Supplement of Hydrol. Earth Syst. Sci., 24, 2017–2041, 2020
https://doi.org/10.5194/hess-24-2017-2020-supplement

Hydrology and
Earth System
Sciences
Open Access

Supplement of

A crash-testing framework for predictive uncertainty assessment when forecasting high flows in an extrapolation context

Lionel Berthet et al.

Correspondence to: Lionel Berthet (lionel.berthet@developpement-durable.gouv.fr)

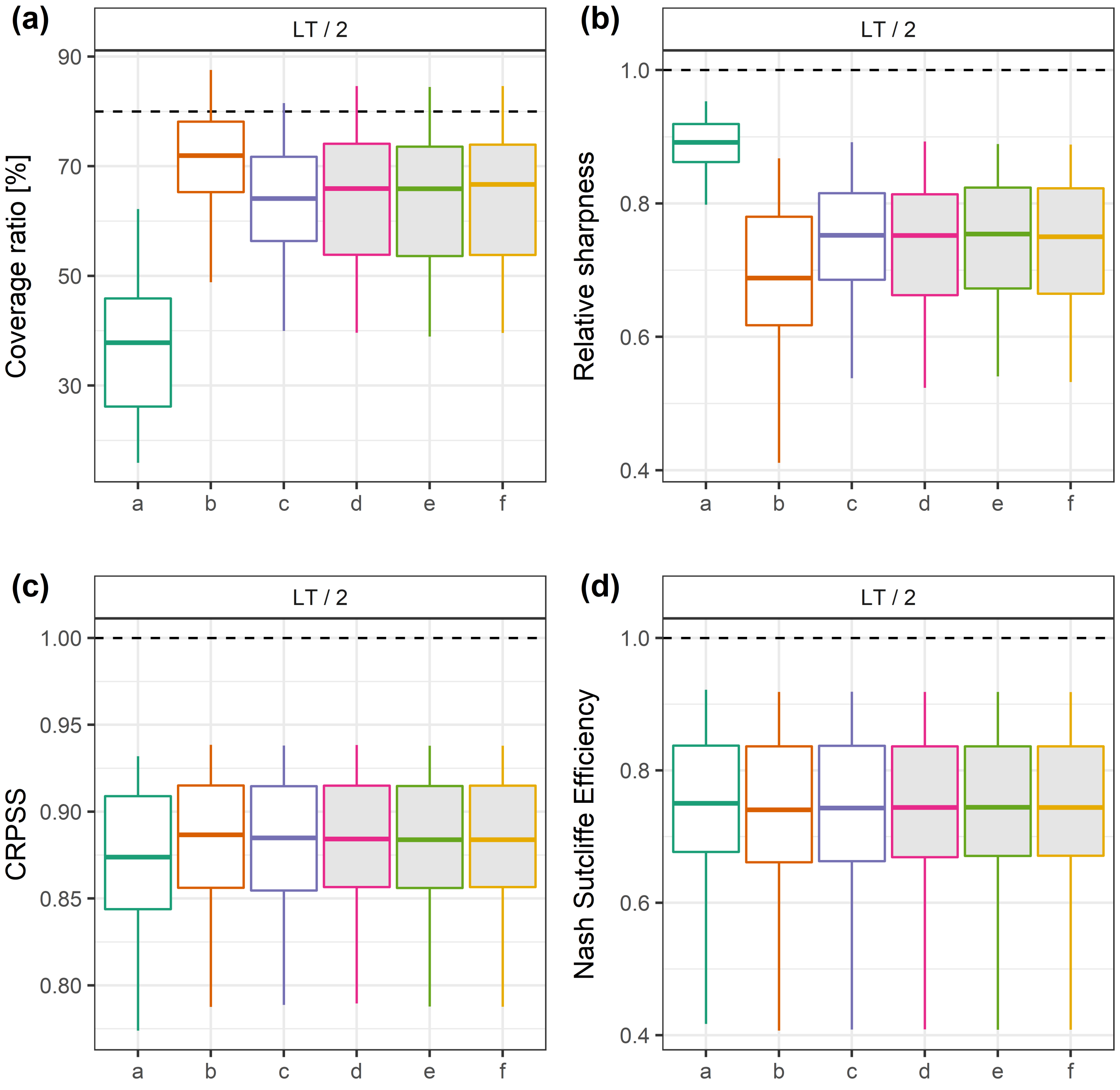

Figure S1. Distributions of coverage rate, relative sharpness, CRPSS and NSE values over the catchment set on control data set D3, obtained with the different transformations tested (the filled boxplots are related to calibrated transformations), for lead time LT / 2 (same as Fig. 13).

