

Water scarcity has become a crucial issue in France: during the 2019-2020 drought, several areas experienced drinking water supply interruption; water transportation through tank trucks was needed in areas that are usually regional *water towers*, like Jura or Southern Vosges. Reliable low-flow forecasting tools would allow decision-makers to design and implement public policies that would help anticipating and mitigating such crisis, especially when they result from long-term hydrological dynamics in catchments. In many regions, these long-term phenomena are borne by aquifers, whose level is monitored by a large network of piezometers available all over the French territory. Can piezometric data help improving low-flow simulation through a simple calibration adaptation ?

1 Introduction: piezometry in low-flow modelling

Aquifers are a key element of water cycle, especially during droughts, thanks to their water storage capacity: they store recharge water during rainy periods and they contribute to streamflow during rainless ones. The alternation between recharge and discharge periods of aquifers is generally seasonal in temperate regions, at least for the ones that are connected to the surface water cycle, but it is also affected by longer climatic variability. Severe droughts often result from an accumulation of rainfall deficit over several years, which causes aquifers to reach extreme low levels.

Therefore, measuring the water table level of aquifers – through piezometers – gives information about the water storage content of the corresponding catchment. To write it in a simpler way, a catchment beginning its dry season with low aquifers will probably have a drier trajectory than the same catchment with high piezometric level. From a hydrological modelling perspective, piezometry is an information that can be used to evaluate the states reflecting water content, at a given time step, of the model.

So-called *physically-based* distributed models generally compute piezometry explicitly, which allows modellers to use piezometric data as calibration or validation data (Vergnes et al., 2020). The situation is less straightforward for lumped conceptual models, whose states do not correspond directly to physical measurable data. Specific models have been developed to simulate only piezometry, without streamflow (Mackay et al., 2014), but we are dealing here with rainfall-runoff models: to represent piezometry while modelling river flow, it is necessary to identify a particular state in the model structure that can correspond to piezometry. It is the principle of the Gardénia model (Thiery, 1988, 2014), which *ground* reservoir is identified to aquifer storage and is calibrated using piezometric data.

In this work, we incorporate piezometric data into GR6J, a daily lumped conceptual hydrological model, through a composite calibration combining streamflow and piezometry, without changing the model's structure. Simulation performance are compared to the reference calibration.

2 Material and methods

This study was conducted using a dataset of 107 catchments in Metropolitan France, on which long time series – more than 20 years – of piezometry and streamflow data are available. For each catchment, the local hydrogeological context was analysed and one to five relevant piezometers – which monitor aquifers that are connected to surface water bodies – were selected. All in all, the dataset comprises 160 catchment / piezometer pairs, for which available data have at least ten years in common. Rainfall and potential evaporation data were taken from SAFRAN reanalysis (Vidal et al., 2010).

The analysis uses a modified version of the daily lumped conceptual rainfall-runoff model GR6J (Pushpalatha et al., 2011), to which one state was added to explicitly simulate a piezometric anomaly. The model is composed of three conceptual reservoirs (the production one, the routing one and the exponential one); their levels are the main states of the model and since they represent water contents, we examined whether one of them could be used to represent groundwater content and thus, piezometry.

