Author's response for 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 represents the main hydrological processes occurring within the watersheds from precipitations to groundwater flows as shown in ***Erreur ! Source du renvoi introuvable.****. AquiFR accounts for spatial heterogeneity by using different spatial scales. The atmospheric fields 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) uses these atmospheric variables to solve the energy and surface water budget at the land-atmosphere interface at a 5 minute 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 the vegetation (in France from 0.2 to 3 m) 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 basins, these two fluxes are transferred to the MARTHE (Thiéry, 2015a) or EauDyssée (Saleh et al., 2013) groundwater models. These models simulate the transfer to the unsaturated zone, groundwater flows within and between the aquifer layers, the routing 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, 2018a). Each karstic system is represented by a lumped reservoir model solved at a daily time scale. Conceptual approaches are preferred for simulating karstic systems. Indeed, 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.*** *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 of SURFEX is taken into account. Therefore, a preliminary step illustrated by* ***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 ***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. 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 to 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 on 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."*

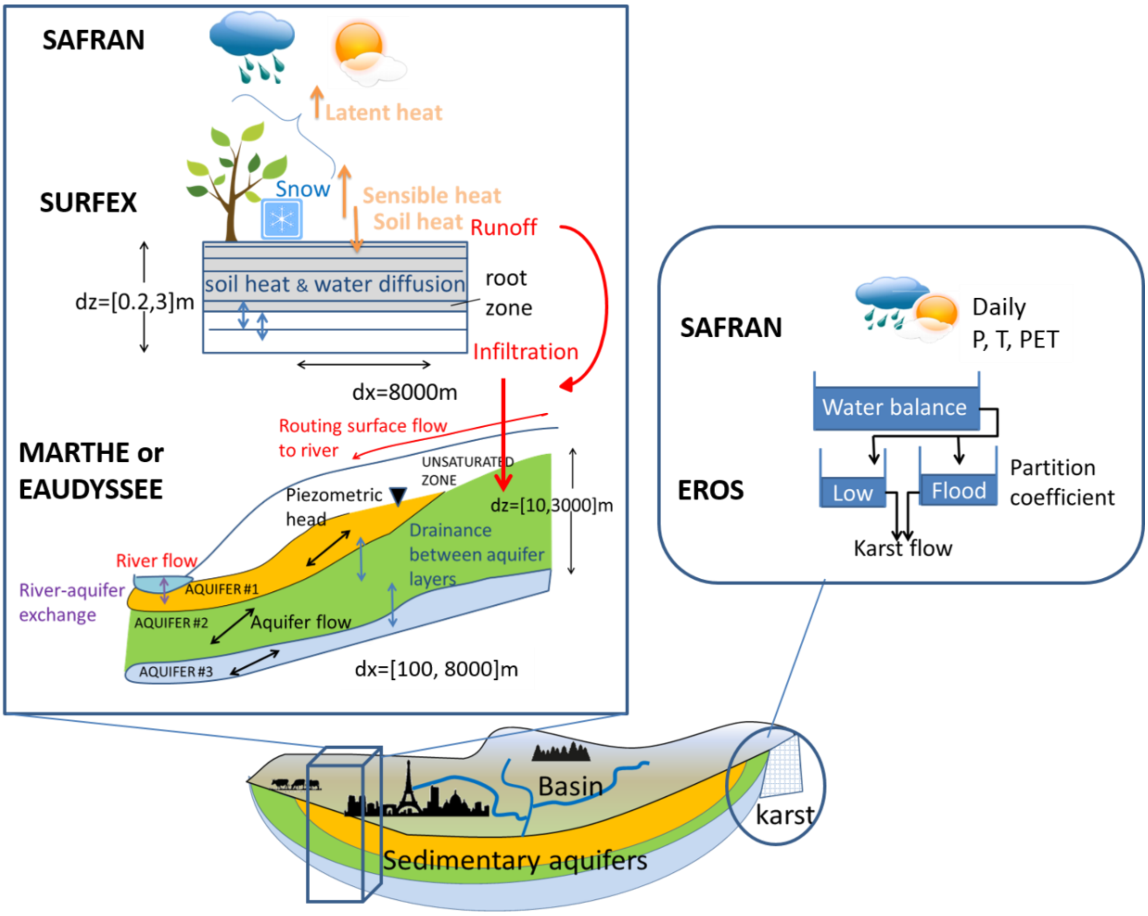


Figure 1: Scheme of the AquiFR physical system. The simulation of the watersheds depends on its hydrogeologic characteristics. For sedimentary basins, the transfer of water within the watersheds is estimated by MARTHE or EauDyssée. It accounts for flows in the unsaturated zones, to (red thin arrow) and 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.

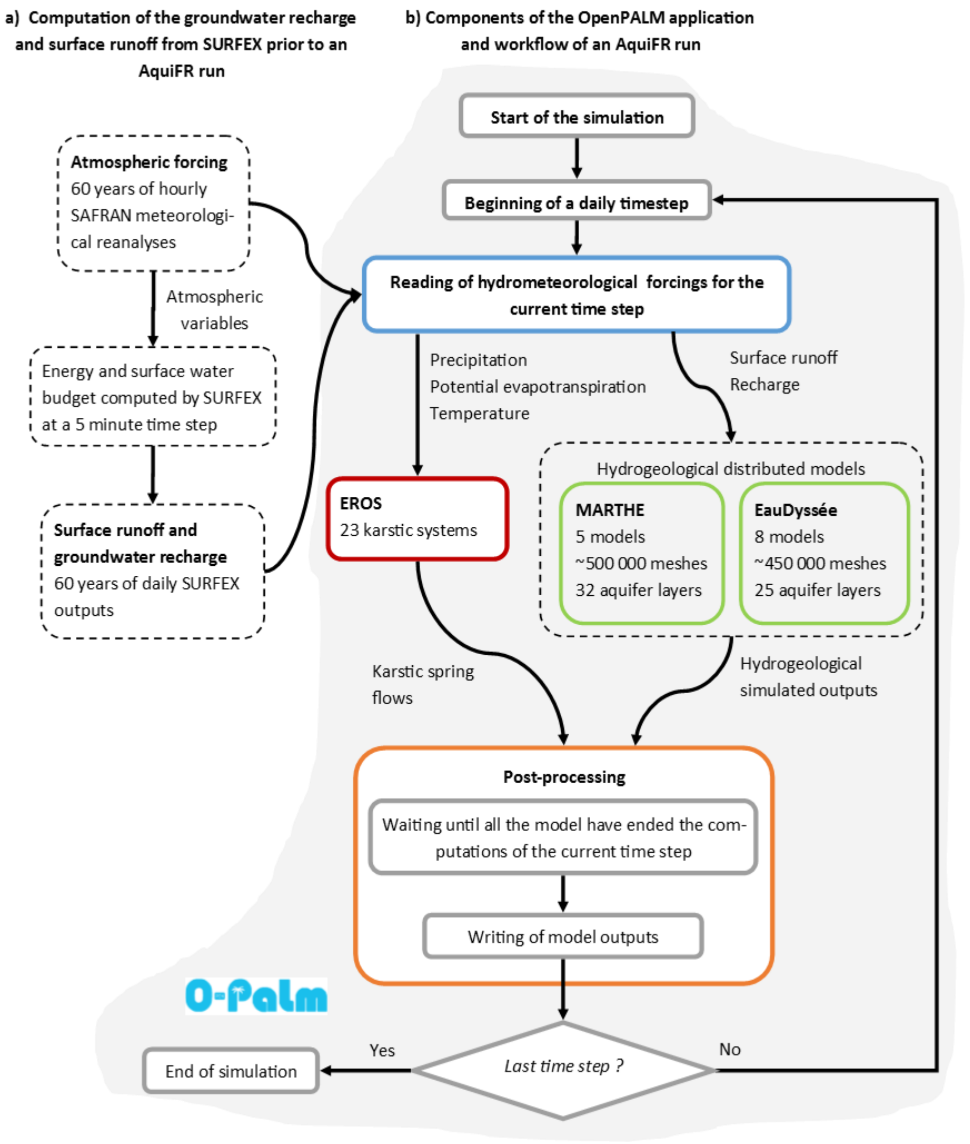


Figure 2: Scheme of the numerical implementation of AquiFR. (a) SAFRAN and SURFEX are run separately, as well as the processes that extract the daily surface runoff and groundwater recharge at 8 km resolution on a daily time step over the full 60 year period. (b) The components implemented within the 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.

Anonymous Referee #2

General comments

Comment: *1 It is not clear if the runoff is calculated with a common code and/or grid (P15L17 suggest that it is not the case), but nothing is explained about how the river routing is performed in the different basins.*

Response: Yes, indeed, some information was lacking. Surface runoff is computed using the land surface scheme of SURFEX on an 8 km resolution grid. This 8 km resolution grid corresponds to the grid provided by the SAFRAN atmospheric analysis. The surface runoff is then routed to the river by the hydrogeological models, with their own spatial resolution (varying from 100 m to 8 km). To be clearer, in the revised manuscript, the section 2 is entirely modified. It includes a new scheme (see Figure 1 of the present document) presenting the physical interaction between the modules and the main processes accounted for the estimation of the water flows, and on a new version of the former Figure 1 (see Figure 2 of the present document) that better presents the technical connection between the module. The computation of river routing is described in section 2.4 for Marthe and 2.3 for EauDyssée.

Comment: *2. Are the rivers connected bidirectionally with the groundwater? P6L28-29 suggests it can be done, but is it done?*

Response: Yes rivers are connected in both direction in the MARTHE and EauDyssée models. A sentence is added in the text (see answer above). This information is also now provided in section 2.3 (EauDyssée) and 2.4 (MARTHE) that presents the hydrogeological models and is shown in the new Figure 1.

Comment: *3. EROS are lumped models that simulate karst in a simpler way. This is reasonable. It is mentioned that AquiFR will be used for climate change studies, but it is not mentioned how the calibrations of these lumped models will hold in a changing climate.*

Response: A new section "3. Methodology" is added to the manuscript for describing the models, the calibrations and the statistical criterias used for the evaluation. A subsection of this new section "3. Methodology" is now devoted to the calibration of the models. It is now stated that "*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 riverflows*." This new section is presented at the end of this document.

It is true that part of the uncertainty of the impact of climate change on the karst systems is linked to the hydrological model and to its calibration. But it is beyond the scope of this article to discuss such uncertainty.

Comment: *4. It is not clear if there is a bidirectional coupling between the aquifer and the soil (SURFEX). P6L15 says it can be done. Figure 1 shows and arrow that goes from the post-processing to SAFRAN/SURFEX, but it is not clear what it means. Is there a bidirectional coupling between soil and aquifer? Is SURFEX just a forcing or at each time step it is updated with information coming from the aquifers?*

Response: Thanks to stress out this important point. Although capillary rise can be accounted for in SURFEX, in the current version of AquiFR, no bidirectional coupling between the aquifer and the soil is taken into account. This is no clearly stated in section 2: "*In this version 1.2 of AquiFR, no feedback from groundwater to the soil of SURFEX is taken into account. Therefore, a preliminary step illustrated by Figure 2a is to estimate groundwater recharge and surface runoff with SURFEX taking* *into account the atmospheric forcing from SAFRAN prior to an OpenPALM run".* In section 2.2 presenting SURFEX, it is now stated: "*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.* " . The former Figure 1 was modified to better explain what are the exchanged data between each modules within the AquiFR platform.

Comment: *5. It seems that all the models have been recalibrated in order to be able to use the recharge coming from SURFEX. However P13L9-11 confuses me on this point. Have the models been recalibrated in order to use SURFEX as forcing?*

Response: Yes, some information on the need of such calibration was stated page 4 lines 14-17, and is now even made clearer: "*the combined use of SURFEX and SAFRAN provides a consistent set of hydro-meteorological data over an 8 km resolution grid over France, including groundwater recharge and surface runoff from SURFEX, as well as potential evapotranspiration, precipitation, and temperature from SAFRAN. The use of these SURFEX 8 km resolution fluxes made necessary the recalibration of the hydrogeological models included in the platform*". Indeed, it was found that most often, there are some differences between the fluxes estimated by SURFEX and by the original water balance scheme using P/PET, mostly in terms of dynamic. Such differences affected some comparisons between observed and simulated heads, either positively or negatively. To give more information on this recalibration, a new subsection is now added in the new "Methodology" section 3 presented at the end of this document.

Comment: *6. It is not discussed if the recalibrated models forced by SURFEX perform better or worse than the same models, calibrated with P/ETP data and using P/ETP data as forcing. What is the impact of using SURFEX as forcing? Having a homogeneous forcing has value, but does it have downsides?*

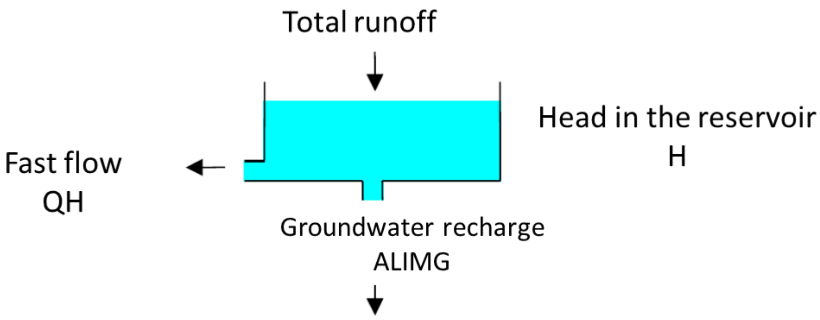
Response: This is a good question, and it is true that no information was given in the first version of the article. Overall, the statistical results obtained with SURFEX were similar to those obtain with the original version. A sentence is now added to stress out this point (see answer above), and some information is added in Table 1. Such result is indeed disappointing, as SURFEX is a more physical model and is more demanding computationally. It is one objective of the AquiFR project to improve such results.

Comment: *7. How good is the partition of surface runoff and drainage of SURFEX, in general? This is a key input for the whole system, but it is not validated, not even discussed in the paper. As far as I understand SURFEX may have some empirical parameters in order to determine surface runoff. Has this been calibrated? I would like to see a discussion (and data if possible) on the quality of the SURFEX recharge, as it is the main input for the hydrogeological models used in AquiFR.*

Response: The SURFEX partition of the surface runoff and drainage may differ from those calculated by the original models. However, it is difficult to distinguish which of the two is closest to the truth, since the truth is unknown, and as, after recalibration, the statistical results obtained by the two versions are similar. It was necessary to modify the partition between surface runoff and drainage only for the Somme basin by using the total runoff. Comments on this point are now added in section 3 (provided at the end of the text) as well as in Table 1. Detailed information is provided in a report accessible online (Habets et al., 2017).

