



Supplement of

Impact of bias adjustment strategy on ensemble projections of hydrological extremes

Paul C. Astagneau et al.

Correspondence to: Paul C. Astagneau (paul.astagneau@slf.ch)

The copyright of individual parts of the supplement might differ from the article licence.

S1 Dataset

Table S1: List of the 87 catchments used in the study.

River name	Station
Birse	Soyhières, Bois du Treuil
Albula	Tiefencastel
Thur	Jonschwil, Mühlau
Kander	Hondrich
Kleine Emme	Emmen
Emme	Emmenmatt
Glatt	Rheinsfelden
Broye	Payerne, Caserne d'aviation
Areuse	Boudry
Wigger	Zofingen
Sense	Thörishaus, Sensematt
Simme	Oberwil
Töss	Neftenbach
Kleine Emme	Werthenstein, Chappelboden
Plessur	Chur
Lorze	Frauenthal
Ergolz	Liestal
Sitter	St. Gallen, Bruggen / Au
Dünnern	Olten, Hammermühle
Venoge	Ecublens, Les Bois
Murg	Frauenfeld
Allaine	Boncourt, Frontière
Reuss	Andermatt
Ilfis	Langnau
Birse	Moutier, La Charrue

River name	Station
Verzasca	Lavertezzo, Campiöi
Landwasser	Davos, Frauenkirch
Murg	Murgenthal, Walliswil
Werdenberger Binnenkanal	Salez
Rheintaler Binnenkanal	St. Margrethen
Inn	St. Moritzbad
Grande Eau	Aigle
Rom	Müstair
Suze	Sonceboz
Emme	Eggiwil, Heidbüel
Calancasca	Buseno
Promenthouse	Gland, Route Suisse
Gürbe	Belp, Mülimatt
Liechtensteiner Binnenkanal	Ruggell
Seyon	Valangin
Schächen	Bürglen, Galgenwäldli
Seez	Mels
Aubonne	Allaman, Le Coulet
Mentue	Yvonand, La Mauguettaz
Luthern	Nebikon
Areuse	St-Sulpice
Lorze	Zug, Letzi
Necker	Mogelsberg, Aachsäge
Murg	Wängi
Saltina	Brig
Cassarate	Pregassona
Suhre	Oberkirch
Sitter	Appenzell
Chamuerabach	La Punt-Chamues-ch
Aabach	Hitzkirch, Richensee
Scheulte	Vicques
Worble	Ittigen
Veveyse	Vevey, Copet
Langeten	Huttwil, Häberenbad

River name	Station
Minster	Euthal, Rüti
Ova dal Fuorn	Zernez, Punt la Drossa
Goldach	Goldach, Bleiche
Aach	Salmsach, Hungerbühl
Breggia	Chiasso, Ponte di Polenta
Alp	Einsiedeln
Orbe	Le Chenit, Frontière
Riale di Pincascia	Lavertezzo
Grosstalbach	Isenthal
Sionge	Vuippens, Château
Dischmabach	Davos, Kriegsmatte
Goneri	Oberwald
Magliasina	Magliaso, Ponte
Biber	Biberbrugg
Allenbach	Adelboden
Ova da Cluozza	Zernez
Rein da Sumvitg	Somvitg, Encardens
Chli Schliere	Alpnach, Chilch-Erli
Krummbach	Klusmatten
Glatt	Herisau, Zellersmühle
Poschiavino	La Rösa
Sellenbodenbach	Neuenkirch
Grossbach	Einsiedeln, Gross
Riale di Roggiasca	Roveredo, Bacino di compenso
Parimbot	Ecublens, Eschiens
Rietholzbach	Mosnang, Rietholz
Sissle	Eiken
Reppisch	Dietikon

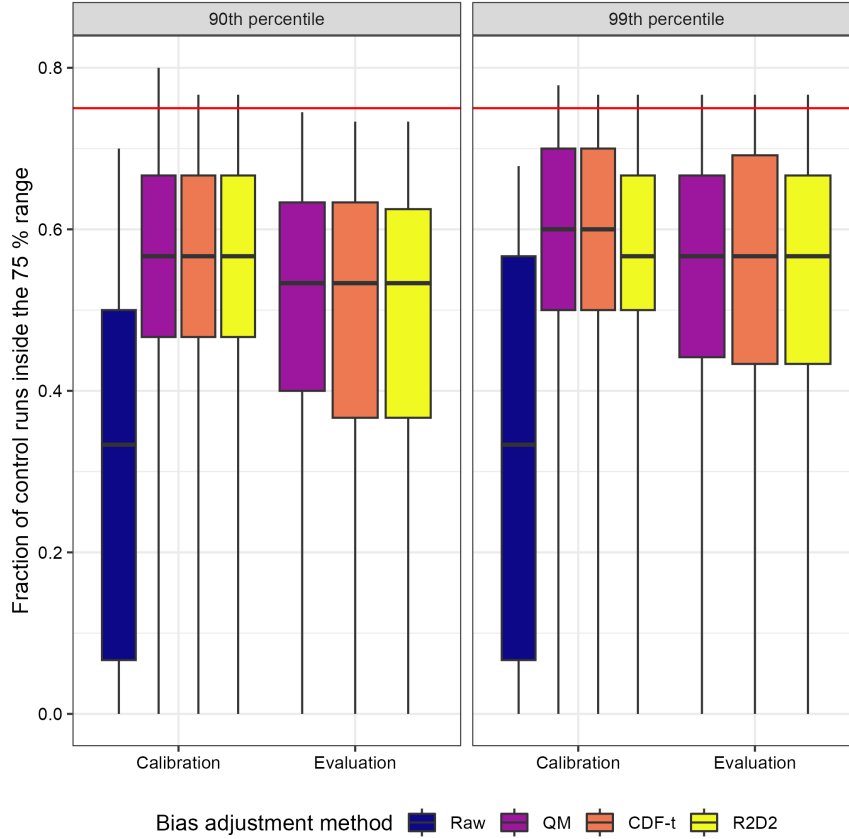


Figure S1. Ability of three bias adjustment methods and the unadjusted ensemble (raw) in reproducing snow water equivalent (control run) statistics for the 87 catchments. The fraction of control runs within the 75 % range was calculated for two percentiles (90th and 99th). The optimum value of the performance criterion is 0.75. QM is the univariate non-change-preserving method. CDF-t is the univariate change-preserving method. R2D2 is the multivariate change-preserving method. All methods were run using the ensemble adjustment approach. Calibration and evaluation combine both climatic sub-periods.

