Full Length Research Paper

Modeling ensemble streamflow: Application to the Senegal River upper the Manantali Dam

Ndione D. M. ¹, Sambou S. ¹, Sane M. L. ¹, Kane S. ¹, Leye I. ¹, Tamba S. ², Diedhiou R. ¹, Cisse M. T. ² and Ngom I. ¹

¹Hydraulic and Fluids Mechanics Laboratory (HFML), Department of Physics, Faculty of Sciences and Technology, Cheikh Anta Diop University (UCAD), BP 5005, Dakar-Fann, Senegal.
²Polytechnic High School of Thies, PB 10, Thies, Senegal.

Received 3 September, 2018; Accepted 15 October, 2018

This study presents an ensemble stream flow simulation with probability of occurrence accounting for errors of the GR4J turned to the Senegal River upper the Manantali dam. Through this approach, probability is associated to simulation at each time, with step and reliability of issues depending on probability scores. Results including the reliability of the results permit decision-makers to judge the reliability of their simulations. Past errors pattern of the model is used to perturb the model for issuing ensemble scenarios rather than classical single deterministic. Basic and reverse Box-Cox transformation is carried out to allow treatment of the errors through weighted multivariate Gaussian distribution. Statistic errors are then used to perturb the hydrological model to obtain ensemble issues (RAW-Ens). Furthermore, a post-processing method (affine kernel dressing) is performed to produce a dressed ensemble (AKD-Ens). Ensembles are evaluated at deterministic and probabilistic scale. Diagrams (attribute and ROC) are also used for this purpose. Evaluating methods reveal through scores that the system is reliable and that dressing method (AKD) improves quality of the raw ensemble drawn from the perturbed model.

Key words: Senegal River, Bafing, Manantali dam, Ensemble simulation, Kernel dressing.

INTRODUCTION

Modeling and forecasting ensemble of hydrological variables is used to help assessment of feasibility of some applications in term of water supply (drinking water, irrigation and hydropower), particularly in dry areas. However, randomness of natural fluids dynamics has motivated the use of probabilistic approaches in hydrology to improve quality of modeling and forecast issues, and consequently, decision making (Bauer Gottwein et al., 2014; Demeritt et al., 2013). Probabilistic modeling based on issuing ensemble scenarios rather than single deterministic are introduced for the assessment of the flow rates particularly in forecasting purpose, estimating soil moisture, groundwater or tributaries from snow layers (Pappenberger et al., 2015). It enables surpassing classical limitations of deterministic processes and has significantly enhanced flood
prediction through the world (Demargne et al., 2013; Cloke et al., 2009; Demeritt et al., 2013; He et al., 2016). Moreover, it has led to optimizing gains on activities depending on water resources and reducing common incidents inherent to floods or other hydro climatic extreme events by forestalling potential risks (Fan et al., 2015; Laiolo et al., 2014).

Indeed, since last decades, developing good hydrological forecasting system considering biases has become a challenge for hydrologists and water managers due to frequent failure of classical forecasting schemes (Ranzi et al., 2009; Pagano et al., 2013). Major public and private hydro meteorological services are moving towards implementing ensemble models and using probabilistic interpretations of modeling issues. This is the case of the National Weather Service (NWS) and the National Oceanic and Atmospheric Administration of the United States of America (NOAA) (Demargne et al., 2013; Cloke and Pappenberger, 2009). Towards the same goal, the HEPEX (Hydrologic Ensemble Prediction Experiment) promotes developing and use of ensemble prediction systems through projects mainly focused on hydrological ensemble forecast (HEF) generally run by meteorological ensemble. Jeong et al. (2005) used both single neural network (SNN) and ensemble neural network (ENN) to improve streamflow forecasting system based on the TANK model at the Daechong multipurpose dam in Korea. Results show that the ENN improves considerably ensemble forecasts. Pappenberger et al. (2011) explain that the European flood alert system (EFAS) provides National Hydrological services with long-term probabilistic flood forecasting via an ensemble prediction system. Beyond efficiency of the probabilistic forecast, they show that ensemble forecast depends on geographical location and catchment spread and that the skill of the forecasted river discharges increases over 10 years forecasting. Pagano et al. (2013) dressed ensemble forecasts of 120 catchment using distribution of the model errors; they found that dressing methods can enhance reliability of raw ensembles. The COSMO-7 deterministic model and the probabilistic COSMO-LEPS have been coupled to the PREVAH hydrological model to help managers of the Sihl River in decision making (Addor et al., 2011).