Comment: *8. In the Somme river you don't use SURFEX's partition between runoff and recharge. It seems, that GARDENIA (no citation is provided) adds them together and makes a new partition. Why? How? This should be explained.*

Response: It is now clearly stated that the partition between surface runoff and groundwater recharge in the Somme basin was biased by SURFEX with an overestimation of surface runoff in the North and an underestimation in the South. GARDENIA is the name of the water balance scheme used originally in MARTHE. But, to avoid confusion, we removed the name, added a reference, and some explanations on how it works: " *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 approach (Thiéry, 2014), for which the parameters were calibrated. Only one reservoir was used, enabling to modify the partition of the surface runoff, and to account for a delay on the groundwater recharge in order to mimic the impact of the deep unsaturated zone*." In details, the reservoir we used is depicted below. H is the head in the reservoir, and is filled by the total runoff from SURFEX. THG is a time transfer coefficient and RUIPER is a partition coefficient that was calibrated. Using such reservoir, not only the partition of the flow between surface runoff and groundwater recharge is modified, but also the dynamic of the flow. This is important in the Somme basin since there is a deep unsaturated zone that is not simulated explicitly in the MARTHE model (see for instance Habets et al., 2010, Multi-model comparison of a major flood in the groundwater-fed basin of the Somme River (France), HESS)



$$ALIMG = H.dt/THG$$

$$QH = H.dt/(THG.RUIPER/H)$$

Figure 3 Partition of the total runoff of Surfex in the MARTHE Somme basin

The figure below presents the comparison of the river flow observed and simulated with Surfex with and without a new partition of the total runoff.

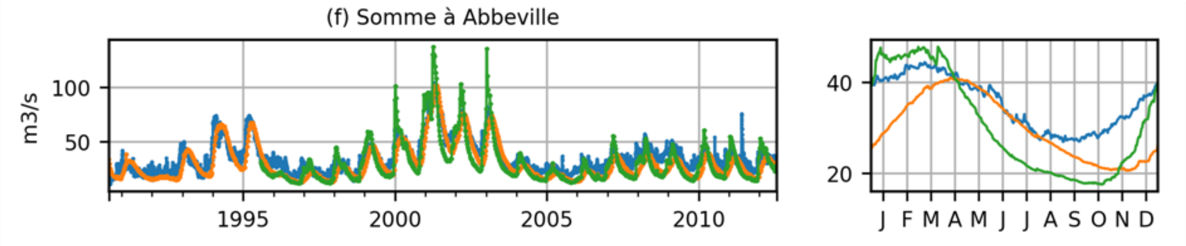


Figure 4 Comparison of the river flows at the outlet of the Somme basin between observations (blue) raw simulation with SAFRAN-SURFEX (orange) and simulation with the total runoff estimated by SAFRAN SURFEX and a partition of the surface runoff and drainage based on the MARTHE original water balance scheme

Comment: *9. Some applications of MARTHE need observed streamflow as an input (boundary conditions). How will you simulate climate change in this area? Why don't you use model streamflow? You simulate it, don't you? You should clarify this point.*

Response: In MARTHE, model streamflow are not simulated outside the simulated aquifer domain. Therefore, if the model does not encompass the entire river basin, boundary conditions are needed to impose flow on these rivers. We used observed streamflow in this version of AquiFR, but it is planned to use a new modelling method based on a lumped-parameter rainfall-runoff model to provide upstream river flows. The text is now modified: *"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 river routing model using a fine scale river network covering all France."*

Indeed, we have a hydrographic network over France, that is used for instance in SIM, but it has a 1km resolution, which is often not enough to match with rivers that are not fully simulated in the hydrogeological models.

In the climate change simulation we have done yet, the hypothesis is to have stable boundary conditions. Therefore, the flow of these not-fully simulated rivers, but also, the sea level, and the surface and groundwater abstractions are expected to be the same as in present day. Of course, it is clear that these hypotheses are not valid, and that the results only provide a first order impact of climate change.

I also have some questions on the cal/val procedure.

Comment: *1. Have you calibrated all the models over the same time period? If no, why? Due to data availability?*

Response: That is correct. The choice was made to calibrate on the same period used by the original model. This ensures that all the data needed are available, and allows comparing fairly with the original models. Please, report to the new section 3.2 provided at the end of the document.

Comment: *2. Do you validate all the models over the whole 60 year period? Do you use the calibration data also for validation? Do you only validate on independent data? The text is not clear to me on this regard and this is a very important issue. Not only for heads, also for streamflow. A model should not be validated using the same data it was used for calibration. If this cannot be avoided, it must be well justified.*

Response: The models are evaluated over the whole 60 year period. As described in the new Methodology section, the calibration procedure was done for each model using the same calibration period that were used to develop each model (see references in Table 1 and (Habets et al., 2017)). The new methodology section helps to better explain this. However, the validation presented in the article covers the 60-year period, restricted to the availability of the observation. Thus, the calibration and validation periods are different, but the validation period encompasses the calibration period. As all the models were not calibrated on the same period, and as the temporal availability of each measurement varies, it was the only way to have a full assessment of the whole AquiFR platform.

Comment: *3. You show the metrics you used for validation, but not for calibration. I guess that each model is calibrated differently, using different tools. Is this the case? This should be commented.*

Response: All the models were calibrated using the same statistical criteria: Efficiency, correlation and ratio for stream flow, and RMSE and biases for piezometric heads. As stated in the article, no automatic calibration tools were used, but only the skill of hydrogeological experts. These two points are now more clearly stated in the new 3. Methodology section : *"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 kartsic systems with karst spring flows only. All the models were recalibrated using the same statitiscal criterias."*

Comment: *4. You also validate using the NSE. Have you considered the KGE? Or even better, the non parametric version of the KGE (Pool et al, 2018)? The KGE allows to separate the contribution of the correlation, the bias and the standard deviation. The non parametric form makes less assumptions on the underlying data distribution so it can be used with different kinds of variables with less problems. Also, the non parametric form is less sensitive to extremes (so you would not need to calculated the sqrt of the streamflow, as you do). I guess it is too late to change this, but you should consider this in the future.*

Response: Thank you for this comment. Indeed, it is true that the KGE has some advantages compared to the NSE. This is a point that we will consider in the future. A sentence is added in the discussion : "*Some statistical scores using less assumptions on the underlying data distribution, such as the non-parametric variant of the Klunge-Gupta efficiency score, could be used to reduce the sensitivity to the extremes (Pool et al., 2018)."*

Comment: *5. Could you explain with more detail what is the NRMSE-BE? Have you substracted the mean and divided by the standard deviation and then calculated the RMSE? Have you removed the seasonal signal? A little bit more detail on this unusual metric should be provided*

Response: The details are now given in the new subsection 3.2 (see the methodology section at the end of this document).

Comment: *1. Which method do you use to calculate the standardized series? Is it parametric or non parametric?*

Response: The calculation of the Standardized Piezometric Level Index is similar to the calculation of the Standardized Precipitation Index (SPI) (McKee et al., 1993). The SPLI is an indicator used in the Monthly Hydrological Survey published each month. Details about its computation are given in Seguin, (2015). Considering a piezometric head time series of N years, the steps are the following:

- Step 1 : the monthly mean observed time series is computed
- Step 2 : constitution of twelve monthly time series (January to December) over the N year period. For each time series of N values, a non-parametric kernel density estimation (KDE) allows estimating the best probability density function (pdf) fitting the observed histograms. The SPI uses a gamma distribution, but time series of piezometric heads show a big variety of histogram. Therefore, the use of a KDE to estimate a pdf fitting the observed histogram is preferred.
- Step 3 : For each month from January to December, the adjusted pdf is projected over the standardized normal distribution using a quantile-quantile projection.

Figure 3 of the present document shows the procedure for the Omiécourt piezometer. The KDE helps to obtain a fit of the probability density function from the observed histogram. The cumulative density function is deduced, and a projection over the standardized normal distribution allows deducing the SPLI.

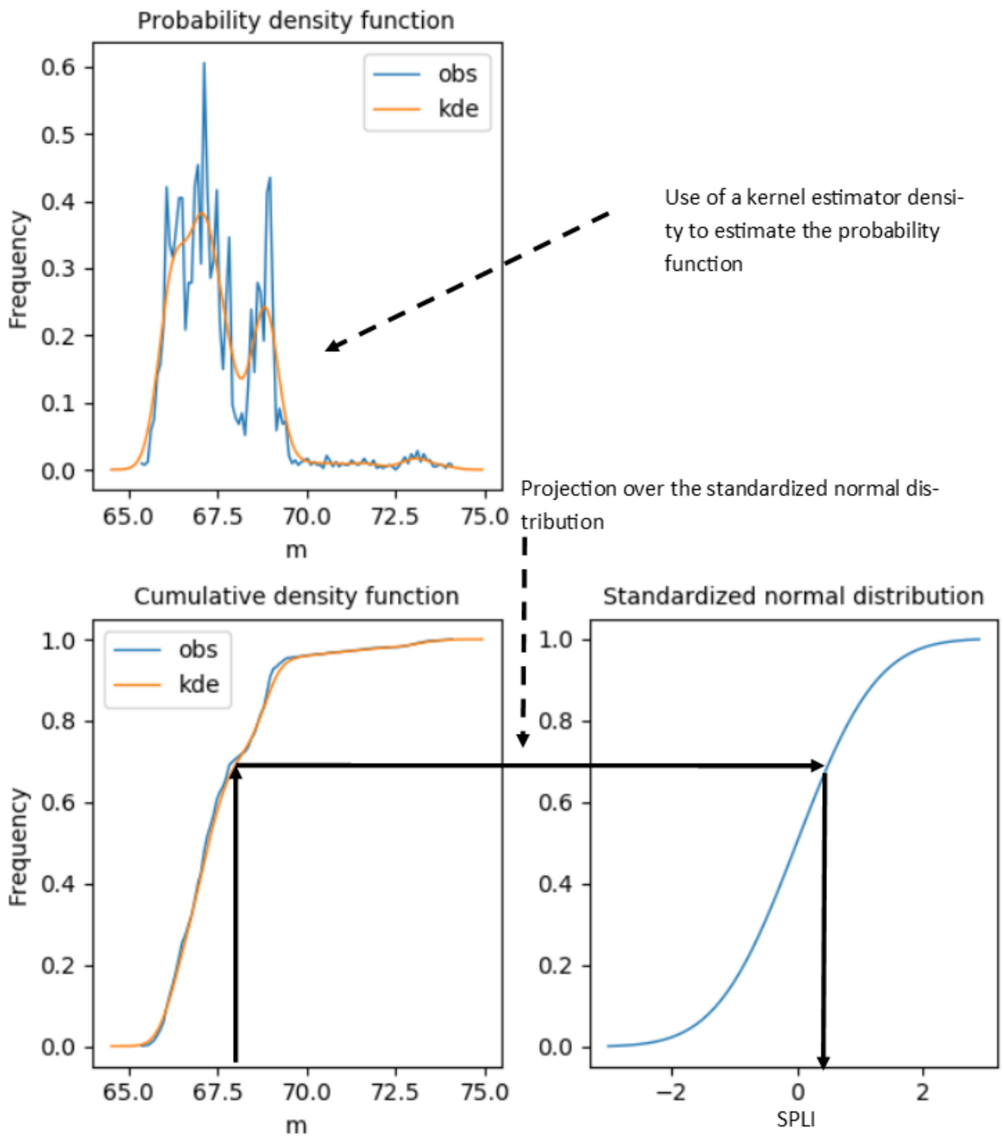


Figure 3: Computation of the SPLI for the Omiécourt piezometer. The probability density function is estimated using kernel density estimator from observed monthly piezometric head values. The estimated cumulative density function is then estimated from the fitted pdf, and a quantile-quantile projection on the standardized normal distribution allows computing the SPLI.

The authors agree that the presentation of the SPLI was not detailed enough. A new presentation is proposed in the new section 3.3 Methodology.

Comment: *2. If it is parametric, which distribution do you fit your data to? Does it fit to all areas equally well?*

Response: As the observed distribution depends on the area where the piezometer is located, a non-parametric KDE is used to estimate the best pdf to fit the observations (Seguin, 2015).

Comment: *3. Figure 10 shows the distribution of the different categories of the SPLI. But some of them are bimodal. I would expect a normal distribution as a standardized variable involves renormalizing the data to a normal distribution. Why these figures don't show a normal distribution?*

Response: Figure 3 of the present document shows the histogram of the observed values which is bimodal. The fitted pdf from KDE is also bimodal, and therefore its projection on the standardized normal distribution will keep this bimodal characteristic.

Comment: *I suggest adding a Methodology section where the cal/val procedure is presented and where the indicators (NRMSE-BE, NSE) and standardizations (SPLI) are presented.*

Response: A new methodology section is added to the revised manuscript, including a presentation of these indicators.

Comment: *Anthropic processes: You take pumping into account for some models. But the subsequent irrigation is not taken into account by SURFEX. Can you comment a little bit more on the current state of anthropic impacts in AquiFR and how this affects the results?*

Response: Most of the groundwater abstraction is used for drinking water. Crop irrigation is not taken into account in the present version of AquiFR. This process can be activated in SURFEX. However, it involves setting up strong hypotheses (where are the irrigated fields, what are the irrigated volume for each field, and when is the irrigation provided) that are beyond the scope of the purpose of the evaluation proposed in this paper. As for the bidirectional coupling between groundwater and SURFEX, this is an option that could be used in the future development of AquiFR.

Specific comments

* P2L6: "Thus, modeling is still a useful tool ...". Well, even with high resolution remote sensing data of storage in aquifers, models would still be useful, as they allow to connect aquifers with the rest of the system (soil, streams, etc.).

Response: The author agree. This sentence is changed into *"At these regional scales, modeling can be a useful tool to provide meaningful information on the groundwater resources (Aeschbach-Hertig and Gleeson, 2012)."*

* PL1: "3 groundwater flow software" -> 3 groundwater flow models.

Response: We try to distinguish software (numerical code) from models (regional models). For example, MARTHE is a hydrogeological modeling software, and 5 models have been developed using this software: the Somme, Poitou-Charentes, Nord-Pas-de-Calais, Basse-Normandie and Alsace models.

"3 groundwater flow software" is replaced by "3 hydrogeological modelling software"

* P4L20: "period.In" -> period. In

* P6L17: "gathersnumerical" should be separated.

* P6L29: coupledto should be separated.

Response: All these remarks are now corrected.

* P7L18: "set of rivers organized in sub-basins". Is this the basis of the acronym? I guess it is in its French form. Maybe it would be better to just put the French name.

Response: It is the english translation of the French acronym. We kept the french name, and as a consequence, we do the same for all the other acronyms that is SAM and MARTHE.