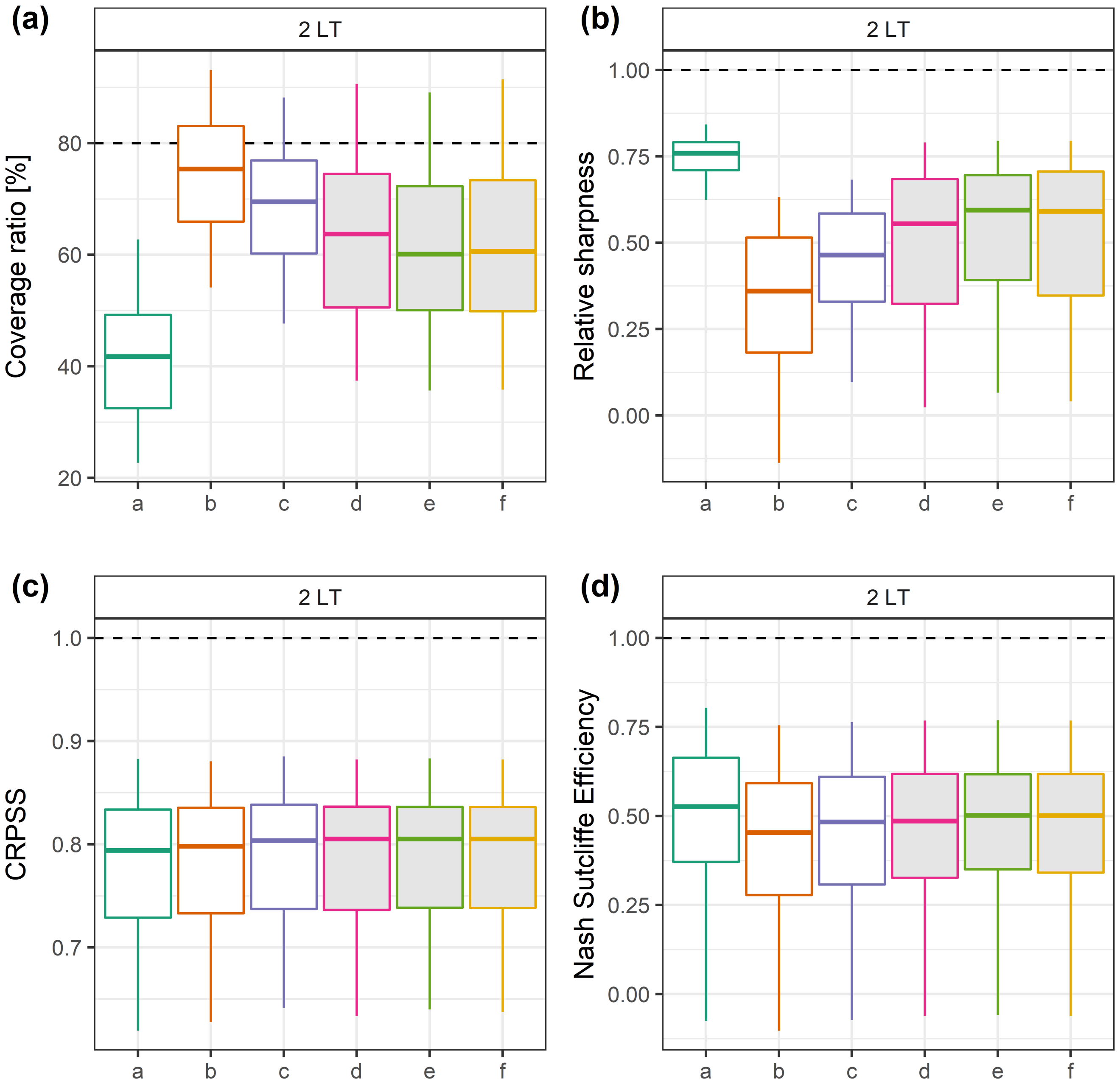

Figure S2. Distributions of coverage rate, relative sharpness, CRPSS and NSE values over the catchment set on control data set D3, obtained with the different transformations tested (the filled boxplots are related to calibrated transformations), for lead time 2 LT (same as Fig. 13).