S2 Snow water equivalent simulations

S3 Spatial differences in bias adjustment high-flow performance

We investigate whether there are spatial differences in high-flow performance between the different bias adjustment methods (Fig. S2). In general, all bias adjustment methods improve the high-flow simulations compared to the raw ensemble for all seasons and most catchments. Specifically, we find a similar spatial pattern for the three bias adjustment methods: the highest performance is achieved for low-elevation catchments (western and northern Switzerland) and the lowest performance for high-elevation catchments (southern, south-eastern and eastern Switzerland). There is an opposite pattern in spring (MAM), i.e. the highest performance is achieved for high-elevation catchments and the lowest performance for low-elevation catchments. On

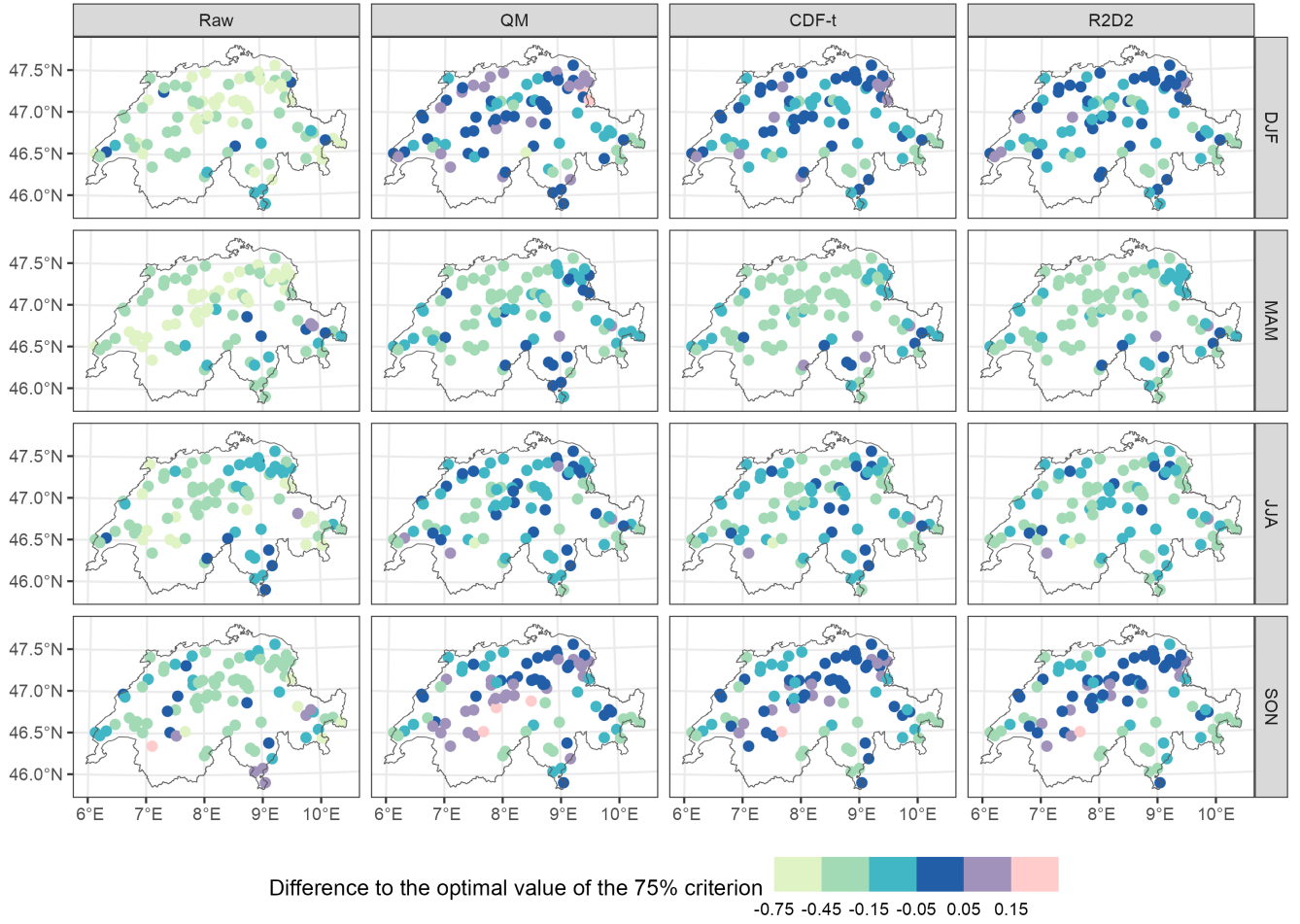


Figure S2. Maps of high-flow performance for the three bias adjustment methods and the unadjusted ensemble (raw) in reproducing high-flow from the control runs for the 87 catchments. The fraction of control runs inside the simulated 75 % confidence interval was calculated for four seasons (December/January/February, March/April/May, June/July/August, September/October/November) and for the 99th percentile. The optimum value of the performance criterion is 0.75 (dark blue colour). QM is the univariate non-change-preserving method, CDF-t the univariate change-preserving method, and R2D2 the bivariate change-preserving method. All methods were run using the ensemble adjustment approach. The results are shown for one evaluation sub-period (1991–2020).

10 the one hand, we find only marginal spatial differences in high-flow performance between univariate (CDF-t) and bivariate (R2D2) adjustments for all seasons, thus indicating no particular a priori influence of inter-variable properties on high-flow performance. On the other hand, the univariate non-change-preserving bias adjustment method (QM) performs better than the other two methods for most catchments. These differences are consistent across all seasons.