* P8L14: GARDENIA (citation needed). You should also explain how GARDENIA works.

Response: We decided not to provide the name GARDENIA and only to keep the reference to the simplified water balance scheme that is implemented in MARTHE. It is effectively true that this water balance scheme is the same in the GARDENIA software, but it is fully implemented in the MARTHE software and it is now part of MARTHE. The new sentence is: *"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)."*

* P9L17: observationsat should be divided.

* P11L8: It sensitivity -> Its sensitivity.

Response: All these remarks are corrected.

* P12L18-20: So you validate on the same stations you used for calibration. Do you?

Response: Yes we do.

* P12L33-P13L1. You calculate the sqrt to avoid an excecive influence of extremes. Is this the case? You should explain it.

Response: Yes. It is explained P12L17 to P12L20. We add a brief reminder about this.

* P13L10-11: Here you imply that you didn't recalibrate the models in order to use SURFEX as forcing. But earlier it seems you did. Did you?

Response: Yes we did. See previous answer and section 3.2

* P13L16-29: I would move this into the introduction.

Response: We decided to keep this part in the discussion in order to better highlight the choice of gathering several models in AquiFR as previously shown in section 2.

* P14L6: Which periods were used for calibration?

Response: Periods for calibration were those initially used for calibrated each model independently. This is now better explained in the new Methodology section included in the revised manuscript. This particular part of the discussion

* Fig1: What do the arrows mean? What fluxes are send to the post-processing and what is send back to SAFRAN/SURFEX? I would add labels to the arrows.

Response: The arrows illustrated the flux exchanges. The new proposed scheme better explains this.

* Fig7: Put the legend outside of the first plot.

Response: The legend is now outside of the first plot.

* Fig10: Being standardized values, I would expect a normal distribution, but on three cases it is bimodal.

Response: A new explanation of the SPLI indicators in the new methodology section helps to understand this.

* Fig11b: difficult to see the circles.

Response: The background SPLI map is now more transparent in order to better highlight the circles.

* Fig12: Why are the x-axis time scales so different? Is it related to data availability? Which is the calibration period?

Response: x-axis time scales are different because the axis limits are related to the observed data availability which is different for each karstic system. The calibration period corresponds to these axis limits. In order to be consistent with the evaluation of AquiFR, all the 60-year time serie is now shown.

New section 3 Methodology

3 Methodology

3.1 The regional models implemented in the AquiFR platform

AquiFR aims at covering all groundwater resources in France. **Erreur ! Source du renvoi introuvable.** 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 **Erreur ! Source du renvoi introuvable.**).

Some regions are simulated by two spatialized models (**Erreur ! Source du renvoi introuvable.**): 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. **Erreur ! Source du renvoi introuvable.** 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 time-dependent boudary conditions 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 based on tax reporting. Pumping concerns drinking water, irrigation, and industrial use. The quality of the dataset as well as its temporal extension varied for each regional modelling, although the later 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.

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 river routing 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 were usually computed using less physical methods and atmospheric local data (precipitation, temperature and potential evapotranspiration) that differ from the physically based approach using SURFEX and the SAFRAN analysis. As a result, in order to be consistent with the estimation of the groundwater recharge estimated by SAFRAN-SURFEX, most of the regional models were recalibrated based on the newfluxes (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 **Erreur ! Source du renvoi introuvable.**) 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 criteria. 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 in 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. It improved the simulation of the river flows using the SURFEX total runoff. Once the new partition was estimated, the aquifer 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 the models are similar with the original water balance fluxes and the ones simulated by SAFRAN-SURFEX, 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 calibration is given in Habets et al. (2017).

3.3 Evaluation criteria of the 60-year 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) 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 is strongly affected by the biases. Therefore, we computed a RMSE bias-excluded in order to better assess the simulation in terms of amplitude and synchronization. Moreover, this RMSE bias-excluded is normed with respect to the observed standard deviation for each observation in order to account for the differences of variability between the numerous wells to help spatial comparison or aggregation. This normed RMSE bias-excluded (NRMSE_BE) is expresed as follows:

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

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 coefficient 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. It 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.

The annual discharge ratio criterion helps to compare the mean simulated and observed river flows as follows:

$$Ratio = \frac{\overline{Q_{sim}}}{\overline{Q_{obs}}}, \quad (4)$$

with $\overline{Q_{sim}}$ and $\overline{Q_{obs}}$ the mean siumlated and observed river flows respectively.

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). It is currently used in France for the Monthly Hydrological Survey (MHS) (Office International de l'Eau, 2019). This MHS provides monthly information to the policymakers and the public on the hydrological state of groundwater. 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 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, a non-parametric kernel density estimator allows estimating the best probability density function fitting the histogram of monthly values. At last, for each month from January to December, a projection over the standardised normal distribution using a quantile-quantile projection allows deducing 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 allows representing wetter and drier periods in a similar way all over the simulated domain.

Relevante changes made in the manuscript

This part lists all the relevant changes made in the manuscript. It does not include corrections of typos as well as reformulations of sentences to improve the language that were included in the revised manuscript. The page, line and figure numbers correspond to the new revised manuscript:

P1L5 : " Longuevergne" was corrected in "Longuevergnes"

P1L11: " Ecole Normale Supérieure" was corrected in "Ecole normale supérieure".

P5L1 – P6L14: This paragraph of section 2 was modified for improving the description of the AquiFR platform. Former Figure 1 was replaced by two new figures: Figure 1 for the description of the physical processes taking place in AquiFR and Figure 2 showing the workflow of an AquiFR run.

P9L1 – P12L6: A new methodology section is added to present the regional models included in the AquiFR platform, the calibration of the models and the evalutation criteria used in the study.

P12L20: The number of gauging stations used to evaluate the simulated river flows of AquiFR was increased from 228 in the initial manuscript to 362. This increase concerns the EauDyssée Seine model as shown in the new Figure 15. With these new observations, the spatial distribution of river flow observations is more representative. The new observations were selected from the HYDRO database (http://hydro.eaufrance.fr/).

P12L27: After a carefull check of the results, a few percentages presented in the results section of the initial manuscript were adjusted in the revised manuscript. They are listed below and concern minor changes that do not affect the overall results. Here, the value of 40% of the biases under 2 m was modified to 42%.

P12L29: 61% was replaced by 62%

P13L7 – L8: A mistake was detected in the computation of the criteria of the Omiécourt piezometer. The new RMSE_BE value is 0.93 instead of 0.81 while the bias is equal to -0.86 m instead of -0.15 m. The comments were changed accordingly and the changes are reported in Table 2.

P13L25 – L26: Percentages for the spatial distribution of the SPLI criteria were slightly adjusted. The new sentence is: "*20% of the NSE are greater than 0.7 and 56% greater than 0.5, while 12% are lower than zero. In parallel, 65% of the correlation are greater than 0.7."*

The old sentence was: *"20% of the NSE scores are greater than 0.7 and 55% greater than 0.5, while 11% are lower than zero. In parallel, 64% of the correlation scores are greater than 0.7"*

P14L13: After a carefull check of the results, percentages of categorized SPLI for the simulated piezometers were corrected with respect to the original manuscript. The new sentence is: *"19% (29%) of the simulated (observed) piezometers are in normal conditions, 46% (31%) are in moderately wet conditions, 16% (13%) are in wet conditions, and 16% (18%) are in extremely wet conditions."*

The old sentence was: "*17% (29%) of the simulated (observed) piezometers are in normal conditions, 50% (32%) are in moderately wet conditions, 14% (12%) are in wet conditions, and 16% (18%) are in extremely wet conditions"*

P15L11: In order to be more representative of the results, the number of gauging stations for evaluating river flows was increased from 228 to 362. New gauging stations concern the Seine model. New results for NSE are given. The new sentence is: "*96% of the NSE using the square root of the* *daily karstic spring flows are greater than 0.7. Regarding the results of Figure 15c, for rivers in continuous aquifers, 34% of the NSE are greater than 0.7. Moreover 63% of these NSE are greater than 0.5 while 18% are negatives.*"

The previous results in the original manuscript were: "*80% of the NSE score using the square root of the daily river flows are greater than 0.8. Regarding the results of Figure 14c, for rivers in continuous aquifers, 27% of the NSE scores are greater than 0.7, 58% are greater than 0.5, and 22% are negatives.*"

P25 Figure 1: A new Figure 1 is proposed to present the physical processes included in AquiFR with the associated legend.

P26 Figure 2 : A new Figure 2 is proposed to show the linkage between the different components of AquiFR and the workflow of an AquiFR run.

P39 Figure 15 : Figure 15 was modified in order to account for the 362 gauging stations.

P41 Table 1: A new column is added for specifying if the models were recalibrated or not.

P58 Table 2: Statistical scores were corrected for the Omiécourt piezometer.

Marked-up manuscript

This marked-up manuscript does not include the figures but only the legends. The new figures can be found in the revised manuscript.

The AquiFR hydrometeorological modelling platform as a tool for improving groundwater resource monitoring over France: evaluation over a 60~~-~~-year period.

Jean-Pierre Vergnes[1], Nicolas Roux[2], Florence Habets[3,4], Philippe Ackerer[5], Nadia Amraoui[1], François Besson[6], Yvan Caballero[1], Quentin Courtois[7], Jean-Raynald de Dreuzy[7], Pierre Etchevers[6], Nicolas Gallois[8], Delphine J. Leroux[2], Laurent ~~Longuevergnes[7]~~Longuevergne[7], Patrick Le Moigne[2], Thierry Morel[9], Simon Munier[2], Fabienne Regimbeau[6], Dominique Thiéry[1], Pascal Viennot[8]

[1]BRGM French Geological Survey, Water Environment and Ecotechonologies Division, Orléans Cedex 2, France
[2]GAME/CNRM, Météo France, CNRS, Toulouse, France
[3]CNRS/Sorbonne University, UMR 7619 Métis, Paris Cedex 5, France
[4]Laboratoire de Géologie, Ecole ~~normale supérieure~~Normale Supérieure, PSL Research University, CNRS UMR 8538, 24 rue Lhomond, 75005 Paris, France
[5]LHYGES, UMR 7517 CNRS, EOST / Université de Strasbourg, France
[6]METEO-FRANCE/DCSC, Toulouse, France
[7]Géosciences Rennes, UMR 6118, CNRS, Université de Rennes I, Rennes, France
[8]Centre de Géosciences, MINES ParisTech, Fontainebleau, France
[9]CERFACS, Toulouse, France

Correspondence to: Jean-Pierre Vergnes (jp.vergnes@brgm.fr)

Abstract. The new AquiFR hydrometeorological modelling platform was developed to provide short to long-term forecasts for groundwater resource management in France. ~~The present~~This study aims to describe and assess this new tool over a long~~-term~~ period of 60 years. This platform gathers in a single numerical tool ~~different~~several hydrogeological models covering much of the French metropolitan area. Eleven aquifer systems are simulated through spatially distributed models using either the MARTHE groundwater modelling software or the EauDyssée hydrogeological platform. Twenty-three karstic systems are simulated by a lumped ~~models~~reservoir approach using the EROS software. AquiFR computes the groundwater level, the groundwater surface water exchanges, and the river flows~~ at multiple river gauging stations.~~. A simulation covering a 60~~-~~-year period from 1958 to 2018 is achieved in order to evaluate the performance of this platform. The 8 km resolution SAFRAN meteorological ~~reanalysis~~analysis provides the atmospheric variables needed by the SURFEX land surface model in order ~~tocompute~~to compute surface runoff ~~that are~~and groundwater recharge used by ~~all~~ the hydrogeological models. The assessment is based on ~~a wide range of selected~~more than 600 piezometers ~~as well as~~and more than 300 gauging stations corresponding to simulated rivers and outlets of karstic systems. For the simulated piezometric heads, ~~40~~42% and 60% of the absolute biases are lower than 2 m and 4 m respectively. The Standardized Piezometric Level Index (SPLI) was computed to assess the ability of AquiFR to identify extreme events such as groundwater ~~flooding~~floods or droughts in the long-term ~~simulations~~simulation over a set of piezometers used for groundwater resource management. ~~55~~56% of the Nash-Sutcliff ~~scores~~coefficient calculated between the observed and simulated SPLI time series are greater than 0.5. ~~Further work will focus on the use~~The quality of

~~this~~the results makes it possible to consider using the platform for ~~short-term to~~ real-time monitoring and seasonal forecasts ~~in an operational mode and~~ of groundwater resources as well as for climate change impact ~~assessment~~assessements.

1 Introduction

Groundwater is the most important freshwater resources on ~~the~~ Earth~~, but it~~. It is ~~still poorly known. Groundwater is indeed located at some depth below the soil~~widely used for drinking water, agricultural, and ~~therefore this resource~~industrial use. Knowing the spatial and temporal evolutions of groundwater and being able to predict its future evolution over short to long-term periods are essential to manage water resources management 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 ~~gravimetry~~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). ~~Thus~~Therefore, modeling ~~is still~~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.

An increasing number of numerical weather prediction models includes a representation of groundwater (Barlage et al., 2015; Sulis et al., 2018). Nevertheless, such representations are not detailed enough to be used to monitor or to forecast the groundwater resources. This is the reason why some dedicated approaches aim at providing groundwater level forecasts at the well scale with lumped models (Prudhomme et al., 2017) or neural networks (Amaranto et al., 2018; Dudley et al., 2017; Guzman et al., 2017).

At the regional scale, only a few modelling approaches use spatially distributed models to monitor and forecast the groundwater resource. Henriksen et al. (2003) presented the development of national hydrogeological models in Denmark aiming at gathering competencies from research ~~organisms~~organizations and water agencies and establishing a national overview of the present state and future trends of groundwater resources. An integrated groundwater/surface water hydrological model covering a spatial extension of about 43 000 km^2 with a 1 km grid resolution was then developed by making full use of the available data (Henriksen et al., 2003). The modelling system is composed of 11 regional sub-models. This model has been regularly updated, as reported by Højberg et al. (2013) who used local studies in relation with active stakeholders to include local data to improve the national model. The Danish model is planned to be used for real time monitoring (He et al., 2016) and climate change studies (Højberg et al., 2013).

In the Netherlands, national and regional water authorities decided to build the Netherlands Hydrological Instrument (NHI) which couples various physical models for all parts of the water system in order to support long-term plans for sustainable water use and safety under changing climate conditions (De Lange et al., 2014). The model was developed by research institutions but local knowledge has been adopted in cooperation with the national water boards (Højberg et al., 2013). It aims to be a model for long-term national policymaking and real-time forecasting for daily water management.

