Response to Referee 1

We thank R1 for this detailed review, which enabled us to significantly improve our article. Enclosed please find a detailed explanation of the revisions we made based on R1's comments. For convenience, comments are in bold and our responses are in italic. Revisions made in the manuscript are presented in italic with grey background.

General comments: This paper describes a new sophisticated method to adjust and disaggregate daily RCM output to hourly values, which are usually necessary to force energy-balanced based land surface models. Such a method is an interesting and useful addition to the field. The work is therefore relevant, although the applied reanalysis data set used as reference is very specific and the performance of the method with other observational datasets first needs to be demonstrated. After the authors investigated the impact of the grid point selection, the "ultimate quantile mapping" and the transferability in time, it would also be interesting to know the impact of the weather regime consideration or not. This is just a wish and since the paper is already long enough I understand if the authors want to cut this point. Therefore, I recommend publishing the paper once the authors addressed at least the points listed below:

We thank the reviewer for this review, please see our specific responses to each point below.

*Concerning the remark: « **it would also be interesting to know the impact of the weather regime consideration or not.** », we also think this would be interesting. We haven't looked deeply into this, but as shown in Driouech et al. (2010), the frequency of weather regimes changes in a warmer climate, contributing significantly to the change in precipitation. If we adjust a model irrespective of the regimes, the adjustment may result from a compensation between regime-dependent systematic errors. It is therefore wiser, if sampling permits, to correct this error for each regime separately. As a result, the conditional systematic error for each regime for present climate is, by construction, zero. But since the model regime frequencies are not exactly the same as the observed ones during the training period, the full-sample systematic error is not zero. Our method is thus a compromise : we slightly degrade present climate in oder to expect less bias in future climate. However, in order to keep the manuscript as short as possible, we did not develop this point further.*

Specific comments:

L80: ADAMONT stands for what?

ADAMONT is the name of one of the projects which funded this study, which was then used to name the method. We thus prefer to not give any sense to the name chosen.

L123: daily RCM model outputs

This was included (Line 126).

L143: 4 weather regimes: Can they be named or described somehow? If I understand right, this means that every day in the reference period has been categorized in one of the four weather types, which is valid for all massifs for this day? Is this already implemented in ADAMONT for Europe as a kind of look-up table? The 4 weather regimes are based on quite old study. What about the consideration of 5 weather regimes as proposed in a recent study (http://onlinelibrary.wiley.com/doi/10.1002/2017GL074188/epdf) ?

There are 4 different regimes defined for each season and it is now explained (Line 144-147) : « RCM weather regimes were determined based on the synoptic fields of the GCM model used as boundary condition for the RCM. In Michelangeli et al. (1995) and Driouech et al. (2010), only regimes for the winter season are defined. We chose to apply the same method to determine weather regimes for the other seasons as well.» For the winter season, regimes have been given a traditional name (the number is arbitrary): 1 = Zonal, 2 = Atlantic Ridge, 3 = Blocking, 4 = Greenland Anticyclone. For other seasons, they don't have any name, we applied the same algorithm as for winter.
Indeed, every day in the reference period has been categorized into one of the four weather types for each season, which is valid for all massifs for this day. This is now clarified (Line 142-144) : « The ERA-Interim reanalysis (Dee and Uppala, 2009) was used to infer weather regimes corresponding to each observation date and for all observation points. ».
Our choice of regimes is based on the study of Michelangeli et al. (1995), which has served as basis for other studies such as Driouech et al. (2010). Moreover, the initial method has been updated here by computing weather regimes for the different seasons, and by computing the clusters based on ERA-40 (as in previous studies), but using ERA-Interim to infer the weather regime corresponding to each observation date and for all observation points. We could use more regimes, but this would endanger the robustness of our results, because of the too limited number of data used to infer quantile values if too many weather regimes are considered, some corresponding to a small number of days. Four regimes is found to be satisfying for this study, as it ensures a sufficiently large size of the datasets for quantile mapping. We have rephrased the end of the paragraph (Line 149-153) : « This number is a compromise between accuracy of the correction and robustness of the percentile estimation (more regimes can be used, such as in Ummenhofer et al. (2017)). On the other hand this relatively small number of regimes ensures a sufficiently large size of the datasets used for quantile mapping (which are, as described below, further partitioned into 4 seasons DJF, MAM, JJA, SON). »
A new Figure was added to represent the different regimes used (Line 153-154 & Fig. 1) : « Figure 1 represents the different regimes used in this study. »

L146: Integration: Would "Aggregation" not be the better term?

Yes, this would be better. We changed the word « integration » to « aggregation » and « integrated » to « aggregated » (Line 155 and caption of Table 1).

L148: 6 am to 6 am the next day: Is this UTC or local time?

This is UTC. We have now changed the sentence to introduce this precision (Line 157): « (from 6:00 UTC to 6:00 UTC the next day) ».

L150: daily mean: 6 am to 6 am?

No, this is UTC. This was included (Line 159-161).

L152: Ok, you calculate the 99 percentiles, but what do you mean with "99 percentiles + 0.5 % and 99.5 % quantiles" ?

The percentiles (1 %, 2 %, …, 99%) are calculated, and we also calculate the 0.5 % quantile and the 99.5 % quantile. This sentence was changed to (Line 162-163): « The quantile values (the 99 percentile values as well as the 0.5 % and 99.5 % quantile values)... »

L160: For RCM values greater than the 99.5 % quantile, a constant adjustment based on the value of this last quantile is applied in order to allow for new extremes.

This precision was added (Line 170-172).

L170: A further criterion can be applied: Did you apply it or not?

Yes, this sentence was corrected (Line 181) : « A further criterion is applied... ».

L174: a random draw: This in contradiction to the desired "consecutive time slices" described above!

R1 is right. This point was not clear enough. We have rephrased the paragraph as follows (Line 185-191) : « For the first RCM date, a random draw amongst all available observational dates is performed, then the dates are browsed through chronologically until one meets all the requirements outlined above. This analogous day is then used in the following step for all variables. If the following analogue day in the observations still meets all requirements, it is selected as analogue for the following day in the RCM (to ensure as far as possible consecutive time slices). A new random draw is only performed once the analogue fails to meet all requirements described above. »

L175: browsed through: in which direction?

They are browsed through chronologically. This is now specified (Line 186) : « (…) then the dates are browsed through chronologically until one meets all the requirements outlined above. »

L186: RCM adjusted daily minimum and maximum: It should be mentioned before, that RCM often provide daily minimum and maximum temperatures.

This is now mentioned in point n°3 (Line 156-158) : « for temperature, the daily minimum and maximum values (from 6:00 UTC to 6:00 UTC the next day) are selected (RCMs generally offer daily minimum and maximum temperature). »

L201: Equation 4: T_h_RCM(24h, i-1) is not available for the first day. What to take then?

We thank R1 for this remark. Indeed, this is a point that we did not describe in detail in the article. This is now included (Line 218-224) : « For specific cases, i.e. for the first day where T_{RCM} (24h, i − 1) *does not exist, or if the determinant of our system is too close to zero (< 0.1), or in the case where a < 0, a much simpler equation is used, in which we only ensure that final mimimum and maximum daily values correspond to the RCM adjusted minimum and maximum values, by solving:*

$a = (Tmax_{RCM}^{d,adj}(i) – Tmin_{RCM}^{d,adj}(i)) / ((Tmax_{OBS}^{h} – Tmin_{OBS}^{h})$
$b = Tmax_{RCM}^{d,adj}(i) – a\ Tmax_{OBS}^{h}$*. »*

L206-209: X_h_SAF should be replaced with X_h_OBS!

It is now corrected (Line 229-230).

L231: Definition of snow year is missing!

This is now defined (Line 251-253) : « The resulting adjusted hourly time series for each variable are obtained for each snow year (from the 1st of August to the 31st of July of the following year) ».

L260: all massifs: Should it not be one massif, since the calculation is done by massif?

No, this criterion on the wet/dry analogue days is applied to all massifs in order to « maximise the consistency between massifs after the adjustment process », as indicated in the text. Please keep in mind, however, that this criterion is a second order criterion for the selection of analogous days, the first order criterions being the month of the year, the weather regime and whenever possible, consecutive time slices for consecutive RCM dates.

L282: Replace Method with ADAMONT

The section title was changed to « ADAMONT method evaluation » (Line 306).

L300: I guess a given altitude level means a +/-150 m wide elevation band?

We have added a clearer description of elevation bands earlier in the article (Line 267-269) : «SAFRAN data are available for elevation bands with a resolution of 300 m, i.e. altitude levels 600, 900, 1200, 1500 m etc. are typically considered, making it possible to extract meteorological information at these altitude levels, or in-between using altitude interpolation. »

L304-372: References to the corresponding tables and figures would help a lot.

The reviewer is correct that it could ease reading, but given that Tables and Figures are introduced in the Results section, referring to them earlier in the manuscript would recquire major changes to their description. Indeed, this would alter Figures and Tables order, and lead to the need to introduce them fully in the Methods sections, before the introduction of their detailed content, which is a problem too. Given these condiserations, we chose to not introduce the Tables and Figures there.

L391-393: Please mention that the good agreement for snow depth is due to the fact that the difference in winter precipitation is small (see Fig. 5)!

This is now indicated (Line 441-444) : « Smaller autumn and winter precipitation biases lead to a good agreement between the magnitude of average snow depth in the different adjusted RCM simulations and the results obtained using the reanalysis as meteorological input (as noted in Fig. 4). »

L403 & L412: Are there no noteworthy differences between massifs?

The large biases and RMSEs values obtained when using raw RCM simulations compared to adjusted simulations are features common to all massifs. It is now more clearly indicated (Line 426-428) : « This highlights the large biases and RMSEs values obtained when using raw RCM simulations compared to adjusted simulations, a feature common to all massifs (Figs. 5-6 and Supplementary Information)». So is the fact that the longer learning period 1980-2010 generally presents smaller biases and RMSEs. The word « generally » in this sense encompasses the analysis across all massifs.

L415: smaller than 150 kg m−2 per month: This should also be expressed in percentage!

OK. This information was added (Line 439-441) : « Biases of the adjusted simulations remain smaller than 150 kg m $^{-2}$ per month in absolute value, corresponding to up to 90% depending on the massif and altitude »

L422: biases never exceed 50 cm: This should also be expressed in percentage!

This was added (Line 450-451) : « For snow depth, the biases never exceed 50 cm, which corresponds to up to 50% depending on the altitude and the massif »

L430: Fig. 5 & 6

Indeed (it is now Figs. 6-7). This was corrected (Line 460-461).

L533: as found by Lafaysse (2011)

Done (Line 562).

L551: TSS are generally better for massifs of the Northern Alps: Could you please provide some percentage range!

This information was added (Line 580-582) : « TSS are generally better for massifs of the Northern Alps (0.25 to 0.6) than the Southern Alps (0.1 to 0.4, Supplementary Information) ».

L564: Why not Figs. 10-13?

Yes (now Figs. 11-14), this error was corrected (Line 593).

L575-577: Please give a reference for this statement!

This was added (Line 603-604) : « (…) when most observations from mountain stations are not available (Gobiet et al., 2015) ».

L602: biases for precipitation

This was included (Line 630) .

L622-623: The new method ADAMONT is able to statistically adjust daily regional climate model projections and to provide hourly. . .

This sentence was rephrased accordingly (Line 651-653).

L647: ultimate quantile mapping: Should be again explained in more detail for the conclusion section.

We thank R1 for this suggestion. This was added (Line 676-679) : « the ultimate quantile mapping applied to snowfall and rainfall (i.e., after a first quantile mapping on total precipitation, an additional quantile mapping against the observational dataset is applied for daily cumulated adjusted RCM rainfall and snowfall separately) »

Table 2: For a better understanding, the configuration with N=0 should be also labeled as such.

OK. This has now been included in Table 2, Figs 4-16 and in the Supplementary Information.

Figure 1: An additional small map with numbered massifs (e.g. right of the elevation color bar) would give the reader a possibility to geographically locate the massifs listed in Table 3, where the same number needs to be inserted.

This is now included in Fig. 3 and Table 3.

Figure 3 (top left): Why is the 1800 m elevation not considered?