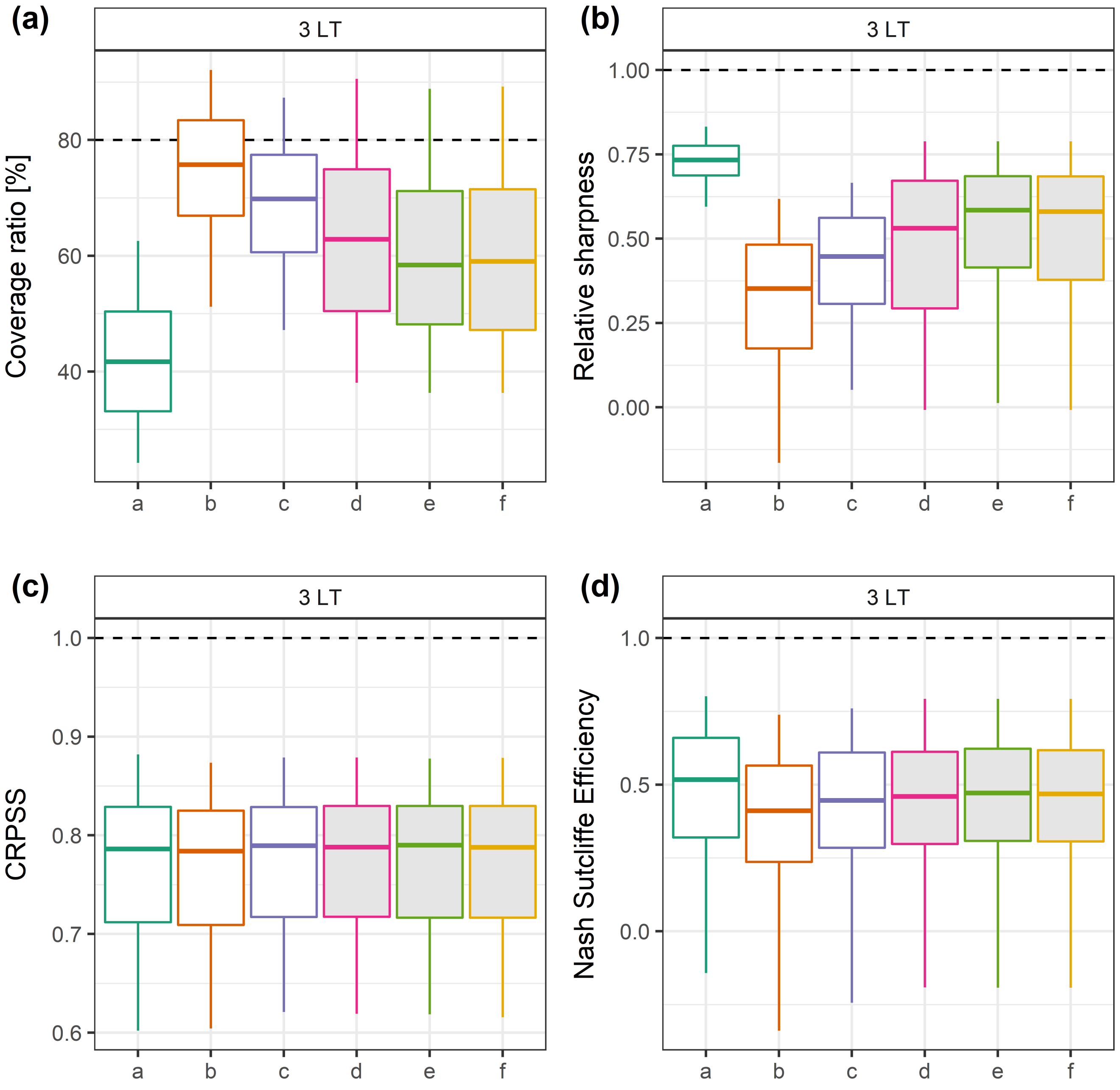

Figure S3. Distributions of coverage rate, relative sharpness, CRPSS and NSE values over the catchment set on control data set D3, obtained with the different transformations tested (the filled boxplots are related to calibrated transformations), for lead time 3 LT (same as Fig. 13).

