Response to reviewer hess-2019-166
The AquiFR hydrometeorological modelling platform as a tool for improving groundwater resource monitoring over France: evaluation over a 60 year period.

Anonymous Referee #1

Main Comments

Comments: - *The presentation of the AquiFR Hydrometeorological Modelling Platform is neither well described nor structured. Especially the first two paragraphs of Section 2, intended to be an introduction of the newly developed model, lacks a detailed description of the connection between different compartments. This part of the text should be closely connected to a meaningful (!) scheme of the AquiFR platform. I highly recommend replacing Figure 1 with a more detailed scheme and using this as a central theme guiding the reader through Section 2.*

Response: Thanks for this comment. In order to improve this section, we added a paragraph that presents the physical connection between the compartments as well as a new scheme (see Figure 1 below). Moreover, we replaced the former Figure 1 by a more detailed scheme with a detailed description of the time step during an AquiFR simulation (see Figure 2 below). The description of the AquiFR platform is modified in consequence in section 2 in the revised manuscript:

"The AquiFR hydrometeorological modelling platform allows representing the main hydrological processes occurring within the watersheds from precipitations to groundwater flows as shown in ***Erreur ! Source du renvoi introuvable.Erreur ! Source du renvoi introuvable.****. AquiFR accounts for spatial heterogeneity by using different spatial scales. The atmospheric forcing from SAFRAN and the estimation of the surface water budget fluxes by SURFEX are provided on an 8 km resolution grid. The SAFRAN meteorological analysis (Quintana Segui et al., 2008) provides hourly precipitation (rainfall and snowfall), temperature, relative air humidity, wind speed and downward radiations. The SURFEX land surface model (Masson et al., 2013) needs these atmospheric variables to solve the energy and surface water budget at the land-atmosphere interface at a 5-minutes time step. SURFEX estimates the spatial partition of the flow between surface runoff and groundwater recharge. It accounts for different soil and vegetation types and uses a diffusion scheme to represent the transfer of heat and water through the soil. The soil in SURFEX is represented by a multilayer approach. Its depth varies according to vegetation (in France from 0.2 to 3m) and is partly accessible to plant roots. Deep soil infiltration constitutes groundwater recharge flux. Surface runoff can occur according to saturation excess or infiltration excess.*

The simulation of the watersheds depends on its hydrogeologic characteristics. For sedimentary basin, these two fluxes are transferred to the MARTHE (Thiéry, 2015) or EauDyssée (Saleh et al., 2013) groundwater models. These models simulate transfer to the unsaturated zone, groundwater flows within and between the aquifer layers, transfer of surface runoff to and within rivers, and river-aquifer exchanges. They also account for the numerous groundwater abstractions within the river basins. The temporal resolution is daily and the spatial resolution varies from 100 m to a maximum of 8000 m. The depth of the deepest aquifer layer can reach locally about 1000 meters.

Karstic aquifer systems are simulated through a conceptual reservoir modelling approach using the EROS software (Thiéry, 2018). Each karstic system is represented by a lumped-parameter reservoir model solved at a daily time scale. Conceptual approaches are preferred for simulating karstic systems since their heterogeneities make it difficult to use a physically-based approach. EROS uses the *daily precipitation, snow, temperature and potential evapotranspiration provided by SAFRAN to compute karstic spring flows.*

Technically, the AquiFR hydrogeological modelling platform was developed using the OpenPALM coupling system (Buis et al., 2005; Duchaine et al., 2015). OpenPALM allows the easy integration of high-performance computing applications in a flexible and scalable way. It was originally designed for oceanographic data assimilation algorithms, but its application domain extends to multiple scientific applications. In the framework of OpenPALM, applications are split into elementary components that can exchange data. The AquiFR platform is an OpenPALM application that currently gathers 5 components. ***Erreur ! Source du renvoi introuvable.*** *Figure 2 shows the linkage between these components and the workflow of an AquiFR run. In this version 1.2 of AquiFR, no feedback from groundwater to the soil is taken into account. Therefore, a preliminary step illustrated by Figure 2****Erreur ! Source du renvoi introuvable.****a is to estimate groundwater recharge and surface runoff with SURFEX accounting for the atmospheric forcing from SAFRAN prior to an OpenPALM run. This preliminary step gives access to 60 years of daily groundwater recharge and surface runoff on a regular 8 km resolution over all the French metropolitan area.*