This paper aims at simulating ensemble flow discharge of Senegal River upper the Manantali dam using the GR4J (Rural Engineering with four parameters issuing daily simulation) model in order to integrate commonly deterministic bias in further simulations (Makhlof, 1994; Caseri, 2017) by issuing probabilistic results. This approach allows streamlining results from classical hydrological modeling by considering the model uncertainty (Davilolo et al., 2013; Seo et al., 2010; Pagano et al., 2013; Henri et al., 2015). The ensemble stream flow is obtained by perturbing the input model with errors pattern cyclically designed on the basis of yearly pseudo periodic behavior of flow discharge at the Senegal River. Biases are used in designing the perturbation scheme from 14 running years of the model calibrated through two years. The members of the error scheme are then statistically enhanced using a multivariate Gaussian kernel without loss of original characteristics (Sharma et al., 2002) using forward and reverse Box-Cox. Then, the new statistical error pattern is used as perturbation for GR4J model already calibrated to produce ensemble simulation for the period ranging from 1965 to 1976. The corresponding ensemble in this paper is referred to as RAW-Ens. Indeed, multiple scenarios from models enable quantifying uncertainties in the forecasting processes through assessment of probabilities of categorized members of the ensemble (Sivillo et al., 1997; Keune et al., 2014; Krzysztofowicz, 2001). Data used for this study are from the OMVS (Organization of the Development of the Senegal River) and are composed of daily flows gauges between 1963 and 1976 at the Bafing flow gauge, precipitation and evapotranspiration of corresponding period of study. In addition, affine kernel dressing method (Jha et al., 2015; Lucatero et al., 2017; Silverman, 1998; Broecker., 2008) is used to dress the raw ensemble leading one other ensemble referred to as AKD-Ens. Goodness of the modeling schemes and their results are assessed on one hand through commonly used statistical criterions such the mean error, mean absolute error and the root mean squared error of ensembles; and on the other hand, through probability scores such as the Brier score, the rank probability score, the continuous rank probability. In addition, diagrams such as the attribute diagram and the relative operating diagram (ROC) are used as verification tool. Attribute diagram is used to check for reliability, resolution and skill of modeling systems. The ROC curve is exploited to investigate how hits vary against false alarm rates for each ensemble modeling scheme issuing the probabilistic simulations.

MATERIALS AND METHODS

Study area

The Bafing is the main tributary of Senegal River, that is, the second longer river of West Africa with catchment area of about 343 000 km² (Michel, 1973). The Bafing area is estimated at about 38000 km² and reach both Malian and Guinean areas (Sane et al., 2017) with its length estimated at about 670 km. In addition, it provides about 60% of the Senegal River inflow which slope estimated at about 5 m/km (Maiga et al., 1995). The main source of the Bafing catchment is the mountains of Fouta Djallon in Guinea (Figure 1). The multifunctional (hydropower, low flow support for irrigation and navigation and flood mitigation) dam of Manantali was set up in the Malian part of the Bafing catchment and provides about 12% of electric energy consumption in Senegal (Bader and Albergel, 2015). Local climate is composed of two types: sub-Guinean in Southern of the Bafing and Sudanese in Northern part (Sane et al., 2017). The data used in this study are from the OMVS (Organization for the Development of the Senegal River) database and are composed of daily flows gauged from 1963 to 1976 at the Bafing-Makana station upper the Manantali dam.
**GR4J model and raw error pattern**

GR4J model is a global conceptual rainfall-runoff hydrological model set up by the French Research Center namely the CEMAGREF. It is briefly described below without detailing its operating mode. The GR4J (Rural Engineering with four parameters issuing daily simulation) model involves four parameters: the maximum capacity of production store referred to as $X_1$ (mm), the groundwater exchange coefficient noticed by $X_2$ (mm), the maximum capacity of routing store by $X_3$ (mm) and the time base of unit hydrograph $X_4$ (mm). Inputs model daily are constituted of observed flows discharges, rainfall and daily potential evapotranspiration (Traore et al., 2010; Perrin et al., 2003; Valéry et al., 2014). An error pattern is drawn to perturbation of the GR4J model to issuing ensemble simulations. Thus, the GR4J has been calibrated to set parameters, and then runs over 12 in addition to the two-calibrating period at daily time step. After run, cyclical biases of which period is 365 days are settled to build the raw error pattern. Error pattern used in the ensemble generation are inferred from years daily simulations noticed by $\hat{Q}_j$. Through the running years of the model already calibrated, overall obtained data will be 365×$N_i$ (Wilks, 2006). Let $\hat{Q}_{ij}$ be a matrix constituted of $i$ rows representing daily simulations and $j$ columns that represents yearly simulation time series.

Furthermore, $\hat{Q}_j$ represents a row vector containing simulated flow discharge of same day through the entire period of study and $\hat{Q}_i$ represents a column vector containing 365 simulation of a given year. In this study, simulated flows are settled so that rows represent the years and the columns define the dates of the corresponding years. So, they can be written as:

$$\hat{Q}_j^T = [Q_{1,j}, Q_{2,j}, Q_{3,j}, \ldots, Q_{365,j}]$$

(1)

Simulations in overall years are written in matrix form such as $\hat{Q}^T$ in which $N_i$ rows correspond to years and columns to the calendar dates. Data are arranged to have good scheme for further statistical error generation. Finally, overall flow discharges from the GR4J through $N_i$ years are drawn in matrix form: $\hat{Q}^T = [\hat{Q}_{ij}]^T$. For more understanding of the process, daily flow discharges are schemed as (Wilks, 2006; Vincendon, 2011; Wu 2011; Fard and Hajighaie-Keshteli (2018a,b,c; Fard et al., 2018a,b,c):

$$[\hat{Q}]^T = [\hat{Q}_{ij}]^T$$

(2)

In a similar structure, simulated flows are compared with associated...
observations noticed by [Q], thereafter, bias are extracted at each date by calculating the differences between simulated and observed flow discharges. Bias is in same scheme used as raw error pattern in the purpose of ensemble producing and is given by:

$$\delta = ([\bar{Q}] - [Q])^T$$  \hspace{1cm} (3)

