Responses to comments from anonymous Referee 2

On “Should altitudinal gradients of temperature and precipitation inputs be inferred from key parameters in snow-hydrological models?” by D. Ruelland (HESS-2019-556)

Referee’s comment

The article analyzes the sensitivity of a snow accounting procedure and hydrological modeling results to the evaluation of temperature and precipitation in space and time in mountainous catchments. The study is based on a set of 20 catchments in the French Alps and two hydrological models. The author evaluates the interplay between the lapse rate, snow routine and hydrological model parameters.

I found this is a clear and interesting paper. I have a few suggestions for improvement detailed below, some of which are quite major and requiring new calculations. I suggest considering the paper for possible publication in HESS after major revision.

Authors’ response

I would like to thank the referee for the time spent in reviewing the initial paper and making interesting suggestions. Most of them were judged useful even though I did not agree with all comments. In any cases, I provided a point-by-point response to the reviewer’s comments and tried to bring modifications to the manuscript accordingly.

Detailed comments

Referee’s comment

1. I found that the literature review could have been more exhaustive, to better stress the originality of the work compared to existing studies on similar or close topics. Some recent works could be discussed, for example the work by Le Moine et al. (2015) on the link between snow and hydrological sub-models in model parameterization, some studies on using snow data to calibrate hydrological models (Besic et al. 2014, Henn et al. 2016, Riboust et al. 2019), some studies with physical approaches to estimate lapse rates (Rahman et al. 2014, Zhang et al. 2015, Naseer et al. 2019). The review could also be extended on how gauge undercatch factors are estimated. The author should further discuss to which extent the proposed approach is original compared to these past findings.

Authors’ response and modifications to manuscript

I thank the referee for these additional references, some of which I did not know. Note however that some recent references have not been yet published at the moment where the current paper was conceived and written (e.g. Naseer et al., 2019; Riboust et al., 2019), which makes it difficult to provide an up-to-date literature.

Most of the proposed references were judged useful. Consequently, they were cited in the text and added to the reference list.

Following the referee comment, the following paragraph was added in the introduction section:

“…Several studies proposed approaches to estimate lapse rates based on physically-based or conceptual models on specific catchments. Zhang et al. (2013) showed that the runoff simulation results involving snowmelt and rainfall runoff were highly sensitive to the temperature and precipitation lapse rates in a Tibetan catchment. Rahman et al. (2014) calibrated the SWAT model in a snow-dominated basin in the Swiss Alps and found also that temperature lapse rate was significantly important for hydrological performance. Naseer et al. (2019) considered a dynamic lapse rate based on a vertical profile of temperature in a catchment in Japan and succeeded to improve the precipitation phase in a distributed hydrological modelling framework. Henn et al. (2016) investigated the value of snow data to constrain the inference of precipitation from streamflow, using lumped hydrologic models and an elevation-band snow model in a Californian basin. Their results suggested that multiple types of hydrologic observations, such as streamflow and SWE, may help to constrain the water balance of high-elevation basins. Le Moine et al. (2015) proposed a calibration strategy where the parameters of both an interpolation model and a daily snow-hydrological model are jointly inferred in a multi-variable approach applied in a catchment in the French Alps. Using a hydro-meteorological modelling chain involving 31 calibrated parameters, they showed the potential of using different types of observations (rain gauges, snow water equivalent measurements and streamflow data) to help assess temperature and precipitation lapse rates according to different weather types. These examples encourage testing whether an inverse modelling approach based on calibrated constant lapse rates can perform well with parsimonious conceptual models applied in numerous basins."

The recent reference from Riboust et al. (2019) was mentioned later in the introduction:
"…Moreover, other authors (Franz and Karsten, 2013; He et al., 2014; Riboust et al., 2019) showed that adding snow data information to the calibration procedure enabled the identification of more robust snow parameter sets by making the snow models less dependent on the rainfall-runoff model with which they are coupled.
It was also acknowledged in section 4.1 about the snow accounting routine (se answer to the referee comment #7.

Some references were also acknowledged in the conclusion section:
"…Accurate estimate of these parameters greatly helps in determining the form of precipitation and spatial distribution of temperature and precipitation, and are critical for snow cover and runoff modelling in high mountain catchments, as already reported in other regions (Zhang et al., 2013; Naseer et al., 2019)."

Referee's comment

2. Section 2.1: It would be useful to add a figure showing the distributions of mean precipitation and temperature over the set of gauges, to give an idea of the variability across the study domain.

Authors' response and modifications to manuscript

The other reviewer also made this comment. Details about the "estimated" climatology of the study period are now provided in Table 1 (see below), which has been modified to include mean annual temperature (T), total precipitation (P), snowfall fraction (S) and streamflow (Q) for each basin. Note however that T, P and S values are very delicate to provide since they necessarily rely on approximations depending on the method used to distribute temperature and precipitation (in link with the paper issue). This is why they were not included in the initial submitted paper. As indicated in the modified caption of Table 1, catchment areal temperature, total precipitation and snowfall fraction were estimated after calibrating local altitudinal gradients over 2000–2016 using the snow-hydrological inverse approach proposed in the current paper (see Test #4 in Table 5).

Table 1 Streamflow gauging stations and main catchment characteristics. Percentages of glacierized area were estimated from the World Glacier Inventory (NSIDC, 2012). Mean annual precipitation (P), snowfall fraction (S) and temperature (T) were estimated after calibrating local altitudinal gradients over 2000–2016 using the snow-hydrological inverse approach proposed in the current paper (see Test #4 in Table 5).

Station	River	Area (km²)	Glacierized area (%)	Elevations (m.a.s.l.) Min	Max	Mean annual precip. (P) (mm/yr)	Snowfall fraction (S) (%)	Mean annual temp. (T) (°C)	Mean annual streamflow (Q) (mm/yr)
Barcelonnette	Ubaye	549	0	1132	3308	802	48	1.9	521
Lauzet-Ubaye	Ubaye	946	0	790	3308	947	44	3.0	654
Beynes	Asse	375	0	605	2273	920	16	8.7	344
Saint-André-Les-Alpes	Issole	137	0	931	2392	965	24	6.8	481
Villar-Lourbière	Séveraisse	133	4	1023	3623	1561	47	2.3	1317