*These water fluxes are then accessible by the OpenPALM application that includes the three hydrogeological modelling components, the pre-processing component, and the post-processing component as shown in Figure 2****Erreur ! Source du renvoi introuvable.****b. All these components exchange data during the parallel execution of a single OpenPALM run. At each daily time step, a first pre-processing component retrieves both the atmospheric forcing and the SURFEX groundwater recharge and surface runoff at the beginning of the time step. Then, the EauDyssée, MARTHE and EROS modelling software compute the evolution of the simulated hydrogeological variables for the current time step for each groundwater model independently. Then, a last post-processing component synchronizes the simulation (it waits until all the models have ended their computations for the current time step) and collects the individual outputs of each model to write comprehensive outputs for the entire domain. At last, a signal is sent by the post-processing component in order to allow the platform to compute the next time step.*

The use of OpenPALM allows running each instance of the models in parallel over several processors. The 60 year simulation presented in this study needs approximately 1.5 days of computation time on a high-performance computer. The following subsections present a brief description of the components integrated within the OpenPALM application in AquiFR."

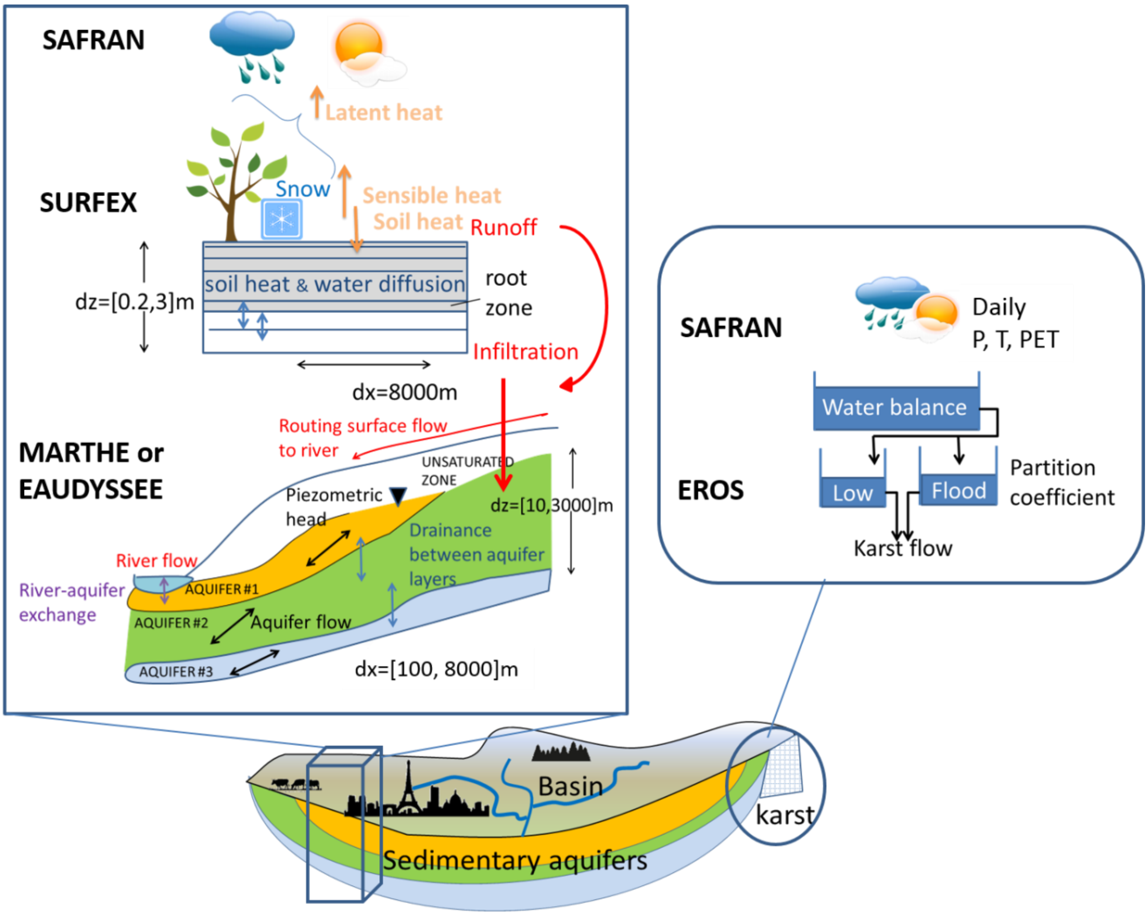

Figure 1: Scheme of the AquiFR physical system. The simulation of the watersheds depends on its hydrogeologic characteristics. For sedimentary basins, the transfer of water within the watersheds is estimated by MARTHE or EauDyssée. It accounts for flows in the