Figure 1 : modified GR6J scheme

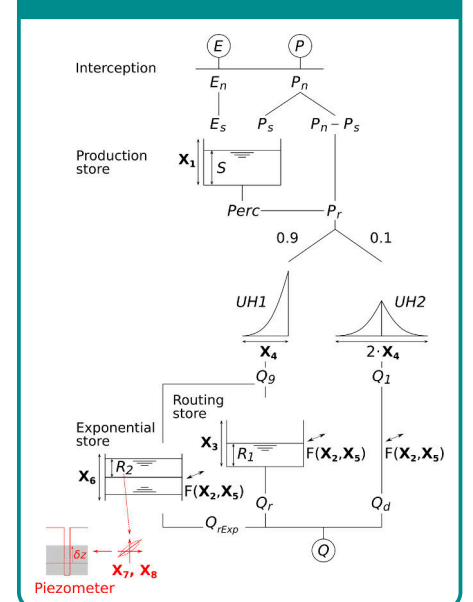
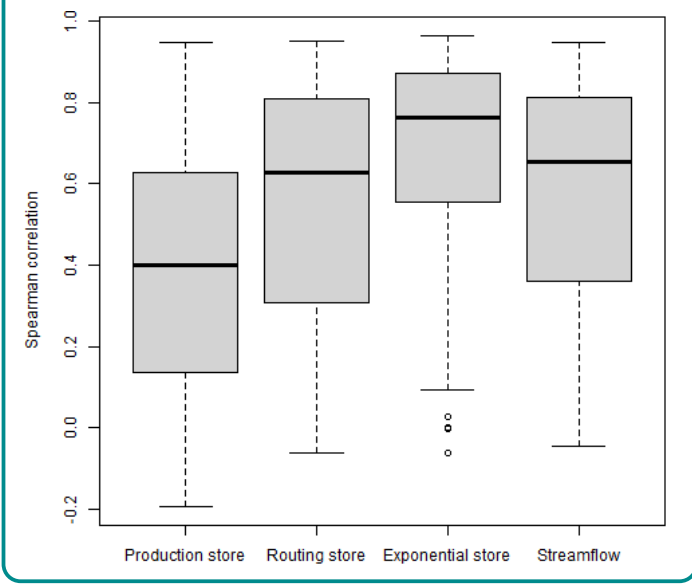


Figure 2 : distributions of correlation between model states and piezometry



The original version of GR6J – in black on figure 1 – was calibrated on the 107 catchments, using the Nash-Sutcliffe efficiency criterion with square root transformation of streamflow and all available hydrometric data on the 1958-2018 period – on which SAFRAN climatic data is available. Then, correlations between piezometry and monthly means of the three reservoirs' levels, plus simulated streamflow, were computed for each of the 160 catchment / piezometer pairs. Results are shown on figure 2.

It appears that the exponential reservoir is the most correlated with piezometry, in the majority of cases with a linear relationship between the two variables. To avoid scale issues, piezometric data are converted to anomalies δz , using the following reduction, σ_z being the standard deviation and $\bar{z}$ the mean of each series:

$$\delta z = \frac{z - \bar{z}}{\sigma_z}$$

We can now adapt the model to simulate piezometry as an output variable, through a simple linear function of exponential store level; it adds two new parameters, X_7 and X_8 :

$$\delta z = \frac{1}{X_7} \left(\frac{Exp}{X_6} + X_8 \right)$$

X_7 is positive – typically between 0 and 5 – and X_8 is non-negative – between 0 and 20. As shown in red on figure 1, there is no direct change on the streamflow modelling chain, but calibrating the model on both observed streamflow and piezometry can lead to a different set of parameters than usual calibration performed on streamflow only and therefore, taking in account piezometry can change streamflow modelling.

A composite criterion taking in account streamflow and piezometry is used. For streamflow, usual Nash-Sutcliffe efficiency with square-root transformation is kept; for piezometry, by analogy with NSE, we use $ZError$, defined by the following formula:

$$ZError = 1 - \sum_{t=1}^T (\delta z_{sim}(t) - \delta z_{obs}(t))^2$$

Like NSE, it is equal to 1 in case of perfect simulation and the higher it is, the closer simulated and observed piezometric anomalies are. The composite criterion is a linear combination of these two criteria, with a weight α between 0 and 1:

$$Crit(\alpha) = \alpha ZError + (1 - \alpha) NSE$$

To assess the impact of the new calibration scheme on model performance, a split-sample test is performed: for each catchment / piezometer pair, available observed data are divided into two equal periods P_1 and P_2 , with a concern for contemporaneity between observed piezometry and streamflow. Calibration is performed on P_1 and the model is then evaluated on P_2 , using the parameter set from the calibration, and inversely. For the sake of simplicity, Nash-Sutcliffe efficiency with square-root transformation is used as validation criterion. Other validation criteria more focused on low-flows – such as NSE with cubic root transformation – were tried but they did not highlight different trends.

