water

MDPI

Article

Uncertainty Estimation in Hydrogeological Forecasting with Neural Networks: Impact of Spatial Distribution of Rainfalls and Random Initialization of the Model

Nicolas Akil [1,2,*], Guillaume Artigue [1,*], Michaël Savary [2,*], Anne Johannet [1] and Marc Vinches [1]

[1] HydroSciences Montpellier, Univ. Montpellier, IMT Mines Ales, IRD, CNRS, 30100 Ales, France; anne.johannet@mines-ales.fr (A.J.); Marc.vinches@mines-ales.fr (M.V.)
[2] AQUASYS, 2 Rue de Nantes, 44710 Port-Saint-Père, France
[*] Correspondence: nicolas.akil@aquasys.fr (N.A.); guillaume.artigue@mines-ales.fr (G.A.); michael.savary@aquasys.fr (M.S.)

Abstract: Neural networks are used to forecast hydrogeological risks, such as droughts and floods. However, uncertainties generated by these models are difficult to assess, possibly leading to a low use of these solutions by water managers. These uncertainties are the result of three sources: input data, model architecture and parameters and their initialization. The aim of the study is, first, to calibrate a model to predict Champagne chalk groundwater level at Vailly (Grand-Est, France), and, second, to estimate related uncertainties, linked both to the spatial distribution of rainfalls and to the parameter initialization. The parameter uncertainties are assessed following a previous methodology, using nine mixed probability density functions (*pdf*), thus creating models of correctness. Spatial distribution of rainfall uncertainty is generated by swapping three rainfall inputs and then observing dispersion of 27 model outputs. This uncertainty is incorporated into models of correctness. We show that, in this case study, an ensemble model of 40 different initializations is sufficient to estimate parameter uncertainty while preserving quality. Logistic, Gumbel and Raised Cosine laws fit the distribution of increasing and decreasing groundwater levels well, which then allows the establishment of models of correctness. These models of correctness provide a confidence interval associated with the forecasts, with an arbitrary degree of confidence chosen by the user. These methodologies have proved to have significant advantages: the rigorous design of the neural network model has allowed the realisation of models able to generalize outside of the range of the data used for training. Furthermore, it is possible to flexibly choose the confidence index according to the hydrological configuration (e.g., recession or rising water table).

Keywords: neural networks; uncertainty; hydrogeology; probability; probability density function; model; rainfall

check for updates

Citation: Akil, N.; Artigue, G.; Savary, M.; Johannet, A.; Vinches, M. Uncertainty Estimation in Hydrogeological Forecasting with Neural Networks: Impact of Spatial Distribution of Rainfalls and Random Initialization of the Model. *Water* **2021**, *13*, 1690. https://doi.org/10.3390/w13121690

Academic Editor: Renato Morbidelli

Received: 28 April 2021
Accepted: 14 June 2021
Published: 18 June 2021

1. Introduction

Water is an essential resource for life on Earth but also a hazard, through its scarcity during droughts or its abundance during floods. Water-related risks sometimes cause damage and fatalities and have a strong impact on water supply, agriculture and industries. The current climate change context has causes the rise of extreme phenomena frequency and duration [1]. Moreover, water demand is growing in developed countries due to change in water uses, thus becoming a major issue. Predictive systems can be used in order to manage and to prevent these hydro(geo)logical risks. Among the available solutions to forecast groundwater level or river discharge, two stand out. The first consists in using physically based models, which are supposed to represent a deep knowledge of the study basin. Unfortunately, this level of knowledge is often difficult to reach because of the heterogeneity and anisotropy of hydro-systems. Besides, these models require meteorological forecasts whose reliability, at the necessary space and time scales, can be insufficient. The second consists in statistical modeling, among which artificial neural networks (ANN) are widely represented. This family of models does not require a strong knowledge about the system behavior, but does require a database representing the assumed relationship between input and output data. Moreover, ANN do not necessarily require forecasts of their inputs to provide output forecasts, as long as the lead-time remains below or close to the response time of the system. A specific model, the multilayer perceptron, is known to be able to identify any nonlinear but differentiable function thanks to the universal approximation property [2]. The choice of ANN models thus relies on poor knowledge of the underground processes of the area and on very low operational calculation times. This avoids making an uncertain hypothesis about the hydrosystem. As any model will, ANN generate uncertainties that are difficult to quantify and to communicate, the absence of which would optimize the decision-making process for end users. Especially, as is often the case in hydrology and hydrogeology, if the decision is based on threshold crossing, ambiguities in decisions are not acceptable. Thus, uncertainties can lead to mistakes and inconsistencies in decision-making. For these reasons, [3] focus on the key issues for modelers, especially the issue of how far the model has been able to capture the catchment behavior and [4] focuses on the origins of the uncertainties. [4] described three main origins for these uncertainties: (i) input data, especially noise and non-measured spatial variability of these data, (ii) oversimplified structure and (iii) parameters determination. The origins of input data uncertainties are mainly related to data quality, input representativity in the basin related to spatial distribution, environmental conditions and errors from measurements (resolution, measuring instruments) [4,5]. The uncertainties in the model parameters are found in training performance, as well in initialization during the training step [4]. Bayesian Model Averaging was developed for parameter uncertainty estimation [6] and applied to the Pô streamflow in Italy. Uncertainty is represented as a forecast interval with a certain probability of correctness [6]. As long as the chosen model generates multiple outputs with an ensemble model, an interval of uncertainty can be drawn. The uncertainty related to various input variables can thus be addressed, for example, the spatial distribution of rainfalls. The latter was approached with a Bayesian Forecasting System [7], coupled to a precipitation forecast, with interesting results [8], indicating that the uncertainty, estimated by the prediction interval delivered by the ensemble model, could be improved by post processing. Therefore, Bayesian models can be an alternate solution to estimate model uncertainties. This approach is therefore an effective method for approximating the uncertainty of the various hyperparameters of a model.

In the present paper, we propose a methodology to estimate the uncertainty generated by both the neural network model itself and by the non-measured spatial heterogeneity of rainfall. This work was carried out on the Champagne chalk aquifer (Northern France) as a case study, at a 10-day time-step. Predictions are achieved for up to 20 days (two time-steps). A reliable enough ANN model to forecast Champagne chalk groundwater level is built, thanks to a rigorous variable and complexity selection process [9,10] helped by the application of regularization methods [10], mainly cross-validation [11] and early stopping [12]. The uncertainties due to the model parameters are then estimated [13]. Even though the noise in rainfall inputs is assumed to be limited due to the quality of rain gauges and to the 10 day time-step, the spatial variability of rainfall is significant, and the number of rain gauges may be insufficient to properly represent this variability. An original method to assess this uncertainty is to perform permutations and substitutions of the available rain gauges, simulating a spatial variation of rainfall and thus giving an ensemble of forecasts. These new forecasts are finally embedded in the assessment and the representation of uncertainties method.

The article is organized as follows: the material and methods section presents the neural networks and the method used for model design. Then, the target basin used to implement this method, the Champagne chalk groundwater basin, the quality criteria used and, finally, the method used to estimate the uncertainties are described. A correctness model implementation method is proposed, allowing the display of a confidence interval, choosing an a priori confidence index. Then the design method deployed to realize the neural network model is detailed in Section 3. Section 4 goes on to present the results of the prediction and an uncertainty estimation. Section 5 proposes a discussion and some paths of improvement, before the general conclusion.

2. Material and Methods

2.1. Neural Network Models

2.1.1. Definitions

An artificial neuron is a mathematical operator that first calculates its potential, i.e., the weighted sum of its inputs with its parameters, and secondly its output, applying a nonlinear transformation to its potential.

Neurons are combined inside a network following an architecture, which is built according to the targeted function: classification or regression. Neurons can be organized in layers of two types: (i) output layer, whose outputs are those of the model, and associated to measured values, and (ii) hidden layers, whose outputs are not associated to measured values [14].

One of the most common forms of neural network architecture in hydrology is the feed-forward model "multilayer perceptron" (MLP), for which the universal approximation property has been shown by [2,15]. This property states that the model (Figure 1) is able to approximate any differentiable function with an accuracy depending on the number of hidden neurons. The other property of this architecture is parsimony. This means that the model needs less parameters to describe phenomena, compared to other statistical nonlinear models [16]. This comes from the calculation of the output, which depends non-linearly on the inputs and the parameters. Parsimony is even more valuable when the number of input variables increases. For these reasons, this model is used in this study.

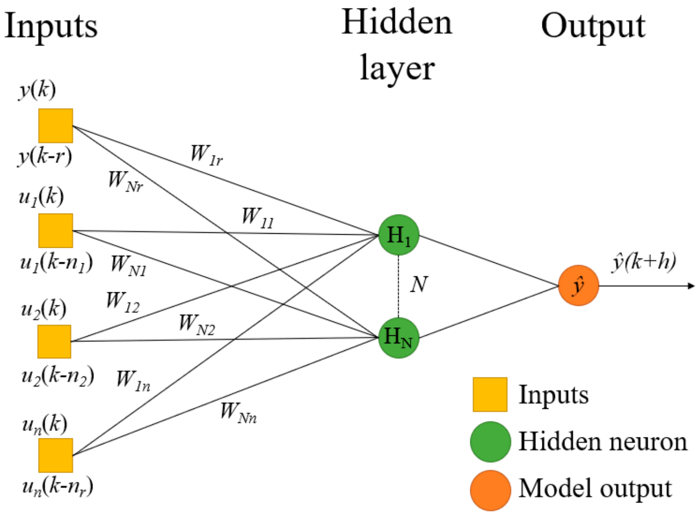


Figure 1. Multilayer Perceptron representation, with x_i, the exogenous variables; **W**, the matrix of parameters; y, the measured output; $\hat{y}$, the predicted output; r, the order of the model, n_r, the input window width; H_i (i = 1 to N) the hidden neurons; N, the number of hidden neurons; k, the discrete time and h, the lead time [13].

2.1.2. Role of Time in Neural Networks Models

Assuming the crucial role of time in forecasting, a neural network can have a dynamic behavior, according to its architecture [17].

In the first case, the static character of the model implies that time has no functional role and that input variables are all exogenous (1). The model is a finite impulse response filter.

$$\hat{y}(k) = \varphi(\mathbf{u}(k), \ldots, \mathbf{u}(k - n_r + 1), \mathbf{W}) \tag{1}$$

where *ŷ(k)* is the estimated output at the discrete time *k*; u the input vector; φ the nonlinear function implemented by the neural model; n_r the sliding time-window size that defines the length of the necessary history of exogenous data; and W the matrix of parameters.

In the second case, the recurrent model uses the result of the simulation at previous time-steps in addition to the exogenous variables (2). The model is thus an infinite impulse response filter and presents a dynamic behavior.

$$\hat{y}(k) = \varphi\left(\hat{\mathbf{y}}(k-1), \ldots, \hat{\mathbf{y}}(k-r); \mathbf{u}(k), \ \ldots, \ \mathbf{u}(k-n_r+1), \ \mathbf{W}\right) \tag{2}$$

with the same notations as previously stated, and r, the order of the recurrent model; that is to say, the number of previous output values applied at the input of the model.

The recurrent model allows the simulation of a dynamic function: it is used when the noise added by output measurements is supposed to be higher than that affecting inputs. Practically, recurrent models must also be used when real-time data are unavailable [18].

In the third case, the feed-forward model, the recurrent inputs are substituted by the measurement of the output variables at previous time steps (3). The model is thus a finite impulse filter. It is static, rigorously speaking, but thanks to the addition of observed state variables as exogenous input variables, it is able to simulate dynamic behavior.

$$\hat{y}(k) = \varphi(\mathbf{y}(k-1), \ldots, \ \mathbf{y}(k-r); \mathbf{u}(k), \ \ldots, \ \mathbf{u}(k-n_r+1), \ \mathbf{W}) \tag{3}$$

where *y(k)* is the observed output of the simulated system at the discrete time *k*.

Feed-forward models are used if the noise due to the measurement of the output variables is low, or lower than the noise on inputs [17,19].

2.1.3. Training and Overfitting

Training a neural network consists of calculating the parameters set by minimizing a cost function, measuring the error between observed and simulated values. This stage uses a training rule applied on a subset of the database: the training set.

Afterward, the quality of the model is assessed on another subset: the test set. The test is used to assess the property of "generalization" of the model. It is never used during training, nor for optimization of the architecture.

Ref. [20] showed that the training error is not a relevant estimator of the test error. Indeed, the training error diminishes when the complexity (number of free parameters) increases, while the test error (the variance) increases. This phenomenon is called "bias/variance dilemma", and indicates that a too complex model perfectly fits the training data, also including the random noise contained by these data. This model is thus unable to correctly generalize to unknown data. Conversely, a too simple model will not be able to adapt to the signal. This leads to a high bias and a low variance.

2.1.4. Regularization Methods

Regularization methods can be used to prevent overfitting. The cross-validation method [11] is used in this study. It allows assessment of the generalization error and provides the cross-validation score that assesses the variance. It is used to select the variables and the complexity of the model (see Section 3).

The second regularization method used in this study is early stopping. Early stopping [12] considers that training the model too much increases the amplitude of the parameters and is equivalent to increasing the complexity. It thus stops training before the generalization capacity decreases. For this purpose, a subset of the database, called the "stop set" must be defined.

As shown by [21,22], the ensemble model can be used to reduce the uncertainty due to the initialization of parameters before training. At each time-step, the output is calculated as (4):

$$\hat{y}_M^k = Median_X\left(\hat{y}_i^k\right) \tag{4}$$

where $\hat{y}^k_M$ is the output of the ensemble model at time-step k; y^k_i is the output of one member of the ensemble at time-step k; and Median$_X$ represents the median calculated on the outputs of a set of X models. The choice of the number of models in the ensemble depends on the application.

2.1.5. Model Design

The first step in the model design is to split the database into several subsets: training set, stop set and test set. Various test sets can be chosen depending on the modelling purpose. For example, the most intense event can be chosen when floods are targeted.

The second step consists in choosing the relevant kind of model regarding the role of time (Section 2.1.2).

The third step is to realize cross-correlations between input data and between input and output data. These correlations allow the obtaining of a response time and a memory effect duration for each input variable. This gives a first overview of the hydro-system dynamics and allows preselecting the input windows width of the model and a relevant lead-time (Figure 1).