unsaturated zones, to (red thin arrow) and in the rivers, in (black arrows) and between (blue arrows) aquifer layers, as well as the exchange between the river and the aquifer (purple arrow). The temporal resolution is daily and the spatial resolution varies from 100 m to a maximum of 8000 m. The depth of the deepest aquifer layer can reach locally about thousand meters. The 8 km spatial partition of the flow between surface runoff and groundwater recharge (red thick arrows) is estimated by the SURFEX land surface scheme. It solves the water and energy budget at a 5 minutes time step. It accounts for the local type of vegetation and soil, the presence of snow, and a multilayer soil that can reach a depth of 3 meters. The atmospheric forcing is provided by SAFRAN. For the karstic systems, the EROS conceptual model is used. It represents each karstic system as lumped basins based on a reservoir approach at a daily time scale. The incoming atmospheric forcing is provided by SAFRAN.

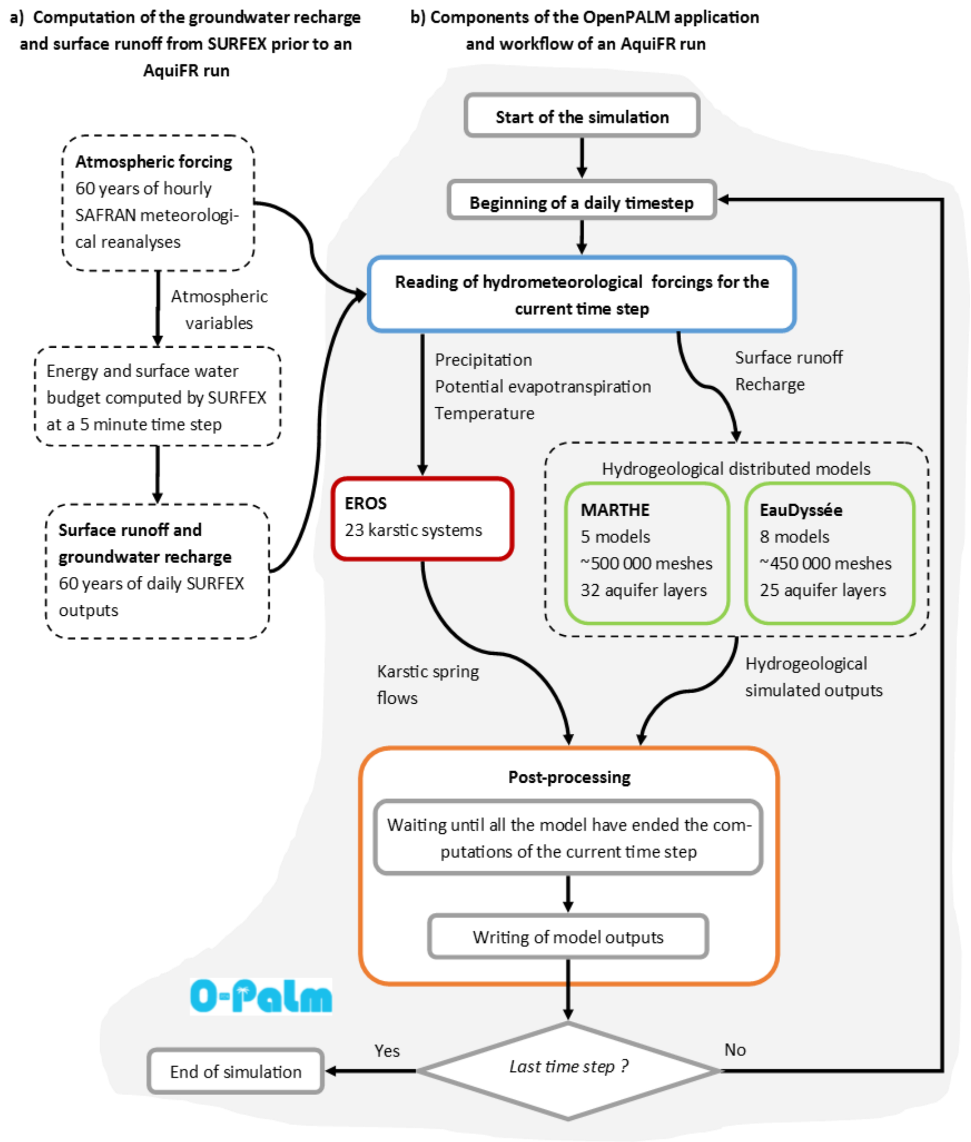

Figure 2: Scheme of the numerical implementation of AquiFR. (a) SAFRAN and SURFEX are run separately, as well as the processes that extract the daily surface runoff and groundwater recharge at 8 km resolution on a daily time step over the full 60 year period. (b) The components implemented within the coupling system O-Palm are presented. Pre-processing in blue gives access to the surface runoff and groundwater recharge as well as atmospheric forcing to the 3 groundwater models for the current time steps. Then, each hydrogeologic software runs all of their models for the current time step. The fluxes and state variables are then transferred daily to the post-processing, that writes the model outputs and manage the following time step.

