Responses to comments from anonymous Referee 1

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

This paper presents a nicely conceived study where several alternative approaches for distributing temperature and precipitation are compared in a mountain region (French Alps). In addition to standard interpolation approaches based on inverse-distance weighting and Kriging, the author explores the possibility of optimizing lapse rates as part of a snow-hydrologic-model calibration procedure. Results of a split-sample test show that the latter approach provides improved results for the target variables considered during calibration, that is, fractional snow cover (FSC) from MODIS, streamflow, and the water balance. Also, optimizing the temperature and precipitation distribution algorithm with hydrologic data results in the temperature and precipitation fields being colder and wetter than those obtained by using only in-situ measurements of temperature and precipitation, respectively; this result agrees with expectations, especially since the considered ground-based network is not representative of high elevations in the study catchments.

I enjoyed reading the manuscript and I think it represents an interesting contribution for HESS, especially because the investigated topic is a clear open issue in mountain hydrology. I do have several general and specific comments, which I attach below. Overall, I think that the revision is feasible.

Authors' response

I would like to sincerely thank the referee for the time and effort he/she spent in reading the initial manuscript and for making many clear, pertinent and constructive suggestions for improvement. This helped a lot to re-write the paper.

General comments

Referee's comment

1. in my understanding, the author essentially compares two main strategies: the first is to regionalize temperature and precipitation based on in-situ data and various interpolation/extrapolation schemes based on IDW or Kriging (Section 3); the second is to ''embed" part of the distribution process into the snow-hydrologic models via lapse rates that will correct a first-guess distribution based on IDW (see Section 4, Table 2 and 5 and Equations 9 and 10). While the first approach is independent from hydrologic data like fractional snow cover and streamflow, the second does take advantage of these data to adjust some of the distribution parameters. In my understanding, the main point of the paper is that the second strategy is superior to the first, especially since adjusting snow parameters rather than precipitation-distribution parameters does not allow the model to significantly improve its performance (see Table 5 and Fig. 5). Unless I am missing something here, this improvement was however assessed based on the same hydrologic variables that were used to calibrate the snow-hydrologic models, rather than on independent measurements of the two variables of interest: temperature and precipitation. This left me wondering if this experiment shows that "calibrating the local gradients using an inverse snow-hydrological modelling framework" improves actual temperature and precipitation estimates, or if it shows that it improves hydrologic predictions. In principle, one would expect the obtained altitudinal gradients to be both more effective in terms of hydrologic predictions and in terms of temperature and precipitation, but the improvement obtained by ''embedding" part of the distribution process into the snow-hydrologic models is quantified in terms of modeling skills for fractional snow cover, streamflow, and the water balance (Figs. 5 to 8) rather than for independent estimates of temperature and precipitation. If independent data of temperature and precipitation at high elevations are not available, then I would recommend the *author to clarify the extent to which these results apply to temperature and precipitation in addition to hydrologic variables.*

Authors' response and modifications to manuscript

I agree that an improved fit for hydrologic variables may not automatically mean that the model is also better representing weather patterns of temperature and precipitation. Since independent data of temperature and precipitation at high elevations were indeed not available, it was not possible to clarify the extent to which these results apply to temperature and precipitation in addition to hydrologic variables. As a result and following the relevant referee comment and argumentation, the following text has been added in the Section 6.2. Recommendations:

"….However the differences in the two compared approaches are worth discussing. The first is to regionalize temperature and precipitation based on in-situ data and various interpolation/extrapolation schemes based on IDW or Kriging; the second is to "embed" part of the distribution process into the snow-hydrologic models via calibrated lapse rates correcting a first-guess distribution based on IDW. While the first approach is independent from hydrological data like fractional snow cover and streamflow, the second does take advantage of these data to adjust some of the distribution parameters. The second strategy prove superior to the first, especially since calibrating distribution parameters rather than adjusting snow parameters allowed the models to significantly improve their performance. This improvement was however assessed based on the same hydrological variables that were used to calibrate the snow-hydrologic models, rather than on independent measurements of temperature and precipitation. This left wondering if improving hydrologic predictions by calibrating the local gradients using an inverse snow-hydrological modelling framework also improves actual temperature and precipitation estimates. In principle, one would expect the obtained altitudinal gradients to be both more effective in terms of hydrologic predictions and in terms of temperature and precipitation, but the improvement obtained by "embedding" part of the distribution process into the snow-hydrologic models was only quantified in terms of modelling skills. An improved fit for hydrologic variables may not automatically mean that the model is also better representing weather patterns of temperature and precipitation. A good example is that the optimized lapse rates (Fig. 10) can locally be quite different between the two hydrologic models considered. Since independent data of temperature and precipitation at high elevations are not available, we were not able to clarify the extent to which these results apply to temperature and precipitation in addition to hydrologic variables."

