



**HAL**  
open science

## Strategies for hydrologic ensemble generation and calibration: On the merits of using model-based predictors

Anne-Laure Tiberi-Wadier, Nicole Goutal, Sophie Ricci, Philippe Sergent, Maxime Taillardat, François Bouttier, Céline Monteil

### ► To cite this version:

Anne-Laure Tiberi-Wadier, Nicole Goutal, Sophie Ricci, Philippe Sergent, Maxime Taillardat, et al.. Strategies for hydrologic ensemble generation and calibration: On the merits of using model-based predictors. *Journal of Hydrology*, 2021, 599, pp.126233. 10.1016/j.jhydrol.2021.126233. hal-03520689

**HAL Id: hal-03520689**

**<https://hal.science/hal-03520689>**

Submitted on 11 Jan 2022

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

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

This is an author's version of a published peer-reviewed article.  
The accepted article was published online in Journal of Hydrology on 24 March 2021  
under the doi:10.1016/j.jhydrol.2021.126233  
This submitted version is hereby made freely available 6 months after publication in  
in accordance with French law regarding Government funded research  
(loi du 7 octobre 2016 pour une République Numérique)

## Strategies for hydrologic ensemble generation and calibration: on the merits of using model-based predictors

Anne-Laure Tiberi-Wadier<sup>a,\*</sup>, Nicole Goutal<sup>b</sup>, Sophie Ricci<sup>c</sup>, Philippe  
Sergent<sup>d</sup>, Maxime Taillardat<sup>e</sup>, François Bouttier<sup>f</sup>, Céline Monteil<sup>g</sup>

<sup>a</sup>*Cerema Eau Mer et Fleuves, Plouzané, France*

<sup>b</sup>*Saint-Venant Hydraulics laboratory, EDF R&D, Chatou, France*

<sup>c</sup>*CECI, CNRS UMR 5318/CERFACS, Toulouse, France*

<sup>d</sup>*Cerema Eau Mer et Fleuves, Margny-Lès-Compiègne, France*

<sup>e</sup>*Météo-France, Toulouse, France*

<sup>f</sup>*CNRM, Toulouse University, Météo-France and CNRS, Toulouse, France*

<sup>g</sup>*EDF R&D LNHE - Laboratoire National d'Hydraulique et Environnement, Chatou, France*

---

### Abstract

This paper investigates the hydrometeorological chain with an ensemble approach. The objective is the generation of Hydrologic Ensemble Forecasts (HEF) on the Odet catchment (France, Brittany), using the Quantile Regression Forest (QRF) method usually applied for the ensemble calibration of meteorological forecasts. First, a Global Sensitivity Analysis (GSA) in the distributed MORDOR-TS model is carried out taking into account uncertainty in forecasted rain with AromeEPS-RR1 and in model parameters. GSA highlights the role and importance of the different hydrologic model parameters during rain events and allows to only take into account the most influent parameters for the generation of an Hydrologic Ensemble Forecast (HEF).

Three strategies for the generation of HEF are then compared. First (i), a raw ensemble is built with a model-based only approach using the deterministic forecast rainfall Expert-RR3 and perturbed model parameters, without further statistical calibration. Then, the QRF calibration method is used to generate two ensembles of quantiles: (ii) the observation-based approach uses only predic-

---

\*Corresponding author

Email address: Anne-Laure.Tiberi-Wadier@cerema.fr (Anne-Laure Tiberi-Wadier)

tors that are independent from hydrology, whereas (iii) the combined model and observation approach combines these predictors with statistics of the raw hydrologic ensemble (mean, standard deviation). This latter approach was shown to outperform the previous ones, enhancing the importance of the choice of the predictors in the QRF method. In the prospect of using the hydrologic ensemble as input for hydraulic simulation, the Ensemble Copula Coupling method (ECC) and a trajectory smoothing procedure is then applied on (iii). This step slightly deteriorates the reliability of hourly streamflows, yet Continuous Ranked Probability Score (CRPS) and forecast skills on the cumulated or maximum streamflows are improved.

*Keywords:* Ensemble Prediction System, Global Sensitivity Analysis, Hydrologic Ensemble Forecast, Quantile Regression Forest, MORDOR-TS, Odet catchment

*2010 MSC:* 00-01, 99-00

---

## 1. Introduction

In spite of the great efforts that have been put into land surface models coupled to hydrologic models, deterministic simulation and forecast of streamflow remains limited mostly due to the stochastic nature of precipitation and the complexity of meteorological and hydrologic processes. Thus an ensemble approach is favored; it provides a probabilistic hydrologic forecast needed by decision support systems dealing with risk-based stakes in real-time (low and high flow) as well as by water resources management actors (hydropower production, irrigation, navigation, tourism). Forcing hydraulic models with forecasted hydrologic inflow allows to extend forecast lead time at stations where security and production are at stake. Before use for application, the performance of the ensemble forecast should be assessed (Jolliffe and Stephenson, 2003) in terms of reliability and resolution, with respect to observations. The reliability refers to the statistical consistency between the ensemble and the observations. The resolution describes the ability of the ensemble to discriminate situations leading or not to an event.

### 1.1. Hydrometeorological uncertainties

The objective of ensemble forecasts is to cover and represent the uncertainties existing in the prediction chain. Numerous and various sources of uncertainties along the hydrometeorological simulation chain lead to uncertainty in discharge simulation and forecast. Three main sources of hydrometeorological uncertainties are acknowledged in the literature (Bourgin, 2014; Zappa et al., 2011; Thibault et al., 2016; Demargne et al., 2014): atmospheric forcing observation and prediction, hydrologic model initial condition, hydrologic model structure and parameters. According to the classification proposed by Krzysztofowicz (1999), atmospheric forcing is referred to as input uncertainty, whereas the other sources are referred to as hydrologic uncertainties.

On the one hand, input uncertainty, i.e meteorological uncertainty, can be accounted for by an Ensemble Prediction System (EPS), which accounts for

30 uncertainties in initial conditions and model physics in Numerical Weather Pre-  
31 diction (NWP) models. The most straightforward technique to issue Hydrologic  
32 Ensemble Forecasts (HEF) is to use EPS as input for a hydrologic model, thus  
33 producing an ensemble of discharges. Numerous studies and operational applica-  
34 tions have been conducted on this topic (Cloke and Pappenberger (2009), Papp-  
35 enberger et al. (2016)). On the other hand, hydrologic uncertainty is also  
36 taken into account to issue HEF considering model parameters uncertainties  
37 Dietrich et al. (2009) or using a multi-model approach (Hopson and Webster,  
38 2010; Velázquez et al., 2011; Thiboult et al., 2016; Bellier, 2018). Since 2004,  
39 research, operational and user communities gathered around the HEPEx ini-  
40 tiative (Hydrologic Ensemble Prediction EXperiment, [www.hepex.org](http://www.hepex.org)), which  
41 aims at advancing the science and practice of hydrological ensemble prediction  
42 and demonstrating their utility in decision making (Thielen et al., 2008; Schaake  
43 et al., 2010).

#### 44 *1.2. Uncertainty quantification for Hydrologic Ensemble Forecasts (HEF) gen-* 45 *eration*

46 The performance of the raw hydrologic ensemble is linked to how the different  
47 sources of uncertainties are accounted for in the ensemble generation. The HEF  
48 system should be built taking into account major sources of uncertainties (both  
49 atmospheric and hydrologic) with associated ranges of uncertainty. To that end,  
50 identifying and ranking sources of uncertainty is necessary. Given assumptions  
51 on uncertain parameters, this can be achieved with a sensitivity analysis (SA).  
52 A review of SA methods is available in Iooss and Lemaître (2015).

53 Global Sensitivity Analysis (GSA) (Saltelli, 2008) allows to quantify the  
54 contribution of model inputs to its outputs. It implies the integration of an en-  
55 semble of simulations from which sensitivity indices are computed; for instance  
56 Sobol indices based on variance-based methods (Efron and Stein, 1981). GSA is  
57 widely applied in hydrology. In Emery et al. (2016), a GSA is achieved in order  
58 to highlight the key parameters impacting the river-routing scheme Total Runoff  
59 Integrating Pathways (TRIP) that simulates river water height and discharge on

60 the Amazon catchment. Garambois et al. (2013) analyzed MARINE hydrologic  
61 model sensitivity during flash-floods. In Michon and Castaings (2017), a GSA  
62 on the Kling-Gupta Efficiency (KGE, Gupta et al. (2009)) of the MORDOR-TS  
63 hydrologic model was carried out on several French catchments.

### 64 *1.3. Statistical calibration of ensembles*

65 Ensemble approaches aim at representing the uncertainty along a simulation  
66 chain. However, the total uncertainty is rarely fully covered and raw EPS are  
67 known to be underdispersive and biased (Hamill and Colucci, 1997; Schaake  
68 et al., 2010). A post-processing step on the ensemble precipitation forecast  
69 (post-processing with respect to meteorology, pre-processing with respect to  
70 hydrology) should thus be achieved before using them as input for rainfall-  
71 runoff models and issuing HEF. Similarly, HEF should be post-processed to  
72 account for uncertainty in EPS as well as in hydrologic model parameters. The  
73 improvement of EPSs' and HEFs' reliability and resolution relies on a statistical  
74 post-processing, named ensemble calibration.

75 Statistical ensemble calibration is an active field of research in meteorology  
76 and hydrology. It exploits the relation between the past previsions and their  
77 corresponding observations (Wilks, 2018) to correct the forecast. Parametric  
78 and non-parametric calibration methods are reported in the literature. Para-  
79 metric methods rely on an a priori assumption for the output data distribution  
80 which parameters are identified by the calibration algorithm. In contrast, non-  
81 parametric methods are data-based only. Wilks (2018) describes the state of  
82 the art of statistical postprocessing of meteorological ensemble forecasts. A re-  
83 view of statistical ensemble calibration methods used in the field of hydrology  
84 is available in Li et al. (2017).

85 Two widely used parametric methods are Ensemble Model Output Statistics  
86 (EMOS that is a regression method) (Gneiting et al., 2005) and the Bayesian  
87 Model Averaging method (BMA that is an ensemble dressing method) (Raftery  
88 et al., 2005). Both EMOS and BMA provide the entire predictive distribution for  
89 the output variable. In the field of EPS calibration, EMOS technique was used

90 in Taillardat et al. (2016, 2019); Bellier (2018) for calibration of temperature and  
91 precipitation ensemble forecast. In the field of ensemble streamflow calibration,  
92 EMOS was used in Bellier (2018); Hemri et al. (2015), and BMA was used in  
93 Duan et al. (2007).

94 Popular non-parametric methods are: rank histogram recalibration (Hamill  
95 and Colucci, 1997), quantile regression (Bremnes, 2004), individual ensemble-  
96 member adjustments (Van Schaeybroeck and Vannitsem, 2015) or statistical  
97 learning methods, also called machine learning methods (Hastie et al., 2009).  
98 While non-parametric methods require very large training data sets, they are  
99 flexible, data-adaptive, and adapted to non-linearities in the input-output re-  
100 lation (Wilks, 2018). The non-parametric Quantile Regression Forest (QRF)  
101 method proposed by Meinshausen (2006) is a statistical learning approach. QRF  
102 provides an estimation of desired quantiles for the output data, but not the en-  
103 tire distribution, as opposed to EMOS or BMA.

#### 104 *1.4. Scope of the paper*

105 The objective of this paper is the implementation and assessment of an  
106 HEF system for small to medium size catchments taking into account hydro-  
107 logic model parameters' uncertainty. The first part of the study is dedicated to  
108 the analysis and classification of uncertainties in the distributed MORDOR-TS  
109 model (Garçon, 1996; Garavaglia et al., 2017; Rouhier et al., 2017) with GSA  
110 in order to identify the most significant sources of uncertainties to take into  
111 account into the ensemble generation. The GSA is carried out with respect to  
112 uncertainty in hydrologic model parameters and EPS using forecasted precip-  
113 itation provided by Arome Ensemble Prediction System <sup>1</sup> (Seity et al., 2011;  
114 Bouttier et al., 2012; Raynaud and Bouttier, 2016; Bouttier et al., 2016), when  
115 available.

116 The second part of the study is dedicated to the HEF generation and en-

---

<sup>1</sup>This product, denoted by AromeEPS-RR1 in the following, provides a forecast of 1-hour  
rainfall cumul, with a maximum lead-time of 45 hours, updated every day.

117 semble calibration. Different strategies are compared : a model-based approach  
 118 where the raw ensemble comes from ensemble hydrologic simulation is imple-  
 119 mented, then two approaches are implemented with QRF ensemble calibration  
 120 method, with different predictors.

121 Figure 1 summarizes the general process flow of the study. The two main  
 122 objectives of the study correspond to the grey boxes.



Figure 1: General process flow of the study.

123 The paper is organized as follows. Section 2 presents the Odet catchment on  
 124 which the study is carried out, the hydrologic model as well as the atmospheric  
 125 and hydrologic data sets. Methods for GSA and QRF are detailed in section  
 126 3. Section 4 presents the experimental settings for the GSA and the framework  
 127 for the generation and ensemble calibration of the HEF. Section 5.1 discusses  
 128 the results of the GSA for the MORDOR-TS hydrologic model and section 5.2  
 129 presents the comparative skills of the raw and calibrated hydrologic ensembles.  
 130 Conclusion and perspectives are given in Section 6.



131 **2. Material: models and data**

132 *2.1. Modeling the hydrology of the Odet catchment with MORDOR-TS*

133 *2.1.1. The Odet catchment*

134 The Odet river shown in Fig. 2 is a French coastal river located in Western  
135 Brittany. It flows through the city of Quimper, then South to the sea. Its mouth  
136 is located at Plaisance where astronomical tide ranges between 1.40 m and 5.55  
137 m. The Jet and Steir rivers are two tributaries of the Odet river. The Odet  
138 catchment area is 720 km<sup>2</sup> and the Odet river is of about 60 km long. Quimper  
139 is often subject to flooding in urban areas, resulting from the combination of  
140 two phenomena: excessive rainfall that contribute to the rivers, and high tides  
141 increased by storm surges impacting the flow up to Quimper. The distributed  
142 MORDOR-TS rainfall-runoff model is built on three upstream sub-catchments  
143 (Fig. 2, Tab. 1), with outlets at respectively Tréodet (Odet river), Kerjean (Jet  
144 river) and Ty-Planche (Steir river).

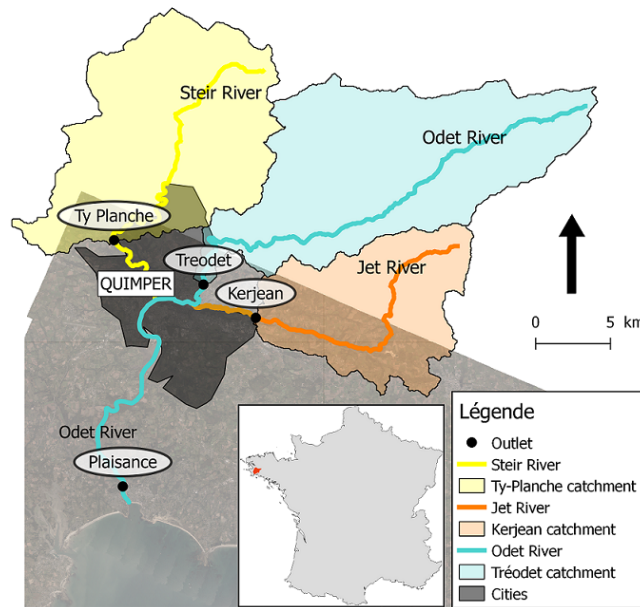


Figure 2: Odet catchment

Table 1: Characteristics of the Tréodet, Kerjean, Ty-Planche sub-catchments

Sub-catchment	Tréodet	Kerjean	Ty-Planche
Elevation of the source	175 m	200 m	100 m
Total river length (km)	37	21	23
Catchment area (km <sup>2</sup> )	205	107	179
Mean flow (m <sup>3</sup> /s)	4.8	2.27	3.79
10-year flow (m <sup>3</sup> /s)	55	19	39
50-year flow (m <sup>3</sup> /s)	75	25	53
Highest flow recorded 12-2000 (m <sup>3</sup> /s)	110	46.6	81
Second highest flow recorded 12-2013 (m <sup>3</sup> /s)	91.5	17.6	42.7
Mean rainfall (mm/yr)	743	672	671

145 *2.1.2. The MORDOR-TS distributed conceptual rainfall-runoff model*

146 The MORDOR-TS model (Garçon, 1996; Garavaglia et al., 2017; Rouhier  
147 et al., 2017) dedicated to water resource management is implemented on each  
148 sub-catchment of the Odet catchment. MORDOR-TS is a spatialized and con-  
149 tinuous conceptual rainfall-runoff hydrologic model that connects the mesh cells  
150 according to the hydrographic network. At each time step, the production is  
151 calculated for each cell and then routed to simulation points on the mesh. The  
152 structure of the production module is presented in Fig. 3 and MORDOR-TS's  
153 hydrologic parameters are described in Tab. 2. The production module takes  
154 spatially distributed precipitation (P) and temperature (T) as input data and  
155 adjusts the water balance through two coefficients  $c_p$  and  $k_{min}$ . This latter pa-  
156 rameter is involved in the calculation of the actual evapotranspiration  $AET$ .  
157 The production module is then composed of six conceptual reservoirs; two for  
158 ice and snow (not active here), and four others: a superficial reservoir U of ca-  
159 pacity  $U_{max}$ , an intermediate reservoir L of capacity  $L_{max}$  of which filling level  
160 is driven by the parameter  $ev_L$ , an evaporating reservoir Z of capacity  $Z_{max}$  and  
161 a deep reservoir N of which filling level is driven by the parameter  $lk_N$ . Three  
162 fluxes components are transferred from the production module to the routing  
163 module: area runoff  $Q_s$ , subarea runoff  $Q_v$  and base runoff  $Q_b$ . The parameter

164  $k_r$  determines the ratio of the water feeding reservoir N and the subarea runoff.  
 165 The routing module propagates the water production of each cell into the hy-  
 166 drographic network. The transfer function is based on the 1D diffusive wave  
 167 model, with celerity  $Cel$  and diffusion  $Diff$  coefficients independent from the  
 168 runoff (Hayami, 1951).