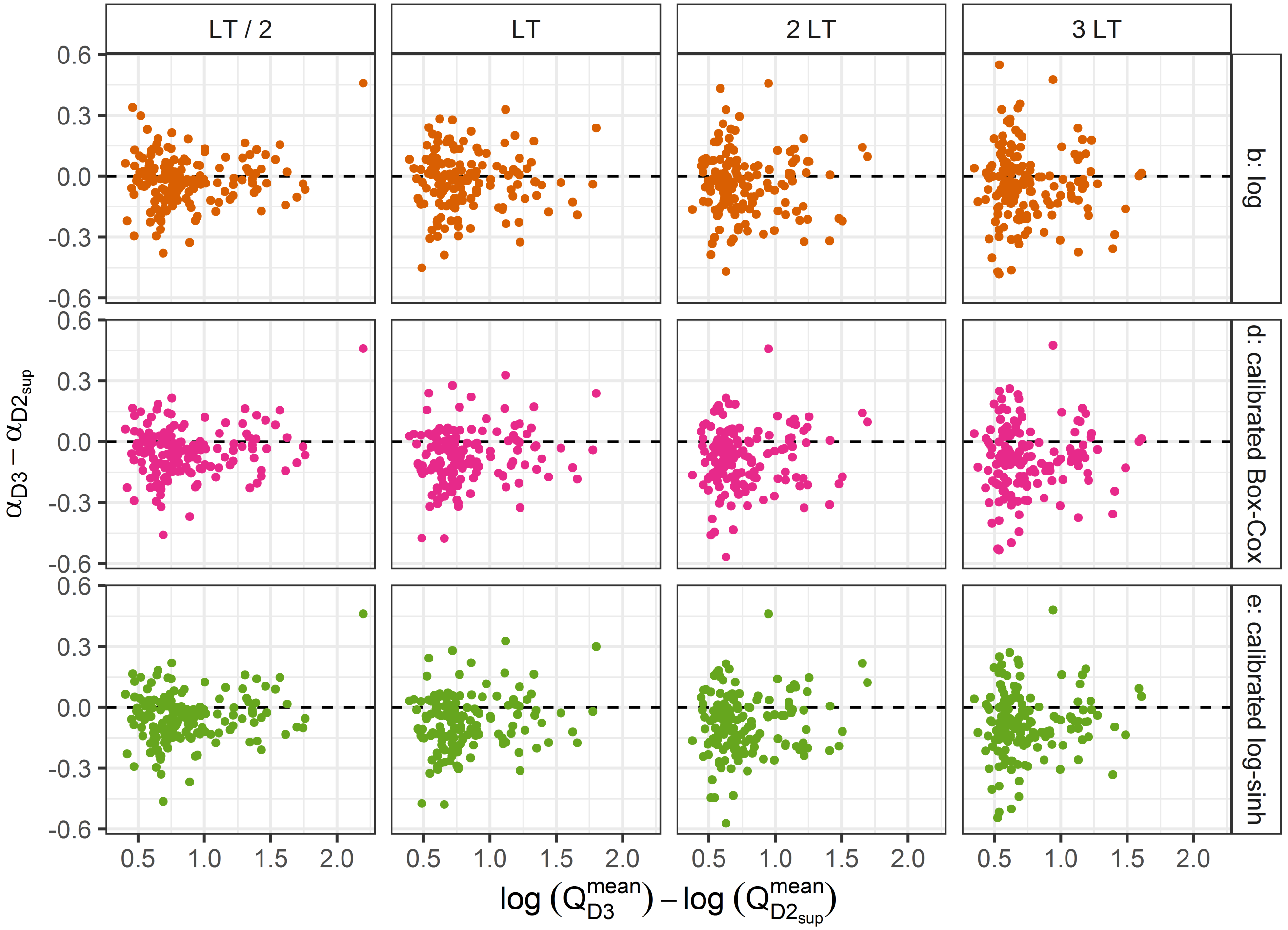

Figure S4. Investigating the performance loss in an extrapolation context (Sect. 3.3): reliability loss as a function of the rarity of the D3 events (ratio of the mean forecasted discharge D3 to mean forecasted discharge $D2_{sup}$ over the catchment set. Similar results were obtained for the three other transformations.