In the United KingdomsKingdom (UK), Pachocka et al. (2015) used a numerical model representingto compute the piezometric head evolution of the three most important UK unconfined aquifers as separate layersusing a finite difference scheme. These three unconfined aquifer basins were discretized usinginto a 5 km resolution grid with a finite-difference scheme. The layers are and connected to thea river network and the. The model was tested against 37 gauging stations distributed across the country. Good fit to the observations was obtained in a steady state run. This study seems to be the first step toward a system that will be used for water management studies and climate impact studies.

Another study covering a wide domain corresponding to the major part of the United States (6.3 billion of km^2) was carried out by Maxwell et al. (2015). A 3- dimensional hydrogeological model (ParFlow) was used at a 1 km grid resolution in steady state run. This model has four layers over the first meter of soil, and then a fifth layer from 1 to 100 meter depths. The computation time was one week on high-performance computer for a steady-state simulation. Thus, thoughwhile this study confirms the possibility of running a 3D groundwater model at fine resolution over a very large territory, it is still difficult to consider its application for operational water management purpose.

Other examples include the Texas Water Development Board that has implemented several sub-models to help monitor groundwater resources at the state scale (more than 500 000 km^2) (Texas Water Development Board, 2018)) or New Zealand where a nationwide groundwater recharge model is currently under development (Westerhoff et al., 2018).

In France, the hydrometeorological model SAFRAN-ISBA-MODCOU (Habets et al., 2008) that is used for long-term reanalyses (Vidal et al., 2010) as well as real time monitoring and forecast (Coustau et al., 2015; Singla et al., 2012; Thirel et al., 2010)(Coustau et al., 2015) and forecast (Singla et al., 2012; Thirel et al., 2010) includes an explicit representation of two aquifer systems. However, the representation of these aquifer systems is rather coarse and is mostly used to have a realistic representation of the river base flow (Rousset et al., 2004) rather than providing consistent information on groundwater resources. Vergnes et al. (2012) developed a hydrogeological model dedicated to climate modelling that was first applied over France and then at theon a global scale (Vergnes and Decharme, 2012). However, only a single layer at the resolution of approximately 10 km over France was considered. This approach is still too coarse to be used for groundwater management over France.

The need to have a national scale consistent representation of groundwater resources in France clearly appeared during the project EXPLORE2070 leadled by the French environment ministry that aimed at providing climate projections of the evolution of the water resource in France including groundwater (Stollsteiner, 2012). Indeed, severalSeveral regional hydrogeological models were used in this project, together with downscaled climate change projections. The results were difficult to analyse due to the differences in the way the surface water balance was calculated (either lumped-parameters model or soil-vegetation-atmosphere scheme), in the initialization methods, and in the way the evolutions were estimated. Moreover, in the meantime, several regional groundwater models were developed independently by research institutions in close relationship with the stakeholders for regional water management purposes or climate impact studies (Amraoui et al., 2014; Croiset et al., 2013; Douez, 2015; Habets et al., 2010; Monteil et al., 2010; Vergnes and Habets, 2018).

In such context, the AquiFR project was built to capitalize these developments in order to provide monitoring and forecasts of groundwater resources in France, as well as long-term reanalyses and future projections. The project associates research teams in hydrogeological, numerical, and atmospheric fields. It is funded by a national stakeholder in charge of the water resource, the French Agency for Biodiversity (AFB). The main idea of AquiFR is to include existing hydrogeological models developed with different groundwater modelling software and to connect them with real time atmospheric analysis and weather forecasts for producing relevant information for water resource management through a single numerical tool. This project also encourages new developments over areas where no groundwater models currently exist. To achieve these objectives the AquiFR hydrogeological modelling platform was developed. The main objectives of this paper are to describe this platform, to evaluate its performance against observations, and to prove its suitability and robustness for operational and research purposes.

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. The project associates research teams in hydrogeological, numerical modelling, and atmospheric fields. It is funded by a national stakeholder in charge of the water resource, the French Agency for Biodiversity (AFB). The main idea of AquiFR is to include existing hydrogeological models developed with different groundwater modelling software and to connect them with real time atmospheric analysis and weather forecasts for producing relevant information for water resource management through a single numerical tool. This project also encourages new developments over areas where no groundwater models currently exist. To achieve these objectives the AquiFR hydrogeological modelling platform was developed. The main objectives of this paper are to describe this platform, to evaluate its performance against observations, and to prove its suitability and robustness for operational and research purposes.

In its present form, the AquiFR system includes 3 groundwater flowhydrogeological modelling software covering 11 sedimentary aquifers and 23 karstic systems: the EauDyssée hydro(geo)logical numerical platform (Saleh et al., 2013), the MARTHE groundwater flow software (Thiéry, 2015a) and the EROS lumped model software used for karstic systems (Thiéry, 2018a). These softwaressoftware are embedded in an application developed with the OpenPALM coupling system (Duchaine et al., 2015)(Duchaine et al., 2015). All these models cover an area of about 149 000 km^2 and contains up to 10 overlaid aquifer layers. Prior to real time monitoring and forecast, AquiFR need to be assessed on a long-time period, which is reported onpresented in the present study. The evaluation is carried out over a 60--year period from 1958 to 2018 at a daily time step. This long-term simulation provides a unique insight on the long-term evolution of groundwater in France, as most of the groundwater data are available over about 30 years. This long-term simulation can then be used to characterize the daily situation compared to past events. The SAFRAN meteorological reanalysisanalysis (Vidal et al., 2010), available over the French metropolitan area at an 8 km resolution, allow to supply the meteorological variables to the SURFEX land surface model (Masson et al., 2013) which evaluates the water balance over the French metropolitan area. It must be stressed that the hydrogeological models could have been classically fed with the SAFRAN reanalysisanalysis precipitation, potential evapotranspiration, and temperature data using their own water balance calculation. TheHowever, the combined use of

SURFEX and SAFRAN provides a consistent set of hydro-meteorological data over an 8 km resolution grid over France, including groundwater recharge and surface runoff from SURFEX, as well as potential evapotranspiration, precipitation, and temperature from SAFRAN. The use of these SURFEX 8 km resolution fluxes made necessary the recalibration of the hydrogeological models included in the platform.~~ More details can be found in Habets et al. (2017).~~

A wide range of gauging stations and piezometers were selected in order to perform the evaluation of the simulated piezometric heads, river flows and karstic spring flows. This evaluation allows to ~~assess the performance of the platform to~~ identify extreme events such as groundwater floods or droughts over a long-term period. In this paper, a detailed description of the ~~datasets~~AquiFR platform and ~~the different~~its components ~~of AquiFR are~~ is presented in section 2. Section 3 provides information on the regional models, their calibration and the statistical criteria used to evaluate their performance. Section 4 presents the assessment of the long-term simulation based on comparison with observations of river ~~flow~~flows, karstic spring ~~flow~~flows and piezometric ~~head observations, which is~~ heads. The results are then discussed in section ~~4. Conclusions are drawn in section~~ 5, prior to the conclusions.

2 The AquiFR Hydrometeorological Modelling Platform

~~The AquiFR hydrogeological modelling platform was developed using the OpenPALM coupling system (Buis et al., 2005; Duchaine et al., 2015)~~The AquiFR hydrometeorological modelling platform represents the main hydrological processes occurring within the watersheds from precipitations to groundwater flows as shown in Figure 1. AquiFR accounts for spatial heterogeneity by using different spatial scales. The atmospheric fields 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) uses these atmospheric variables to solve the energy and surface water budget at the land-atmosphere interface at a 5 minute 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 the vegetation (in France from 0.2 to 3 m) 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 basins, these two fluxes are transferred to the MARTHE (Thiéry, 2015a) or EauDyssée ~~. 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: one component that retrieves the fluxes previously simulated by the SURFEX land surface~~

model (Masson et al., 2013), one component for each model: the EauDyssée (Saleh et al., 2013) groundwater models. These models simulate the transfer to the unsaturated zone, groundwater flows within and between the aquifer layers, the routing 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, 2018a). Each karstic system is represented by a lumped reservoir model solved at a daily time scale. Conceptual approaches are preferred for simulating karstic systems. Indeed, 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. and MARTHE (Thiéry, 2015a) groundwater modelling software, the EROS software (Thiéry, 2018a), and a last component that synchronizes each model at each time step and gathers all the data for post processing. All these elements are connected together and exchange data during the parallel execution of a single OpenPALM executable. Figure 1 shows the basic diagram of AquiFR as it is developed in OpenPALM. The use of OpenPALM allows to run 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.

Prior to an AquiFR run, atmospheric forcing from an independent simulation of a numerical weather forecast model or from observed database are provided to the SURFEX land surface scheme (Masson et al., 2013). SURFEX computes the energy and surface water budget at the land-atmosphere interface and simulates surface runoff and groundwater recharge needed by the hydrogeological models in the AquiFR platform. Atmospheric forcing corresponds to precipitation, temperature, relative air humidity, wind speed and downward radiations. For the present study, the SAFRAN meteorological reanalysis is used.

The workflow of an AquiFR time step occurs in the following way. First, both the atmospheric forcing and the SURFEX groundwater recharge and surface runoff are retrieved at the beginning of the time step. Then, within the AquiFR platform, EauDyssée and MARTHE use the SURFEX fluxes to compute the simulated piezometric heads, river flows, and groundwater-surface water exchanges over regional aquifers at different spatial resolutions depending on the hydrogeological model. These groundwater-surface water exchanges include both the stream-groundwater interactions and the overland groundwater overflows. In parallel, the EROS software computes karstic spring flows at the outlets of the karstic systems. Inputs are precipitations (snow and rainfall), temperature and potential evapotranspiration provided by the meteorological forcing inputs. Once each hydrogeological model reaches the end of the current time step, the synchronization component retrieves the current time step outputs for post-processing and allow the platform to compute the next time step.

~~The following subsections present a brief description of the components integrated within the OpenPALM application as well as the models currently included in AquiFR.~~

~~2.1 The SAFRAN meteorological reanalysis~~

~~The SAFRAN analysis system~~The AquiFR platform is an OpenPALM application that currently gathers 5 components. 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 of SURFEX is taken into account. Therefore, a preliminary step illustrated by Figure 2Figure 1a 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 2Figure 1b. 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. 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 to 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 on 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.

2.1 The SAFRAN meteorological analysis

The SAFRAN meteorological analysis is a mesoscale atmospheric analysis system for surface variables. It provides meteorological forcing data over France on an 8 by 8 km grid at the hourly time step using observed data and atmospheric simulations. Originally intended for mountainous areas, it was later extended to cover France (Quintana-Seguí et al., 2008). SAFRAN ~~analyses~~estimates eight variables: ~~the~~ rainfall, snowfall, incoming solar and atmospheric radiations, cloudiness, ~~2 m~~ air temperature~~, 2 m~~ and relative humidity 2 m above ground, and ~~10 m~~ wind speed~~.~~ 10 m. Potential evapotranspiration can also be computed from these atmospheric variables. SAFRAN is based on climatic zones where the atmospheric variables only vary according to the topography. More than 600 homogeneous climate zones are defined over France. The average area for each zone is about 1000 km^2 so that each one contains one surface meteorological station and at least two rain gauges. SAFRAN uses all the observations available to ~~analyse~~estimate each atmospheric variable except for radiations. For each variable, values are assign to given altitudes using an optimal interpolation method. The analyses are computed every 6 hours and an interpolation is made to an hourly time step~~ and radiation~~. Radiation fluxes are computed using a radiative transfer scheme. The daily precipitation rates are estimated using a wide range of daily rain gauges and converted to hourly data using the evolution of the air relative humidity. The vertical profiles of the atmospheric parameters are then computed in each climatic zone and the values are spatially interpolated over the 8 km grid as a function of the altitude within each climatic zone. Further details on the SAFRAN analysis system can be found in ~~Quintana-Seguí et al. (2008)~~Quintana-Seguí et al. (2008) and Vidal et al. (2010).

2.2 The SURFEX modelling platform

SURFEX is a modelling platform aiming ~~to simulate~~at simulating the water and energy fluxes at the interface between the surface and the atmosphere (Masson et al., 2013). SURFEX is built to be coupled to forecast and climate models. It includes databases, interpolation schemes, and ~~then can be used~~several physical options that allow its use over different spatial and temporal scales. SURFEX gathers several physical schemes in a single platform, allowing the simulation of the urban surfaces and the main components of the water cycle: sea and ocean, lake, vegetation and soil.~~ In the present study, SURFEX is used in offline mode in order to provide groundwater recharge and surface runoff to the AquiFR platform.~~

Mis en forme : Anglais (États-Unis)

Land surface processes are taken into account using the Interaction between Soil Biosphere and Atmosphere (ISBA) ~~land surface scheme.~~(Noilhan and Planton, 1989) land surface scheme. ISBA uses a short list of parameters depending on vegetation and soil types. The ~~tempral~~temporal evolution of the soil water and energy budget is computed using a multilayer soil scheme based on the explicit resolution of the one-dimension Fourier law as well as the mixed form of the Richards equation~~.~~ (Boone et al., 2000; Decharme et al., 2013). Surface/groundwater capillary exchanges can be explicitly taken into account (Vergnes et al., 2014) as well as the vertical root profile in the soil (Braud et al., 2005).

In the present study, no bidirectional coupling between the soil of SURFEX and the aquifers is accounting for. 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. The soil column thickness represented in each 8 km resolution grid cell varies from 0.20 to 3 meters according to the land cover. It corresponds mostly 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).

2.3 The EauDyssée groundwater modelling software

The EauDyssée modelling platform ~~gathersnumerical~~gathers numerical modules representing several hydrological processes, the most important being the aquifer module based on the Simulation ~~of Multilayer Aquifers (~~des Aquifères Multicouches SAM (multilayer aquifer system) regional groundwater ~~model~~modelling software (Ledoux et al., 1989) and the ~~riverrouting~~river routing scheme based on the Routing Application for Parallel ~~Computation~~computatIon of Discharge (RAPID) model (David et al., 2011).

~~The~~ SAM ~~model~~ computes the evolution of the piezometric heads of multilayer aquifers using a finite difference numerical scheme to solve the groundwater diffusivity equation with a square grid discretization. Groundwater horizontal flows are two-dimensionals ~~while~~and vertical flows through aquitards are ~~represented by leakage.~~taken into account. Therefore, unconfined and confined aquifers can be represented. SAM was successfully used to predict groundwater and surface water flows in different basins of various scales and hydrogeological contexts: the Seine basin ~~(Viennot, 2009)~~(Viennot, 2009), the Somme basin (Habets et al., 2010), the Loire basin (Monteil et al., 2010) or the Rhine basin (Thierion et al., 2012; Vergnes and Habets, 2018).