Val-des-Prés	Durance	207	0	1360	3059	836	54	0.9	688
Briançon	Durance	548	1	1187	3572	844	51	1.7	714
Argentière-la-Bessée	Durance	984	3	950	4017	1014	52	2.1	765
Embrun	Durance	2170	2	787	4017	990	48	2.9	693
Espinasses	Durance	3580	1	652	4017	964	45	3.4	654
Villeneuve-d'Entraunes	Var	132	0	926	2862	989	37	4.8	650
Val-d'Isère	Isère	46	9	1831	3538	1245	63	-1.5	1119
Bessans	Avérole	45	12	1950	3670	1399	66	-2.4	1311
Taninges	Giffre	325	0	615	3044	2031	36	4.7	1771
Vacheresse	Dranse d'Abondance	175	0	720	2405	1669	29	4.9	1088
La Baume	Dranse de Morzine	170	0	690	2434	1636	32	4.7	1285
Dingy-Saint-Clair	Fier	223	0	514	2545	1649	26	6.5	1243
Allèves	Chéran	249	0	575	2157	1486	23	6.9	819
Mizoën	Romanche	220	9	1057	3846	1205	56	0.8	978
Allemond	L'Eau Dolle	172	2	713	3430	1460	46	2.7	1164

Referee's comment

3. Section 2.2: Reference could be given to the work by Leleu et al. (2014).

Authors' response and modifications to manuscript

I could not access this reference from the journal "*La Houille Blanche*", although I contacted the authors to obtain a hard copy. As a result, I could not judge if it was appropriate to reference this work here. However, Section 2.2 deals mainly with the streamflow series gathered from the French hydrological database (www.hydro.eaufrance.fr) for 20 catchments, as indicated in the text. In other publications, the French hydrological database is usually acknowledged this way by indicating the web site from which data (and metadata) can be freely accessed.

Referee's comment

4. Table 1: Please explain the meaning of abbreviations in the last column. Is this information useful here?

Authors' response and modifications to manuscript

The information was indeed not very useful. It has been removed from the revised Table 1 (see Table above).

Referee's comment

5. Section 3.3: The author calculates the efficiency criteria on precipitation values. However, the criteria may be strongly influenced by a few large rainfall events, which may not be representative of the average characteristics of precipitations. It may be useful to consider computing the efficiency criteria on transformed precipitation (e.g. root square transformation) to avoid putting too much weight on outlier values. Would this change something in results?

Authors' response and modifications to manuscript

The Jack-Knife cross-validation procedure on precipitation series is usually performed on RMSE (e.g. Kyriadis et al., 2001; Le Moine et al., 2013; Yang et al., 2018 to cite just a few). I could not find in the literature an example where RMSE was applied based on a root square transformation of precipitation in such a procedure. Of course, it can be argued than computing the objective function (here RMSE) on transformed precipitation may lead to different interpolation parameters. However, it can also be assumed that large rainfall events are critical for elevation/precipitation regressions when looking at the optimized surrounding gauges to consider in the KED and IED methods. This means that using an efficiency criterion on transformed precipitation may also put less weight on large rainfall events to compute the regressions.

Following the referee comment, the JK cross-validation was re-run with IED using $RMSE_{sqrt}$ on daily, monthly and yearly precipitation series. No significant differences could be found in the optimized interpolation parameters as shown in the following Table 1. Only the number of optimized number of surrounding neighbours changed slightly at the daily time scale from 17 to 15. This cannot be not judged significant as this parameter presents a low sensitivity between 12 and 20 neighbours due to the intrinsic compromise looking on the whole study domain.

Table 1 Cross-validation of the IED method with RMSE or $RMSE_{sqrt}$ (root square transformation of precipitation) against yearly, monthly and daily series from precipitation gauges over the period 2000–2016. The values of $n(u)$ and ω represent the interpolation parameters, which were optimised using the leave-one-out procedure.

	Current paper (RMSE as EC)			Alternative test ($RMSE_{sqrt}$ as EC)		
	Efficiency criterion	IED parameters		Efficiency criterion	IED parameters	
	RMSE	$n(u)$	ω	$RMSE_{sqrt}$ (RMSE)	$n(u)$	ω
Yearly	150.31 mm/year	12	3	2.13 (150.31)	12	3
Monthly	22.20 mm/month	12	2	1.05 (22.2)	12	2
Daily	2.90 mm/day	17	2	0.51 (2.91)	15	2

For these different reasons, and also because the $RMSE_{sqrt}$ is difficult to interpret since units are not allowed, the usual RMSE criterion was kept in the article as efficiency criteria on precipitation series. Note also that the two other criterions (MAE and NSE which were not used for optimization) are not presented anymore in Table 4 as they seemed to cause confusion in the result interpretation (see answer to the Referee's comment #10 below).

Referee's comment

6. L261: The name "RMSE" given to the normalized RMSE is a bit confusing. The author may choose another name, e.g. NRMSE.

Authors' response and modifications to manuscript

There was an error in the text, which is now corrected. In fact the models were cross-validated against the usual RMSE (root mean square error) without any normalization, as follows:

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{N} (Vpre_i - Vobs_i)^2 / N} \tag{1}$$

where $Vpre_i$ and $Vobs_i$ are the predicted and observed variables respectively at time scale i and N the total number of time steps.

Referee's comment

7. Section 4.1: Some modifications in this snow module were recently proposed by Riboust et al. (2019), to account for snow-covered area. This should be shortly commented, to better explain how the proposed approach compares to this existing work.

Authors' response and modifications to manuscript

Agreed. This is now commented in the Section 4.1, as follows:

"In the original version of CEMANEIGE, fractional snow-covered area (FSC) is calculated as follows:

$$FSC_i(t) = \min\left(\frac{SWE_i(t)}{SWE_{th}}, 1\right) \tag{11}$$

where *SWE* is the quantity of snow accumulated on the catchment in snow water equivalent (a state variable of the model, in mm), and SWE_{th} is the model's melting threshold. SWE_{th} is calculated as being equal to 90% of mean annual solid precipitation on the catchment considered (Valéry et al., 2014). Alternative approaches have been proposed to account for the hysteresis that exists between *FSC* and *SWE* during the accumulation and melt phases (Riboust et al., 2019). However, introducing such a hysteresis adds two additional free parameters to the SAR. Instead, SWE_{th} was fixed to 40 mm since preliminary sensitivity analyses showed that this value gave very satisfactory *FSC* values when compared to the MODIS observations in the studied catchments."

Referee's comment

8. Fig.3: Maybe add the meaning of the key variables (at least inputs/output) in the figure caption. If UZL is the threshold for the upper output, maybe the arrow should stop at the level of this output.

Authors' response and modifications to manuscript

Agreed. The meaning of the key variables has been added in the Figure caption, as follows: "*R*, *M*, *PE* and *Q* stand for rainfall, melt, potential evapotranspiration and streamflow, respectively". Following the referee comment, the arrow for *UZL* was also slightly modified in the figure.