3 Results

51 values of α have been tested, from 0 to 1 with a 0.02 step. Figure 3 shows the number of catchments with a significant increase ($\Delta NSE > 0.05$) or decrease ($\Delta NSE < -0.05$) of the validation criterion, with respect to the reference value obtained for $\alpha = 0$. It appears that for values of α greater than 0.3, there are more catchments on which the model performance are deteriorated than the ones on which it is improved. At most, performances are increased on 13 catchments out of 107, for $\alpha = 0.04, 0.06, 0.16$ or 0.22 .

To evaluate the model performances on the whole set of catchments, a Wilcoxon-Mann-Whitney test was used to compare the set of validation criteria obtained, with $\alpha = 0$ for NSE and $\alpha = 1$ for $ZError$, to the same set obtained for other α values. Results are shown on figure 4.

Figure 3 : number of catchments with a significant change in validation criterion

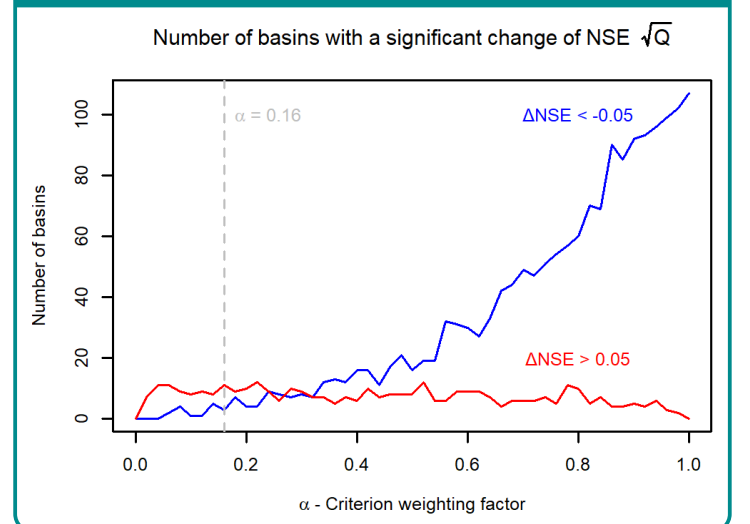
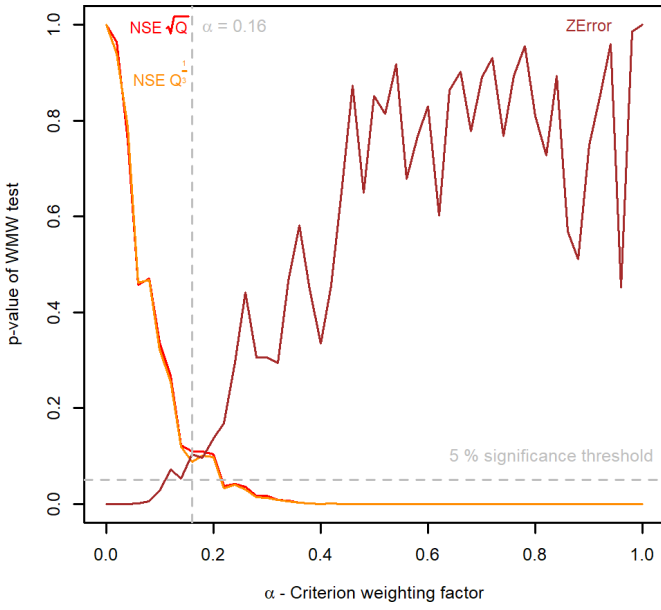


Figure 4 : result of the Wilcoxon-Mann-Whitney test



The statistical test compares the medians of the criteria sets. It appears that for $\alpha > 0.2$, performance are significantly deteriorated for streamflow; for $\alpha > 0.14$, the model is able to simulate piezometry with no significant difference with respect to $\alpha = 1$. Therefore, the value $\alpha = 0.16$ seems to be a good compromise between performance in simulating streamflow and piezometry.