Comment: - *"SURFEX is a modelling platform aiming to simulate the water and energy fluxes at the interface between the surface and the atmosphere" (Page 6, Line 5); "MARTHE embeds single to multilayer aquifers, hydrographic networks and the exchanges with the atmosphere (rainfall, snow and evapotranspiration) for the computation of the soil water balance" (Page 7, Line 24); "Snow accumulation, snow melting and pumping is taken into account" (EROS software; Page 7, Line 24) How do you deal with redundant parameters and processes which originally are elements in several of the models (e.g. evapotranspiration)?*

Response: Indeed, some information was missing. The MARTHE hydrogeological software includes different options that can be used to generate surface runoff or groundwater recharge. It includes its own computation of the soil water balance, including evapotranspiration, surface runoff and recharge. It can also directly receive surface runoff and recharge from an independent model, that is SURFEX in our case. This is this second option that is used in AquiFR, and this is now stated explicitly in the text.

The EROS software is not connected to SURFEX, and is directly connected to SAFRAN, this is now more clearly explained in the new paragraph of section 2 and appears clearly in Figure 1.

Comment: - *The SURFEX modelling platform: You are using the SURFEX model to calculate groundwater recharge and surface runoff. How do you address the specific karstic features (e.g. Epikarst, fast recharge components) in your model?*

Response: Specific karstic features are not taken into account in the SURFEX land surface model. Epikarst and fast recharge components could affect the simulation of karstic flows in SURFEX. This is why EROS is not connected to SURFEX and instead uses directly the atmospheric forcing from SAFRAN. The new paragraph in section 2 and the additional scheme better present the multilayer aspect of SURFEX and the main characteristics of the way the runoff and infiltration are computed. A few words about this are now added to the manuscript (section 2.2):

"The soil column thickness represented in each 8 km resolution grid cell varies from 20 centimeters to 3 meters according to the land cover in France and mostly corresponds to the root zone layer (Decharme et al., 2013). Thus, the recharge provided by SURFEX is the vertical flux leaving the bottom of the soil column of each grid cell. Further details on ISBA can be found in Decharme et al. (2013)."

Comment: - *Why do you present the quality criteria in section 3 (Results)? I would like to have more information about the evaluation of the model quality: a) general descriptions of the applied criteria, b) information about the calculation (equations) and references: e.g. How do you define bias and how do you exclude the bias from the calculation of the normalized RMSE?*

Response: In order to clarify the quality criteria used in section 3, a new Methodology section is now included in the revised manuscript. This methodology section includes 3 subsections:

3.1 The regional models implemented in the AquiFR platform

3.2 Calibration of the hydrogeological models

3.3 Evaluation criteria of the 60 years long-term simulation

This last subsection includes a general description of the applied criteria, that is bias, Nash-Sutcliff coefficient, normalized RMSE bias-excluded, and SPLI indicator. This new section is presented at the end of the present document.

Comment: - *The Numbering of the sections should be adapted. Section 4 is entirely missing*.

Response: It is now corrected in the revised manuscript.

Secondary Comments

Comment: *Introduction: The beginning of the Introduction has been kept general. I would like to have more information on "but is still poorly known" (Page 2, Line 2) and I do not understand what you mean by "Groundwater is indeed located at some depth below the soil" (Page 2, Line 2).*

Response: We agree about this. The beginning of the introduction was changed in order to be more explicit about the context:

"Groundwater is the most important freshwater resources on Earth. It is widely used for drinking water, agricultural, and industrial use. Knowing the spatial and temporal evolutions of the groundwater and being able to predict its future evolution over short to long term periods are essential to manage water resources and anticipate climate change impacts. However, groundwater is characterized by a strong spatial heterogeneity making its monitoring difficult. Thus, it is mostly monitored through well networks that can give information only at specific locations (Aeschbach-Hertig and Gleeson, 2012; Fan et al., 2013). Remote sensing gravimeters can provide large scale estimates of groundwater storage changes (Long et al., 2015) but it is not suited for regional scale studies (Longuevergne et al., 2010). Therefore, modeling can be a useful tool to provide meaningful information on the groundwater resources (Aeschbach-Hertig and Gleeson, 2012) at different spatial scales and different temporal periods in the past or in the future."

