The Cryosphere Discuss.,
https://doi.org/10.5194/tc-2017-184-AC1, 2018

Interactive comment on "Distributed vs. semi-distributed simulations of snowpack dynamics in alpine areas: case study in the upper Arve catchment, French Alps, 1989–2015" *by* Jesús Revuelto et al.

Jesús Revuelto et al.

jesus.revuelto@meteo.fr

Received and published: 22 January 2018

Author's comments:

Below we provide a detailed response to all comments and indicate resulting changes in the manuscript. Please, note that lines referred in this response are these of the manuscript tracked with changes. Additionally to changes referred in this response other changes have been accomplished to improve the final manuscript. For instance the manuscript title has been changed to better describe our study. Now the title is:

“Multi-criteria evaluation of snowpack simulations in complex alpine terrain with two spatialization approaches”. Also some sections of the article have been reduced and others include further information; please check the manuscript with tracked changes to see them.

Reviewer’s 1 comments: 2.1 Potential model deficiencies R1 I understand that the authors want to establish a long data set for validation, which may have led to the decision to use SAFRAN with 14 years of meteorological forcing. However, after reading this manuscript I think SAFRAN does not vary precipitation amounts in one elevation/aspect category within the massif scale (i.e. 1000 km2). I am not sure if such a rough meteorological forcing is helpful to establish differences between a high resolution distributed model (250 m) and a semi-distributed model. I would suggest that a numerical weather model with a fine resolution (e.g. AROME with a 2.5 km resolution, Queno et al., 2016; or Vionnet et al., 2016) may be used as well for a few available years to show more clearly the potential of high-resolution modelling.

Author’s response (A): The SAFRAN meteorological reanalysis is used in this manuscript because it is still the best meteorological dataset available in the French Alps to drive spatialized snowpack simulations over a long period. The potential of the AROME Numerical Weather Prediction (NWP) model has already been investigated in other papers (Quéno et al., 2016; Vionnet et al., 2016). The spatial pattern of precipitation is more accurate with the kilometric resolution of AROME than with the massif scale of SAFRAN. However, some significant biases occur in AROME since the atmospheric forcing is only made of successive forecasts. No precipitation observations are assimilated to correct the precipitation amount, contrary to SAFRAN. As a result, both papers exhibit a better skill of snowpack simulations in terms of SWE and snow depth when SAFRAN is used as forcing. In the case of the Mont-Blanc massif, unpublished results (Vionnet et al, in prep.) were obtained for the “Mer de Glace” glacier for winter 2011/2012 and are provided here. Figure 1 shows the glacier winter surface mass balance (SMB) simulated with Crocus and driven by three meteorological forcing: AROME at 2.5 km downscaled to 0.5 km (green), an experimental version of AROME at 0.5 km resolution (blue) and SAFRAN spatially distributed at 0.5 km resolution (red). This figure also includes winter SMB observations (black dots) and their associated uncertainty. At high elevations where there is no observation assimilated by SAFRAN, the winter SMB is underestimated with SAFRAN and is more realistic with AROME. However, at lower elevations ($\sim$ below 2500 m) the winter SMB is highly overestimated with AROME and more realistic with SAFRAN thanks to the assimilation of observations. Therefore, we decided to use the atmospheric forcing from SAFRAN in our study since AROME still suffers from limitations in the Mont Blanc area. One of the goals of this paper is also to study the impact of shadowing in high spatial resolution snowpack simulations regardless the future possible availability of a higher resolution meteorological forcing. Therefore, it makes sense to perform this analysis with the same meteorological dataset in both simulations and to choose the best and longest dataset currently available for that purpose.

Figure 1 caption (complete): Glacier Winter Surface Mass Balance observed and simulated for the "Mer de Glace" in 2011-2012 snow season. SAF correspond to Crocus simulations with SAFRAN meteorological forcing, ARO_0p5 to and ARO_2p5D to Crocus simulations driven respectively by an experimental version of AROME at 0.5 km and AROME 2.5 km downscaled at 0.5 km and Observations correspond to punctual observations. For each mass balance a linear tendency line has been included.

Obviously, taking benefit from high resolution NWP models in snowpack simulations is still a major research perspective. This still requires the development of a high-resolution distributed analysis combining observations and AROME forecast and the development of downscaling methods to fill the gap between their kilometric resolution and the 250 m resolution which is necessary for an accurate representation of slopes in alpine environments. Please refer to response to the comment 2.3 for the corresponding changes in the manuscript.