It can be seen more precisely on figure 5, which shows the performance distribution on the catchment set for the different α values, with respect to reference criterion values – the red dotted lines, $\Delta NSE = 0$ for streamflow and median ZError for $\alpha = 1$ for piezometry. Performance in streamflow simulation visibly decrease above $\alpha = 0.3$ and for piezometry, under $\alpha = 0.12$. Here again, $\alpha = 0.16$ seems to be an acceptable compromise, even though nearby values between 0.12 and 0.3 would give a similar result. The median ZError is 0.71 with more than $\frac{3}{4}$ of piezometers above 0.3, which is a noteworthy performance.

Figure 5 : distributions of model validation performance for different α values

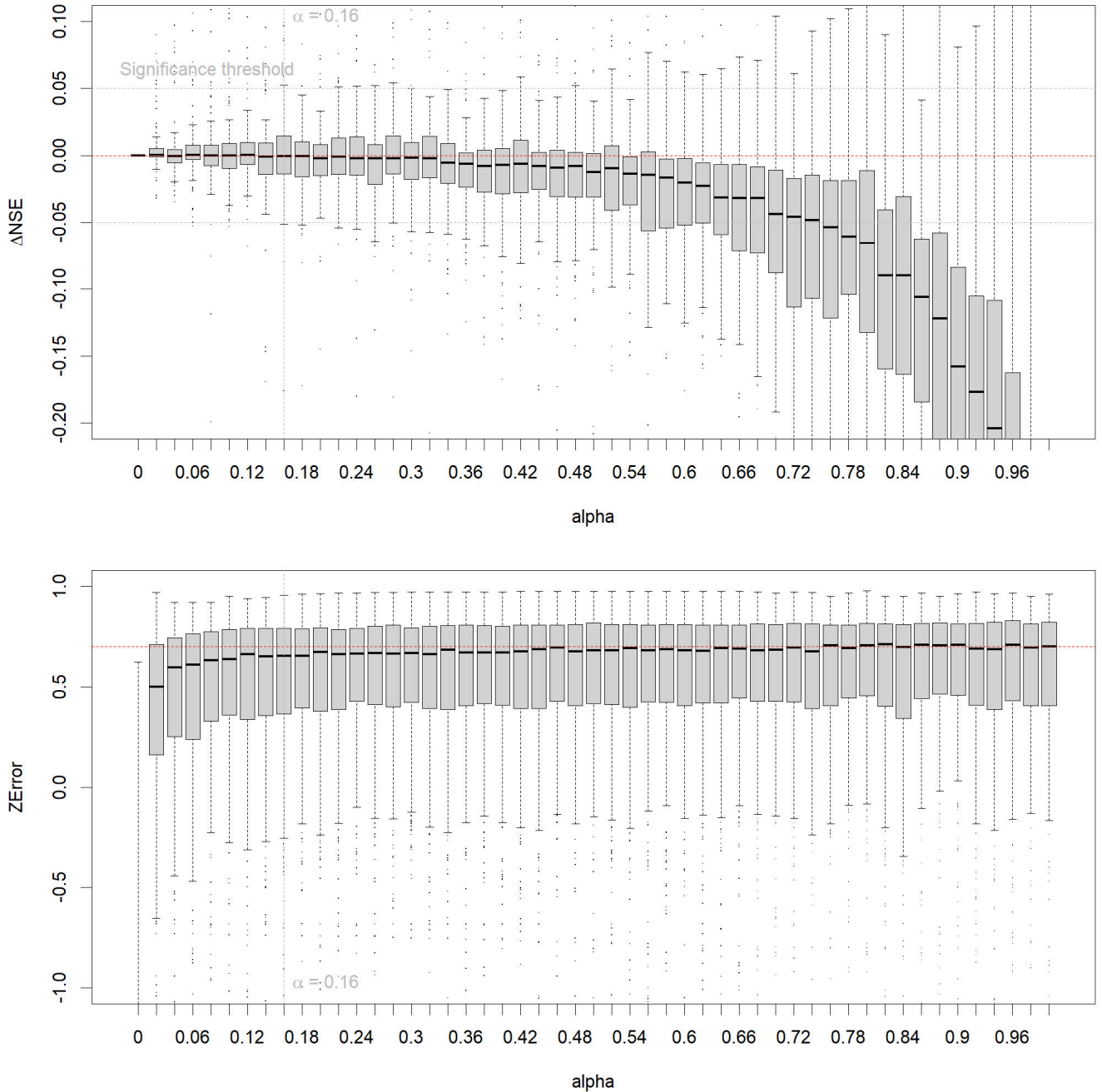
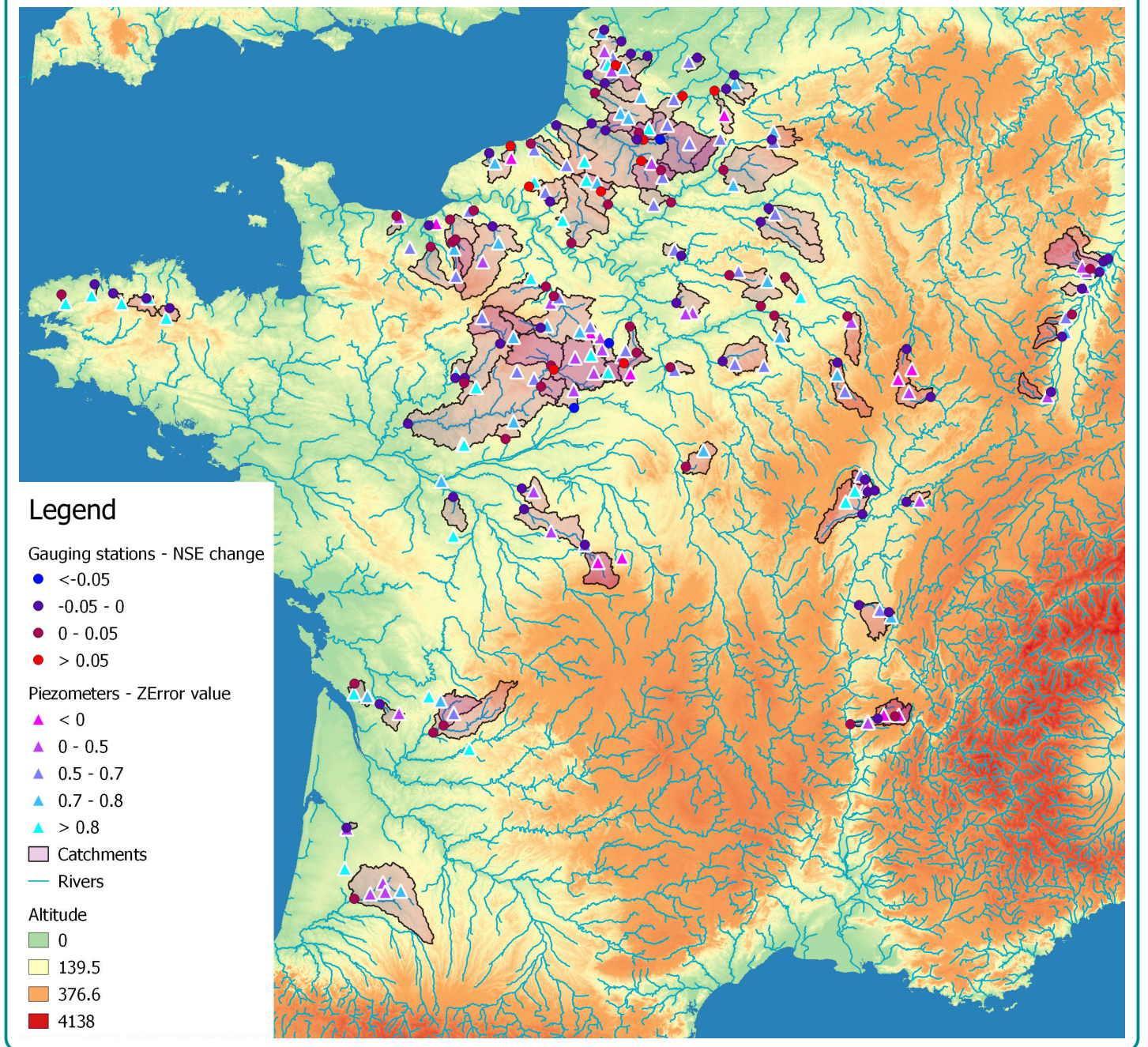