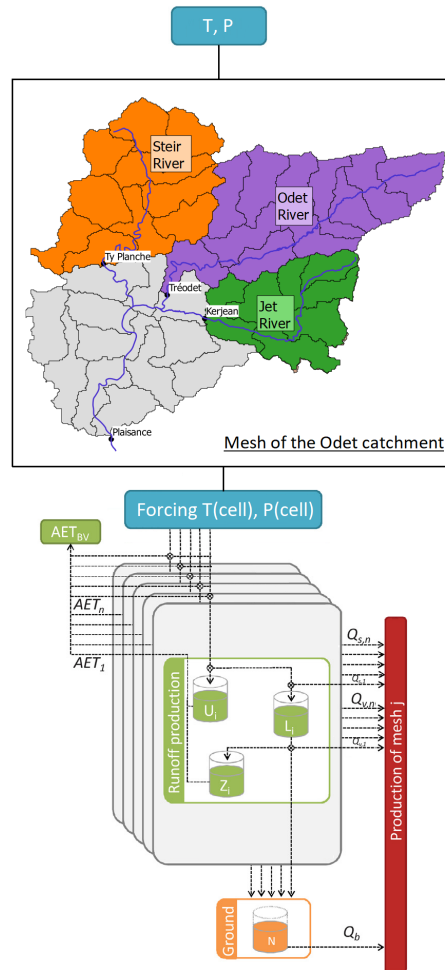


Figure 3: Structure of the MORDOR-TS model over the Odet catchment. Adapted from Rouhier et al. (2017)

Table 2: Description of the parameters of MORDOR-TS hydrologic model

Parameter	Unit	Description	Module
$c_p$	-	Precipitation multiplicative correction factor	Water balance
$k_{min}$	-	Maximum seasonal crop coefficient	
$U_{max}$	mm	Maximum capacity of the root zone U	Runoff production
$L_{max}$	mm	Maximum capacity of the hillslope zone L	
$ev_L$	-	Outflow exponent of storage L	
$Z_{max}$	mm	Maximum capacity of the capillarity storage Z	
$k_r$	-	Runoff coefficient	
$lk_N$	mm.h <sup>-1</sup>	Outflow coefficient of storage N	
$Cel$	m.s <sup>-1</sup>	Wave celerity	Routing module
$Diff$	m <sup>2</sup> .s <sup>-1</sup>	Wave diffusion	

169 The hydrologic calibration <sup>2</sup> of the 10 previously described parameters of  
170 MORDOR-TS (Tab. 2) is achieved after a one-year spinup, using Banque Hy-  
171 dro streamflow observations, ANTILOPE rainfall and SAFRAN surface tem-  
172 perature forcing described in Sect. 2.2. The hydrologic calibration is carried out  
173 with respect to a multi-objective function using the caRamel genetic algorithm  
174 (Le Moine et al., 2015; Monteil et al., 2019). The multi-objective function gath-  
175 ers three scores : (i) Nash-Sutcliffe Efficiency (NSE, Nash and Sutcliffe (1970))  
176 over the entire time serie, (ii) NSE over the inter-annual daily regime and (iii)  
177 NSE over the empirical cumulative distribution.

178 The hydrologic calibration is first achieved over a 6-year period (01/01/2007-  
179 12/31/2013) and validated over the period 01/01/2014-05/31/2017. Over the  
180 validation period, scores values are of the same order of magnitude as over the  
181 calibration period, which validates the hydrologic calibration. In the following,  
182 the deterministic simulation with calibrated hydrologic parameters is referred to

---

<sup>2</sup>In this paper, the term *calibration* refers either to the calibration of the parameters of the hydrologic model MORDOR-TS or to the statistical ensemble calibration with the Quantile Regression Forest (QRf) method. To avoid confusion, the calibration of the hydrologic parameters is always denoted as *hydrologic calibration*. The statistical calibration with QRF is denoted as *ensemble calibration* when necessary.

183 as the *reference simulation*. In a second time, a set of hydrologic calibrations is  
184 achieved over 2-year periods from 01/01/2007 to 05/31/2017 in order to estimate  
185 the uncertainty in each parameter, required to perform the GSA and to generate  
186 the raw HEF (methodology fully described in Sect. 4.2.1).

## 187 2.2. Data sets

188 Depending on their availability, different data sets are used in the study.

### 189 2.2.1. Observed data

190 Spatially distributed observed rainfall, temperature data and flow discharge  
191 are available at hourly time step from January 2006 to May 2017 on the Odet  
192 catchment.

- 193 • Rain data: The ANTILOPE rainfall product is a combination of radar and  
194 gauge rainfall data from Météo-France (Champeaux et al., 2009). The grid  
195 resolution is 1 km.
- 196 • Temperature data: Surface temperature is extracted from SAFRAN re-  
197 analysis (Vidal et al., 2010). The grid resolution is 8 km.
- 198 • Discharge data: River flow discharge time series at Tréodet, Kerjean and  
199 Ty-Planche are extracted from the French national archive (Banque hydro,  
200 <http://www.hydro.eaufrance.fr>, Leleu, Isabelle et al. (2014)).

### 201 2.2.2. Forecasted data

202 AromeEPS-RR1 ensemble rain forecast product is used to describe uncer-  
203 tainty in rainfall for the GSA. As this product is available over a limited period  
204 only, the deterministic Expert-RR3 rain forecast product is used by default, in  
205 particular for the generation and the ensemble-calibration of the HEF.

206 *AromeEPS-RR1*. AromeEPS-RR1 uses the regional atmospheric model AROME  
207 described in Seity et al. (2011) with ensemble perturbations documented in  
208 Bouttier et al. (2012); Raynaud and Bouttier (2016); Bouttier et al. (2016),  
209 accounting for uncertainties in initial conditions, boundary conditions, surface

210 conditions and the model physics. Its large scale boundary conditions are pro-  
211 vided by the global PEARP ensemble (Descamps et al., 2015). AromeEPS-RR1  
212 ensemble is composed of 12 equiprobable members and covers a 1800x1700km  
213 <sup>2</sup> Western European domain that encompasses the Odet catchment with a hor-  
214 izontal grid at 2.5 km resolution. AromeEPS-RR1 is operational since the end  
215 of 2016. A limited period of 112 days in early 2016 was made available *a poste-*  
216 *riori* for this study in order to cover former rain events on the Odet catchment.  
217 It provides a daily forecast at 21:00 UTC, with hourly output over a 45-hour  
218 forecast range.

219 *Expert-RR3*. Expert-RR3<sup>3</sup> is a 3-hours deterministic rainfall accumulation fore-  
220 cast specified by human experts on the basis of numerical forecasts from atmo-  
221 spheric models, with a 72 hours forecast range. These data are used opera-  
222 tionally for flood forecasting in the French governmental services and available  
223 over 2011-2014 for this study.

---

<sup>3</sup>RR3 provide a forecast of 3-hour rainfall cumul, with a maximum lead-time of 72 hours,  
updated every 15 minutes

224 **3. Methods**

225 *3.1. Global Sensitivity Analysis (GSA)*

In the present study, the GSA stands in the computation of Sobol' indices (Sobol, 2001), with the assumption that the input aleatory variables are independant. Sobol' indices apportion the variance of the output  $Y = f(X)$  with  $X = (X_1, X_2, \dots, X_k)$ , to the variation of different inputs  $X_1, \dots, X_k$  on their uncertainty domain. With the assumption that the variance of  $Y$  is finite and the input variables are independant, the Hoeffding decomposition (Hoeffding, 1948) leads to the definition of the Sobol'indices :

$$1 = \sum_i S_i + \sum_i \sum_{j>i} S_{i,j} + \dots + S_{1,2,3,\dots,k} \quad (1)$$

226 where

227 •  $S_i = \frac{V_i}{V(Y)}$  is the first order Sobol index of  $X_i$  and represents the normal-  
 228 ized elementary contribution of  $X_i$  to  $V(Y)$ ,

229 •  $S_{i,j} = \frac{V_{i,j}}{V(Y)}$  is the second order Sobol index of  $X_i$  and  $X_j$  and represent  
 230 the normalized contribution due to interactions between  $X_i$  et  $X_j$  to  $V(Y)$ ,

231 and so on.

The total Sobol' index  $S_{T_i}$ , gathering all contributions related to  $X_i$  is then defined as

$$S_{T_i} = S_i + \sum_{j \neq i} S_{i,j} + \sum_{j \neq i, k \neq i, j < k} S_{i,j,k} + \dots + S_{1,2,3,\dots,k} = \sum_{l \in \#i} S_l \quad (2)$$

232 where  $\#i$  are all the subsets of  $\{1, \dots, k\}$  including  $i$ .

233 Sobol' indices thus measure the influence of the different independant im-  
 234 puts  $X_1, X_2, \dots, X_k$  and their interactions on the output  $Y = f(X)$ . The more  
 235 sensitive the model response is to an input parameter, the larger its associated  
 236 Sobol index. Usually only first and total Sobol indices are computed. The first  
 237 order indices are useful for factor prioritization (FP) and provides the parame-  
 238 ter(s) that most significantly control(s) the most the output variance, whereas

239 the total order indices are helpful in factor fixing (FF), to determine which pa-  
240 rameters can be fixed without consequences on the output (Saltelli, 2008). In  
241 practice, the estimation of Sobol' indices is generally achieved with a stochastic  
242 estimation using an ensemble of model output realizations (Saltelli and Annoni,  
243 2010). The computation of Sobol' indices is here achieved using the python  
244 modules OpenTURNS (<http://openturns.org/>) and OT-Batman (T.Roy et al.,  
245 2018).

### 246 *3.2. Quantile Regression Forest (QRF)*

247 *Non parametric regression with QRF method.* The Quantile Regression For-  
248 est (QRF) method (Meinshausen (2006)) technique is detailed in Zamo et al.  
249 (2014); Taillardat et al. (2019). The principle of the method is the aggregation  
250 of meteorological or hydrological situations according to their forecasts, with the  
251 assumption that close forecasts predictors lead to close observations. In that  
252 way, this method can be linked to the analog method (Hamill and Whitaker,  
253 2006; Zalachori et al., 2012; Delle Monache et al., 2013). QRF is a non-parametric  
254 and non-linear regression, which consists in building random forests from binary  
255 decision trees given a set of predictors (Breiman, 2001). Contrary to random  
256 forests that approximate the conditional mean, QRF estimates the full condi-  
257 tional distribution of the response variable.

258 *Choosing the predictors.* A wide range of predictors are available for the QRF  
259 method. They can obviously stem from the ensemble forecast : statistics (en-  
260 semble mean, variance or percentiles) of the variable to calibrate or other vari-  
261 ables within the ensemble, as well as other characteristics of the forecast, such  
262 as the time or day (providing a modality instead of a value). Predictors can  
263 also be described from real observation of the forecasted variable. The choice  
264 of the predictors represents a key element in the implementation of the QRF  
265 calibration method. In the following, a learning sample is formed by choosing a  
266 set of predictors and picking the associated observations.



267 *Building a binary decision tree.* A binary decision tree is built by iteratively  
 268 splitting the learning sample into two groups. For a quantitative predictor, the  
 269 split is done according to a threshold, while for a qualitative predictor, the split  
 270 is done according to the modality. The predictor and the splitting criteria are  
 271 chosen to minimize the variability of the associated observations in the resulting  
 272 two groups. Each resulting group is then itself split in two following the same  
 273 algorithm, until a stopping criterion is reached (for example a minimum number  
 274 of data in the sub-groups, or an insufficient decrease of variance). Each final  
 275 group is called leaf and contains a set of observations, also called predictand.  
 276 The splitting algorithm is illustrated in Fig. 4 with two predictors p1 and p2  
 277 ranging between 0 to 1. In this example, the tree has three leaves with associated  
 278 observations (rain or discharge for our purpose).

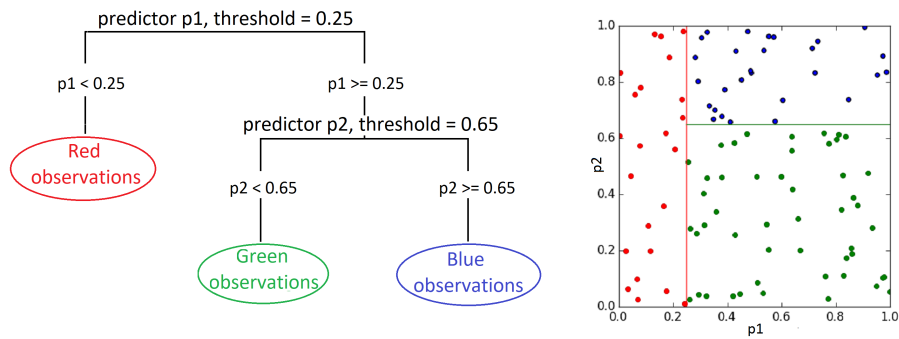


Figure 4: Illustration of binary decision tree with 2 predictors p1 and p2.

279 *Building a forest.* Breiman (1996) proposes to improve the robustness of the  
 280 prediction issued from a decision tree by selecting different learning samples  
 281 to build several trees and form a forest. Since this approach would require a  
 282 large amount of data, usually not available in practice, bootstrap sampling is  
 283 often used. Randomly choosing a subset of predictors for each split of each tree,  
 284 as suggested by Breiman (2001), enhances the independence of the trees and  
 285 consequently forms a *random* forest.

286 *How to use QRF to calibrate a new ensemble forecast ?*. In operational mode,  
287 when a new ensemble forecast is available for calibration, its associated pre-  
288 dictors are computed and run through each tree of the previously constructed  
289 forest. The terminal leaf predictand values are used to estimate the predictive  
290 CDF and quantiles of the calibrated forecast.

291 *Merits and limitation of the QRF method*. It should be noted that no a priori  
292 assumption is made on the distribution of the variable to calibrate neither before  
293 nor after calibration. Also, since the terminal leaves are composed of observa-  
294 tions, the ensemble calibration is bound to output physically consistent values  
295 (for example, no negative amount of precipitation). The associated drawback is  
296 that by construction, the predicted CDF is unable to predict values outside of  
297 the observation range within the learning sample. This may be limiting when  
298 dealing with extreme values but may be overcome when working with anomalies  
299 or fitting a parametric function to the CDF (Taillardat et al., 2019). It should  
300 finally be noted that one of the main drawback of QRF method stands in the  
301 need of a large data set.

302 *Reconstruction of the calibrated members*. In the following, the Ensemble Cop-  
303 ula Coupling (ECC, Schefzik et al. (2013)) method is used to reorder the post-  
304 processed quantiles to recreate time-series, but this method yields unrealistic  
305 jumps when applied to HEF streamflows. The Trajectory Smoothing (TS) pro-  
306 cedure proposed by Bellier (2018) was thus applied to preserve temporal cor-  
307 relation consistency, in the perspective of using the ensemble of reconstructed  
308 streamflow time series as forcing to hydraulic ensemble simulations. It should be  
309 noted that, as opposed to Ensemble Copula Coupling, the Trajectory Smoothing  
310 procedure modifies the post-processed ensemble.