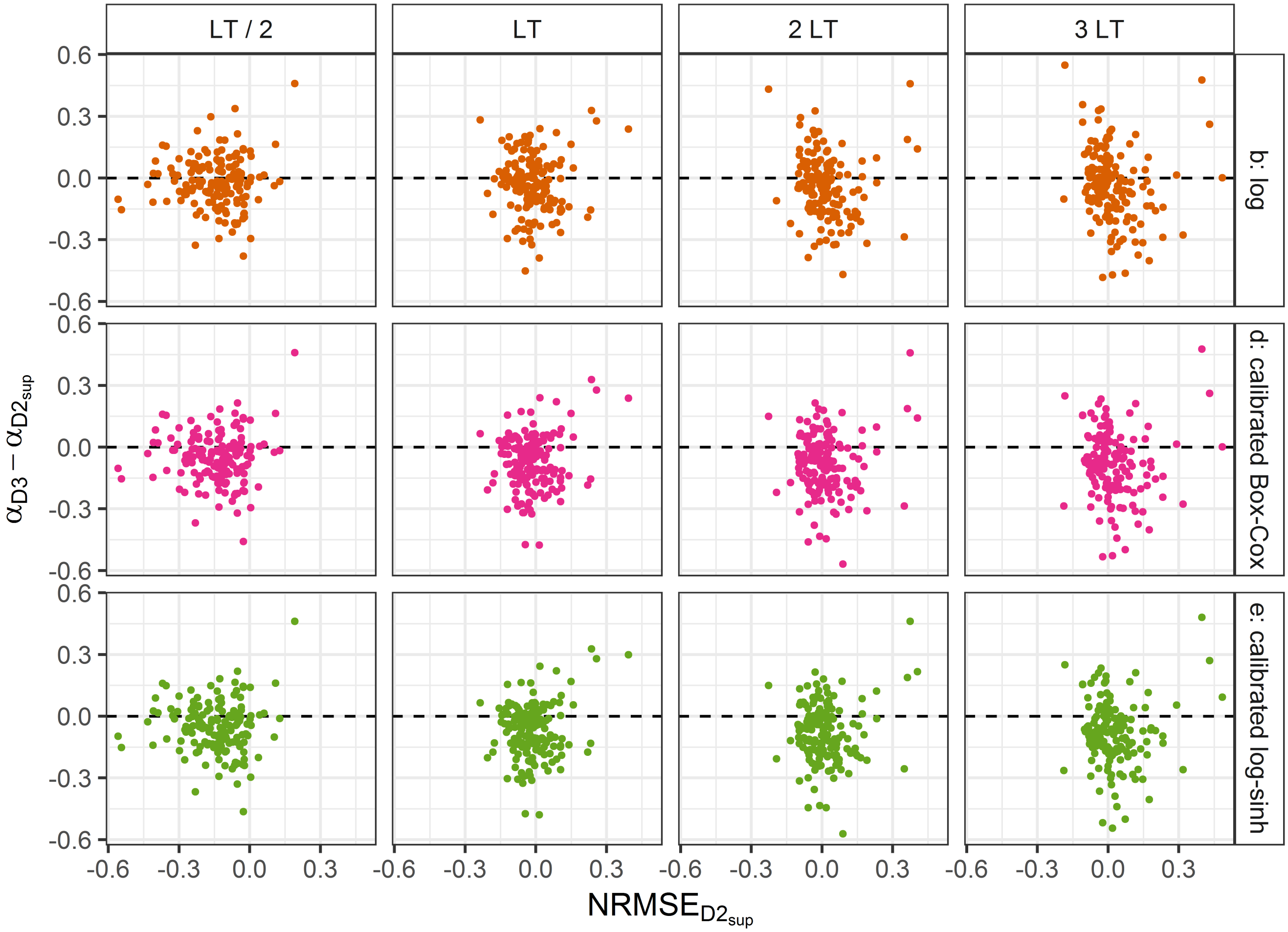

Figure S5. Investigating the performance loss in an extrapolation context (Sect. 3.3): reliability loss as a function of the relative accuracy of the deterministic forecasts on $D2_{sup}$. A normalised RMSE was used to facilitate the visual representation, as in Lobligeois et al. (2014): the normalization allows to compare RMSE values from different catchments. Again, no clear trend is seen, which means that the goodness-of-fit during the calibration phase cannot be used as an indicator of the robustness of the uncertainty estimation in an extrapolation context. Similar results were obtained for the three other transformations.