Figure 6 : map of the results for $\alpha = 0.16$



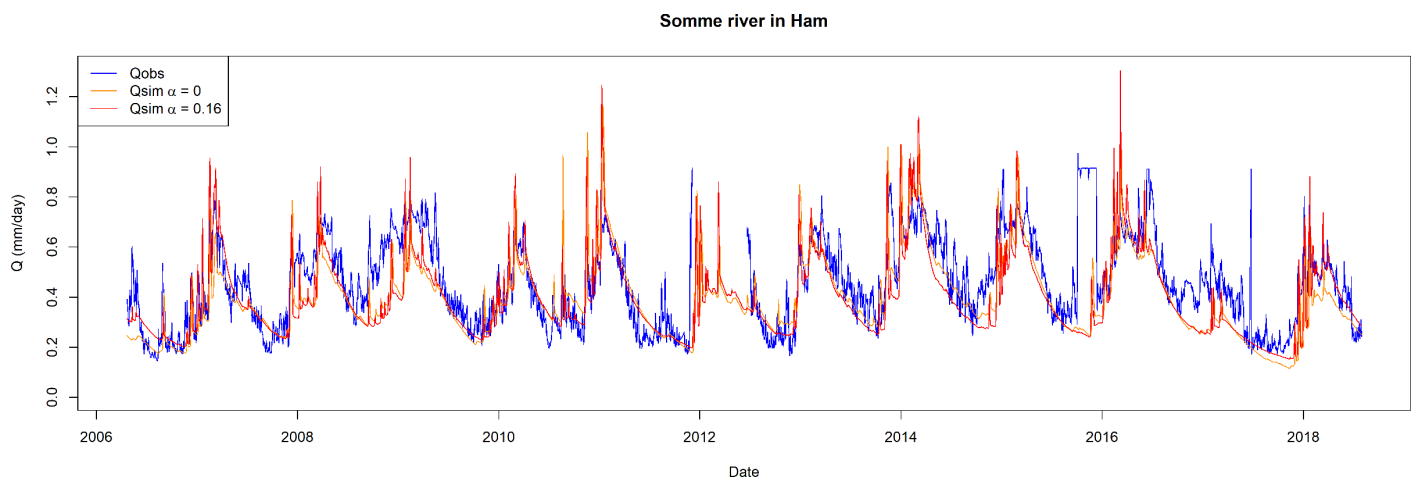
The geographical configuration of the results can be seen on figure 6. For each catchment, the color of gauging station on the map shows the improvement of the best NSE among the available piezometers. For each piezometer, the best ZError among catchments that are coupled with the piezometer is highlighted.

No obvious spatial pattern emerges from the map, but we can see that surface aquifers that are closely linked to surface water bodies, like in Brittany or in the Saone basin, have high performance in simulating piezometry without improving streamflow simulation. Results are very heterogeneous in the Beauce region, where catchments are very influenced by a large slow aquifer and thus where we would have expected significant gains from taking in account piezometric data. This can be mostly explained by the low performance of the original model due to water balance issues: indeed, due to the very flat topography of the region and the huge part of groundwater flows in water transport, topographic

catchments can be really different from real ones and therefore, the model has *too much* or *too little* water to correctly simulate streamflow. The same issue is visible in the Bièvre glacial plain in the Rhone valley.