The RAPID software is a river routing model based on the Muskingum routing scheme (David et al., 2011). It can be ~~coupledto~~coupled to groundwater and land surface models. ~~Volume~~Volumes and river ~~flow~~flows are computed along a river network discretized into square grid-cells to ease the simulation of the ~~exchange~~exchanges with groundwater. River-groundwater exchanges are taken into account in both directions.

2.4 The MARTHE groundwater modelling software

~~The~~The Modélisation d'Aquifères par un maillage Rectangulaire en régime Transitoire pour le calcul Hydrodynamique des Ecoulements MARTHE (Modelling Aquifers with Rectangular cells, Transport and Hydrodynamics) computer code is the hydrogeological ~~model~~modelling software from the French Geological Survey (BRGM) (Thiéry, 2015a, 2015c, 2015b). MARTHE embeds single to multilayer aquifers~~,~~ and hydrographic networks ~~and the exchanges with the atmosphere (rainfall, snow and evapotranspiration) for the computation of the soil water balance~~. It is designed for 2D or 3D modelling of flows and mass transfers in aquifer systems, including climatic, human influences and possible geochemical reactions. Groundwater flow is computed by a ~~3-D~~3D finite volume approach to solve the hydrodynamic equation based on the Darcy's law and mass conservation, using irregular rectangular grids, with the possibility of nested grids. ~~Based~~River flows are simulated based on a kinematic wave approach~~, flow in river networks~~ that is fully coupled to groundwater flow. Groundwater-river exchanges are taken into account in both directions.

Other options are available and can be integrated to the simulation: mass transfer for pollutants in water, temperature effects, impact of salinity, degradation of pollutants, transfers in the unsaturated zone and geochemical reactions.

This software ~~has been~~is widely used for groundwater resources management in France: for example in the Somme River basin (Amraoui et al., 2014), in the Poitou-Charentes region (Douez, 2015), in the Basse-Normandie region (Croiset et al., 2013) or in the Aquitaine sedimentary basin (Saltel et al., 2016). It ~~has~~is also ~~been~~ used in other environmental fields such as pollutant infiltration in unsaturated zones (Herbst et al., 2005; Thiéry et al., 2018) or for the simulation of pollution plume coming from a contaminated area.

2.5 The EROS software

~~The~~The Ensemble de Rivières Organisés en Sous-bassins EROS (set of rivers organized in sub-basins) numerical code is a ~~hydro-climatic rainfall-river flow-piezometric head~~ distributed ~~model~~reservoir modelling software dedicated to large river systems (Thiéry, 2018a; Thiéry and Moutzopoulos, 1992). It allows the simulation of river flow or karstic spring flow and piezometric head in heterogeneous river basins. These river basins are ~~described~~represented in EROS as a cluster of elementary lumped-parameter hydrological models connected with each other. For each sub-model, a hydroclimatic lumped model ~~allows to compute~~computes the local river discharge at the outlet of the sub-model and the piezometric head in the underlying water table. Each sub-model simulates the main mechanisms of the water cycle through simplified physical laws (Thiéry, 2015d). Snow accumulation, snow melting and pumping ~~is~~are taken into account. The total river flow at the outlet of each sub-basin is computed from the upstream tree of sub-basins.

EROS was initially developed to simulate regional ~~watershed~~watersheds avoiding the complexity of ~~a more complex~~ spatially physically based model~~ without sacrificing the performance.~~. In the framework of ~~the~~ AquiFR~~ project~~, this software was ~~used and~~ adapted in order to simulate in a single ~~run~~instance 23 karstic systems as independent sub-~~models, making the simulation of these karstic systems less expensive in terms of computational burden with respect to the use of 23 distinctive~~ models (Thiéry, 2018b). It is not connected to SURFEX but directly to SAFRAN as described in Figure 1.

~~2.6~~3 Methodology

3.1 The regional models implemented in the AquiFR platform

Mis en forme : Titre 2

AquiFR aims at covering all groundwater resources in France. Figure 3~~ 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). These hydrogeological 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, each regional model was recalibrated based on the SURFEX fluxes (Habets et al., 2017). Generally speaking, observations and periods of calibration were the same as those initially used to develop each model. Hydrodynamic parameters, including hydraulic conductivities and specific yields, were modified based on hydrogeological expertise in order to obtain the best fit between simulated and observed river flow and piezometric heads. It should be mentioned that for the MARTHE Somme basin model, in order to get an accurate simulation, at each time step, the choice was made to divide the total runoff provided by SURFEX into surface runoff and groundwater recharge using the GARDENIA water balance scheme inside the MARTHE computer code.~~

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 4).

Some regions are simulated by two spatialized models (Figure 4): 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 3).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 againstwith the SURFEX fluxes. It corresponds to the four EauDyssée sub-models over the Seine basin and the MARTHE models for the Somme and Basse- Normandie regions.MARTHE models. Figure 4 also shows the 23 karstic systems (median catchment area of 99 km2) simulated by EROS. (Thiéry, 2018b) (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)(Courtois, 2018; Marçais et al., 2017) and integrated in the near future.

Groundwater withdrawals are integrated as input datatime-dependent boudary conditions 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 (that is about 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. based on tax reporting. Pumping concerns drinking water, irrigation, and industrial use. The quality of the data setdataset as well as its temporal extension varied for each regional modelling, although the latterlater does not exceed 20 years. Further details on regional models can be found in the references listed in Table 1.Table 1. To extend the pumping estimation to the 1958-2018 period, a monthly mean annual cycle is used for the years without data. Some spatialized models also need to prescribe time dependant boundary conditions. For the 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 models it is required to providebasins are partially represented. Therefore, they need to prescribe time-dependent boundary conditions at the upstream river flow from 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 cycles are calculated from the observed period and applied to the simulated period not covered by observed data.observed river flow. In the near future, a new method basedthe 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 will be implemented in or by the RAPID river routing 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 were usually computed using less physical methods and atmospheric local data (precipitation, temperature and potential evapotranspiration) that differ from the physically based approach using SURFEX and the SAFRAN analysis. As a result, in order to be consistent with the estimation of the groundwater recharge estimated by SAFRAN-SURFEX, most of the regional models were recalibrated based on the newfluxes (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,

Mis en forme : Non Exposant/ Indice

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 criteria. 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 in 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. It improved the simulation of the river flows using the SURFEX total runoff. Once the new partition was estimated, the aquifer 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 the models are similar with the original water balance fluxes and the ones simulated by SAFRAN-SURFEX, although locally, they may be better ~~estimate these upstream flows.~~ 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 calibration is given in Habets et al. (2017).

3.3 Evaluation criteria of the 60-year 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:

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

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) 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 is strongly affected by the biases. Therefore, we computed a RMSE bias-excluded in order to better assess the simulation in terms of amplitude and synchronization. Moreover, this RMSE bias-excluded is normed with respect to the observed standard deviation for each

Mis en forme : Anglais (États-Unis)

Mis en forme : Anglais (États-Unis)

observation in order to account for the differences of variability between the numerous wells to help spatial comparison or aggregation. This normed RMSE bias-excluded (NRMSE_BE) is expresed as follows:

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

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 coefficient 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. It 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.

The annual discharge ratio criterion helps to compare the mean simulated and observed river flows as follows:

$$Ratio = \frac{\overline{Q_{sim}}}{\overline{Q_{obs}}}, \quad (4)$$

with $\overline{Q_{sim}}$ and $\overline{Q_{obs}}$ the mean siumlated and observed river flows respectively.

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). It is currently used in France for the Monthly Hydrological Survey (MHS) (Office International de l'Eau, 2019). This MHS provides monthly information to the policymakers and the public on the hydrological state of groundwater. 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 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, a non-parametric kernel density estimator allows estimating the best probability density function fitting the histogram of monthly values. At last, for each month from January to December, a projection over the standardised normal distribution using a quantile-quantile projection allows deducing 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 allows representing wetter and drier periods in a similar way all over the simulated domain.

4 Results

The long-term simulation was carried out over a 60--year period from the August 1, 1958 to July 31, 2018 at a daily time step using the SAFRAN meteorological reanalysesanalyses. The mean precipitations corresponding to the simulated domain of Figure 4 and averaged over the 60--year period is equal to 743 mm/year. The surface water budget is then computed by SURFEX from the SAFRAN outputs. The mean simulated effective rainfalltotal runoff is partitioned between 163 mm/year of groundwater recharge and 60.5 mm/year of surface runoff. Thus, the groundwater abstractions represent about 25% of the groundwater recharge.

The evaluation of this simulation wasis made using the numerous in situ datasets available in France. Observed piezometric heads over France are available in the “Accès aux Données sur les Eaux Souterraines” (ADES) database (http://www.ades.eaufrance.fr/). 639 observation boreholes covering the regional models included in AquiFR were selected,domain corresponding to both confined and unconfined aquifers, and with at least 10 years of continuous time series., were selected. Figure 5 shows the temporal evolution of the number of daily measurements along the 60--year period. Starting in 1958, only a few measures are available. Starting from 1970, the number of wells increases slowly to reach about 100 in 1990. Then the number of daily measurements quickly increases to reach more than 450 in 2010. This number remains stable then, except for the last year (2018) with a decrease because the datasets were not yet fully transmitted. In situ daily river flow observationsat 228observations at 362 gauging stations were also selected for evaluating the daily simulated river flows from the Hydro database (http://hydro.eaufrance.fr/).

34.1 Piezometric head

In order to evaluate the quality of the simulation, two statistical critera are used: the bias and a normalised Root Mean Square Error (RMSE) bias-excluded (NRMSE-BE). Figure 6a shows the spatial distribution of the bias scores for the 639 observed piezometers selected over the simulated regional multilayer aquifers.. A positive (negative) value means that the simulation overestimates (underestimates) the mean observed piezometric head with respect to the observationwhile a negative value means the opposite. The north of the Loire river basin, corresponding to the Beauce region, shows a significant underestimation of the mean observed groundwater level. Elsewhere, no significant patterns appear. Figure 6b summarizes these results with the accumulated distribution of the absolute biases for all the piezometers. 4042% and 60% of the absolute biases are lower than 2 m and 4 m, respectively.

The computation of the RMSE scores is very affected by these biases. Therefore, we compute a RMSE bias-excluded score in order to avoid these biases and 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 allows to take into

Mis en forme : Anglais (États-Unis)

Mis en forme : Anglais (États-Unis)

Mis en forme : Anglais (États-Unis)

Mis en forme : Anglais (États-Unis)

account the differences of variability between the observed points and to better compare them with each other. Figure 7a shows the spatial distribution of these normalised RMSE bias-excluded (NRMSE-BE) scores and Figure 6b summarizes these results with the and accumulated distribution of the NRMSE-_BE for all the piezometers. 16% and 6162% of these scores arethe wells obtain a value lower than 0.6 and 1 respectively, while 88% of them arehave a value lower than 2. Some piezometers that were affected by important biases in Figure 6a exhibit however good NRMSE-_BE scores, in particular over the Loire river basin and in the northern Poitou-Charentes region, meaning that the temporal evolution is well simulated.

Five examples of simulated and observed daily evolution of piezometric heads are shown in Figure 8. These piezometers are encircled in Figure 3 and statistical scores are available in Table 2. They were chosen to characterize different hydrogeological contexts. The first piezometer named Omiécourt is located in the Chalk aquifer of the Somme River basin. The temporal evolution of the groundwater level is characterized by multiyear cycles well captured by the model. However, the simulation displays annual cycles that are not observed. These annual cycles explainIt explains why the NRMSE-_BE score is equal to 0.8193 while the bias score is equal to -0.1586 m. The two piezometers named Ruffec and Le Bec hellouin correspond to limestone aquifers and are located in the Poitou-Charente region and near the coast of the English Channel respectively. The first one is characterized by large annual cycles with wide amplitudes. The model is able to reproduce these annual cycles (correlation of 0.82) but with an underestimation of the peaks leading to a negative bias score of -1.44 m. The Le Bec hellouin piezometer is characterized by both multiyear and annual cycles that are captured by the model except, although between the 2005 and 2015 years where the simulated groundwater level is underestimated with respect to the observation. The piezometer named Farceaux is located in a chalk aquifer in the Seine River basin. It is characterized by a systematic bias of about -8.3 m. Otherwise, the multiyear and annual cycles are well reproduced by the model, which is confirmed by the NRMSE-_BE score equal to 0.52. The last example corresponds to a piezometer for which the model cannot reproduce the strong seasonal decrease of level occurring each year. Such behaviours in the observation are likely due to groundwater withdrawals that are not well prescribed in the model near this observation well.

34.2 The StandardizedStandardised Piezometric Level Index

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. 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 is normalized so that wetter and drier periods can be represented in a similar way all over the French national territory. The SPLI is currently used in France for the Monthly Hydrological Survey (MHS) (Office International de l'Eau, 2019). This MHS provides monthly information to the policymakers and the public on the hydrological state of groundwater. The SPLI is categorized into seven classes summarized in Table 3 from the driest to the wettest conditions. According to Seguin and Klinka (2016), a set of piezometers were chosen in order to compute the SPLI indicator in this MHS with the following characteristics: a continuous time series withatwith at least 15 years and no impact of pumping wells. Among the 639 selected observation wells in Figure 6 and Figure 7, 103 contribute to the calculation of the SPLI in the MHS.

The correlation and the Nash-Sutcliffe model efficiency score (NSE)Figure 9 and Figure 10 show the spatial distribution of the NSE and the correlation respectively computed between the observed and simulated SPLI indicator for the 103 selected piezometers. 20% of the NSE are greater than 0.7 and 56% greater than 0.5, while 12% are lower than zero. In parallel, 65% of the correlation are greater than 0.7.

Figure 11 (Nash and Sutcliffe, 1970) are used to evaluate the SPLI indicator. A negative NSE means that the mean observed signal is a better predictor than the model. A NSE above 0.7 is generally accepted as a reasonable estimate of the signal dynamic, however depending on the hydrogeological and climate context of the basin. It is often applied to compare observed and simulated river flows but can be used for other variables such as the SPLI indicator. It sensitivity to high-frequency fluctuations makes its use for comparing groundwater levels less obvious. Figure 8 and Figure 9 show the spatial distribution of the NSE criterion and the correlation scores respectively computed between the observed and simulated SPLI indicator for the 103 selected piezometers of the groundwater domain. 20% of the NSE scores are greater than 0.7 and 55% greater than 0.5, while 11% are lower than zero. In parallel, 64% of the correlation scores are greater than 0.7.