*Ensemble evaluation metrics*. Various tools are available to evaluate proba-  
bilistic forecasts and are well described in the literature. In this study, two  
widely used verification measures for ensemble forecast are used. The Contin-  
uous Ranked Probability Score (Matheson and Winkler (1976), Hersbach (2000),

Gneiting and Raftery (2007)) assesses reliability and resolution simultaneously. It is negatively oriented: the lower the better. The Rank Histogram (Talagrand et al. (1997), Hamill and Colucci (1997), Anderson (1996)) is useful to assess reliability only. A reliable ensemble implies that each rank is filled with the same probability, so the rank histogram is flat. Candille and Talagrand (2005), then Delle Monache et al. (2006) and Taillardat et al. (2016) introduced the notion of norms of a rank histogram with  $K+1$  ranks. In particular,  $\|\cdot\|_2$  can be defined as :

$$\|\cdot\|_2 = \sqrt{\sum_{i=1}^{K+1} (f_i - \frac{1}{K+1})^2} \quad (3)$$

311 where  $f_i$  represents the frequency of observations in the  $i$ th rank. For a perfectly  
 312 reliable ensemble system,  $\|\cdot\|_2$  is equal to zero.

313 **4. Experimental settings**

314 *4.1. General ensemble workflow*

315 The first part of the study consists in a Global Sensitivity Analysis (GSA)  
 316 applied to the distributed MORDOR-TS model, taking into account (i) un-  
 317 certainties in rainfall and hydrologic model parameters (GSA-Arome), or (ii)  
 318 only uncertainty in hydrologic model parameters (GSA-hydro). This last GSA  
 319 without uncertainty in rainfall aims at assessing the consistency of the results  
 320 when a deterministic rainfall (ANTILOPE) is taken into account in place of an  
 321 EPS (AromeEPS-RR1). The implementations of the two GSAs are described in  
 322 Sect. 4.2. It should be noted that the use of AromeEPS-RR1 for GSA-Arome is  
 323 preceded by a QRF ensemble-calibration step. These two GSAs studies validate  
 324 the methodology used for HEF generation in the second part of the study: hy-  
 325 pothesis on the choice of uncertain variables, associated statistical distributions  
 326 and related hyper parameters.

327 The second part of the study is dedicated to the HEF generation and cal-  
 328 ibration over a three-year (2011-2014) period over which there was significant  
 329 hydrological events. Only uncertainty in hydrologic parameters is considered  
 330 since the AromeEPS-RR1 product is not available over this period. The raw  
 331 HEF ensemble is then calibrated with the QRF method.

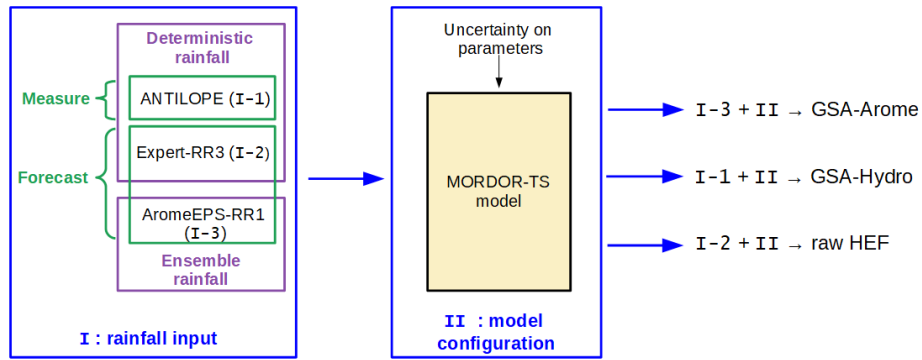


Figure 5: Configurations for GSA-Arome, GSA-hydro and raw HEF

332 Fig. 5 presents the different model settings for the three hydrologic raw  
333 ensembles built in the study: GSA-Arome, GSA-hydro and raw HEF. Tab 3  
334 presents the required data needed for each step of the study :

- 335 • The GSA-Arome requires the integration of an ensemble of hydrology  
336 simulations over  $P_{GSA-Arome}$  from 12/24/2015 to 03/15/2016. The atmo-  
337 spheric forcing is provided by AromeEPS-RR1 ensemble calibrated with  
338 QRF against ANTILOPE rainfall observed data.
- 339 • The GSA-hydro is carried out over two periods. It requires the integration  
340 of an ensemble of hydrology simulations over the period  $P_{GSA-hydro}$  that  
341 is either  $P_{GSA-Arome}$  or a sub period of  $P_{HEF}$  (12/23/2013-12/26/2013).  
342 The deterministic atmospheric forcing is provided by ANTILOPE for both  
343 periods.
- 344 • The raw HEF generation is achieved using atmopsheric forcing from the  
345 Météo-France Expert-RR3 deterministic forecast (as AromeEPS-RR1 is  
346 not available) over  $P_{HEF}$  (09/01/2011-06/01/2014). HEF ensemble stream-  
347 flows are calibrated with QRF against Banque Hydro streamflow obser-  
348 vations, using ANTILOPE observations, Expert-RR3 rain forecast and  
349 Banque Hydro streamflow observations as predictors.

350 In both GSAs and HEF, hydrology simulations start from a spin-up forced by  
351 ANTILOPE and SAFRAN observed data.

#### 352 4.2. GSAs for MORDOR-TS model

353 The GSA-Arome and GSA-hydro for streamflow on the Odet catchment  
354 are carried out with respect to the forecasted runoff at the outlet of each of  
355 the subcatchments Tréodet, Kerjean and Ty-Planche. Both GSA take into  
356 account uncertainties that relate to a set of 10 parameters for MORDOR-TS  
357 (Sect. 4.2.1). GSA-Arome (Sect. 4.2.2) is performed for each lead-time with a  
358 cycled procedure and takes also into account uncertainties that relate to rain  
359 forcing, considering an integer that represents the index within the 12-member

Table 3: Data for the GSA and QRF calibration of rainfall and streamflow

Type	Name	QRF			QRF HEF
		AromeEPS-RR1	GSA-Arome	GSA-hydro	
	Period	$P_{GSA-Arome}$	$P_{GSA-Arome}$	$P_{GSA-hydro}$	$P_{HEF}$
Observed data	rainfall	ANTILOPE	X	X	X
	temperature	SAFRAN		X	X
	streamflow	BanqueHydro			X
Forecasted data	rainfall	AromeEPS-RR1	X	X	
	rainfall	Expert-RR3			X

360 calibrated AromeEPS-RR1 ensemble, drawn from a uniform distribution  $U_{[1,12]}$ .  
361 In GSA-hydro (Sect. 4.2.3), ANTILOPE deterministic rain is used to force all  
362 model runs within the ensemble while hydrologic parameters vary.

#### 363 4.2.1. Hydrologic parameters' distributions

364 The 10 MORDOR-TS parameters are drawn from uniform distributions with  
365  $V_{min}$  and  $V_{max}$  extreme values, shown in Tab 4 and determined from a set of  
366 2-year of hydrologic calibrations. The lower bound (respectively upper bound)  
367 of the uniform distribution is chosen as the minimum (respectively maximum)  
368 of the values reached in the different hydrologic calibrations (Sect. 2.1.2).

Table 4: Extreme values for MORDOR-TS model parameters' uniform PDFs

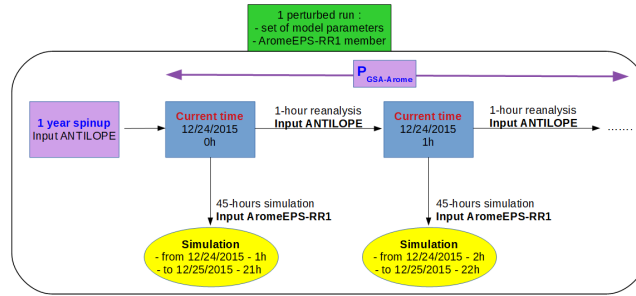
Paramètre	Tréodet	Tréodet	Kerjean	Kerjean	Ty-Planche	Ty-Planche
	min	max	min	max	min	max
$c_p$ (-)	0.98	1.21	0.97	1.13	0.99	1.18
$k_{min}$ (-)	0.18	1.11	0.14	1.03	0.32	0.79
$U_{max}$ (mm)	75	196	30	111	39	116
$L_{max}$ (mm)	64	153	30	493	74	298
$ev_L$ (-)	3.95	4.00	1.50	3.98	3.38	4.00
$Z_{max}$ (mm)	47	138	77	256	72	482
$k_r$ (-)	0.10	0.28	0.10	0.41	0.10	0.30
$lk_N$ (mm.h <sup>-1</sup> )	-5.8	-5.2	-6.2	-5.5	-5.7	-5.4
$Cel$ (km.h <sup>-1</sup> )	0.44	0.56	0.45	1.10	0.50	0.55
$Diff$ (km <sup>2</sup> .h <sup>-1</sup> )	1	659	1	5000	329	731

369 *4.2.2. GSA-Arome implementation*

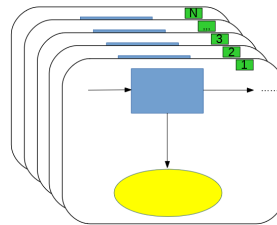
370 *Ensemble calibration of AromeEPS-RR1.* AromeEPS-RR1 ensemble is calibrated  
371 with the QRF method before its use in GSA-Arome. In order to increase the vol-  
372 ume of data available for AromeEPS-RR1 ensemble calibration over the limited  
373 period  $P_{GSA-Arome}$  in 2016, the calibration is carried out without discrimi-  
374 nating the lead-times. The ensemble gathers all AROME cells within a sub-  
375 catchment (63, 36 and 51 for Tréodet, Kerjean and Ty-Planche respectively);  
376 thus forming a new ensemble with (12\*number of cells) members. The size of  
377 the learning sample is 45 (lead-times)  $\times$  112 (days). The ensemble is calibrated  
378 against the ANTILOPE observations, averaged over the whole subcatchment.  
379 For each subcatchment, the predictors are the percentiles 10, 50 and 90 of the  
380 raw AromeEPS-RR1 hourly rainfall, surface humidity, surface temperature and  
381 the moment of the day of the lead-time (morning, afternoon, evening, night).  
382 The evaluation of the calibration is achieved with a leave-one out method: each  
383 element of the training sample is alternatively used for validation, while the  
384 other are used for learning. This AromeEPS-RR1 calibration strategy relies  
385 on the fact that the error in rainfall intensity is homogeneous over a small  
386 catchment, and that it only depends on the moment of the day. The loss of  
387 predictability as the lead time increases is assumed to be negligible. It is also  
388 assumed that the rainfall intensity data are non correlated over time. This  
389 assumption would not be valid for temperature or streamflow. For rainfall,  
390 a 6h correlation is suspected and could be taken into account to improve the  
391 robustness of the calibration, especially if used to calibrate new data.

392 *GSA-Arome.* The GSA-Arome study is carried out in an operational framework  
393 with hourly updated AromeEPS-RR1 forecast. For that purpose, the ensemble  
394 is built as shown in Fig. 6(a) over a spin-up, a re-analysis and a forecast period.  
395 Each ensemble member is associated to a realisation of the set of hydrologic  
396 parameters and of AromeEPS-RR1. Each hydrologic simulation is run over a  
397 1-year spin-up and a 1-hour re-analysis with ANTILOPE rainfall, then run over  
398 a 45-hour forecast AromeEPS-RR1 rainfall. This is cycled hourly as the rain

399 product is updated hourly. The GSA-Arome indices are computed hourly for  
 400 the 45 lead times over the  $N$  cycled simulations (Fig. 6(b)). Here, to ensure the  
 401 convergence of the Sobol' indices estimation, the GSA-Arome study is achieved  
 402 with an ensemble of 192000 MORDOR-TS simulations.



(a) 1 cycled run



(b)  $N$  cycled runs

Figure 6: Cycled runs for GSA-Arome

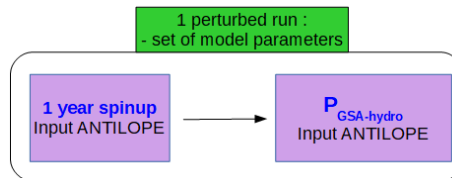


Figure 7: Continuous runs for GSA-hydro



403 *4.2.3. GSA-hydro implementation*

404 The GSA-hydro study uses the observed rainfall ANTILOPE as input for  
405 every member, thus the notion of lead-time is no longer relevant. Each member  
406 is initiated with a 1-year spin-up, then integrated over  $P_{GSA-hydro}$  using its  
407 associated set of hydrologic parameters as illustrated in Fig. 7. To ensure the  
408 convergence of the Sobol' indices estimation, the GSA-hydro Sobol indices are  
409 computed hourly over 154000 MORDOR-TS simulations.

410 *4.3. Generation and ensemble-calibration of the HEF*

411 It was not possible to apply the QRF calibration on the ensemble of stream-  
412 flows from the MORDOR-TS simulations that were used for the GSA-Arome  
413 over  $P_{GSA-Arome}$ . Indeed, this period is too short, the learning sample for the  
414 QRF method is thus too small, especially given that it should be further re-  
415 duced to account for temporal correlation of streamflow. Another ensemble was  
416 thus generated, over a longer period  $P_{HEF}$  (2011-2014) over which significant  
417 hydrological events occurred, but without considering uncertainty related to the  
418 rainfall since the AromeEPS-RR1 product was not available over this period.  
419 Expert-RR3 deterministic forecast rain was used to generate the raw HEF. The  
420 HEF generation and calibration is achieved with a hydrologic ensemble of 99  
421 members. The raw HEF is generated in an operational framework with hourly  
422 updated RR3 forecast. For that purpose, the ensemble is built as shown in  
423 Fig. 8(a), similarly to that for GSA-Arome over a spin-up, a re-analysis and a  
424 forecast period, except that the deterministic forecast Expert-RR3 is used in  
425 place of AromeEPS-RR1. This is cycled hourly as the RR3 forecast rain prod-  
426 uct is updated several times per hour. Only the predominant parameters that  
427 were previously identified by the GSA are taken into account for the ensemble  
428 generation of raw HEF. The forecasts are calibrated and evaluated against ob-  
429 servations over 27 months, from October 2011 to June 2014, excluding summer  
430 months (July, August and September). A cross-validation method is used: each  
431 month of the calibration period is alternatively used for validation, while the 26  
432 other months are used for learning. It should be noted that data over a 10-day

433 period before and after the validation month was removed from the learning  
 434 sample to avoid auto-correlation between learning and validation data.

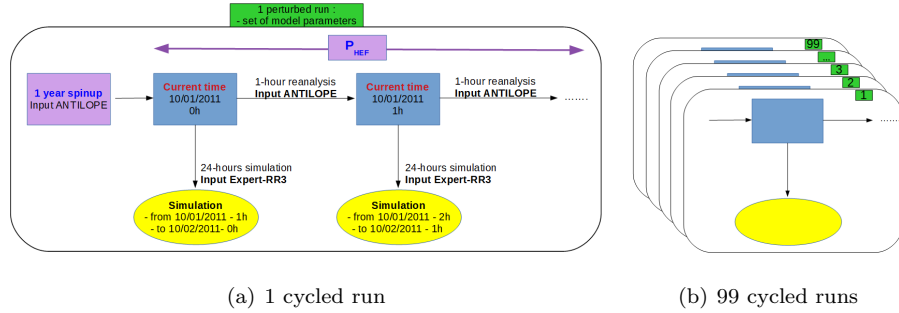


Figure 8: Cycled runs for HEF

435 *Ensemble calibration strategies.* Different strategies to generate hydrologic en-  
 436 sembles are implemented as shown in Fig. 9. The model-based-only approach  
 437 relies on the integration of MORDOR-TS forced by observed ANTILOPE and  
 438 forecast RR3 rain product with perturbed hydrologic parameters; this leads to  
 439 the raw ensemble denoted as raw HEF. Then, two approaches are implemented  
 440 with QRF calibration method. QRF aims at constructing a relation between  
 441 chosen predictors and corresponding observations, which further allows to es-  
 442 timate desired quantiles of the observed quantity. All available predictors for  
 443 QRF are presented in Tab 5. In the observation-based approach, predictors for  
 444 the QRF method use no information from the hydrologic simulations. Strictly  
 445 speaking, this is not a calibration strategy as the raw ensemble is not used;  
 446 QRF is used to generate calibrated quantiles of an observed variable. This  
 447 leads to the ensemble denoted as QRF-nohydro quantiles. In the combined  
 448 model-observation approach, predictors for QRF calibration uses the mean and  
 449 standard deviation of the raw ensemble. The resulting ensemble is denoted as  
 450 QRF-hydro quantiles. After applying the Ensemble Copula Coupling (ECC)  
 451 method to reorder the calibrated QRF-hydro quantiles, the Trajectory Smooth-  
 452 ing (TS) procedure was finally applied, leading to QRF-hydro-TS ensemble.

453 The impact of smoothing the reconstructed hydrologic time-series from QRF  
 454 quantiles is here investigated in the prospect of using the hydrologic ensemble  
 455 as input for hydraulic simulation. It should be noted that the ECC method can  
 456 not be applied on QRF-nohydro quantiles, since it requires the availability of  
 457 the raw ensemble as dependence template.