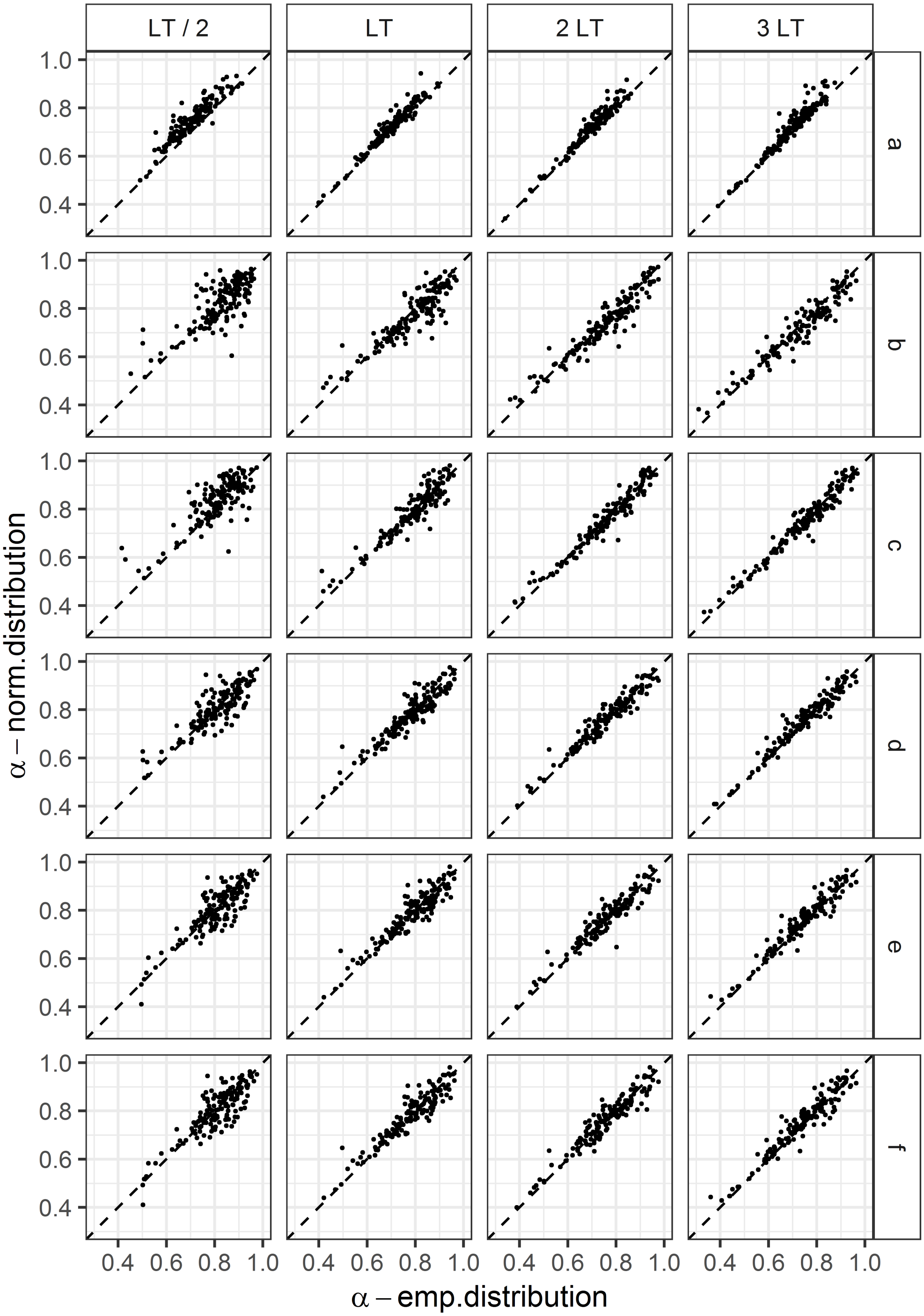

Figure S6. Empirical-based versus distribution-based approaches (Sect. 4.2). Scatter plots of the reliability α-index over the catchment set: x-axis, using the empirical residuals distribution observed on the training data set; y-axis: using the Gaussian distribution fitted on the previous one. The values obtained for all catchments are well distributed on both sides of the bisector: there is no systematic behaviour; for some transformations, it is slightly better to choose the theoretical Gaussian quantiles, while