Referee's comment

2. The point above is particularly important since hydrologic models may suffer from several sources of conceptual and parametric uncertainties, some of which are visible in the interesting Figure 8. It follows that an improved fit for hydrologic variables may not automatically mean that the model is also better representing weather patterns of temperature and precipitation. A good example here is that the obtained lapse rates (Fig. 9) can locally be quite different between the two hydrologic models considered. To me, this may challenge the idea that this approach could be used to "infer local altitudinal gradients from a sparse network of gauges based on key parameters in the snow-hydrological models" (L 592ff). It does suggest that the method improves hydrologic predictions, but implications for actual temperature and precipitation are more elusive to me and should be discussed more extensively.

Authors' response and modifications to manuscript

Agreed. The sentence (L 592 ff) has been removed and an entire paragraph (based on the argumentation from the referee in his/her general comment #1 and #2 has been added in the Section 6.2. Recommendations (see answer to the preceding comment).

- We removed the sentence " we thus suggest using the proposed modelling framework to infer local altitudinal gradients from a sparse network of gauges based on key parameters in the snow-hydrological models"

Referee's comment

3. Related to this, both hydrologic models were used in lumped mode (L340), even if several other modeling approaches explicitly account for spatial variability in hydrologic processes (e.g., raster-based models). At least some discussion on this point would be interesting.

Authors' response and modifications to manuscript

Agreed. This is part of an on-going research. To address the referee comment, the following text has been added at the end of the conclusion section (6.3. Prospects):

"…Finally, it is worth mentioning that spatial variability was only considered along five elevation bands in each catchment since preliminary tests showed no improvement in the hydrologic predictions when applying the SAR in a full distribution mode. However, the SAR was not designed to explicitly account for topographic effects (slope, aspect and shading) on snow redistribution, accumulation and melt (see e.g. Frey and Holzmann, 2015). A grid-based temperature-index model could thus be implemented to include potential clear-sky direct solar radiation at the surface, thus considering both the seasonal variations of melt rates and the geometric effects on melt attributable to terrain (see e.g. Hock, 1999). It would thus be interesting to assess whether accounting for the influence of such effects can further improve the daily hydrologic predictions at the basin scale."

Referee's comment

4. Spatial variability was considered in the snow model, which was implemented along five elevation bands in each catchment. This model does include all fundamental snow processes, but in my understanding does not include a specific provision for wind drift. Relying on FSC from MODIS may sometimes lead to confounding effects in this regard, where wind-driven accumulation and erosion is mistakenly assumed as due to precipitation or melt. Was this somehow taken into account here, or could the author suggest how to include this in the framework?

Authors' response and modifications to manuscript

The snow accounting routine (SAR) is only based on (semi-)distributed temperature and precipitation inputs to simulate the main snow processes related to accumulation and melt. As a result, it does not include a specific provision for wind drift within each catchment, which would probably require an additional distributed input regarding wind speed and direction. Given the challenge to distribute temperature and precipitation from a sparse gauge network, distributing wind (from even more scarce measures than for temperature/precipitation) may be unrealistic at the spatio-temporal scales considered (daily analysis in the French Alps over the period 1998–2016). It would also probably require applying the SAR in a full distribution mode. But, even doing so, it is unlikely that accounting for wind-driven accumulation and erosion would have a significant impact on the streamflow and FSC simulations at the basin scale, because it can be assumed that these processes are somewhat averaged at the basin scale. Although these aspects are worth discussing, it seems difficult to integrate them into the text without adding weight and making the discussion too large and complex. However, in link with the preceding referee comment, note that the following text (dealing also with spatial variability) has been added at the end of the conclusion section:

"…Finally, it is worth mentioning that spatial variability was only considered along five elevation bands in each catchment since preliminary tests showed no improvement in the hydrologic predictions when applying the SAR in a full distribution mode. However, the SAR was not designed to explicitly account for topographic effects (slope, aspect and shading) on snow redistribution, accumulation and melt (see e.g. Frey and Holzmann, 2015). A grid-based temperature-index model could thus be implemented to include potential clear-sky direct solar radiation at the surface, thus

considering both the seasonal variations of melt rates and the geometric effects on melt attributable to terrain (see e.g. Hock, 1999). It would thus be interesting to assess whether accounting for the influence of such effects can further improve the daily hydrologic predictions at the basin scale."

Referee's comment

5. I am also interested in the different outcomes of this analysis for precipitation between the daily and the annual time scales (table 4). Maybe one key to interpret this result is that summer vs. winter precipitation patterns are different, and the in-situ network might be more representative of the former than of the latter (or vice versa). I am thinking to convective precipitation here, which sometimes show significantly different elevational gradient from stratiform or orographic precipitation. Some more discussion on precipitation regimes could be interesting in this paper.