S4 Streamflow simulations with another hydrological model

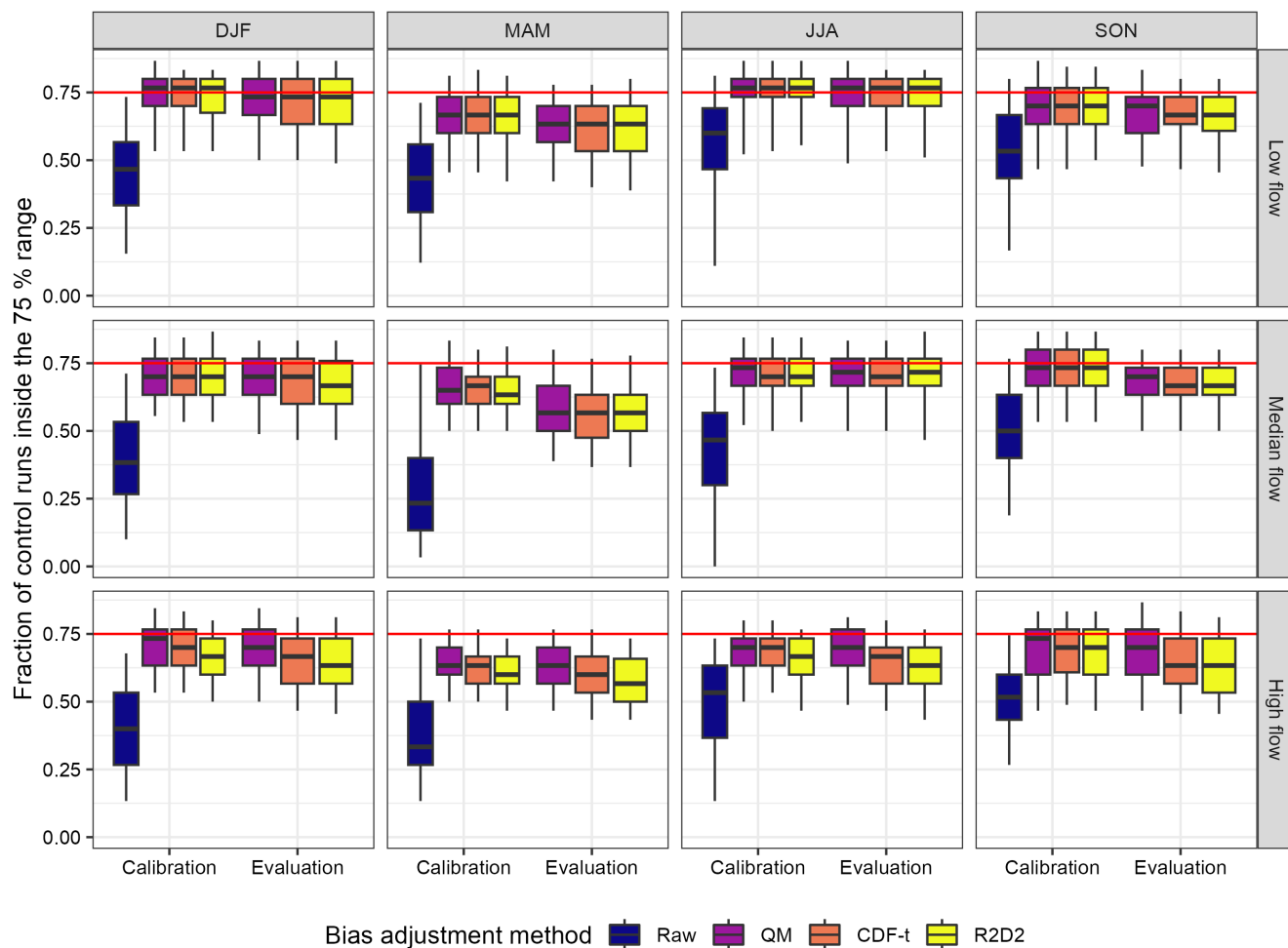


Figure S3. Ability of three bias adjustment methods and the unadjusted ensemble (raw) in reproducing streamflow statistics of the control runs for the 87 catchments. Streamflow was simulated with the Cemaneige-GR5J model (Le Moine, 2008; Valéry et al., 2014; Coron et al., 2020). The fraction of control runs within the 75 % range was calculated for four seasons (December/January/February, March/April/May, June/July/August, September/October/November) and three streamflow percentiles (1st, 50th and 99th). The optimum value of the performance criterion is 0.75. QM is the univariate non-change-preserving method. CDF-t is the univariate change-preserving method. R2D2 is the multivariate change-preserving method. All methods were run using the ensemble adjustment approach. Calibration and evaluation combine both climatic sub-periods.

