The first thing we want to mention is that thanks to the reviewers' comments, we have become aware of two general aspects must be improved in the final manuscript. The first aspect is the necessity of smoothing the general tone of the manuscript, since it is transmitting an impression, which is far from our sincere intention. Secondly, it seems we have not explained properly the main target of the paper: the aim was not to find the best error model (neither the best hydrological model) which yields the best performance for the FB basin (see our **reply RC1#6** or our **reply RC2#10** to reviewer-2 for more details, among others, those concerning the metrics for reliability and resolution). The objective of this paper is to construct the full general additive error model developed in Schoups and Vrugt (2010) with the recommendations of Evin et al. (2013) within a "strict" Bayesian joint inference framework. We want to be "strict" following the reflection of Todini and Mantovan (2007): "*Statistical scientists will have a very low regard for the hydrological sciences if we, the hydrologists, pretend to use statistical techniques, but then deliver theoretically incorrect answers and results*". Of course the WLS-AR1-PP and other post-processing solutions can work, but they eliminate potential positive interactions between the two models (hydrological and error) during the parameter inference process and, for this reason, we have not considered them in our research.

Therefore, our initial hypothesis was that, by "merging" both mentioned approaches, we could perform Bayesian joint inference on our models. However, after the construction of the Bayesian joint inference framework, we saw that problems remained. This major setback led us to thinking about the basics of error modeling. We understood the mistake looking at the most basic error model (**Figure 1a** of original manuscript), which underlies to the simple Least Squares method (SLS). The joint inference, to be statistically correct, should assume the existence of the joint probability distribution of the variable to be predicted and its deviation from its observation (the error). Consequently, the relationship between the marginal and conditional distributions of this joint distribution must be taken into account in the inference process. The Total Laws define this relationship, resulting in a reduction of the degrees of freedom in the inference problem. We think nobody questioned these issues in the past, perhaps because with SLS error model it was not needed, since SLS meets TLs by its own hypotheses. But with more complex error models, we should take up those old Laws again.

Once we put on board the Total Laws, we wanted to highlight in one case study the problems that can arise without their application, some of them found in previous researches. With this objective in mind, we think we do not need to unnecessarily enlarge the paper with more case studies, because with FB basin we have already found the following spurious problems:

1- Meaningless enlargement of the uncertainty bands with both CRR and GR4J hydrological models; this problem was also found by Schoups and Vrugt (2010), Evin et al. (2013, 2014) and Scharnagl et al. (2015)
2- Non-identifiable autocorrelation parameter with CRR model; this problem was also found by Schoups and Vrugt (2010) in their second case study, and by Scharnagl et al. (2015) with their Likelihood2.
3- Very high, but identifiable, value for the autocorrelation parameter, with GR4J model; Evin et al. (2013, 2014) also reported this feature. In our opinion and in

some cases, the non-suitability of the AR(1) model could also overlap with the spurious effect.

4- Spurious correlation among hydrological parameters (not shown in manuscript, but shown for GR4J model, in **reply RC2#11**). In our knowledge, none of previous publications has reported this issue.

Our research shows for a limited number of examples how these spurious effects on FB basin appears, when TLs are neglected in a strict joint inference. We have also shown how these effects disappear when TLs are enforced. The magnitude and type of these spurious problems depends on the error and hydrological models, as our manuscript shows. Of course, the basin is also a factor but we do not want to enlarge unnecessarily the paper with a large number of examples. Our paper makes a negative empirical demonstration and actually one fail is enough!

There is not a biunivocal relationship of the kind: "violation of TLs ⇔ inference shows spurious problems". In "Conclusions" of the revised manuscript we will highlight this possible misunderstanding. For example, WLS inference (<u>2 free error parameters</u>) does not exhibit any problem when TLs are neglected. Similarly, Schoups and Vrugt (2010) in their first case study inferred the SEP distribution but without considering skewness parameter (<u>4 free error parameters</u>) and they found no problems. Probably, the problems also do not appear for FB basin, in joint inferences which consider neither skewness nor kurtosis parameters, as in Evin et al. (2014) (<u>3 free error parameters</u>). But we found that including the skewness parameter in the joint inference of the first case study of Schoups and Vrugt (2010) (<u>5 free error parameters</u>), the MCMC inference does not converge. By only applying TLs (this case is not shown in original manuscript) or only applying recommendation of Evin et al. (2013) (this latter case is our **GL++NTL**) inference easily converges, but with the spurious enlargement of the uncertainty band. It is necessary the joint application of both measures to allow the convergence and to avoid spurious effects; this is made in our **GL++** error model.

Therefore, it seems that one pattern emerges: the greater or lesser complexity of the inferred error structure (in our case represented by a bivariate pdf) seems to predetermine the occurrence (or not) of spurious problems during the strict joint inference, when parameters are not properly constrained. FB basin is an "easy" basin, therefore its error structure is not complex: until 4 error parameters are admitted without constrictions and without showing spurious problems. Basins which are more difficult to model yield a more complex error structure: in these cases, the bivariate distribution modeling would admit less than 4 <u>free</u> error model parameters.

TLs could be understood as one of those possible parameter restrictions to avoid the spurious problems during the strict joint inference. However, in our opinion, TLs enforcement is more than an ad-hoc method to restrict the error model parameters. Meeting TLs is, theoretically, a statistical requirement which eventually produces the convenient error parameter restriction. TLs are fulfilled in classical inferences (with SLS method) and we understand that these laws are perfectly transposable to any inference which involves to the inferred variables and its errors.

References

Todini, E. and Mantovan, P.: Comment on:'On undermining the science?'by Keith Beven, Hydrol. Process., 1638(January), 1633–1638, doi:10.1002/hyp.6670, 2007.