Authors' response and modifications to manuscript

I sincerely did not see how the analysis of seasonal patterns of precipitation (see Figure 1 below) can explain why interpolation performance is improved by the external drift at the annual (and also monthly) time scale, while it is not at the daily time scale. According to me, this only shows that the correlation between precipitation and topography increases with the increasing time aggregation as already reported in other studies (e.g., Bárdossy and Pegram, 2013; Berndt and Haberlandt, 2018). The elevation-dependency of precipitation thus depends significantly on the accumulation time. For instance, if precipitation events do not occur exactly the same day within the surrounding neighboring gauges, the correlation between precipitation and topography may be weak at the daily time scale, whereas it may be more significant at the monthly, seasonal or annual time scale…

The two preceding sentences were added in the Section 5.1 to try to better explain the different outcomes for precipitation between the daily and the annual time scales (Table 4).

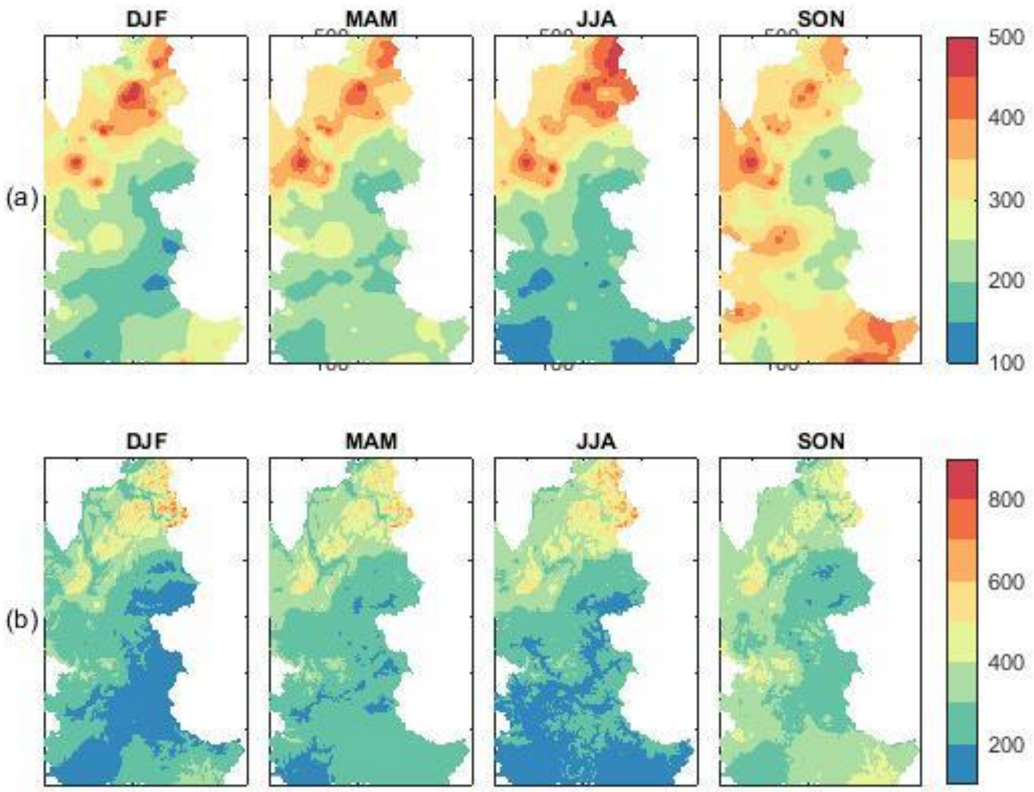

Fig. 1 Seasonal patterns of precipitation with (a) IDW applied to in-situ network (without elevation depend) and (b) IED applied to in-situ network (with elevation dependency via external drift). Values are in mm per season (i.e. 3 months).

Specific comments

Referee's comments in grey – Authors' responses and modifications to manuscript in blue

Line 149: is this because no gap was originally present in the dataset, or because these gaps were filled? If the second, maybe briefly mention how.

These meteorological gauges were selected because no gap was originally present in their original time series from the 1st September 2000 to the 31st of August 2016. The text has been modified to make the purpose clearer.

Section 2.1: a histogram with the elevation distribution of in-situ stations may be helpful, along with more details about the climatology of the study period (annual mean temperature and precipitation, annual runoff etc). Doing so may help the author to set the context of the analysis, especially for non-local readers.

Done. Figure 1 was modified to incorporate elevation distributions of in-situ stations, DEM and basins (see below). 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 these 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).