Referee's comment

9. L376-378: This is a point I did not understand in the proposed methodology. By introducing this criterion WB in the objective function, the author forces the model to close the water balance in the sense of Budyko. This is quite successful when looking at results shown in Fig. 6, since no data lies outside the boundaries of balance closure in the plot. However, I do not understand the physical rationale behind putting this constraint. There are many catchments where the water balance cannot be closed in the Budyko sense for good reasons, mainly because of underground water exchanges. The author artificially constrains the models using WB. I think a more classical bias criterion would be better to consider instead.

Authors' response and modifications to manuscript

The rationale behind using WB in the objective function was to enhance the parameter identifiability without decreasing the model efficiency. I agree that there are many catchments where the water balance cannot be closed in the Budyko sense for good reasons due notably to inter-catchment groundwater exchanges (IGE). There are also several bad reasons for which water balance is not closed at the basin scale: errors in the precipitation volumes, wrong estimate of potential evapotranspiration, inaccurate knowledge of the catchment area, etc. Since the paper deals with the lapse rates of the temperature and precipitation inputs, it can be assumed that using a more classical objective function (i.e. without WB) may also lead to optimize the lapse rates while water balance is not closed for the above "bad" reasons (errors in the precipitation volumes, wrong estimate of potential evapotranspiration). In the paper, two models are used: HBV which considers the catchments as closed systems, and GR4J which allows for potential IGE (and/or wrong water balance estimates) to be considered via its X2 parameter. Consequently, the Budyko constraint left room for not completely close the water balance with GR4J.