Comment: *Page 2, Line 28: "[: : :] as separate layers discretized using a 5 km resolution grid [: : :]" – The word separate is confusing here. Please, rephrase this sentence and maybe the next one as well. Also point out that the different layers are not connected to each other but to the river network.*

Response: The authors agree that this part was not clear. It is modified in the revised manuscript as follows: *"In the United Kingdom (UK), Pachocka et al. (2015) used a numerical model to compute the piezometric head evolution of the three most important UK unconfined aquifers using a finite difference scheme. These three unconfined aquifer basins were discretized into a 5 km resolution grid and connected to a river network. The model was tested against 37 gauging stations distributed across the country."*

Comment: *Page 3, Line 25: I do not understand how the AquiFR project can provide monitoring of groundwater resources. Please, elaborate this.*

Response: AquiFR is expecting to help monitoring the groundwater resources since it is planned to be used on real-time, in order to provide each day a present state of groundwater on the simulated domain. The sentence was modified as follows:" *In such context, the AquiFR project was initiated to capitalize these developments in order to provide real time monitoring (Coustau et al., 2015); and forecasts (Singla et al., 2012; Thirel et al., 2010) of groundwater resources in France, as well as long-term reanalyses and future projections"*

Comment: *The SAFRAN meteorological reanalysis: I am not sure if "analyses eight variables" (Page 5, Line 26) and "analyses each atmospheric variables" (Page 5, Line 30) are suitable expressions. Although, Quintana-Seguí et al. (2008) uses the same expression. I think estimates or calculates would be more suitable here.*

Response: The authors agree. We changed "analyses" by "estimates".

Comment: *Page 6, Line 6: The sentence needs to be rephrased.*

Response: The sentence is now: "*SURFEX is built to be coupled to forecast and climate models. It includes databases and interpolation scheme and several physical options that allows to use it at different spatial and temporal scales*"

Comment: *Page 6, Line 9: "[: : :] SURFEX is used in offline mode [: : :]" If this part of the sentence is useful information for the reader, elaborate it. Otherwise, I would delete it.*

Response: This part was modified. The new sentence gives more information on the coupling between SURFEX and the aquifers. The part "offline mode" is deleted: "*In the present study, no bidirectional coupling between the soil of SURFEX and the aquifers is taken into account. Thus, a one-*

way coupling from the soil of SURFEX to the aquifer is taken into account in order to provide groundwater recharge and surface runoff to the AquiFR platform"

***Comment**: Page 7, Line 18: "hydro-climatic rainfall-river flow-piezometric head distributed model" is the direct translation of the expression used in Thiéry (2018a). Don't you think "reservoir model" is also a correct description of the model?*

Response: We agree with the reviewer. We replaced this expression by "distributed reservoir model"

Comment: *Page 9, Line 22: Please, erase the brackets and use a different expression (e.g. wise versa) instead.*

Response: The text is now "*A positive value means that the simulation overestimates the mean piezometric head with respect to the observation while a negative value means the opposite*."

***Comment:** Page 12, Line 21: Please, consider rephrasing the sentence "They were kept [: : :]"*

Response: The new sentence is now: "*The present study used the same observed datasets to evaluate the river discharges simulated with the AquiFR platform over the 1958-2018 period*"

***Comment:** Page 13, Line 1: Please, consider splitting this sentence.*

Response: This sentence is now split: "*Regarding the results of Figure 15c, for rivers in continuous aquifers, 27% of the NSE scores are greater than 0.7. Moreover 58% of these NSE scores are greater than 0.5 while 22% are negatives.*"

***Comment:** Page 14, Line 11: Why did you (re)calibrate a few of the catchment/karst models and others not? You are proposing an inverse calibration tool - How did you calibrate the models after connecting them to SURFEX?*

Response: All the karst models were calibrated using the SAFRAN atmospheric forcing. Almost all the distributed models included in AquiFR were calibrated using the SAFRAN-SURFEX fluxes. Some models were not recalibrated either because the results were good enough with the new fluxes, or because additional changes are expected. Each distributed model was developed independently and calibrated with different periods of calibration. The calibration was based on trial-and-error method over the same period that was used to develop them. To better address such question, a subsection on the calibration is now presented in the new section "3. Methodology" at the end of the present document, and the Table 1 provides information on the calibration.