It is considered, but it corresponds to the same ALADIN RCM grid point as for 1500 m, so we don't see the line corresponding to 1800 m. We included the following sentence in the caption of Fig. 4 to indicate this : « In this case the 1500 m and 1800 m lines are similar. »

Figure 3 (top right): I guess the time period of the SAFRAN reanalysis is 1980-2010. Please give this information in the legend or in the figure caption.

This is now included in the figure caption (Fig. 4).

Figure 3 (caption L3): different elevations considered (900-2400 m. . .)

Done (Fig. 4).

Figure 8: I guess the time period of the SAFRAN reanalysis is 1980-2010. Please give this information in the legend or in the figure caption.

This is now included in the figure caption (Fig. 9).

Figure 10: Scale of the y-axis for the two elevations should be the same for comparability. The y-axis labeling in the 2. and the 4. column is missing. Should be like Figure 11.

We are afraid that if we use the same scale for the y-axis for 1200 m and 2100 m, some curves for 1200 m won't be readable anymore. The values at 1200 m and 2100 m can be very different, especially for precipitation and even more for snow depth. This is why we would prefer to keep the scales for the y-axis as is. However, we have changed the labeling of the y-axis as proposed by R1 (Fig. 11).

Response to Referee 2

We thank R2 for this helpful review. Enclosed please find a detailed explanation of the revisions we made based on R2's comments. For convenience, comments are in bold and our responses are in italic. Revisions made in the manuscript are presented in italic with grey background..

Review report for manuscript "The method ADAMONT v1.0 for statistical adjustment of climate projections applicable to energy balance land surface models" by Verfaillie et al. (2017) This study introduces the method ADAMONT v1.0 to adjust and disaggregate daily climate projections from a regional climate model against an observational dataset at hourly time resolution. The method makes use of a refined quantile mapping approach for statistical adjustment and an analogous method for sub-daily disaggregation. The method is capable of producing adjusted hourly time series of temperature, precipitation, wind speed, humidity, and short- and longwave radiation, which can in turn be used to drive any energy balance land surface model (e.g. a fully distributed energy and water balance hydrologic model). The observational dataset used here is the SAFRAN meteorological reanalysis, which covers the entire French Alps split into 23 massifs, within which meteorological conditions are provided for several 300 m elevation bands. In order to evaluate the skills of the method itself, it is applied to the ALADIN-Climate v5 RCM using the ERA-Interim reanalysis as boundary conditions, for the time period from 1980 to 2010. The authors find the disaggregation method to preserve inter-variable dependency structures although it performed well for temperature compared to precipitation. The manuscript is well organized and the analyses methods are well thought out, except a few points. Please find below a few comments which could help you to improve your manuscript on the way to publication.

We thank the reviewer for this review, please see our specific responses to each point below.

Major comment: Line 1 – 64: The authors introduce the need for bias-correction of RCM outputs but completely fail to address the many flaws of bias-adjustment which have been well detailed in Ehret at al 2012: "Should we apply bias correction to global and regional climate model data?" Most impact studies are now utilizing convection permitting models at <4km resolution to overcome some of these limitations. Also, the authors have to specifically state that the results of the quantile mapping are sensitive to data sets used and adjustment method as well. Thus, there is a wide array of uncertainties associated with these kinds of studies.

The reviewer is correct that bias-correction is not a perfect solution, but it is still a necessary step when using regional climate model data for impact studies (Maraun 2016), be it convection permitting or not. In addition, while a few studies have recently emerged using non-hydrostatic high-resolution model approaches targeting summertime processes such as convection-driven events (e.g. Ban et al., 2015, Giorgi et al., 2016, https://www.hymex.org/cordexfps-convection/wiki/doku.php?id=modellist), in some areas impact studies have only marginally employed such models and most existing studies extensively rely on 10-km resolution regional climate models such as those employed in EURO-CORDEX. For example, studies addressing snow in mountainous areas have only in a few cases employed high resolution non-hydrostastic models (e.g. Musselmann et al., 2017), mostly for upstream research and process studies rather than for impact studies, which require very low biases because of the threshold effects at play in snowpack processes. We therefore believe that, even though future studies will increasingly

employ high resolution convection permitting regional climate models, many impact studies will be carried out using hydrostatic models as part of large-scale projects such as EURO-CORDEX and beyond. Furthermore, as indicated above, convection-permitting models are not immune of biases (Prein et al., 2015) and will require appropriate adjustment for being used in impact assessments.
Concerning the sensitivity of quantile mapping to the data sets used and adjustment method, we have now added the following sentence to account for this (Line 64-66) : « Furthermore, the performance level of quantile mapping methods is sensitive to the observation data set used and the detailed characteristics of their implementation, which requires specific attention. »

Ban, N., J. Schmidli and C. Schär, 2015: Heavy precipitation in a changing climate: Does short-term summer precipitation increase faster?. Geophys. Res. Lett., 42, 1165-1172.

Giorgi F., C. Torma, E. Coppola, N. Ban, C. Schär and S. Somot, 2016: Enhanced summer convective rainfall at Alpine high elevations in response to climate warming, Nature Geoscience, 9, 584-590.

Musselman, K.N, M.P. Clark, C. Liu, K. Ikeda, and R. Rasmussen, 2017: Slower snowmelt in a warmer world, Nature Climate Change, 7, 214-219.

Prein, A.F., W. Langhans, G. Fosser, A. Ferrone, N. Ban, K. Goergen, M. Keller, M. Tölle, O. Gutjahr, F. Feser, E. Brisson, S. Kollet, J. Schmidli, N.P.M. van Lipzig, and R. Leung, 2015: A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges, Rev. Geophys., 53, 323-361.

Minor Comments:

Abstract: I could not tell for which RCP(s) the adjustment was made just by reading the abstract. Please make the abstract a standalone section.

No RCP was used. In this article, we only focus on the evaluation for the recent period 1980-2010, as indicated in the abstract (Line 11-13) : « In order to evaluate the skills of the method itself, it is applied to the ALADIN-Climate v5 RCM using the ERA-Interim reanalysis as boundary conditions, for the time period from 1980 to 2010. »

What is "ADAMONT"?

ADAMONT is the name of one of the projects which funded this study. There is no meaningful definition beyond this name.

Line 145 – 160: what do you mean by integration? Just use something like "aggregation" for easy understanding. Tmax/Tmin is taken from 6am to 6am? This is not clear at all. When did you take the max and min specifically?

We thank R2 for this remark.
We changed the word « integration » to « aggregation » and « integrated » to « aggregated » (Line 155 and caption of Table 1). Maximum and minimum values are calculated from 6 am to 6 am, and only for temperature. For other variables, the daily mean (from 6 am to 6 am, this information has now been included) or the last value of each day is used.
We have slightly changed this paragraph to make it clearer (Line 156-161) : « for temperature, the daily minimum and maximum values (from 6:00 UTC to 6:00 UTC the next day) are selected (RCMs generally offer daily minimum and maximum temperature). For wind speed and humidity, the last value of each day (at 6:00 UTC) is selected (in order to be comparable to an instantaneous value), and for precipitation and radiation, the daily mean (6:00 UTC to 6:00 UTC) is used. »

Line 335: The authors should clearly state that the RMSE and mean bias were used to evaluate model performance in terms of reproducing amounts while FAR, POD etc. for occurrence.

OK, even though we don't evaluate model performances, but rather the performances of the ADAMONT method.
This is now stated (Line 359-362) : « – the root mean square error (RMSE) and the mean bias over the evaluation period, computed over seasonal integration periods based on the SAFRAN and the adjusted RCM datasets (to evaluate the method performance in terms of reproducing amounts);
– scores specific to the detection of occurrence of precipitation events (...) »

Manuscript prepared for Geosci. Model Dev.
with version 2015/04/24 7.83 Copernicus papers of the LaTeX class copernicus.cls.
Date: 18 September 2017

The method ADAMONT v1.0 for statistical adjustment of climate projections applicable to energy balance land surface models

Deborah Verfaillie[1], Michel Déqué[2], Samuel Morin[1], and Matthieu Lafaysse[1]
[1]Météo-France - CNRS, CNRM UMR 3589, Centre d'Études de la Neige, Grenoble, France
[2]Météo-France - CNRS, CNRM UMR 3589, Toulouse, France
Correspondence to: Deborah Verfaillie (deborah.verfaillie@meteo.fr)

Abstract.

We introduce the method ADAMONT v1.0 to adjust and disaggregate daily climate projections from a regional climate model against an observational dataset at hourly time resolution. The method uses a refined quantile mapping approach for statistical adjustment and an analogous method for sub-daily disaggregation. The method produces ultimately adjusted hourly time series of temperature, precipitation, wind speed, humidity, and short- and longwave radiation, which can in turn be used to force any energy balance land surface model. While the method is generic and can be employed on any appropriate observation time series, here we focus on the description and evaluation of the method in the French mountainous regions. The observational dataset used here is the SAFRAN meteorological reanalysis, which covers the entire French Alps split into 23 massifs, within which meteorological conditions are provided for several 300 m elevation bands. In order to evaluate the skills of the method itself, it is applied to the ALADIN-Climate v5 RCM using the ERA-Interim reanalysis as boundary conditions, for the time period from 1980 to 2010. Results of the ADAMONT method are compared to the SAFRAN reanalysis itself. Various evaluation criteria are used for temperature, precipitation, but also snow depth, which is computed by the SURFEX/ISBA-Crocus model using the meteorological driving data from either the adjusted RCM data, or the SAFRAN reanalysis itself. The evaluation addresses in particular the time transferability of the method (using various learning/application time periods), the impact of the RCM grid point selection procedure for each massif/altitude band configuration, and the inter-variable consistency of the adjusted meteorological data generated by the method. Results show that the performance of the method is satisfactory, with similar or even better evaluation metrics than alternative methods. However, results for air temperature are generally better than for precipitation. Results in terms of snow depth are satisfactory, which can be viewed as indicating a reasonably good inter-variable consistency of the meteorological data produced by the method. In terms of temporal transferability (evaluated over time periods of 15 years only), results depend on the learning period. In terms of RCM grid point selection technique, the use of a complex RCM grid points selection technique, taking into account horizontal but also altitudinal

proximity to SAFRAN massif centre points/altitude couples, generally degrades evaluation metrics for high altitudes, compared to a simpler grid point selection method based on horizontal distance.

1 Introduction

Projections of future climate change in terms of meteorological conditions and their impacts are requested for many scientific and societal applications (IPCC, 2013, 2014a, b, c). For a given socio-economic or greenhouse-gas concentration scenario, these projections generally concern future temperature and precipitation, and associated extreme events, and are usually generated using the outputs of global climate models (GCMs) and regional climate models (RCMs). However, GCMs and RCMs suffer from biases compared to local observations (Christensen et al., 2008; Rauscher et al., 2010; Kotlarski et al., 2014). Raw climate projections must therefore be adjusted (Déqué, 2007; Themeßl et al., 2011; Gobiet et al., 2015; Maraun, 2016), before they can be used as such (meteorological conditions), or in order to drive specific impact models. Various downscaling and adjustment methods have been developed (Maraun et al., 2010; Teutschbein and Seibert, 2012, 2013). They all require an observation dataset which (i) meets the data requirements of the application and (ii) is sufficiently long and reliable to be used to infer the relationships between the observations and the raw climate projections during the observation time period. Several approaches, such as the analog method, search for relationships between observed large-scale predictors (generally from reanalyses) and observed local-scale predictands (Vrac et al., 2007a; Dayon et al., 2015). In contrast, model output statistics approaches calibrate model outputs against observations, with various levels of complexity, such as scaling methods (linear scaling, local intensity scaling, variance scaling, ...), delta-change methods (e.g., Abegg et al., 2007; Hantel and Hirtl-Wielke, 2007; Schmucki et al., 2014) and distribution mapping methods (e.g., Boe et al., 2007; Déqué, 2007; Gobiet et al., 2015; Olsson et al., 2015). The latter include quantile mapping, which is considered as an efficient and easy to implement adjustment method (Themeßl et al., 2011; Teutschbein and Seibert, 2012; Maurer and Pierce, 2014; Gobiet et al., 2015). The main advantage of this method is that it adjusts deviations in the shape of the distribution, and is thus able to adjust deviations not only for the mean but the entire probability distribution function (Themeßl et al., 2011). Moreover, the adjustment is not strictly restricted to the range of observed values in the reference period, which is the case for example for methods based on analog weather patterns (e.g., Déqué, 2007; Themeßl et al., 2011; Rousselot et al., 2012; Dayon et al., 2015), provided that values based on the lowermost and uppermost quantiles are handled appropriately (Gobiet et al., 2015). It can thus be used for evaluation of climate extremes or projections at the end of the 21st century, as long as the probability associated with these events is robustly estimated from a long enough sample. The main limits of quantile mapping are the assumption of time-invariant model deviation to observations on which it is based, and the fact that the temporal properties of the model are not adjusted. If the model has a chronological behaviour which differs from the observations (too chaotic or too persistent), this will not be adjusted (Déqué, 2007). Moreover, quantile mapping does not guarantee the spatial and inter-variable consistency, in contrast to e.g. the analog method. Furthermore, the performance level of quantile mapping methods is sensitive to the observation data set used and the detailed characteristics of their implementation, which requires specific attention.