Nevertheless, I decided to follow the referee comment because sensitivity analyses to the objective function are far beyond the paper issue and because other readers may not be convinced by the proposed constraint without such an in-deep demonstration. I re-run all the simulations with a more classical objective function (i.e. without using the WB constraint) based only on NSE_{FSC} and NSE_{sqrtQ}. The results were the same as regards to the modelling distribution performances between the various tests (see Figures 1 and 2 below). Changing the objective function thus did not change the main findings of the paper notably as regards to the interest of calibrating the temperature and precipitation lapse rates via a parsimonious 2-parameter SAR. Obviously, the water balance was not closed systematically and it was not interesting anymore to present the Budyko graphs for the different tests. This led also to deteriorate the general parameter identifiability (see Figure 3 below). This particularly affected the identifiability of the *X2* parameter for the reasons explained above.

However, the ranking of parameter identifiability in between tests did not change and the parsimonious 2-parameter SAR still led to the best parameter identifiability, while remaining among the best-performing models. Finally, the optimized lapse rates were slightly changed: the precipitation gradients were notably found more similar between the two hydrological models tested (see Figure 4 below). This is probably the most important reason that convinced me to renounce to the WB constraint in the objective function (even though I still believe that this constraint is both original and efficient).

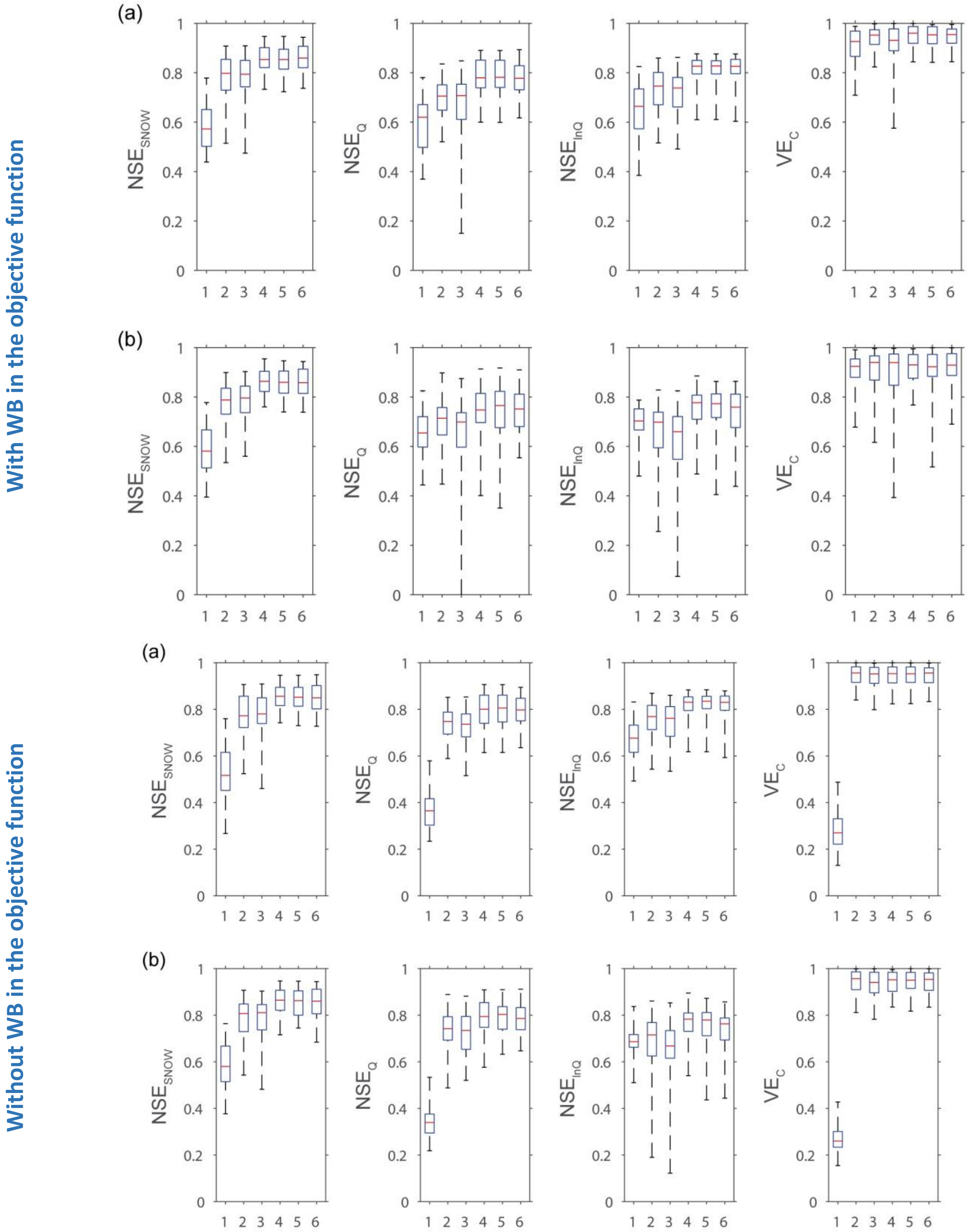
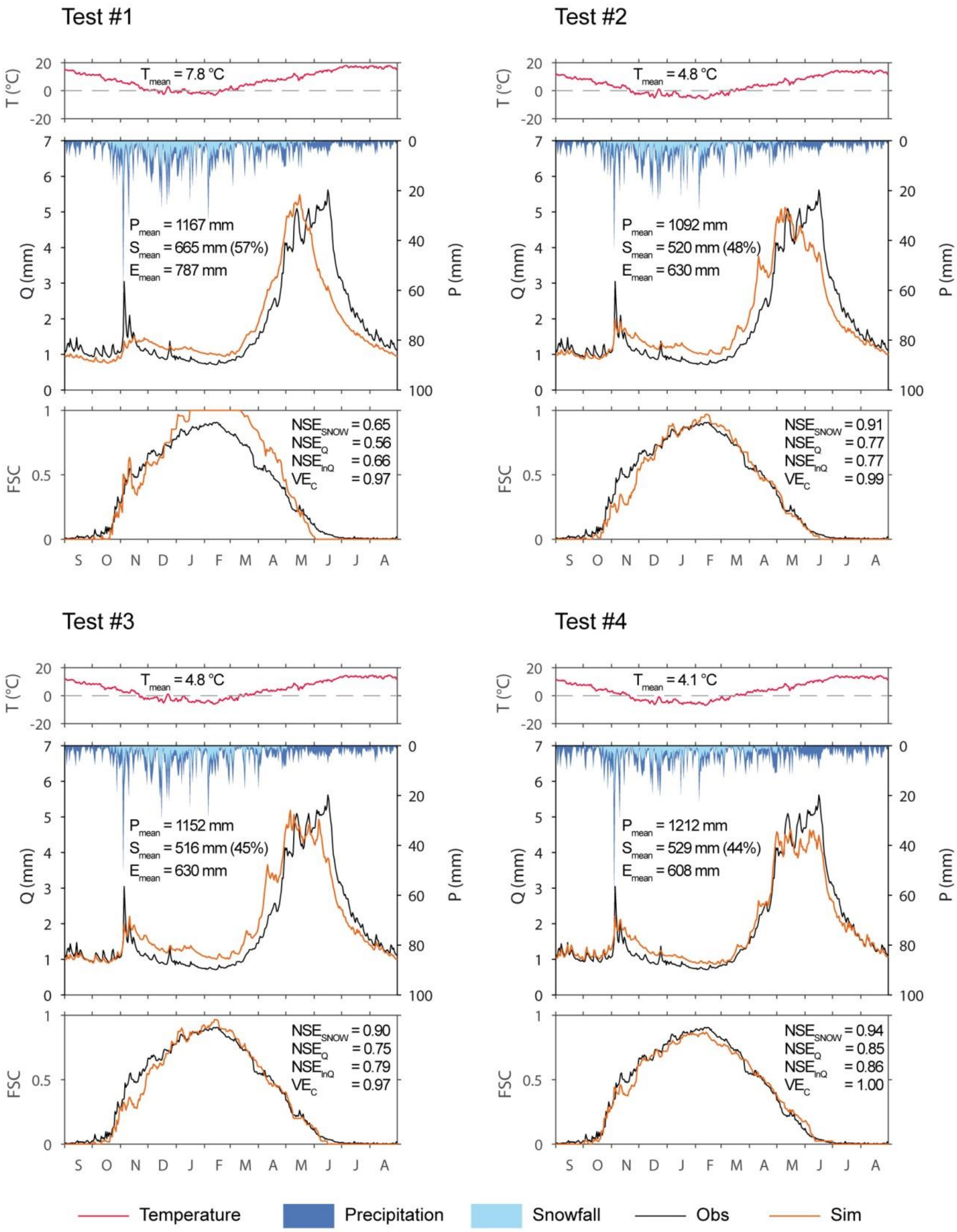