It is not surprising that the few piezometers in the dataset that are in karstic aquifers – in the Berry limestones or in the upper Meuse basin – do not improve streamflow simulation and have low ZError. Most of the catchments were performance are significantly improved – bright red dots – are in the chalky part of Paris basin, but the results are far from being homogeneous in this region. Figure 6 shows the simulated and observed hydrographs of the Somme river in Ham, one the most performant catchments – $\Delta NSE = 0.15$. The simulation of low-flows is a bit improved in some parts of the hydrograph, but the result is hardly visible to the naked eye. However, the piezometry simulation – not shown here – is clearly closer than what was obtained with the original calibration.

Figure 6 : simulated and observed hydrographs of the Somme river in Ham



4 Conclusion: should we be disappointed?

After seeing the insufficiency of the GR6J model in simulating low-flows, especially for catchments encountering long and slow droughts, we examined whether an adapted calibration scheme, incorporating piezometric data as new information, would improve the model performance. It appears that the change in streamflow modelling score is generally minimal, which can be seen as disappointing. However, the model turned out to be unexpectedly talented in piezometry simulation.

Despite its small number of parameters, GR6J is, to

some extent, affected by equifinality, i.e. several parameter sets can give the same modelling performance. Among these sets, in many cases, one exists that is able to correctly simulate piezometry. By sitting the model on more data, more information about the catchment, we can make it more trustworthy for operational users and develop other uses from piezometry simulation.

Anyway, this new calibration scheme is a departure point for studies on adapting the model structure, which is so far unable to correctly represent both long-term (Fowler et al., 2020) and short-term dynamics. Further developments are also needed for catchments influenced by several aquifers.

References

- K. Fowler, W. Knoben, M. Peel, T. Peterson, D. Ryu, M. Saft, K.-W. Seo, and A. Western. 2020. Many Commonly Used Rainfall-Runoff Models Lack Long, Slow Dynamics: Implications for Runoff Projections. *Water Resources Research* 56, 5 (2020), e2019WR025286. <https://doi.org/10.1029/2019WR025286>
- J.D. Mackay, C.R. Jackson, and L. Wang. 2014. A lumped conceptual model to simulate groundwater level time-series. *Environmental Modelling and Software* 61 (2014), 229–245. <https://doi.org/10.1016/j.envsoft.2014.06.003>
- R. Pushpalatha, C. Perrin, N. Le Moine, T. Mathévet, and V. Andréassian. 2011. A downward structural sensitivity analysis of hydrological models to improve low-flow simulation. *Journal of Hydrology* 411, 1 (2011), 66–76. <https://doi.org/10.1016/j.jhydrol.2011.09.034>
- D. Thiéry. 1988. Forecast of changes in piezometric levels by a lumped hydrological model. *Journal of Hydrology* 97, 1 (1988), 129–148. [https://doi.org/10.1016/0022-1694\(88\)90070-4](https://doi.org/10.1016/0022-1694(88)90070-4)
- D. Thiéry. 2014. *Logiciel GARDÉNIA, version v8.2. Guide d'utilisation*. BRGM, Orléans, France. <https://www.brgm.fr/sites/default/files/documents/2020-11/logiciel-gardenia-v8-2-rp-62797-fr-notice.pdf> BRGM report RP-62797-FR.
- J.-P. Vergnes, N. Roux, F. Habets, P. Ackerer, N. Amraoui, F. Besson, Y. Caballero, Q. Courtois, J.-R. de Dreuzy, P. Etchevers, N. Gallois, D. J. Leroux, L. Longuevergne, P. Le Moigne, T. Morel, S. Munier, F. Regimbeau, D. Thiéry, and P. Viennot. 2020. The AquifR hydrometeorological modelling platform as a tool for improving groundwater resource monitoring over France: evaluation over a 60-year period. *Hydrology and Earth System Sciences* 24, 2 (2020), 633–654. <https://doi.org/10.5194/hess-24-633-2020>
- J.-P. Vidal, E. Martin, L. Franchistéguy, M. Baillon, and J.-M. Soubeyroux. 2010. A 50-year high-resolution atmospheric reanalysis over France with the Safran system. *International Journal of Climatology* 30, 11 (2010), 1627–1644. <https://doi.org/10.1002/joc.2003>