### Processing of error of the model

#### Box-Cox transformation

The Box-Cox transformation is commonly used to statistically normalize a data set. This operator is applicable either to univariate or multivariate data. Indeed, bias (errors) obtained characterized by the difference between simulations and observations in time series modeling are not often normal. Also, the Box-Cox transformation is applied to error pattern including the mean error. So, to generate more likelihood errors, multivariate Gaussian kernel density function is performed. Normalization is done to allow the use of Gaussian kernel error for augmenting members of the error pattern. Based on this, the Box-Cox inverse is applied to restore the initial structure of the error pattern. In this study, application of the Box-Cox transformation on the error matrix $[\delta]$ lead to normalized data referred to as $[\delta']$ with random residual. The transformation is given by the following formula (Bickel and Doksum, 1981 Sakia, 1992):

$$[\delta'] = \begin{cases} 
\frac{\text{sgn}([\delta]) ([\delta])^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\
\frac{\text{sgn}([\delta]) \log ([\delta])}{\lambda} & \text{if } \lambda \neq 0 
\end{cases}$$  \hspace{1cm} (4)

In dressing methods such as the best member method (Fortin et al., 1988; Hemri et al., 2015), data in matrix form requires estimated kernel density of the raw biases from the GR4J run needed to generate likelihood error pattern. Indeed, through covariance matrix and mean of the raw error pattern, a kernel density function is defined to enhance members the error pattern (estimate). The covariance matrix of data from the Box-Cox transformation is given by:

$$\Sigma_{\delta'} = E[(\delta'\delta'^T - \mu\mu^T)]$$  \hspace{1cm} (5)

Where, $\sigma_{m,n} = E((\delta'_{m} - \mu_m)(\delta'_{n} - \mu_n))$, $\mu_n = \frac{1}{N_t} \sum_{i=1}^{N_t} \delta'_{ln}$ and $\mu = [\mu_1, \mu_2, \mu_3, \ldots, \mu_N]$. After applying the Box-Cox processing, statistical ensemble is generated from the Gaussian multivariate probability density function which mean $\mu$ and covariance matrix $\Sigma_{\delta'}$. Generated error pattern is noticed by $e_{k}$ and follows theoretically to the same low distribution as the normal data from forward Box-Cox application with statistical characteristics $\mu$ and $\Sigma_{\delta'}$. The multivariate probability density function of the smooth ensemble is given as follow (Wilks, 2002):

$$F(\delta') = 2\pi^{-N_t/2} |\text{det}\Sigma|^{-1/2} \exp\left[-\frac{1}{2} (\delta' - \mu)^T \Sigma^{-1} (\delta' - \mu)\right]$$  \hspace{1cm} (6)

A statistical likelihood error pattern is produced using the inverse of the $\Phi$ linear transformation of the raw error pattern $\delta'$. The linear transformation satisfying our target is defined by $\Phi = A^{1/2} E$. It is important to highlight that eigenvalue-eigenvector matrixes have same information as the $\Sigma_{\delta'}$ symmetric matrix (Wilks, 2006). Let $(\delta' - \mu)$ centralized error pattern of zero mean and covariance matrix equal to the identity matrix $I_d$ in accordance to the central limit theorem; thus transformed data is obtained such:

$$\delta'' = \Phi(\delta' - \mu) + \mu \sim N(\mu, \Sigma_{\delta''})$$  \hspace{1cm} (7)

Reverse transformation of Equation 7 is used to generate studied statistical errors pattern referred to as $[\delta]$, in which members will be enhanced according to the following equation (Wilks, 2002; Thomas et al., 2007; Sharma et al., 2002):

$$[\delta] = \Phi^{-1}[\delta''] - \mu + \mu \sim (\mu, \Sigma_{\delta'})$$  \hspace{1cm} (8)

Larger member of errors is generated to increase the number of members in the ensemble stream flow simulation being set up. The number of the ensemble members is enhanced from $N_t$ to $N=50$ which $N > N_t$.

#### Ensemble dressing

It consists of transforming ensemble issues into a continuous distribution function. In dressing methods, independent and larger ensemble is drawn through the raw issued ensemble members (Brocker and Smith, 2008). It is done using an adequate kernel dressing (Gogonet, 2013; Wang et al., 2005). In dressing process, ensemble members for given simulation dates are dressed by a Gaussian kernel in which mean and variance are either or not fitted or scaled by appropriate fitting coefficients (Silverman, 1998; Broecker and Smith, 2008). In this paper, two dressing methods that overcome the shortcoming of the Gaussian density function (Brocker and Smith, 2008) are tested. One of them is the Silverman rule of Thumb (SVMRT). The SVMRT consists of smoothing some discrete ensemble members. For each date, original data is replaced by one Gaussian kernel with non-fitted mean and fitted variance. In this paper, ensemble from this method is noticed by SVMEns. The Gaussian kernel density of an ensemble with parameters denoted $\Sigma$ and $\bar{X}$ can be written as:

$$K_d(Ens) = 2\pi^{-N_t/2} |\text{det}\Sigma|^{-1/2} \exp\left[-\frac{1}{2} (\text{Ens} - \bar{X}/\Sigma)^2\right]$$  \hspace{1cm} (9)