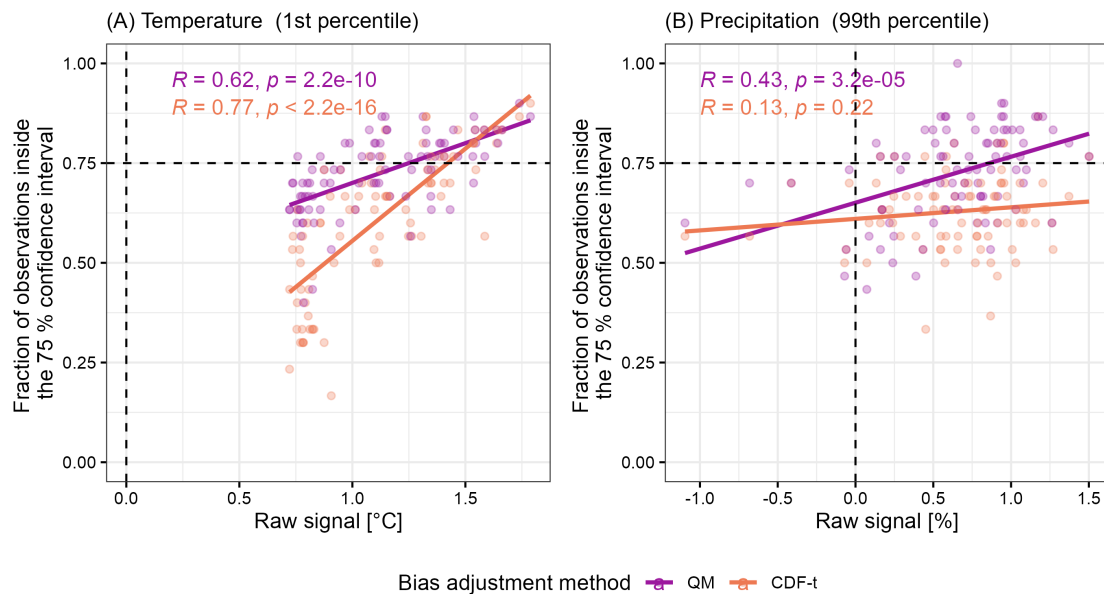


Figure S4. Relationship between the raw signal between sub-periods and (A) temperature (1st percentile) performance; (B) precipitation (99th percentile) performance, for 87 catchments. Performance is assessed with the fraction of observations falling inside the simulated 75 % confidence interval. The signal is the difference (absolute for temperature and relative for precipitation) between the percentile value of the sub-period P2 and the percentile value of the sub-period P1. The results are shown with a linear regression. QM is the non-change-preserving bias adjustment method and CDF-t the change-preserving bias adjustment method. The results are shown for the ensemble adjustment option and the second evaluation period only (see Fig. S5, S6, S7 and S8 for the monthly relationships and the two evaluation periods).

While the analysis of ensemble variability clearly shows that individual-member adjustments alter the variability of the ensemble, it does not explain why ensemble adjustments combined with the change-preserving method lead to a drop in performance for high precipitation and low temperatures (Fig. 4). To investigate this question, we now examine the relationship between the performance of the adjustments and the raw signal between the two sub-periods used for the calibration/evaluation experiment. Our hypothesis is that the change-preserving method will theoretically preserve the signal of the raw ensemble (compared to the non-change-preserving method) and will therefore be more efficient if there is a strong signal.

We find that the performance of the change-preserving method correlates with the raw signal for low temperatures (Fig. S4A) but not for high precipitation (Fig. S4B). The lower the signal of the raw ensemble for low temperatures, the lower the performance of the change-preserving adjustments. The performance of the non-change-preserving method also correlates with the raw signal but we observe the largest differences in performance between the two bias adjustment methods for the lowest temperature signals (around +0.75 °C). For the largest temperature signals (around +1.5 °C), the two bias adjustment methods are comparable in terms of performance.