Climate projections in mountainous regions, which are motivated by a broad range of geophysical, environmental and societally relevant scientific challenges (Martin et al., 1994; Beniston, 1997; Jomelli et al., 2009; Castebrunet et al., 2014; Piazza et al., 2014; Schmucki et al., 2014; Lafaysse et al., 2014; Boulangeat et al., 2014; Thuiller et al., 2014; Castebrunet et al., 2014; Francois et al., 2015; Spandre et al., 2016), are particularly sensitive to the quality of the adjustment method. Indeed, regional climate model resolutions typically between 10 and 50 km are not sufficient to capture the fine-scale processes and thresholds at play. Resolving altitude dependencies is critical, especially for snow-related issues (because of the temperature dependency of the snow/rain transition). Furthermore, not only temperature and precipitation act on the snowpack, but a broader range of meteorological conditions and their diurnal variations. As a consequence, considering only adjusted daily temperature and precipitation would miss some of the non-linear response of the snowpack. Such phenomena cannot be addressed using delta-change methods, which by definition apply fixed changes to an observed time series, conserving its statistical persistence properties and seasonality (e.g., Abegg et al., 2007; Hantel and Hirtl-Wielke, 2007; Schmucki et al., 2014; Marty et al., 2017) although those could evolve significantly under changed climate conditions.

Here we introduce the ADAMONT v1.0 method, to adjust climate model projections in order to provide hourly adjusted meteorological conditions for past and future conditions based on climate model output and observational datasets. Although it could be applied for GCM output, it was primarily designed to process RCM output. Indeed, raw regional climate projection data are increasingly made available, e.g. the World Climate Research Program (WCRP) Coordinated Regional Downscaling Experiment (CORDEX, Giorgi et al. (2009)), whose aim is to improve and distribute regional climate modelling worldwide. Its European branch, EURO-CORDEX (Jacob et al., 2014), gathers regional climate simulations over Europe from 30 different modelling groups at 50 km (EUR-44) and 12.5 km (EUR-11) resolution. On the observation side, the use of surface meteorological reanalysis is a powerful alternative to station observation data to provide the necessary observational dataset (Berg et al., 2015). Indeed, the process by which such reanalyses are generated addresses the time and space variations of the meteorological conditions, and by design they consist of gap-free and complete time series. Here we describe the use of the ADAMONT method based on RCM model output comparable to EURO-CORDEX and on the mountain meteorological reanalysis SAFRAN. SAFRAN was developed specifically to address the needs of snowpack numerical simulations in mountainous regions, and contains hourly time series of temperature, precipitation, wind speed, humidity, and short- and longwave radiation for so-called massifs (ranging between 500 and

2,000 km^2 in the French Alps) by elevation steps of 300 m (Durand et al., 2009a, b). Here, quantile mapping is applied using daily outputs from a given RCM for all the variables provided in the SAFRAN reanalysis. Following a subdaily disaggregation step based on analog days selection from the reanalysis itself, these hourly adjusted fields are then used to force the SURFEX/ISBA-Crocus (Vionnet et al., 2012) model over the French Alps. We evaluate the performance of the ADAMONT method, by applying it to the ALADIN-Climate v5 RCM (Colin et al., 2010) forced by the ERA-Interim reanalysis (Dee et al., 2011) over the period 1980-2010. Sect. 2 describes the models used and the evaluation approach. Sects. 3 and 4 contain the results and their discussions, respectively, and general conclusions are drawn in Sect. 5.

2 Models and methods

2.1 Description of the ADAMONT method

ADAMONT is primarily a quantile mapping adjustment method (Déqué, 2007; Gobiet et al., 2015). In general, quantile mapping is considered one of the most efficient bias adjustment methods available (Themeßl et al., 2011; Maurer and Pierce, 2014; Gobiet et al., 2015). It consists in adjusting the quantiles of the simulated historical distribution based on the quantiles of the observed distribution. The main issues with quantile mapping relate to the assumption of time-invariant model biases, the fact that temporal properties of the RCM are untouched by the adjustment method and that the spatial and inter-variable consistency is not guaranteed. Moreover, Driouech et al. (2009) showed that for mid-latitude climates, such as that in Morocco, quantile mapping adjustment can vary for different weather regimes, because model biases vary in different regimes. Similarly, Addor et al. (2016) demonstrated the sensitivity of quantile mapping adjustment to circulation biases over the Alpine domain. Additionally, the frequency of weather regimes may change in a changing climate (Boe et al., 2006; Cattiaux et al., 2013). To improve the stationarity of our method in a changing climate, weather regimes are thus taken into account ~~in our method~~, i.e. quantile adjustment functions are computed and applied depending on the weather regime.

Assuming the availability of a gap-free meteorological observational dataset at hourly time resolution consisting of one or several geographical locations considered sharing similar large scale meteorological conditions, and daily RCM model outputs covering the geographical domain of interest, the statistical adjustment method ADAMONT consists in the following steps:

1. RCM grid point selection: For each observation point, a RCM grid point is selected, by minimizing the following distance:

$$\sqrt{(\Delta x)^2 + (\Delta y)^2 + (N \times \Delta z)^2}, \quad (1)$$

where Δx, Δy and Δz represent the longitudinal, latitudinal and vertical distances (in km) between the observation point and the RCM grid points, and N is referred to as the elevation factor. Values of 0, 50 and 100 were tested, but only results using a value of 0 and 50 (N50) are reported in this study. The factor N is a scaling factor between horizontal and vertical distances, allowing to take into account the strong dependence of meteorological variables (mainly precipitation and temperature) on altitude (e.g., Gottardi et al., 2012; Kotlarski et al., 2012).

2. Weather regime computation: Each day of the RCM and observational records are clustered into different daily weather regimes based on the geopotential height at 500 hPa, following Michelangeli et al. (1995), similar to the method described in Driouech et al. (2010). Weather regimes clusters ~~are~~ were previously computed on the basis of ~~a~~ the large scale meteorological reanalysis ERA-40 (Uppala et al., 2005). ~~consistent with the observational dataset (in our case, ERA-Interim reanalysis, Dee and Uppala, 2009), and used to infer the weather regime for each date of the RCM dataset based on the synoptic fields of the GCM model used as boundary condition for the RCM.~~ The ERA-Interim reanalysis (Dee and Uppala, 2009) was used to infer weather regimes corresponding to each observation date and for all observation points. RCM weather regimes were determined based on the synoptic field of the GCM used a boundary condition for the RCM. In Michelangeli et al. (1995) and Driouech et al. (2010), only regimes for the winter season are defined. We chose to apply the same method to determine weather regimes for the other seasons as well. A classifiability and reproducibility analysis performed by Michelangeli et al. (1995) showed that 4 weather regimes can reasonably be chosen for Europe. This number ~~also~~ is a compromise between accuracy of the correction and robustness of the percentile estimation (more regimes can be used, such as in Ummenhofer et al. (2017)). On the other hand this relatively small number of regimes ensures a sufficiently large size of the datasets used for quantile mapping (which are, as described below, further partitioned into 4 seasons DJF, MAM, JJA, SON). Figure 1 represents the different regimes used in this study.

3. ~~Integration~~ Aggregation from hourly to daily observations: The observational data are ~~integrated~~ aggregated from hourly to daily time resolution, depending on the variable considered (see Table 1) : for temperature, the daily ~~(6 am to 6 am the next day)~~ minimum and maximum values (from 6:00 UTC to 6:00 UTC the next day) are selected (RCMs generally offer daily minimum and maximum temperature). For wind speed and humidity, the last value of each day (at ~~6 am~~ 6:00 UTC) is selected (in order to be comparable to an instantaneous value), and for precipitation and radiation, the daily mean (6:00 UTC to 6:00 UTC) is used.

4. Computation of quantile distributions: The quantile~~s~~ values (the 99 percentile~~s~~ values as well as the 0.5 % and 99.5 % quantile~~s~~ values) of the observational dataset and corresponding RCM grid point distributions are calculated for each variable, each season (DJF, MAM, JJA, SON)

and each of the four weather regimes, for a reference (also referred to as learning) time period when both datasets are available.

5. Quantile mapping: Quantile mapping is then applied to the entire RCM dataset for the application time period, taking into account the season and the weather regime. A linear interpolation is used for quantile values between the quantiles values specifically computed (~~99 percentiles + 0.5 % and 99.5 % quantiles~~ the 99 percentile values as well as the 0.5 % and 99.5 % quantile values). For RCM values greater than the 99.5 % quantile, a constant adjustment based on the value of this last quantile is applied in order to allow for new extremes. For precipitation, it can happen that for low quantiles, the probability of precipitation is lower in the RCM than in the observation dataset (i.e. several null values in the RCM, which can correspond to different positive values in the observational data). In this case, a random draw is performed amongst the observation values within the same quantile.

6. Selection of analogue date for sub-daily disaggregation: For each day in the RCM dataset, an analogous date is chosen in the observational dataset, matching the following criteria: the month and the weather regime must be the same as in the RCM dataset, and whenever possible, consecutive time slices are chosen in the observational dataset in order to avoid artificial jumps in the final data linked to the choice of analogues. A further criterion ~~can be~~ is applied to ensure that the weather situations are even more comparable between the RCM date and the analogous date from the observational record, based on precipitation consistency (wet vs. dry conditions). A threshold of 1 kg m^{-2} day^{-1} on total precipitation is applied to partition dates between dry and wet conditions. For ~~each~~ the first RCM date, a random draw amongst all available observational dates is performed, then the dates are browsed through chronologically until one meets all the requirements outlined above. This analogous day is then used in the following step for all variables. If the following analogue day in the observations still meets all requirements, it is selected as analogue for the following day in the RCM (to ensure as far as possible consecutive time slices). A new random draw is only performed once the analogue fails to meet all requirements described above.