Then, in order to prevent overfitting, the architecture of the model is optimized using the cross-validation method. For this optimization, the output is that of the ensemble model. This architecture selection is carried out by adjusting the following hyperparameters using cross-validation:

- The window widths of the different (exogenous) input variables (n_r in Equations (1)–(3)).
- The "order" of the model, corresponding to the window width of the estimated (or observed, if the model is a feed-forward) targeted variable (output variable), for previous time-steps, applied at the input of the model (r in Equations (2) and (3)).
- The number of neurons in the hidden layers: N.

2.2. Study Area: The Champagne Chalk Groundwater Basin

2.2.1. Field Study Presentation

- Location Located in Northern France, in the Grand-Est region, the Champagne chalk groundwater basin area is estimated at 5927 sq.km. It corresponds mainly to the drainage of the rivers Marne and Aube, delimited by piezometric ridges characterized as follows: chalk limit on the eastern part, tertiary rocks on the western part, other hydrogeological basins on the northern limit, the Seine river for the southern part and, as a bedrock, marlstones [23]. Elevation varies from 40 to 286 m.a.s.l. (Figure 2).
- Water use Water is mainly used for tap water production and agriculture [23]. Annual water withdrawals via studied piezometer made on average between 2012 and 2017 are 17,393 m^3, however, showing a decreasing trend [24]. Water is also used for agriculture, with 61.5% of groundwater withdrawal for irrigation in 2017 (against 38.5% for tap water production) in Vailly (location of the studied piezometer) and neighboring towns [25].
- Climate The climate of the basin corresponds to a transition climate between oceanic and semi-continental climates. The mean annual rainfall varies from 640 to 820 mm, measured on 22 meteorological stations calculated on the 1981–2010 period [26]. The recharge is estimated at 160 mm/y [23].
- Geology and groundwater behavior This basin is mainly composed of chalk, and limestones to a lesser proportion, with sands and clay along the hydrographic network [27,28]. Intense shallow fracturing, mainly caused by climate action, has developed a significant permeability especially near the hydrographic network. Groundwater recharge time in the champagne basin is estimated at 100 days in our study piezometer (Craie à Vailly (nouveau)) [29], and the underground levels can increase from 6 m to 25 m [23,30]. Groundwater levels, especially in the Barbuise catchment area, which is close to the study piezometer, are influenced by the shallow water [27]. Consequently, the Barbuise river discharge is strongly correlated to piezometric levels at Craie à Vailly [27,29].

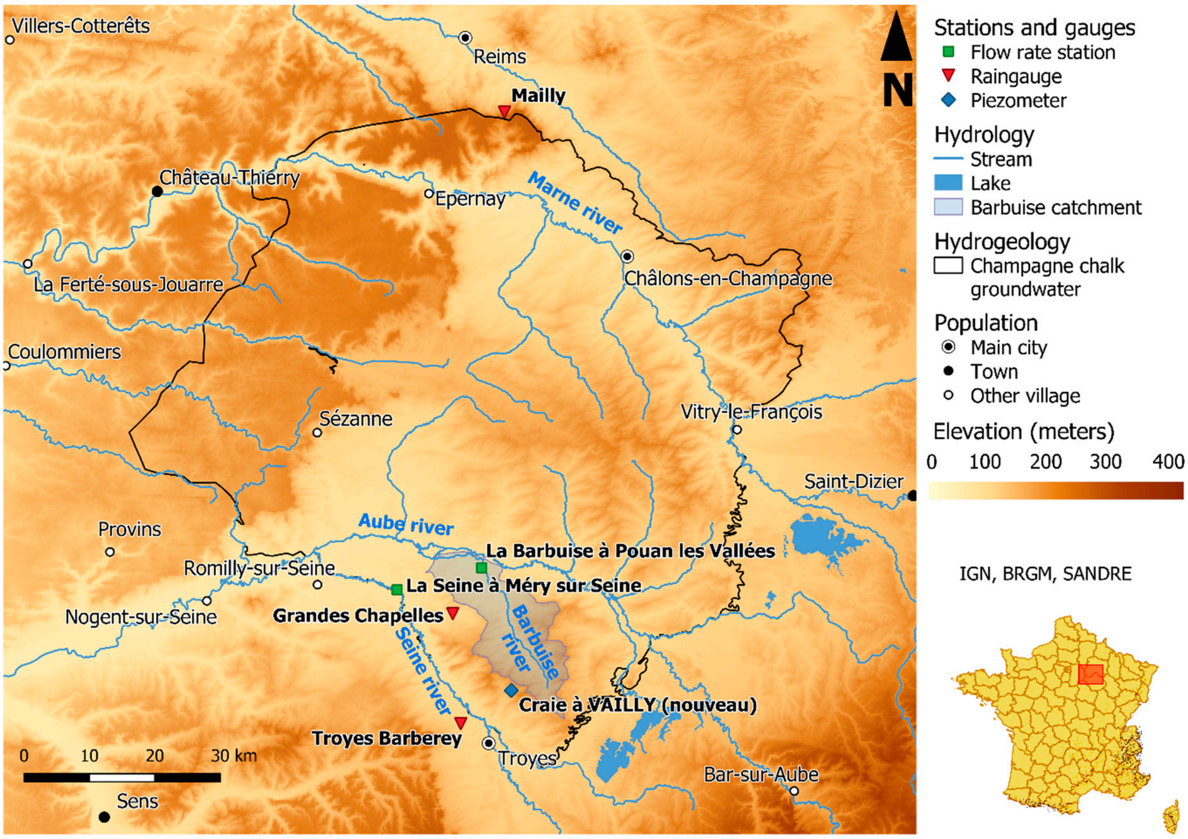


Figure 2. Map of the Champagne chalk groundwater basin (IGN, BRGM, BD Carthage).

2.2.2. Database Presentation

Meteorological data are provided at three stations:

1. Troyes-Barberey (R_{TB}) (precipitation and potential evapotranspiration),
2. Grandes-Chapelles (R_{GC}) (precipitation),
3. Mailly (R_{MA}) (precipitation)

Groundwater levels are available at the Craie à Vailly piezometer (L_{CV}), and two discharge stations are located in Pouan-les-Vallées in Barbuise catchment (D_{BP}) and at Méry-sur-Seine in the Seine catchment (D_{SM}) (Table 1).

Table 1. Simple statistics on Champagne chalk time series from 1977 to 2018.

Station Name	Measured Variable	Unit	Time Step	Max Value	Min Value	Median	Average
Craie at Vailly (L_{CV})	Level	m.a.s.l.	10 days	134.75	109.75	119.95	120.558
Barbuise at Pouan les Vallées (D_{BP})	Discharge	$m^3.s^{-1}$	10 days	4.50	0.00	0.67	0.836
Seine at Méry-sur-Seine (D_{SM})	Discharge	$m^3.s^{-1}$	10 days	182.2	5.95	25.61	35.71
Grandes-Chapelles (R_{GC})	Rain	mm	10 days	131.2	0.0	15.9	19.87
Troyes-Barberey (R_{TB})	Rain	mm	10 days	86.4	0.0	13.2	17.29
Mailly (R_{MA})	Rain	mm	10 days	138.8	0.0	17.0	21.50
Troyes-Barberey (PET)	Potential Evapotranspiration	mm	10 days	64.7	0.0	19.0	21.20
Bassin (I)	Irrigation	$m^3.ha^{-1}.month^{-1}$	month	833.9	0.6	176.0	280.1

Irrigation (I) information was calculated from cultivated area and agricultural data. The cultivated area can be estimated thanks to (RPG2017 from IGN [31]). Agricultural data contains monthly water demand for each crop type (OUGC84). Therefore, it is possible to establish the monthly water demand by crop type for the total area of the basin. Monthly irrigation needs were then resampled at 10-days sampling rate.

Data ranges from 1977 to 2018 at a 10-day time-step, with a gap between August 1991 and January 1995 and another one from January 2014 to May 2014. Smaller gaps exist in data but never exceed two months, as in April 1985, December 1995, and early 2003 and 2005. Data from October to December 1990 have been set apart due to potential errors in groundwater level measurements, because the data are constant at 116.65 m.a.s.l. during this period. Even though short gaps (one or two time steps) were filled by simple interpolations, the other more important gaps were not filled due to the lack of information about water table variability during these gaps. This is particularly true during periods of extreme levels, such as during the test set, which has been shortened for this reason. Irrigation data are provided on Appendix A Figure A1.

Cross correlations between inputs and output were calculated to better understand the behavior of the basin. These provide information on input-output relationships by showing the response time (mean delay between rain peak and discharge/groundwater level peak, in number of time steps) and the memory effect [32]. They help define the reasonable lead-time that can be reached. Correlation analysis is synthetized in Table 2.

Table 2. Correlation analysis. Diagonal shows the memory effect (in number of time-steps) when simple correlation is calculated (orange). When cross-correlation is calculated, blue cells show memory effect and green cells show response time. NC means that correlation score is always under 0.2, showing a very weak correlation, leading to possible misinterpretations of the memory effect.

	L_{CV}	ΔL_{CV}	D_{BP}	D_{SM}	R_{GC}	R_{TB}	R_{MA}	PET	I
L_{CV}	17	6	2	5	21	22	15	12	16
ΔL_{CV}	15	4	−4	0	1	1	1	4	8
D_{BP}	17	1	11	3	2	6	7	9	12
D_{SM}	19	4	15	5	1	1	1	3	7
R_{GC}	NC	2	NC	3	0	0	0	0	27
R_{TB}	NC	2	NC	3	1	1	0	0	27
R_{MA}	NC	3	NC	3	0	1	0	0	33
PET	19	11	16	9	NC	NC	NC	8	4
I	22	14	19	12	NC	NC	NC	11	7

Regarding the first line of Table 2 showing the response times of all variables over water levels at Craie at Vailly (L_{CV}) piezometer, it appears that the shortest response time is two time-steps for the discharge at Barbuise at Pouan les Vallées (D_{BP}). This indicates that, statistically, the discharges at Barbuise at Pouan les Vallées have a greater influence on the groundwater at Craie at Vailly after two time-steps delay. And that this response time is the shorter. This confirms the quick interaction between surface water and groundwater. Regarding the impact of surface water on both the water quality and the groundwater level, the two time-step lead-time was thus chosen. In this way, a lead-time of 20 days (two time-step) is considered as a good compromise between model accuracy and end users' needs. A shorter lead-time would reduce the interest of the forecast for the end users, while a longer lead-time would require the availability of the Barbuise at Pouan les Vallées discharge forecast. Thus, this lead time ensures that available inputs explain the output.

2.3. Quality Criteria

It is important to use impartial criteria to evaluate the quality of result. This study investigates two modeling goals: groundwater level prediction and uncertainty quantification. As a result, two kinds of criteria will be used, as presented below.

2.3.1. Quality of Fitting and Prediction

The Pearson's correlation coefficient allows quantifying of the linear relation between two variables. It varies between -1 to +1, and is the covariance divided by the product of the standard deviations of the two variables (5).

$$r^2 = \frac{cov(u,\ y)}{\sigma_u \sigma_y} \tag{5}$$

where $cov(u,\ y)$ is the covariance between variables u and y, σ_u and σ_y, respectively, are their standard deviation.

- The persistency criterion

This criterion was proposed by [33] for prediction (6). It must be close to 1. A 0 value represents the score of the naive forecasting (prediction value = the actual value), and a negative value means that the forecasting is even worse than the naive forecasting.

$$C_P = 1 - \frac{\sum_k^n (y_{k+h} - \hat{y}_{k+h})^2}{\sum_k^n (y_k - y_{k+h})^2} \tag{6}$$

where y_k is the measured value at the discrete time k, $\hat{y}_{k+h}$ is the simulated value at the discrete time k, y_{k+h} is the observed value at the discrete time k+h, h, the lead time, and n the number of samples of the considered dataset.

2.3.2. Uncertainties Quantification

The following criteria assume that several models are available for prediction, and that a prediction interval can therefore be defined between the largest and smallest of the forecast values, at each time step.

- Prediction Interval Coverage Probability

This criterion expresses the empirical probability that the prediction interval contains the measured value (7). It represents a kind of accuracy of the predicted value. In the (7) equation, $f(.)$ is the function of belonging to the prediction interval [34,35].

$$C_{PICP} = \frac{1}{n}\sum_k^n f\left(y_k, \hat{y}_k^{min}, \hat{y}_k^{max}\right) \tag{7}$$

with the same notations as before, and $f(y_k) = 1$ if $y_k \in \ [\hat{y}_k^{min}; \hat{y}_k^{max}]$, else $f(y_k) = 0$

- Mean Prediction Interval

The Mean Prediction Interval, C_{MPI}, is the average of all the results set of the interval of prediction calculated at each time-step. It quantifies the mean scattering of the prediction [34], following (8).

$$C_{MPI} = \frac{1}{n}\sum_{k=1}^n \left(\hat{y}_k^{max} - \hat{y}_k^{min}\right) \tag{8}$$

with the same notations as before, y_k the measured value at the discrete time k and $\hat{y}_k^{max}$ and $\hat{y}_k^{min}$ the upper and lower bounds of the forecast interval.

- Prediction Confidence Criterion

The Prediction Confidence Criterion, C_{PC}, is a ratio quantifying the performance of a predictor for providing a prediction having the highest empirical probability of lying within the smaller prediction interval (9). It is simply defined by the ratio between the two previous criteria [13].

$$C_{PC} = \frac{C_{PICP}}{C_{MPI}} = \frac{\frac{1}{n}\sum_{k}^{n} f\left(y_k, \hat{y}_k^{min}, \hat{y}_k^{max}\right)}{\frac{1}{n}\sum_{k=1}^{n}\left(\hat{y}_k^{max} - \hat{y}_k^{min}\right)} \tag{9}$$

with the same notations as before, y_k is the measured value at the discrete time k and $\hat{y}_k^{max}$ and $\hat{y}_k^{min}$ the upper and lower bounds of the forecast interval at discrete time k.