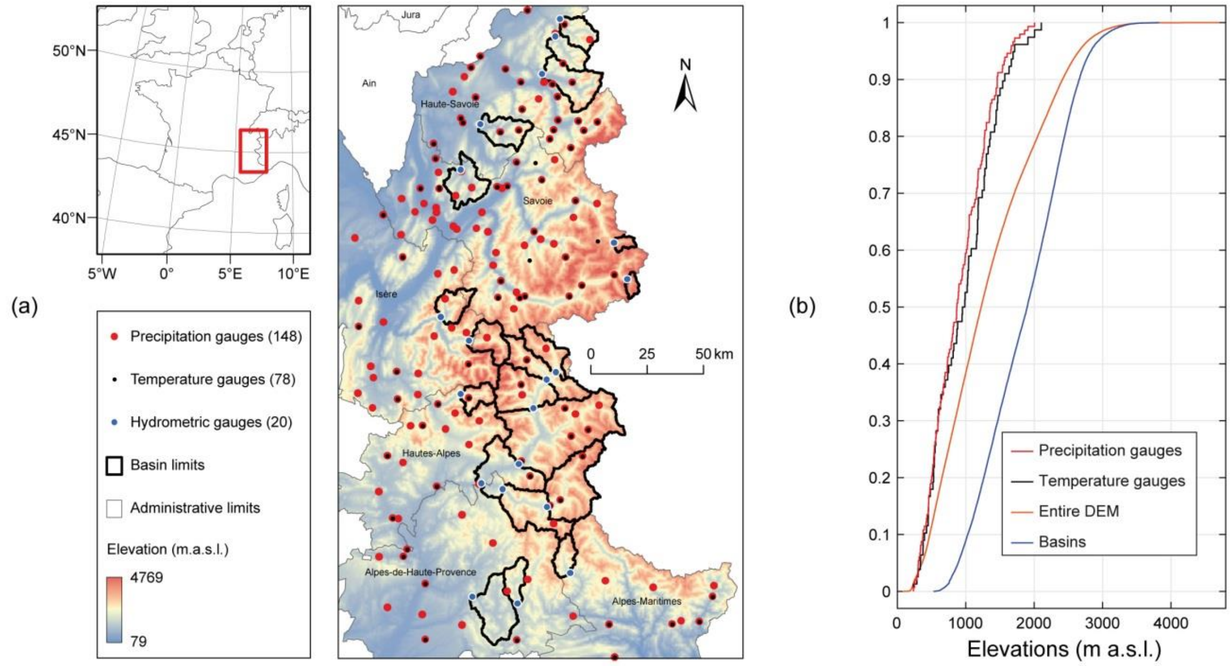

Fig. 1 Study area and data: (a) Location of the selected precipitation, temperature and streamflow stations, as well as elevations from a SRTM digital elevation model (DEM resampled to a grid with 0.5x 0.5 km cells) in the French Alps; (b) Elevation distributions of in-situ stations, DEM and basins.

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	Glacierized area	Elevations (m.a.s.l.)		Mean annual precip. (P)	Snowfall fraction (S)	Mean annual temp. (T)	Mean annual streamflow (Q)
		(km^2)	(%)	Min	Max	(mm/yr)	(%)	(°C)	(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

Section 2.2: is any of these catchments glacierized? If so, how were glaciers considered in this framework? If not, may glaciers hamper the applicability of this method in other regions, especially with regard to the mass-balance-closure term in Eq. 12 and Fig. 6?

During the catchment selection process, we tried to minimize possible interactions with non-snow related processes that could also influence streamflow. Therefore, we tried avoiding glacierized basins, basins with known inter-catchment groundwater flows, and catchments with documented flow diversions. However, it must be acknowledged that some basins are partly glaciated. Table 1 now includes the percentage of glaciated areas estimated for each catchment from the World Glacier Inventory (NSIDC, 2012).

As the majority of basins had negligible glacierized areas (see Table 1), no specific glacier model was activated. This led to ignore the late summer contribution of glacier melt to river discharge in the three basins having 9–12% glacierized areas, which did not affect significantly the mass-balance-closure term at the annual scale. However, it is worth mentioning that more important contribution of ice melt would require a glacier component to not hamper the use of WB in the objective function.

The last paragraph was inserted at the end of the 4.3.2 Section to highlight on the fact that, without activating a glacier model, the applicability of the mass-balance-closure term may be hampered on catchments with an important contribution of ice melt.

Section 2.3: the approach by Gascoin et al. 2015 was, to my knowledge, developed in the Pyrenees, a mountain range with significantly lower elevations than the Alps. How was the method adapted for the French Alps? Is the performance similar to that originally published by Gascoin et al. 2015 in a different mountain range?