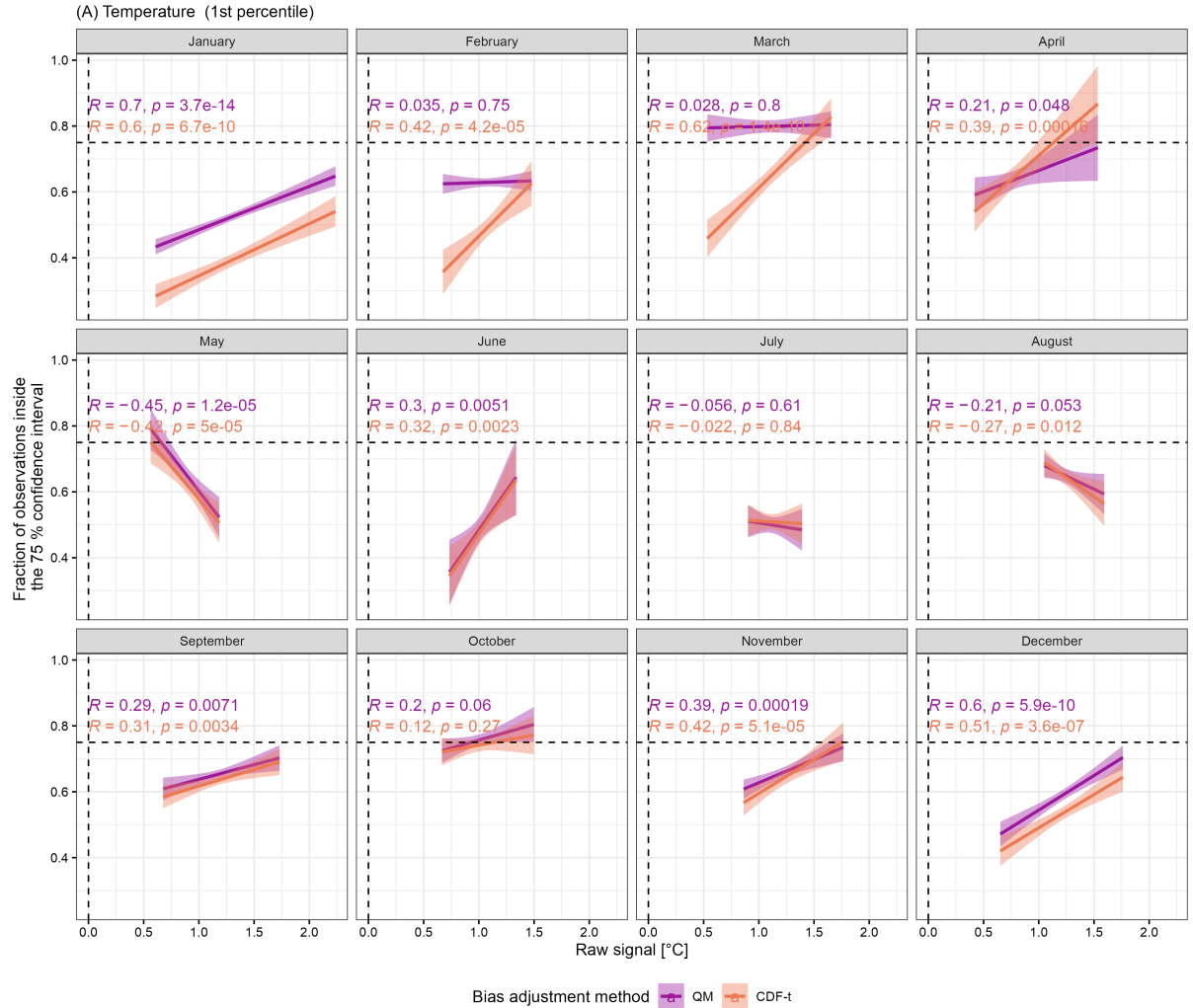


Figure S5. Relationship between temperature (1st percentile) performance (evaluation sub-period P1) and the raw signal between sub-periods, for 87 catchments.. Performance is assessed with the fraction of observations falling inside the simulated 75 % confidence interval. The signal is the difference (absolute) between the percentile value of the sub-period P2 and the percentile value of the sub-period P1. The results are shown with a linear regression (line) with the 95% confidence interval (bandwidth). QM is the non change-preserving bias adjustment method and CDF-t is the change-preserving bias adjustment method. The results are shown for the ensemble adjustment option.

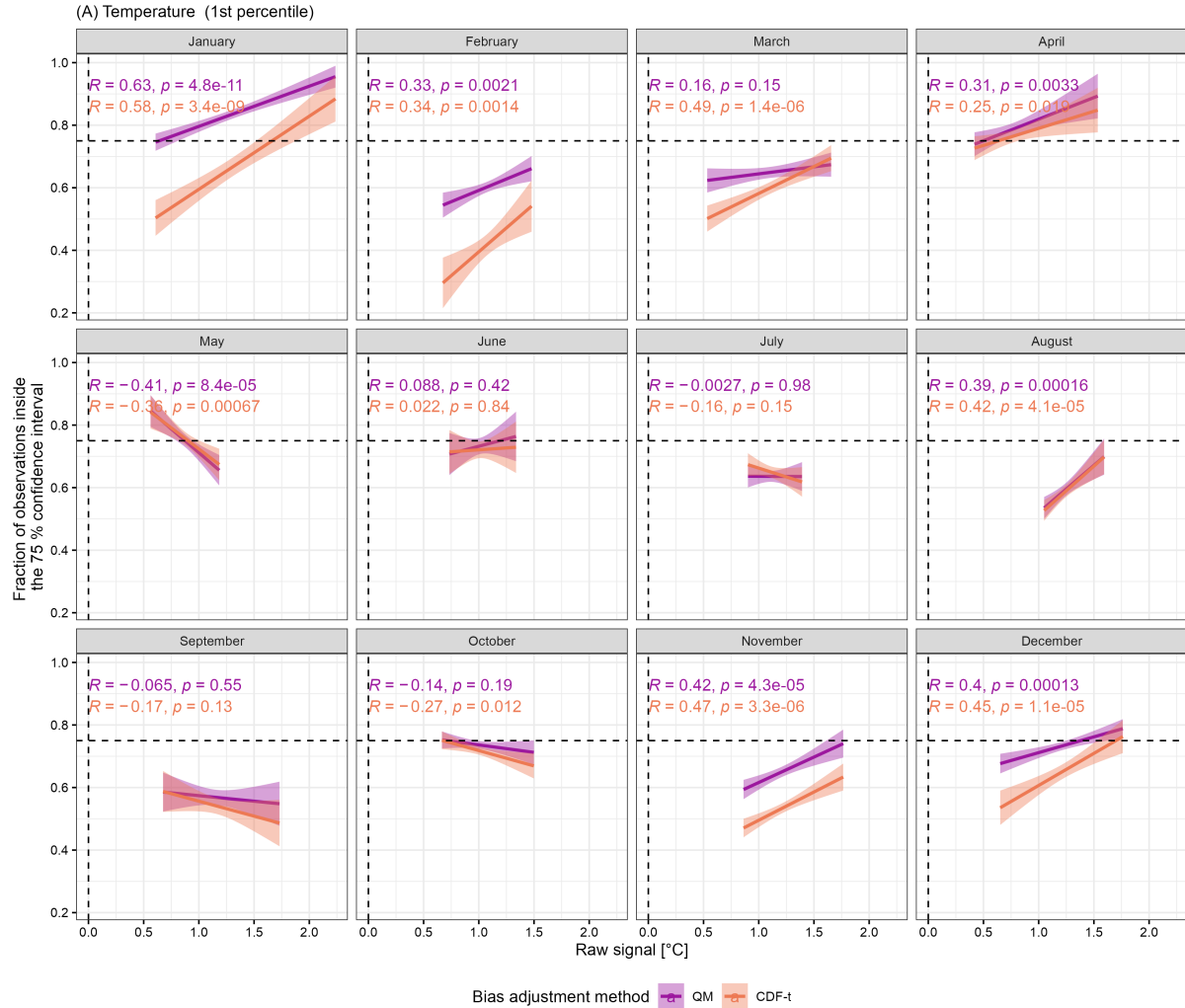


Figure S6. Relationship between temperature (1st percentile) performance (evaluation sub-period P2) and the raw signal between sub-periods, for 87 catchments.. Performance is assessed with the fraction of observations falling inside the simulated 75 % confidence interval. The signal is the difference (absolute) between the percentile value of the sub-period P2 and the percentile value of the sub-period P1. The results are shown with a linear regression (line) with the 95% confidence interval (bandwidth). QM is the non change-preserving bias adjustment method and CDF-t is the change-preserving bias adjustment method. The results are shown for the ensemble adjustment option.