The other method is the Affine Kernel Dressing (AKD). In this method, five parameters are involved in order to dress a new statistical ensemble using estimate mean and variance for a Gaussian kernel (Bröcker et al., 2008; Sharma et al., 2002). The basis of the AKD dressing is to smooth the raw ensemble from the ensemble modeling scheme. The smoothing involves five free parameters that either scales or shifts the mean or the variance of the raw ensemble in addition to the Silverman factor (Silverman, 1986; Bröcker and Smith, 2008, Fathollahi Fard et al., 2018a,b; Navid et al., 2018) such as:

$$\bar{X}_d = r_1 + r_2[\bar{X}] + a[X]$$  \hspace{1cm} (10)

$$\Sigma_d = (4/3K)^{0.4} (S_1 + S_2a^2 \text{Var}[X])$$  \hspace{1cm} (11)

Parameters in the AKD process are on one hand, fitted through the raw ensemble already settled. Ensemble from the AKD process in which parameters are fitted minimizing the continuous rank
probability scores (CRPS) is called in this paper the AKDEns fit. In this method, fitting parameters involve interposing a logarithmic barrier to ensure positive variance for the Gaussian kernel. The logarithmic barrier is defined as:

$$\overline{CRPS}_d = CRPS ([X], Obs) \ast [1 + 0.01 \ast \max \theta - \log \min \text{Var} (X_d)]$$

(12)

The AKD processing leads to new continuous distribution function from the $K$ member of the raw ensemble probability by $p(y/\text{Ens}, \theta)$ so that.

$$p(y/\text{Ens}, \theta) = \left(1/K\right) \left(\sum_{i=1}^{K} K_d(y - X_d/i)\right)$$

(13)

$y =$espectations; $\theta=$ estimated parameters ($a, r_1, r_2, s_1, s_2$) of the AKD post-processing and $\text{Ens}$= raw ensemble from the model.

**Ensemble verification**

**Brier score (BS) and its decomposition**

This score was set up by Brier in 1950 and determines the mean squared errors between predictive probability $p_i$ at time period $i$ ($i = 1, ..., N$) and the binary observations $o_i$ with value 1 in case when observation exceeds a probability threshold $p_i$ and 0, otherwise, at time period (Addor et al., 2011). The Brier score is given by following formula:

$$BS = \sum_{i=1}^{N} (p_i - o_i)^2$$

(14)

Mathematically, the lower the Brier score value the best the ensemble modeling system. Thus, optimum Brier score corresponds to zero for perfect modeling. Perfect modeling means that all expected scenarios are observed. The BS can be decomposed into three components: reliability, resolution and uncertainty (Murphy, 1973; Talagrand et al., 1997; Atger, 2004).

$$\frac{1}{K} \sum_{k=0}^{K} N_k (p_k - o_k)^2 - \frac{1}{K} \sum_{k=0}^{K} N_k (\hat{o}_k - \bar{o})^2 + \bar{o}(1 - \bar{o})$$

(15)

Where $\bar{o}$ represents total frequency of the events.

Reliability measures difference between simulated ensemble and the mean observation. The resolution estimates skillfulness of an ensemble modeling system in considering lower or higher probability events. Simply, the BS characterizes skill of a model in covering more cases. Wide resolution means possibly clustering observations to categories with considerable difference. Uncertainty in the BS decomposition measures the variability of the data. Generally, uncertainty in the BS decomposition does not impact on reliability of the model.

**Rank probability score (RPS)**

The RPS is based on comparison between cumulative distribution functions of simulations and observations in subsequent percentiles. Thereafter, the observations cumulative distribution function is represented by a cumulative function with probability density taking the value of 1 when the probability threshold is exceeded and 0 otherwise (Epstein, 1969; Murphy, 1971). The RPS is obtained by summing the Brier score at different percentiles (10%, ... 30%, ... 90%). This approach is useful for the purpose of categorizing river flows. For example: low flow (<10%), medium flows (10% < Q < 90%) and high flows (> 90%) (Zalachori, 2013). Thus, different probability states can be obtained for an ensemble according to chosen percentiles. Let $N$ be the spread of an ensemble forecast in which Brier score is calculated for $M$ categories of the ensemble determined for a percentile. The rank probability score at time $i$ is given by:

$$RPS_i = \frac{1}{M-1} \sum_{m=1}^{M} \left( \sum_{i=1}^{m} (f_{i,j} - o_{i,j})^2 \right)$$

(16)

$f_{i,j}$ denotes the forecast probability of category $m$ at time $i$, $o_{i,j}$ represents binary variable which value 1 when the category $m$ is observed and 0 if associated category is not observed at time $i$. Thus, the RPS over the simulation period is obtained through following formula:

$$RPS = \frac{1}{N} \sum_{i=1}^{N} RPS_i$$

(17)

**Continuous rank probability score (CRPS) and its decomposition**

The continuous rank probability score is an extension to infinity of the discrete rank probability score. The CRPS determines global quality of an ensemble issue. In CRPS assessment, the numbers of category are stretched to infinity. Further, when the $M$ numbers of category is well stretched in equation of the RPS, the CRPS is obtained (Matheson and Winkler, 1976; Hersbach, 2000; Zalachori, 2013). This score measures the distance between the ensemble and the observations. It can be defined as integral over infinite limit interval of the Brier score. Variables involved in estimating the CRPS are: distribution function of the simulated values and other one distribution function with values of 0 when the simulated values are lower than observed one and 1 otherwise (Heaviside function).