The referee is right. The approach by Gascoin et al. (2015), itself inspired by previous works from Parajka and Blöschl (2008) in Austria, and Gafurov and Bárdossy (2009) in Afghanistan, was adapted for the French Alps. To do so, the MODIS snow-products were gap-filled in various mountain ranges (including the Pyrenees and the French Alps). The missing data (due mainly to cloud obscuration) were less important in the Pyrenees (46%) than in the Alps (52%) for the same period (2000–2016). As a result, for the temporal deduction by sliding the time filter, we allowed the window size to be incremented up to 9 days in the Alps (versus up to 6 days in the Pyrenees) to account for the differences in cloud obscuration. Moreover, Gascoin et al. (2015) used an adjacent spatial deduction as a second step: each no-data pixel was reclassified as snow (no-snow) if at least five of the eight adjacent pixels were classified as snow (no-snow). We did not use this step considering it was a too rough approximation. Finally, we validated our adapted method on the two mountain ranges. Validation based on confusion matrices with 1 image/month (i.e. about 200 cloud-free images over the studied period) showed that the gap-filling technique applied to the MODIS snow-products led to the reconstruction of validation images with average accuracies of 98% in the Pyrenees and 94% in the Alps.

To address the referee comment, more details regarding the gap-filling method and validation have been given through additional text and a new figure (see below) in the Section 2.3.

"MOD10A1 (Terra) and MYD10A1 (Aqua) snow products version 5 were downloaded from the National Snow and Ice Data Center for the period 24 February 2000–1 January 2017. This corresponds to 6157 dates among which 98.8% are available for MOD10A1 and 85.8% for MYD10A1 since Aqua was launched in May 2002 and became operational in July 2002. These snow products are derived from a Normalised Difference Snow Index (NDSI) calculated from the near-infrared and green wavelengths, and for which a threshold was defined for the detection of snow (Hall et al., 2006, 2007). Cloud cover represents a significant limit for these products, which are generated from instruments operating in the visible-near-infrared wavelengths. As a result, the grid cells were gap filled to produce daily cloud-free snow cover maps of the study area. The different classes in the original products were first merged into three classes: no-snow (no snow or lake), snow (snow or lake ice), no-data (clouds, missing data, no decision, or saturated detector). The missing values were then filled according to a gap filling algorithm inspired by techniques proposed in the literature (Parajka and Blöschl, 2008; Gafurov and Bárdossy, 2009; Gascoin et al., 2015). The algorithm works in three sequential steps:

(i) Aqua/Terra combination: for every pixel, if no-data was found in MOD10A1 then the value from MYD10A1 was used instead. Priority was given to MOD10A1 because MYD10A1 was found to be less accurate (see Gafurov and Bárdossy, 2009).

(ii) Temporal deduction by sliding time filter: a no-data pixel was reclassified as snow (no-snow) if the same pixel was classified as snow (no-snow) in both the preceding and following grids. The preceding and following grids were searched within a sliding temporal window, whose size was incremented up to 9 days in order to reduce the remaining fraction of no-data pixels to below 12% at least (Fig. 2a). It should be noted that three periods of gaps in an upper time window (11, 13 and 18 days) were present in the data because of technical failures of the MODIS sensor. In these cases, a longer time deduction was used beforehand to specifically fill these periods.

(iii) Spatial deduction based on elevation and neighbourhood filter: for each date and each pixel, a 3x3 neighbourhood spatial filter was used to account for the elevation and the data in the neighbouring pixels to fill the remaining no-data pixels. Two configurations were considered: either the central pixel has no-data and the algorithm tries to attribute a neighbouring value, or the central pixel has a value that can be assigned to some of its neighbours. The two configurations were repeated until there were no more gaps (Fig. 2a).

The resulting database consists of 5844 binary (snow/no-snow) maps at 500 m spatial resolution for the period 2000–2016 (16 hydrological years, from the 1st of September 2000 to the 31st of August 2016). As a synthesis of these maps, snow cover durations over the study area are presented in Fig. 2c.

In order to validate the gap-filling technique, a daily snow product with less than 10% of no-data pixels was selected for each month of the studied period. These images were "blackened" (i.e. with 100% no-data pixels), before applying the algorithm over the entire period to fill all gaps, including validation images. Filling accuracy was estimated for each image removed by computing confusion matrices which compared the pixels of the removed validation images and the filler reconstructions of these images. Validation based on confusion matrices with 1 image/month showed that the gap-filling technique applied to the MODIS snow-products led to the reconstruction of images with average accuracies of 94% (Fig. 2b). The mean monthly accuracies show greater ease in filling gaps in summer than in winter due to the differences in cloud obscuration. However it should be noted that the actual accuracy of the MODIS gap-filling technique is necessarily greater than that of the validation procedure, in which many quality images needed to fill the gaps were missing."