The left part of Figure 10 focuses on five examples of observed and simulated temporal evolutions of the SPLI indicator. These piezometers correspond to the ones shown in Figure 8 and are part of the selected piezometers used for the MSH. Background colours correspond to the classification of the SPLI from the driest (red) to the wettest (blue) hydrological conditions as shown in Table 3. Table 2 presents the related NSE and correlation scores. The NSE scores computed for these SPLI time series are all greater than or equal to 0.6 except for the Bourdet piezometer characterized by a NSE score equal to -0.51. This lower score may be due to a lack in the inputs of the model input, such as the absenceunderestimation of withdrawal data in its vicinity.

The SPLI is a statistical calculation with the particularity to centre the temporal evolution of the piezometric head on zero, hence removingThe SPLI, as a frequency indicator, does not account for the potential biases between the observed and simulated groundwater levels. This is the reason why the systematic biases found in Figure 8 do not appear in the monthly SPLI evolutionscomparisons in Figure 11, in particular for the Farceaux piezometer. Moreover, the SPLI indicator normalized the amplitude of both the observed and simulated temporal evolution of groundwater levels. The right part of Figure 11 describesshows the histograms of the simulated (in percentage of the simulated red) and observed (in blue) and observed (in black) monthly SPLI values distributed against thefor each classes of Table 3. This classification isThe histograms are similar for both the observed and simulated SPLI ofat the Ruffec and Le Bec Hellouin piezometers. The occurrences of the wetter conditions are well reproduced for the Omiécourt piezometer but the model tends to overestimate the importance of underestimate the driestnumber of moderately dry conditions when compared to the observations (26% and 18% of "moderately dry" events for the simulation and the observation respectively and 4.5% and 7.2% of "very dry" events for the simulation and the observation respectively). For the Farceaux piezometer, the model underestimates the occurrences of the driest events and overestimates the occurrences of the wetter events.. Despite the poor scores obtained for the Bourdet piezometer, in particular for the correlation scorescorrelations, the distribution of all the monthly SPLI values with respect to the classes of Table 4Table 3 is similar for both the observation and the simulation.

The MHS published every month in France for water resources management includes the calculation of the SPLI. As an example, Figure 12a shows the observed SPLI values calculated for the 103 selected piezometers for June 2016. We chose this specific month since it follows large precipitation events that leads to floods in the Seine and Loire basins (Philip et al., 2018). Figure 12b shows the simulated SPLI values computed for this specific month. The model reproduces the overall pattern of normal and wet conditions but tends to overestimate the importance of the moderately wet conditions: 1719% (29%) of the simulated (observed) piezometers are in normal conditions, 50% (3246% (31%) are in moderately wet conditions, 14% (1216% (13%) are in wet conditions, and 16% (18%) are in extremely wet conditions. The background map of Figure 12b showshows the SPLI computed in the cells of the whole outcropping domain. These SPLI values were computed with respect to a 30 years reference period from 1981 to 2010, which might lead to differences between the simulated SPLI map and observed values. This map shows a large area of extremely wet conditions located in the south of the Loire River, which refers to the extreme event episode of rainfall from the end of May 2016.

34.3 River flow and karstic spring flow

The 23 karstic systems simulated by the EROS model are evaluated against gauging stations located at the outlet of the corresponding karstic systems. All these gauging stations were also used to calibrate the model (Thiéry, 2018b). Figure 13 compares shows the comparison on the 60-year period of the observed and simulated monthly river flows for four examples of karstic systems located in Figure 3. There is a tight agreement between the observation and the simulation. The NSE scores of the square root of the daily riverkarstic spring flows are given for each example. Using the square root of the daily riverkarstic spring flow allows to attenuateattenuating the importance of the flood peaks characterizing these small karstic systems and enables to better evaluate the riverkarstic spring flow simulation. Such transformation is necessary because of the excessive sensitivity of the NSE criteria to extreme values in a river flow time series (Legates and McCabe Jr, 1999). For these four examples, all the NSE scoresOther statistical scores using less assumptions on the underlying data distribution, such as the non-parametric variant of the Klunge-Gupta efficiency (Pool et al., 2018), could be used to reduce the sensitivity to the extremes. For these four examples, all the NSE are greater than 0.7.

The distributed models included in the AquiFR modelling platform integrate river networks and the simulation of river flows on each river grid cell. 228 gauging stations were selected to evaluate the simulated river flows. These gauging stations correspond to the ones that were used for calibrating and evaluating each model (see references in Table 1). They were kept to evaluate the AquiFR platform (Habets et al., 2017). Figure 13 compares362 gauging stations were selected to evaluate the simulated river flows. Figure 14 shows the comparison between the observed and simulated daily river discharges for four of them: the Charente River, the Somme River, the Seine River and the Loire River. The locations of these gauging stations is shown in Figure 3. The Charente River and the Somme River correspond to watersheds with areas lower than 10 000 km^2 and the Loire River and the Seine Rivers to regional watersheds with areas greater than 80 000 km^2. Three statistical scores were selected to evaluate the performance of the models: the NSE, the correlation and the annual discharge ratio criterion (Ratio = Q_{sim}/Q_{obs}). TheseThe statistical scores are summarized in Table 4 for the gauging stations of Figure 13.. The simulated river flow of the Charente River underestimates the peak floods, which leads to a ratio score of 0.81. The river flows of the Seine River and the Loire River are well reproduced with daily NSE scores of 0.86 and 0.9 respectively. The Somme river flow is also well reproduced with a NSE score equal to 0.69 due to a lower ratio score of 0.92.

Figure 15a shows the spatial distribution of NSE scores for the 228362 gauging stations helping to evaluate the distributed hydrogeological models (circles) and the 23 karstic systems (stars). The NSE scores calculated for the karstic systems correspond to the square root of the daily riverkarstic spring flows. in order to attenuate the importance of the flood peaks characterizing these small karstic systems, as explained above. The corresponding accumulated distributions are shown for the 23 karstic systems and for the 228362 gauging stations in Figure 15b and Figure 15c respectively. 8096% of the NSE score using the square root of the daily riverkarstic spring flows are greater than 0.87. Regarding the results of Figure 15c, for rivers in continuous aquifers, 2734% of the NSE scores are greater than 0.7, 58%. Moreover 63% of these NSE are greater than 0.5, and 22 while 18% are negatives.

5 Discussion

The results obtained overon the 1958-2018 period demonstrates the feasibility and the utility to gather several regional models developed separately in severaldifferent research institutes into a single numerical tool to provide simulations of the water resource at a daily time step at the national scale. It was shown that the AquiFR platform is able to reproduce the evolution of the observed hydrological variables, including piezometric levels and river flows, with reasonable statistical scores. Some regions are nevertheless better reproduced than others. For example, the Loire and the Nord-Pas-de-Calais regionsregion exhibit a poor NRMSE- BE scores in Figure 6 than the other regions.. Part of the error is linked to the estimation of the groundwater recharge by SURFEX essentially because the. Indeed, the regional groundwater models were developed and calibrated independently using various methods and data to compute groundwater recharge (see references in Table 1). For consistency. As a consequence, the development of AquiFR reinforcesreinforced the need to calibrate these models based on the SURFEX forcing fields. This work was accomplished for most of the models included in AquiFR except for some of them, including the Loire River basin (Habets et al., 2017). The use of an inverse model as the one proposed by Hassane Maina et al. (2017) should help to improve such calibration.

Starting from scratch with an integrated method could have prevented the burden of maintaining each model separately and handling the different outputs of the models. Such method was applied by Kollet et al. (2018) over the North Rhine-Westphalia domain using Parflow-CLM by integrating all the physical processes related to groundwater/surface water into a single numerical tool. De Lange et al. (2014) used an approach closer to AquiFR by coupling five physically different models for different water domains with different concepts, different temporal and spatial scales, and different national and regional databases altogether embedded into the National Hydrological Instrument NHI. These two models are used for integrated water management and policymaking issues. The areas covered by these models are 22 500 km^2 and 41 500 km^2 respectively. According to Figure 3, the BDLISA database references regional sedimentary aquifer systems (in green) and alluvial aquifers (in blue) in France both covering an area of about 355 000 km^2. Reaching the complexity of a fine-tuned regional model, including the geometrical, geological and physical contexts, in a single integrated numerical tool covering such large territory would be time consuming to build, to calibrate, and to evaluate. ~~Itwould~~It would also ~~necessitate~~demand big resources ~~in~~of computational power. An attempt was made by Vergnes et al. (2012) to simulate~~ in~~ a single integrated model groundwater over the French metropolitan territory. Even though the results obtained were good enough to be used for large-scale climate applications, it was not fitted for operational water management: only one layer was defined at the coarse resolution of about 10 km, no pumping were defined, no calibration was achieved and very simplified parameterizations were used.

To overcome this difficulty, the choice was made for AquiFR to bring together different models developed independently. Currently, the area covered by the platform is equal to around 133 000 km^2 with a number of layers that can reach 10 layers for some models, and these numbers will increase in the future with the addition of new models. This multi-model approach allows to promote the share of knowledge in hydrogeology and to gather the competencies accumulated in the different research institutes involved in AquiFR in water resource modelling. Moreover, thanks to the evolutive approach of the OpenPALM coupling software, the platform facilitates the addition of new software and new models.

Results from AquiFR show a global view of the performance of the AquiFR platform but are characterized by uncertainties. These uncertainties are mainly related to the calibration of the models and to the lack of some input data like groundwater abstractions. Indeed, the regional models have been calibrated over a shorter period compared to the long-term simulation~~.~~, as stated in the section 3.2. Consequently, ~~they do not take into account in the ~~the calibration process may have not included the extreme climatic conditions ~~known over~~encountered on the 1958-2018 period. Other uncertainties may be related to the choices of the resolution, the databases used, the geometry of the models, and more generally the representation of the physical processes in the hydrogeological software. Some regions are better monitored than others and the global view of the performance of the AquiFR platform is certainly affected by this. Moreover, the chosen method of evaluation based on statistical scores such as NSE could also be improved. Indeed, some authors report that the use of more realistic upper and lower benchmarks for each simulated basins could improve the judgement of model performances with respect to the climate and hydrogeological context of the basins (Pappenberger et al., 2015; Seibert et al., 2018). ~~In~~At last, in order to diminish ~~these~~ uncertainties, a long-term calibration effort using a denser observation network ~~should~~could be undertaken to improve the AquiFR performance.

The AquiFR platform can be seen as an improvement of the SIM hydrometeorological tool for operational water management purpose (Habets et al., 2008). These two systems share common points as the SURFEX land surface model or the groundwater component of the EauDyssée platform. However, the SIM tool uses coarse hydrogeological modelling ~~and~~with less aquifer layers or no river loss to the aquifer. It mainly focuses on operational forecasts of river flows and monitoring of soil humidity.

The AquiFR platform is intended to focus also on the forecast of groundwater levels for the multilayer aquifers and karstic systems described in Figure 4. To achieve this goal, the SPLI indicator will be calculated to provide forecasts of extreme events. For this purpose, AquiFR is able to produce different representation of this indicator and to compare it with other variables depending on the need of the stakeholders. For example, Figure 16 compares the simulated time series of daily mean groundwater recharge, stream-groundwater exchanges budget, monthly mean piezometric head and SPLI averaged over the chalk aquifer of the Somme model. It gives a global view of the past states of groundwater related to climatic extreme events, such as the severe flood of the 2001 year, characterized by groundwater flooding and sustained stream-to-groundwater exchanges (Amraoui and Seguin, 2012).

6 Conclusions

This study introduces the AquiFR hydrogeological modelling platform aiming to provide at the French national scale short-term to seasonal hydrological ~~forecast~~forecasts as well as real-time monitoring for daily water management and long-term simulations for climate impact studies. It was developed using a coupling software in order to gather ~~inside the same numerical tool different softwares for different water domains~~different software and several models covering a set of French multilayer aquifers. Daily surface runoff and groundwater recharge ~~prrovided~~provided by the SURFEX land surface model ~~forced by~~using the SAFRAN meteorological analysis were used ~~to feed AquiFR~~ for simulating the daily evolution of groundwater levels and river flows of French regional multilayer aquifers and karstic systems ~~over~~on the 1958-2018 period.

The results confirm the feasibility of gathering independent hydrogeological model developed in different research institutes into the same coupling platform. All these models were initially developed and calibrated on shorter ~~period~~periods with heterogeneous geological and meteorological databases. Some of these models were recalibrated against the SAFRAN-SURFEX ~~and SAFRAN~~ fluxes. The evaluation of the 1958-2018 long-term simulation shows a good comparison with the observations available on the same period. It confirms the relevance of using AquiFR as a tool for ~~future~~ long-term impact studies. The evaluation of the SPLI indicator also shows that AquiFR could be used in an operational context ~~for predicting~~to monitor future ~~extreme~~ events ~~in a similar way than the MSH produces each month in France to quantify the hydrological state~~and be part of the ~~aquifers~~Monthly Hydrological Survey, provided the necessary caution in terms of communication of model uncertainties and performances.

The advantage of this platform lies in its modularity. AquiFR encourages the development of groundwater modelling where it is missing and, more generally, it has the potential to be a valuable tool for many applications in water resource management and research studies~~.~~, as for instance climate change studies and seasonal forecasts. In the future, more regional ~~spatial~~ models developed with MARTHE or EauDyssée will be included in order to extend the cover of AquiFR, as the Tarn and Garonne aquifer (Figure 4). A new software will be included in order to simulate bedrock aquifers located in Britany ~~(see Figure 3)~~ (Courtois, 2018). A new modelling method based on a lumped-parameter rainfall-runoff model will be used to provide upstream river flows as boundary conditions for the MARTHE ~~model~~models that required it. ~~Another project will be to add a~~

~~river network covering all the French territory.~~ Assessment of the seasonal forecast of the groundwater resource is now ~~on~~in progress (Roux, 2018). Since errors in the initial conditions can alter significantly the skill of the forecast, dedicated studies on data assimilation to improve initial state conditions are also done in parallel (Hassane Maina et al., 2017).

Acknowledgements.

The authors are grateful for financial support from the French Agency for Biodiversity (AFB) and the French Ministry of Ecological and Solidarity Transition. The authors are also grateful to the research institutions that provide computer access, data, models and software for making this project feasible (BRGM, CERFACS, Météo France, Mines ParisTech, Geosciences Rennes). Special thanks come to Claire Magand, Thimotée Leurent and Bénédicte Augeard for supporting this project with AFB, and Nathalie Dörfliger who supports this project with BRGM. Finally, the authors wish also to thank Eric Martin and Jean-Michel Soubeyroux that were involved at the beginning of this project.

References

Aeschbach-Hertig, W. and Gleeson, T.: Regional strategies for the accelerating global problem of groundwater depletion, Nat. Geosci., 5(12), 853–861, doi:10.1038/ngeo1617, 2012.