R1: I am a bit surprised that an author who measured snow depth in alpine terrain in high resolution suggests a model approach that defines SCA to be zero, when modelled snow depth in a pixel is below 0.15 m, and one otherwise. 15 cm mean snow depth is not sufficient to cover large parts of a 250 m pixel in complex terrain. This may be true after a new snow fall, but not when a seasonal snow cover is melting out. While this is a common model approach, there were new SCA parameterisations developed in the past years to account for heterogeneous snow depth distribution based on terrain characteristics (e.g. Helbig et al., 2015; Cristea et al., 2017). This model evaluation has a large emphasis on SCA. In such a case I think a state-of-the-art SCA parameterisation is necessary. This is also necessary for a 250 m resolution grid since snowdepth varies largely at smaller scales (e.g. Trujillo et al., 2007). To my opinion, the benefit of terrain shading in the distributed approach may be overridden by these model deficiency.

A: We agree that considering 0.15 m as the threshold for considering a simulated pixel as snow covered or not, may be disputable if inferring full snow cover in the pixel, especially considering the high spatial heterogeneity of snowpack in mountain areas. This is also true for the choice of threshold used to convert SCA derived from MODIS observations to a binary map for use of the selected metrics. However, most of the works which studied the snow covered area over large domains use a threshold for considering a pixel as snow covered or not , regardless the spatial scale(e.g. Cristea et al., 2017 at the very high resolution of 3 m, Gascoin et al., 2015 at the resolution of 500 m). Furthermore, we demonstrate in our paper that the strict validity of this relationship is secondary in the context of the evaluations performed in this paper. Indeed, we calculated the evaluation metrics for several SD and SCA threshold and the results are now included in Table 2 in the revised manuscript. The calculations were obtained for both spatialization approaches. This table shows that the evaluation metrics are only slightly sensitive to the choice of these thresholds and that for every threshold and metrics the ranking of the two approaches remains the same. The selected thresholds are those leading to the best scores. Note also that the snow depth threshold of 15 cm is consistent with other studies (e.g., Figure 4 in Gascoin et al. 2015). Testing fractional SCA parametrizations hypotheses would require higher spatial resolution simulation, which could equally be argued to involve a higher uncertainty in relation to the approach involved for spatial distributing (e.g. precipitation quantities must be changed depending on the topographic characteristics). Moreover fractional SCA area hypotheses would also require a higher spatial resolution database to test the results regarding observations of snow presence/absence (Cistera et al., 2017) or snow depth (Helbig et al., 2015), commonly derived from LiDAR observations. Additionally, testing higher spatial resolution snowpack simulations also requires significantly more computational time. An increase on the spatial resolution from 250 m to about 20-30 m pixel size would be the only alternative that may potentially reproduce a realistic snow distribution in mountain terrain (Deems et al., 2006, Trujillo et al., 2007, Revuelto et al., 2014) which is beyond the scope of the present study. The following changes were consequently performed in the manuscript (please also check manuscript with tracked changes for smaller changes). Lines 565-573, “Table 2 shows the SCA simulation results estimated based on 0.1, 0.15 and 0.2 m snow depth thresholds compared with the various UWS thresholds tested, for the 2008–09 and 2009–10 snow seasons (average snow accumulations) and for both spatialization approaches. This table shows that the evaluation metrics are only slightly sensitive to the choice of these thresholds and that for every threshold and metrics the ranking of the two approaches remains the same. In light of the sensibility test results we selected a 0.15 m snow depth threshold for the simulations and 0.35 SCA threshold for MODImLab UWS product for classifying a pixel as snow-covered. Lines 744-757: “Crocus simulates the energy and mass exchanges with soil and atmosphere and also within the snowpack layers, but it does not simulate small scale topographic effects on snow depth distribution (Revuelto et al., 2016a). Given the fact that the final objective of this study is to compare two simulation approaches and that one of them would not allow an appropriate parametrization of topographic control on snow distribution (semi-distributed approach), we have not considered novel approaches for distributing snow based on terrain parameters (Cristera et al., 2017, Helbig et al., 2015), which may also require a higher spatial resolution for accounting topographic effect on snow distribution (Deems et al., 2006, Trujillo et al., 2007). Hence, we decided to simulate snowpack evolution with a spatial scale in which satellite observations were available over a long time period with a suitable temporal resolution, which lead to select 250m spatial resolution simulations (same as MODImLab products). Moreover this spatial resolution provides an appropriate representation of slopes for future applications forecasting snow avalanches with expert systems (MEPRA, Lafaysse et al., 2013)."