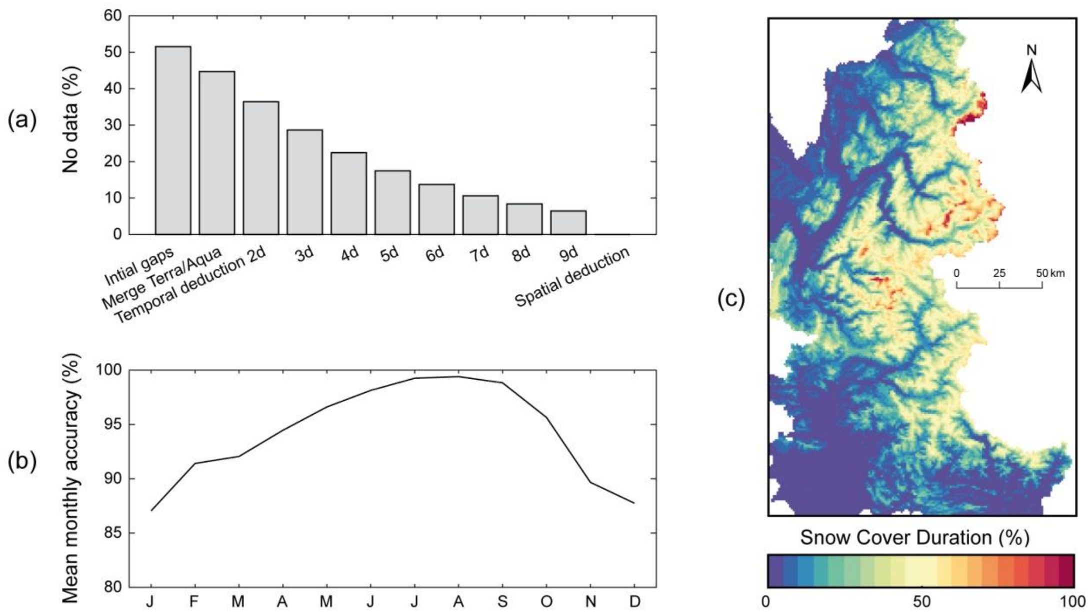

Fig. 2 Results of gap-filling applied to MODIS snow products: (a) Evolution of the number of pixels classified as no-data (e.g., clouds) during the gap-filling procedure; (b) Mean monthly accuracies according to validation based on confusion matrices with 1 image/month, i.e. ~200 cloud-free images over the 2000–2016 period; (c) Snow cover duration based on gap-filled MODIS snow products over the 2000–2016 period.

Line 263ff: was mean precipitation computed across the whole study region? Might doing so exclude more localized precipitation events in favor of more widespread stratiform events?

In the initial submission, parameters were only estimated for days with at least 1 mm mean precipitation, i.e. approximatively 41% of the daily sample (whereas parameters were calculated for all months and years since there were no locations with dry months or dry years in the dataset). The 1 mm mean precipitation was indeed computed across the whole study region. The initial motivation was to limit the effect of precipitation zero values during the optimization of the daily local regression between precipitation and elevation for external drift. Doing so might indeed bias the optimization of the *n(u)* surrounding observations during the leave-one-out cross-validation, thus potentially affecting the computation of the external drift with the KED and IED technique when further used to interpolate daily precipitation in the study region (Fig. 5 – Fig. 4 in the initial submission).

Following the referee comment, the leave-one-out cross-validation was re-run with the whole set of daily samples using the four interpolation techniques (IDW, ORK, KED and IED). As a result, the RMSE, MAE and NSE values changed for the cross-validation of the interpolation methods against precipitation daily series (see new values in Table 4). However, interestingly, this did affect neither the ranking between methods nor the optimized interpolation parameters. For instance, 17 surrounding neighbors were still found during optimization to compute altitudinal gradients of precipitation based on the daily, local linear regression with KED and IED. This means that the initial results (optimized parameters and daily interpolated precipitation) were not affected by the 41% subset. In other words, the subset did not exclude more localized precipitation events in favor of more widespread events. However, as it revealed unnecessary, the sentence about it in the section 3.3 was deleted and the RMSE, MAE and NSE scores were changed in Table 4 accordingly to the computation of the cross-validation with the all daily precipitation samples.

Title of Section 4: ASSESSSMENT -> ASSESSMENT.
Corrected in the revised manuscript.

Section 4: a table with the list of all parameters considered by the snow and hydrologic models would be helpful, including an explicit statement of which parameters where calibrated. Some of these parameters are only mentioned at the very end of the manuscript (Section 5.3).
The referee probably missed Table2 that lists all (fixed or calibrated) parameters regarding the snow accounting routine (SAR) and Table3 that lists all calibrated parameters regarding the two hydrological models. The two models were run with the SAR on top. Since the SAR parameters were calibrated differently (from 2 to 5 free parameters) in each modelling experiment, it was necessary to introduce two separated tables for the parameters. Note however that these two tables are presented in two consecutive sections, making them theoretically understandable.