Amaranto, A., Munoz-Arriola, F., Corzo, G., Solomatine, D. P. and Meyer, G.: Semi-seasonal groundwater forecast using multiple data-driven models in an irrigated cropland, J. Hydroinformatics, 20(6), 1227–1246, doi:10.2166/hydro.2018.002, 2018.

Amraoui, N. and Seguin, J.-J.: Simulation par modèle maillé de l'impact d'épisodes pluvieux millénaux sur les niveaux de la nappe de la craie et les débits du fleuve Somme. Rapport final, BRGM/RP-61864-FR, Orléans~~.~~, 2012.

Amraoui, N., Castillo, C. and Seguin, J.-J.: Évaluation de l'exploitabilité des ressources en eau souterraine de la nappe de la craie du bassin de la Somme. Rapport final, BRGM/RP-63408-FR, Orléans~~.~~, 2014.

Barlage, M., Tewari, M., Chen, F., Miguez-Macho, G., Yang, Z.-L. and Niu, G.-Y.: The effect of groundwater interaction in North American regional climate simulations with WRF/Noah-MP, Clim. Change, 129(3), 485–498, doi:10.1007/s10584-014-1308-8, 2015.

Bessière, H., PIcot, J., Picot, G. and Parmentier, M.: Affinement du modèle hydrogéologique de la Craie du Nord-Pas-de-Calais autour des champs captants de la métropole Lilloise. Rapport final, BRGM/RP-63689-FR, Orléans~~.~~, 2015.

Boone, A., Masson, V., Meyers, T. and Noilhan, J.: The Influence of the Inclusion of Soil Freezing on Simulations by a Soil–Vegetation–Atmosphere Transfer Scheme, J. Appl. Meteorol., 39(9), 1544–1569, doi:10.1175/1520-0450(2000)039<1544:TIOTIO>2.0.CO;2, 2000.

Braud, I., Varado, N. and Olioso, A.: Comparison of root water uptake modules using either the surface energy balance or potential transpiration, J. Hydrol., 301(1–4), 267–286, doi:10.1016/j.jhydrol.2004.06.033, 2005.

Mis en forme : Police :+Corps (Times New Roman)

Buis, S., Piacentini, A. and Déclat, D.: PALM: a computational framework for assembling high-performance computing applications, Concurr. Comput. Pract. Exp., 18(2), 231–245, doi:10.1002/cpe.914, 2005.

Courtois, Q.: A hillslope-based aquifer model of free-surface flows in crystalline regions, p. 15, Saint-Malo. [online] Available from: https://www.irisa.fr/sage/jocelyne/CMWR2018/pdf/CMWR2018_paper_429.pdf, 2018.

Coustau, M., Rousset-Regimbeau, F., Thirel, G., Habets, F., Janet, B., Martin, E., de Saint-Aubin, C. and Soubeyroux, J.-M.: Impact of improved meteorological forcing, profile of soil hydraulic conductivity and data assimilation on an operational Hydrological Ensemble Forecast System over France, J. Hydrol., 525, 781–792, doi:10.1016/j.jhydrol.2015.04.022, 2015.

Croiset, N., Wuilleumier, A., Bessière, H., Gresselin, F. and Seguin, J.-J.: Modélisation des aquifères de la plaine de Caen et du bassin de la Dives. Phase 2 : construction et calage du modèle hydrogéologique, BRGM/RP-62648-FR, Orléans.,, 2013.

David, C. H., Habets, F., Maidment, D. R. and Yang, Z.-L.: RAPID applied to the SIM-France model, Hydrol. Process., 25(22), 3412–3425, doi:10.1002/hyp.8070, 2011.

De Lange, W. J., Prinsen, G. F., Hoogewoud, J. C., Veldhuizen, A. A., Verkaik, J., Oude Essink, G. H. P., van Walsum, P. E. V., Delsman, J. R., Hunink, J. C., Massop, H. Th. L. and Kroon, T.: An operational, multi-scale, multi-model system for consensus-based, integrated water management and policy analysis: The Netherlands Hydrological Instrument, Environ. Model. Softw., 59, 98–108, doi:10.1016/j.envsoft.2014.05.009, 2014.

Decharme, B., Martin, E. and Faroux, S.: Reconciling soil thermal and hydrological lower boundary conditions in land surface models, J. Geophys. Res. Atmospheres, 118(14), 7819–7834, doi:10.1002/jgrd.50631, 2013.

Douez, O.: Actualisation 2008-2011 du modèle maillé des aquifères du Jurassique en Poitou-Charentes. Rapport final, BRGM/RP-64816-FR, Orléans.,, 2015.

Duchaine, F., Jauré, S., Poitou, D., Quémerais, E., Staffelbach, G., Thierry Morel, T. and Gicquel, L.: Analysis of high performance conjugate heat transfer with the OpenPALM coupler, Comput. Sci. Discov., 8(1), 015003, doi:10.1088/1749-4699/8/1/015003, 2015.

Dudley, R. W., Hodgkins, G. A. and Dickinson, J. E.: Forecasting the Probability of Future Groundwater Levels Declining Below Specified Low Thresholds in the Conterminous U.S., JAWRA J. Am. Water Resour. Assoc., 53(6), 1424–1436, doi:10.1111/1752-1688.12582, 2017.

Fan, Y., Li, H. and Miguez-Macho, G.: Global Patterns of Groundwater Table Depth, Science, 339(6122), 940–943, doi:10.1126/science.1229881, 2013.

Guzman, S. M., Paz, J. O. and Tagert, M. L. M.: The Use of NARX Neural Networks to Forecast Daily Groundwater Levels, Water Resour. Manag., 31(5), 1591–1603, doi:10.1007/s11269-017-1598-5, 2017.

Habets, F., Boone, A., Champeaux, J. L., Etchevers, P., Franchistéguy, L., Leblois, E., Ledoux, E., Moigne, P. L., Martin, E., Morel, S., Noilhan, J., Quintana Seguí, P., Rousset-Regimbeau, F. and Viennot, P.: The SAFRAN-ISBA-MODCOU hydrometeorological model applied over France, J. Geophys. Res., 113(D6), D06113, doi:10.1029/2007JD008548, 2008.

Habets, F., Gascoin, S., Korkmaz, S., Thiéry, D., Zribi, M., Amraoui, N., Carli, M., Ducharne, A., Leblois, E., Ledoux, E., Martin, E., Noilhan, J., Ottlé, C. and Viennot, P.: Multi-model comparison of a major flood in the groundwater-fed basin of the Somme River (France), Hydrol Earth Syst Sci, 14(1), 99–117, doi:10.5194/hess-14-99-2010, 2010.

Habets, F., Amraoui, N., Caballero, Y., Thiéry, D., Vergnes, J.-P., Morel, T., Le Moigne, P., Roux, N., Dreuzy, J.-R. D., Ackerer, P., Maina, F., Besson, F., Etchevers, P., Regimbeau, F. and Viennot, P.: Plateforme de modélisation hydrogeologique nationale AQUI-FR : Rapport final de 1ère phase. (The national hydrogeological plateform Aqui-FR. Report of the first phase), ONEMA. [online] Available from: https://www.metis.upmc.fr/~aqui-fr/Rapport_fin_phase1_Aqui-FR_VF.pdf, 2017.

Hassane Maina, F., Delay, F. and Ackerer, P.: Estimating initial conditions for groundwater flow modeling using an adaptive inverse method, J. Hydrol., 552, 52–61, doi:10.1016/j.jhydrol.2017.06.041, 2017.

He, X., Stisen, S., Wiese, M. B. and Henriksen, H. J.: Designing a Hydrological Real-Time System for Surface Water and Groundwater in Denmark with Engagement of Stakeholders, Water Resour. Manag., 30(5), 1785–1802, doi:10.1007/s11269-016-1251-8, 2016.

Henriksen, H. J., Troldborg, L., Nyegaard, P., Sonnenborg, T. O., Refsgaard, J. C. and Madsen, B.: Methodology for construction, calibration and validation of a national hydrological model for Denmark, J. Hydrol., 280(1–4), 52–71, doi:10.1016/S0022-1694(03)00186-0, 2003.

Herbst, M., Fialkiewicz, W., Chen, T., Pütz, T., Thiéry, D., Mouvet, C., Vachaud, G. and Vereecken, H.: Intercomparison of Flow and Transport Models Applied to Vertical Drainage in Cropped Lysimeters, Vadose Zone J., 4(2), 240–254, doi:10.2136/vzj2004.0070, 2005.

Højberg, A. L., Troldborg, L., Stisen, S., Christensen, B. B. S. and Henriksen, H. J.: Stakeholder driven update and improvement of a national water resources model, Environ. Model. Softw., 40, 202–213, doi:10.1016/j.envsoft.2012.09.010, 2013.

Kollet, S., Gasper, F., Brdar, S., Goergen, K., Hendricks-Franssen, H.-J., Keune, J., Kurtz, W., Küll, V., Pappenberger, F., Poll, S., Trömel, S., Shrestha, P., Simmer, C. and Sulis, M.: Introduction of an Experimental Terrestrial Forecasting/Monitoring System at Regional to Continental Scales Based on the Terrestrial Systems Modeling Platform (v1.1.0), Water, 10(11), 1697, doi:10.3390/w10111697, 2018.

Korkmaz, S.: Modélisation des régimes de crue des systèmes couplés aquifères-rivières, thesis, Paris, ENMP, 1 January. [online] Available from: http://www.theses.fr/2007ENMP1495 (Accessed 30 November 2018), 2007.

Ledoux, E., Girard, G. and De Marsily, G.: Spatially distributed modelling: conceptual approach, coupling surface water and groundwater, in Unsaturated Flow in Hydrologic Modeling: Theory and Practice., vol. 275, pp. 435–454, H.J. Morel-Seytoux., 1989.

Legates, D. R. and McCabe Jr, G. J. M.: Evaluating the use of “goodness-of-fit” Measures in hydrologic and hydroclimatic model validation, Water Resour. Res., 35(1), 233–241, doi:10.1029/1998WR900018, 1999.

Long, D., Longuevergne, L. and Scanlon, B. R.: Global analysis of approaches for deriving total water storage changes from GRACE satellites, Water Resour. Res., 51(4), 2574–2594, doi:10.1002/2014WR016853, 2015.

Longuevergne, L., Scanlon, B. R. and Wilson, C. R.: GRACE Hydrological estimates for small basins: Evaluating processing approaches on the High Plains Aquifer, USA, Water Resour. Res., 46(11), doi:10.1029/2009WR008564, 2010.

Marçais, J., de Dreuzy, J.-R. and Erhel, J.: Dynamic coupling of subsurface and seepage flows solved within a regularized partition formulation, Adv. Water Resour., 109, 94–105, doi:10.1016/j.advwatres.2017.09.008, 2017.

Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Belamari, S., Barbu, A., Boone, A., Bouyssel, F., Brousseau, P., Brun, E., Calvet, J.-C., Carrer, D., Decharme, B., Delire, C., Donier, S., Essaouini, K., Gibelin, A.-L., Giordani, H., Habets, F., Jidane, M., Kerdraon, G., Kourzeneva, E., Lafaysse, M., Lafont, S., Lebeaupin Brossier, C., Lemonsu, A., Mahfouf, J.-F., Marguinaud, P., Mokhtari, M., Morin, S., Pigeon, G., Salgado, R., Seity, Y., Taillefer, F., Tanguy, G., Tulet, P., Vincendon, B., Vionnet, V. and Voldoire, A.: The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes, Geosci. Model Dev., 6(4), 929–960, doi:10.5194/gmd-6-929-2013, 2013.

Maxwell, R. M., Condon, L. E. and Kollet, S. J.: A high-resolution simulation of groundwater and surface water over most of the continental US with the integrated hydrologic model ParFlow v3, Geosci Model Dev, 8(3), 923–937, doi:10.5194/gmd-8-923-2015, 2015.

McKee, T. B., Doesken, N. J. and Kleist, J.: The relationship of drought frequency and duration to time scales, in Proceedings of the 8th Conference on Applied Climatology, vol. 17, pp. 179–183, American Meteorological Society Boston, MA., 1993.

Monteil, C., Flipo, N., Poulin, M., Habets, F., Krimissa, M. and Ledoux, E.: Assessing the contribution of the main aquifer units of the Loire basin to river discharge during low flow. [online] Available from: https://hal-mines-paristech.archives-ouvertes.fr/hal-00800812 (Accessed 8 January 2018), 2010.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, J. Hydrol., 10(3), 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.

Noilhan, J. and Planton, S.: A Simple Parameterization of Land Surface Processes for Meteorological Models, Mon. Weather Rev., 117(3), 536–549, doi:10.1175/1520-0493(1989)117<0536:ASPOLS>2.0.CO;2, 1989.

Noyer, M. L. and Elsass, P.: Modélisation de la salure profonde au droit et en aval du Bassin Potassique - Rapport final, BRGM/RP-54389-FR, Orléans.,, 2006.

Office International de l'Eau: Bulletin de situation hydrologique de janvier 2019, Office International de l'Eau, France. [online] Available from: https://www.eaufrance.fr/publications/bsh/2019-01 (Accessed 7 February 2019), 2019.

Pachocka, M., Mansour, M., Hughes, A. and Ward, R.: Challenges of modelling a complex multi-aquifer groundwater system at a national scale: case study from the UK, pp. 216–220, Colorado School of Mines, Colorado, USA. [online] Available from: http://igwmc.mines.edu/conference/Mod2015/MM15_Proceedings.pdf (Accessed 24 July 2018), 2015.

Pappenberger, F., Ramos, M. H., Cloke, H. L., Wetterhall, F., Alfieri, L., Bogner, K., Mueller, A. and Salamon, P.: How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction, J. Hydrol., 522, 697–713, doi:10.1016/j.jhydrol.2015.01.024, 2015.

Philip, S., Kew, S. F., Jan van Oldenborgh, G., Aalbers, E., Vautard, R., Otto, F., Haustein, K., Habets, F. and Singh, R.: Validation of a Rapid Attribution of the May/June 2016 Flood-Inducing Precipitation in France to Climate Change, J. Hydrometeorol., 19(11), 1881–1898, doi:10.1175/JHM-D-18-0074.1, 2018.

Prudhomme, C., Hannaford, J., Harrigan, S., Boorman, D., Knight, J., Bell, V., Jackson, C., Svensson, C., Parry, S., Bachiller-Jareno, N., Davies, H., Davis, R., Mackay, J., McKenzie, A., Rudd, A., Smith, K., Bloomfield, J., Ward, R. and Jenkins, A.: Hydrological Outlook UK: an operational streamflow and groundwater level forecasting system at monthly to seasonal time scales, Hydrol. Sci. J., 62(16), 2753–2768, doi:10.1080/02626667.2017.1395032, 2017.