7. Sub-daily disaggregation: The adjusted RCM dataset is disaggregated from a daily integration period into an hourly time step by using the hourly observational data from each analogous date chosen in the previous step to reconstruct the daily cycle of the data:

$$X^h_{RCM}(i) = a \times X^h_{OBS} + b, \tag{2}$$

where $X^h_{RCM}(i)$ is the hourly adjusted RCM value of the variable X and X^h_{OBS} is the hourly observational value of the same variable from the chosen analogous date (step 6). Different criteria are chosen to calculate a and b, depending on the variable considered (Table 1). For the disaggregation of RCM adjusted temperature from daily to hourly (Table 1), a compromise must be made between obtaining minimum and maximum daily values as close as possible to RCM adjusted daily minimum and maximum and minimizing the possible jump in adjusted values between consecutive days. This is achieved by minimising the function:

$$Q(\alpha) = [T^h_{RCM}(1h,i) - T^h_{RCM}(24h,i-1)]^2 + \alpha[Tmin^h_{RCM}(i) - Tmin^{d,adj}_{RCM}(i)]^2 + \alpha[Tmax^h_{RCM}(i) - Tmax^{d,adj}_{RCM}(i)]^2, \qquad (3)$$

where $T^h_{RCM}(1h,i)$ and $T^h_{RCM}(24h,i-1)$ are the hourly adjusted RCM temperature values at the first time step of day i and at the last time step of day $i-1$, $Tmin^h_{RCM}(i)$ and $Tmax^h_{RCM}(i)$ are the hourly minimum and maximum adjusted RCM temperature values respectively, and $Tmin^{d,adj}_{RCM}(i)$ and $Tmax^{d,adj}_{RCM}(i)$ are the daily minimum and maximum adjusted RCM temperature values respectively (Fig. 2). α is a parameter which can be tuned to balance the importance of the minimisation of differences between daily and hourly RCM minima and maxima and the minimisation of the jump between two consecutive days. For a value of α of zero, there would be no jump in values between consecutive days, but the values of $Tmin^h_{RCM}(i)$ and $Tmax^h_{RCM}(i)$ could be far from the values of $Tmin^{d,adj}_{RCM}(i)$ and $Tmax^{d,adj}_{RCM}(i)$. For an infinitely large value of α, the minimum and maximum hourly and daily values would match, but the jump between consecutive days could be significant. Sensitivity tests yielded an optimal value of 2 for α. Following eq. 2, eq. 3 transforms into:

$$Q(\alpha,a,b) = [a \times T^h_{OBS}(1h) + b - T^h_{RCM}(24h,i-1)]^2 + \alpha[a \times Tmin^h_{OBS} + b - Tmin^{d,adj}_{RCM}(i)]^2 + \alpha[a \times Tmax^h_{OBS} + b - Tmax^{d,adj}_{RCM}(i)]^2. \qquad (4)$$

By searching for the local minima $\delta Q/\delta a = 0$ and $\delta Q/\delta b = 0$, a and b can be determined, and the hourly adjusted RCM temperature can be obtained following eq. 2. For specific cases, i.e. for the first day where $T^h_{RCM}(24h,i-1)$ does not exist, or if the determinant of our system is too close to zero (< 0.1), or in the case where $a < 0$, a simpler equation is used, in which we only ensure that final mimimum and maximum daily values correspond to the RCM adjusted minimum and maximum values, by solving:

$$a = \frac{Tmax^{d,adj}_{RCM}(i) - Tmin^{d,adj}_{RCM}(i)}{Tmax^h_{OBS} - Tmin^h_{OBS}} \qquad (5)$$

$$b = Tmax^{d,adj}_{RCM}(i) - a \times Tmax^h_{OBS}. \qquad (6)$$

This procedure is only applied for temperature, because the use of the maximum and minimum criterion can lead to important jumps between consecutive days, which is not the case for other variables (Table 1). For humidity, eq. 2 is solved using $b = 0$ and $a = X^{d,adj}_{RCM}(i)/X^h_{OBS}(24h,i)$, so that the hourly adjusted RCM value and the hourly observational value at the last time step of day i ($X^h_{OBS}(24h,i)$) are equal. For wind speed, the same calculation as for humidity is applied, except if $a > 1$ (i.e., $X^{d,adj}_{RCM}(i) > X^h_{OBS}(24h,i)$). If so, $b = X^{d,adj}_{RCM}(i) - X^h_{OBS}(24h,i)$

is calculated. For humidity and wind speed, if $X^h_{OBS}(24h,i) \leq 10^{-10}$, $a=0$. For precipitation and radiation, $b=0$ and $a = X^{d,adj}_{RCM}(i)/X^h_{OBS}(mean,i)$, so that the mean hourly adjusted RCM value and the mean hourly ~~SAFRAN~~ observation value of day i are equal. For solar radiation, if $X^h_{OBS}(mean,i) \leq 10^{-10}$, $a=0$. For precipitation, if this is the case, $a=1$.

8. Snow/rain partitioning: Total precipitation is separated into rainfall and snowfall based on hourly adjusted temperature (a threshold of 1 °C is used for the transition from snow to rain). As mentioned above, inter-variable consistency is not guaranteed by quantile mapping. Given the importance of the consistency between temperature and precipitation in many applications and in particular in mountainous areas, given that precipitation and temperature are corrected independently from each other (step 5), and because the adjustment can differ for the different precipitation phases, the relationship between temperature and precipitation phase may be modified by quantile mapping, so that the adjusted rain and snow distributions may lose consistency. To avoid this, Olsson et al. (2015) separated temperature data into wet and dry days before adjustment. In our case an additional quantile mapping against the observational dataset is applied for daily cumulated adjusted RCM rainfall and snowfall separately. Hourly adjusted RCM rainfall and snowfall (a_2) are then determined by applying the ratio between daily rainfall or snowfall after quantile mapping (A_2) and daily rainfall or snowfall before quantile mapping (A_1) to the hourly rainfall or snowfall before quantile mapping (a_1):

$$a_2 = a_1 \times \frac{A_2}{A_1} \tag{7}$$

If $A_1 = 0$ and $A_2 = 0$, then $a_2 = 0$. If $A_1 = 0$ and $A_2 \neq 0$, then $a_2 = A_2$.

9. Final adjusted dataset: The resulting adjusted hourly time series for each variable are obtained for each snow year (from the 1st of August to the 31st of July of the following year), matching the format of the observational dataset.

2.2 SAFRAN reanalysis and application of ADAMONT method using SAFRAN

Although the ADAMONT method is highly generic and can be applied using any hourly-resolution observational dataset, in the following we focus on the use of ADAMONT using the SAFRAN reanalysis data as an observational dataset. We first describe SAFRAN, then we present specific features of the ADAMONT method when using SAFRAN as the observational dataset.

The SAFRAN system is a regional scale meteorological downscaling and surface analysis system (Durand et al., 1993), which provides hourly data of temperature, precipitation amount and phase, specific humidity, wind speed, and shortwave and longwave radiation for each mountain region (or "massif") in the French Alps (23 massifs, as illustrated in Fig. 3) but also in the French and Spanish Pyrenees and Corsica. Unlike traditional reanalyses, SAFRAN does not operate on a grid, but on French mountain regions subdivided into different polygons known as massifs. Massifs (Durand et al., 1993, 1999) correspond to regions ranging approximately between 500 and 2,000 km^2 for which meteorological conditions are assumed to be spatially homogeneous but vary with altitude. SAFRAN data are available for elevation bands with a resolution of 300 m, i.e. altitude levels 600, 900, 1200, 1500 m etc. are typically considered, making it possible to extract meteorological information at these altitude levels, or in-between using altitude interpolation.. It was used by Durand et al. (2009b) to create a meteorological reanalysis over the French Alps by combining the ERA-40 reanalysis (Uppala et al., 2005) with various meteorological observations including in situ mountain stations, radiosondes and satellite data. It was complemented after the end of the ERA-40 reanalysis (2002) by large-scale meteorological fields from the ARPEGE analysis, so that it now spans the period from 1959 to 2016, making it one of the longest meteorological reanalyses available in the French mountain regions.

When the ADAMONT method is applied using the SAFRAN reanalysis, only one geographic coordinate is used for each massif, corresponding to the center of the massif (see Fig. 3). However, for each massif several altitude levels are considered, which means that depending on the N factor considered different RCM grid points may be selected for a given massif and altitude. Also, in order to maximise the consistency between massifs after the adjustment process, the dry/wet analogue day criterion used for the time disaggregation of RCM adjusted variables into hourly variable is computed generally for the entire SAFRAN dataset, here in the 23 French Alps massifs. This means that a day is considered dry when the average of all daily precipitation data is below 1 kg m^{-2} day^{-1}, and wet if it falls above the threshold for all massifs and all altitude levels (from an observational perspective), and for all corresponding adjusted RCM grid points (from an adjusted RCM perspective).

2.3 SURFEX/ISBA-Crocus model

Crocus (Brun et al., 1989, 1992; Vionnet et al., 2012) is a detailed snowpack model within the SURFEX externalised surface module (Masson et al., 2013). It enables the computation of the exchanges of energy and mass between the snow surface and the atmosphere (radiative balance, turbulent heat and moisture fluxes, ...), but also between the snowpack and the ground underneath. Similarly to most land surface models, it requires sub-diurnal (ideally hourly) meteorological forcing data including air temperature, humidity, incoming longwave and shortwave radiation, wind speed, as well as rain and snow precipitation. The one-dimensional multilayer physical snow scheme Crocus is able to simulate the evolution of the snowpack over time, by accounting for several processes occurring in the snowpack, such as thermal diffusion, phase changes, metamorphism, etc. The SAFRAN-Crocus model chain has been operationally used for more than 20 years for avalanche hazard forecasting and extensively evaluated over the alpine domain in particular against snow depth observation stations (Durand et al., 1999, 2009b; Lafaysse et al., 2013). Here we apply the Crocus model using either the SAFRAN reanalysis itself, or adjusted fields from a given RCM using the ADAMONT method, in order to compute and compare snow conditions using either driving data. This is both a proof-of-concept of the applicability of the ADAMONT method to generate data appropriate to driving land surface model, and a mean to assess the intervariable consistency of the ADAMONT output given that Crocus is simultaneously sensitive to all meteorological fields and potentially disturbed by inconsistencies in the forcing dataset.

2.4 ADAMONT method evaluation

To evaluate the ADAMONT method, it was applied to the Météo France ALADIN RCM forced by ERA-Interim over the time period from 1980 to 2010. This RCM was run at 12.5 km resolution and we use the daily time resolution output, which is consistent with typical output from EURO-CORDEX RCMs. This simulation was then adjusted against the SAFRAN reanalysis. The spatial domain (2,200 x 2,200 km, centred on France, see Fig. 3) is deliberately smaller than EURO-CORDEX (5,000 x 5,000 km domain covering all of Europe, Fig. 3) although both are on the same order of magnitude, in order to place more emphasis on the method skills than on the output of the RCM itself, especially in terms of chronology. Indeed, the smaller the domain, the more it is constrained by its driving large-scale model (be it a GCM or a reanalysis) (Alexandru et al., 2007).

Performance indicators described below were computed for temperature and total precipitation, but also for the snow depth, which integrates all the meteorological variables considered in the ADAMONT method. Focus was hereby placed on evaluating the ability of the method to correctly represent integrated outputs computed using SURFEX/ISBA-Crocus from meteorological variables adjusted independently of each other. This is often applied to river discharge for downscaling methods used for hydrological applications (e.g., Lafaysse et al., 2014; Olsson et al., 2015).

The method was applied for all 23 massifs of the French Alps and all elevation bands (Fig. 3), totalling 187 massif/altitude configurations. Performance indicators, described below, were either computed spanning all configurations, or focusing on a given altitude level (1200 m and 2100 m) and/or a subset of massifs (the Vercors massif was taken as an example, and computations were also performed separately for the Northern and Southern Alps, respectively).

We specifically tested the following aspects of the method:

- RCM grid points neighbour selection techniques ($N = 0$ or $N = 50$)
- Learning period: Split-sample evaluation was performed using three different learning and application periods (1980-1995, 1995-2010 and 1980-2010), by evaluating the results on an evaluation period different from the learning period (1995-2010 for simulations with the learning period 1980-1995 and vice-versa). These two sub-periods correspond to markedly different climate conditions in the French Alps (Reid et al., 2015). For simulations using the entire learning period 1980-2010, the evaluation period was 1980-2010. This case with a 30 years

learning period corresponds to the typical duration of the learning period when the method is applied for climate projections.