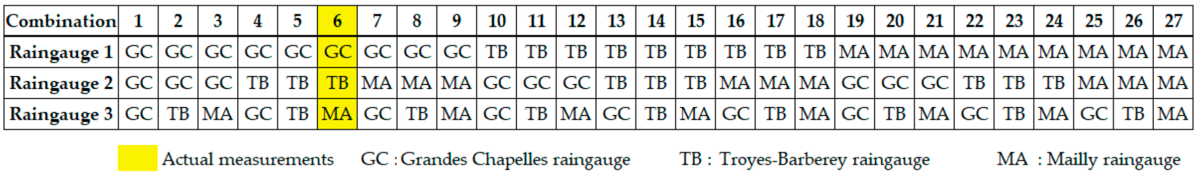
2.4. Uncertainties Linked to the Initialization of the Parameters and to the Spatial Variability of the Rains

2.4.1. Variability Due to the Initialization of Parameters

As presented in Section 2.1.4, the implementation of an ensemble model makes possible a reduction in the variability of the outputs caused by the random initialization of the model parameters during the training step. However, the number of members in the ensemble model needs to be determined in order to sufficiently reduce this variability. The purpose is to obtain the smallest prediction interval, measured by the MPI criterion (Mean Prediction Interval) [34], whereas this interval includes a maximum of the observed values, measured by PICP criterion (Prediction Interval Coverage Probability) on the subset of interest, considered as a whole [34,35]. Therefore, the optimal number of random initializations (members of the ensemble) can be determined thanks to the calculation of the C_{PC} criterion (Prediction Confidence Criterion) [13] defined in Section 2.3.2, that synthetizes both criteria.

2.4.2. Spatial Rainfall Variability

In order to approach the uncertainty caused by the spatial variability of rainfalls, usually poorly described by a small number of rain gauges, once the training is over, we propose to run permutations and substitutions of rain gauges applied to the model inputs. This allows generation of another ensemble model of possible forecasts obtained in the test. As three rain gauges are available for the studied basin, 27 combinations, shown in Figure 3, composed the ensemble model. This new kind of ensemble model, whose members differ by the combination of used rain gauges, is denoted as ensemble$_{\text{-RG}}$. All the permutations made give a range of outputs that can be considered as a prediction interval related to the spatial variability of rain.



Combination	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Raingauge 1	GC	GC	GC	GC	GC	GC	GC	GC	GC	TB	TB	TB	TB	TB	TB	TB	TB	TB	MA	MA	MA	MA	MA	MA	MA	MA	MA
Raingauge 2	GC	GC	GC	TB	TB	TB	MA	MA	MA	GC	GC	GC	TB	TB	TB	MA	MA	MA	GC	GC	GC	TB	TB	TB	MA	MA	MA
Raingauge 3	GC	TB	MA	GC	TB	MA	GC	TB	MA	GC	TB	MA	GC	TB	MA	GC	TB	MA	GC	TB	MA	GC	TB	MA	GC	TB	MA

🟨 Actual measurements GC : Grandes Chapelles raingauge TB : Troyes-Barberey raingauge MA : Mailly raingauge

Figure 3. Combinations of rainfall input data with three rain gauges.

2.5. Estimation of Empirical Confidence Intervals Using Probability Density Functions

2.5.1. Method

The purpose of this section is to present the method used to estimate a confidence interval for predictions achieved by the model. The original process consists in:

- Establishing the frequencies of appearance of the water level classes histogram; this is then considered as an empirical probability density function (pdf) of the data;
- Fitting a theoretical well-known pdf, for example the normal one, to the empirical pdf by adjusting its parameters. If necessary, thanks to the Expectation-Maximization algorithm (EM) [36,37], the theoretical pdf can be a composition of several pdfs of the same type, each one having different parameters; this composite pdf is called the target pdf. The algorithm provides the constituent parameters of the theoretical

elementary theoretical pdfs as well as the weights that enable them to be assembled to fit the target pdf;

- Starting from target pdf, determining a probability of occurrence of the measured value inside the predicted interval for each class;
- For a given confidence index (for example 95%), and for each class, supposing the data verify the constraints of a normal law and establishing a model of "correctness" using the erf (error function). This provides the estimated error associated to each class;
- Finally, drawing the possible errors on the water chart.

The Expectation-Maximization algorithm follows two different steps: the expectation step and the maximization step. The expectation step consists in defining an expected value for log-likelihood parameters of the target pdf. The maximization step consists in obtaining parameters that maximize the expected value, using an iterative process [36].

The Expectation-Maximization process is applied four times to four categories of samples: (i) samples with a negative slope ($y_i^k - y_i^{k+1} > 0$), (ii) samples with a positive slope ($y_i^{k+1} - y_i^k > 0$), (iii) samples included inside the MPI and (iv) samples not included inside the MPI. Discriminating between negative and positive slopes is useful in considering the current conditions in the system: decreasing water levels are the sign of a draining of groundwater whereas increasing water levels are the sign of a refill of groundwater stock. This corresponds to two different physical behaviors that should not be mixed to capture the guiding factors of the system.

2.5.2. Chosen Probability Density Functions

Nine probability density functions were chosen (Table 3) (Equations (10)–(18)).

Table 3. Equations of nine probability density functions.

Name of *pdf* Law	Formula	Eq.	References
Normal	$\mathcal{N}(\overline{x},\, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\overline{x})^2}{2\sigma^2}}$	(10)	[38,39]
Gumbel	$\mathcal{Gum}(\overline{x},\beta) = \frac{e^{-\frac{(x-\overline{x})}{\beta}}\, e^{-e^{-\frac{(x-\overline{x})}{\beta}}}}{\beta}$	(11)	[40]
Laplace	$\mathcal{Lap}(\overline{x},\, b) = \frac{1}{2b} e^{-\frac{(\lvert x-\overline{x}\rvert)}{b}}$	(12)	[41]
Raised Cosine	$\mathcal{Cosr}(\overline{x},\, s) = \frac{1}{2s}\left(1 + \cos\left(\frac{(x-\overline{x})}{s}\, \pi\right)\right)$	(13)	[42,43]
Cauchy	$\mathcal{Cau}(x_0,\, a) = \frac{1}{\pi a\left(1+\left(\frac{x-x_0}{a}\right)^2\right)}$	(14)	[44,45]
Logistic	$\mathcal{Logist}(\overline{x},\, s) = \frac{e^{-\frac{(x-\overline{x})}{s}}}{s\left(1+e^{-\frac{(x-\overline{x})}{s}}\right)^2}$	(15)	[46]
Slash	$\mathcal{Sla}(x) = \frac{\varphi(0)-\varphi(x)}{x^2}$	(16)	[47]
Bhattacharjee	$\mathcal{Bhat}(\overline{x},\sigma_x,\, a) = \frac{1}{2a}\left(\Phi\left(\frac{x-\mu+a}{\sigma}\right) - \Phi\left(\frac{x-\mu-a}{\sigma}\right)\right)$	(17)	[48]
Huber	$\mathcal{Hub}(z) = \frac{1}{2\sqrt{2\pi}\left(\Phi(z)-\frac{\phi(z)}{z-\frac{1}{2}}\right)}\, e^{-\rho_z(x)}$	(18)	[49,50]

where x is the variable, $\overline{x}$ its average, x_0, its median, σ its standard deviation, σ^2 its variance, a, b, β and s scale parameters, φ the normalized normal distribution, ϕ the normal law, Φ the cumulative normal law, z the degree of robustness and ρ_z the Huber loss. The Huber loss depends on the degree of robustness and can be written following Equation (19) [49].

$$\rho_z(x) = \begin{cases} \frac{1}{2}x^2 & ,\lvert x\rvert \le z \\ z\lvert x\rvert - \frac{1}{2}z^2 & ,\lvert x\rvert > z \end{cases}, \tag{19}$$

3. Model Design

3.1. Definition of Subsets for Training Testing, Stop and Cross-Validation

As presented previously in Section 2.1.4, it is necessary to define the training, stop and test sets. For this purpose, it is usual to split the database into three subsets. To be consistent with the needs of the end users, the test set is chosen as the driest period of the database: the 1988–1990 period. It will be used to assess the quality of the model generalization. Requesting that the model will be able to predict correctly the extreme dry period of the database is a strong requirement to ensure generalization capacity. The 2011–2013 period is used for the stop set (12th subset). The remainder is the training set. Cross validation is used to select input variables and complexity, as presented in Section 2.1.5. To run cross-validation, we also have to define the cross-validation subsets inside the training set. Each one must contain a sufficient amount of data. A length of three years (108 samples based on the 10-day time step) is thus chosen for each cross-validation subset. Thus, 13 cross-validation subsets are defined as shown in Table 4. They are used in validation of each one its turn, and for each turn all other cross-validation subsets are used in training (T on Table 4). Therefore, the training set is divided in two kinds of subset: 12 training subsets and a cross-validation subset. At the end of the 13 sequences of training, 13 cross-validation scores are calculated and the resulting global cross-validation score is calculated as follows (20):

$$S_{CV} = \frac{1}{D} Median\left(C_{P_q}\right) \tag{20}$$

with D=13 the number of cross validation subsets, C_P the score of persistency, and q the number of the considered subset.

Table 4. Split subsets; T is for training, V for validation, S for stop, T$_\mathrm{e}$ for test; C_{P_q} is for the score calculated on the q subset.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Scores
V	T	T	T	T	T	T	T	T	T	T	S	T	T	T$_\mathrm{e}$	C_{P_1}
T	V	T	T	T	T	T	T	T	T	T	S	T	T	T$_\mathrm{e}$	C_{P_2}
T	T	V	T	T	T	T	T	T	T	T	S	T	T	T$_\mathrm{e}$	C_{P_3}
T	T	T	V	T	T	T	T	T	T	T	S	T	T	T$_\mathrm{e}$	C_{P_4}
															…
T	T	T	T	T	T	T	T	T	T	V	S	T	T	T$_\mathrm{e}$	$C_{P_{11}}$
T	T	T	T	T	T	T	T	T	T	T	S	V	T	T$_\mathrm{e}$	$C_{P_{13}}$
T	T	T	T	T	T	T	T	T	T	T	S	T	V	T$_\mathrm{e}$	$C_{P_{14}}$
														Median	

3.2. Choice of the Model and Complexity Selection

A single hidden layer model is used because of its lower complexity whereas its performance is sufficient for the modelling objectives.

According to the recommendations made in Section 2.1.2, the chosen predictor is the feed-forward model because the measurement of water level is accurate, which is not the case for the rainfall-field estimation. Indeed, not only are the rain gauges inaccurate, but the rain also has a spatial variability not sufficiently described by the three available rain gauges. Therefore, it can be assumed that the noise affecting the output of the process (the water table) is lower than the noise affecting the inputs (the rain).

Hyperparameters window-width ranges for rains, irrigation, and shallow water discharge, were chosen using correlation analysis as shown in Table 5, suggested by [10].

For training, the Levenberg-Marquardt algorithm, which is a second order learning method [51,52], has been used with 100 epochs for each experience.

Several architectures were tried with different complexities, and for each one the cross-validation score was calculated in order to select the best. Table 3 synthetizes the investigated architectures and the selected model for the two time-steps' lead-time (justified in Section 2.2), in order to simulate groundwater levels at a middle term in a drought context. The selected model is also shown in Figure 4. Table 5 presents the selected hyperparameters during the design stage.

Table 5. Selected architecture using the design procedure synthetized in 2.1.5.

Model Element	Selected Hyperparameters		Tested Range Values	
Order	r (L_{CV})	3	(3–6)	(8–14)
Exogenous input window-widths	n_1 (I)	8	(7–10)	
	n_2 (PET)	12	(9–12)	(9–12)
	n_3 (D_{SM})	5	(2–5)	
	n_4 (D_{BP})	5	(2–5)	(2–5)
	n_5 (R_{GC})	2	(1–4)	(7–12)
	n_6 (R_{TB})	2	(1–4)	(7–12)
	n_7 (R_{MA})	3	(1–4)	(7–12)
Number of hidden neurons	N	3	(2–10)	(2–10)

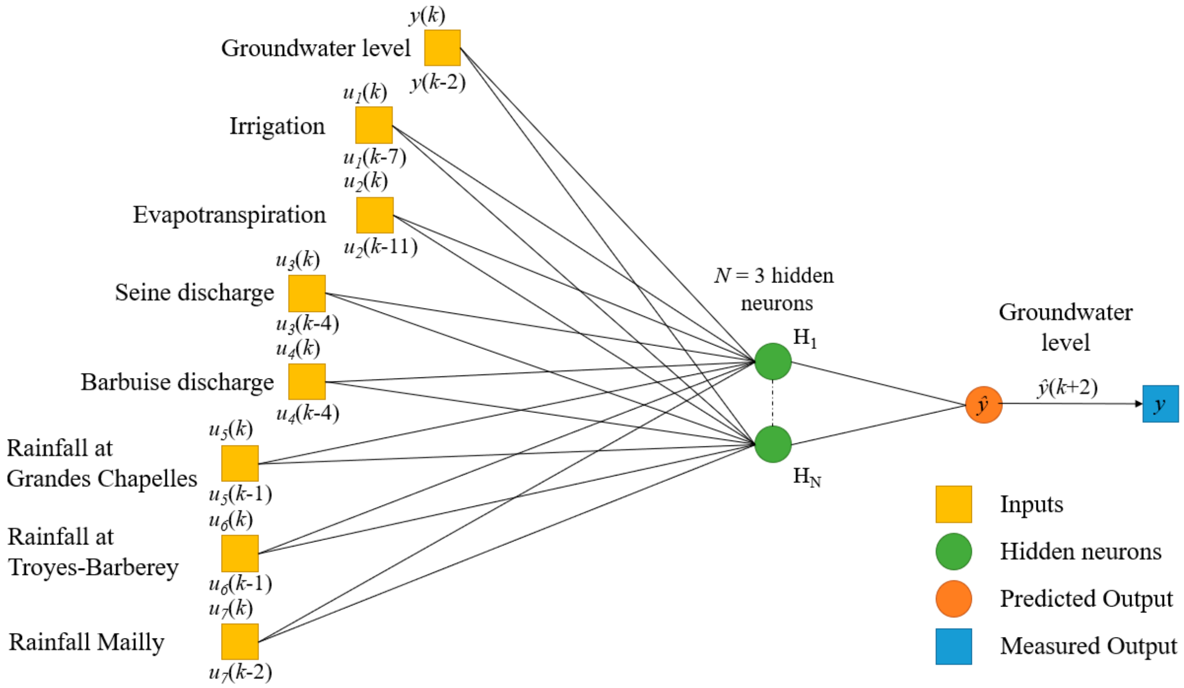


Figure 4. Selected architecture.