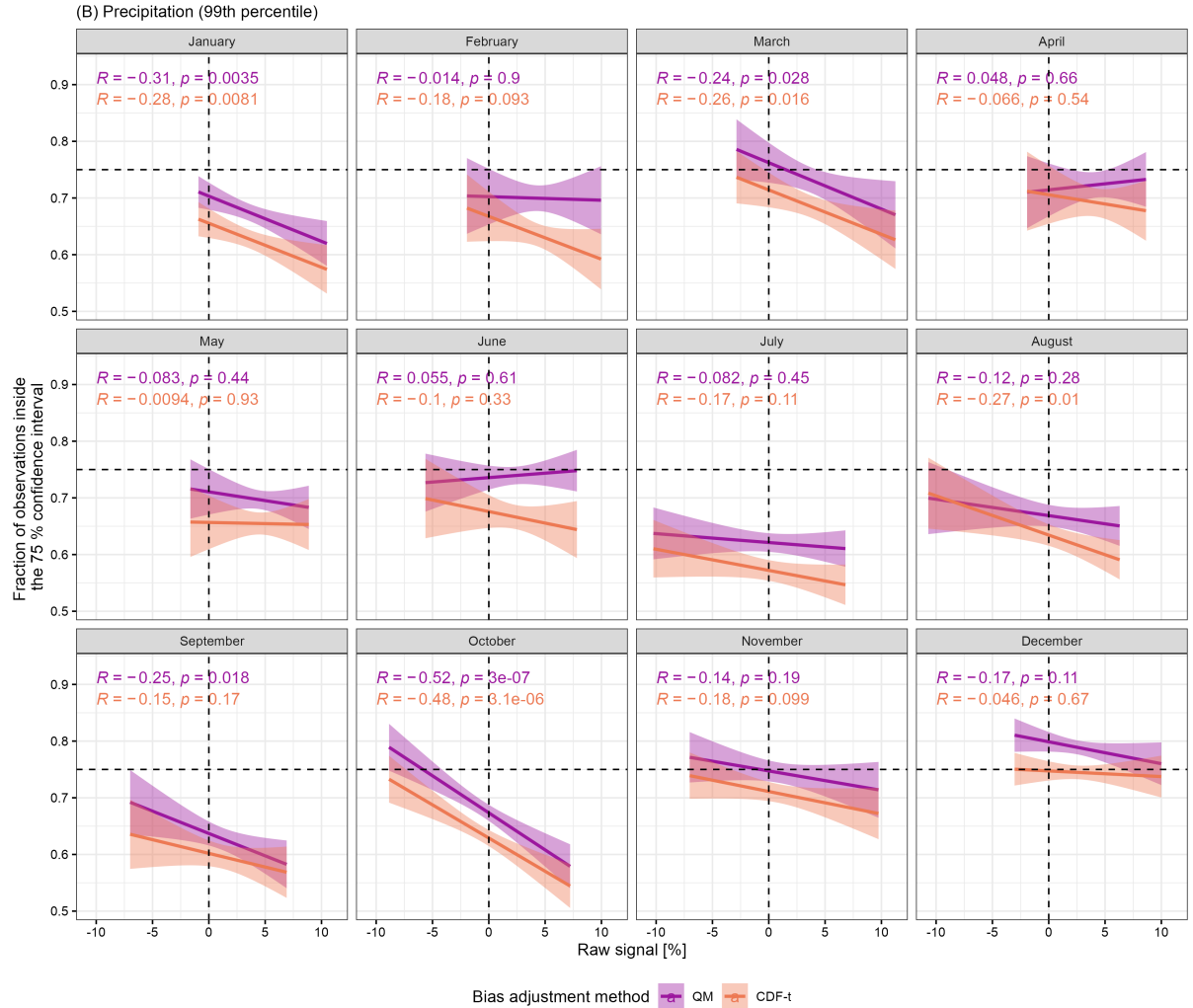


Figure S7. Relationship between precipitation (99th percentile) performance (evaluation sub-period P1) and the raw signal between sub-periods, for 87 catchments.. Performance is assessed with the fraction of observations falling inside the simulated 75 % confidence interval. The signal is the difference (relative) between the percentile value of the sub-period P2 and the percentile value of the sub-period P1. The results are shown with a linear regression (line) with the 95% confidence interval (bandwidth). QM is the non change-preserving bias adjustment method and CDF-t is the change-preserving bias adjustment method. The results are shown for the ensemble adjustment option.