Fig. 1 Boxplots (showing 0.00, 0.25, 0.50, 0.75 and 1.00 percentiles) of the efficiency distributions obtained in validation by the (a) GR4J and (b) HBV9 models combined with the snow model according to six different tests (see Table 5) to account for elevation dependency in the T and P inputs on the 20 snow-affected Alpine catchments.

With WB in the objective function



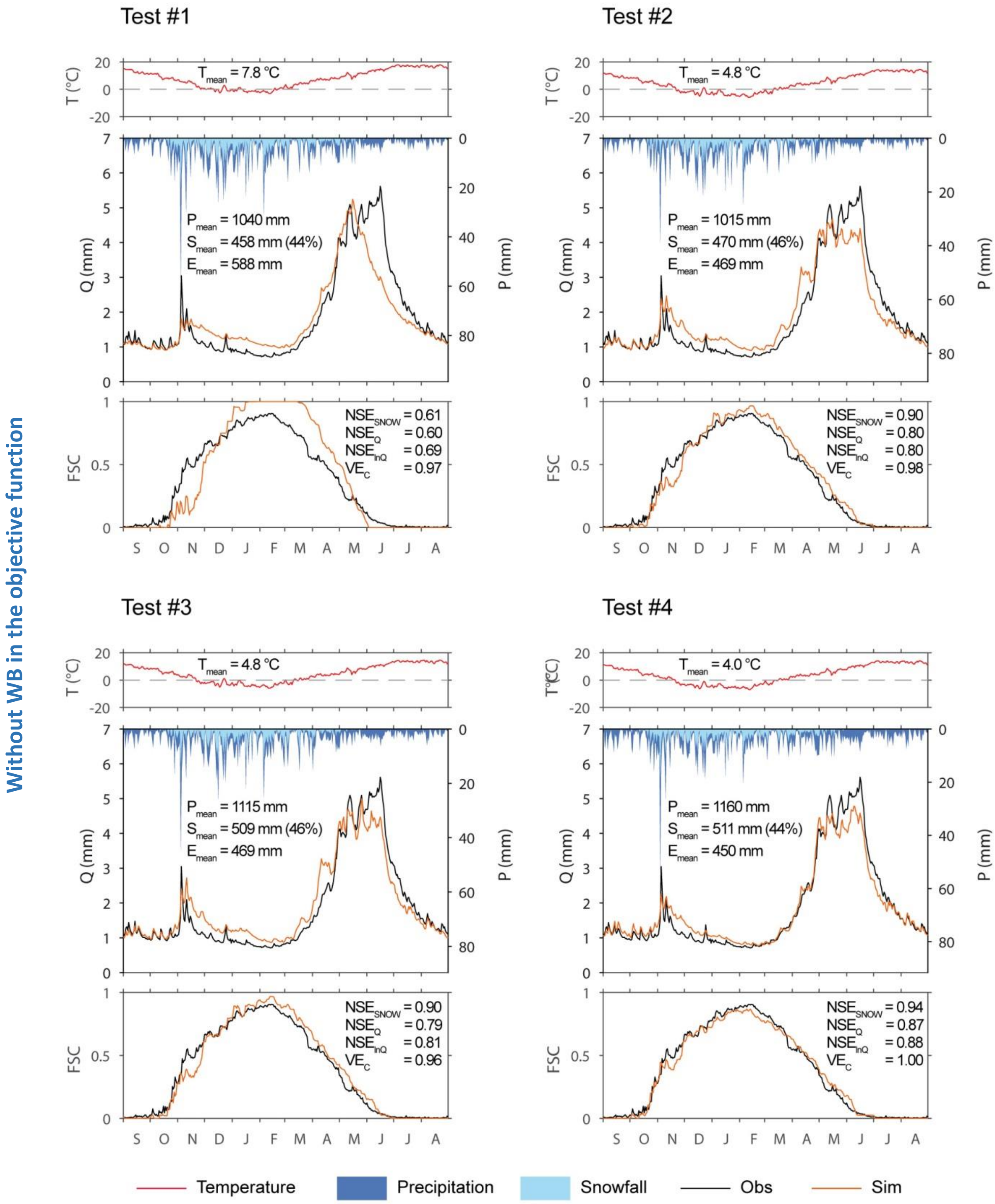
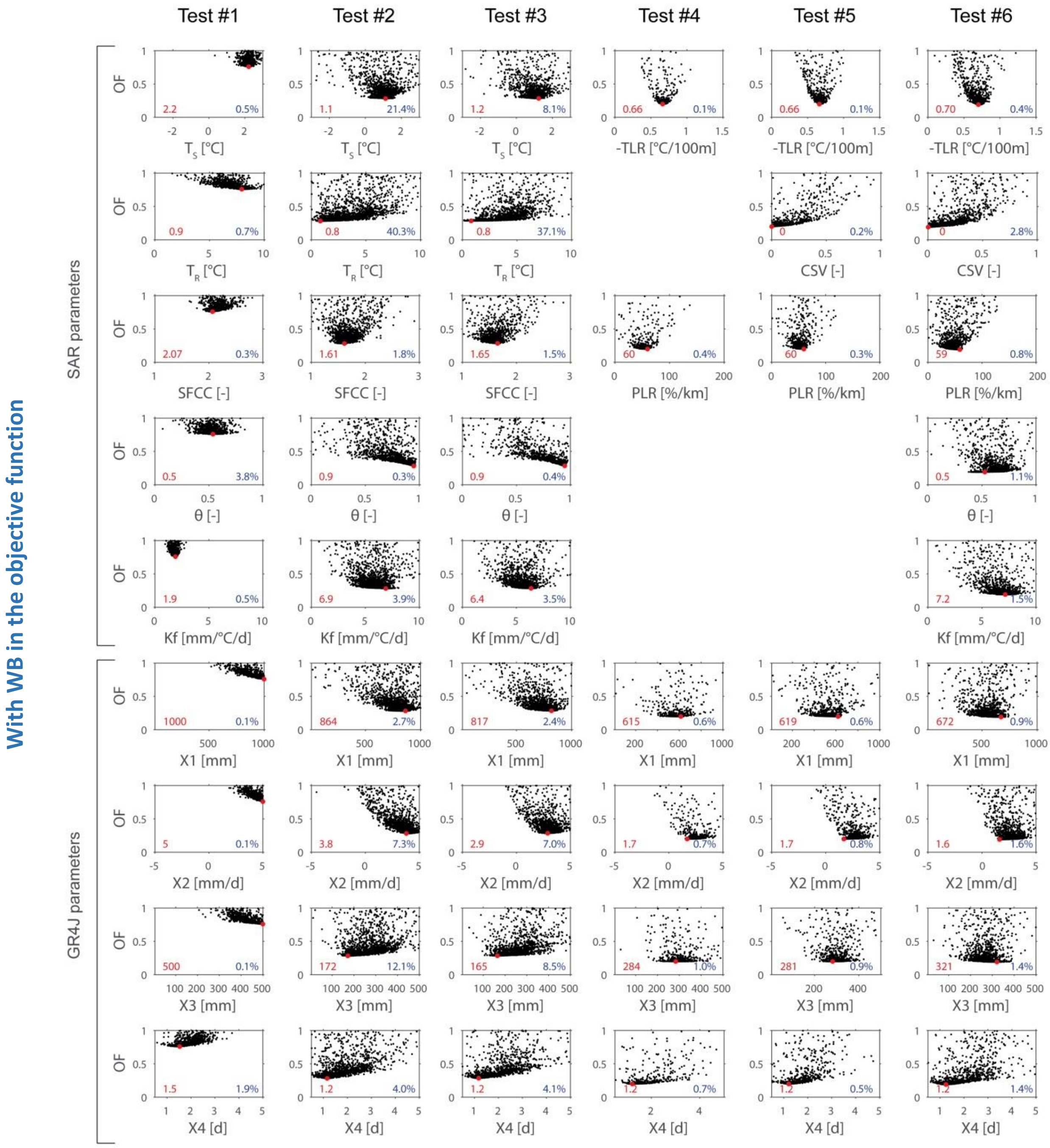


Fig. 2 Comparison of snow-hydrological simulations with elevation dependency according to Tests #1 to #4 (see Table 5) with GR4J for the Durance at Serre-Ponçon. The graphs show mean inter-annual time-series of temperature, precipitation, streamflow and fractional snow cover at the catchment scale in validation over the period 2008–2016. T_{mean}, P_{mean} and S_{mean} stand for mean annual temperature, precipitation, and snowfall, respectively. The efficiency criterions NSE_{SNOW}, NSE_{Q}, NSE_{lnQ} and VE_{C} are computed from continuous (not mean seasonal) series over 2008–2016.


