
Fig. 4 and **Table 5** that the best results are provided by the feed-forward model. This is usual because the feedforward model uses the previous observations of the modelled variable in input. The recurrent model is usually not as efficient but exhibits better dynamics, which is also frequently observed [4]. The static model presents an acceptable performance, being able to generate 63% of the peak discharge.

Table 5. The models performances on the test set

Model	*R^2*	*SPPD %*	*P_D(0.5h)*
Static	0.83	63,3	1
Recurrent.	0.89	78.5	0
Feed-Forward	0.99	99.3	1

After having verified that the models are convenient, it is possible to apply the KnoX method. The extracted contributions are presented in **Table 5**.

Regarding the rainfalls, one can note that in general, SRDT is the station with the highest contribution. The contributions do not change significantly for *Mialet* through all the models. BDC and *Mialet* are probably affected by their location close to the border of the basin whereas SRDT is close to the middle of the basin.

Regarding the balance between the state variables and the rainfalls, it appears that when the previous observed discharge is used as an input variable, it brings almost 50

% of the contribution to the output. This observation means that the model does not pay enough attention to rain inputs and this could be the reason of the sensitivity to parameters initialization. Beside this, it also appears that the state variables in the static model have lesser contribution than they do in the other two models. In general, from the static model to the feed-forward one, the total contributions of the state variables are respectively 45%, 61 % and 65 %, where the biggest parts are imputed to the previous observed discharge (feed-forward). These observations are fully consistent and the results seem highly interpretable.

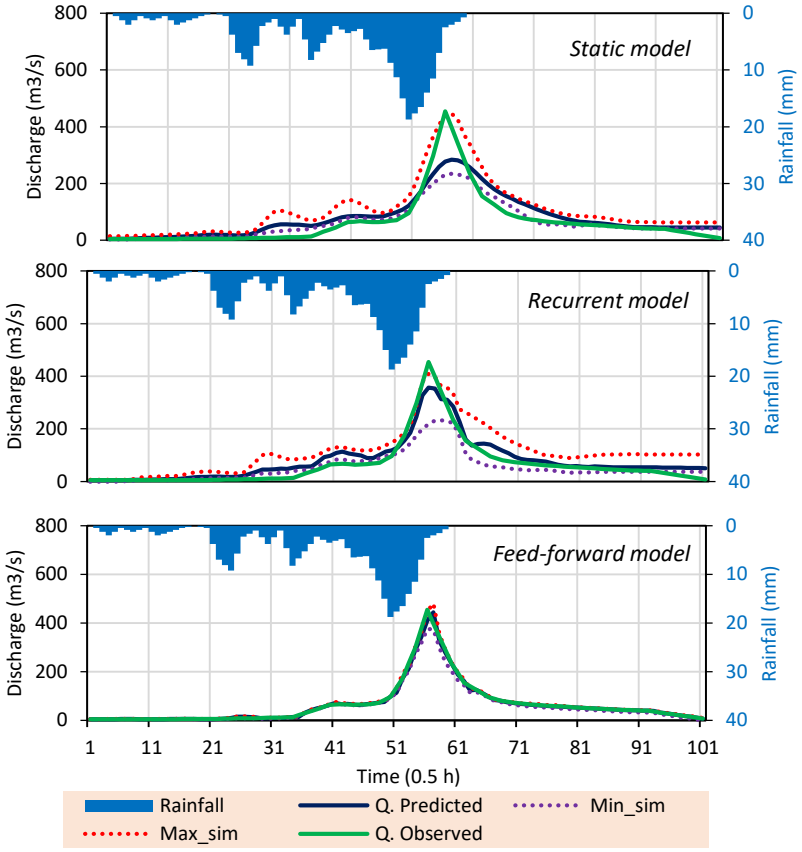


Fig. 4. Hydrographs for the test set. Min_sim and Max_sim correspond to the minimum and maximum values of the ensemble model. Q is the median of the 20 members of the ensemble.

Table 6. Contributions (P_A) for the variables, from each model, expressed in %.

Name of variable	***Static***	***Recurrent***	***Feed-forward***
BDC	13 %	12 %	5 %
SRDT	31 %	17 %	22 %
Mialet	11 %	11 %	9 %
Cumulated rainfall	31 %	20 %	12 %
Previous Q. obs	--	--	45 %

Previous Q. calc	--	25 %	--
bias	14 %	16 %	8 %

4 Interpretation

These results show how the kind of model can modify the contribution of explanatory variables on an observed phenomenon. Thus, some kind of models must be preferred when it comes to represent physical relations. It is also shown that the mean cumulative rainfall used here as a state variable plays a great role in models where the previous discharge is not used as input. This state variable seems to have a great interest in hydrologic modelling. The value of the bias, surprisingly, seems to have a role. It is usually interpreted as the base flow. Nevertheless, its behavior is consistent: it shows more involvement when the previous observed discharges are not used as input; then by complementarity with the humidity information, it guides the models to acceptably approximate the real discharge information.