Equation of the continuous Rank Probability Score (CRPS) is formulated as follows (Marty, 2013; Casati et al., 2008; Gogonel et al., 2014):

$$\text{CRPS} = \frac{1}{N} \sum_{i=1}^{N} \int_{x=+\infty}^{+\infty} \left( F_i^1(x) - F_i^0(x) \right)^2 \text{dx}$$

(18)

$$\overline{CRPS} = \int_{-\infty}^{+\infty} \text{BS}(x) \text{dx}$$

Where $F_i^1(x)$ is d expectation distribution function at time $i$. $F_i^0(x)$ is the heaviside function and $N$ is number of time step.

**Attribute diagrams**

Attribute diagram is the plot of predictive probabilities against frequency of occurrences. It allows assessing if predictive probability agrees with observed frequency. Attribute diagram gives reliability, resolution and the uncertainty of an ensemble system with respect to the observation and is concretely based on analysis of underlying plot relatively to some graphical characteristics (1:1 line of perfect simulation, area of skill and resolution). Perfect simulation leads to superposition of the curve with the first bisector.
Uncertainties are defined by the distances between point of classes for given threshold and the first bisector (Achter, 2004). The resolution estimates how much the forecasts differ from the climatological mean probabilities of the event, and how the systems get it right. Thus, attribute diagram is used in this study to assess how implemented ensemble simulation scheme performs.

**ROC curve**

The ROC (Relative Operating Characteristic curve) curve allows evaluation of separation of means of conditional distribution for simulated data and is a popular tool for decision making. ROC curve is obtained by plotting the false alarm rate against hit rate. The ROC is positive when the curve is beyond the 1:1 line with area under the ROC (AUC) of 0.5. Perfect modeling corresponds to AUC of 1. The AUC defines the probability of successfully discriminating an event from a nonevent (Mason, 1982; Atger, 2001). Here, expectation relevant to the modeling scheme is referred to as \( f \) and is defined as a probability exceeding a threshold value noticed by \( p_x \) : \( p (x \geq p_x) \). Relatively, \( f_0 \) defines the non-expected event characterizing probability of \( x < p_1 \). Let characterized on one hand be expected and observed events (Hit) by \( f_{so} = \sum_{i=1}^{N}(f_0 = 1) \) and by \( f_{sno} = \sum_{i=1}^{N}(f_0 = 0) \), expected and unobserved events (False alarm). On the other hand, it is represented by \( f_{bno} = \sum_{i=1}^{N}(f_0 = 1) \) as the non-expected and observed events (Miss) and by \( f_{bno} = \sum_{i=1}^{N}(f_0 = 0) \) as the non-expected and unobserved events (Correct rejection). The false alarm and hit rates can be defined as follow:

\[
F_{ar} = \frac{f_{sno}}{f_{bno} + f_{sno}} \\
H_r = \frac{f_{so}}{f_{b0} + f_{s0}}
\]

Hit and false alarm rates are calculated using the following contingency Table 1.

**RESULTS AND DISCUSSION**

**Processed errors for ensemble making**

Errors from past simulations of GR4J model already calibrated and runs upper the Manantali dams are settled in matrix form with rows corresponding to variables and columns to observations. Estimate parameter (\( \lambda = 0.24 \)) of the Box-Cox is found using the least squared method at significant level of 0.5. The \( p \)-value and the statistic of the estimation are respectively 0.143 and 0.902. Normalization is done aimed at extending the number of members through multivariate Gaussian kernel. Result gives statistical error pattern with a random residual. Members of the ensemble errors are extended from 15 to 50. Reverse Box-Cox transformation is further applied to restore initial properties (statistic characteristics) of the errors pattern. The set of 50 member’s errors is used to perturb flows in calibrating the GR4J model further runs.

**Plots of the Raw-Ens and AKD-Ens**

The raw ensemble with the post-processed one are assessed here. In this study ensemble less than the raw ensemble (RAW-Ens) and the ensemble from dressing method (AKD-Ens) are plotted and brief exploratory analysis of ensemble is made concerning the representativeness of the two ensembles. It is important to highlight the fact that width of ensemble makes representation difficult using common computers. Then, in this study it is done reducing time span of the overall time series. Indeed, direct manipulation of ensemble simulation allows discovering related shortcoming. Thus, ensembles are represented between 01/01/1965 and 31/12/1968 (Figures 3 and 4). Exploratory analysis of both RAW-Ens and AKD-Ens ensembles hydrograms show ensemble can be assumed to likely represent observations and it is important to notice that hydrogram of the observation goes up and down ensemble. Thus, good representation can be assumed.