- Rain/snow quantile mapping: The method was applied with (base case) or without ("no corr") the last adjustment step operating on the rainfall and snowfall separately.
- Raw RCM data: Raw RCM simulations, without any adjustment, were considered for some of the variables (temperature and precipitation only) and compared to adjusted results. This can not be used in the case of snow depth, because daily resolution RCM output cannot be employed to run Crocus.
- The impact of using 6-hour input RCM data instead of daily data was also tested, but yielded similar results (not shown). Only results based on daily RCM input are presented because GCM/RCM outputs are often available at this time step on data distribution platforms such as the one of EURO-CORDEX

The following indicators were analysed for temperature, total precipitation and snow depth:

- the seasonal average time series from 1980 to 2010 in the SAFRAN and the adjusted RCM datasets;
- the mean annual cycle over 2 distinct periods: 1980-1995 and 1995-2010 in the SAFRAN and the adjusted RCM datasets;
- the mean value for each elevation band over the evaluation period in the SAFRAN and the adjusted RCM datasets;
- the correlation and the ratio of standard deviations between time series of the SAFRAN and the adjusted RCM datasets for each variable and as a function of the integration window (from 1 day to several years) over the evaluation period;
- the cumulated probability density function (PDF) of daily variables over the evaluation period in the SAFRAN and the adjusted RCM datasets;
- the root mean square error (RMSE) and the mean bias over the evaluation period, computed over seasonal integration periods based on the SAFRAN and the adjusted RCM datasets (to evaluate the method performance in terms of reproducing amounts);
- scores specific to the detection of occurrence of precipitation events in the SAFRAN and the adjusted RCM datasets over the evaluation period: the probability of detection (POD = $n_{hh}/(n_{hh}+n_{hd})$), the false alarm rate (FAR = $n_{dh}/(n_{dh}+n_{hh})$), the probability of false detection (POFD = $n_{dh}/(n_{dh}+n_{dd})$) and the true skill score (TSS = POD - FAR), where n_{hh} is the number of days which are wet in the SAFRAN and wet in the adjusted RCM, n_{dd} the

number of days which are dry in the reanalysis and dry in the adjusted RCM, n_{hd} the number of days which are wet in the reanalysis but dry in the adjusted RCM and n_{dh} the number of days which are dry in the reanalysis but wet in the adjusted RCM (a threshold of 1 kg m^{-2} d^{-1} was considered for the occurrence of precipitation);

- scores for the duration and persistence of precipitation events over the evaluation period (Wilby et al., 1998; Boe et al., 2006): the relative error on the probability of a dry day (EPD = $(n_d^R - n_d^S)/n_d^S$), the relative error on the probability of a dry day following a dry day (EPDD = $(n_{d-d}^R/n_d^R - n_{d-d}^S/n_d^S)/(n_{d-d}^S/n_d^S)$), the relative error on the probability of a wet day following a wet day (EPHH = $(n_{h-h}^R/(n-n_d^R) - n_{h-h}^S/(n-n_d^S))/(n_{h-h}^S/(n-n_d^S))$) and the relative error on the mean duration of wet periods (EHD = $(hdur^R - hdur^S)/hdur^S$), where n_d^R and n_d^S are the number of dry days in the adjusted RCM and in SAFRAN respectively, n_{d-d}^R and n_{d-d}^S the number of dry days following a dry day in the adjusted RCM and in SAFRAN respectively, n_{h-h}^R and n_{h-h}^S the number of wet days following a wet day in the adjusted RCM and in SAFRAN respectively, n is the total number of days, and $hdur^R$ and $hdur^S$ the duration of wet periods in the adjusted RCM and in SAFRAN respectively. A threshold of 1 kg m^{-2} d^{-1} was considered for the occurrence of precipitation.

These indicators are classically employed (e.g., Boe et al., 2006; Vrac et al., 2007b; Lafaysse, 2011; Kotlarski et al., 2014) to assess:

1. the ability of a model/method to reproduce the statistical characteristics of the observed meteorological variables (through the RMSE, the mean bias, the ratio of standard deviations, the duration and persistence of precipitation events and the cumulated PDFs) and their spatial variability (through the mean values at each elevation band and the analysis of different massifs);

2. its capacity to reproduce the low frequency variability of the observations, i.e. their chronology (through the analysis of seasonal average time series, the correlation as a function of the integration window, the detection of precipitation events);

3. its temporal transferability, i.e. its ability to reproduce the observed variables over a period different from the learning period (through the use of split-sample evaluation, the analysis of the mean annual cycle over two distinct periods, the seasonal average time series);

4. its inter-variable consistency, which is assessed here by applying the evaluation indicators to snow depth, an integrated output of the Crocus model.

When available, we compare the indicators with the same criteria applied to analog resampling based or transfer function algorithms by Lafaysse (2011) and Lafaysse et al. (2014), and for other downscaling and adjustment methods by Vrac et al. (2012) and Olsson et al. (2015).

Table 1 outlines the input and output variables of Crocus. Table 2 presents a summary of the different configurations used for the evaluation.

3 Results

3.1 Spatial variability and statistical characteristics of the variables

This section provides the evidence needed to assess the performance of the ADAMONT method applied to a RCM driven by a global reanalysis (ERA-Interim) using the SAFRAN meteorological reanalysis as the observational dataset in the French Alps. Adjusted RCM data are compared to SAFRAN itself. Adequate performance of the method is attained when the two datasets match most.

Figure 4 presents the location of the Vercors massif and its average temperature, precipitation and snow depth for each elevation band, for the evaluation period in the SAFRAN/Crocus reanalysis as well as adjusted RCM. The shape of the mean altitudinal evolution of all three variables is well represented compared to SAFRAN, which is also the case for other massifs (see Supplementary Information). The computed temperature values are very similar to the one in SAFRAN. It is less the case for precipitation, with over- or underestimation depending on the learning period (Fig. 4) and the massif considered (Supplementary Information). Despite the differences in the magnitude of average precipitation in the adjusted RCM compared to SAFRAN, the magnitude of average snow depth in the different adjusted RCM simulations is remarkably close to the results obtained using the reanalysis as meteorological input, with slight differences depending on the massif (see Supplementary Information). For all variables and all massifs, the difference between simulations using the two RCM grid points neighbour selection techniques ($N = 0$ or $N = 50$) is smaller than the difference induced by using different learning periods.

Figs. 5-7 display the mean bias and the RMSE for each raw and adjusted RCM simulation compared to SAFRAN, for temperature, precipitation and snow depth, for the Vercors massif. Additionally, Table 3 presents the corresponding scores at the annual time scale compared to mean values, for the adjusted RCM L. 1980-2010 simulation, for each massif in the French Alps and for the Northern and Southern Alps, at 1200 m a.s.l. and 2100 m a.s.l.. This highlights the large biases and RMSEs values obtained when using raw RCM simulations compared to adjusted simulations, a feature common to all massifs (Figs. 5-6 and Supplementary Information).

For temperature, biases of the adjusted RCM simulations vary with elevation and for the different massifs (Fig. 5, Table 3 and Supplementary Information), but lie always within 1 K. Biases are generally smaller in autumn (SON) than for other seasons. RMSEs also vary with elevation and massifs, and can differ significantly between simulations using the two different RCM grid points neighbour selection techniques. For elevations above $\approx$ 2100 m a.s.l., stronger biases and higher RMSEs are found for simulations using the selection technique accounting for altitude differences ($N = 50$), especially in summer (JJA) than for other seasons. Temperature biases and RMSEs values also depend on the learning period considered, the longer learning period 1980-2010 generally presenting smaller biases and RMSEs (Fig. 5 and Supplementary Information).

For precipitation, biases generally vary with altitude (Fig. 6, Table 3 and Supplementary Information), but less than for temperature (Fig. 5, Table 3 and Supplementary Information). Biases of the adjusted simulations remain smaller than 150 kg m^{-2} per month in absolute value, corresponding to up to 90% depending on the massif and altitude, and are generally stronger in summer. Smaller autumn and winter precipitation biases lead to a good agreement between the magnitude of average snow depth in the different adjusted RCM simulations and the results obtained using the reanalysis as meteorological input (as noted in Fig. 4). RMSEs values generally increase with altitude. Using different RCM grid points neighbour selection techniques has less impact on precipitation scores than for temperature, except that the $N = 50$ configuration yields more variability in scores with altitude. This is due to the choice of different grid points for different altitudes of a single massif, because precipitation is spatially more variable than temperature. The influence of the learning period on scores is also visible.

For snow depth, the biases never exceed 50 cm, which corresponds to up to 50% depending on the altitude and the massif (Fig. 7, Table 3 and Supplementary Information). The biases are smaller in autumn than for other seasons, similar to temperature (Fig. 5 and Supplementary Information). Summer biases at high altitudes are almost always negative, which cannot always be explained by a combination of positive biases in temperature and/or negative biases in precipitation, indicating the possible impact of other variables on snow depth (such as longwave radiation for example). RMSE values increase with altitude, due to the effect of increased snow accumulation with altitude. Using the $N = 50$ configuration generally degrades scores at high elevations, similar to the effect on temperature.

For precipitation and snow depth, simulations without the ultimate quantile mapping on snowfall and rainfall are also presented (by definition it has no impact on temperature). It is clear from Figs. 6-7 and the Supplementary Information that without this ultimate correction (no corr), biases in precipitation and snow depth are much stronger and RMSEs much higher than when this correction is applied.

Fig. 8 represents the ratio of standard deviations between each adjusted RCM simulation and SAFRAN for temperature, precipitation and snow depth and as a function of the integration window (from 1 day to several years), over the learning period. Ratios are displayed for the Vercors massif, for altitudes of 1200 m a.s.l. and 2100 m a.s.l.. If this ratio is lower than 1, it means that adjusted RCM simulations have a smaller standard deviation (i.e. variability) than SAFRAN. For temperature, the ratio of standard deviations is very close to 1 for integration windows of 1 day to a few months. It varies more for longer integration windows of 1 year or more. The differences between the two altitudinal levels considered or between massifs are limited (Fig. 8 and Supplementary Information). Similarly, choosing different learning periods or different grid points neighbour selection techniques has little effect on the ratio of standard deviations. For precipitation, ratios of standard deviations are also close to 1 (generally between 0.8 and 1.2) for integration windows of 1 day to 1 month. This result is similar to ratios of variance between daily RCMs adjusted with a Cumulative Distribution Function-transform and observations for the Mediterranean region in Vrac et al. (2012). For integration windows of 1 month or more, the ratios vary more, with under- or overestimation of variance depending on the massif, the learning period and the grid points neighbour selection technique considered (Fig. 8 and Supplementary Information). For snow depth, the ratio does not vary until 1 month of integration approximately (Fig. 8 and Supplementary Information), and shows larger variations for higher values. Some differences can be noted for different altitudes, and different massifs, but also for different learning periods and the two grid points neighbour selection techniques considered.

Fig. 9 presents the cumulated probability density functions (PDFs) of daily temperature, precipitation and snow depth at 1200 m a.s.l. and 2100 m a.s.l. for the Vercors massif. The distributions of daily temperature of adjusted RCM simulations are remarkably close to the distribution of SAFRAN (Fig. 9 and Supplementary Information). The agreement is better than the one observed in Lafaysse (2011) and Lafaysse et al. (2014) between the different configurations of analog-based and transfer functions algorithms and SAFRAN for the Durance basin (see Fig.F.2 in Lafaysse (2011), and Fig.5 in Lafaysse et al. (2014)). A similar agreement was observed in Olsson et al. (2015) between two configurations of a distribution-based scaling method and observations in Finland. Only small differences are observed for different altitudes or different massifs (Fig. 9 and Supplementary Information), and the choice of the learning period or the grid points neighbour selection technique has almost no impact on the PDF. For precipitation, the PDFs of adjusted RCM simulations are also very close to the PDF of SAFRAN, with a slight overestimation or underestimation of moderate to high precipitation, depending on the learning period, occurring for most massifs (Fig. 9 and Supplementary Information). This result is similar to that observed in Lafaysse (2011) for the Durance basin (see Fig.11.7 therein). As for temperature, altitude and massif location have only a small impact on the distribution, as well as the grid points neighbour selection technique considered. The distribution of snow depth, on the other hand, depends more on the massif considered and the altitude (Fig. 9 and Supplementary Information). As for precipitation, the moderate to high snow depth values seem to be slightly overestimated or underestimated for most massifs, depending on the learning period. The choice of the grid points neighbour selection technique has also slightly more impact on snow depth PDFs than for temperature and precipitation. The fact that PDFs for temperature and precipitation are very close to the ones of SAFRAN is a logical consequence of using a quantile mapping approach. That it is also true for snow depth indicates that even if they are treated separately, the inter-variable consistency of the meteorological fields generated using our method is in general appropriate.