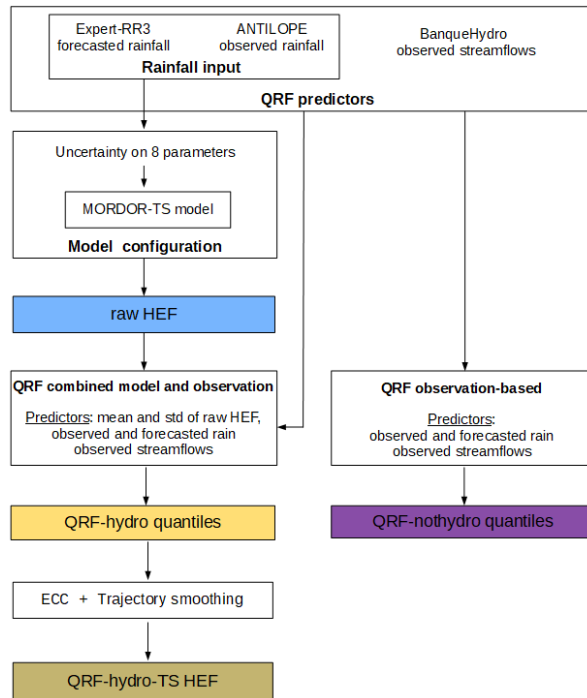


Figure 9: Strategies for Hydrologic Ensemble Forecast generation. Data used for hydrologic model integration as well as QRF predictors are shown. The resulting four HEFs or ensemble of quantiles are indicated in colored boxes.

458 *Computation of CRPS and RH for HEF assessment.* HEF assessment is first  
 459 achieved with univariate criteria with CRPS and RH computed for streamflow  
 460 over each catchment and for each lead-time. As suggested by Bellier (2018),  
 461 cumulated and maximum quantities are also considered, and such quantities  
 462 are denoted as global variables in the following. CRPS and RH are thus com-  
 463 puted for the cumulated and maximum streamflow that respectively relate to

Table 5: Choice of predictors for QRF-hydro and QRF-nohydro HEF calibration strategies.

Name	Unit	Description	QRF-hydro	QRF-nohydro
<i>Mean</i>	m <sup>3</sup> /s	mean of raw ensemble streamflows	X	
<i>Sigma</i>	m <sup>3</sup> /s	standard deviation of raw ensemble streamflows	X	
<i>Month</i>		month of the validation time	X	X
<i>Period</i>		period of validation time (0 am - 6 am, 6 am - 12 am, 0 pm - 6 pm, 6 pm - 12 pm)	X	X
<i>Q<sub>0</sub></i>	m <sup>3</sup> /s	measured streamflow at the current time	X	X
<i>GradQ<sub>0</sub></i>	m <sup>3</sup>	gradient of the measured streamflow at the current time	X	X
<i>M<sub>rain</sub></i>	mm	measured rain over the catchment during the $N_b$ past hours of re-analysis ( $N_b = 24$ hours - lead-time)	X	X
<i>F<sub>rain</sub></i>	mm	forecasted rain over the catchment between the current time and the lead-time	X	X

464 volume and peak flow (Hemri et al., 2015). Following conclusions from Bel-  
465 lier et al. (2017) and Gneiting and Ranjan (2011), the HEF is assessed with  
466 a forecast-based stratification for CRPS and RH computation, focusing on the  
467 (*forecast, observation*) pairs for which the maximum of the ensemble is be-  
468 yond the 90<sup>th</sup> percentile of the associated observation (hourly, cumulated or  
469 maximum measured streamflow), computed for heavy rainfall.

470 **5. Results**

471 *5.1. GSAs for MORDOR-TS model*

472 *5.1.1. Calibration of AromeEPS-RR1*

473 As shown in Fig. 10, ensemble calibration improves AromeEPS-RR1 with  
 474 smaller mean CRPS value and flatter rank histograms than that of the raw  
 475 ensemble (0.119 and 0.1351 respectively for the CRPS). The  $\|\cdot\|_2$  norm for the  
 476 QRF rank histogram ( $1.11 \times 10^{-2}$ ) is smaller than that of the raw ensemble  
 477 ( $3.58 \times 10^{-2}$ ). The calibrated AromeEPS-RR1 data are thus more consistent  
 478 with observations than before calibration, and thus can be used as input of  
 479 GSA-Arome.

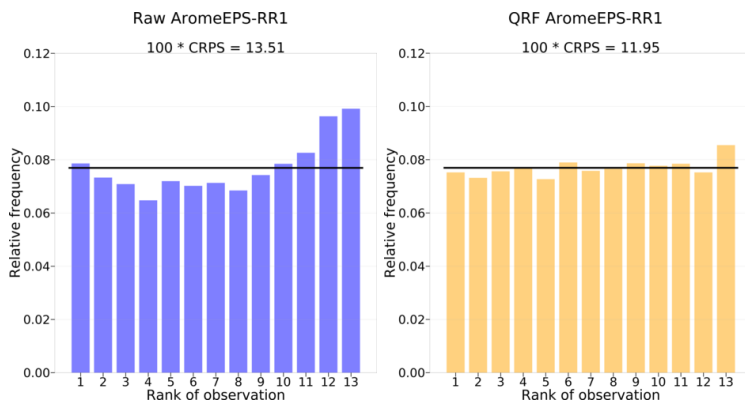
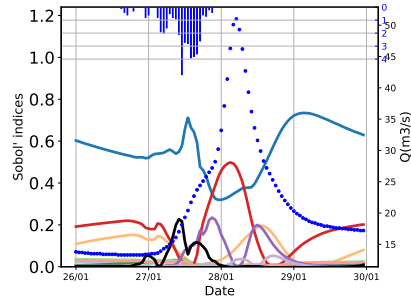


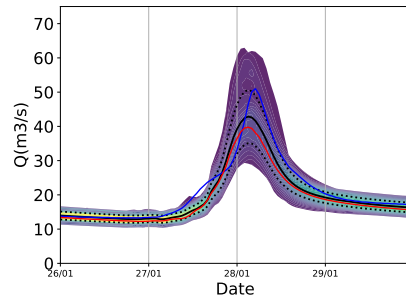
Figure 10: Rank histogram and CRPS for the raw and calibrated AromeEPS-RR1 ensemble.

480 *5.1.2. GSA-Arome Results*

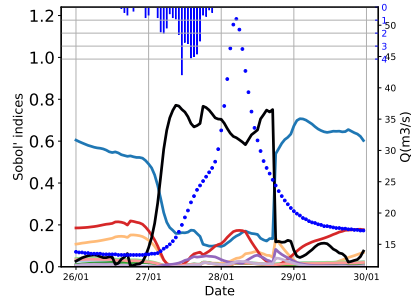
481 Results for the hydrologic GSA-Arome on forecasted streamflow are shown  
 482 here over the subperiod 01/26/2016-01/30/2016 of  $P_{GSA-Arome}$  for 6-hour and  
 483 21-hour lead-times. Time-varying Sobol' indices and streamflow hydrographs  
 484 at Tréodet are shown in Fig. 11. A vertical section of Fig. 11(b) and Fig. 11(d)  
 485 displays the probability density function at the given date. Similar results for  
 486 Kerjean and Ty-Planche can be found as supplementary material in Section 7  
 487 (Fig. 17 and 18).



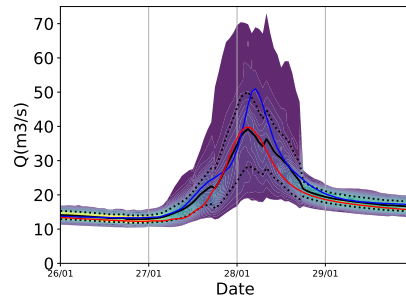
(a) Sobol' indices - 6h lead-time



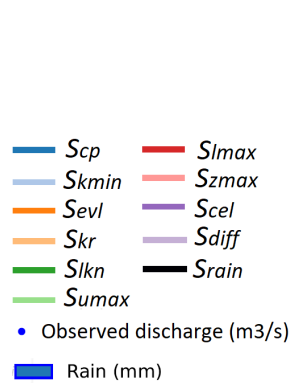
(b) Streamflow hydrographs - 6h lead-time



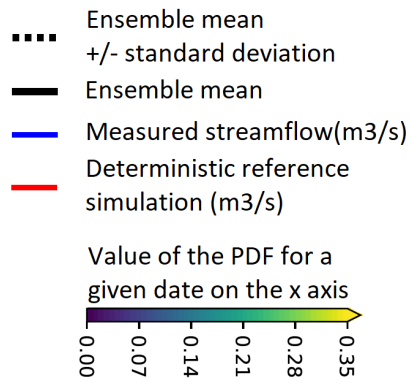
(c) Sobol' indices - 21h lead-time



(d) Streamflow hydrographs - 21h lead-time



(e) Legend for MORDOR-TS Sobol' indices



(f) Legend for streamflow graphs

Figure 11: First order Sobol' indices and associated hydrographs for the event between 01/26/2016 and 01/30/2016 at Tréodet catchment for GSA-Arome. A vertical section of Fig(b) and Fig(d) displays the probability density function at the given date, according to the legend shown in (f). The results are shown for 6-hour ((a) and (b)) and 21-hour ((c) and (d)) lead-times.

488 *Sobol' indices.* The Sobol' indices time-series for GSA-Arome are presented in  
 489 Fig. 11(a) and 11(c) for 6-hour and 21-hour lead times with time on the x-  
 490 axis. The observed discharge is represented by the blue dotted curve (right  
 491 y-axis) and the observed rainfall histogram is represented at the top of the  
 492 panel. It shows that, for both lead times, the simulated streamflow is dominated  
 493 by the precipitation correcting factor  $c_p$  (dark blue curve). This predominance  
 494 decreases in favor of the choice of the AromeEPS-RR1 rain scenario when heavy  
 495 precipitations occur, especially for longer lead-times, since this choice strongly  
 496 influences the total amount of water in the MORDOR-TS reservoirs. Streamflow  
 497 is thus mostly dominated by  $c_p$  and rain that both control the water balance  
 498 of the model. Rainfall mostly matters for lead-times that are greater than the  
 499 concentration time of the catchment. This implies that the construction of  
 500 HEF for short lead times can be achieved only taking into account uncertainties  
 501 that relate to hydrology. To a lesser extent, at the beginning of the event, the  
 502 streamflow is influenced by the capacity of the reservoir  $L_{max}$  (red curve) that  
 503 directly feeds the runoff. The importance of  $L_{max}$  decreases before the end of  
 504 the event; indeed, when the reservoir is full, the overflow directly feeds the runoff  
 505 and its capacity no longer has impact on the simulated streamflow. For 6-hour  
 506 lead-time, the wave celerity  $Cel$  (dark purple curve) has noticeable importance  
 507 when the streamflow gradient is strong (both in increase and decrease phases).  
 508 The runoff coefficient  $k_r$  (light orange curve) has noticeable importance at the  
 509 peak of the event (and beyond). Indeed, when reservoir  $Z$  is full, part of the  
 510 excess amount of water  $k_r \times out_Z$  directly feeds the runoff. Finally, simulated  
 511 streamflow is not sensitive to the two parameters  $k_{min}$  and  $Diff$  over the three  
 512 catchments.

513 *Temporal PDF.* The streamflow probability density functions represented over  
 514 time for GSA-Arome are displayed in 11(b) and 11(d) for 6-hour and 21-hour  
 515 lead times respectively, with time on the x-axis. The observed discharge is plot-  
 516 ted in blue, the deterministic reference simulation with parameters issued from  
 517 the hydrologic calibration presented in Section 2.1.2 is plotted in red and the

518 MORDOR-TS ensemble mean and mean plus or minus standard deviation are  
519 plotted in black (solid and dotted lines). The reference simulation clearly un-  
520 derestimates the flood peak (except for the Kerjean catchment). The ensemble  
521 probability density function underestimates the flood peak since the measured  
522 streamflow is closer to the upper dotted black line in Fig. 11 (ensemble mean  
523 + standard deviation) than to the mean curve; neither the perturbation of the  
524 hydrologic model parameters nor that of the rain scenario allows to overcome  
525 this effect.

### 526 5.1.3. Consistency between GSA-Arome and GSA-hydro results

527 GSA-hydro was carried out over the subperiod of  $P_{GSA-Arome}$  (01/26/2016-  
528 01/30/2016) and over a subperiod of  $P_{HEF}$  (12/23/2013-12/26/2013) and as-  
529 sociated Sobol' indices are shown in Fig. 12 at Tréodet. Results at Kerjean and  
530 Ty-Planche can be found in Fig. 19 and 20 as supplementary material in section  
531 7.

532 For the 2016 event, GSA-hydro (Fig. 12(a)) and GSA-Arome (Fig. 11(a))  
533 at short lead time, for which uncertainty in rainfall is not significant, show  
534 similar results. This allows to rely on GSA analysis when rainfall uncertainty  
535 is neglected for HEF construction at short forecast lead time. Moreover, GSA-  
536 hydro over  $P_{GSA-Arome}$  (Fig. 12(a)) and  $P_{HEF}$  (Fig. 12(b)) show similar results  
537 since these two events correspond to similar weather conditions. This allows to  
538 further the study with HEF construction over  $P_{HEF}$ . Finally, as previously  
539 stated over the three catchments, the two parameters  $k_{min}$  and  $Diff$  have very  
540 low Sobol' indices meaning that these two parameters have thus no impact  
541 on the simulated runoff. These parameters were thus fixed to their hydrologic  
542 calibration value for HEF generation in the following while the other parameters  
543 are drawn from uniform distribution described in Tab. 4.

### 544 5.2. Calibration of the HEF

545 Ensemble reliability and resolution are assessed with univariate and global  
546 RH and CRPS metrics computed over streamflow. Fig. 13 shows rank his-



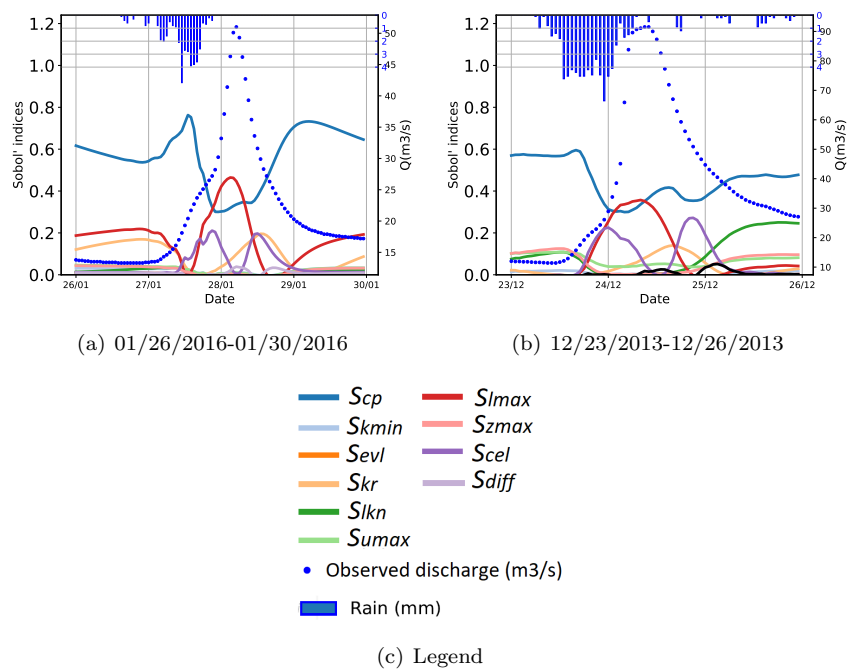


Figure 12: First order Sobol' indices without uncertainty on rain input for the events 01/26/2016-01/30/2016 (12(a)) and 12/23/2013-12/26/2013 (12(b)) at Tréodet catchment

547 tograms for raw HEF, QRF-hydro quantiles, QRF-nohydro quantiles and QRF-  
 548 hydro-TS HEF for 6-hour and 21-hour lead-times with associated CRPS, for  
 549 Tréodet catchment. Similar results for Kerjean (respectively Ty-Planche) are  
 550 given in Fig. 21 (respectively Fig. 22) as supplementary material in Section 7.  
 551 Tab. 6 presents the time-averaged  $\|\cdot\|_2$  norm for RHs and Fig. 14 shows the  
 552 time-varying CRPS for the four ensembles on all catchments.