With WB in the objective function
Test #1
Test #2
Test #3
Test #4
Test #5
Test #6
SAR parameters
GR4J parameters
OF
T_S [°C]
-TLR [°C/100m]
T_R [°C]
CSV [-]
SFCC [-]
PLR [%/km]
θ [-]
Kf [mm/°C/d]
X1 [mm]
X2 [mm/d]
X3 [mm]
X4 [d]

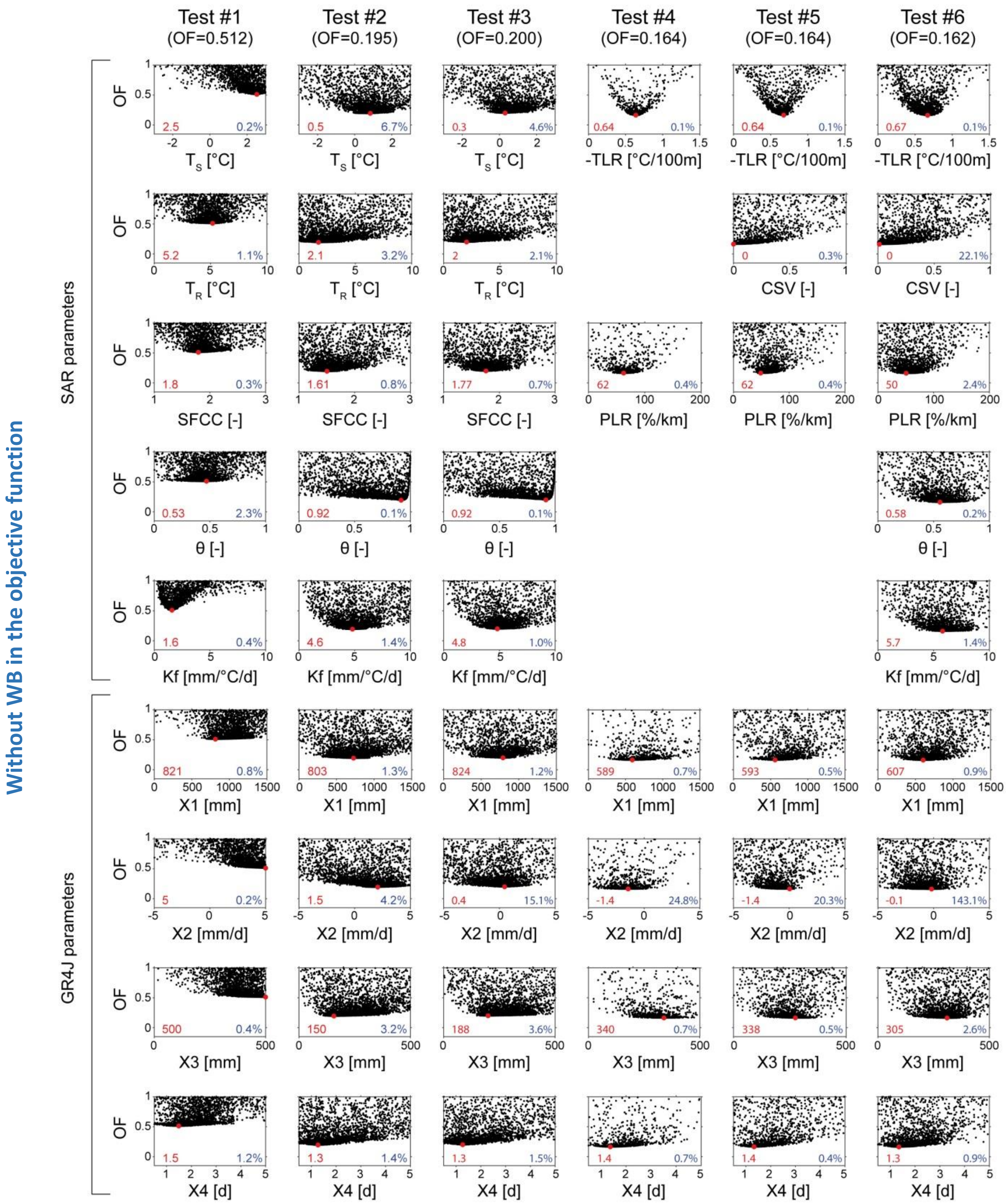


Fig. 3 Parameter sensitivity to the objective function (OF) according to Tests #1 to #6 (see Table 5) with GR4J combined with the snow accounting routine (SAR) on the Durance at Serre-Ponçon. The values and dots in red indicate the optimised calibrated parameters when minimising OF, the black dots represent trials of the SCE-UA optimisation algorithm, and the values in blue are the variation coefficients (in %) of the 20% best parameter solutions compared to the optimised values for each parameter (the lowest value, the easiest parameter identifiability). Note that depending on the tests, the calibrated parameters of the SAR vary from 2 to 5 (see Table 5 and Table 2 in the manuscript), while the GR4J hydrological models has 4 parameters (see Table 3 in the manuscript).

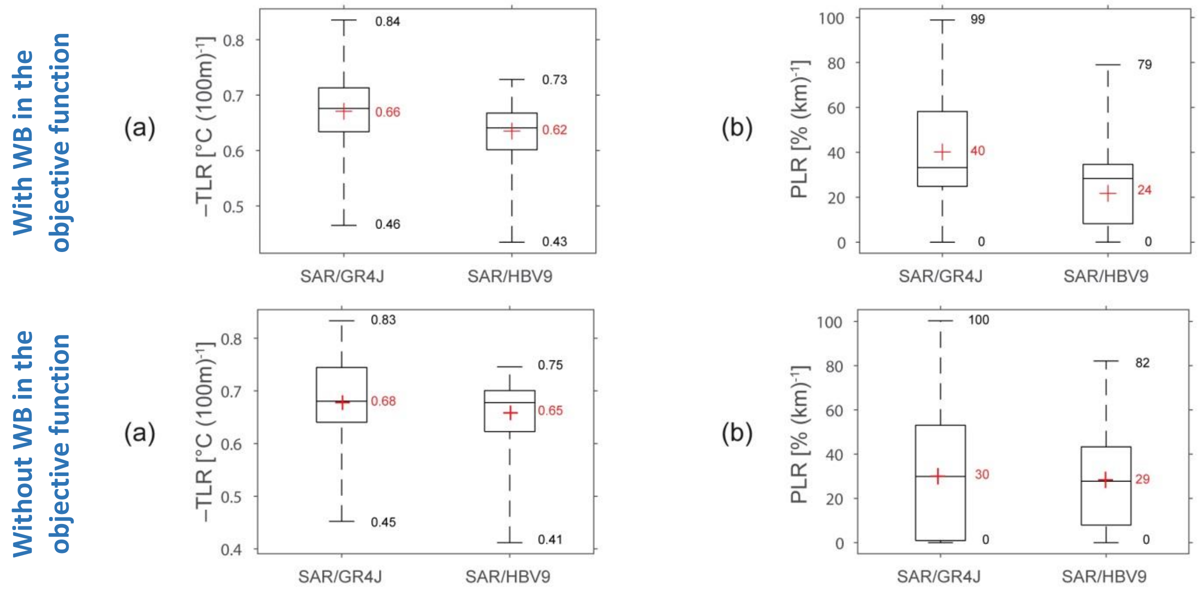


Fig. 4 Boxplots (showing 0.00, 0.25, 0.50, 0.75 and 1.00 percentiles) of the ranges of (a) temperature and (b) precipitation lapse rates calibrated with the 2-parameter SAR (Test #4) in association with the GR4J and HBV9 models on the 20 snow-affected Alpine catchments. The red crosses indicate mean values.