R1: 2.2 Validation The authors chose a quite subjective definition what a good model performance is. This lead to formulations as "appeared reliable" (line 439), "Overall, the ability [: : :] was satisfactory" (line 568ff). For many measures it is not clear to me if a difference in model performance is substantial or negligible. Here I would suggest a more objective procedure integrating a baseline model or quantifying error metrics in relation to year to year differences, for example.

A: We agree with the reviewer that the comparison of scores was not sufficiently based on objective criteria in the initial manuscript. In order to provide a more objective quantification of error metrics (RMSE, MAE and R2) a 100-member bootstrap obtained by random sampling with replacement of the different years of observations (14 for the SCA evaluation and 20 for the SMB evaluation) was generated. This bootstrap allows us assessing the uncertainty of each score. Thus, we can compare the scores of the semi-distributed and distributed simulations by a classical t-student test: the associated p-value obtained from the bootstrapped sample allows us to accept or reject the null hypothesis (i.e. equality of scores). Tables 3 and 8, which previously showed only error metrics averages for the whole period, now show the average value and the standard deviation of RMSE, MAE and R2 obtained from the bootstrapped sample. Error metrics in bold note p-values lower than 0.01 (99% confidence interval for rejecting null hypothesis). The standard deviations of the RMSE and the MAE for the snow cover area are lower than the difference between the scores of both approaches. As a result, the p-value allows us to reject the null hypothesis with a 99% of confidence interval and that the skills of distributed and semi-distributed simulations are not statistically equivalent. Conversely, the standard deviation of the R2 for the snow cover area is high compared to the difference between the scores of the two spatial discretizations. As a result, the high p-value indicated that the null hypothesis should be accepted and that these scores are not statistically different between both approaches. More generally, when looking at all scores for SCA and SMB, we can conclude objectively from this new analysis that there is a slight but significant added value of the distributed discretization. Finally, formulations as these noted by the reviewer have been removed in final manuscript version. Please check the manuscript with tracked changed to see how are described the bootstrapping and the t-student test. Result section now includes the following lines describing the results obtained: Lines 602-619: "Error estimates for the SCA simulated for the whole study site and for north and south aspects (Tables 3, 4 and 5) were lower for the distributed simulations compared with the satellite observations. RMSE and MAE standard deviations obtained from the bootstrapping (Table 3) are lower than the difference between scores for both approaches. The p-values for these two error metrics are lower than 0.01 and thus the null hypothesis is rejected with a 99% confidence interval and the skills of distributed and semi-distributed simulations are not statistically equivalent. Conversely the R2 standard deviation of the SCA is high compared to the difference between the scores of both approaches. As a result, the high p-value indicated (in this case above 0.05) that the null hypothesis should be accepted and that these scores are not statistically different between both approaches. R2, MAE and RMSE average values for high (2006-2008 snow seasons, Table 4) and low (2011-2013 snow seasons, Table 5) levels of snow accumulation also show the better capacity of distributed simulations to reproduce SCA evolution. The t-student test has demonstrated that RMSE and MAE results for both approaches are not statistically equivalent and that for all aspects and periods, the distributed simulations presents lower errors. We can conclude that this latter approach significantly better reproduce the SCA evolution. The differences in the error metrics (RMSE and MAE) between distributed and semi-distributed simulations are significant for both, north and south aspects but higher for north aspect. However, it must be highlighted that for the whole catchment and for any aspect, the null hypothesis can be accepted based on the R2 value between distributed and semi-distributed approaches. This means that the added value of the distributed approach is not visible on this criterion." Lines 703-710: "Table 8 shows RMSE, MAE and R2 means and standard deviations obtained from the 100-member bootstrap sample. For most of the error metrics, the standard deviations are lower than score differences and the p-values are low enough to reject null hypothesis. This way results obtained with both simulation approaches are statistically different. In winter, the SMB simulations show similar results. For both glaciers, lower RMSE and MAE are obtained for distributed simulations and better R2 for semi-distributed simulations. Oppositely during summer all scores show better results for the distributed approach. The annual SMB also exhibit better results for distributed simulations."