***Comment:** Page 14, Line 18: What do you mean by "However, the SIM tool uses coarse hydrogeological modelling [: : :]"?*

Response: In SIM, only few aquifers are simulated explicitly with the MODCOU hydrogeological model: the Seine and the Rhône aquifer basins (Habets et al., 2008). These two models correspond to outdated versions that have not been upgraded since. Thus, in SIM, the Seine aquifers are described by only 3 aquifer layers rather in AquiFR, 6 layers are accounted for as well as the river loss to the aquifer. More details regarding this point is now added in the article: *" However, the SIM tool uses coarse hydrogeological modelling with less aquifer layers or no river loss to the aquifer. It mainly focuses on operational forecasts of river flows and soil humidity."*

Comment: *Figure 2/3: Karst springs or Karst instead of Karsts*

Response: It is corrected.

MINOR COMMENTS AND TYPOGRAPHICAL ERRORS

Comment: - *Page 1, Line 27: to compute*

Comment: - *Page 1, Line 28: that is used*

Comment: - *Page 2, Line 1: on Earth*

Comment: - *Page 2: Line 15: research organizations (?)*

Comment: - *Page 2, Line 27: United Kingdom (UK)*

Comment: - *Page 3, Line 2: delete though*

Comment: - *Page 3, Line 13: on a global scale*

Comment: - *Page 3, Line 17: led by the*

Comment: - *Page 3, Line 18: delete Indeed*

Comment: - *Page 3, Line 25: the AquiFR project was initiated*

Comment: - *Page 3, Line 27: numerical modelling (?)*

Comment: - *Page 4, Line6: I am not sure if reported on the present study is a suitable expression: presented by?*

Comment: - *Page 4, Line 20: period. In*

Comment: - *Page 5, Line 26: eight variables: rainfall, snowfall*

Comment: - *Page 5, Line 26: air temperature and relative humidity 2 m (above ground) and wind speed 10 m above ground.*

Comment: - *Page 5, Line 29: two rain gauges. SAFRAN*

Comment: - *Page 6, Line 2: zone. Further*

Comment: - *Page 6, Line12: temporal*

Comment: - *Page 6, Line18: gathers numerical*

Comment: - *Page 6, Line 23: Horizontal groundwater flow (?)*

Comment: - *Page6, Line 24: leakage. Therefore*

Comment: - *Page6, Line 29: coupled to*

Response: Thanks for all these corrections. They are now corrected in the revised manuscript.

Comment: - *Page 7, Line 15: Thiéry et al, 2018 – a or b?*

Response: Thiéry et al., 2018 is the correct citation ; it corresponds to the reference Thiéry, D., Amraoui, N. and Noyer, M.-L.: Modelling flow and heat transfer through unsaturated chalk – Validation with experimental data from the ground surface to the aquifer, J. Hydrol., 556, 660–673, doi:10.1016/j.jhydrol.2017.11.041, 2018

The citations Thiéry 2018a, and Thiéry 2018b correspond to the references with only Thiéry in single author:

Thiéry, D.: Logiciel ÉROS version 7.1 - Guide d'utilisation. Rapport final, BRGM/RP-67704-FR, Orléans., 2018a.

Thiéry, D.: Modélisation hydrologique globale des débits de 23 sources karstiques avec le logiciel ÉROS. Rapport final, BRGM/RP-67723-FR, Orléans., 2018b.

Comment: - *Page 8, Line 25: Is there a number missing in the brackets?*

Response: yes, we wanted to gives the estimation of 16 mm/year in billion of m3 per year (that is 2.4 billion of m3 per year). Thank you for this correction.

Comment: - *Page 9, Line 16: observations at*

Comment: - *Page 9, Line 26: 2m and 4 m, respectively.*

Comment: - *Page 10, Line 2: with at least*

Comment: - *Page 11, Line 19: model input instead of inputs of the model*

Comment: - *Page 11, Line 29: delete one of the two dots*

Comment: - *Page 12, Line 4: shows*

Comment: - *Page 12, Line 8: which refers to the extreme rainfall event at the end of May 2016.*

Comment: - *Page 12, Line 11: Better: Figure 12 shows two plots comparing*

Comment: - *Page 12, Line 21: same here*

Comment: - *Page 13, Line 7: Nevertheless, some regions are*

Comment: - *Page 13, Line 9: than the other regions (cf. Fig. 5).*

Comment: - *Page 13, Line 25: It would also demand big resources of computational power.*

Comment: - *Page 13, Line 26: to simulate a*

Response: All these elements are corrected in the revised manuscript.