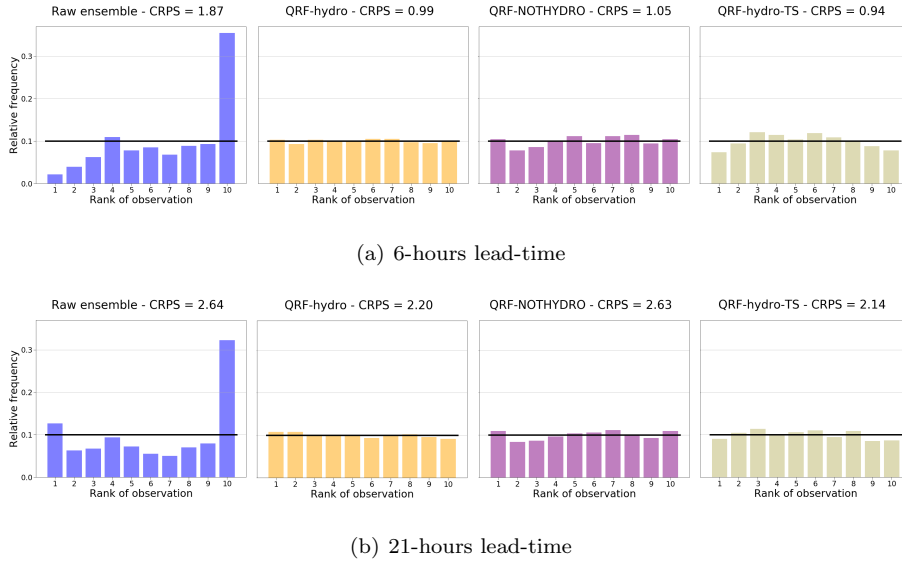


Figure 13: Rank histograms for raw HEF, QRF-hydro quantiles, QRF-nohydro quantiles and QRF-hydro-TS HEF for 6-hours and 21-hours lead-times for Treodet. The associated CRPS is given at the top of the panel.

Table 6: Time-averaged  $\|\cdot\|_2$  RHs norm over lead-times ranging from 1 to 24 hours for raw HEF, QRF-hydro quantiles, QRF-nohydro quantiles and QRF-hydro-TS HEF, for Tréodet, Kerjean and Ty-Planche.

Catchment	$10^{-2}\ \cdot\ _2$	$10^{-2}\ \cdot\ _2$	$10^{-2}\ \cdot\ _2$	$10^{-2}\ \cdot\ _2$
	raw	QRF-hydro	QRF-nohydro	QRF-hydro-TS
Tréodet	25.8	1.8	3.3	4.5
Kerjean	15.8	1.8	2.8	4.0
Ty-Planche	9.3	1.9	3.6	3.1

553 The raw ensemble is biased and underdispersive over all three catchments,

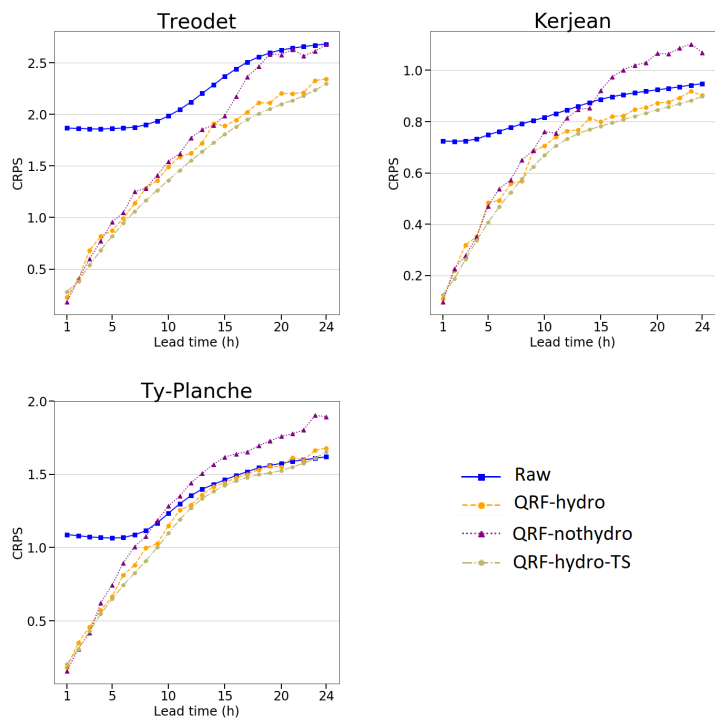


Figure 14: Time-varying CRPS for raw HEF, QRF-hydro quantiles, QRF-nohydro quantiles and QRF-hydro-TS HEF over lead-time for Tréodet, Kerjean and Ty-Planche.

554 and tends to underestimate the flood peak values. QRF-hydro and QRF-  
555 nohydro calibrated ensembles display flatter RH with significantly smaller  $\|\cdot\|_2$   
556 values. The use of hydrology-related predictors for calibration leads to the most  
557 satisfying results in terms of reliability. The TS smoothing procedure slightly  
558 deteriorates reliability. This is confirmed with statistical tests for the detection  
559 of slope, convexity or waves in the rank histograms (Jolliffe and Primo, 2008;  
560 Zamo, 2016) that lead to the rejection of the flatness hypothesis for the raw and  
561 the QRF-hydro-TS ensembles. It should be noted that reliability improvement  
562 remains significant as lead-time increases as shown in Fig. 13. Ensemble cali-  
563 bration considerably improves CRPS values for QRF-hydro and QRF-nohydro  
564 as shown in Fig. 14. Yet, the improvement decreases as the lead-time increases,  
565 especially when no hydrology-related predictors are used. The TS smoothing  
566 procedure further improves the CRPS values as this strategy provides less dis-  
567 persive trajectories, a priori centered on observations after calibration step. In  
568 conclusion, QRF-hydro calibration leads to the best reliability. It should be  
569 noted that TS slightly degrades reliability, yet provides far better results than  
570 the raw ensemble. The CRPS simultaneously assesses for reliability and res-  
571 olution. Since the reliability of QRF-hydro-TS is degraded and the CRPS is  
572 improved against QRF-hydro, that means that the TS procedure improves the  
573 resolution of the ensemble.

574 The cumulated and maximum streamflow over the 24 lead-times are com-  
575 puted for the four ensembles. For both quantities, the rank histogram,  $\|\cdot\|_2$   
576 associated norm and CRPS are computed and shown in Fig. 15 and in Tab. 7.  
577 Similar results for Kerjean (respectively Ty-Planche) are given in Fig. 23 (re-  
578 spectively Fig. 24) as supplementary material in Section 7.3.

579 Ensemble calibration improves the rank histogram for both cumulated and  
580 maximum streamflow. The rank histogram for cumulated and maximum stream-  
581 flow shows that the raw ensemble is underdispersive and biased as it tends to  
582 underestimate high cumulated and maximum streamflows for all three catch-  
583 ments. For Tréodet, all calibrated ensembles show overdispersive rank histogram  
584 for cumulated streamflow while it is rather flat for Kerjean and Ty-Planche when

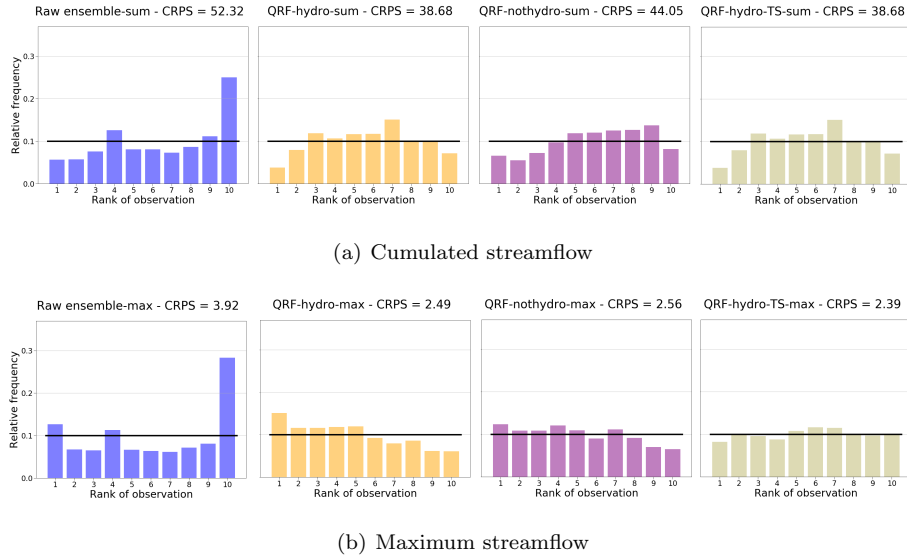


Figure 15: Rank histograms for the cumulated and the maximum streamflows over the 24 hours lead-time for the four ensembles for Tréodet catchment. The associated CRPS is given at the top of the panel.

Table 7: Time-averaged  $\|\cdot\|_2$  RHs norm for cumulated and maximum streamflow cumulated over 24 lead-times for the four ensemble for Tréodet, Kerjean and Ty-Planche catchments.

		$10^{-2}\ \cdot\ _2$	$10^{-2}\ \cdot\ _2$	$10^{-2}\ \cdot\ _2$	$10^{-2}\ \cdot\ _2$
Catchment		raw	QRF-hydro	QRF-nohydro	QRF-hydro-TS
Cumulated Stramflow	Tréodet	17.1	9.4	8.8	9.4
	Kerjean	13.3	6.9	11.1	6.9
	Ty-Planche	9.1	6.8	9.5	6.8
Maximum streamflow	Tréodet	20.5	8.6	6.0	3.3
	Kerjean	16.8	6.6	4.8	3.9
	Ty-Planche	20.1	4.0	3.6	6.7

585 hydrology-related predictors are used. It should be noted that TS has no im-  
 586 pact on rank histogram for cumulated streamflow, meaning that it preserves  
 587 the volume of water within each simulated member. The rank histogram for  
 588 maximum streamflow is further improved when trajectory smoothing is used at  
 589 Tréodet and Kerjean but not at Ty-Planche. The merits of ensemble calibration  
 590 on reliability is confirmed by the  $\|\cdot\|_2$  norm values that are significantly reduced  
 591 for cumulated and maximum streamflow ensembles. Ensemble calibration also  
 592 significantly improves CRPS values, especially when hydrology-related predic-  
 593 tors are used for the three catchments. For the three catchments, the CRPS  
 594 computed for maximum streamflow is minimized when trajectory smoothing is  
 595 applied. To conclude, ensemble calibration with hydrology related predictors  
 596 globally improves reliability and CRPS values for both maximum and cumu-  
 597 lated streamflow and TS procedure brings a slight additional improvement for  
 598 the maximum streamflow.

### 599 5.2.1. Importance of the QRF predictors

600 The a priori choice of the predictors is a key element in the QRF calibration.  
 601 The a posteriori usefulness of the predictors is assessed as the loss in the mean-  
 602 squared error of the whole forest if the predictor is randomly permuted without  
 603 replacement: the values of the given predictor is a random sample of the original  
 604 values. The predictor is of great (little) importance if the mean-squared error  
 605 does (not) significantly increase when the predictor is randomly permuted.

606 Fig. 16 displays the log-importance of the QRF-hydro predictors for 2-hour  
 607 and 16-hour lead-times for Tréodet. Since the QRF calibration is achieved with  
 608 a cross-validation approach, a forest is built for each of the 27 months of  $P_{HEF}$ ,  
 609 and the importance of each predictor is thus computed over each of the 27  
 610 forests. For short lead-times, the most important predictor is the measured  
 611 streamflow  $Q_0$ ; that is consistent with the fact that hourly streamflows are  
 612 strongly correlated in time. The second most important predictor is the gradient  
 613 of the measured streamflow  $GradQ_0$  that accounts for the flow dynamics. When  
 614 the lead-time increases, rain-related predictors become predominant. When

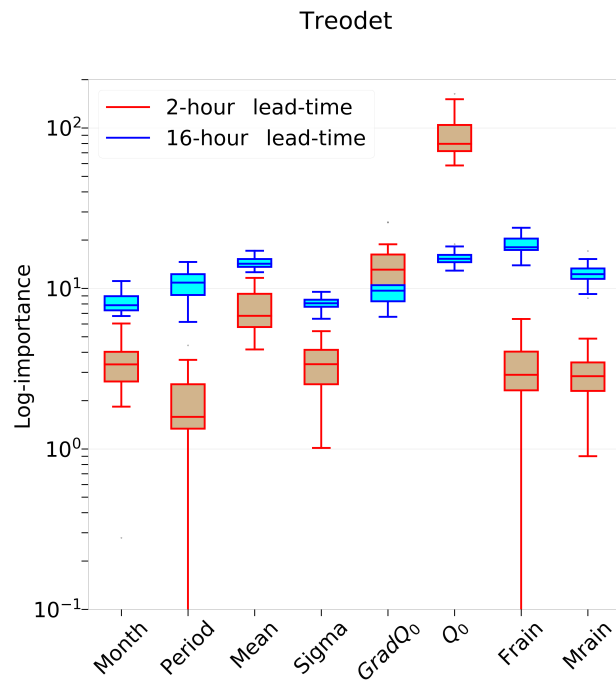


Figure 16: Log-importance of QRF-hydro predictors for the 2-hour and 16-hour lead-times at Tréodet. The box-plot is built for the measure of importance over 27 forests. The box extends from the lower to upper quartile values of the data, with a line at the median. The whiskers extend from the box to show the range of the data.

615 the lead-time exceeds the catchment concentration time (about 15 hours for  
616 Tréodet), the cumulated forecasted rain since the beginning of the run  $F_{rain}$ ,  
617 is the most important predictor, before  $Q_0$ . It also predominates the impact  
618 of the measured rain  $M_{rain}$ , meaning that the measured rainfall before the run  
619 has less importance than forecasted rainfall between the current time and the  
620 lead-time.

621 For the short and long lead-times, the raw ensemble mean and the standard  
622 deviation (*Mean* and *Sigma*) have moderate influence (with a slight predom-  
623 inance of *Mean* over *Sigma*). This is consistent with the fact that the QRF-  
624 nohydro ensemble, which has not been built with these predictors, presents  
625 comparatively less favourable performances than the QRF-hydro ensemble. It  
626 should finally be noted that the importance of the different predictors tend  
627 to homogenise as the lead-time increases, and the interquartile range tends to  
628 decrease.



## 629 **6. Discussion and perspectives**

630 In this study, the hydrometeorological chain was investigated with an en-  
631 semble approach at a catchment scale, with the aim of issuing a statistically  
632 coherent Hydrological Ensemble Forecast (HEF) for up to 24 hours for the  
633 Odet cathment in France. The predominance of uncertainty sources is as-  
634 sessed with a global sensitivity analysis, first taking into account uncertainty in  
635 rain using AromeEPS-RR1, then only focusing on hydrological parameters from  
636 MORDOR-TS model. Predominant parameters were identified and taken into  
637 account in the HEF generation.

638 Because of the lack of AromeEPS-RR1 availability, the HEF is generated  
639 without taking into account uncertainty in rain forecast. However, it should  
640 be pointed out that for smaller lead-times than the concentration time of the  
641 catchments, the conclusions of this study are expected to be similar, as demon-  
642 strated by the comparative GSA without uncertainty in rainfall: the Sobol  
643 index associated to the rain is negligible at 6-hour lead-time. It was shown that  
644 the raw hydrological ensemble is underdispersive and underestimates observed  
645 streamflows, especially large values. The model-based-only approach is thus  
646 not sufficient to generate well-calibrated ensembles. The QRF calibration strat-  
647 egy is also applied for quantiles generation, with two different approaches: an  
648 observation-based-only approach, where the hydrologic model is not used and a  
649 combined model-observation approach where both observed and simulated infor-  
650 mation from the raw ensemble are used as predictors. The merits of the QRF  
651 calibration are assessed with reliability and resolution metrics computed for  
652 time-varying, cumulated and maximal streamflow. This article demonstrates,  
653 on the Odet catchment, that forest-based techniques, often used for the cal-  
654 ibration of EPS, are also suitable for hydrologic ensemble calibration. QRF  
655 calibration provides an improvement of the reliability and CRPS values over  
656 the three studied sub-catchments. It was shown that better forecast skills are  
657 obtained when hydrology-related predictors are used. This highlights that the  
658 choice of the predictors is of great importance for the QRF calibration strategy

659 to be successful. Moreover, it also shows that statistically learning from obser-  
660 vation fails to substitute for solving partial derivative equations when it comes  
661 to representing and forecasting the dynamics of the catchment. It should also  
662 be kept in mind that the reconstruction step following QRF-calibration requires  
663 the availability of the raw HEF in the Ensemble Copula Coupling method. The  
664 reconstructed HEF should finally be smoothed out before use as input for a  
665 hydraulic simulation with a hydrometeorological chain. The reliability of the  
666 smoothed ensemble is slightly deteriorated for hourly streamflow, but CRPS  
667 and global skills are mainly improved.

668 This strategy paves the way for an operational HEF system. A straightfor-  
669 ward perspective is to apply this study over a period for which AromeEPS-RR1  
670 is available, hydrological events occurred, and that is long enough for QRF cali-  
671 bration to be applied on streamflows. This would allow to fully consider sources  
672 of uncertainty, especially those that relate to precipitation. Taking into account  
673 a larger variation of the corrective rain factor in MORDOR-TS jointly to using  
674 EPS could also be considered. This methodology could *a priori* be applied and  
675 assessed to other catchments for consistency check. The choice of predictors  
676 may be catchment dependant. For instance, the predictors  $M_{rain}$  and  $F_{rain}$   
677 are closely related to the concentration time of the catchment: the period (in  
678 hours) over which the mean of measured or forecasted rain is computed may be  
679 adjusted, hence reduced for catchments that are subject to flash floods. The pre-  
680 dictor *Month* may also be more significant as it allows to discriminate autumn  
681 flash floods from other events.