The capacity to reproduce the duration and persistence of precipitation events is shown in Fig. 10. The ratio between the number of dry days and the number of rainy or snowy days is very correctly reproduced for every massif and altitude (Fig. 10 and Supplementary Information), the relative error on the probability of a dry day being lower than 5%. This feature was also observed by Lafaysse (2011) in his study of the Durance basin (see Fig.11.10 therein). The persistence of dry and rainy/snowy events is generally underestimated (up to about -30%), which was also the case in Lafaysse (2011), even though the error depends on the massif and the altitude considered. In general, errors on the persistence of precipitation events are larger in massifs of the Southern Alps than the Northern Alps (Supplementary Information). Using different learning periods and different grid points neighbour selection techniques has an impact on scores, but this is small compared to the influence of the massif or the altitude.

3.2 Mean seasonal variations

Fig. 11 represents the mean annual cycle of temperature, precipitation and snow depth for the different adjusted RCM simulations vs. the SAFRAN/Crocus reanalysis, for the period 1980-1995 and 1995-2010, for the Vercors massif at 1200 m a.s.l. and 2100 m a.s.l.. The mean annual cycle of temperature is very well reproduced for every massif and altitude (Fig. 11 and Supplementary Information). Using different grid points neighbour selection techniques has a limited impact on the mean annual cycle. For precipitation, the mean annual cycle is relatively well reproduced (Fig. 11 and Supplementary Information). The choice of grid points neighbour selection technique can have slightly more influence on the results than for temperature. For snow depth, the annual cycle is remarkably well reproduced, with peak snow depth in the core of winter (JFM), and no snow or reduced amounts in late summer months (JAS) (Fig. 11 and Supplementary Information). As for temperature, the impact of the grid points neighbour selection technique is very limited.

3.3 Interannual variability

The chronology of time series of seasonal averages of temperature, precipitation and snow depth from 1980 to 2010 is shown in Figs. 12-14, for the Vercors massif at 1200 m a.s.l. and 2100 m a.s.l., in SAFRAN and the adjusted RCM. Temperature RCM time series are similar to SAFRAN, with an interannual variability which is well reproduced (Fig. 12 and Supplementary Information). Some significant differences appear when using different learning periods, as already noted in Sect. 3.2. Using different grid points neighbour selection techniques has an impact on the time series of temperature which is generally smaller than the influence of the learning period. However, as already noted in Sect. 3.1, the agreement between observed and simulated time series is degraded for high altitudes under the spatial and altitudinal ($N = 50$) grid points neighbour selection technique. The interannual variability of precipitation is also well reproduced for most massifs and altitudes (Fig. 13 and Supplementary Information), especially given that the only forcing of the RCM comes from ERA-Interim reanalysis at the boundaries of the RCM domain. It is slightly less well reproduced in summer (JJA), as observed by Lafaysse (2011) for the analog resampling based transfer function algorithm DSCLIM (Pagé et al., 2009) and the Durance basin (see Fig.10.1 therein). Differences between simulations using different learning periods mostly appear in summer (JJA). The use of different grid points neighbour selection techniques has a rather limited impact on time series of precipitation, whose magnitude depends on the massif and the altitude (Fig. 13 and Supplementary Information). For snow depth, the interannual variability is well reproduced in winter (DJF) and correctly reproduced in intermediate seasons (MAM and SON). Summer snow depths are generally underestimated, as already noted in Sect. 3.1, but represent a small portion of the annual snow accumulation. Likewise, adjusted data using the spatial and altitudinal ($N = 50$) RCM grid points selection technique can be degraded at high altitudes, similarly to temperature.

Fig. 15 displays the temporal correlation between each adjusted RCM simulation and SAFRAN over the evaluation period for temperature and precipitation and as a function of the integration window (from 1 day to several years). Correlations are displayed for the Vercors massif, for altitudes of 1200 m a.s.l. and 2100 m a.s.l.. Additionally, Table 3 presents the same correlation values at the same altitudes, for an integration window of 1 year, and for the adjusted RCM L. 1980-2010 simulation only, for every massif of the French Alps, and for the Northern and Southern Alps. Snow depth values were not included because of their cumulative nature. Correlations for temperature are very high (always above 0.8) for all massifs and altitudes until an integration window of a few months to 1 year (Fig. 15, Table 3 and Supplementary Information), ~~similar to~~ as found by Lafaysse (2011) (see Fig.F.21 therein). The differences between learning periods are negligible. As already observed in Sect. 3.1 and for the time series above, the correlation is clearly degraded for high altitudes (above ≈2100 m a.s.l.) in simulations using the $N = 50$ grid points selection technique. Precipitation also yields satisfactory correlation values (always above 0.4) until a few months integration window, which vary depending on the massif considered(Fig. 15 and Supplementary Information). Correlations are generally similar or even better than the ones observed in Lafaysse (2011) for various statistical downscaling models and different configurations of the ALADIN RCM (see Fig.12.10 therein). The use of the $N = 50$ grid points neighbour selection technique increases or decreases correlation values depending on the massif and the altitude considered. The choice of learning period has a limited effect on correlation, at least up to integration windows of a few months. Correlations are higher at the scale of the Northern and Southern Alps than at the massif scale (Table 3). This scale dependence of precipitation downscaling skill was also illustrated by Gangopadhyay et al. (2004) and Mezghani and Hingray (2009).

Scores for the detection of precipitation events are presented in Fig. 16, for the Vercors massif, for altitudes of 1200 m a.s.l. and 2100 m a.s.l.. The scores vary depending on massifs and altitude, but a general pattern emerges (Fig. 16 and Supplementary Information). The POD is the highest, with values between 0.55 and 0.8, very similar to Lafaysse (2011) (see Figs. 11.14 and 12.8 therein). The FAR is rather low (always below 0.5), as well as the POFD, below 0.2. TSS are generally better for massifs of the Northern Alps (0.25 to 0.6) than the Southern Alps (0.1 to 0.4, Supplementary

Information), where PODs are lower and FAR much higher. Such results indicate that the method performs well in detecting precipitation events. Using different learning periods has a rather limited impact on the detection of precipitation. The choice of the grid points selection technique has a limited influence at low to mid-altitudes, which increases above ≈2100 m a.s.l..

4 Discussion

This section discusses the main limits of the method described and evaluated here, and the limits of the evaluation method itself.

4.1 Transferability in time

The temporal transferability of the ADAMONT method, i.e. its capacity to apply adequately to a period which is different from the learning period, can be evaluated from results in Sects. ~~3.1, 3.3 and 3.2~~ 3.1, 3.2 and 3.3.

Figs. ~~12-14, 11~~ 11-14 and Supplementary Information reveal some significant differences when using different learning periods. This feature is generally most visible in summer. It denotes a limit in the temporal transferability of the ADAMONT method, which was also the case in Lafaysse (2011) for the analog-based and transfer functions algorithms (see Figs. 11.11 and 11.12 therein). Using the longer learning period 1980-2010 yields better results, most probably due to the fact that, in this case, the learning and evaluation periods are the same, but also the fact that the learning period is longer.

There are some limits in the conclusions which can be drawn from this transferability assessment. First, reanalysis data used here as forcing for the RCM (ERA-Interim) or for statistical adjustment and evaluation purposes (SAFRAN reanalysis) are heterogeneous in time (Sterl, 2004; Vidal et al., 2010). These heterogeneities are especially marked in summer in the SAFRAN reanalysis, when most observations from mountain stations are not available (Gobiet et al., 2015). Secondly, variations which will occur in the future climate are expected to be much stronger than the variations which can be tested on our evaluation period. Issues related to the time transferability of the adjustment approach may be amplified when applied in the context of climate projections, but their relative impact will probably be lower than shown here given the magnitude of the expected changes.

4.2 Impact of the spatial selection technique

The impact of the RCM grid point selection technique is illustrated in Sects. 3.1 and 3.3. Indeed, Figs. 5-7, 12-14, 15 and Supplementary Information show a clear degradation of scores for elevations above ≈ 2100 m a.s.l. using a selection criterion explicitly accounting for the altitude difference ($N = 50$). This is linked to the scarcity of high altitude grid points in ALADIN compared to SAFRAN, resulting in grid points being selected several tens of kilometres from the centre point of most SAFRAN massifs (see Fig. 4 and Supplementary Information for the location of selected grid points). The impact of this issue depends on the location of massifs relative to high-altitude grid points in ALADIN. For example, most Southern Alps massifs are affected, except the southernmost massifs of Ubaye, Alpes Azur and Mercantour (Supplementary Information), which are located less than 15 km from high-altitude points. This shows that, although it seems appealing to select RCM grid points at elevations matching the elevation of the observation dataset, rather than using RCM grid points with a potentially large elevation difference (hence leading to stronger adjustment requirement), in practice the results are far more homogeneous and quantitatively generally equivalent or better when concentrating only on the horizontal distance between the RCM grid points and the observation dataset.

4.3 Inter-variable consistency

The lack of explicitly enforced inter-variable consistency of the quantile mapping method can be a major disadvantage. As we focus on a mountainous region for the evaluation and future use of the method the consistency between temperature and precipitation phase is crucial. The impact of this ultimate correction is assessed in Sect. 3.1. Figs. 6-7 and Supplementary Information show that without this ultimate correction (no corr), biases for precipitation are much stronger and RMSEs much higher than with this ultimate correction, highlighting its importance.

The inter-variable consistency of the ADAMONT method is indirectly assessed by applying the evaluation metrics described above to an integrated output of the Crocus model, the snow depth, which is computed from meteorological variables adjusted independently from each other. As mentioned above, snow depth results are generally satisfying, which tend to indicate a good inter-variable consistency. Performance indicators for snow depths are often consistent with temperature and precipitation indicators, even though they cannot always be explained by these two variables alone (for example the analysis of biases in Sect. 3.1), indicating the probable influence of other variables not directly analysed here such as longwave radiation.

4.4 Limits of the evaluation method

The spatial consistency of the ADAMONT method has not been evaluated other than by using spatial averages. In future studies, it would be necessary to test it by evaluating spatial correlations (for example using metrics described in Kotlarski et al. (2014)), or by using integrated variables requiring spatial variability, such as snow cover area or river discharges.

In this study, we evaluated the method using only the ALADIN-Climate RCM. However, Olsson et al. (2015) showed that the choice of RCM could have a significant impact on the evaluation of the performance of the adjustment method. Evaluation using another RCM could thus prove useful, even though we would have to use RCM outputs run on the same spatial domain as the ALADIN-Climate RCM in order to compare them.

5 Conclusions

~~The new method to statistically adjust regional climate model projections ADAMONT is introduced, which provides hourly~~ The new method ADAMONT is able to statistically adjust daily regional climate model projections and to provide hourly adjusted outputs of temperature, precipitation, wind speed, humidity and short- and longwave radiation necessary to force energy balance land surface (impact) models. The method processes daily outputs from an RCM and adjusts them against a sub-daily (hourly, typically) observational dataset. The method was evaluated using outputs from the ALADIN-Climate RCM driven by ERA-Interim reanalysis for the time period 1980 - 2010, using the SAFRAN meteorological reanalysis in the French Alps as an observation dataset. The direct outputs of the ADAMONT method, namely temperature and total precipitation, as well as an indirect output, namely snow depth, computed by the Crocus model from meteorological variables corrected independently of each other, were evaluated. The impact of the learning period was tested, as well as the method to select RCM grid points corresponding to each observational point. The evaluation addressed four main concerns: (1) the ability of the ADAMONT method to reproduce the spatial (especially altitudinal) variability and the statistical characteristics of SAFRAN variables, (2) its ability to reproduce the low frequency variability, i.e. the chronology, of SAFRAN, through the analysis of the interannual variability and the annual cycle of adjusted variables (3) the temporal transferability of the method, and (4) its inter-variable consistency.