Comment: - *Page 14, Line 8: into account the*

Response: "into account the" instead of "into account in the"

Comment: - *Page 15, Line 13: more regional models?*

Response: "more regional model" instead of "more regional spatial model"

Comment: - *Page 15, Line 18: in progress*

Comment: - *Figure 15: (b) Somme*

Response: All these elements are corrected in the revised manuscript.

3 Methodology

3.1 The regional models implemented in the AquiFR platform

AquiFR aims at covering all groundwater resources in France. Figure 2 shows the main aquifers covering France classified by geological type as defined in the French hydrogeological reference system BDLISA (https://bdlisa.eaufrance.fr/). The current version of AquiFR gathers 13 spatially distributed models corresponding to regional single or multilayer aquifers (Table 1 and Figure 3).

Some regions are simulated by two spatialized models (Figure 3): the Somme and the Basse-Normandie basins are covered by MARTHE and EauDyssée models, and the chalk aquifer of the Seine basin is covered by both the EauDyssée Seine model and four EauDyssée sub-models (Marne-Loing, Marne-Oise, Seine-Eure, and Seine-Oise regional models, see Figure 4). This allows a multi-model approach, which can be useful for forecast and climate change impact studies. For these regions, the results presented in this paper correspond to the models that were considered as the best calibrated with the SURFEX fluxes. It corresponds to the four EauDyssée sub-models over the Seine basin and the Somme and Basse Normandie MARTHE models. Figure 3 also shows the 23 karstic systems (median catchment area of 99 km2) simulated by EROS (Thiéry, 2018b) as well as the hard rock aquifer in Britany that will be simulated using a hillslope model (Courtois, 2018; Marçais et al., 2017) and integrated in the near future.

Groundwater withdrawals are integrated as input data in the spatially distributed models. On annual average and with respect to the total surface area of the simulated domain, it corresponds to about 16 mm/year (2.4 billion of cubic meters per year) distributed in more than 16 000 grid cells. Data on groundwater pumping are provided by the regional water agencies on the basis of tax reporting. Pumping concerns drinking water, agriculture, and industrial use. The quality of the data set as well as its temporal extension varied for each regional modelling, although the latter does not exceed 20 years. Further details on regional models can be found in the references listed in Table 1. To extend the pumping estimation to the 1958-2018 period, a monthly mean annual cycle is used for the years without data. River routing is performed based on kinematic wave approach in MARTHE and by the RAPID model based on the Muskingum approach (David et al., 2011) in EauDyssée. River-groundwater exchanges are in both directions for all the models. Each regional model uses its own river network at its own resolution. Most of the simulated domains encompass the entire river basins corresponding to the simulated rivers. Only the Alsace and the Poitou-Charentes basins are partially represented. Therefore, they need to prescribe time dependent boundary conditions at the upstream of some rivers based on river flow observations. If the observed data don't cover the full period, the missing values are filled by the daily mean annual observed river flow. In the near future, the advantage to have the atmospheric forcing and surface fluxes over the entire domain will be used to estimate the upstream flow based either on a lumped-parameter rainfall-runoff model integrated in the MARTHE computer code or by the RAPID model using a fine scale river network covering all France.

3.2 Calibration of the hydrogeological models

The original hydrogeological regional models were developed independently most often based on stakeholder requests. The water budgets in these models were usually computed using less physical methods and atmospheric local data (precipitation and temperature) that differ from the physically-based approach using SURFEX and SAFRAN. As a result, in order to be consistent with the estimation of the groundwater recharge estimated by SURFEX, most of the regional models were recalibrated based on the SURFEX fluxes (Habets et al., 2017). This recalibration effort was not undertaken for the Alsace and Loire models since both of them will be soon updated and then recalibrated.

Periods of recalibration were the same as those initially used to develop and calibrate each model (see references in Table 1), in order to facilitate the comparison between the recalibrated models and the initial models. Hydrodynamic parameters, including hydraulic conductivities and specific yields, were modified based on hydrogeological expertise in order to obtain the best fit between observations and simulations. The calibration was made only on the piezometric heads, except for the MARTHE Somme model for which piezometric heads and riverflows were accounted for, and for the karstic systems with karst spring flows only. All the models were recalibrated using the same statistical criterias. A comparison between the initial water budget of the models and the SURFEX fluxes was performed as a first step to estimate the need for recalibration of each model.