5 Conclusion

Prediction of flash flood events is a very challenging task in the *Cévennes* range. It was previously realized using neural networks but sometimes appeared difficult to understand because of the specific behaviors of the models. In order to be able to improve these models, the present work takes steps to better understand the processes involved in such events. To this end, the KnoX method, developed to extract information from a neural network model was applied to the *Gardon de Mialet* Basin. The obtained results show that by using relevant variables properly combined on whatever the network used here, efficient model can be built out. Besides, the KnoX method allows to see how the variables are handled by the model to approximate the phenomenon. There has been evidence that the variables do not express themselves in the same way through the different models used. As it is understandable, sometimes, the choice for a model is commanded by the situations in presence. The information extracted from the network can probably be used to compare to some physical meaningful characteristics of watershed or events, such as the Thiessen polygons, the response time, the cross correlation etc. It provided also some guidelines to deal with the sensitivity of the model to the parameter's initialization.

6 Aknowledgement

The authors thank the METEO-France weather agency, the SPGD flood-forecasting agency for providing rainfall datasets. Our gratitude is extended to Bruno Janet for the stimulating collaboration shared with the SCHAPI Unit, and to Roger Moussa and Pierre Roussel-Ragot for the helpful discussions and support. The constant effort made by Dominique Bertin and the Geonosis Company to enhance and develop the neural network software RNF Pro are thereby acknowledged as well.

7 References

1. Rouzeau, M., Xavier M., and Pauc, J.C. 2010. “Retour d’expériences des inondations survenues dans le departement du Var les 15 et 16 juins 2010.” http://cgedd.documentation.developpement-durable.gouv.fr/documents/cgedd/007394-01_rapport.pdf.
2. J. Roberts, Stephen, and Will Penny. 1981. Neural Networks: Friends or Foes? Sensor Review. Vol. 17. London: MCB University Press.
3. Toukourou M., Johannet A., Dreyfus G.,Ayral P.A. 2011. Rainfall-runoff Modeling of Flash Floods in the Absence of Rainfall Forecasts: the Case of "Cévenol Flash Floods", App. Intelligence, 35 2,178-189.
4. Artigue, G,et al. 2012. "Flash Flood Forecasting in Poorly Gauged Basins Using Neural Networks: Case Study of the Gardon de Mialet Basin (Southern France)". NHESS, 12(11): 3307-24.
5. Oussar, Yacine, and Gérard Dreyfus. 2001. "How to Be a Gray Box: Dynamic Semi-Physical Modeling." Neural Networks 14 (9): 1161-72. https://doi.org/10.1016/S0893-6080(01)00096-X
6. Johannet, Anne, B Vayssade, and Dominique Bertin. 2007. "Neural Networks: From Black Box towards Transparent Box - Application to Evapotranspiration Modelling." Int. Journal of Comp. Int. 24 (1): 162.
7. Kong-A-Siou, L., et al, S.: KnoX method, or Knowledge eXtraction from neural network model. Case study on the Lez karst aquifer (southern France), J. Hydrol., 507, 19–32.
8. Darras, T., et al. 2015. Identification of spatial and temporal contributions of rainfalls to flash floods using neural network modelling: case study on the Lez basin (southern France) Hydrol. Earth Syst. Sci., 19, 4397–4410, 2015
9. Darras, T., Johannet, A., Vayssade, B., Kong-A-Siou, L. and Pistre, S. (2014). Influence of the Initialization of Multilayer Perceptron for Flash Floods Forecasting: How Designing a Robust Model, (ITISE 2014), Ruiz, IR, Garcia, GR Eds, 687-698.
10. Dreyfus, G. 2005. Neural networks, methodology and applications, Springer, Berlin.
11. Hornik, Kurt, Maxwell Stinchcombe, and Halbert White. 1989. "Multilayer Feedforward Networks Are Universal Approximators." Neural Networks 2 (5): 359-66.
12. Barron, A R. 1993. "Universal Approximation Bounds for Superpositions of a Sigmoidal Function." IEEE Trans. Inf. Theor. 39 (3): 930-45. https://doi.org/10.1109/18.256500.
13. Nerrand, O., P. Roussel-Ragot, L. Personnaz, G. Dreyfus, and S. Marcos. 1993. "Neural Networks and Nonlinear Adaptive Filtering: Unifying Concepts and New Algorithms." Neural Comp 5 (2): 165-99.
14. Geman, Stuart, Elie Bienenstock, and René Doursat. 1992. "Neural Networks and the Bias/Variance Dilemma." Neural Computation 4 (1): 1-58.
15. Sjöberg, J., et al. 1995. "Nonlinear Black-Box Modeling in System Identification: A Unified Overview." Automatica 31 (12): 1691-1724.
16. Stone, M. 1976. "Cross-Validatory Choice and Assessment of Statistical Predictions (With Discussion)." Journal of the Royal Statistical Society: Series B (Methodological) 38 (1): 102-102.
17. Dietterich T.G., 2000. Ensemble Methods in Machine Learning, in J. Kittler and F. Roli (Ed.), First Int. Workshop on Multiple Classifier Systems, LNCS, p. 1-15, New York: Springer Verlag.
18. Nash, J.E., and J.V. Sutcliffe. 1970. "River Flow Forecasting through Conceptual Models Part I - A Discussion of Principles." Journal of Hydrology 10 (3): 282-90.
19. Jenkins, G.-M., Watts, D.-G. 1969 “Spectral analysis and its applications”, Holden-Day, - 525 pages
20. Mangin, Alain. 1984. "Pour Une Meilleure Connaissance Des Systèmes Hydrologiques à Partir Des Analyses Corrélatoire et Spectrale." Journal of Hydrology 67 (1-4): 25-43.