Performance scores are always better for adjusted RCM simulations than for raw RCM simulations, which highlights the need for such adjustment and demonstrates the skill of the method. In general, the performance of the ADAMONT method concerning temperature is better than for precipitation. However, evaluation indicators for precipitation are generally similar or even better than the indicators evaluated in Lafaysse (2011) and Lafaysse et al. (2014) for other types of algorithms (analog-based or transfer functions). Snow depth yields good results, considering its integrated nature, i.e. the fact that it was computed from variables corrected independently. The impact of the learning period depends on the evaluation indicator considered, and must be considered when applying the method. The best solution is probably to choose the longest possible learning period. For precipitation and snow depth, the importance of the ultimate quantile mapping applied to snowfall and rainfall (i.e., after a first quantile mapping on total precipitation, an additional quantile mapping against the observational dataset is applied for daily cumulated adjusted RCM rainfall and snowfall separately) is unambiguously demonstrated. Using a grid points selection technique relying on spatial but also altitudinal proximity between SAFRAN massif centre points and RCM grid points either had no impact on the performance indicators or degraded them for altitudes higher than 2100 m a.s.l.. As a consequence, the simple spatial grid points neighbour selection technique will be retained for future applications of the method.

The ADAMONT method is generic and can be applied to any observational dataset. Its application using the SAFRAN reanalysis as the observation dataset is somewhat a specific case, initially tailord for French mountainous regions (Durand et al., 2009a). However, beyond the French mountain regions, the method could be applied in France using the SAFRAN-France gridded reanalysis (Vidal et al., 2010). A Spanish version of SAFRAN was also developed recently (Quintana-Seguí et al., 2017).The method could be applied to other observational datasets or meteorological reanalyses, such as ERA-Interim surface fields (Dee et al., 2011) or MESCAN (Soci et al., 2016).

6 Code availability

The code of the ADAMONT v1.0 method is available as an open git repository after free registration at https://opensource.cnrm-game-meteo.fr/projects/adamont. The version used for this article is available at https://opensource.cnrm-game-meteo.fr/projects/adamont/repository?rev=ADAMONT-v1.0.

The version of the open source code of SURFEX/ISBA-Crocus used in this study is available as a specific branch of an open svn repository, after free registration, at https://opensource.cnrm-game-meteo.fr/projects/surfex. For reproductibility of results, the version used in this work is tagged as http://svn.cnrm-game-meteo.fr/projets/surfex/tags/ADAMONT-1.0.

Acknowledgements. This study benefited from funding from the French Ministry for Ecology (MEEM) through the GICC program and ONERC, in the framework of the ADAMONT project. It also forms part of the Interreg project POCTEFA/Clim'Py. CNRM/CEN is part of LabEX OSUG@2020 (ANR10 LABX56). We thank S. Bernus, P. Lassegues, F. Besson and V. Gouget at Météo-France for helping making the method more robust. We thank two anonymous reviewers for useful and constructive comments.

References

Abegg, B., Agrawala, S., Crick, F., and de Montfalcon, A.: Climate change impacts and adaptation in winter tourism, in: Climate Change in the European Alps, edited by Agrawala, S., pp. 25–60, OECD Paris, doi:10.1787/9789264031692-en, 2007.

Addor, N., Rohrer, M., Furrer, R., and Seibert, J.: Propagation of biases in climate models from the synoptic to the regional scale: Implications for bias adjustment, J. Geophys. Res. - Atmos., 121, 2075–2089, doi:10.1002/2015JD024040, 2016.

Alexandru, A., de Elia, R., and Laprise, R.: Internal variability in regional climate downscaling at the seasonal scale, Mon. Weather Rev., 135, 3221–3238, doi:http://dx.doi.org/10.1175/MWR3456.1, 2007.

Beniston, M.: Variations of snow depth and duration in the Swiss Alps over the last 50 years: links to changes in large-scale climatic forcings, Climatic Change, 36, 281–300, doi:10.1023/A:1005310214361, 1997.

Berg, P., Bosshard, T., and Yang, W.: Model consistent pseudo-observations of precipitation and their use for bias correcting regional climate models, Climate, 3, 118–132, doi:10.3390/cli3010118, 2015.

Boe, J., Terray, L., Habets, F., and Martin, E.: A simple statistical-dynamical downscaling scheme based on weather types and conditional resampling, J. Geophys. Res. - Atmos., 111, doi:10.1029/2005JD006889, 2006.

Boe, J., Terray, L., Habets, F., and Martin, E.: Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies, Int. J. Climatol., 27, 1643–1655, doi:10.1002/joc.1602, 2007.

Boulangeat, I., Georges, D., Dentant, C., Bonet, R., Van Es, J., Abdulhak, S., Zimmermann, N., and Thuiller, W.: Anticipating the spatio-temporal response of plant diversity and vegetation structure to climate and land use change in a protected area, Ecography, 37, 1230–1239, doi:10.1111/ecog.00694, 2014.

Brun, E., Martin, E., Simon, V., Gendre, C., and Coléou, C.: An energy and mass model of snow cover suitable for operational avalanche forecasting, J. Glaciol., 35, 333 – 342, 1989.

Brun, E., David, P., Sudul, M., and Brunot, G.: A numerical model to simulate snow-cover stratigraphy for operational avalanche forecasting, J. Glaciol., 38, 13 – 22, 1992.

Castebrunet, H., Eckert, N., Giraud, G., Durand, Y., and Morin, S.: Projected changes of snow conditions and avalanche activity in a warming climate: the French Alps over the 2020-2050 and 2070-2100 periods, The Cryosphere, 8, 1673–1697, doi:10.5194/tc-8-1673-2014, 2014.

Cattiaux, J., Douville, H., and Peings, Y.: European temperatures in CMIP5: origins of present-day biases and future uncertainties, Clim. Dynam., 41, 2889–2907, doi:10.1007/s00382-013-1731-y, 2013.

Christensen, J., Boberg, F., Christensen, O., and Lucas-Picher, P.: On the need for bias correction of regional climate change projections of temperature and precipitation, Geophys. Res. Lett., 35, doi:10.1029/2008GL035694, 2008.

Colin, J., Déqué, M., Radu, R., and Somot, S.: Sensitivity study of heavy precipitation in Limited Area Model climate simulations: influence of the size of the domain and the use of the spectral nudging technique, Tellus A, 62, 591–604, doi:10.1111/j.1600-0870.2010.00467.x, 2010.

Dayon, G., Boé, J., and Martin, E.: Transferability in the future climate of a statistical downscaling method for precipitation in France, Journal of Geophysical R, 120, 1023–1043, doi:10.1002/2014JD022236, 2015.

Dee, D. and Uppala, S.: Variational bias correction of satellite radiance data in the ERA-Interim reanalysis, Quart. J. Roy. Meteor. Soc., 135, 1830–1841, doi:10.1002/qj.493, 2009.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kålberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quart. J. Roy. Meteor. Soc., 137, 553–597, doi:10.1002/qj.828, 2011.

Déqué, M.: Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values, Global Planet. Change, 57, 16–26, doi:10.1016/j.gloplacha.2006.11.030, 2007.

Driouech, F., Déqué, M., and Mokssit, A.: Numerical simulation of the probability distribution function of precipitation over Morocco, Clim. Dynam., 32, 1055–1063, doi:10.1007/s00382-008-0430-6, 2009.

Driouech, F., Déqué, M., and Sánchez-Gómez, E.: Weather regimes—Moroccan precipitation link in a regional climate change simulation, Global Planet. Change, 72, 1–10, doi:10.1016/j.gloplacha.2010.03.004, 2010.

Durand, Y., Brun, E., Mérindol, L., Guyomarc'h, G., Lesaffre, B., and Martin, E.: A meteorological estimation of relevant parameters for snow models, Ann. Glaciol., 18, 65–71, 1993.

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

Durand, Y., Giraud, G., Laternser, M., Etchevers, P., Mérindol, L., and Lesaffre, B.: Reanalysis of 47 Years of Climate in the French Alps (1958–2005): Climatology and Trends for Snow Cover, J. Appl. Meteor. Climat., 48, 2487–2512, doi:10.1175/2009JAMC1810.1, 2009a.

Durand, Y., Giraud, G., Laternser, M., Etchevers, P., Mérindol, L., and Lesaffre, B.: 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. Meteor. Climat., 48, 429–449, doi:10.1175/2008JAMC1808.1, 2009b.

Francois, B., Hingray, B., Creutin, J., and Hendrickx, F.: Estimating water system performance under climate change: influence of the management strategy modeling, Water Resour. Manag., 29, 4903–4918, doi:10.1007/s11269-015-1097-5, 2015.

Gangopadhyay, S., Clark, M., Werner, K., Brandon, D., and Rajagopalan, B.: Effects of spatial and temporal aggregation on the accuracy of statistically downscaled precipitation estimates in the upper Colorado River basin, J. Hydrometeorol., 5, 1192–1206, doi:http://dx.doi.org/10.1175/JHM-391.1, 2004.

Giorgi, F., Jones, C., and Asrar, G. R.: Addressing climate information needs at the regional level: the CORDEX framework, World Meteorol. Organ., 58, 175–183, 2009.

Gobiet, A., Suklitsch, M., and Heinrich, G.: The effect of empirical-statistical correction of intensity-dependent model errors on the temperature climate change signal, Hydrol. Earth Syst. Sc., 19, 4055–4066, doi:10.5194/hess-19-4055-2015, 2015.

Gottardi, F., Obled, C., Gailhard, J., and Paquet, E.: Statistical reanalysis of precipitation fields based on ground network data and weather patterns: Application over French mountains, Journal of Hydrology, doi:10.1016/j.jhydrol.2012.02.014, 2012.

Hantel, M. and Hirtl-Wielke, L.-M.: Sensitivity of Alpine snow cover to European temperature, Int. J. Climatol., 27, 1265–1275, doi:10.1002/joc.1472, 2007.

IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)], Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp, 2013.

IPCC: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Field, C.B., V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (eds.)], Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1132 pp, http://www.ipcc.ch/report/ar5/wg2/, 2014a.

IPCC: Climate change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Barros, V.R., C.B. Field, D.J. Dokken, M.D. Mastrandrea, K.J. Mach, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (eds.)], Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 688 pp, http://www.ipcc.ch/report/ar5/wg2/, 2014b.

IPCC: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Edenhofer, O., R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schlömer, C. von Stechow, T. Zwickel and J.C. Minx (eds.)], Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1435 pp, http://www.ipcc.ch/report/ar5/wg3/, 2014c.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., Braun, A., Colette, A., Deque, M., Georgievski, G., Georgopoulou, E., Gobiets, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kroner, N., Kotlarski, S., Kriegsmann, A., Martin, E., Meijgaard, E. V., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.-F., Teichmann, C., Valentini, R., Vautard, R., Weber, B., and Yiou, P.: EURO-CORDEX: new high-resolution climate change projections for European impact research, Regional Environmental Change, 14, 563–578, doi:10.1007/s10113-013-0499-2, 2014.

Jomelli, V., Brunstein, D., Déqué, M., Vrac, M., and Grancher, D.: Impacts of future climatic change (2070–2099) on the potential occurrence of debris flows: a case study in the Massif des Ecrins (French Alps), Climatic Change, 97, 171–191, doi:10.1007/s10584-009-9616-0, 2009.

Kotlarski, S., Bosshard, T., Lüthi, D., Pall, P., and Schär, C.: Elevation gradients of European climate change in the regional climate model COSMO-CLM, Climatic change, 112, 189–215, doi:10.1007/s10584-011-0195-5, 2012.

Kotlarski, S., Keuler, K., Christensen, O. B., Colette, A., Deque, M., Gobiet, A., Goergen, K., Jacob, D., Luthi, D., Meijgaard, E. V., Nikulin, G., Schar, C., Teichmann, C., Vautard, R., Warrach-Sagi, K., and Wulfmeyer, V.: Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble, Geoscientific Model Development, 7, 1297–1333, doi:10.5194/gmd-7-1297-2014, 2014.

Lafaysse, M.: Changement climatique et régime hydrologique d'un bassin alpin. Génération de scénarios sur la Haute-Durance, méthodologie d'évaluation et incertitudes associées, Ph.D. thesis, Université Toulouse III-Paul Sabatier, 2011.

Lafaysse, M., Morin, S., Coleou, C., Vernay, M., Serca, D., Besson, F., Willemet, J.-M., Giraud, G., and Durand, Y.: Towards a new chain of models for avalanche hazard forecasting in French mountain ranges, including low altitude mountains, in: Proceedings of International Snow Science Workshop Grenoble–Chamonix Mont-Blanc, pp. 162–166, CEN, 2013.