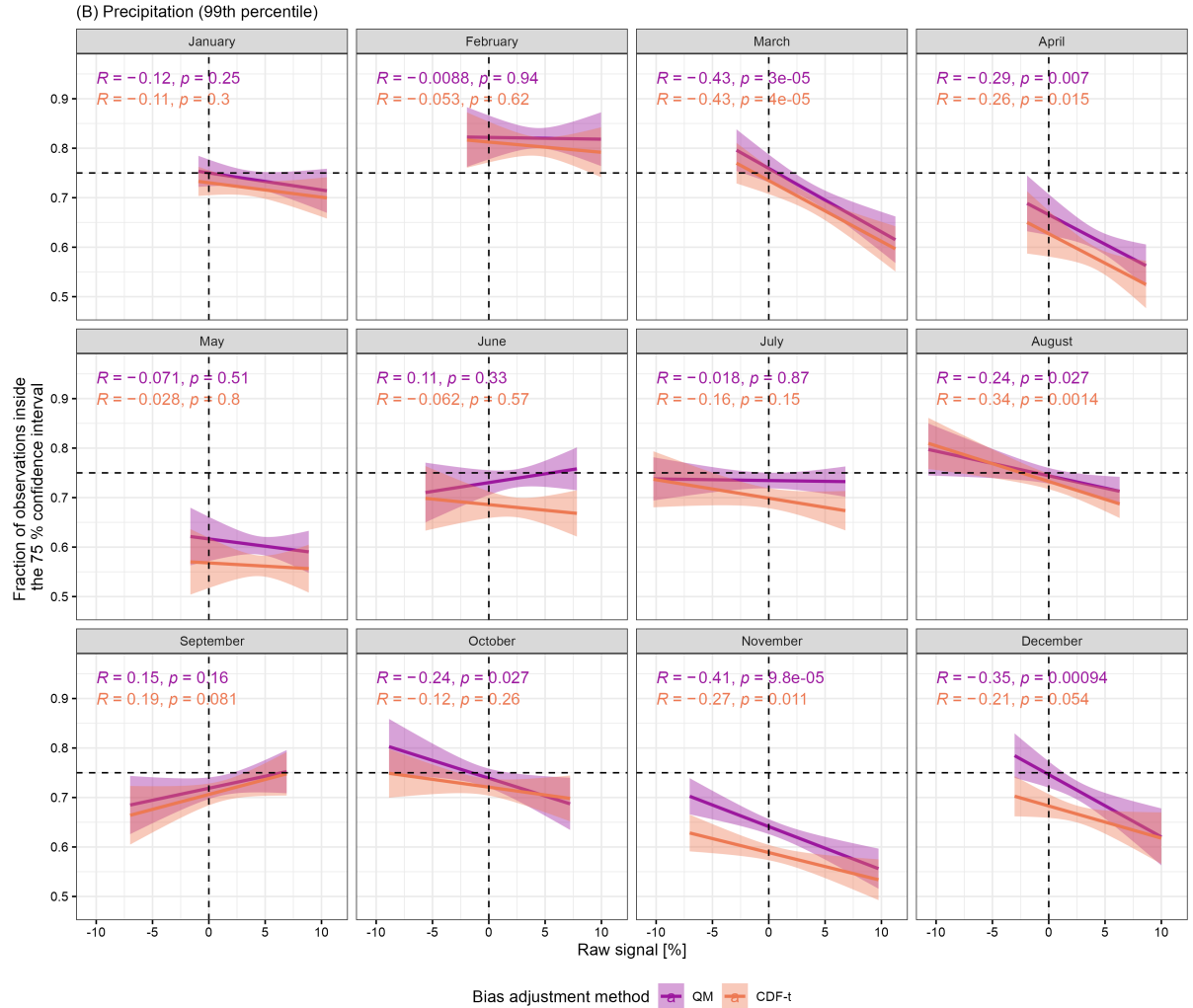


Figure S8. Relationship between precipitation (99th percentile) performance (evaluation sub-period P2) and the raw signal between sub-periods, for 87 catchments. Performance is assessed with the fraction of observations falling inside the simulated 75 % confidence interval. The signal is the difference (relative) between the percentile value of the sub-period P2 and the percentile value of the sub-period P1. The results are shown with a linear regression (line) with the 95% confidence interval (bandwidth). QM is the non change-preserving bias adjustment method and CDF-t is the change-preserving bias adjustment method. The results are shown for the ensemble adjustment option.

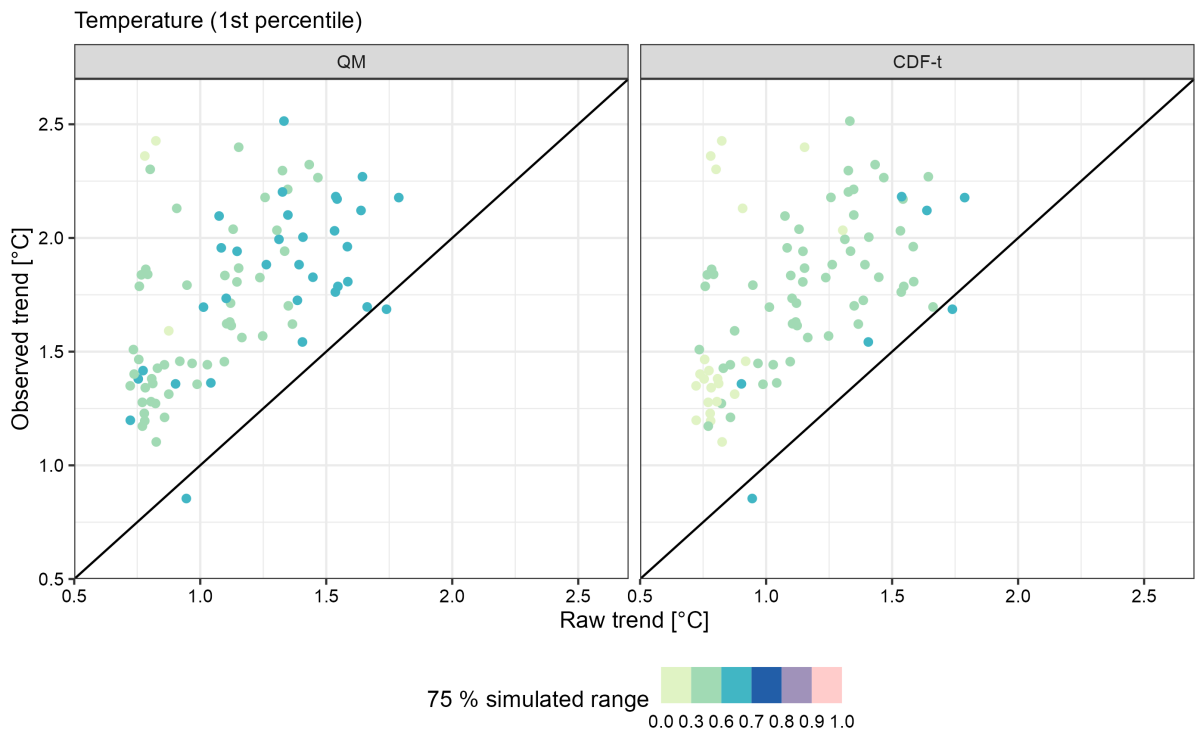


Figure S9. Comparison between observed and raw temperature signals (1st percentile) with regards to performance (evaluation sub-period P1) for 87 catchments. Performance is assessed with the fraction of observations falling inside the simulated 75 % confidence interval. The signal is the difference (absolute) between the percentile value of the sub-period P2 and the percentile value of the sub-period P1. QM is the non change-preserving bias adjustment method and CDF-t is the change-preserving bias adjustment method. The results are shown for the ensemble adjustment option.