Referee's comment

10. Table 4: There is a strong drop in the NSE criterion for temperature when going from monthly to daily time steps for IDW and ORK. How this drop can be explained?

Authors' response and modifications to manuscript

The drop in the NSE criterion for temperature was in fact when going from monthly (or daily) to yearly time scale for IDW and ORK. As it can be seen from the following figure 5, the NSE criterion is very sensitive to the number of considered time step, and further on the range of sampled temperatures (which is quite different at the yearly versus monthly time step). As a result, the NSE values between the different methods should be compared only for a given time scale, and not in between time scales.

On the opposite, the RMSE criterion (which was used as objective function in the JK cross-validation) is better representative for the comparison of temperature (whatever the time scale) since units are directly comparable. Following the referee comment, and as NSE values seemed to cause confusion in the result interpretation, only the RMSE values are now presented in Table 4.

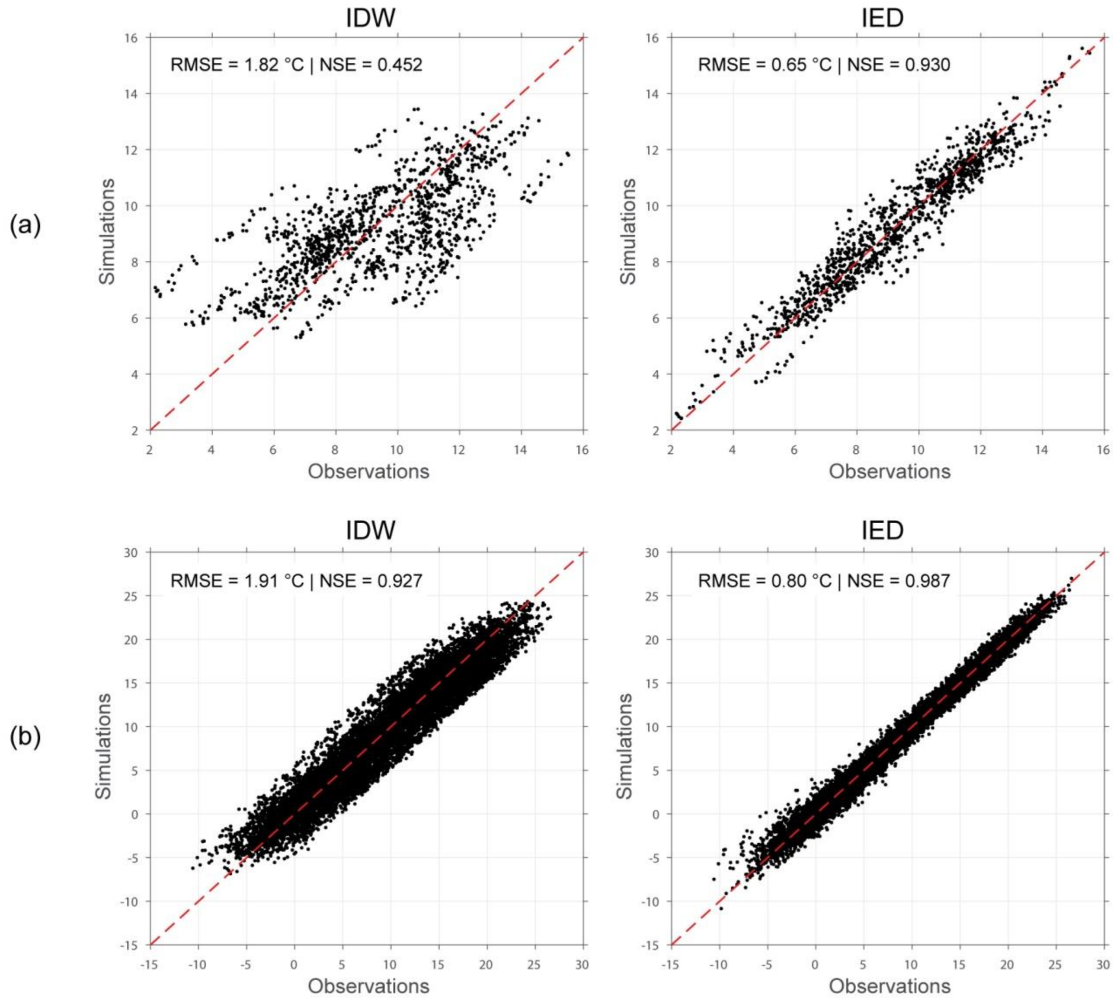


Fig. 5 Cross-validation of the IDW and IED methods against (a) yearly and (b) monthly series from temperature gauges over the period 2000–2016. The NSE criterion is very sensitive to the number of considered time step, and further on the range of sampled temperatures (which is quite different at the yearly versus monthly time step). As a result, there is drop in NSE values for temperature when going from monthly to yearly time step, particularly with the IDW method. Therefore, the NSE values between the different methods should be compared only for a given time step, and not in between time steps. On the opposite, the RMSE criterion (used as objective function for cross-validation) is better representative for the comparison of temperature whatever the time step.

Referee's comment

11. L472-476: I think this result is the consequence of using WB in the objective function. As mentioned above, this constraint is artificial and potentially counterproductive for the efficiency of the model.

Authors' response and modifications to manuscript

This result was indeed the consequence of using WB in the objective function. Please note however that this constraint was not counterproductive for the efficiency of the model as it can be seen clearly from Figures 1 and 2 of the revision notes: with or without WB in the objective function, the hydrological predictions are significantly improved as regards to the efficiency criterions when using a SAR targeting for the temperature and precipitation lapse rates (Tests #4, #5 and #6).

Following the referee comment, a more classical objective function was used (i.e. without the WB term in the OF). Obviously, the water balance was not closed systematically and it was not interesting anymore to present the Budyko graphs for the different tests. The Budyko graphs and associated comments were therefore removed from the manuscript.

Referee's comment

12. L510-516: I find this a bit contradictory with the WB constraint. If the author makes the hypothesis that underground water exchanges between catchments may play a key role, why does the author constrain water balance not to account for such exchanges in the optimization phase?

Authors' response and modifications to manuscript

I do not understand why this comment would be contradictory with the WB constraint. As explained above, inter-catchment groundwater exchanges (IGE) are not the only reason why the water balance may not be closed in the Budyko sense. Other reasons (maybe more important) may play a key role such as errors in the precipitation volumes or wrong estimate of potential evapotranspiration. Since the paper deals with the optimization of temperature (impacting snow accumulation and melt, but also evapotranspiration estimates) and precipitation gradients, constraining the water balance in the objective function aimed mainly at enhancing the parameter identifiability (see Fig. 3 of the revision note) without deteriorating the modelling efficiency (see Figs. 1 & 2 of the revision note). While the HBV model considers the catchments as closed systems, GR4J allows potential IGE via its X2 parameter. Therefore, the Budyko constraint left room for not completely close the water balance with GR4J, as it was commented in the submitted paper.