4. Results

4.1. Optimal Number of Members in Ensemble Models

Once the architecture is selected as presented in 2.1.5 and in 3.1.2, ensemble forecasts are calculated between 3 to 120 members in each (respectively 3, 5, 10, 20, 30, 40, 50, 60, 80, 100 and 120 forecasts). C_{PC}, Prediction Confidence Criterion, is calculated for each ensemble, allowing definition of the optimal number of members, i.e., the number of members whose parameters are randomly initialized (X in Equation (4)).

Figure 5 presents the evolution of the C_{PC} versus the number of members in the ensemble models. Schematically, the curve can be approximated by two straight lines whose intersection is at around 40 members. The first line decreases when the number of members in the ensemble increases, corresponding to a stage where the MPI increases. The second line corresponds to a plateau that indicates the stability of the two criteria that make up the $C_{PC.}$ The intersection of two lines corresponds to the minimal number of initializations for which the gain of ensemble starts to become stationary. We thus propose this value (40 members in Figure 5) as the number X of members. Although the C_{PC} could possibly be enhanced by using more members, the cost-benefit ratio (especially regarding calculation time) pleads in favor of this choice.

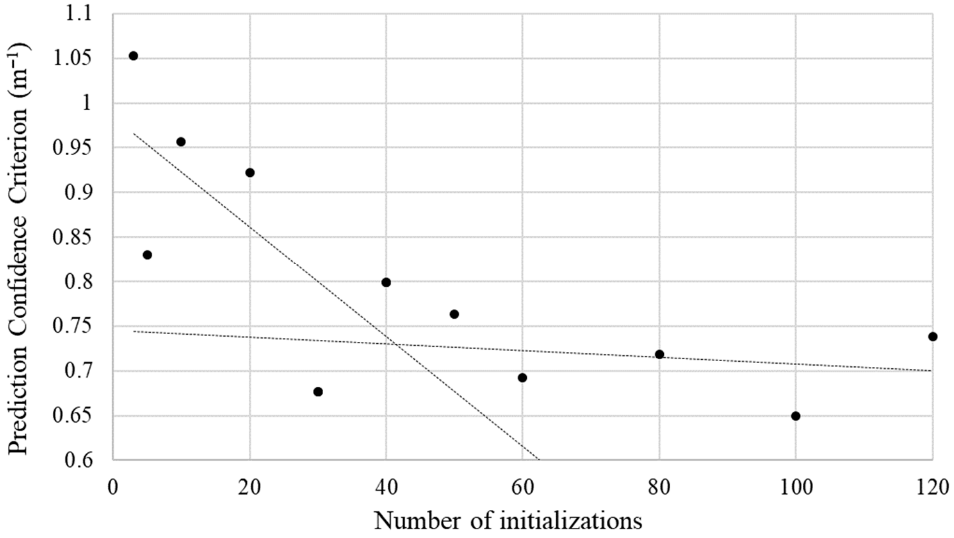


Figure 5. C_{PC} cross-validation score of ensemble models as a function of the number of members (X).

4.2. Prediction Results

The cross-validation persistency score for the architecture reaches C_p = 0.65 and the test score is C_p = 0.40. The performances are thus lower on the test set than on the other sets during cross validation. This is consistent with the choice of the test set, which corresponds to the drier period of the database. Nevertheless, the quality of the forecast presented in Figure 6, made for the year with the driest summer in the entire database, shows that: (1) the model is capable of generalizing to periods of extreme behavior, (2) confirms the interest in being able to visualize uncertainty, so that the manager can analyze the most uncertain parts of the limnigram. As a reminder, the three last months of 1990 were not considered due to possible errors in piezometric levels.

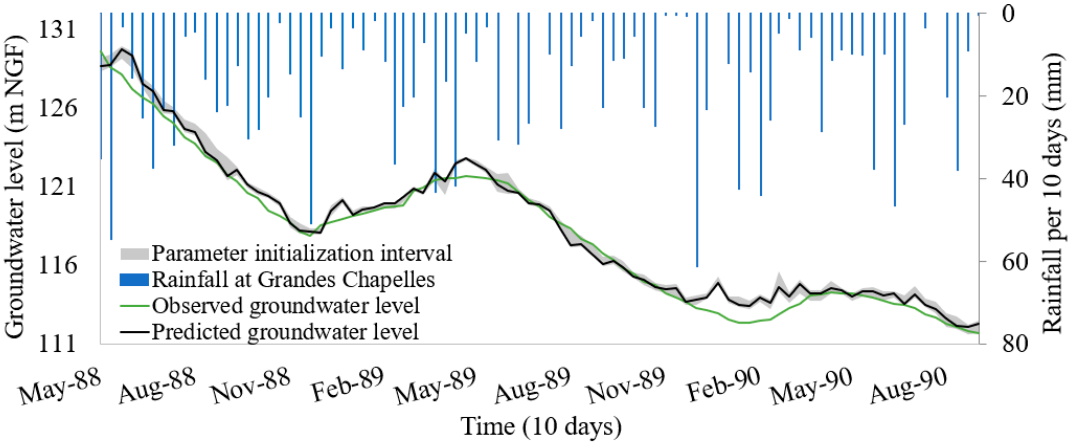


Figure 6. Prediction of groundwater levels at 20 days lead-time on the test set. The three last months of 1990 were not taken into account due to possible errors in piezometric levels. C_p = 0.40; C_{PICP} = 0.39; C_{MPI} = 0.62 m.

Figure 6 also shows the prediction interval for the test event. In this case, P_{ICP} = 0.39, M_{PI} = 0.62 m, and C_{PC} = 0.63 m^{-1}. It appears that the prediction is fairly close to the measurement except for the early spring of 1990, for which the forecast level is not low enough. On the other hand, the grey band showing the uncertainty is very thin and does not contain enough observed values (P_{ICP} = 0.39) to be able to inspire confidence in the end users.

4.3. Representation of Uncertainties Caused by the Initialization Parameters

4.3.1. Theoretical Composite pdf for Four Distributions

The empirical *pdf*s of training and stop sets are represented in one *pdf* using classes of observed water levels of 20-cm-wide. This interval is the result of a compromise between the class width and the number of samples in each class.

Remember that we have chosen to represent four distributions: two for the sign of the slope of the groundwater evolution, and two for the belonging or non-belonging of the observed sample to the prediction interval. For each distribution, a theoretical composite *pdf* is built as presented in Section 2.5.2.

To illustrate the procedure, let us consider the specific distribution of water levels with a decreasing slope, using a Normal law as theoretical law. The obtained theoretical composite *pdf* is presented in (Figure 7).

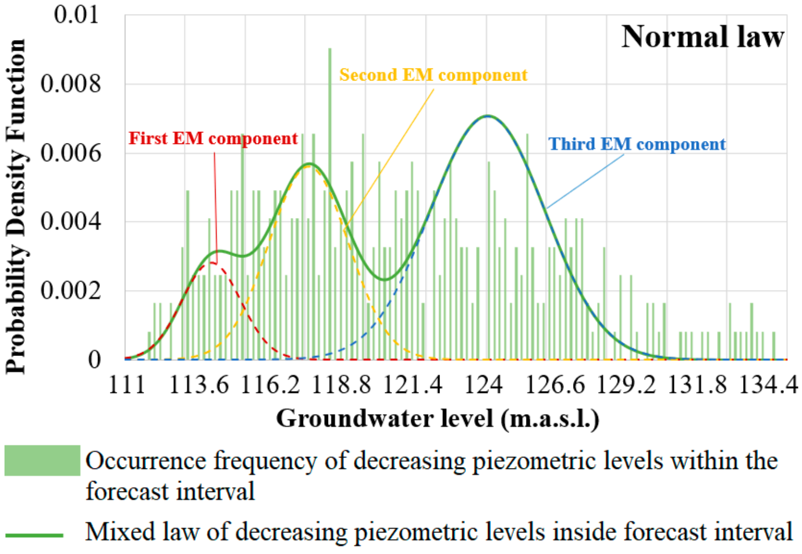


Figure 7. Theoretical composite *pdf* of piezometric levels, obtained using Esperance Maximization (EM) algorithm using a Normal law and applied to observed values having a decreasing slope and being inside the prediction interval. Each elementary normal law is represented with a different color and denoted as ”EM component.”

One can note that the water levels from 115 to 125 m.a.s.l show the highest frequencies whereas above 132 m.a.s.l. or under 115 m.a.s.l. observed groundwater values are underrepresented.

According to the Figure 7, the Esperance Maximization (EM) algorithm provides the parameters for each elementary distribution. For the Normal law, these parameters are the mean, the variance and the lambda, which is the maximum amplitude of each elementary distribution. The theoretical mixed *pdf*, fitted by measured groundwater level distribution classes, is obtained by the sum of the three components, as shown in Figure 7. This process is repeated for each of the three other groundwater levels configurations, and for the nine different studied laws (presented in Table 3).

Let us now examine which of the theoretical laws presented in Table 3 provides the best C_{PICP} on the Train+Stop dataset. To this end Figure 8 shows the representation, explained in Figure 7, for each one of the theoretical laws, regarding the two distributions of values inside the prediction interval (green) or outside this interval (red). Table 6 shows the correlations between the empirical distribution and the theoretical laws, for the measured groundwater levels inside the prediction interval $\left(r_{in}^2\right)$ and outside the prediction interval $\left(r_{out}^2\right)$. Best correlations are shown in green and worst in red. It appears in Table 6 that the best adjustment is achieved by the Raised Cosine theoretical law.

Table 6. Pearson's correlation coefficients between the distributions with negative slope of groundwater evolution, and the theoretical composite *pdf*; r^2_{in} applies to measurements inside prediction interval, r^2_{out} applies to measurements outside the prediction interval.

Law	Normal	Gumbel	Laplace	Raised Cosine	Cauchy	Logistic	Slash	Bhatta-Charjee	Huber
r^2_{in}	0.62	0.74	0.66	0.76	0.69	0.75	0.64	0.64	0.50
r^2_{out}	0.68	0.74	0.67	0.77	0.70	0.75	0.72	0.75	0.65

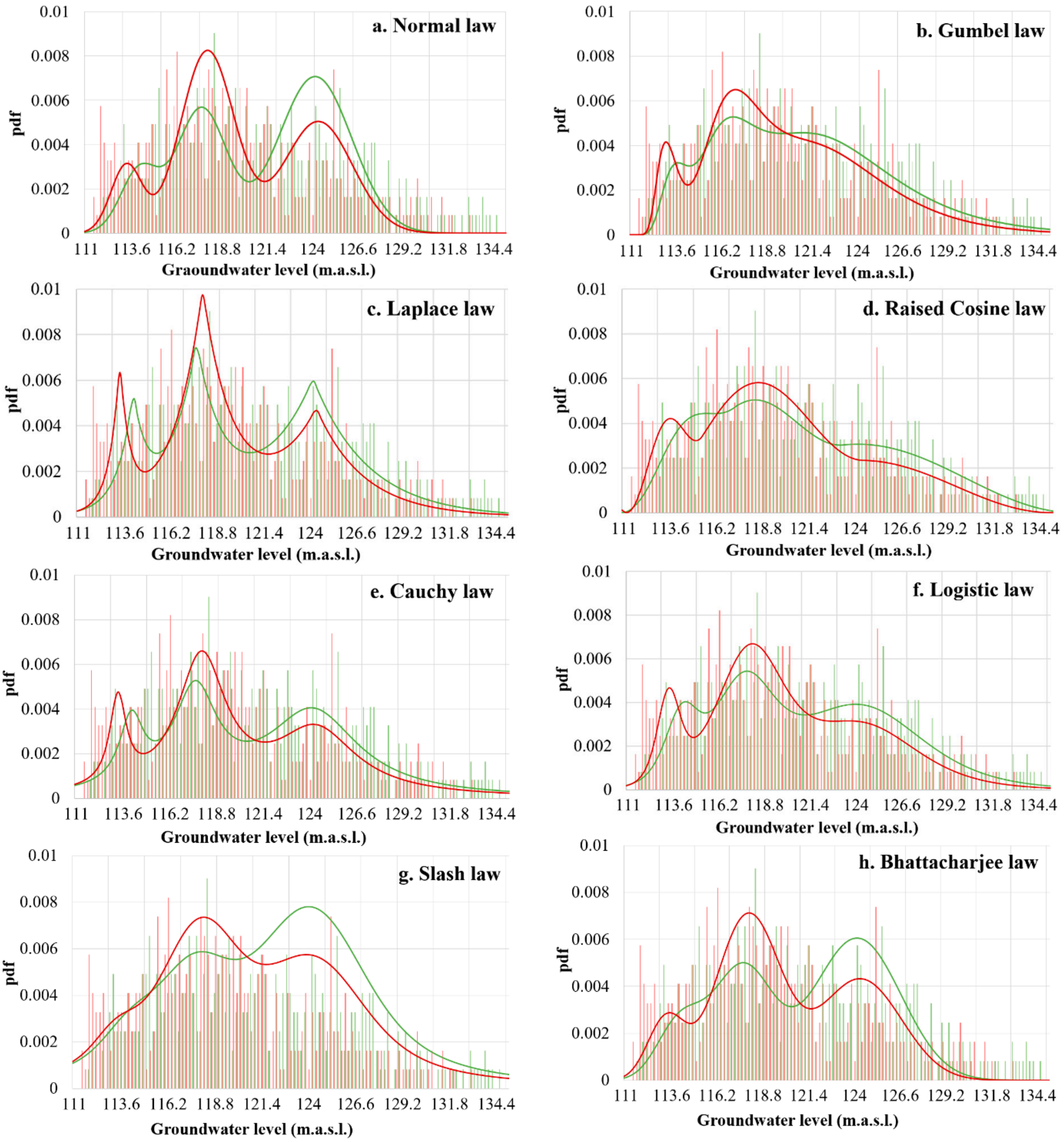


Figure 8. *Cont.*

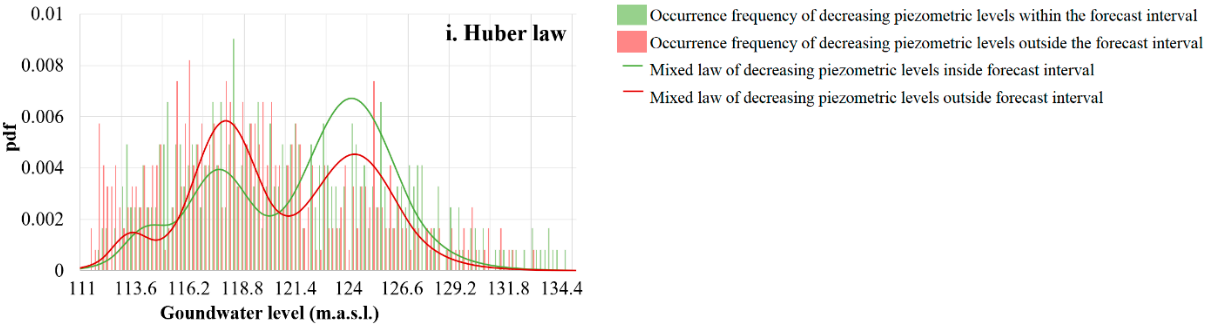


Figure 8. Composite *pdf* for decreasing measured groundwater distributions with (**a**) Normal law, (**b**) Gumbel law, (**c**) Laplace law, (**d**) Raised Cosine law, (**e**) Cauchy law, (**f**) Logistic law, (**g**) Slash law, (**h**) Bhattacharjee law and (**i**) Huber law.