R1: 2.3 Differences and similarities to previous papers Last year this group of authors published two papers validating a SAFRAN/AROMECrocus model (Queno et al., 2016; Vionnet et al., 2016). I would like to see more clearly what the advance of this manuscript is and how findings in those papers compare to this study (e.g. effects of meteorological forcing using SAFRAN). Some differences are obvious (250 m vs. 2.5 km resolution), long-term dataset with a spatial focus, but should be clearly mentioned in the Introduction.

A: Following the reviewer recommendation, we included a new paragraph in the introduction. This paragraph describes the main characteristics of the simulations obtained in Vionnet et al., 2016 and Queno et al., 2016. Additionally, the main objectives of this work are contextualized within the findings of these two works. The next lines have been included in the introduction (lines 147-180): "Recent studies have assessed the impact of high-resolution atmospheric forcing from the Numerical Weather Prediction system AROME (Seity et al., 2010) on distributed snowpack simulations with Crocus. Queno et al., (2016) and Vionnet et al., (2016) compared simulations at a 2.5 km spatial resolution forced by AROME forecasts or by SAFRAN reanalysis (Durand et al., 2009). These works demonstrated that the geographical patterns simulated by the AROME-Crocus model chain are realistic and more detailed than the SAFRAN-Crocus model chain over large areas (the Pyrenees and French Alps). Nevertheless these studies also exhibit some significant biases in meteorological and snow variables with the AROME-Crocus chain which do not assimilate any meteorological observation, in particular precipitation. As a result, Queno et al., (2016) and Vionnet et al., (2016) exhibit a better skill of snowpack simulations when SAFRAN is used as forcing. They conclude that the potential of the high spatial resolution atmospheric forcing from the NWP system will be more beneficial in snowpack simulations with the development of a high-resolution distributed analysis combining observations and AROME forecast and the development of downscaling methods to fill the gap between their kilometric resolution and the resolution required to capture the variability of slopes and aspects in alpine environments. Moreover the impact of topographic effects on snowpack simulations (implemented in the snowpack model) has not yet been assessed in detail. At present, the implementation of terrain shadowing effects on Crocus snowpack model (achieved on distributed simulations) has not been analyzed in complex alpine terrain. It is therefore necessary to compare distributed and semi-distributed snowpack simulations with a spatial resolution that enables a detailed representation of alpine terrain."

R1: 2.4 Communicating and presenting results I had large difficulties with understanding this manuscript. Statements in the text were sometimes not documented in the presented Figures and Tables. For example, I do not see in Figure 6 a higher consistency between observations and simulations in the winter in relation to the summer, which is stated in line 470. In Table 6 and 7 the ASSD values are higher for the distributed approach, but oppositely stated in line 506. Furthermore, I do not see that the distributed version is simulating the Surface mass balance in the winter similarly well compared to the semi-distributed version (lines 757ff), while in Table 8 the RMSE for Mer-de-Glace and WSMB is much worse.

A: The way of communicating results was clarified in the revised paper. Moreover the new statistical framework included comparing the scores of the distributed and semi-distributed simulations of SCA and SMB allows obtaining more robust conclusion. The last remark about previous version of Table 8 regarding the WSMB was due to an error on the values previously shown. Now this table presents the bootstrapping averages and standard deviation of error metrics for both approaches, for Mdg the difference between the semi-distributed and the distributed approaches for WSMB is smaller than for Arg Moreover the MAE of both approaches for the WSMB in Mer de Glace is not significantly different. Thus the statement has been deleted in the revised manuscript.

R1: 3 Technical comments The following comments show that some methods are not described in a sufficient way in order to be able to replicate results. Lines 231ff: Please add more details on how the coarse SAFRAN categories were interpolated to the fine 250 m grid. The given citation (Vionnet et al., 2016) does not describe the procedure, but refers to another citation. Since energy balance models are quite sensitive to input data, the chosen interpolation methods should be presented here to understand the differences between the distributed and semi-distributed model approach.