**Deterministic characteristics and scores**

Both the deterministic and the probabilistic evaluation criteria were studied. The dressed ensemble gives better results even if the raw one remains good ensemble. The mean error (Me) has been significantly improved: 0.519 for RAWEns to 0.026 for the AKD-Ens. The mean absolute error (EnsMae) has been improved by 27.5% and the ensemble root mean squared error (EnsRmse) by 32.64%. These ensembles characteristics with additional probability scores are presented in Tables 1 and 2. Results show that the affine kernel dressing can be used to improve quality of simulations from an ensemble modeling scheme. Thus, for more reliable
Table 2. Deterministic characteristics of the ensembles with BS, RPS and CRPS.

| Ensemble | EnsMe | EnsMae | EnsRmse | BS    | RPS    | CRPS   |
|----------|-------|--------|---------|-------|--------|--------|
| RAW-Ens  | 0.519 | 0.771  | 8.018   | 0.1166| 0.2002 | 0.5824 |
| AKD-Ens  | 0.026 | 0.559  | 3.947   | 0.0842| 0.2095 | 0.4095 |

Table 3. Performance characteristics of assessed ensembles.

| Ensemble | BSReli | Res | Unc | BSpot | CRPSReli | CRPSpot | AUC   |
|----------|--------|-----|-----|-------|----------|---------|-------|
| RAW-Ens  | 0.0036 | 0.1058 | 0.2222 | 0.1164 | 0.2798  | 0.3026  | 0.9018 |
| AKD-Ens  | 0.0113 | 0.1020 | 0.2222 | 0.1202 | 0.0208  | 0.3888  | 0.9224 |

Simulation of flow discharges upper the Manantali dam, this approach can be used. From both performed ensembles, probabilistic issues in bin observations are produced using tercile clustering. For tercile simulation, three categories of data simulation are drawn for each calendar date according to a probability threshold and the first category is retained with its associated probability. Probabilistic simulation leads to acceptable Brier scores for the two schemes (RAW-Ens and AKD-Ens). The best Brier Score is 0.0842 (AKD-Ens), whereas the best RPS and CRPS are respectively 0.2002 (RAW-Ens) and 0.3855 (AKDEns) and 0.4090.40955 (AKD-Ens). Detailed results concerning the ensembles probability scores including reliabilities and spots characterizing the difference between uncertainties and resolutions through BS and CRPS for each ensemble are also shown in Tables 2 and 3. It is important to keep in mind that for probabilistic evaluation, the closer to zero the scores the better the system is. If scores are closer to one, model is not reliable and the use of such system is discouraged. Between the two ensembles, quality of simulation varies according to the evaluating method. Consequently, dressing method remains improving method for probabilistic simulation.

**Attribute diagrams of the ensemble schemes**

Attribute diagram show simultaneously reliability, resolution and uncertainty of one performed ensemble. In the attribute diagram, frequency of issued probabilities is plotted against binned observations. For perfect modeling, the plot of frequencies against binary observation fits the first bisector. Interpretation of the reliability diagram is based on exploratory analysis of the behavior of observed frequency against probability plot in comparison to the first bisector. In addition, it is also based on analysis of the spread of the resolution band and position of the curve relatively to the lines delineating the skill and the resolution (Hamill, 2006; Wilks, 2006) of the model. Attribute diagrams of the two models are presented in Figure 5. Good reliability is assigned to both models (Figure 4a and b). The affine kernel dressing method diminish the reliability (because distance between Probability of simulation vs Observed Relative Frequency has been increased) and resolution of the raw ensemble according to the Brier score, while overall reliability through the continuous rank probability score and discrimination have been improved. In comparison to Figure 2, both ensemble modelling systems are reliable and can be used for water planning at the Manantali dam.

**Roc plots**

Discrimination of an ensemble from a system can be assessed by analysis of ROC curves. Performing the ROC curve is based on determining the success rates (Hits) and failure rates cases for an implemented ensemble modeling system. For a set of probability exceedances ranging from 0.1 to 1, the sum of number of simulated and observed events and sum of number of simulated and non-observed events are calculated. For each probability threshold, strictly lower events are considered as non-expected, then are classified into two categories: non-expected and non-observed on one hand and non-expected and observed on the other. The results for the different systems are shown in Figure 6. For two systems (RAW-Ens and AKD-Ens), Hits prevail on False alarm rates except for the AKDEns. Their AUC are over the line of skill. Also, the systems have good skill in reproducing flow discharges upper the Manantali dam. Ensemble systems considered as skillful in probability simulation of flows is based on the fact that Hits prevails on False rates for most of the probability thresholds. Balance sheet analysis can be used to visually assess skill of a model with regard to bins which Hit rates are beyond or down the 1:1 line of no skill. Hence, both systems lead to more success than failures and can be used to produce probabilistic simulation of flow discharges at the Senegal River. Elsewhere, through ROC curves, decision making can be improved particularly when...
faced with extreme events of which threshold value is fixed (drought or flood). Once more, the dressing method improves issues from the scheme providing the RAW-Ens. The AUC has been improved by 2.3% and is
greater than 0.5 that corresponds to the 1:1 line of the ROC diagram. Besides, for failing system, the ROC curve goes down the 1:1 line of no skill.