For measured water levels having negative slopes (Figure 8), C_{PICP} = 0.51, meaning that the probability of the interval of prediction containing the observed value is similar to the probability of it not containing the observed value, whatever the groundwater level and theoretical composite *pdf* law selected. We can also notice that Normal, Bhattacharjee and Huber laws have the same kind of pattern whereas Logistic, Gumbel and Raised Cosine laws have similar shapes. This can be explained by the fact that Slash, Bhattacharjee and Huber are derived from the Normal law. Logistic, Gumbel and Raised Cosine laws seem to fit well with observed groundwater level distribution inside the prediction interval, having a Pearson's correlation coefficient over 0.74 for measured water levels inside the prediction interval, whereas other laws provide correlations ranging from 0.50 to 0.69 (Table 6 and Figure 8).

Raised Cosine, Gumbel, Logistic and Bhattacharjee laws have the best linear correlation with groundwater level distribution when they are outside the prediction interval, with a correlation from 0.74 to 0.77.

For measured water levels having positive slopes (Figure 9), C_{PICP} = 0.24, meaning that the observed groundwater levels outside the prediction interval are more numerous than groundwater levels inside the prediction interval. Pearson's correlation coefficients between the distribution of groundwater levels and the composite theoretical laws are shown in Table 7.

Table 7. Pearson's correlation coefficients between the distributions with positive slope of groundwater evolution and the theoretical composite *pdf*; r^2_{in} applies to measurements inside prediction interval, r^2_{out} applies to measurements outside the prediction interval.

Law	Normal	Gumbel	Laplace	Raised Cosine	Cauchy	Logistic	Slash	Bhatta-Charjee	Huber
r^2_{in}	0.43	0.41	0.47	0.44	0.45	0.44	0.43	0.43	0.42
r^2_{out}	0.52	0.54	0.54	0.66	0.55	0.61	0.63	0.57	0.52

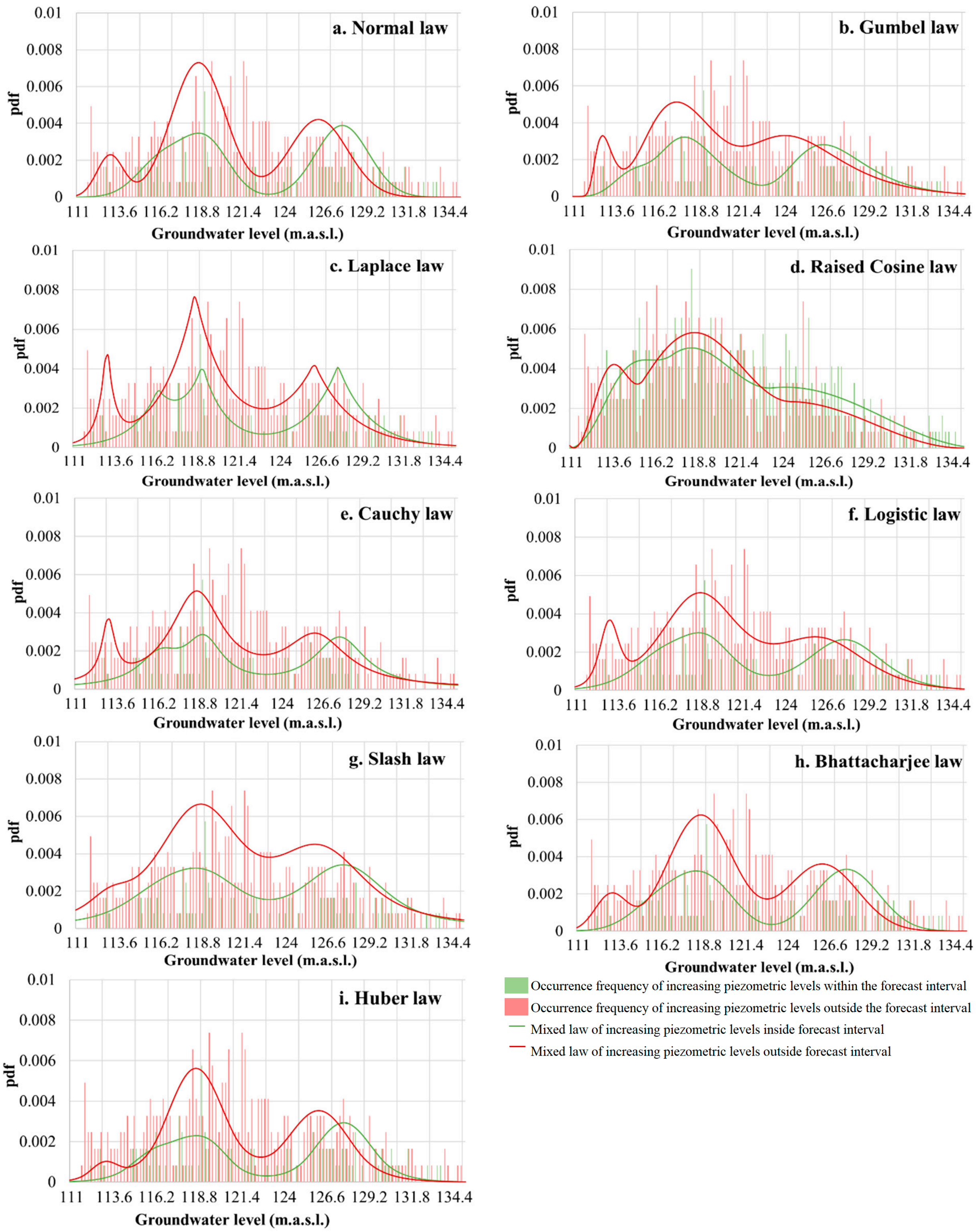


Figure 9. Mixed *pdf* for increasing groundwater distributions. with (**a**) Normal law, (**b**) Gumbel law, (**c**) Laplace law, (**d**) Raised Cosine law, (**e**) Cauchy law, (**f**) Logistic law, (**g**) Slash law, (**h**) Bhattacharjee law and (**i**) Huber law.

As seen previously, Normal, Slash, Bhattacharjee and Huber laws have a similar theoretical composite *pdf*. Curves representing observed values outside the prediction interval present a large bell around 119 m.a.s.l. for all of these four laws. This produces a lower correlation than for other laws (except for Slash law, which is smoother), with r^2 varying between 0.52 and 0.57. Slash, Raised Cosine and Logistic laws seem to provide the best correlated composite *pdf*, with a correlation between 0.61 and 0.66. On the other hand, the composite *pdf* representing observed values inside the prediction interval has two flared "peaks" at 117 m.a.s.l. and 127 m.a.s.l. for the nine laws. However, correlations are low due to the small frequencies of increasing groundwater levels inside prediction interval. Laplace and Cauchy laws appear to be the laws with the best fit, with a correlation above 0.45. Raised Cosine and Logistic correlations reach only 0.44.

This illustrates the model's difficulties obtaining relevant prediction intervals when the slope of groundwater levels is positive.

4.3.2. Error Margins

The last step consists in calculating the models of correctness. This consists in calculating the probability that the measured value lies within the prediction interval, and in adding errors around this probability for each class, following the procedure presented in 2.5.1. This is done for both increasing and decreasing measured groundwater level distributions. For each distribution, the probability that the measured value belongs to the interval of prediction is calculated by numerical integration, and the possible associated error is deduced using the number of samples inside the considered class using the *erf* (error function, the inverse of the Normal law), supposing that the data follows a normal distribution, and for a predefined confidence index. Then, for each class, a model of "correctness" is established and provides a confidence interval around the probability (shadow band around the probability in Figures 10 and 11). To provide the charts in Figures 10 and 11, a predefined confidence index of 0.95 is chosen, and several specific considerations are adopted:

- It is supposed that the distribution of samples inside a class follows a Normal Distribution,
- When a class contains no sample, for example, the class around 135 m.a.s.l., the error is maximum and is divided into two parts: 50% above 50% underneath the probability.
- When a class contains very few samples (less than three), this class is not considered for ${r_C}^2$ and M_E calculations.

Considering for example the distributions of increasing measured groundwater levels (Figure 10), all models of correctness show a poor correlation with the C_{PICP} (each cross is a C_{PICP} calculated thanks to the ensemble model), always having a Pearson's correlation under 0.3. This is consistent with the high dispersion of C_{PICP}. However, the percentage of these C_{PICP} included inside the calculated error margin seems to be a better indicator of the quality of the model of correctness. In this case, six laws have more than 75% of C_{PICP} inside the error margin. Pearson's correlation coefficients and the error margin indicator for Cauchy and Slash law's models of correctness are the highest, with, respectively, 0.25 and 75% and 0.22 and 80.3% values (Table 8).

Except for the Cauchy and Slash models of correctness, the probability increases for highest groundwater values (over than 130 m.a.s.l). For all models of correctness, the probability of a correct prediction varies from 0.2 to 0.4. The low probability of correctness still shows the difficulty of the model in forecasting increasing groundwater levels, with a low prediction interval.

Table 8. Error margin (E_M), and Pearson's correlation coefficients (r_C^2) between the model of correctness and the empirical C_{PICP} calculated for each 20-cm groundwater levels having a positive slope.

Law	Normal	Gumbel	Laplace	Raised Cosine	Cauchy	Logistic	Slash	Bhatta-charjee	Huber
r_C^2	0.15	0.04	0.24	0.19	0.25	0.24	0.22	0.16	0.20
E_M	76.3%	73.7%	72.4%	76.3%	75.0%	73.7%	80.3%	76.3%	77.6%

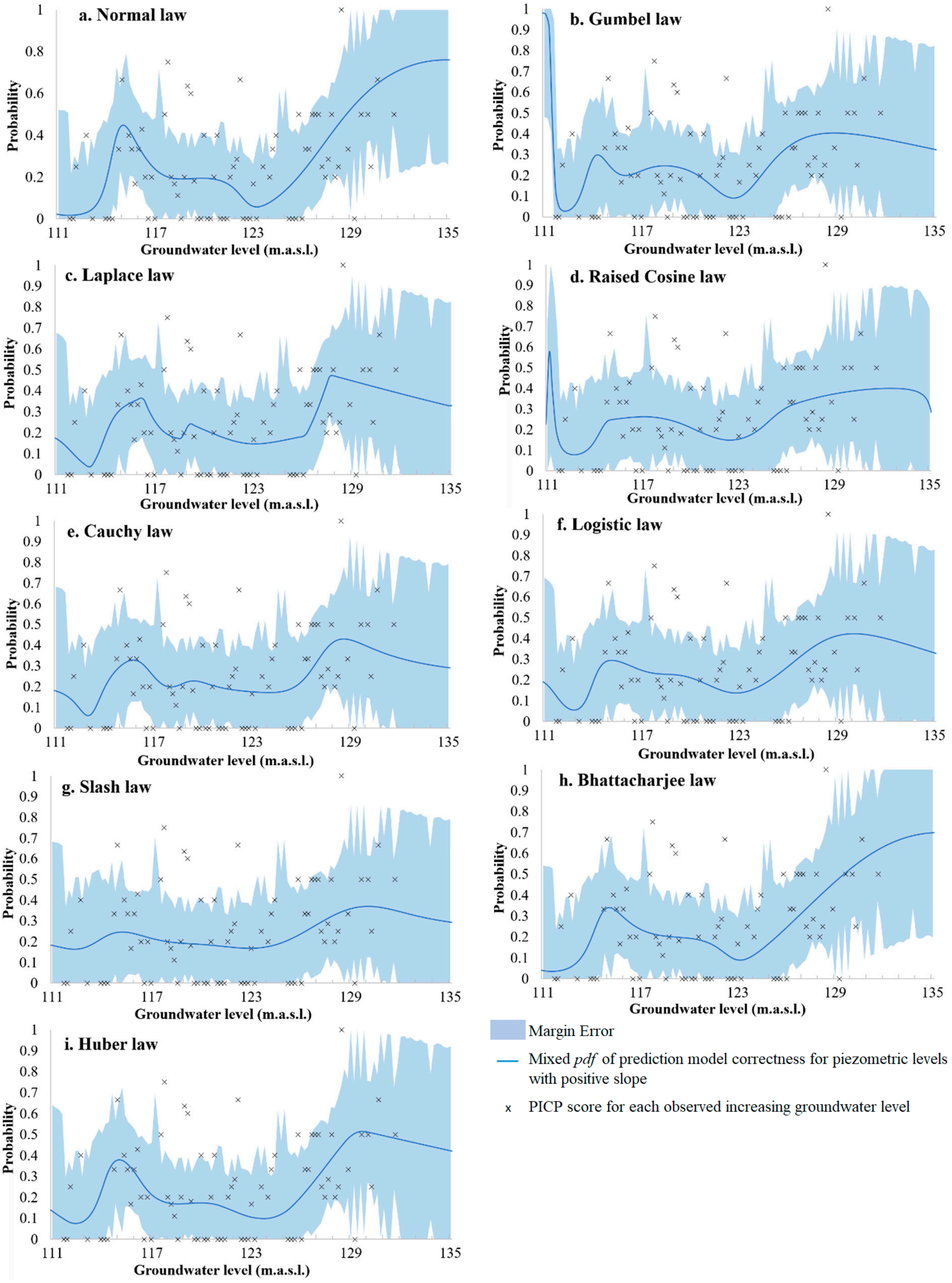


Figure 10. Models of correctness and Error margin (M_E), calculated from composite *pdf* of increasing levels and using a confidence index of 95% for (**a**) Normal law, (**b**) Gumbel law, (**c**) Laplace law, (**d**) Raised Cosine law, (**e**) Cauchy law, (**f**) Logistic law, (**g**) Slash law, (**h**) Bhattacharjee law and (**i**) Huber law.