682 Finally, the hydrometeorological simulation chain could be extended to hy-  
683 draulics, using forecasted calibrated streamflows as inputs to a hydraulic model  
684 to provide Hydraulics Ensemble Forecast. GSA on discretized water level and  
685 discharge in the river would be carried out with respect to rain-, hydrology- and  
686 hydraulic- related sources of uncertainty, now also considering uncertain friction  
687 and river bathymetry. Major sources uncertainties in the meteo-hydro-hydraulic  
688 chain could then be corrected in real time with assimilation of observed water  
689 level relying on the multi-physics ensemble simulation system.

690 *Acknowledgements.* The authors greatfully thank the French national service for  
691 flood forecasting (SCHAPI) for supporting this study, EDF-DTG for providing  
692 MORDOR-TS hydrologic model as well as Rémy Garçon and Matthieu Le Lay  
693 for their active support on the model, Fabrice Zaoui (EDF R&D) for developping  
694 the Python API interface to MORDOR-TS and Charles Perrin and Julie Viatgé  
695 (INRAE) for fruitful discussions on the GRP hydrologic model. The authors also  
696 acknowledge the Pôle de Calcul et de Données Marines (PCDM) for providing  
697 DATARMOR storage and computational resources (<http://www.ifremer.fr/pcdm>).

<sup>698</sup> **7. Supplementary material**

<sup>699</sup> *7.1. Sobol' indices at Kerjean and Ty-Planche*

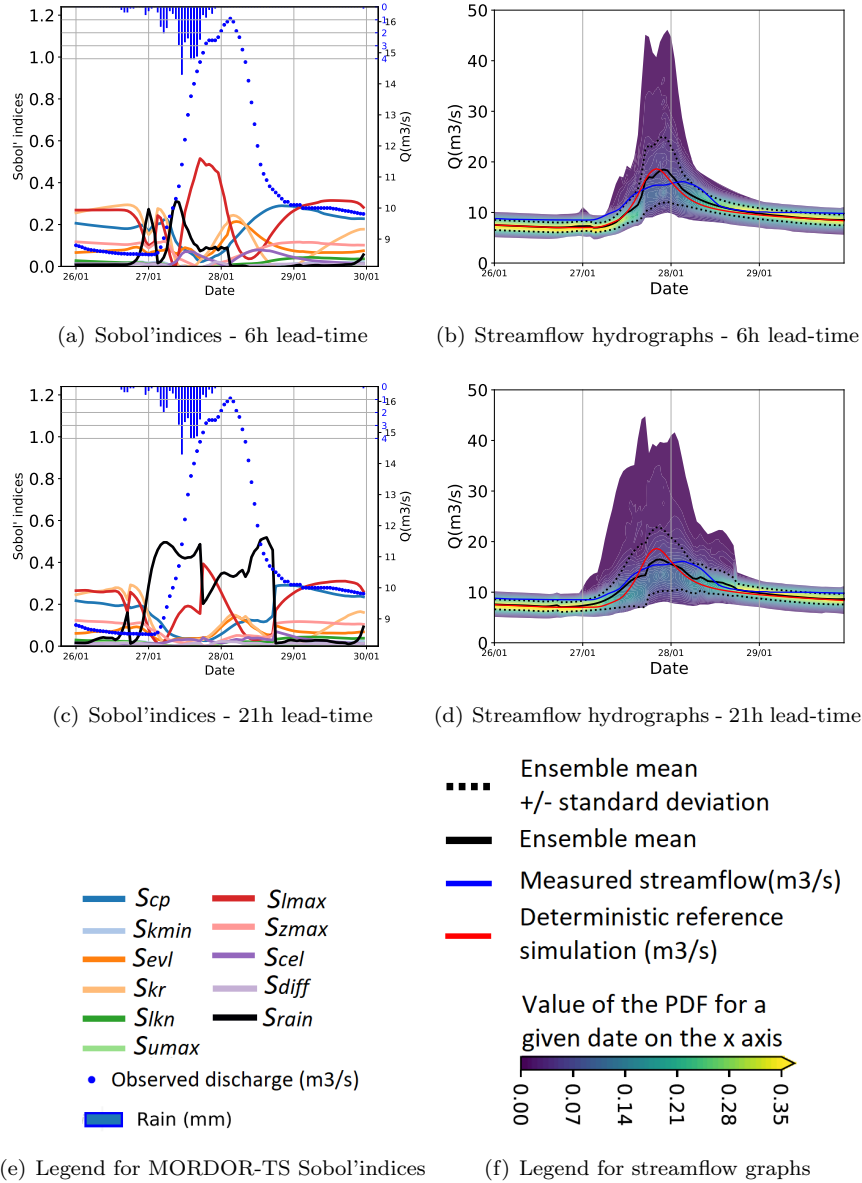
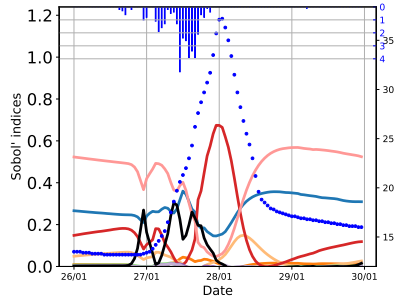
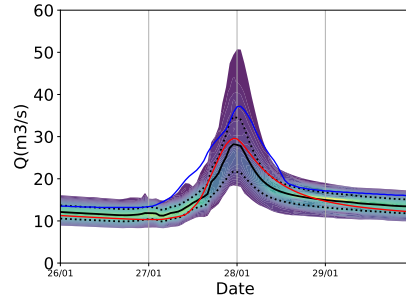


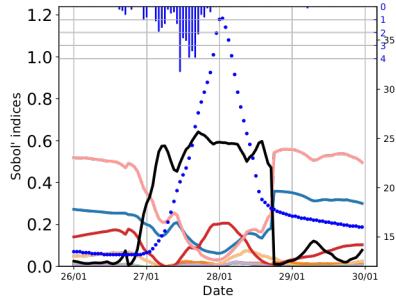
Figure 17: First order Sobol' indices and associated hydrographs for the event between 01/26/2016 and 01/30/2016 at Kerjean catchment for GSA-Arome. A vertical section of Fig(b) and Fig(d) displays the probability density function at the given date, according to the legend shown in (f). The results are shown for 6-hour ((a) and (b)) and 21-hour ((c) and (d)) lead-times.



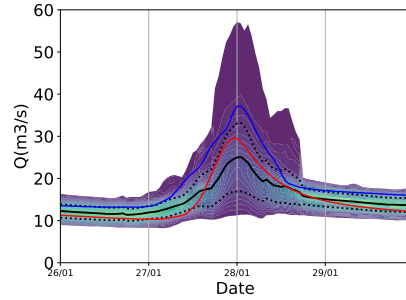
(a) Sobol' indices - 6h lead-time



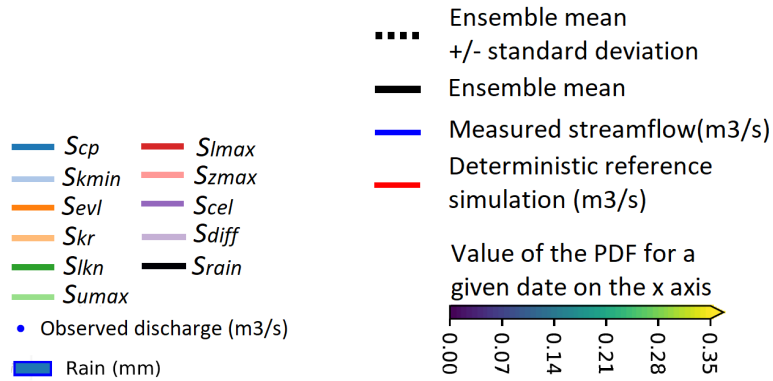
(b) Streamflow hydrographs - 6h lead-time



(c) Sobol' indices - 21h lead-time



(d) Streamflow hydrographs - 21h lead-time



(e) Legend for MORDOR-TS Sobol' indices

(f) Legend for streamflow graphs

Figure 18: First order Sobol' indices and associated hydrographs for the event between 01/26/2016 and 01/30/2016 at Ty-Planche catchment for GSA-Arome. A vertical section of Fig(b) and Fig(d) displays the probability density function at the given date, according to the legend shown in (f). The results are shown for 6-hour ((a) and (b)) and 21-hour ((c) and (d)) lead-times.

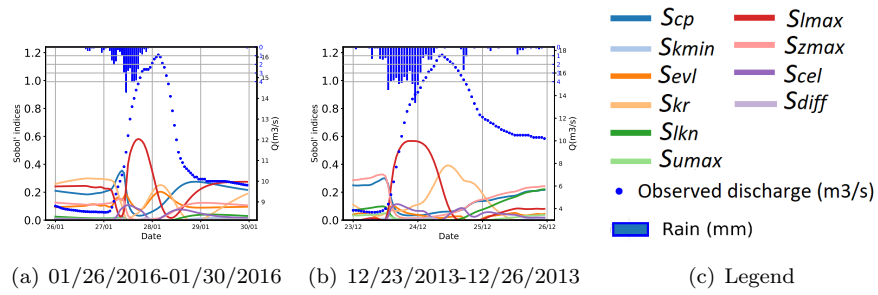


Figure 19: First order Sobol' indices without uncertainty on rain input for the events 01/26/2016-01/30/2016 (19(a)) and 12/23/2013-12/26/2013 (19(b)) at Kerjean catchment

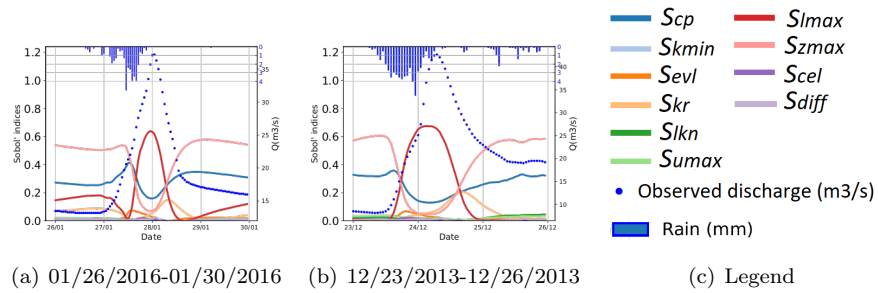


Figure 20: First order Sobol' indices without uncertainty on rain input for the events 01/26/2016-01/30/2016 (20(a)) and 12/23/2013-12/26/2013 (20(b)) at Ty-Planche catchment

700 7.2. Rank Histograms for Kerjean and Ty-Planche catchments

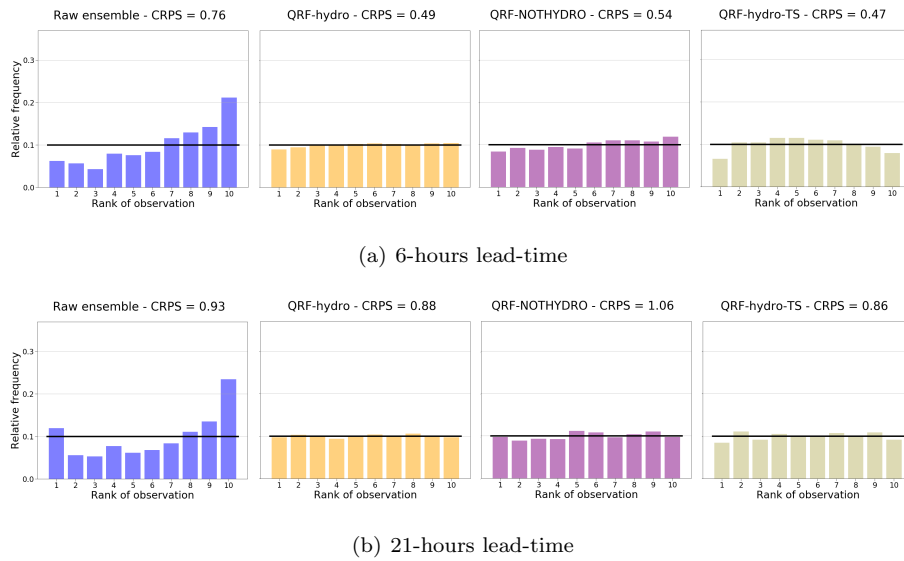
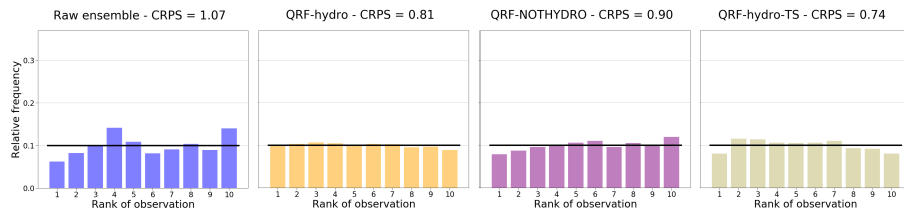
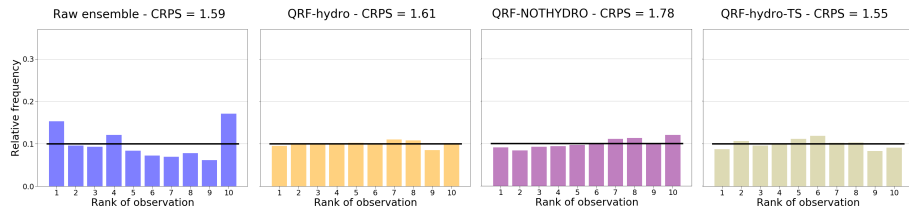


Figure 21: Rank histograms for raw HEF, QRF-hydro quantiles, QRF-nohydro quantiles and QRF-hydro-TS HEF for 6-hours and 21-hours lead-times for Kerjean. The associated CRPS is given at the top of the panel.





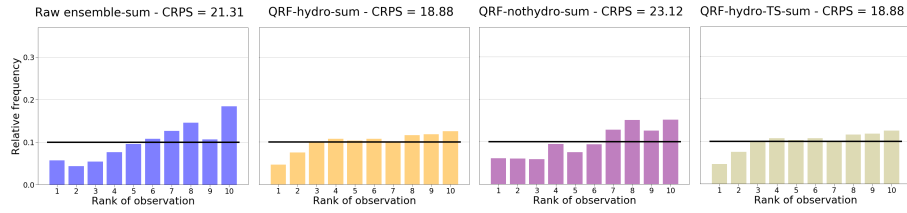
(a) 6-hours lead-time



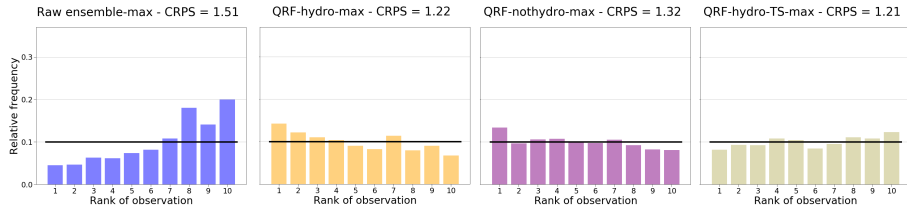
(b) 21-hours lead-time

Figure 22: Rank histograms for raw HEF, QRF-hydro quantiles, QRF-nothydro quantiles and QRF-hydro-TS HEF for 6-hours and 21-hours lead-times for Ty-Planche. The associated CRPS is given at the top of the panel.

701 7.3. Global skills for Kerjean and Ty-Planche catchments

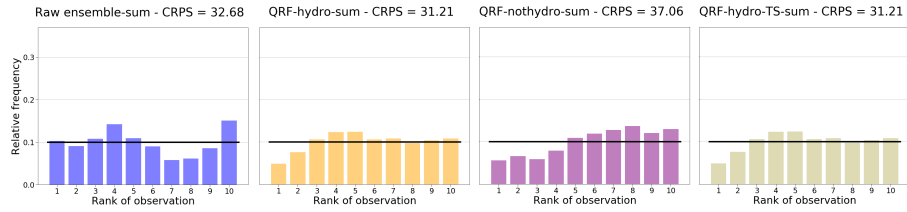


(a) Cumulated streamflow

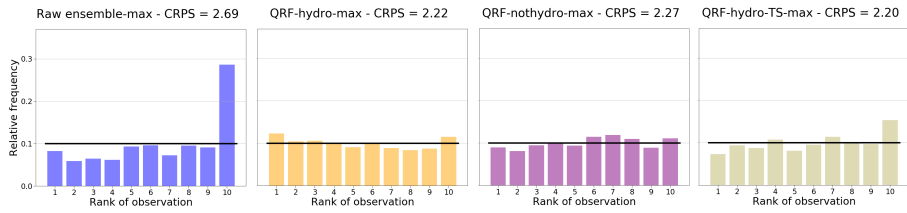


(b) Maximum streamflow

Figure 23: Rank histograms for the cumulated and the maximum streamflows over the 24 hours lead-time for the four ensembles for Kerjean catchment. The associated CRPS is given at the top of the panel.