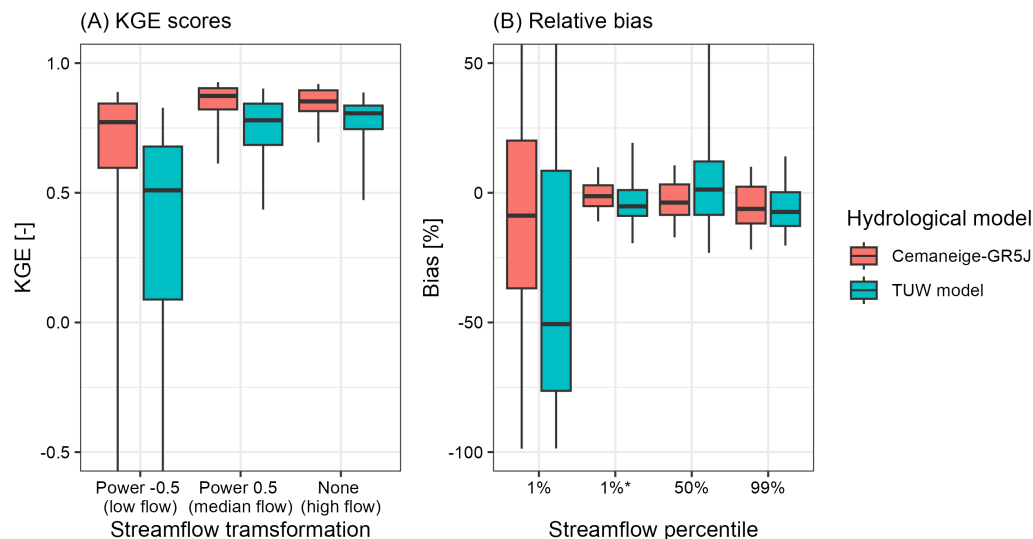


Figure S10. Hydrological model streamflow performance for the 87 catchments (extrapolation: 2011-2019). 1%* refers to the relative bias calculated on the 1st streamflow percentile normalized by the streamflow mean.

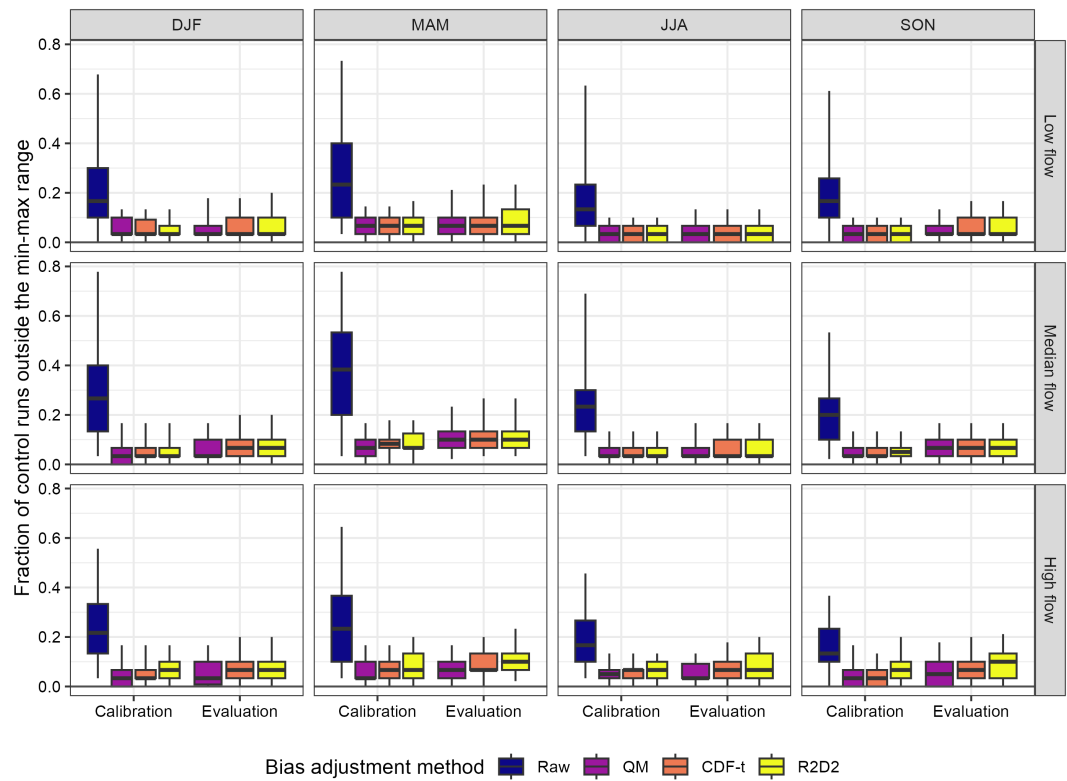


Figure S11. Ability of the three bias adjustment methods and the unadjusted ensemble (raw) to reproduce streamflow statistics of the control runs (streamflow time series simulated by the hydrological model with observed precipitation and temperature inputs) for the 87 catchments. The fraction of control runs outside the simulated min-max confidence interval was calculated for four seasons (December/January/February, March/April/May, June/July/August, September/October/November) and three streamflow percentiles (1st, 50th and 99th). The optimum value of the performance criterion is 0.

References

Coron, L., Delaigue, O., Thirel, G., Dorchies, D., Perrin, C., and Michel, C.: airGR: Suite of GR Hydrological Models for Precipitation-Runoff Modelling. R package version 1.7.6., <https://doi.org/10.15454/EX11NA>, 2020.

35 Le Moine, N.: Le bassin versant de surface vu par le souterrain : une voie d'amélioration des performances et du réalisme des modèles pluie-débit ?, Ph.D. thesis, UPMC, Cemagref, <https://hal.inrae.fr/tel-02591478>, 2008.

Valéry, A., Andréassian, V., and Perrin, C.: 'As simple as possible but not simpler': What is useful in a temperature-based snow-accounting routine? Part 2 – Sensitivity analysis of the Cemaneige snow accounting routine on 380 catchments, *Journal of Hydrology*, 517, 1176–1187, <https://doi.org/10.1016/j.jhydrol.2014.04.058>, 2014.