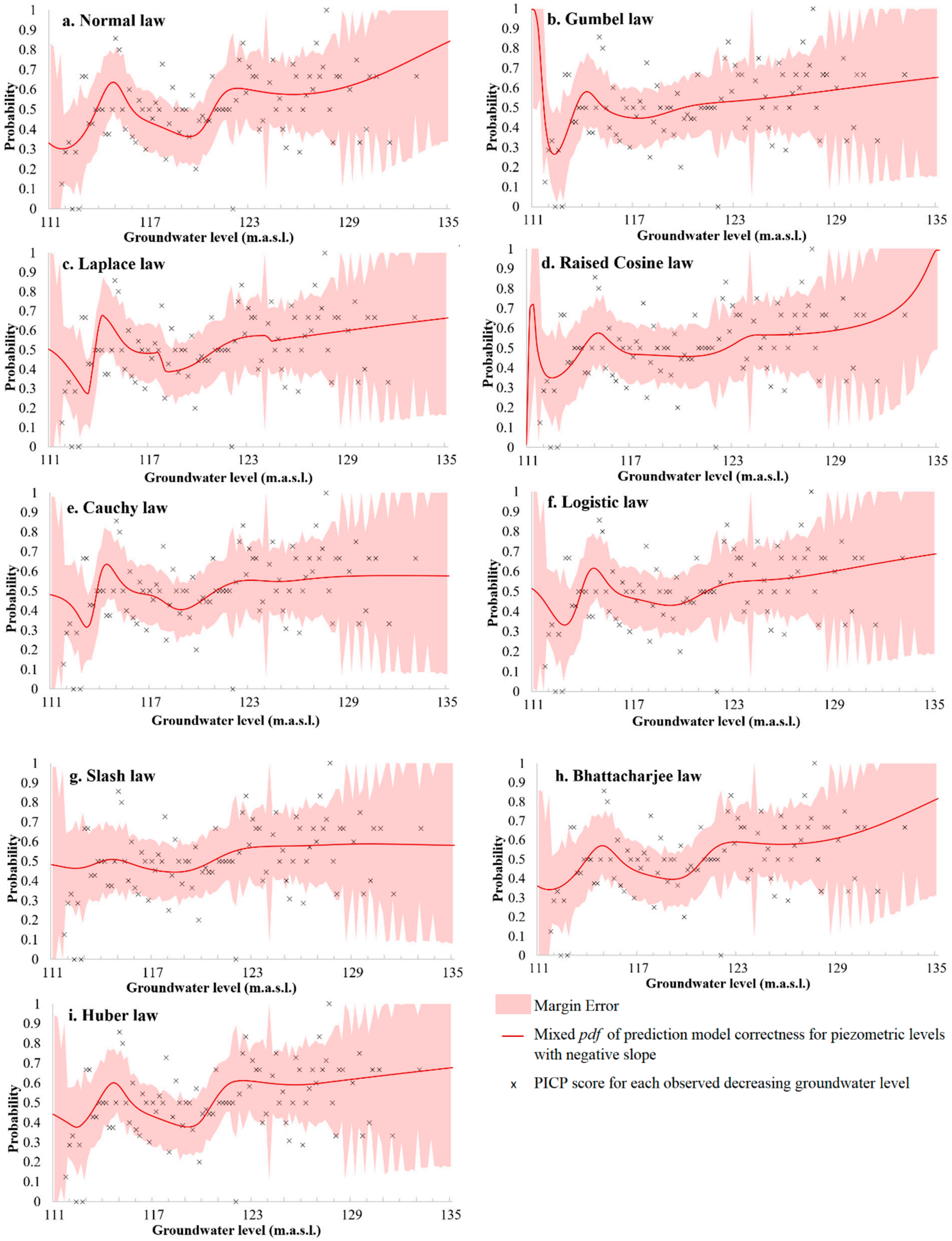


Figure 11. Models of correctness and Error margin (M_E), calculated from composite *pdf* of decreasing levels and using a confidence index of 95% for (**a**) Normal law, (**b**) Gumbel law, (**c**) Laplace law, (**d**) Raised Cosine law, (**e**) Cauchy law, (**f**) Logistic law, (**g**) Slash law, (**h**) Bhattacharjee law and (**i**) Huber law.

The same kind of models of correctness are made for decreasing groundwater levels, shown in Figure 11.

Correlations between the C_{PICP} and the model of correctness are still low, with an average value around 0.2 for all laws. The highest correlation comes from the model of correctness provided by the Slash mixed *pdf*. However, the crosses representing the C_{PICP} inside the error margins, reaching more than 79% for Logistic law (Table 9), are slightly higher than the ones obtained for decreasing groundwater levels (Table 8). Figure 11 shows that the models of correctness of the nine laws have a similar shape, with a stagnation of probability for observed piezometric levels above 125 m.a.s.l. The probability of correctness, for each law and each groundwater level, is above or equal to 0.5.

Table 9. Error margin (M_E), and Pearson's correlation coefficients (r_C^2) between the model of correctness and the empirical C_{PICP} calculated for each class of 20-cm groundwater levels having a negative slope.

Law	Normal	Gumbel	Laplace	Raised Cosine	Cauchy	Logistic	Slash	Bhatta-charjee	Huber
r_C^2	0.02	0.21	0.16	0.21	0.23	0.23	0.30	0.28	0.27
M_E	74.4%	70.3%	70.3%	72.5%	70.3%	79.1%	71.4%	73.6%	70.3%

However, these models of correctness only consider the uncertainty linked to the parameter's initialization before training. The uncertainty linked to the spatial representativeness of rain measurements is the next step, in order to consider and draw the two major origins of uncertainty.

4.4. Determination of Spatial Distribution of Rainfall Uncertainty

In order to assess the uncertainty linked to the spatial representativeness of rain measurements, we propose the operation of permutations and substitutions of data between rain gauge inputs. Doing this, following Figure 3, we obtain 27 different datasets. As the method previously applied requires a subset devoted to its assessment, independent of training, test and stop sets, we reused the cross-validation process in order to estimate the uncertainty in the 13 subsets used in cross-validation, which will then be used in cross-uncertainty assessment. This method also has the advantage of producing a sufficient number of values in a validation situation, since 27 permutations are performed on the rain gauges for each one of the 13 cross-uncertainty assessment sets. Added to the variability due to the initialization of the parameters, which is also considered, this method generates an ensemble model integrating the two types of variability: that due to rainfall and that due to the initialization of the parameters (27*40 members for each validation set).

Based on this ensemble, the prediction values are calculated: the median, higher and lower values, allowing the definition of a prediction interval for each time step and for each configuration. Thus, we find that the C_{PICP} equals 0.49 for the test set, shown in Figure 12. This was 0.39, considering only variability due to the parameter's initialization. Regarding the prediction interval, it logically became wider when including the rainfall variability, with C_{MPI} criterion of 0.73 m and a C_{PC} of 0.68 m^{-1}.

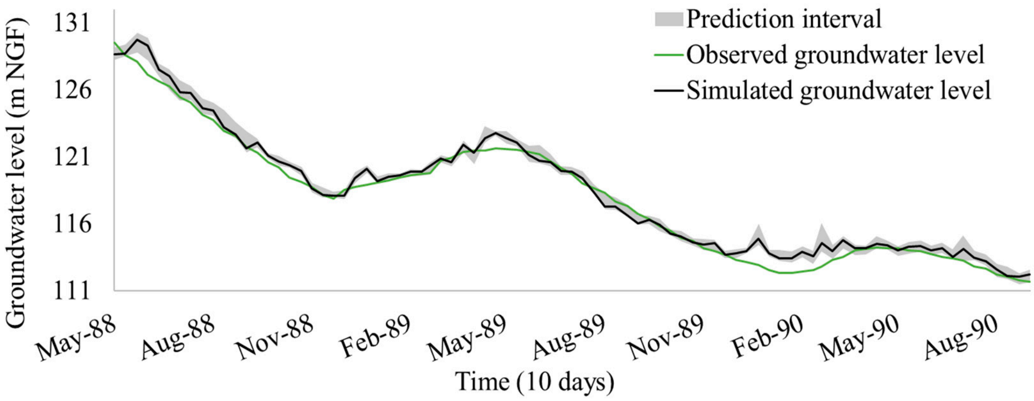


Figure 12. Groundwater level forecasting with 20 days' lead-time. Grey band shows the variability due to the parameter's initialization and to the rain spatial variability. C_p = 0.40; C_{PICP} = 0.49; C_{MPI} = 0.73 m.

4.5. Impact of the Spatial Distribution of Rainfall Uncertainty on the Model of Correctness

A new model of correctness is built as before, using the new ensemble model including the uncertainty due to the parameter's initialization, and the variability of rainfalls.

In order not to lengthen the presentation of the work, only the results of the laws with the set of best fits are presented, i.e., Gumbel, Raised Cosine, Logistic and Slash laws. Table 10 presents the correlations and indexes in a similar way to the presentation of Tables 6–9.

Table 10. Pearson's coefficients of correlation (r^2) between the measured water level distribution and the composite pdf; correlations between the model correctness and the empirical C_{PICP} calculated for each class of 20-cm groundwater levels (r_C^2); and Error Margins (EM).

Groundwater Level Class	Criteria	Laws			
		Gumbel	Raised Cosine	Logistic	Slash
Positive Slope	r_{in}^2	0.47	0.59	0.55	0.55
	r_{out}^2	0.52	0.56	0.53	0.54
	r_C^2	0.19	0.33	0.20	0.15
	$\%EM$	60.5%	61.8%	60.5%	63.2%
Negative Slope	r_{in}^2	0.78	0.79	0.77	0.72
	r_{out}^2	0.74	0.79	0.77	0.71
	r_C^2	0.52	0.41	0.49	0.47
	$\%EM$	70.3%	72.5%	71.4%	67.0%

As outlined above, the permutations of rainfall logically increase the forecast interval. Consequently, more observed values of water level belong to this interval. The Pearson correlation coefficients presented in Table 10 are therefore significantly higher than those of Tables 8 and 9. The Raised cosine law is that which clearly generates the highest correlations among the different composite pdfs. It will be thus used hereafter.

4.6. Definition of a Confidence Interval

The model of correctness is calculated using a confidence index. This defines a confidence interval for each measured water level regardless of its configuration (positive or negative slope, inside or outside the prediction interval). A predefined confidence index thus provides a confidence interval.

Using the Raised Cosine's models of correctness, which presented the best fit, we have varied the confidence index from 0.60 to 0.95 by steps of 0.5. The corresponding C_{PICP} are gathered in Table 11, and Figure 13 shows the confidence interval obtained for the confidence index of 0.90.

Table 11. Evolution of PICP and MPI scores following confidence index.

Confidence Index	C_{PICP} (Train + Test Datasets)	C_{PICP} (Test Set)	C_{MPI} (m)	C_{MPI} (m) (Without Extreme Values)
0.60	0.60	0.42	2.51	2.20
0.65	0.65	0.45	2.80	2.45
0.70	0.70	0.52	3.18	2.78
0.75	0.75	0.58	3.75	3.28
0.80	0.81	0.62	4.66	4.08
0.85	0.86	0.68	6.14	5.37
0.90	0.91	0.81	8.47	7.42
0.95	0.95	0.94	17.44	15.27

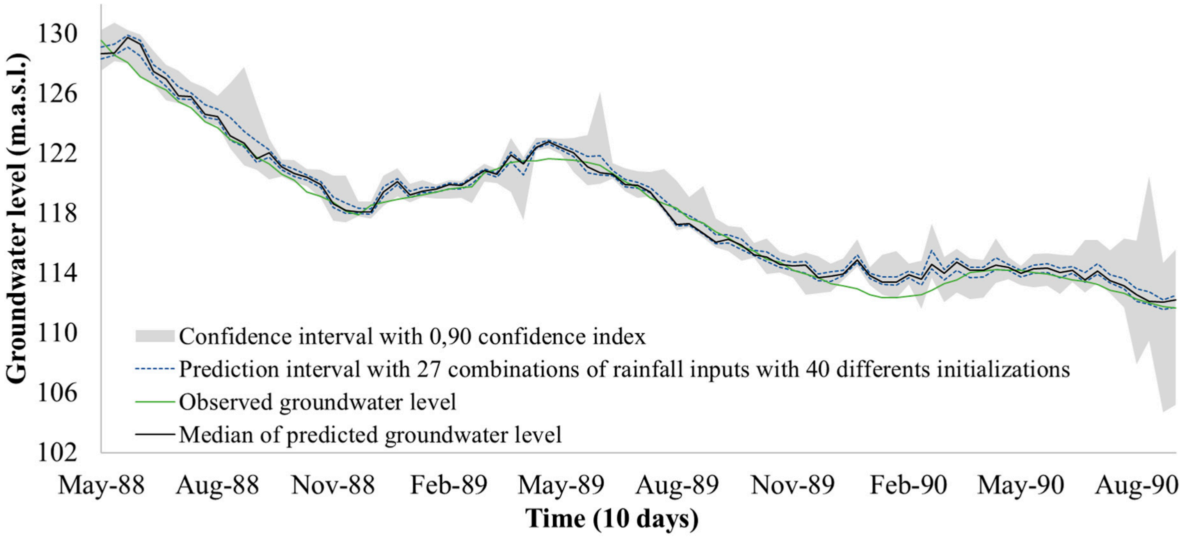


Figure 13. Groundwater level forecasting for 20 days' lead-time and confidence interval calculated with 0.90 confidence index. $C_p = 0.40$; $C_{PICP} = 0.81$; $C_{MPI} = 8.47$ m.