Line 321: this should in fact be evaporation to me, since there is no transpiration in the snow module (correct?)
There is no transpiration in the snow module. However, PE is computed based on the temperature-based formulation proposed by Oudin et al. (2005) which explicitly mentioned the "evapotranspiration" term ("…towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling"). The formula aims at estimating the maximum amount of evapotranspirable water taking into account the meteorological context and for a plant cover corresponding to grass. As a result, it is common to use the term potential evapotranspiration when one refers to the Oudin formula.

Section 4.3.1: where the two periods similar in terms of snow conditions and streamflow, as well as mean temperature and mean precipitation across the in-situ network?
Mean annual precipitation increased by around 17% between the two periods (1104 mm/yr vs. 1294 mmm/yr), while mean annual temperature were stable (9.2 °C vs. 9.3 °C) across the in-situ network presented in Section 2.1. Although the second period was generally wetter, this hides differences in between the catchments. At the basin scale, the differences between the two periods ranged from -10% to 15% for precipitation, -0.5 °C to +0.5 °C for temperature, and -11% to +50% for streamflow. These details are now indicated in Section 4.3.1.

As far as the snow conditions are concerned, differences in mean annual snowfall between the two periods ranged from -10% to +50% according to the best-performing simulations. As these values were obtained through simulations, we feel it is not correct to mention them in Section 4.3.1. We hope the above values regarding precipitation, temperature and streamflow between the two periods are sufficient to address the referee comment.

Eq 12 and Section 4.3.2: does the third component of the OF assume that interannual variability in subsurface storage is negligible? This might not be an issue in the studied area, but it may be worth mentioning this in case interested readers would like to apply this approach somewhere else. In fact, results in Section 5.3 do suggest that interannual sub-surface dynamics are worth discussing.

The referee is right. The third component (WB) of the objective function assumes that interannual variability in subsurface storage is negligible. This has been indicated in the text as required in case readers would like to apply the approach somewhere else. To address another referee comment, it was also highlighted that, without activating a glacier model, the applicability of the mass-balance-closure term may be hampered on catchments with an important contribution of ice melt.

Section 5.3 does not suggest that "interannual" sub-surface dynamics are worth discussing. Rather, it suggests that groundwater exchanges may occur by gaining/loosing water from/to neighboring basins during the year. GR4J can account for part of these exchanges through its X2 parameter whereas HBV9 considers catchments as closed systems. Despite these differences, snow-hydrological predictions are significantly improved for both HBV9 and GR4J. To address the referee comment, the following sentence was slightly rephrased in the 5.4 section:

"…The differences between the two models may be due to the GR4J ability to gain (or loose) water from inter-catchment groundwater flows through its X2 parameter (see section 5.3.), <u>unlike HBV9 which considers the catchment as a closed system</u>. On the other hand, HBV9 relies on more parameters for production and transfer, thus enabling to compensate differently for the errors in the precipitation volumes…"

Section 5.1: do statistics reported in Fig. 4 and at lines 406ff consider areas outside the studied catchments too, including Italy and Switzerland? It might be better to report statistics for the French Alps only here since this is where data were available to this study.

Yes. To address the referee comment, a geographical mask (representing only the six French administrative departments from which data were selected) was applied on the initial maps. Statistics were re-computed accordingly. Note that only means have changed since the minimum and maximum values were still in the masked maps. Note that the same mask was also applied to the DEM in Fig. 1 to make coherent the presentation.

Section 5.2: the first paragraph of this section and Table 5 should be moved to the Methods. It should also be clarified that each re-calibration mode included hydrologic parameters too (correct?)

We acknowledge that this first paragraph and Table 5 could somehow be moved to the Methods. However, as indicated in the text, "for sake of brevity, here we only present the results we obtained with the datasets interpolated with the IDW and IED procedures, since cross-validation at the daily time scale showed that they slightly outperformed the ORK and KED methods, respectively". This means that the short "methodological" paragraph and associated Table 5 are strongly linked to the results presented in the previous section 5.1 and cannot be moved to the Methods since they depend on these results. Note also that all the modelling experiment (snow-accounting routine, hydrological models and calibration/validation methods) are fully described in the Methods. Here, we further combine this modelling experiment to part of the results of the interpolation methods to introduce the tests to account for elevation dependency in the T and P inputs via the modelling experiment. Since the tests (#1-6) are rather complex and since their results are immediately presented and discussed in the Section 5.2, we believe that reading is probably easier in that way.

We hope these arguments will convince the referee.