(a) Cumulated streamflow



(b) Maximum streamflow

Figure 24: Rank histograms for the cumulated and the maximum streamflows over the 24 hours lead-time for the four ensembles for Ty-Planche catchment. The associated CRPS is given at the top of the panel.

702 **8. Bibliographie**

703 **References**

704 Anderson, J.L., 1996. A Method for Producing and Evaluating Probabilis-  
705 tic Forecasts from Ensemble Model Integrations. *Journal of Climate* 9,  
706 1518–1530. URL: [https://doi.org/10.1175/1520-0442\(1996\)009<1518:  
707 AMFPAE>2.0.CO;2](https://doi.org/10.1175/1520-0442(1996)009<1518:AMFPAE>2.0.CO;2), doi:10.1175/1520-0442(1996)009<1518:AMFPAE>2.0.  
708 CO;2. number: 7.

709 Bellier, J., 2018. Prévisions hydrologiques probabilistes dans un cadre multi-  
710 varié: quels outils pour assurer fiabilité et cohérence spatio-temporelle? Ph.D.  
711 thesis. Université Grenoble Alpes. URL: [https://tel.archives-ouvertes.  
712 fr/tel-01950725](https://tel.archives-ouvertes.fr/tel-01950725).

713 Bellier, J., Zin, I., Bontron, G., 2017. Sample Stratification in Verifica-  
714 tion of Ensemble Forecasts of Continuous Scalar Variables: Potential Ben-  
715 efits and Pitfalls. *Monthly Weather Review* 145, 3529–3544. URL: [http:  
716 //journals.ametsoc.org/doi/10.1175/MWR-D-16-0487.1](http://journals.ametsoc.org/doi/10.1175/MWR-D-16-0487.1), doi:10.1175/  
717 MWR-D-16-0487.1. number: 9.

718 Bourgin, F., 2014. Comment quantifier l’incertitude prédictive en modélisation  
719 hydrologique ? : Travail exploratoire sur un grand échantillon de bassins  
720 versants. PhD Thesis. AgroParisTech. URL: [http://www.theses.fr/  
721 2014AGPT0016/document](http://www.theses.fr/2014AGPT0016/document).

722 Bouttier, F., Raynaud, L., Nuissier, O., Ménétrier, B., 2016. Sensitiv-  
723 ity of the AROME ensemble to initial and surface perturbations during  
724 HyMeX. *Quarterly Journal of the Royal Meteorological Society* 142, 390–  
725 403. URL: [https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/  
726 qj.2622](https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2622), doi:10.1002/qj.2622. number: S1.

727 Bouttier, F., Vié, B., Nuissier, O., Raynaud, L., 2012. Impact of Stochas-  
728 tic Physics in a Convection-Permitting Ensemble. *Monthly Weather Review*

729 140, 3706–3721. URL: <http://journals.ametsoc.org/doi/abs/10.1175/>  
730 [MWR-D-12-00031.1](http://journals.ametsoc.org/doi/abs/10.1175/MWR-D-12-00031.1), doi:10.1175/MWR-D-12-00031.1. number: 11.

731 Breiman, L., 1996. Bagging predictors. *Machine Learning* 24, 123–  
732 140. URL: <http://link.springer.com/10.1007/BF00058655>, doi:10.  
733 1007/BF00058655. number: 2.

734 Breiman, L., 2001. Random Forests. *Machine Learning* 45, 5–32. doi:10.1023/  
735 A:1010933404324.. number: 4.

736 Bremnes, J.B., 2004. Probabilistic wind power forecasts using local quantile  
737 regression. *Wind Energy* 7, 47–54. URL: [https://onlinelibrary.wiley.](https://onlinelibrary.wiley.com/doi/abs/10.1002/we.107)  
738 [com/doi/abs/10.1002/we.107](https://onlinelibrary.wiley.com/doi/abs/10.1002/we.107), doi:10.1002/we.107. number: 1.

739 Candille, G., Talagrand, O., 2005. Evaluation of probabilistic prediction systems  
740 for a scalar variable. *Quarterly Journal of the Royal Meteorological Society*  
741 131, 2131–2150. URL: <http://doi.wiley.com/10.1256/qj.04.71>, doi:10.  
742 1256/qj.04.71. number: 609.

743 Champeaux, J.L., Dupuy, P., Laurantin, O., Soulan, I., Tabary, P., Soubey-  
744 roux, J.M., 2009. Les mesures de précipitations et l'estimation des  
745 lames d'eau à Météo-France : état de l'art et perspectives. *La Houille*  
746 *Blanche* , 28–34 URL: <http://www.shf-lhb.org/10.1051/lhb/2009052>,  
747 doi:10.1051/lhb/2009052. number: 5.

748 Cloke, H., Pappenberger, F., 2009. Ensemble flood forecasting: A review. *Jour-*  
749 *nal of Hydrology* 375, 613–626. URL: [https://linkinghub.elsevier.com/](https://linkinghub.elsevier.com/retrieve/pii/S0022169409003291)  
750 [retrieve/pii/S0022169409003291](https://linkinghub.elsevier.com/retrieve/pii/S0022169409003291), doi:10.1016/j.jhydrol.2009.06.005.  
751 number: 3-4.

752 Delle Monache, L., Eckel, F.A., Rife, D.L., Nagarajan, B., Searight, K., 2013.  
753 Probabilistic Weather Prediction with an Analog Ensemble. *Monthly Weather*  
754 *Review* 141, 3498–3516. URL: [http://journals.ametsoc.org/doi/abs/](http://journals.ametsoc.org/doi/abs/10.1175/MWR-D-12-00281.1)  
755 [10.1175/MWR-D-12-00281.1](http://journals.ametsoc.org/doi/abs/10.1175/MWR-D-12-00281.1), doi:10.1175/MWR-D-12-00281.1. number: 10.

- 756 Delle Monache, L., Hacker, J.P., Zhou, Y., Deng, X., Stull, R.B.,  
757 2006. Probabilistic aspects of meteorological and ozone regional en-  
758 semble forecasts. *Journal of Geophysical Research: Atmospheres*  
759 111. URL: [https://agupubs.onlinelibrary.wiley.com/doi/abs/10.](https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005JD006917)  
760 1029/2005JD006917, doi:10.1029/2005JD006917. number: D24.
- 761 Demargne, J., Wu, L., Regonda, S.K., Brown, J.D., Lee, H., He, M., Seo,  
762 D.J., Hartman, R., Herr, H.D., Fresch, M., Schaake, J., Zhu, Y., 2014.  
763 The Science of NOAA's Operational Hydrologic Ensemble Forecast Service.  
764 *Bulletin of the American Meteorological Society* 95, 79–98. URL: <http://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-12-00081.1>, doi:10.  
765 1175/BAMS-D-12-00081.1, doi:10.  
766 1175/BAMS-D-12-00081.1. number: 1.
- 767 Descamps, L., Labadie, C., Joly, A., Bazile, E., Arbogast, P., Cébron,  
768 P., 2015. PEARP, the Météo-France short-range ensemble prediction  
769 system. *Quarterly Journal of the Royal Meteorological Society* 141,  
770 1671–1685. URL: [https://rmets.onlinelibrary.wiley.com/doi/abs/10.](https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2469)  
771 1002/qj.2469, doi:10.1002/qj.2469. number: 690.
- 772 Dietrich, J., Schumann, A.H., Redetzky, M., Walther, J., Denhard, M., Wang,  
773 Y., Pfützner, B., Büttner, U., 2009. Assessing uncertainties in flood forecasts  
774 for decision making: prototype of an operational flood management system  
775 integrating ensemble predictions. *Natural Hazards and Earth System Sci-*  
776 *ence* 9, 1529–1540. URL: [http://www.nat-hazards-earth-syst-sci.net/](http://www.nat-hazards-earth-syst-sci.net/9/1529/2009/)  
777 9/1529/2009/, doi:10.5194/nhess-9-1529-2009. number: 4.
- 778 Duan, Q., Ajami, N.K., Gao, X., Sorooshian, S., 2007. Multi-model ensem-  
779 ble hydrologic prediction using Bayesian model averaging. *Advances in*  
780 *Water Resources* 30, 1371–1386. URL: [https://linkinghub.elsevier.](https://linkinghub.elsevier.com/retrieve/pii/S030917080600220X)  
781 [com/retrieve/pii/S030917080600220X](https://linkinghub.elsevier.com/retrieve/pii/S030917080600220X), doi:10.1016/j.advwatres.2006.  
782 11.014. number: 5.
- 783 Efron, B., Stein, C., 1981. The Jackknife Estimate of Variance. *The Annals*

784 of Statistics 9, 586–596. URL: <https://doi.org/10.1214/aos/1176345462>,  
785 doi:10.1214/aos/1176345462. number: 3.

786 Emery, C.M., Biancamaria, S., Boone, A., Garambois, P.A., Ricci, S., Ro-  
787 choux, M.C., Decharme, B., 2016. Temporal Variance-Based Sensitivity Anal-  
788 ysis of the River-Routing Component of the Large-Scale Hydrological Model  
789 ISBA–TRIP: Application on the Amazon Basin. *Journal of Hydrometeo-*  
790 *rology* 17, 3007–3027. URL: [http://journals.ametsoc.org/doi/10.1175/](http://journals.ametsoc.org/doi/10.1175/JHM-D-16-0050.1)  
791 [JHM-D-16-0050.1](http://journals.ametsoc.org/doi/10.1175/JHM-D-16-0050.1), doi:10.1175/JHM-D-16-0050.1. number: 12.

792 Garambois, P.A., Roux, H., Larnier, K., Castaings, W., Dartus, D.,  
793 2013. Characterization of process-oriented hydrologic model behavior  
794 with temporal sensitivity analysis for flash floods in Mediterranean catch-  
795 ments. *Hydrology and Earth System Sciences* 17, 2305–2322. URL:  
796 <https://www.hydrolog-earh-syst-sci.net/17/2305/2013/>, doi:10.5194/  
797 [hess-17-2305-2013](https://www.hydrolog-earh-syst-sci.net/17/2305/2013/). number: 6.

798 Garavaglia, F., Le Lay, M., Gottardi, F., Garçon, R., Gailhard, J., Pa-  
799 quet, E., Mathevet, T., 2017. Impact of model structure on flow sim-  
800 ulation and hydrological realism: from a lumped to a semi-distributed  
801 approach. *Hydrology and Earth System Sciences* 21, 3937–3952. URL:  
802 <https://www.hydrolog-earh-syst-sci.net/21/3937/2017/>, doi:10.5194/  
803 [hess-21-3937-2017](https://www.hydrolog-earh-syst-sci.net/21/3937/2017/). number: 8.

804 Garçon, R., 1996. Prévion opérationnelle des apports de la Durance à  
805 Serre-Ponçon à l’aide du modèle MORDOR. Bilan de l’année 1994-1995.  
806 *La Houille Blanche* , 71–76URL: [http://www.shf-lhb.org/10.1051/lhb/](http://www.shf-lhb.org/10.1051/lhb/1996056)  
807 [1996056](http://www.shf-lhb.org/10.1051/lhb/1996056), doi:10.1051/lhb/1996056. number: 5.

808 Gneiting, T., Raftery, A.E., 2007. Strictly Proper Scoring Rules, Pre-  
809 diction, and Estimation. *Journal of the American Statistical Associa-*  
810 *tion* 102, 359–378. URL: [http://www.tandfonline.com/doi/abs/10.1198/](http://www.tandfonline.com/doi/abs/10.1198/016214506000001437)  
811 [016214506000001437](http://www.tandfonline.com/doi/abs/10.1198/016214506000001437), doi:10.1198/016214506000001437. number: 477.

- 812 Gneiting, T., Raftery, A.E., Westveld, A.H., Goldman, T., 2005. Cali-  
813 brated Probabilistic Forecasting Using Ensemble Model Output Statistics  
814 and Minimum CRPS Estimation. *Monthly Weather Review* 133, 1098–  
815 1118. URL: <https://doi.org/10.1175/MWR2904.1>, doi:10.1175/MWR2904.  
816 1. number: 5.
- 817 Gneiting, T., Ranjan, R., 2011. Comparing Density Forecasts Using Threshold-  
818 and Quantile-Weighted Scoring Rules. *Journal of Business & Economic Statis-*  
819 *tics* 29, 411–422. URL: [http://www.tandfonline.com/doi/abs/10.1198/](http://www.tandfonline.com/doi/abs/10.1198/jbes.2010.08110)  
820 [jbes.2010.08110](http://www.tandfonline.com/doi/abs/10.1198/jbes.2010.08110), doi:10.1198/jbes.2010.08110. number: 3.
- 821 Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decompo-  
822 sition of the mean squared error and NSE performance criteria: Impli-  
823 cations for improving hydrological modelling. *Journal of Hydrology* 377,  
824 80 – 91. URL: [http://www.sciencedirect.com/science/article/pii/](http://www.sciencedirect.com/science/article/pii/S0022169409004843)  
825 [S0022169409004843](http://www.sciencedirect.com/science/article/pii/S0022169409004843), doi:[https://doi.org/10.1016/j.jhydrol.2009.08.](https://doi.org/10.1016/j.jhydrol.2009.08.003)  
826 003. number: 1.
- 827 Hamill, T.M., Colucci, S.J., 1997. Verification of Eta–RSM Short-Range  
828 Ensemble Forecasts. *Monthly Weather Review* 125, 1312–1327. URL:  
829 [http://journals.ametsoc.org/doi/abs/10.1175/1520-0493%281997%](http://journals.ametsoc.org/doi/abs/10.1175/1520-0493%281997%29125%3C1312%3AVOERSR%3E2.0.CO%3B2)  
830 [29125%3C1312%3AVOERSR%3E2.0.CO%3B2](http://journals.ametsoc.org/doi/abs/10.1175/1520-0493%281997%29125%3C1312%3AVOERSR%3E2.0.CO%3B2), doi:10.1175/1520-0493(1997)  
831 125<1312:VOERSR>2.0.CO;2. number: 6.
- 832 Hamill, T.M., Whitaker, J.S., 2006. Probabilistic Quantitative Precipitation  
833 Forecasts Based on Reforecast Analogs: Theory and Application. *Monthly*  
834 *Weather Review* 134, 3209–3229. URL: [http://journals.ametsoc.org/](http://journals.ametsoc.org/doi/abs/10.1175/MWR3237.1)  
835 [doi/abs/10.1175/MWR3237.1](http://journals.ametsoc.org/doi/abs/10.1175/MWR3237.1), doi:10.1175/MWR3237.1. number: 11.
- 836 Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learn-  
837 ing. Springer. URL: [https://web.stanford.edu/~hastie/Papers/ESLII.](https://web.stanford.edu/~hastie/Papers/ESLII.pdf)  
838 [pdf](https://web.stanford.edu/~hastie/Papers/ESLII.pdf).
- 839 Hayami, S., 1951. On The Propagation Of Flood Waves. Technical Report.

840 Disaster prevention research institute. URL: <https://repository.kulib.kyoto-u.ac.jp/dspace/bitstream/2433/123641/1/b01p0n000p01.pdf>.

841

842 Hemri, S., Lisniak, D., Klein, B., 2015. Multivariate postprocessing techniques for probabilistic hydrological forecasting. *Water Resources Research* 51, 7436–7451. URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014WR016473>, doi:10.1002/2014WR016473. number: 9.

843

844

845

846 Hersbach, H., 2000. Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. *Weather and Forecasting* 15, 559–570. URL: <http://journals.ametsoc.org/doi/abs/10.1175/1520-0434%282000%29015%3C0559%3ADOTCRP%3E2.0.CO%3B2>, doi:10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2. number: 5.

847

848

849

850

851 Hoeffding, W., 1948. A Class of Statistics with Asymptotically Normal Distribution. *The Annals of Mathematical Statistics* 19, 293–325. URL: <https://doi.org/10.1214/aoms/1177730196>, doi:10.1214/aoms/1177730196. number: 3.

852

853

854

855 Hopson, T.M., Webster, P.J., 2010. A 1–10-Day Ensemble Forecasting Scheme for the Major River Basins of Bangladesh: Forecasting Severe Floods of 2003–07\*. *Journal of Hydrometeorology* 11, 618–641. URL: <http://journals.ametsoc.org/doi/abs/10.1175/2009JHM1006.1>, doi:10.1175/2009JHM1006.1. number: 3.

856

857

858

859

860 Iooss, B., Lemaître, P., 2015. A Review on Global Sensitivity Analysis Methods, in: Dellino, G., Meloni, C. (Eds.), *Uncertainty Management in Simulation-Optimization of Complex Systems*. Springer US, Boston, MA. volume 59, pp. 101–122. URL: [http://link.springer.com/10.1007/978-1-4899-7547-8\\_5](http://link.springer.com/10.1007/978-1-4899-7547-8_5), doi:10.1007/978-1-4899-7547-8\_5.

