Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-166-RC2, 2019

Hydrology and Earth System Sciences
Discussions

Interactive comment on "The AquiFR hydrometeorological modelling platform as a tool for improving groundwater resource monitoring over France: evaluation over a 60 year period" *by* Jean-Pierre Vergnes et al.

Anonymous Referee #2

Received and published: 22 July 2019

General comments ==========

This article presents a new platform to simulate groundwater in France. The platform puts together two different hydrogeological models and one lumped model system for karst aquifers. The objective is to be able to simulate the whole of France with one set of tools. At this stage, about half of France is implemented. This aims at being and extension of the SIM France hydrometeorological model. This is a very interesting approach to national-scale groundwater modeling and thus deserves publication. Furthermore, AquiFR will probably play an important role in water management and

research in France in the future, so the paper is needed.

The main strength of the AquiFR approach (a sort of confederation of models) is also its weaknesses (how do you generate homogeneity from such an heterogeneous set of models?). AquiFR is not one model, it is a set of models that simulate different aquifers using the same forcing. This has allowed to put a large scale high resolution hydrogeological model together much faster than if a single model was run from scratch, which is very valuable. We should not forget that these models have a long history behind them and thus have had different feedback cycles with stakeholders. That would be very difficult to do from scratch. However, the very nature of AquiFR makes interpreting the results difficult, as different systems are simulated with different models and, thus, the results will not be spatially homogeneous. However, the model is validated almost as if it was one model, without discussing in detail the role of each of the models. In order to compensat for the heterogeneity, all the models are forced with the recharge that comes from SURFEX, which seems appropriate, but there are some exceptions, for example, in one case the partition between surface runoff and recharge is recalculated with GARDENIA or in another the model is also fed by streamflow observations. Achieving homogeneity is difficult with so many different models and approaches. The problem is that, as the text is now, all this is a little bit confusing and it is very difficult to get a clear picture of how the model works.

As I wrote in the first paragraph, the article is valuable and deserves to be published, but first it must undergo an improvement of the text in order to clarify how the model works, how it has been calibrated and how it has been validated.

The article explains the model architecture, but not with enough detail. I have many questions on this topic that should be clarified in the text:

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.

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

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.

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?

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?

6. It is not discussed if the 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?

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.

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.

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.

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

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

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.

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.

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.

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.

Using the SPLI is appropriate, but you don't detail enough how you calculate it.

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

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

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 an standardized variable involves renormalizing the data to a normal distribution. Why these figures don't show a normal distribution?

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.

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?

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.).

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

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

* P6L17: "gathersnumerical" should be separated.

* P6L29: coupledto should be separated.

* 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.

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

* P9L17: observationsat should be divided.

* P11L8: It sensitivity -> Its sensitivity.

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

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

* 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?

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

* P14L6: Which periods were used for calibration?

* 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.

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

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

* Fig11b: difficult to see the circles.

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

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-166, 2019.