**Conclusion**

Simulating or forecasting ensemble hydro climatic variables is an ingenious way to quantify uncertainties related to forecasting and simulation processes. In order to overcome difficulties of bias inherent to hydro climatic modeling, scientists have been oriented towards probability theories in order to consider modeling uncertainties through issuing probabilistic values. Through ensembles processing, expectations are associated to a probability of occurrence, and then naturally of non-occurrence. This paper thus aims at producing probabilistic simulation of flow discharge at Senegal River upper the Manantali dam. The main ensembles called the raw ensemble is set up by perturbing the input flow of the GR4J model. Furthermore, three other ensembles are also set up through post-processing tools used in ensemble forecast systems processing. Inputs are perturbed using past errors pattern of the GR4J runs upon 14 years. The error pattern in this study is designed on the basis of the periodic behavior of Senegal River flow over 365 days. Box-Cox transformation is used to normalize the raw error pattern in order to generate wide members through a multivariate Gaussian kernel. Reverse Box-Cox transformation is also applied to restore initial statistical characteristics of the errors pattern. Drawn perturbations are applied to recorded flow discharges used in calibrating the GR4J model. From this process, ensemble streamflow is performed in which members are 51 including the ensemble mean. A post-processing referred to as Affine kernel dressing has been applied to the raw ensemble (RAWEns) from turned GR4J with perturbed input (AKD-Ens). Ensemble
forecast verification tools are used to evaluate the reliability, resolution and skill of the tested ensemble modeling systems schemes. The results are satisfactory for two systems. Globally, the dressing method improves in this case of study the raw ensemble. Both ensembles are reliable and can be used to help in decision making at the Manantali dam though predominance of hit on false alarm, rates.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

ACKNOWLEDGEMENTS

The authors appreciate the Senegal UCAD IRD and the Organization for the Development of the Senegal River (OMVS) for their contributions. Also worthy of thanks are the reviewers of this paper for the valuable suggestions given during the review process.

REFERENCES

Addor N, Jaun S, Fundel F, Zappa M (2011). An operational hydrological ensemble prediction system for the city of Zurich (Switzerland): skill, case studies and scenarios. Hydrology and Earth System Sciences 15:2327-2347.

AFTER F (2004). Estimation of the reliability of ensemble-based probabilistic forecasts. Quarterly Journal of the Royal Meteorological Society 130:627-646.

Bader JC, Albergel J, Atger F (2015). Objective Stochastic Closed-Loop Supply Chain Network Design with Social Considerations. Applied Soft Computing 71:505-525.