861

862

863

864

865 Jolliffe, I.T., Primo, C., 2008. Evaluating Rank Histograms Using Decompositions of the Chi-Square Test Statistic. *Monthly Weather Review* 136, 2133–2139. URL: <https://doi.org/10.1175/2007MWR2219.1>, doi:10.1175/2007MWR2219.1. number: 6.

866

867

868



- 869 Jolliffe, I.T., Stephenson, D.B., 2003. Forecast verification: a practitioner's  
870 guide in atmospheric science. Wiley, Chichester.
- 871 Krzysztofowicz, R., 1999. Bayesian theory of probabilistic forecasting via  
872 deterministic hydrologic model. *Water Resources Research* 35, 2739–  
873 2750. URL: <http://doi.wiley.com/10.1029/1999WR900099>, doi:10.1029/  
874 1999WR900099. number: 9.
- 875 Le Moine, N., Hendrickx, F., Gailhard, J., Garçon, R., Gottardi, F., 2015.  
876 Hydrologically Aided Interpolation of Daily Precipitation and Tempera-  
877 ture Fields in a Mesoscale Alpine Catchment. *Journal of Hydrometeorol-  
878 ogy* 16, 2595–2618. URL: [http://journals.ametsoc.org/doi/10.1175/  
879 JHM-D-14-0162.1](http://journals.ametsoc.org/doi/10.1175/JHM-D-14-0162.1), doi:10.1175/JHM-D-14-0162.1. number: 6.
- 880 Leleu, Isabelle, Tonnelier, Isabelle, Puechberty, Rachel, Gouin, Philippe,  
881 Viquendi, Isabelle, Cobos, Laurent, Foray, Anouck, Baillon, Martine, Ndimi,  
882 Pierre-Olivier, 2014. La refonte du système d'information national pour la ges-  
883 tion et la mise à disposition des données hydrométriques. *La Houille Blanche*  
884 , 25–32 URL: <https://doi.org/10.1051/lhb/2014004>, doi:10.1051/lhb/  
885 2014004. number: 1.
- 886 Li, W., Duan, Q., Miao, C., Ye, A., Gong, W., Di, Z., 2017. A review on sta-  
887 tistical postprocessing methods for hydrometeorological ensemble forecast-  
888 ing. *WIREs Water* 4, e1246. URL: [https://onlinelibrary.wiley.com/  
889 doi/abs/10.1002/wat2.1246](https://onlinelibrary.wiley.com/doi/abs/10.1002/wat2.1246), doi:10.1002/wat2.1246. number: 6 eprint:  
890 <https://onlinelibrary.wiley.com/doi/pdf/10.1002/wat2.1246>.
- 891 Matheson, J.E., Winkler, R.L., 1976. Scoring Rules for Continuous Prob-  
892 ability Distributions. *Management Science* 22, 1087–1096. URL: [http:  
893 //pubsonline.informs.org/doi/abs/10.1287/mnsc.22.10.1087](http://pubsonline.informs.org/doi/abs/10.1287/mnsc.22.10.1087), doi:10.  
894 1287/mnsc.22.10.1087. number: 10.
- 895 Meinshausen, N., 2006. Quantile Regression Forests. *Journal of Machine Learn-  
896 ing Research* 7, 17.

- 897 Michon, T., Castaings, W., 2017. Stratégie de calage du modèle hydrologique  
898 semi-distribué MORDOR-SD. Technical Report. TENEVIA.
- 899 Monteil, C., Zaoui, F., Le Moine, N., Hendrickx, F., 2019. Technical note:  
900 the caRamel R package for Automatic Calibration by Evolutionary Multi  
901 Objective Algorithm. Hydrology and Earth System Sciences Discussions  
902 2019, 1–16. URL: [https://www.hydrol-earth-syst-sci-discuss.net/  
903 hess-2019-259/](https://www.hydrol-earth-syst-sci-discuss.net/hess-2019-259/), doi:10.5194/hess-2019-259.
- 904 Nash, J., Sutcliffe, J., 1970. River flow forecasting through conceptual  
905 models part I — A discussion of principles. Journal of Hydrology  
906 10, 282–290. URL: [https://linkinghub.elsevier.com/retrieve/pii/  
907 0022169470902556](https://linkinghub.elsevier.com/retrieve/pii/0022169470902556), doi:10.1016/0022-1694(70)90255-6. number: 3.
- 908 Pappenberger, F., Pagano, T.C., Brown, J.D., Alfieri, L., Lavers, D.A., Berthet,  
909 L., Bressand, F., Cloke, H.L., Cranston, M., Danhelka, J., Demargne, J., De-  
910 muth, N., de Saint-Aubin, C., Feikema, P.M., Fresch, M.A., Garçon, R.,  
911 Gelfan, A., He, Y., Hu, Y.Z., Janet, B., Jurdy, N., Javelle, P., Kuchment, L.,  
912 Laborda, Y., Langsholt, E., Le Lay, M., Li, Z.J., Mannesiez, F., Marchan-  
913 dise, A., Marty, R., Meißner, D., Manful, D., Organde, D., Pourret, V.,  
914 Rademacher, S., Ramos, M.H., Reinbold, D., Tibaldi, S., Silvano, P., Sala-  
915 mon, P., Shin, D., Sorbet, C., Sprokkereef, E., Thiemig, V., Tuteja, N.K.,  
916 van Andel, S.J., Verkade, J.S., Vehviläinen, B., Vogelbacher, A., Wetter-  
917 hall, F., Zappa, M., Van der Zwan, R.E., Thielen-del Pozo, J., 2016. Hy-  
918 drological Ensemble Prediction Systems Around the Globe, in: Duan, Q.,  
919 Pappenberger, F., Thielen, J., Wood, A., Cloke, H.L., Schaake, J.C. (Eds.),  
920 Handbook of Hydrometeorological Ensemble Forecasting. Springer Berlin Hei-  
921 delberg, Berlin, Heidelberg, pp. 1–35. URL: [https://doi.org/10.1007/  
922 978-3-642-40457-3\\_47-1](https://doi.org/10.1007/978-3-642-40457-3_47-1), doi:10.1007/978-3-642-40457-3\_47-1.
- 923 Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Us-  
924 ing Bayesian Model Averaging to Calibrate Forecast Ensembles. Monthly

925 Weather Review 133, 1155–1174. URL: [https://doi.org/10.1175/](https://doi.org/10.1175/MWR2906.1)  
926 [MWR2906.1](https://doi.org/10.1175/MWR2906.1), doi:10.1175/MWR2906.1. number: 5.

927 Raynaud, L., Bouttier, F., 2016. Comparison of initial perturbation meth-  
928 ods for ensemble prediction at convective scale. Quarterly Journal of  
929 the Royal Meteorological Society 142, 854–866. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.2686>,  
930 doi:10.1002/qj.2686, doi:10.1002/qj.2686. number: 695.

932 Rouhier, L., Le Lay, M., Garavaglia, F., Le Moine, N., Hendrickx, F., Mon-  
933 teil, C., Ribstein, P., 2017. Impact of mesoscale spatial variability of cli-  
934 matic inputs and parameters on the hydrological response. Journal of Hy-  
935 drology 553, 13–25. URL: [http://linkinghub.elsevier.com/retrieve/](http://linkinghub.elsevier.com/retrieve/pii/S0022169417305012)  
936 [pii/S0022169417305012](http://linkinghub.elsevier.com/retrieve/pii/S0022169417305012), doi:10.1016/j.jhydrol.2017.07.037.

937 Saltelli, A. (Ed.), 2008. Global sensitivity analysis: the primer. John Wiley,  
938 Chichester, England ; Hoboken, NJ. OCLC: ocn180852094.

939 Saltelli, A., Annoni, P., 2010. How to avoid a perfunctory sensitivity anal-  
940 ysis. Environmental Modelling & Software 25, 1508–1517. URL: [https://](https://linkinghub.elsevier.com/retrieve/pii/S1364815210001180)  
941 [linkinghub.elsevier.com/retrieve/pii/S1364815210001180](https://linkinghub.elsevier.com/retrieve/pii/S1364815210001180), doi:10.  
942 1016/j.envsoft.2010.04.012. number: 12.

943 Schaake, J., Pailleux, J., Thielen, J., Arritt, R., Hamill, T., Luo, L., Martin, E.,  
944 McCollor, D., Pappenberger, F., 2010. Summary of recommendations of the  
945 first workshop on Postprocessing and Downscaling Atmospheric Forecasts for  
946 Hydrologic Applications held at Météo-France, Toulouse, France, 15-18 June  
947 2009. Atmospheric Science Letters 11, 59–63. URL: [http://doi.wiley.com/](http://doi.wiley.com/10.1002/asl.267)  
948 [10.1002/asl.267](http://doi.wiley.com/10.1002/asl.267), doi:10.1002/asl.267. number: 2.

949 Schefzik, R., Thorarinsdottir, T.L., Gneiting, T., 2013. Uncertainty Quan-  
950 tification in Complex Simulation Models Using Ensemble Copula Coupling.  
951 Statistical Science 28, 616–640. URL: [http://projecteuclid.org/euclid.](http://projecteuclid.org/euclid.ss/1386078881)  
952 [ss/1386078881](http://projecteuclid.org/euclid.ss/1386078881), doi:10.1214/13-STS443. number: 4.

- 953 Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bout-  
954 tier, F., Lac, C., Masson, V., 2011. The AROME-France Convective-  
955 Scale Operational Model. *Monthly Weather Review* 139, 976–991.  
956 URL: <http://journals.ametsoc.org/doi/abs/10.1175/2010MWR3425.1>,  
957 doi:10.1175/2010MWR3425.1. number: 3.
- 958 Sobol, I.M., 2001. Global sensitivity indices for nonlinear mathematical models  
959 and their Monte Carlo estimates. *Mathematics and Computers in Simulation*  
960 55, 271 – 280. URL: [http://www.sciencedirect.com/science/article/](http://www.sciencedirect.com/science/article/pii/S0378475400002706)  
961 [pii/S0378475400002706](http://www.sciencedirect.com/science/article/pii/S0378475400002706), doi:[https://doi.org/10.1016/S0378-4754\(00\)](https://doi.org/10.1016/S0378-4754(00)00270-6)  
962 [00270-6](https://doi.org/10.1016/S0378-4754(00)00270-6). number: 1.
- 963 Taillardat, M., Fougères, A.L., Naveau, P., Mestre, O., 2019. Forest-Based  
964 and Semiparametric Methods for the Postprocessing of Rainfall Ensem-  
965 ble Forecasting. *Weather and Forecasting* 34, 617–634. URL: [http://](http://journals.ametsoc.org/doi/10.1175/WAF-D-18-0149.1)  
966 [journals.ametsoc.org/doi/10.1175/WAF-D-18-0149.1](http://journals.ametsoc.org/doi/10.1175/WAF-D-18-0149.1), doi:10.1175/  
967 [WAF-D-18-0149.1](https://doi.org/10.1175/WAF-D-18-0149.1). number: 3.
- 968 Taillardat, M., Mestre, O., Zamo, M., Naveau, P., 2016. Calibrated Ensem-  
969 ble Forecasts Using Quantile Regression Forests and Ensemble Model Out-  
970 put Statistics. *Monthly Weather Review* 144, 2375–2393. URL: [http://](http://journals.ametsoc.org/doi/10.1175/MWR-D-15-0260.1)  
971 [journals.ametsoc.org/doi/10.1175/MWR-D-15-0260.1](http://journals.ametsoc.org/doi/10.1175/MWR-D-15-0260.1), doi:10.1175/  
972 [MWR-D-15-0260.1](https://doi.org/10.1175/MWR-D-15-0260.1). number: 6.
- 973 Talagrand, O., Vautard, R., Strauss, B., 1997. Evaluation of probabilistic predic-  
974 tion systems, in: *Workshop on Predictability, 20-22 October 1997, ECMWF,*  
975 *Shinfield Park, Reading.* pp. 1–26. URL: [https://www.ecmwf.int/node/](https://www.ecmwf.int/node/12555)  
976 [12555](https://www.ecmwf.int/node/12555).
- 977 Thiboult, A., Anctil, F., Boucher, M.A., 2016. Accounting for three sources of  
978 uncertainty in ensemble hydrological forecasting. *Hydrology and Earth Sys-  
979 tem Sciences* 20, 1809–1825. URL: [https://www.hydro1-earth-syst-sci.](https://www.hydro1-earth-syst-sci.net/20/1809/2016/)  
980 [net/20/1809/2016/](https://www.hydro1-earth-syst-sci.net/20/1809/2016/), doi:10.5194/hess-20-1809-2016. number: 5.

- 981 Thielen, J., Schaake, J., Hartman, R., Buizza, R., 2008. Aims, challenges and  
982 progress of the Hydrological Ensemble Prediction Experiment (HEPEX) fol-  
983 lowing the third HEPEX workshop held in Stresa 27 to 29 June 2007. *At-*  
984 *mospheric Science Letters* 9, 29–35. URL: [http://doi.wiley.com/10.1002/](http://doi.wiley.com/10.1002/asl.168)  
985 [asl.168](http://doi.wiley.com/10.1002/asl.168), doi:10.1002/asl.168. number: 2.
- 986 T.Roy, P., Ricci, S., Dupuis, R., Campet, R., Jouhaud, J.C., Fournier, C., 2018.  
987 BATMAN: Statistical analysis for expensive computer codes made easy. *The*  
988 *Journal of Open Source Software* 3, 493. URL: [http://dx.doi.org/10.](http://dx.doi.org/10.21105/joss.00493)  
989 [21105/joss.00493](http://dx.doi.org/10.21105/joss.00493), doi:10.21105/joss.00493. number: 21.
- 990 Van Schaeybroeck, B., Vannitsem, S., 2015. Ensemble post-processing  
991 using member-by-member approaches: theoretical aspects. *Quar-*  
992 *terly Journal of the Royal Meteorological Society* 141, 807–  
993 818. URL: [https://rmets.onlinelibrary.wiley.com/doi/abs/](https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2397)  
994 [10.1002/qj.2397](https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2397), doi:<https://doi.org/10.1002/qj.2397>. eprint:  
995 <https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.2397>.
- 996 Velázquez, J.A., Anctil, F., Ramos, M.H., Perrin, C., 2011. Can a multi-  
997 model approach improve hydrological ensemble forecasting? A study on  
998 29 French catchments using 16 hydrological model structures. *Advances in*  
999 *Geosciences* 29, 33–42. URL: <https://www.adv-geosci.net/29/33/2011/>,  
1000 doi:10.5194/adgeo-29-33-2011.
- 1001 Vidal, J.P., Martin, E., Franchistéguy, L., Baillon, M., Soubeyroux, J.M., 2010.  
1002 A 50-year high-resolution atmospheric reanalysis over France with the Safran  
1003 system. *International Journal of Climatology* 30, 1627–1644. URL: [http:](http://doi.wiley.com/10.1002/joc.2003)  
1004 [//doi.wiley.com/10.1002/joc.2003](http://doi.wiley.com/10.1002/joc.2003), doi:10.1002/joc.2003. number: 11.
- 1005 Wilks, D.S., 2018. Univariate ensemble post-processing, in: Vannitsem, S.,  
1006 Wilks, D.S., Messner, J.W. (Eds.), *Statistical postprocessing of ensemble fore-*  
1007 *casts*. Amsterdam, Netherlands. Elsevier, pp. 49–89.
- 1008 Zalachori, I., Ramos, M.H., Garçon, R., Mathevet, T., Gailhard, J., 2012. Sta-  
1009 tistical processing of forecasts for hydrological ensemble prediction: a com-

- 1010 parative study of different bias correction strategies. *Advances in Science and*  
1011 *Research* 8, 135–141. URL: <http://www.adv-sci-res.net/8/135/2012/>,  
1012 doi:10.5194/asr-8-135-2012.
- 1013 Zamo, M., 2016. Statistical post-processing of deterministic and ensemble wind  
1014 speed forecasts on a grid.
- 1015 Zamo, M., Mestre, O., Arbogast, P., Pannekoucke, O., 2014. A bench-  
1016 mark of statistical regression methods for short-term forecasting of  
1017 photovoltaic electricity production. Part II: Probabilistic forecast  
1018 of daily production. *Solar Energy* 105, 804 – 816. URL: <http://www.sciencedirect.com/science/article/pii/S0038092X14001601>,  
1019 doi:<https://doi.org/10.1016/j.solener.2014.03.026>.
- 1021 Zappa, M., Jaun, S., Germann, U., Walser, A., Fundel, F., 2011. Su-  
1022 perposition of three sources of uncertainties in operational flood fore-  
1023 casting chains. *Atmospheric Research* 100, 246–262. URL: <https://linkinghub.elsevier.com/retrieve/pii/S016980951000342X>, doi:10.  
1024 [1016/j.atmosres.2010.12.005](https://doi.org/10.1016/j.atmosres.2010.12.005). number: 2-3.  
1025