Some models, such as the Seine EauDyssée model, were not recalibrated since they perform equally well with the use of the SURFEX fluxes (see Table 1). In contrast, the MARTHE Somme river basin model was characterized by an excess of surface runoff in the north and a deficit to the south. In order to compensate for this imbalance, the total runoff provided by SURFEX was split into surface runoff and groundwater recharge using the original water balance scheme of MARTHE. This water balance scheme is based on a reservoir for which parameters are calibrated in order to compute the main components of the surface water budget (Thiéry, 2014). Only one reservoir was used, enabling to modify the partition of the total runoff and to account for a delay on the groundwater recharge in order to mimic the impact of the deep unsaturated zone. This improved the simulation of the river flows using the SURFEX total runoff. Once the new partition was estimated, the permeability was recalibrated. The Somme basin is the only one for which only the total runoff from SURFEX was used. For the other basins, the estimation by SURFEX of the partition of the water fluxes between surface runoff and groundwater recharge was used. Overall, the performance of models using the fluxes from SURFEX are similar to the original version, although locally, they may be better or otherwise degraded.

For the karst system software EROS, the models were calibrated based on the SAFRAN atmospheric analysis by using an optimization of the statistical comparison between observed and simulated daily river flows.

More information about the method of recalibration is given in Habets et al. (2017).

3.3 Evaluation criteria of the 60 years long-term simulation

Statistical criteria are used to evaluate the long-term simulation. The bias allows evaluating the relative mean deviation between the observation and the simulation. It is calculated as follows:

$$BIAS = \frac{1}{n}\sum_{i=1}^{n}\left(X_{obs}(t) - X_{sim}(t)\right), \quad (1)$$

with n the number of observed values, $X_{obs}(t)$ and $X_{sim}(t)$ the observed and simulated values respectively at time t.

The Root Mean Square Error (RMSE) score allows estimating the differences between the observed and simulated values. It is often used to compare observed and simulated piezometric heads. However, the computation of the RMSE score is strongly affected by the biases. Therefore, we computed a RMSE bias-excluded score in order to better assess the simulation in terms of amplitude and synchronization. Moreover, this RMSE bias-excluded score is normed with respect to the observed standard deviation for each observation. It takes into account the differences of variability between the observed points and to better compare them with each other. This normed RMSE bias-excluded (NRMSE_BE) is expressed as follow:

$$NRMSE\_BE = \frac{1}{\sigma_{obs}}\sqrt{\frac{\sum_{i=1}^{n}[(X_{sim}(t)-\overline{X_{sim}})-(X_{obs}(t)-\overline{X_{obs}})]^2}{n}} \quad (2)$$

with $\overline{X_{sim}}$ the temporal mean of simulated values over the considered period and σ_{obs} the observed standard deviation.

The Nash-Sutcliffe model Efficiency score NSE (Nash and Sutcliffe, 1970) measures the variance between the observed and simulated values. It is often applied to compare observed and simulated river flows but can be used for other variables. Its sensitivity to high-frequency fluctuations makes its use for comparing groundwater levels less obvious. This criteria is equal to 1 when the model fits perfectly the observations. A NSE above 0.7 is generally accepted as a good estimate of the signal dynamic, however depending on the hydrogeological and climate context of the basin. A negative NSE means that the mean observed signal is a better predictor than the model. The NSE is calculated as follows:

$$NSE = 1 - \frac{\sum_{i=1}^{n}(X_{obs}(t)-X_{sim}(t))^2}{\sum_{i=1}^{n}(X_{obs}(t)-\overline{X_{obs}})^2}, \quad (3)$$

with $\overline{X_{obs}}$ the temporal mean of observed values over the considered period.

One way to evaluate the ability of the simulation to capture extreme events is to use the Standardized Piezometric Level Index (SPLI). The SPLI is an indicator used to compare groundwater level time series and to characterize the severity of extreme events such as long dry period or groundwater overflows (Seguin, 2015). Assessing the ability of the AquiFR modelling platform to reproduce this indicator is important since the main objective of this platform is to predict such extreme events in short-to-long terms hydrogeological forecasts for groundwater management. The SPLI indicator is based on the same principles as the Standardised Precipitation Index (SPI) defined by (McKee et al., 1993) to characterize meteorological drought at several time scales. First, monthly mean time series are computed from a time series of piezometric heads. Then, twelve monthly time series (January to December) are constituted over the N years of the time series period. For each time series of N monthly values, non-parametric kernel density estimation allows estimating the best probability density function (pdf) fitting the histogram of monthly values. At last, for each month from January to December, a projection over the standardized normal distribution using a quantile-quantile projection allows to deduce the SPLI for each value of the monthly mean time series of piezometric heads.

The SPLI values most often range from -3 (extremely low groundwater levels corresponding to a return period of 740 years) to +3 (extremely high groundwater levels). The SPLI allow representing wetter and drier periods in a similar way all over the French national territory.