Fathollahi-Fard AM, Hajighaee-Keshtefi M (2018a). A bi-objective partial interdiction problem considering different defensive systems with capacity expansion of facilities under imminent attacks. Applied Soft Computing 68:343-359. Fathollahi Fard AM, Hajighaee-Keshtefi M. (2018b). A bi-objective green home health care routing problem, Journal of Cleaner Production 200:423-443. Fathollahi Fard AM, Hajighaee-Keshtefi M (2018c). Hybrid optimizers to solve a tri-level programming model for a tire closed-loop supply chain network design problem. Applied Soft Computing 70:701-722. Fathollahi-Fard AM, Hajighaee-Keshtefi M, Moghaddam R (2018b). The Social Engineering Optimizer (SEO). Engineering Applications of Artificial Intelligence 72:267-293. Gogonel AG (2013). Statistical Post-Processing methods and their implementation on the Ensemble Prediction Systems for Forecasting temperature. In the use of the French electric consumption, Paris Descartes University. Laboratoire MAP 5, École Doctorale de Sciences Mathématiques de Paris Centre. Gogonel A, Collet J, Bar-Hen A (2014). Implementation of two statistical methods for Ensemble Prediction Systems in the management of electrical systems. CS-BIGS 5(2):74-87. Hamill TM, Juras J (2006). Measuring forecast skill: is it real skill or is it the varying climatology? Quarterly Journal of the Royal Meteorological Society 132:2905-2923. He M, Whitin B, Hartman R, Henkel A, Fickenschers P, Staggs S, Morin A, Imgarten M, Haynes A, Russo M (2016). Verification of Ensemble Water Supply Forecasts for Sierra Nevada Watersheds. Hydrology 3(4):35. Henri S, Lieniak D, Klein B (2015). Multivariate postprocessing techniques for probabilistic hydrological forecasting. Water Resources Research 51:7436-7451. Jeong D-I, Kim Y-O (2005). Rainfall-runoff models using artificial neural networks for ensemble streamflow prediction. Hydrological Processes 19:3819-3835. Jha SK, Shrestha DL, Stadnyk T, Coulibaly P (2015). Evaluation of ensemble precipitation forecasts generated through postprocessing in a Canadian catchment. Hydrology and Earth System Sciences Discussion. https://doi.org/10.5194/hess-2017-31. Krzyztopfowicz R (2001). The case for probabilistic forecasting in hydrology. Journal of Hydrology 249:2-9. Laio P, Gabellari S, Reborbora N, Rudar R, Ferraris L, Ratto S, Stevenin H, Cauduro M (2014). Validation of the Flood-PROOF probabilistic forecasting system. Hydrological Processes 28:3466-3481. Lucatero D, Madsen H, Refsgaard JC, Kidmose J, Jensen KH (2017). Seasonal streamflow forecasts in the Ahlergaarda catchment, Denmark: effect of preprocessing and postprocessing on skill and statistical consistency. Hydrology and Earth System Sciences Discussion https://doi.org/10.5194/hess-2017-379. Maiga M (1995). Le bassin du fleuve Sénégal : de la Traite négrière au développement sous régional autoctone. Paris, L’Harmattan. 321 p. Makhlouf Z (1994). Compléments sur le modèle pluie-débit GR4J et essai d’estimation de ses paramètres. PhD thesis, University of Paris-Sud. Marty R, Zin I, Obled Ch (2013). Sensitivity of hydrological ensemble forecasts to different sources and temporal resolutions of probabilistic quantitative precipitation forecasts: flash flood case studies in the Cévennes-Vivarais region (Southern France). Hydrological Processes 27:33-44. Mason SJ (1982). A model for assessment of weather forecast. Met Magazine 30:291-303. Michel P (1973). Les bassins des fleuves Sénégal et Gambie. Etude géomorphologique. ORSTOM document, n° 63. ORSTOM, Paris. Murphy AH (1973). A new vector partition of the probability score. Journal of Applied Meteorology 12:595-600. Navid S, Fathollahi Fard AM, Hajighaee-Keshtefi M (2018). Sustainable tire closed-loop supply chain network design: Hybrid metaheuristic algorithms for large-scale networks. Neural Computing and Applications 1-35. Pagano TC, Shrestha DL, Wang QJ, Robertson D, Hapuarachchi P (2013). Ensemble dressing for hydrological applications. Hydrological Processes 27:106–116. Pappenberger F, Ramos M H, Cloke H L, Wetterhall F, Alfieri L, Bogner K, Mueller A, Salamon P (2015). Journal of Hydrology 522:697–713. Pappenberger F, Cloke H, Del Medico M (2011). The impact of weather forecast improvements on large scale hydrology analyzing a Decade of forecasts of the European Flood Alert System.
 Hydrological Processes. 25:1091-1113.
Perrin C, Michel C, Andréassian V (2003). Improvement of a parsimonious model for streamflow simulation. Journal of Hydrology 279(1-4):275-289.
Sakia RM (1992). The Box-Cox transformation technique: a review. The Statistician. 41:169-178.
Sane ML, Sambou S, Ndione DM, Leye I, Kane S, Badi ML (2017). Analyse et traitement des séries de débits annuels et mensuels sur le Fleuve Sénégal en amont du barrage de Manantali : cas des stations de Bafing Makana et Dakka Saidou.
Sivillo JK, Ahlquist JE, Toth Z (1997). An Ensemble Forecasting Primer. Weather and forecasting. 12:809-818.
Sharma A, O’Neill R (2002). A nonparametric approach for representing interannual dependence in monthly streamflow sequences. Water Resources Research 38(7):1-10.
Silverman BW (1998). Density Estimation for Statistics and Data Analysis. London: Chapman & Hall/CRC. ISBN 0-412-24620-1.
Silverman BW (1986). Density Estimation for Statistics and Data Analysis First Edition. Chapman and Hall, London.
Thomas DB, Luk W (2007). Sampling from the Multivariate Gaussian Distribution using Reconfigurable Hardware. International Symposium on Field-Programmable Custom Computing Machines. 0-7695-2940-2/07 $25.00 © 2007 IEEE.
Valéry A, Andréassian V, Perrin C (2014). As simple as possible but not simpler: what is useful in a temperature-based snow-accounting routine? Water Resources Research 50(1):273-284.
Wang X, Bishop CH (2005). Improvement of ensemble reliability with a new dressing kernel. Quarterly Journal of the Royal Meteorological Society 131:965-98.
Wilks DS (2002). Smoothing forecast ensembles with fitted probability distributions. Quarterly Journal of the Royal Meteorological Society 128:2821-2836.
Wilks DS (2006). Statistical Methods in the Atmospheric Sciences. Second Edition. Academic Press. International Geophysics Series 91:627.
Wu L, Seo D-J, Demargne J, Brown J-D, Cong S, Schaake J (2011). Generating of ensemble precipitation forecast from single-valued quantitative precipitation for hydrological ensemble prediction. Journal of Hydrology 399:281-298.
Zalachori I (2013). Prévisions hydrologiques d’ensemble : développements pour améliorer la qualité des prévisions et estimer leur utilité. Thesis. AgroParisTech.

Vincendon B, Ducrocq V, Nuissier O, Vié B (2011). Perturbation of convection-permitting NWP forecasts for flash-flood ensemble forecasting. Natural Hazards Earth System Sciences. 11:1529-1544.
Wang X, Bishop CH (2005). Improvement of ensemble reliability with a new dressing kernel. Quarterly Journal of the Royal Meteorological Society 131:965-98.
Wilks DS (2002). Smoothing forecast ensembles with fitted probability distributions. Quarterly Journal of the Royal Meteorological Society 128:2821-2836.
Wilks DS (2006). Statistical Methods in the Atmospheric Sciences. Second Edition. Academic Press. International Geophysics Series 91:627.
Wu L, Seo D-J, Demargne J, Brown J-D, Cong S, Schaake J (2011). Generating of ensemble precipitation forecast from single-valued quantitative precipitation for hydrological ensemble prediction. Journal of Hydrology 399:281-298.
Zalachori I (2013). Prévisions hydrologiques d’ensemble : développements pour améliorer la qualité des prévisions et estimer leur utilité. Thesis. AgroParisTech.