As expected, one can note in Table 11 that the confidence interval is wider when the confidence index is higher. The confidence interval associated with the highest confidence index (0.95) is very high (17.44 m) and probably not so useful to an end user. Nevertheless, the C_{MPI} decreases more quickly than the confidence index, which allows the manager to choose a compromise according to his requirements.

5. Discussion

5.1. Role of Rain in the Forecast Interval

Considering the two visualisations of the prediction intervals presented in Figures 6 and 12, it appears first of all that the latter are rather weak, in particular if expressed as a percentage of the maximum water table beat (25 m; cf. Table 1), $C_{MPI} = 0.62$ m (2.5%) for the first and, respectively, $C_{MPI} = 0.73$ m (3%) for the second. These small intervals seem to be very accurate, but they do not provide any real added value for the user, since the measured water level does not always fall within this interval. On average, the C_{PICP} provides the probability that the prediction interval from the model contains the measured value; this is 39% for the former and 49% for the latter. Thus, even considering the uncertainty caused by the measurement of rainfall variability (Figure 12), the model is only correct, on average, one time out of two.

Looking more specifically, in the test set in Figure 6, the model is better at predicting recessions. Both predicted values, water levels and confidence interval, are low during recessions. This is also shown by the correlations calculated on the composite pdf when the slopes are negative (Tables 6–9). For this reason, two different configurations have been separated to calculate the pdf: the case where the slope is positive, and the case where the slope is negative.

On further refinement, the measured and predicted water level curves deviate from each other mainly as a result of a strong rainfall pulse (p > 40 mm/day). This suggests that a specific phenomenon occurs in this configuration, which could be related to the influence of the Barbuise River. Indeed, if we refer to Table 2, we note that the response time linking the Discharge of the Barbuise and Pouan les Vallées (D_{BP}) is 2 decades; however, in Figure 6, we can see that the most important rainfall episode of February 1990 (p > 60 mm) influences the water table in less than one decade, its maximum effect appearing at 2 decades. It thus appears that infiltration with faster dynamics occurs during heavy rainfall episodes and that the model has difficulties in representing these fast and rare infiltrations. Moreover, the prediction interval is rather smaller for the responses to these episodes than for the other configurations, both for Figures 6 and 12, suggesting that, during very wet episodes, the spatial variability of the rainfall events, at the decadal step, does not have a great impact on the response. Given the objective of the modelling, which is to predict low water, this double property, errors in prediction and low uncertainty during high rain pulses, can be considered as not being prohibitive.

Even so, in order to improve the representation of an uncertainty that would be more useful to managers, for example that would allow manager to choose the confidence interval that suits him, we have introduced the correctness models, delivering an "error margin" reported for the forecasts in the form of a confidence interval. This error margin is itself controllable by a user-defined confidence index. This addition has the same shortcomings with respect to major rainfall events but allows the manager to adapt the visualisation of the uncertainty to his needs. It could be used for managing pollutant intrusion during floods and for low levels anticipation, by choosing different confidence indices for these different uses.

5.2. Role of the Amount of Data

An important limitation to note is the amount of extreme data. While the range defining the water level classes of the training set has been chosen to contain at least 10 measurements per class, it can be noted that the data for the extreme class is only observed once. There is a lack of extreme data to be able to build reliable *pdf*. In particular, the chosen test set contains the lowest data over the entire history of the database (1997–2018), and the number of samples in the lowest groundwater levels classes is therefore very low; 0 for the lowest, and a few units for the others. Remember that the calculation of the correctness model assigns the maximum uncertainty when there has never been any measured value. This is the case for the dry period of the test set and this explains why the visualized confidence intervals for the summer 1990 are so large while the measured and predicted values are very close.

An improvement could be obtained by choosing different laws of probability in order to minimize the uncertainties in this specific condition. For example, in Figure 11, it can be seen that Gumbel's law gives the smallest error margin. This result is consistent with the application domain of Gumbel's law, which is aimed at distributions with extreme events. The confidence interval thus could be improved by using this specific law for very low and very high water level values.

6. Conclusions

The goal of this paper was to define a generic method able to estimate the uncertainty generated by both the neural network model itself and by the non-measured spatial heterogeneity of rainfalls. This work was carried out on the Champagne chalk aquifer (Northern France), at a 10-day time-step up to 20 days lead time and could be used for the purpose of tap water production and agriculture groundwater management.

A reliable enough ANN model to forecast groundwater level at "Craie à Vailly" piezometer was built, and tested on the driest summer of the entire 40 years database. The uncertainties due to the model parameters were then estimated following a simple method that takes into account the variability caused by the initialization of parameters. Added to this variability, the paper investigated the impact of the non-measured spatial variation on rainfall.

For this purpose, a set of permutations and substitutions of measurements at rainfall stations, combined in an original way with the implementation of the cross-validation process, was proposed and allowed the calculation of forecasts and uncertainties on a test set, never used in training, stopping, nor in cross-uncertainty assessment subsets. This test set contains the data for the most severe drought in the database. From the uncertainties found on this test set, a correctness model was proposed which provides, for the requirement of a global confidence index, a confidence interval to be applied to each forecast value.

Several limitations were identified. The main one is related to the amount of data: when there is no historical data in the range of values considered by the prediction, the uncertainty is maximal. It would be possible to improve the correctness model by choosing a more appropriate statistical law.

This methodology is original and can be deployed on other hydro-systems having other types of surface or subsurface features and different climate contexts. Applications may either make forecasts or propose a confidence interval associated with these forecasts, with the degree of confidence chosen by the user. These methodologies have proved significant advantages: the rigorous design of the neural network model has allowed the realization of a model capable of generalizing to a range of data that exceeded the range of the training set. Furthermore, it is possible to flexibly choose the confidence index according to the hydrological configuration (e.g., recession or rising water table).

Thanks to this methodology, a mid-term groundwater forecast with its own uncertainty can be provided. Our model, specialized for droughts showing as driest events on test dataset, allows for prevention of dry events, which can be anticipated nearly three weeks before, allowing agricultural and water supply end users to anticipate this risk, by issuing, for example, water withdrawal restriction orders or by carrying out water transfers from dams and reservoirs.

Author Contributions: Conceptualization, N.A., G.A. and M.S.; methodology, N.A., G.A., M.S. and A.J.; software, N.A. and A.J.; validation, G.A. and M.S.; formal analysis, N.A.; investigation, N.A.; resources, all authors; data curation, N.A.; writing—original draft preparation, N.A.; writing—review and editing, G.A., M.S., A.J. and M.V.; visualization, N.A.; supervision, A.J. and M.V.; project administration, A.J. and M.V.. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the ARNT (Association Nationale de la Recherche et de la Technologie), grant number 1700720.

Institutional Review Board Statement: Not applicable.

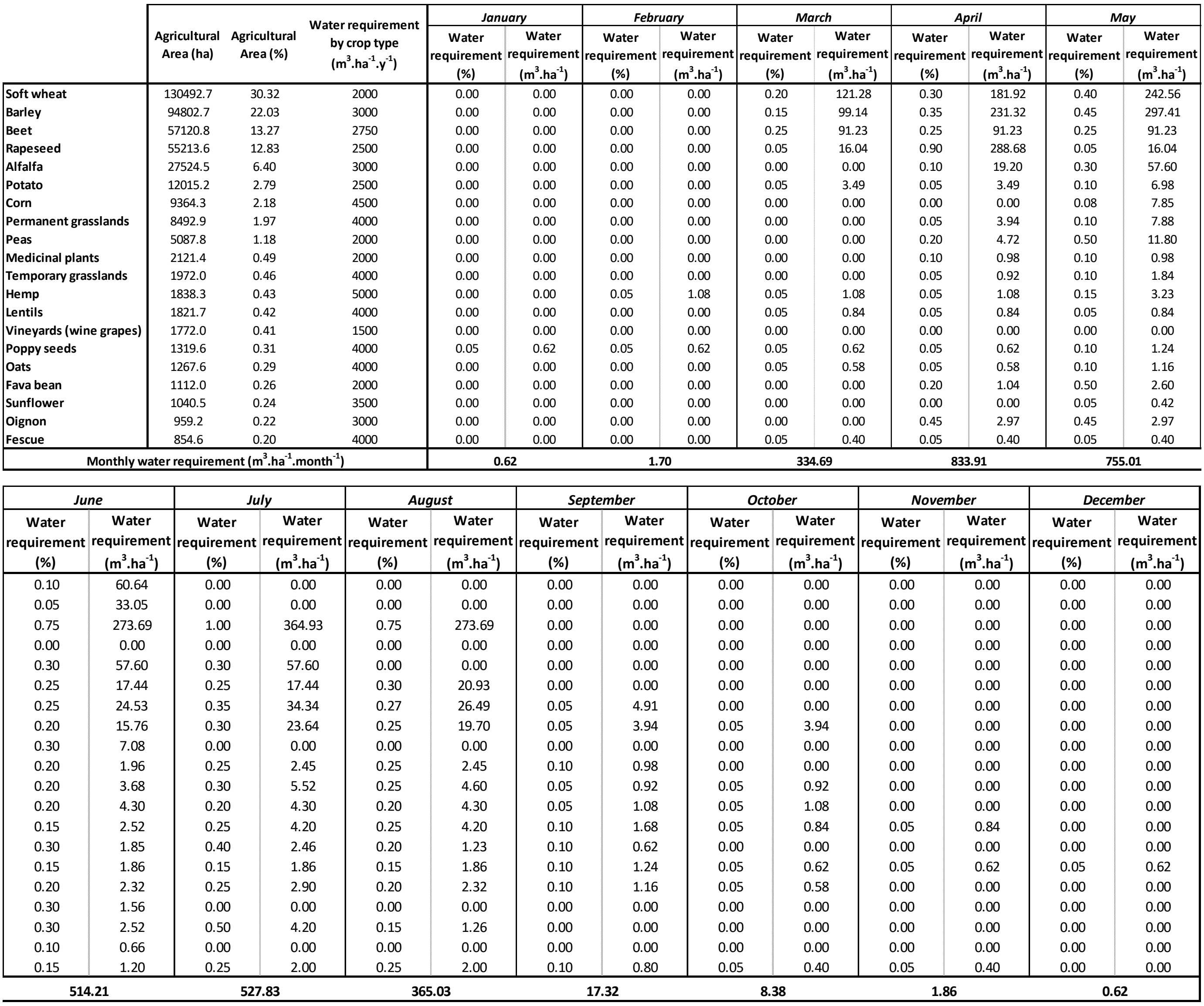
Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable. No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: Authors would like to warmly thank the BRGM, which provided the dataset, J. Nicolas (BRGM) and S. Pistre (Univ. Montpellier) for helpful discussions, IT services of IMT Mines Ales and AQUASYS, for technical support and infrastructure deployed during coronavirus crisis, OUGC84, for sharing agricultural data, GEONOSIS, for modelling software RNF Pro and the ANRT, which contributed to funding this PhD thesis.

Conflicts of Interest: The authors declare no conflict of interest, except for a commercial use. Readers must have to keep in mind that presented results are not generalizable for others basins or others outputs and need additional studies. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A



	Agricultural Area (ha)	Agricultural Area (%)	Water requirement by crop type (m^3.ha^{-1}.y^{-1})	*January*		*February*		*March*		*April*		*May*	
				Water requirement (%)	Water requirement (m^3.ha^{-1})	Water requirement (%)	Water requirement (m^3.ha^{-1})	Water requirement (%)	Water requirement (m^3.ha^{-1})	Water requirement (%)	Water requirement (m^3.ha^{-1})	Water requirement (%)	Water requirement (m^3.ha^{-1})
Soft wheat	130492.7	30.32	2000	0.00	0.00	0.00	0.00	0.20	121.28	0.30	181.92	0.40	242.56
Barley	94802.7	22.03	3000	0.00	0.00	0.00	0.00	0.15	99.14	0.35	231.32	0.45	297.41
Beet	57120.8	13.27	2750	0.00	0.00	0.00	0.00	0.25	91.23	0.25	91.23	0.25	91.23
Rapeseed	55213.6	12.83	2500	0.00	0.00	0.00	0.00	0.05	16.04	0.90	288.68	0.05	16.04
Alfalfa	27524.5	6.40	3000	0.00	0.00	0.00	0.00	0.00	0.00	0.10	19.20	0.30	57.60
Potato	12015.2	2.79	2500	0.00	0.00	0.00	0.00	0.05	3.49	0.05	3.49	0.10	6.98
Corn	9364.3	2.18	4500	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	7.85
Permanent grasslands	8492.9	1.97	4000	0.00	0.00	0.00	0.00	0.00	0.00	0.05	3.94	0.10	7.88
Peas	5087.8	1.18	2000	0.00	0.00	0.00	0.00	0.00	0.00	0.20	4.72	0.50	11.80
Medicinal plants	2121.4	0.49	2000	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.98	0.10	0.98
Temporary grasslands	1972.0	0.46	4000	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.92	0.10	1.84
Hemp	1838.3	0.43	5000	0.00	0.00	0.05	1.08	0.05	1.08	0.05	1.08	0.15	3.23
Lentils	1821.7	0.42	4000	0.00	0.00	0.00	0.00	0.05	0.84	0.05	0.84	0.05	0.84
Vineyards (wine grapes)	1772.0	0.41	1500	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Poppy seeds	1319.6	0.31	4000	0.05	0.62	0.05	0.62	0.05	0.62	0.05	0.62	0.10	1.24
Oats	1267.6	0.29	4000	0.00	0.00	0.00	0.00	0.05	0.58	0.05	0.58	0.10	1.16
Fava bean	1112.0	0.26	2000	0.00	0.00	0.00	0.00	0.00	0.00	0.20	1.04	0.50	2.60
Sunflower	1040.5	0.24	3500	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.42
Oignon	959.2	0.22	3000	0.00	0.00	0.00	0.00	0.00	0.00	0.45	2.97	0.45	2.97
Fescue	854.6	0.20	4000	0.00	0.00	0.00	0.00	0.05	0.40	0.05	0.40	0.05	0.40
Monthly water requirement (m^3.ha^{-1}.month^{-1})				**0.62**		**1.70**		**334.69**		**833.91**		**755.01**	