Regarding the second comment, the parameters of the SAR and the hydrological model were indeed optimized simultaneously, as already indicated in Section 4.3.2. Following the referee comment, it was however repeated and clarified that each re-calibration mode included hydrologic parameters too, by adding the following text in the Table 5 caption: "Note that each calibration tests included also the hydrological parameters of GR4J or HBV9 (the parameter ranges tested are listed in Table 2 for the SAR and in Table 3 for the hydrological models)."

Fig. 6: it seems like all data are within the boundaries given by the water and energy limit. I am not an expert of this approach and was wondering why one should aim to obtain "the least stretched and dispersed cluster". More details on this might be helpful for other readers too.

A closer look to the Figure shows that the dots are systematically within the boundaries given by the water and energy limits only with Test #4. It can also be observed that Test #4 led to the least stretched and dispersed cluster, which was not particularly intended. This is only a finding which adds to the more important fact that the dots are all located within the "physical" limits by the water and energy limits. As this comment seemed to cause confusion, it has been removed. The text has also been modified as follows to try to better explain the differences in water balance between the tests:

"…Water balance is ensured for each year in each catchment only with Test #4. Indeed, unlike the other simulations (Tests #1, #2 and #3), all the dots are within the boundaries given by an upper water limit where $Q = P$ (i.e., $y = Q/P = 1$) and a lower energy limit where $Q = P - PE$ ($Q/P = 1 - PE/P \leftrightarrow y = 1 - 1/x$). This means that annual simulated runoff never exceeds total precipitation and that annual runoff deficit never exceeds total *PE*. Altitudinal temperature and precipitation gradients inferred from snow-hydrological modelling thus lead to more realistic catchment water balance than when they are estimated from gauges using interpolation."

Line 490ff and other similar passages of the manuscript: in fact, this result suggests to me that correcting for precipitation and temperature distribution has a stronger impact on model predictions than adjusting for other snow-related processes like phase partitioning or melt, rather than that "adapting to local snow processes is not indispensable". To me, other processes are important too, but correctly estimating total accumulation is likely the most important one here.

Done. The sentence has been modified as follows (see also other modifications in link with other comments notably in Section 6.2 Recommendations):

'…This suggests that correcting for temperature and precipitation distribution has a stronger impact on model predictions than adjusting for snow-related processes like phase partitioning or melt, and that correctly estimating total accumulation is likely to play a first-order role in the snow-hydrological responses of the studied catchments.'

Section 5.3: I would probably add more details about how parameter identifiability is quantified from Figure 8.

Agreed. Figure 8 (now Figure 9) was slightly modified to indicate more clearly which parameters belong to the SAR or to the hydrological models (GR4J). Moreover, more details have been provided in the figure caption and in the text. Variation coefficients (in %) of the 20% best parameter solutions compared to the optimised values for each parameter have also been introduced in the Figure and in the text to bring more deails about how parameter identifiability can be "quantified" from the Figure.

Line 610 and, earlier, line 490: how were these "physical or general values" obtained?

To make the purpose clearer, the initial sentence has been completed in Section 4.1 as follows:

"….The aim of using this mode was to account for elevation dependency in the T and P inputs from constant, calibrated orographic gradients while fixing the parameters that control snow accumulation and melt to physical or general values: precipitation phase determined based on a linear separation between -1 °C and +3 °C (see USACE, 1956), temperature threshold for snowmelt fixed at 0 °C, degree-day melt factor set at 5 mm. $°C^{-1}.d^{-1}$ (mean general value taken from Hock, 2003)."

The additional references have been also inserted in the reference list.

New references

Frey, S. and Holzmann H.: A conceptual, distributed snow redistribution model, Hydrol. Earth Syst. Sci., 19, 4517–4530, https://doi.org/10.5194/hess-19-4517-2015, 2015.

Gafurov, A. and Bárdossy, A.: Cloud removal methodology from MODIS snow cover product, Hydrol. Earth Syst. Sci., 13, 1361–1373, https://doi.org/10.5194/hess-13-1361-2009, 2009.

Hock, R.: A distributed temperature-index ice- and snowmelt model including potential direct solar radiation. J. Glaciology 45, 101–111, https://doi.org/10.3189/S0022143000003087, 1999.

Hock, R.: Temperature index melt modelling in mountain areas, J. Hydrol., 282, 104–115, https://doi.org/10.1016/S0022-1694(03)00257-9, 2003.

NSIDC: World Glacier Inventory, Version 1. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center. https://doi.org/10.7265/N5/NSIDC-WGI-2012-02, 2012.

USACE: Snow Hydrology: Summary Report of the Snow Investigation. Portland, Oregon, North Pacific Division, Corps of Engineers, U.S. Army, 1956.