However, since sensitivity analyses to the objective function are far beyond the paper issue and because other readers may not be convinced by the proposed constraint without an in-deep demonstration, I renounced to the WB constraint in the objective function (see answers to the referee comment #9) and I re-run all the simulations with a more classical *OF*. Figures and comments were changed accordingly. Please note that it did not change the main findings of the paper.

The following paragraph (and associated new table) was also added in the section 5.3 (Identifiability of the parameters) to further discuss on the IGE issue and suggest the findings of the initial submission using the WB term in the OF:

"…Equifinality is also reduced in Tests #4–6 for the parameters controlling runoff generation and routing (*X1*, *X3* and *X4*). On the opposite, the parameter of the inter-catchment groundwater flows (*X2*) is poorly identifiable with variation coefficients of 24.8%, 20.3% and 143.1% with Test #4, Test #5 and Test #6, respectively. This suggests that inter-catchment groundwater exchanges (IGE) do not play a key role in the studied catchments. Indeed, fixing *X2* to a value of 0 (i.e. without potential IGE) with an alternative GR3J model provided similar mean validation efficiency on the set of catchments as compared to the GR4J associated with the 2-parameter SAR (Table 7). However, other objective functions may result in other findings as far as IGE are concerned. For instance, additional tests (not shown here for brevity sake) confirmed that it was possible to greatly reduce the X2 equifinality without decreasing the model efficiency by adding a water balance term in the objective function to constrain the proportion of years respecting the water and energy balance in the Turc-Budyko non-dimensional graph (see Andréassian and Perrin, 2012). These tests suggested that it remains important to explicitly represent inter-catchment groundwater transfers in association with correcting or scaling factors applied to the precipitation input data to render the distribution between evapotranspiration, streamflow and underground fluxes more realistic, as already reported by Le Moine et al. (2007)."

Table Mean validation efficiency on the set of 20 catchments with the GR4J model and the GR3J model in association with the 2-parameter SAR.

Model	Total number of free parameters	Mean NSE_{SNOW}	Mean NSE_Q	Mean NSE_{lnQ}	Mean VE_C
2-parameter SAR/GR4J	6 (2 + 4)	0.86	0.79	0.82	0.95
2-parameter SAR/GR3J	5 (2 + 3)	0.86	0.78	0.81	0.94

Referee's comment

13. Fig. 8 is interesting. However there are some cases which reveal that the optimum is probably outside the preset parameter range. This is typically the case for Test#1 for parameters X1 to X3. Therefore the ranges should be extended.

Authors' response and modifications to manuscript

I think the referee comment is too categorical here. For Test#1 (and only for Test#1), the parameters *X1* to *X3* indeed reached the maximum allowed range. Please note however that Test#1 serves as a benchmark. As explained in Table 5 and in the text, it differs from the other tests because no elevation dependency in the *T* and *P* inputs are considered. As a result, hydrologic predictions with Test#1 are significantly (and rather logically) outperformed by the other approaches accounting for elevation-dependency (see e.g. Figures 1 and 2 in the revision note). Extending the range of the parameters would be both poorly efficient in improving the simulations and incorrect from a numerical point of view. The referee has to be aware that the parameter ranges were preset to values recommended by the models' authors (Perrin et al., 2003 for GR4J and Beck et al., 2016 HBV9). They have been found after numerous simulations in very different contexts and can be judged as large enough. By the way, it can be seen in Figure 3 of the revision note, that no parameter limits are reached in the other tests, thus suggesting that the preset parameter range are adequate.

To address the referee comment, I only extended the range of the *X1* parameter (from 10-1000 mm to 0-1500 mm) of GR4J to ensure a better correspondence with the UZL parameter range of HBV9. All simulations were re-run with this new range (and also with an objective function without WB, see answer to the referee's comment #9), and Figures and comments were modified accordingly. Please note that this did not change the results (see Figure 1 of the revision note) and the parameters X2 and X3 still reached the maximum allowed range (see Figure 3 in the revision note) with Test#1 (and only for Test #1) for the reasons explained above. The following comment was also added in the beginning of section 5.3:

"…The maximum allowed parameter range is only reached for the parameters *X1* and *X2* with Test #1. This test differs from the others because no elevation dependency in the *T* and *P* inputs are considered. Consequently, hydrologic predictions of Test #1 are significantly outperformed by the other approaches. Extending the parameter ranges beyond the tested values would be both poorly efficient in improving the simulations and incorrect from a numerical point of view since they were set to values recommended by the models' authors. Moreover, no parameter limits were reached in the other tests, thus suggesting that the parameter ranges are adequate…"

Referee's comment

Cited references:

Besic, N., et al. (2014). "Calibration of a distributed SWE model using MODIS snow cover maps and in situ measurements." Remote Sensing Letters 5(3): 230-239.

Henn, B., et al. (2016). "Combining snow, streamflow, and precipitation gauge observations to infer basin-mean precipitation." Water Resour. Res. 52(11): 8700-8723.

Le Moine, N., et al. (2015). "Hydrologically Aided Interpolation of Daily Precipitation and Temperature Fields in a Mesoscale Alpine Catchment." J. Hydrometeorol. 16(6):2595-2618.

Leleu, I., et al. (2014). "Re-founding the national information system designed to manage and give access to hydrometric data." La Houille Blanche(1): 25-32.

Naseer, A., et al. (2019). "Distributed Hydrological Modeling Framework for Quantitative and Spatial Bias Correction for Rainfall, Snowfall, and Mixed-Phase Precipitation Using Vertical Profile of Temperature." J. Geophys. Res.-Atmos. 124(9): 4985-5009.

Rahman, K., et al. (2014). "Streamflow response to regional climate model output in the mountainous watershed: a case study from the Swiss Alps." Environmental Earth Sciences 72(11): 4357-4369.
Riboust, P., et al. (2019). "Revisiting a simple degree-day model for integrating satellite data: implementation of SWE-SCA hystereses." Journal of Hydrology and Hydromechanics 67(1): 70-81.
Zhang, F., et al. (2013). "Snow cover and runoff modelling in a high mountain catchment with scarce data: effects of temperature and precipitation parameters." Hydrol. Processes 29(1): 52-65.

Authors' response and modifications to manuscript

Most of the proposed references were judged useful. Therefore, they were cited in the text and added to the reference list, expect that of Leleu et al. (2014) which I could not find.