A: We included the following complementary information to describe the interpolation technique, (lines 309-318): "As SAFRAN reanalysis provides semi-distributed outputs, the meteorological forcing at hourly time steps was spatially distributed over the 250-m grid DEM using specific routines that accounted for the elevation and aspect of each grid cell. For each cell of the 250-m grid, the spatialization of meteorological variables from the 300-m elevation bands of SAFRAN is based on a linear interpolation between the two closest elevation bands. Only one SAFRAN aspect class is considered for each pixel (nearest-neighbour technique for the aspect) (Vionnet et al., 2016). Therefore, the meteorological input data are similar for all simulations: only minor differences occur because elevation differences (< 300 m) may impact meteorological forcing variables.

R1: Lines 291ff: Please add more details to the conclusion, why a full evaluation of the simulation is possible with the chosen validation data. What is a full evaluation and how do you apply Hanzer et al. (2016) on your dataset?

A: We base our statement about the full evaluation of simulations on the representation Hanzer et al., (2016) introduced and the database they used on their evaluation. From the eight datasets they considered we used four of them, which combined showed a validation dataset quite close to their evaluation. This issue has been clarified in the text as follows (lines 368-377): "To evaluate the simulations in this study we used four datasets based on: in situ snow depth from Météo-France stations; the snow covered area (SCA) from MODIS images; the punctual glacier surface mass balance (SMB); and the glacier equilibrium-line altitude (ELA) from Landsat/SPOT/ASTER. Based on the radar charts presented by Hanzer et al. (2016), shown in their Figure 5, the information available for our study matches four of the datasets exploited in their study (Snow Depth, MODIS, Landsat and Glacier mass balance). These four datasets cover almost the full radar chart space ("optimal" validation dataset), thus providing almost a full evaluation of the simulation performance."

R1: Line 325: I cannot find the name MODImLab in the two references. Is this a software built on these publications? Please clarify.

A: MODImLab software is based on the research of these two publications. This is now stated in the text. The reference to MODImLab user manual has also been added in the manuscript.

R1: Line 329: I cannot find the term unmixing_wholesnow (UWS) in Charrois et al. (2013). Please clarify.

A: We have clarified that the UWS product we present from MODImLab user's manual corresponds to the linear unmixing technique introduced in Charrois et al., (2013)

References:

Charrois, L., Dumont, M., Sirguey, P., Morin, S., Lafaysse, M., and Karbou, F. (2013). Comparing different MODIS snow products with distributed distributed simulation of the snowpack in the French Alps. Proceedings of the International Snow Science Workshop Grenoble – Chamonix Mont-Blanc - 2013 (Grenoble, France), 937-941

Cristea, N. C., Breckheimer, I., Raleigh, M. S., HilleRisLambers, J., and Lundquist, J. D. (2017). An evaluation of terrainâĔŸAREĔĞ based downscaling of fractional snow covered area datasets based on Lidar-derived snow data and orthoimagery. Water Resources Research, 53, 6802–6820,

Durand, Y., Giraud, G., Brun, E., Merindol, L., and Martin, E. (1999). A computer-based system simulating snowpack structures as a tool for regional avalanche forecasting. J. Glaciol. 45, 469–484.

Durand, Y., Laternser, M., Giraud, G., Etchevers, P., Lesaffre, B., and Mérindol, L. (2009). Reanalysis of 44 Yr of Climate in the French Alps (1958–2002): Methodology, Model Validation, Climatology, and Trends for Air Temperature and Precipitation. J. Appl. Meteorol. Climatol. 48, 429–449.

Deems, J. S., Fassnacht, S. R., & Elder, K. J. (2006). Fractal distribution of snow depth from LiDAR data. Journal of Hydrometeorology, 7(2), 285-297.

Gascoin, S., Hagolle, O., Huc, M., Jarlan, L., Dejoux, J.-F., Szczypta, C., Marti, R., and Sánchez, R. (2015). A snow cover climatology for the Pyrenees from MODIS snow products. Hydrol Earth Syst Sci 19, 2337–2351.

Hanzer, F., Helfricht, K., Marke, T., and Strasser, U. (2016). Multilevel spatiotemporal validation of snow/ice mass balance and runoff modeling in glacierized catchments. The Cryosphere 10, 1859–1881.

Helbig, N., van Herwijnen, A., Magnusson, J., and Jonas, T. (2015). Fractional snow-covered áreas parametrization over complex topography. Hydrology and Earth System Sciences, 19(3), 1339-1351.