June		*July*		*August*		*September*		*October*		*November*		*December*	
Water requirement (%)	Water requirement (m^3.ha^{-1})	Water requirement (%)	Water requirement (m^3.ha^{-1})	Water requirement (%)	Water requirement (m^3.ha^{-1})	Water requirement (%)	Water requirement (m^3.ha^{-1})	Water requirement (%)	Water requirement (m^3.ha^{-1})	Water requirement (%)	Water requirement (m^3.ha^{-1})	Water requirement (%)	Water requirement (m^3.ha^{-1})
0.10	60.64	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.05	33.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.75	273.69	1.00	364.93	0.75	273.69	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.30	57.60	0.30	57.60	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.25	17.44	0.25	17.44	0.30	20.93	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.25	24.53	0.35	34.34	0.27	26.49	0.05	4.91	0.00	0.00	0.00	0.00	0.00	0.00
0.20	15.76	0.30	23.64	0.25	19.70	0.05	3.94	0.05	3.94	0.00	0.00	0.00	0.00
0.30	7.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.20	1.96	0.25	2.45	0.25	2.45	0.10	0.98	0.00	0.00	0.00	0.00	0.00	0.00
0.20	3.68	0.30	5.52	0.25	4.60	0.05	0.92	0.05	0.92	0.00	0.00	0.00	0.00
0.20	4.30	0.20	4.30	0.20	4.30	0.05	1.08	0.05	1.08	0.00	0.00	0.00	0.00
0.15	2.52	0.25	4.20	0.25	4.20	0.10	1.68	0.05	0.84	0.05	0.84	0.00	0.00
0.30	1.85	0.40	2.46	0.20	1.23	0.10	0.62	0.00	0.00	0.00	0.00	0.00	0.00
0.15	1.86	0.15	1.86	0.15	1.86	0.10	1.24	0.05	0.62	0.05	0.62	0.05	0.62
0.20	2.32	0.25	2.90	0.20	2.32	0.10	1.16	0.05	0.58	0.00	0.00	0.00	0.00
0.30	1.56	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.30	2.52	0.50	4.20	0.15	1.26	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.10	0.66	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.15	1.20	0.25	2.00	0.25	2.00	0.10	0.80	0.05	0.40	0.05	0.40	0.00	0.00
514.21		**527.83**		**365.03**		**17.32**		**8.38**		**1.86**		**0.62**	

Figure A1. Details of the construction and of the values of the irrigation inputs for the different types of crops as a function of the surface, the water requirement and the season (by months).

References

1. Hartmann, D.L.; Klein Tank, A.M.G.; Rusticucci, M.; Alexander, L.V.; Brönnimann, S.; Charabi, Y.A.-R.; Dentener, F.J.; Dlugokencky, E.J.; Easterling, D.R.; Kaplan, A.; et al. Observations: Atmosphere and Surface. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Cambridge University Press: Cambridge, UK, 2013; Chapter 2; pp. 159–254.
2. Hornik, K.; Stinchombe, M.; White, H. Multilayer Feedforward Networks are Universal Approximators. *Neural Netw.* **1989**, *2*, 359–366. [CrossRef]
3. Wagener, T.; Montanari, A. Convergence of approaches toward reducing uncertainty in predictions in ungauged basins. *Water Resour. Res.* **2011**, *47*. [CrossRef]
4. Bourgin, F. *Comment Quantifier l'incertitude Prédictive en Modélisation Hydrologique? Travail Exploratoire sur un Grand Échantillon de Bassins Versants*; AgroParisTech: Paris, France, 2014.

5. Solomatine, D.P.; Shrestha, D.L. A novel method to estimate model uncertainty using machine learning techniques. *Water Resour. Res.* **2009**, *45*, 1–16. [CrossRef]
6. Biondi, D.; Todini, E. Comparing Hydrological Postprocessors Including Ensemble Predictions into Full Predictive Probability Distribution of Streamflow. *Water Resour. Res.* **2018**, *54*, 9860–9882. [CrossRef]
7. Krzysztofowicz, R. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resour. Res.* **1999**, *35*, 2739–2750. [CrossRef]
8. Biondi, D.; Luca, D.L.D. Performance assessment of a Bayesian Forecasting System (BFS) for real-time flood forecasting. *J. Hydrol.* **2013**, *479*, 51–63. [CrossRef]
9. Wu, W.; Dandy, G.C.; Maier, H.R. Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modelling. *Environ. Model. Softw.* **2014**, *54*, 108–127. [CrossRef]
10. Kong-A-Siou, L.; Johannet, A.; Borrell, V.; Pistre, S. Complexity selection of a neural network model for karst flood forecasting: The case of the Lez Basin (southern France). *J. Hydrol.* **2011**, *403*, 367–380. [CrossRef]
11. Stone, M. Cross-Validatory Choice and Assessment of Statistical Predictions (With Discussion). *J. R. Stat. Soc. Ser. B* **1974**, *38*, 111–147. [CrossRef]
12. Sjöberg, J.; Zhang, Q.; Ljung, L.; Benveniste, A.; Delyon, B.; Glorennec, P.-Y.; Hjalmarsson, H.; Juditskys, A. Nonlinear Black-box Modeling in System Identification: A Unified Overview. *Automatica* **1995**, *31*, 1691–1724. [CrossRef]
13. Akil, N.; Artigue, G.; Savary, M.; Johannet, A.; Vinches, M. Quantification of Neural Network Uncertainties on the Hydrogeological Predictions by Probability Density Functions. In Proceedings of the IOP Conference Series: Earth and Environmental Science, Prague, Czech Republic, 9–13 September 2019; p. 10.
14. Dreyfus, G. *Neural Networks, Methodology and Applications*; Springer: Berlin, Germany, 2005; p. 509.
15. Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural Netw.* **1991**, *4*, 251–257. [CrossRef]
16. Barron, A.R. Approximation bounds for superpositions of a sigmoidal function. In Proceedings of the IEEE International Symposium on Information Theory—Proceedings, San Antonio, TX, USA, 17–22 January 1993; pp. 930–945.
17. Nerrand, O.; Roussel-Ragot, P.; Personnaz, L.; Dreyfus, G.; Marcos, S. Neural Networks and Nonlinear Adaptive Filtering: Unifying Concepts and New Algorithms. *Neural Comput.* **1993**, *5*, 165–199. [CrossRef]
18. Artigue, G.; Johannet, A.; Borrell, V.; Pistre, S. Flash flood forecasting in poorly gauged basins using neural networks: Case study of the Gardon de Mialet basin (southern France). *Nat. Hazards Earth Syst. Sci.* **2012**, *12*, 3307–3324. [CrossRef]
19. Taver, V.; Johannet, A.; Borrell-Estupina, V.; Pistre, S. Feed-forward vs recurrent neural network models for non-stationarity modelling using data assimilation and adaptivity. *Hydrol. Sci. J.* **2015**, *60*, 1242–1265. [CrossRef]
20. Geman, S.; Bienenstock, E.; Doursat, R. Neural Networks and the Bias/Variance dilemma. *Neural Comput.* **1992**, *4*, 1–58. [CrossRef]
21. Darras, T.; Johannet, A.; Vayssade, B.; Long-a-Siou, L.; Pistre, S. Influence of the Initialization of Multilayer Perceptron for Flash Floods Forecasting: How Designing a Robust Model. In *International Work-Conference on Time Series 2014*; Springer: Granada, Spain, 2014; p. 13.
22. Kong-A-Siou, L.; Johannet, A.; Borrell Estupina, V.; Pistre, S. Optimization of the generalization capability for rainfall-runoff modeling by neural networks: The case of the Lez aquifer (southern France). *Environ. Earth Sci.* **2012**, *65*, 2365–2375. [CrossRef]
23. AESN. *Fiche de Caractérisation de la ME HG20—Masse d'eau souterraine HG208 "Craie de Champagne Sud et Centre*; AESN: Nanterre, France, 2015.
24. Météo France. Données Publiques—Observations In Situ. Available online: https://donneespubliques.meteofrance.fr/?fond=rubrique&id_rubrique=26 (accessed on 8 September 2020).
25. BNPE. Base de données des prélèvements en eau, OPR0000035289. Available online: https://bnpe.eaufrance.fr/acces-donnees/codeOuvrage/OPR0000035289/annee/2014 (accessed on 15 January 2021).
26. BNPE. Base de Données des Prélèvements en eau, Vailly (10). Available online: https://bnpe.eaufrance.fr/?q=acces-donnees/codeCommune/10391/annee/2017/etCommunesAdjacentes (accessed on 31 May 2021).
27. Stollsteiner, P. *Connaissance des Ressources Réellement Disponibles sur L'ensemble des Bassins Versants Crayeux*; BRGM: Reims, France, 2013.
28. Kloppmann, W.; Dever, L.; Edmunds, W.M. Residence time of Chalk groundwaters in the Paris Basin and the North German Basin: A geochemical approach. *Appl. Geochem.* **1998**, *13*, 593–606. [CrossRef]
29. Pinault, J.-L.; Allier, D.; Chabart, M. *Prévision des Volumes d'eau Exploitables de 10 Bassins versants en Champagne Crayeuse*; BRGM: Reims, France, 2006.
30. Putot, E.; Verjus, P.; Vernoux, J.F. *Qualification des Piézomètres du Réseau de Bassin Seine Normandie en 2005*; BRGM: Massy, France, 2006.
31. IGN. Registre Parcellaire Graphique (RPG): Contours Des Parcelles et Îlots Culturaux et Leur Groupe de Cultures Majoritaire. Available online: https://www.data.gouv.fr/fr/datasets/registre-parcellaire-graphique-rpg-contours-des-parcelles-et-ilots-culturaux-et-leur-groupe-de-cultures-majoritaire/#_ (accessed on 14 September 2020).
32. Mangin, A. Pour une meilleure connaissance des systèmes hydrologiques à partir des analyses corrélatoire et spectrale. *J. Hydrol.* **1984**, *67*, 25–43. [CrossRef]
33. Kitanidis, P.K.; Bras, R.L. Real-time forecasting with a conceptual hydrologic model: 2. Applications and results. *Water Resour. Res.* **1980**, *16*, 1034–1044. [CrossRef]

34. Shrestha, D.L.; Solomatine, D.P. Machine learning approaches for estimation of prediction interval for the model output. *Neural Netw.* **2006**, *19*, 225–235. [CrossRef]
35. Khosravi, A.; Nahavandi, S.; Creighton, D. A prediction interval-based approach to determine optimal structures of neural network metamodels. *Expert Syst. Appl.* **2010**, *37*, 2377–2387. [CrossRef]
36. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum Likelihood from Incomplete Data via the EM Algorithm. *J. R. Stat. Soc. Ser. B* **1977**, *39*, 1–22.
37. Benaglia, T.; Chauveau, D.; Hunter, D.R.; Young, D.S. Mixtools: An R Package for Analyzing Finite Mixture Models. *J. Stat. Softw.* **2009**, *32*, 29. [CrossRef]
38. De Moivre, A. *The Doctrine of Chances, or, A Method of Calculating the Probability of Events in Play*, III ed.; Chelsea Publishing Company: Londres, UK, 1756; p. 368.
39. Fisher, R.A. On the mathematical foundations of theoretical statistics. *Philos. Trans. R. Soc.* **1922**, *9*, 309–368.
40. Gumbel, E.J. Méthodes graphiques pour l'analyse de débits de crues. *Rev. De Stat. Appl.* **1957**, *5*, 77–89. [CrossRef]
41. Laplace, P.-S. Mémoire sur la probabilité des causes par les évènements. In *Œuvres complètes de Laplace*; Mémoires de l'Academie Royale des Sciences, Ed.; Divers Savan: Paris, France, 1774; Chapter 6; pp. 621–656.
42. Rinne, H. *Location-Scale Distributions Linear Estimation and Probability Plotting Using MATLAB*; Justus Liebig University: Giessen, Allemagne, 2010.
43. Sinha, R.K. A Thought on Exotic Statistical Distributions. *Int. J. Math. Comput. Sci.* **2012**, *6*, 49–52.
44. Cauchy, A.-L. CALCUL DES PROBABILITÉS—Sur les résultats moyens d'observations de même nature, et sur les résultats les plus probables. In *Oeuvres Complètes*; Cambridge University Press: Cambridge, UK, 1853; pp. 94–104.
45. Stigler, S.M. *An Historical Note on the Cauchy Distribution*; Biometrika Trust: Oxford, UK, 1974; pp. 375–380.
46. Verhulst, P.F. Recherches mathématiques sur la loi d'accroissement de la population. In *Nouveaux mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles*; Nabu Press: Bruxelles, Belgique, 1845; Chapter 18; pp. 1–38.
47. Rogers, W.H.; Tukey, J.W. Understanding some long-tailed symmetrical distributions. *Stat. Neerl.* **1972**, *26*, 211–226. [CrossRef]
48. Bhattacharjee, G.P.; Pandit, S.N.N.; Mohan, R. Dimensional Chains Involving Rectangular and Normal Error-Distributions. *Technometrics* **1963**, *5*, 404–406. [CrossRef]
49. Huber, P.J. Robust estimation of a location parameter. *Ann. Math. Stat.* **1964**, *35*, 73–101. [CrossRef]
50. Huber, P.J. *Robust Statistics*; Wiley: Hoboken, NJ, USA, 1981; p. 308.
51. Levenberg, K. A method for the solution of certain non-linear problems in least squares. *Q. Appl. Math.* **1944**, *2*, 164–168. [CrossRef]
52. Marquardt, D.W. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431–441. [CrossRef]