Quintana-Seguí, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L. and Morel, S.: Analysis of Near-Surface Atmospheric Variables: Validation of the SAFRAN Analysis over France, J. Appl. Meteorol. Climatol., 47(1), 92–107, doi:10.1175/2007JAMC1636.1, 2008.

Rousset, F., Habets, F., Gomez, E., Moigne, P. L., Morel, S., Noilhan, J. and Ledoux, E.: Hydrometeorological modeling of the Seine basin using the SAFRAN-ISBA-MODCOU system, J. Geophys. Res., 109(D14), D14105, doi:10.1029/2003JD004403, 2004.

Roux, N.: The Aqui-FR project : the future operational modeling platform for the main regional multilayer aquifers in France, Saint-Malo. [online] Available from: https://www.irisa.fr/sage/jocelyne/CMWR2018/pdf/CMWR2018_paper_429.pdf, 2018.

Saleh, F., Ducharne, A., Flipo, N., Oudin, L. and Ledoux, E.: Impact of river bed morphology on discharge and water levels simulated by a 1D Saint–Venant hydraulic model at regional scale, J. Hydrol., 476, 169–177, doi:10.1016/j.jhydrol.2012.10.027, 2013.

Saltel, M., Wuilleumier, A. and Cabaret, O.: Gestion des eaux souterraines en Région Aquitaine - Développements et maintenance du Modèle Nord-Aquitain de gestion des nappes – Module 1 – Année 5 - Convention 2008-2013. Rapport final, BRGM/RP-65039-FR, Orléans., 2016.

Seguin, J.-J.: Proposition d'un indicateur piézométrique standardisé pour le Bulletin de Situation Hydrologique "Nappes". Rapport final, BRGM/RP-64147-FR, Orléans., 2015.

Seguin, J.-J. and Klinka, T.: Index Piézométrique Standardisé (IPS) - Bilans et comparaisons avec l'indicateur du BSH - Note technique, BRGM/RP-67251-FR, Orléans., 2016.

Seibert, J., Vis, M. J. P., Lewis, E. and Meerveld, H. J. van: Upper and lower benchmarks in hydrological modelling, Hydrol. Process., 32(8), 1120–1125, doi:10.1002/hyp.11476, 2018.

Singla, S., Céron, J.-P., Martin, E., Regimbeau, F., Déqué, M., Habets, F. and Vidal, J.-P.: Predictability of soil moisture and river flows over France for the spring season, Hydrol Earth Syst Sci, 16(1), 201–216, doi:10.5194/hess-16-201-2012, 2012.

Stollsteiner, P.: PROJET Explore 2070 - Evaluation de l'impact du changement climatique - Rapport final, BRGM/RP-61483-FR, Orléans., 2012.

Sulis, M., Keune, J., Shrestha, P., Simmer, C. and Kollet, S. J.: Quantifying the Impact of Subsurface-Land Surface Physical Processes on the Predictive Skill of Subseasonal Mesoscale Atmospheric Simulations, J. Geophys. Res. Atmospheres, 123(17), 9131–9151, doi:10.1029/2017JD028187, 2018.

Texas Water Development Board: Groundwater Division | Texas Water Development Board, [online] Available from: https://www.twdb.texas.gov/groundwater/ (Accessed 6 November 2018), 2018.

Thierion, C.: Modélisation hydrogéologique des aquifères du Bajocien - Bathonien dans le secteur des Bocages Normands, Master 2, Université Pierre et Marie Curie, Ecole des Mines de Paris & Ecole Nationale du Génie Rural des Eaux et des Forêts., 2007.

Thierion, C., Longuevergne, L., Habets, F., Ledoux, E., Ackerer, P., Majdalani, S., Leblois, E., Lecluse, S., Martin, E., Queguiner, S. and Viennot, P.: Assessing the water balance of the Upper Rhine Graben hydrosystem, J. Hydrol., 424–425, 68–83, doi:10.1016/j.jhydrol.2011.12.028, 2012.

Thiéry, D.: Logiciel GARDÉNIA, version 8.2. Guide d'utilisation. Rapport final, BRGM/RP-62797-FR, Orléans, 2014.

Thiéry, D.: Code de calcul MARTHE - Modélisation 3D des écoulements dans les hydrosystèmes - Notice d'utilisation de la version 7.5 (MARTHE: Modeling software for groundwater flows), BRGM/RP-64554-FR, Orléans., 2015a.

Thiéry, D.: Modélisation 3D des écoulements en Zone Non Saturée avec le code de calcul MARTHE - version 7.5, BRGM/RP-64495-FR, Orléans.,, 2015b.

Thiéry, D.: Modélisation 3D du transport de masse et d'énergie avec le code de calcul MARTHE - version 7.5, BRGM/RP-64765-FR, Orléans.,, 2015c.

Thiéry, D.: Validation du code de calcul GARDÉNIA par modélisations physiques comparatives, BRGM/RP-64500-FR, Orléans.,, 2015d.

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.

Thiéry, D. and Moutzopoulos, C.: Un modèle hydrologique spatialisé pour la simulation de très grands bassins : le modèle EROS formé de grappes de modèles globaux élémentaires, pp. 285–295, ORSTOM Editions. [online] Available from: https://hal-brgm.archives-ouvertes.fr/hal-01061971/document (Accessed 10 January 2018), 1992.

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.

Thirel, G., Martin, E., Mahfouf, J.-F., Massart, S., Ricci, S. and Habets, F.: A past discharges assimilation system for ensemble streamflow forecasts over France – Part 1: Description and validation of the assimilation system, Hydrol Earth Syst Sci, 14(8), 1623–1637, doi:10.5194/hess-14-1623-2010, 2010.

Vergnes, J.-P. and Decharme, B.: A simple groundwater scheme in the TRIP river routing model: global off-line evaluation against GRACE terrestrial water storage estimates and observed river discharges, Hydrol Earth Syst Sci, 16(10), 3889–3908, doi:10.5194/hess-16-3889-2012, 2012.

Vergnes, J.-P. and Habets, F.: Impact of river water levels on the simulation of stream–aquifer exchanges over the Upper Rhine alluvial aquifer (France/Germany), Hydrogeol. J., 1–15, doi:10.1007/s10040-018-1788-0, 2018.

Vergnes, J.-P., Decharme, B., Alkama, R., Martin, E., Habets, F. and Douville, H.: A Simple Groundwater Scheme for Hydrological and Climate Applications: Description and Offline Evaluation over France, J. Hydrometeorol., 13(4), 1149–1171, doi:10.1175/JHM-D-11-0149.1, 2012.

Vergnes, J.-P., Decharme, B. and Habets, F.: Introduction of groundwater capillary rises using subgrid spatial variability of topography into the ISBA land surface model, J. Geophys. Res. Atmospheres, 119(19), 2014JD021573, doi:10.1002/2014JD021573, 2014.

Vidal, J., Martin, E., Franchistéguy, L., Baillon, M. and Soubeyroux, J.: A 50-year high-resolution atmospheric reanalysis over France with the Safran system, Int. J. Climatol., 30(11), 1627–1644, doi:10.1002/joc.2003, 2010.

Viennot, P.: Modélisation mathématique du fonctionnement hydrogéologique du bassin de la Seine. Représentation différentiée des aquifères du Tertiaire/Intégration des prélèvements anthropiques, Centre de GéosciencesR071128PVIE, Centre de Géosciences, Fontainebleau.,, 2009.

Viennot, P. and Abasq, L.: Modélisation de la pollution nitrique des grands aquifères du bassin de Seine-Normandie à l'échelle des masses d'eau. Développement des sous modèles hydrogéologiques "Craie" du bassin de la Seine. Rapport final, Centre de Géosciences, Fontainebleau.,, 2013.

Westerhoff, R., White, P. and Rawlinson, Z.: Incorporation of Satellite Data and Uncertainty in a Nationwide Groundwater Recharge Model in New Zealand, Remote Sens., 10(1), 58, doi:10.3390/rs10010058, 2018.

Figure 1~~: Scheme of the AquiFR coupling platform~~: 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 1000 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 manages the following time step.

Figure 3: Main aquifers of France classified by geological type from the BDLISA database (https://bdlisa.eaufrance.fr/). The names of the gauging stations and piezometers shown in Figure 8, Figure 13 and Figure 14 are written.

Figure 4: Map of the regional multilayer aquifers and the karstic systems simulated in AquiFR. The outlines of the models are also shown with colours corresponding to the outcropping aquifers with respect to their geological contexts. Grey areas correspond to models that will be integrated in a near future.

Figure 5: Temporal evolution of the number of piezometric head measurements per day among the 639 selected piezometers over the 1958-2018 simulated period.

Figure 6: (a) Spatial distribution of the ~~bias scores~~biases calculated between the simulated and observed piezometric heads for the 639 selected piezometers. The grey background colour corresponds to the simulated aquifer domain. (b) Accumulated distribution of absolute ~~bias scores~~biases for all the piezometers.

Figure 7: (a) Spatial distribution of NRMSE~~-~~_BE ~~scores~~ calculated between the simulated and observed piezometric heads for the 639 selected piezometers. The grey background colour corresponds to the simulated aquifer domain. (b) Accumulated distribution of NRMSE~~-~~_BE ~~scores~~ for all the piezometers.

Figure 8: Daily observed (dotted blue) and simulated (red) piezometric head variations for the five piezometers encircled in green in Figure 3.

Figure 9: NSE criterion calculated between the observed and simulated SPLI for the 103 selected piezometers.

Figure 10~~: Correlation scores~~: Correlations calculated between the observed and simulated SPLI for the 103 selected piezometers.

Figure 11: (left) Monthly observed (dotted blue) and simulated (red) SPLI indicator variations for the five piezometers encircled in green in Figure 3. Font colours correspond to the classes of Table 3 from the driest (red) to the wettest (blue) intervals. (Right) Histograms in percentage of the SPLI values distributed against the classes of Table 3.

Figure 12~~: SPLI indicators calculated~~: Standardised Piezometric Level Index for the august 2016 month for the (a) observed and (b) simulated piezometers. The SPLI are computed on the period for which the observations are available. (b) The map of the SPLI indicators calculated ~~for all the piezometric heads of the aquifer~~in each grid ~~are also~~cell of the AquiFR domain with a common reference period is shown in the background.

Figure 13: Monthly observed (dotted blue) and simulated (red) river flows of the gauging stations monitoring the four karstic systems encircled in red in Figure 3. NSE ~~scores for~~of the square root of the daily ~~river~~karstic spring flows and drainage area are given in parenthesis for each gauging station.

Figure 14: Daily observed (dotted blue) and simulated (red) river flows for the four gauging stations encircled in yellow in Figure 3.

Figure 15: (a) Spatial distribution of the NSE ~~scores~~ calculated between the observed and simulated karstic spring flows and river flows for the gauging stations monitoring the 23 simulated karstic systems (stars) and the ~~228~~362 selected gauging stations located in the distributed hydrogeological models (circles). Accumulated distribution of NSE ~~scores~~ for (a) the 23 karstic systems and (b) the ~~228~~362 gauging stations of the distributed models are also shown. The NSE ~~scores~~ for karstic systems are computed using the square root of the daily karstic spring flows. The simulated river network is shown in the background.

Figure 16 : (a) Spatial average over the Somme model of daily mean groundwater recharge, (b) daily mean ~~stream~~river-groundwater exchanges budget over the ~~Some~~Somme river network, (c) spatial average over the Somme model of the monthly mean piezometric head and (d) monthly SPLI. Red colors in (b) indicate groundwater to river flows and blue colors stream to groundwater flows. The background colors in (d) correspond to the SPLI classes from Table 3.

Table 1: Short description of the regional multilayer aquifer models available in AquiFR

Software	Model	Number of layers	Number of cells	References	Recalibration
EauDyssée	Basse-Normandie	4	37 667	(Thierion, 2007)	Yes
	Loire	3	37 620	(Monteil et al., 2010)	No
	Marne-Loing	4	66 235	(Viennot and Abasq, 2013)	Yes
	Marne-Oise	2	45 904	(Viennot and Abasq, 2013)	Yes
	Seine	6	41 609	~~(Viennot, 2009)~~(Viennot, 2009)	Not necessary
	Seine-Eure	1	57 306	(Viennot and Abasq, 2013)	In progress
	Seine-Oise	4	87 178	(Viennot and Abasq, 2013)	Yes
	Somme	1	63 226	(Korkmaz, 2007)	No
MARTHE	Alsace	3	40 947	(Noyer and Elsass, 2006)	No
	Basse-Normandie	10	93 800	(Croiset et al., 2013)	No
	Nord Pas-de-Calais	10	226 077	(Bessière et al., 2015)	Yes
	Poitou-Charentes	8	90 084	(Douez, 2015)	Not necessary
	Somme	1	66 924	(Amraoui et al., 2014)	Yes

Tableau mis en forme

Mis en forme : Justifié

Cellules insérées

Table 2: Statistical scores of the comparison between the simulated and observed daily evolution of the piezometers shown in Figure 8.

Piezometer	Model	Time series			SPLI	
		NRMSE-BE	Correlation	Biases (m)	NSE	Correlation
Omiécourt	Somme	0.~~81~~93	0.~~77~~85	-0.~~15~~86	0.~~65~~73	0.~~83~~87
Ruffec	Poitou-Charentes	0.58	0.82	-1.44	0.6	0.79
Le Bec Hellouin	Basse-Normandie	0.57	0.84	-2.76	0.73	0.86
Farceaux	Seine-Oise	0.52	0.86	-8.34	0.67	0.84
Bourdet	Poitou-Charentes	1.02	0.18	-0.72	-0.51	0.24

Table 3: Classification of water table level classes related to the values of the SPLI corresponding to the MSH limits

Classification	SPLI values	Return periods
Very low groundwater level	< -1.28	> 10 dry years

Low groundwater level	between -1.28 and -0.84	Between 10 dry years and 5 dry years
Moderately low groundwater level	between -0.84 and -0.25	Between 5 dry years and 2.5 dry years
Normal groundwater level	between -0.25 and 0.25	Between 2.5 dry years and 2.5 wet years
Moderately high groundwater level	between 0.25 and 0.84	Between 2.5 wet years and 5 wet years
High groundwater level	between 0.84 and 1.28	Between 5 wet years and 10 wet years
Very high groundwater level	> 1.28	> 10 wet years

Table 4: Statistical scores of the comparison between the simulated and observed daily evolution of river discharges shown in Figure 14.

Gauging Station	NSE	Correlation	Ratio
Charente at Jarnac	0.45	0.72	0.81
Somme at Abbeville	0.69	0.86	0.92
Seine at Poses	0.86	0.93	1.01
Loire at Montjean	0.94	0.97	1.05