Lafaysse, M., Morin, S., Coléou, C., Vernay, M., Serça, D., Besson, F., Willemet, J.M., Giraud, G., and Durand, Y. (2013). Towards a new chain of models for avalanche hazard forecasting in French mountain ranges, including low altitude mountains. Int. Snow Sci. Workshop Grenoble-Chamonix Mont-Blanc.

Quéno, L., Vionnet, V., Dombrowski-Etchevers, I., Lafaysse, M., Dumont, M., and Karbou, F. (2016). Snowpack modelling in the Pyrenees driven by kilometric-resolution meteorological forecasts. The Cryosphere 10, 1571–1589.

Revuelto, J., López-Moreno, J.I., Azorin-Molina, C., and Vicente-Serrano, S.M. (2014). Topographic control of snowpack distribution in a small catchment in the central Spanish Pyrenees: Intra- and inter-annual persistence. Cryosphere 8, 1989–2006.

Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C., and Masson, V. (2010). The AROME-France Convective-Scale Operational Model. Mon. Weather Rev. 139, 976–991.

Trujillo, E., Ramírez, J. A., & Elder, K. J. (2007). Topographic, meteorologic, and canopy controls on the scaling characteristics of the spatial distribution of snow depth fields. Water Resources Research, 43(7).

Vionnet, V., Dombrowski-Etchevers, I., Lafaysse, M., Quéno, L., Seity, Y., and Bazile, E. (2016). Numerical Weather Forecasts at Kilometer Scale in the French Alps: Evaluation and Application for Snowpack Modeling. J. Hydrometeorol. 17, 2591–2614.

Vionnet, V., Six, D., Auger, A., Lafaysse, M., Queno, L., Reveillet, M., Dombrowski-Etchevers, I., Thibert, E. and Dumont, M. (in prep). Influence of precipitation datasets on sub-kilometric simulations of snowpack and glacier winter mass balance in alpine terrain.

Interactive comment on The Cryosphere Discuss., https://doi.org/10.5194/tc-2017-184, 2017.

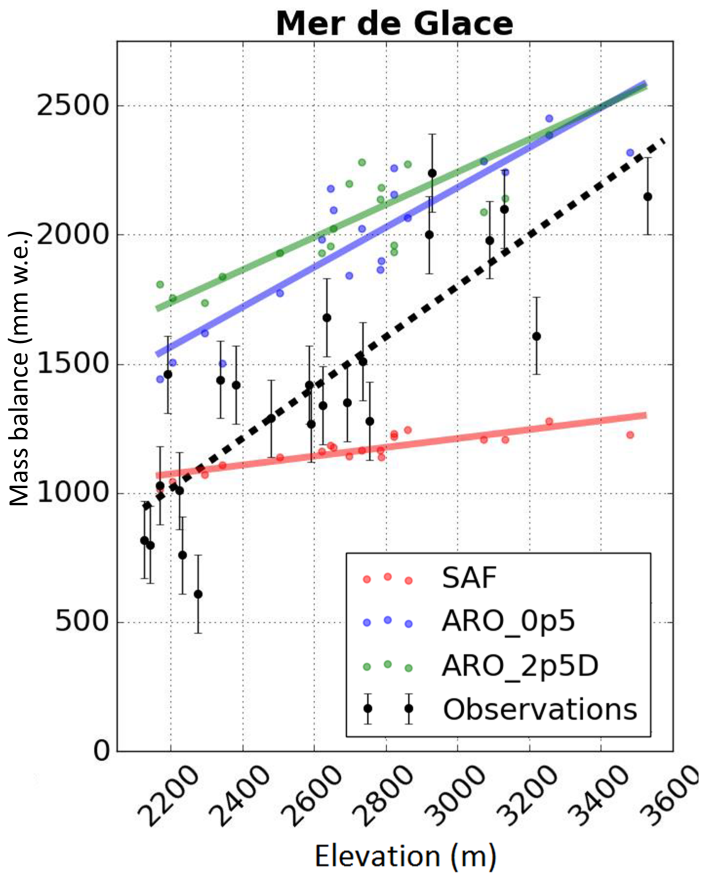

Fig. 1. Glacier Winter Surface Mass Balance observed and simulated for the "Mer de Glace" in 2011-2012 snow season. SAF correspond to Crocus simulations with SAFRAN meteorological forcing, ARO_0p5 to and ARO_