empirical quantiles provide slightly more reliable predictive uncertainty assessment for others. The variability of the distance to the bisector is much lower than the α-index variability obtained among catchments: the choice of the distribution is not the dominant factor to explain the performance.

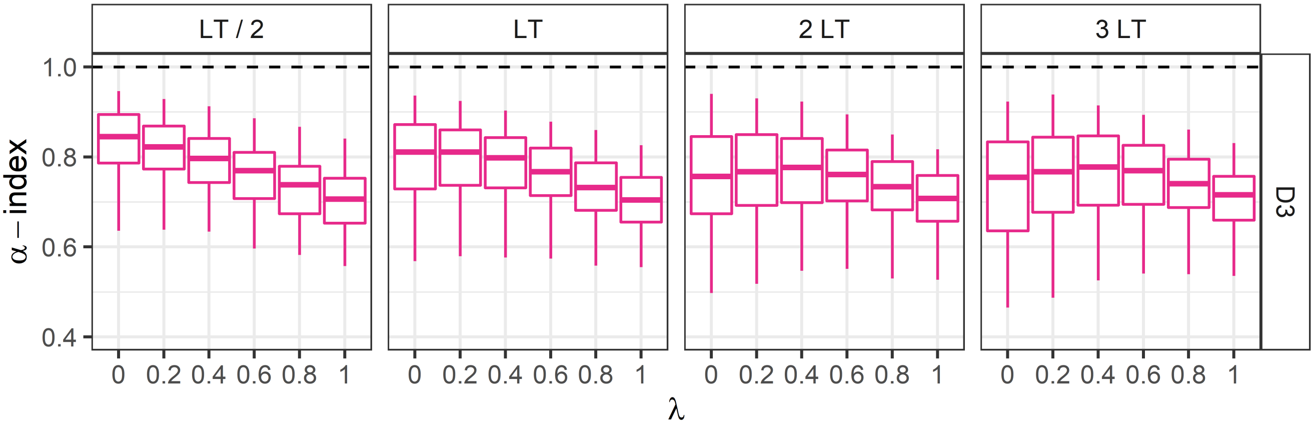

Figure S7. Distributions of the α-index values obtained with different parameter λ values of the Box-Cox transformation on the control data set D3. Best performances are obtained for λ between 0.1 and 0.3.