Lafaysse, M., Hingray, B., Mezghani, A., Gailhard, J., and Terray, L.: Internal variability and model uncertainty components in future hydrometeorological projections: The Alpine Durance basin, Water Resources Research, 50, 3317–3341, doi:10.1002/2013WR014897, 2014.

Maraun, D.: Bias correcting climate change simulations-a critical review, Curr. Clim. Change Rep., 2, 211–220, doi:10.1007/s40641-016-0050-x, 2016.

Maraun, D., Wetterhall, F., Ireson, A., Chandler, R., Kendon, E., Widmann, M., Brienen, S., Rust, H., Sauter, T., Themeßl, M., et al.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, Rev. Geophys., 48, doi:10.1029/2009RG000314, 2010.

Martin, E., Brun, E., and Durand, Y.: Sensitivity of the French Alps snow cover to the variation of climatic variables, Ann. Geophys., 12, 469–477, doi:10.1007/s00585-994-0469-6, 1994.

Marty, C., Schlögl, S., Bavay, M., and Lehning, M.: How much can we save? Impact of different emission scenarios on future snow cover in the Alps, The Cryosphere, 11, 517–529, doi:10.5194/tc-11-517-2017, 2017.

Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Belamari, S., Barbu, A., Boone, A., Bouyssel, F., Brousseau, P., Brun, E., Calvet, J.-C., Carrer, D., Decharme, B., Delire, C., Donier, S., Essaouini, K., Gibelin, A.-L., Giordani, H., Habets, F., Jidane, M., Kerdraon, G., Kourzeneva, E., Lafaysse, M., Lafont, S., Lebeaupin Brossier, C., Lemonsu, A., Mahfouf, J.-F., Marguinaud, P., Mokhtari, M., Morin, S., Pigeon, G., Salgado, R., Seity, Y., Taillefer, F., Tanguy, G., Tulet, P., Vincendon, B., Vionnet, V., and Voldoire, A.: The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of Earth surface variables and fluxes, Geoscientific Model Development, 6, 929–960, doi:10.5194/gmd-6-929-2013, 2013.

Maurer, E. P. and Pierce, D. W.: Bias correction can modify climate model simulated precipitation changes without adverse effect on the ensemble mean, Hydrology and Earth System Sciences, 18, 915–925, doi:10.5194/hess-18-915-2014, 2014.

Mezghani, A. and Hingray, B.: A combined downscaling-disaggregation weather generator for stochastic generation of multisite hourly weather variables over complex terrain: Development and multi-scale validation for the Upper Rhone River basin, J. Hydrol., 377, 245–260, doi:10.1016/j.jhydrol.2009.08.033, 2009.

Michelangeli, P.-A., Vautard, R., and Legras, B.: Weather regimes: Recurrence and quasi stationarity, J. Atmos. Sci., 52, 1237–1256, doi:http://dx.doi.org/10.1175/1520-0469(1995)052<1237:WRRAQS>2.0.CO;2, 1995.

Olsson, T., Jakkila, J., Veijalainen, N., Backman, L., Kaurola, J., and Vehviläinen, B.: Impacts of climate change on temperature, precipitation and hydrology in Finland–studies using bias corrected Regional Climate Model data, Hydrol. Earth Sys. Sci., 19, 3217–3238, doi:10.5194/hess-19-3217-2015, 2015.

Pagé, C., Terray, L., and Boé, J.: dsclim: A software package to downscale climate scenarios at regional scale using a weather-typing based statistical methodology, Tech. rep., Climate Modelling and Global Change–CERFACS, Toulouse, http://www. cerfacs. fr/globc/publication/technicalreport/2009/dsclim doc. pdf, 2009.

Piazza, M., Boe, J., Terray, L., Page, C., Sanchez-Gomez, E., and Deque, M.: Projected 21st century snowfall changes over the French Alps and related uncertainties, Climatic Change, 122, 583–594, doi:10.1007/s10584-013-1017-8, 2014.

Quintana-Seguí, P., Turco, M., Herrera, S., and Miguez-Macho, G.: Validation of a new SAFRAN-based gridded precipitation product for Spain and comparisons to Spain02 and ERA-Interim, Hydrol. Earth Sys. Sci., 21, 2187–2201, doi:10.5194/hess-21-2187-2017, 2017.

Rauscher, S. A., Coppola, E., Piani, C., and Giorgi, F.: Resolution effects on regional climate model simulations of seasonal precipitation over Europe, Clim. Dynam., 35, 685–711, doi:10.1007/s00382-009-0607-7, 2010.

Reid, P., Hari, R., Beaugrand, G., Livingstone, D., Marty, C., Straile, D., Barichivich, J., Goberville, E., Adrian, R., Aono, Y., Brown, R., Foster, J., Groisman, P., Hélaou et, P., Hsu, H.-H., Kirby, R., Knight, J., Kraberg, A., Li, J., Lo, T.-T., Myneni, R., North, R., Pounds, J., Sparks, T., St ubi, R., Tian, Y., Wiltshire, K., Xiao, D., and Zhu, Z.: Global impacts of the 1980s regime shift, Glob. Change Biol., 22, 682–703, doi:10.1111/gcb.13106, 2015.

Rousselot, M., Durand, Y., Giraud, G., Mérindol, L., Dombrowski-Etchevers, I., Déqué, M., and Castebrunet, H.: Statistical adaptation of ALADIN RCM outputs over the French Alps -application to future climate and snow cover, The Cryosphere, 6, 785–805, doi:10.5194/tc-6-785-2012, 2012.

Schmucki, E., Marty, C., Fierz, C., and Lehning, M.: Simulations of 21st century snow response to climate change in Switzerland from a set of RCMs, International Journal of Clim, 35, 3185–3384, doi:10.1002/joc.4205, 2014.

Soci, C., Bazile, E., Besson, F., and Landelius, T.: High-resolution precipitation re-analysis system for climatological purposes, Tellus A, 68, doi:10.3402/tellusa.v68.29879, 2016.

Spandre, P., François, H., George-Marcelpoil, E., and Morin, S.: Panel based assessment of snow management operations in French ski resorts, Journal of Outdoor Recreation and Tourism, doi:10.1016/j.jort.2016.09.002, 2016.

Sterl, A.: On the (in) homogeneity of reanalysis products, J. Climate, 17, 3866–3873, doi:http://dx.doi.org/10.1175/1520-0442(2004)017<3866:OTIORP>2.0.CO;2, 2004.

Teutschbein, C. and Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, J. Hydrol., 456, 12–29, doi:10.1016/j.jhydrol.2012.05.052, 2012.

Teutschbein, C. and Seibert, J.: Is bias correction of regional climate model (RCM) simulations possible for non-stationary conditions?, Hydrol. Earth Syst. Sc., 17, 5061–5077, doi:10.5194/hess-17-5061-2013, 2013.

Themeßl, M. J., Gobiet, A., and Leuprecht, A.: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models, Int. J. Climatol., 31, 1530–1544, doi:10.1002/joc.2168, 2011.

Thuiller, W., Gueguen, M., Georges, D., Bonet, R., Chalmandrier, L., Garraud, L., Renaud, J., Roquet, C., Van Es, J., Zimmermann, N., and Lavergne, S.: Are different facets of plant diversity well protected against climate and land cover changes? A test study in the French Alps, Ecography, 37, 1254–1266, doi:10.1111/ecog.00670, 2014.

Ummenhofer, C., Seo, H., Kwon, Y.-O., Parfitt, R., Brands, S., and Joyce, T.: Emerging European winter precipitation pattern linked to atmospheric circulation changes over the North Atlantic region in recent decades, Geophys. Res. Lett., 44, doi:10.1002/2017GL074188, 2017.

Uppala, S. M., KÅllberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., Mcnally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Quarterly Journal of the Royal Meteorological Society, 131, 2961–3012, doi:10.1256/qj.04.176, 2005.

Vidal, J.-P., Martin, E., Franchistéguy, L., Baillon, M., and Soubeyroux, J.-M.: A 50 year high resolution atmospheric reanalysis over France with the Safran system, Int. J. Climatol., 30, 1627–1644, doi:10.1002/joc.2003, 2010.

Vionnet, V., Brun, E., Morin, S., Boone, A., Martin, E., Faroux, S., Moigne, P. L., and Willemet, J.-M.: The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2, Geosci. Model. Dev., 5, 773–791, doi:10.5194/gmd-5-773-2012, 2012.

Vrac, M., Stein, M., and Hayhoe, K.: Statistical downscaling of precipitation through nonhomogeneous stochastic weather typing, Clim. Res., 34, 169, doi:doi: 10.3354/cr00696, 2007a.

Vrac, M., Stein, M., Hayhoe, K., and Liang, X.-Z.: A general method for validating statistical downscaling methods under future climate change, Geophys. Res. Lett., 34, doi:10.1029/2007GL030295, 2007b.

Vrac, M., Drobinski, P., Merlo, A., Herrmann, M., Lavaysse, C., Li, L., and Somot, S.: Dynamical and statistical downscaling of the French Mediterranean climate: uncertainty assessment, Nat. Hazard. Earth Sys., 12, 2769–2784, doi:10.5194/nhess-12-2769-2012, 2012.

Wilby, R., Wigley, T., Conway, D., Jones, P., Hewitson, B., Main, J., and Wilks, D.: Statistical downscaling of general circulation model output: a comparison of methods, Water Resour. Res., 34, 2995–3008, doi:10.1029/98WR02577, 1998.

Table 1. Variables considered in this study : Variable name, Abbreviation, Input or Output of Crocus, Units, Level and Method of ~~integration~~ aggregation (of the observational dataset from hourly to daily) and disaggregation (RCM adjusted data from daily to hourly). Variables used for the evaluation of the ADAMONT method are highlighted in bold characters. SW = shortwave, LW = longwave.

Variable	Abbreviation	Input/Output	Units	Level	Method
Temperature	Tair	Input	K	2 m	Min, max
Specific Humidity	Qair	Input	kg kg^{-1}	2 m	Last value
Wind speed	Wind	Input	m s^{-1}	10 m	Last value
Rainfall Rate	Rainf	Input	kg m^{-2} h^{-1}	Surface	Mean
Snowfall Rate	Snowf	Input	kg m^{-2} h^{-1}	Surface	Mean
Incident LW Radiation	LWdown	Input	W m^{-2}	Surface	Mean
Incident Direct SW Radiation	DIR_SWdown	Input	W m^{-2}	Surface	Mean
Incident Diffuse SW Radiation	SCA_SWdown	Input	W m^{-2}	Surface	Mean
Snowpack Depth	SNOWDEPTH	Output	m	$<$ Surface	-

Table 2. Name and description of the different configurations used in the evaluation of the ADAMONT method.

Name	Description
SAFRAN reanalysis	Simulation carried out with the SAFRAN reanalysis, over the period considered in the figures (1980-2010, 1980-1995 or 1995-2010)
RCM raw N0	Simulation carried out over the period considered in the figures, with the raw ALADIN RCM (without adjustment)
RCM raw N50	Simulation carried out over the period considered in the figures, with the raw ALADIN RCM (without adjustment), using the spatial and altitudinal RCM grid points neighbour selection technique ($N = 50$)
RCM L. 1980-1995 N0	Simulation carried out over the period considered in the figures, with the ALADIN RCM, and the learning period 1980-1995
RCM L. 1980-1995 N50	Simulation carried out over the period considered in the figures, with the ALADIN RCM, and the learning period 1980-1995, using the spatial and altitudinal RCM grid points neighbour selection technique ($N = 50$)
RCM L. 1980-1995 no corr	Same as RCM L. 1980-1995 N0, but without performing the last quantile mapping for rain and snow
RCM L. 1995-2010 N0	Simulation carried out over the period considered in the figures, with the ALADIN RCM, and the learning period 1995-2010
RCM L. 1995-2010 N50	Simulation carried out over the period considered in the figures, with the ALADIN RCM, and the learning period 1995-2010, using the spatial and altitudinal RCM grid points neighbour selection technique ($N = 50$)
RCM L. 1995-2010 no corr	Same as RCM L. 1995-2010 N0, but without performing the last quantile mapping for rain and snow
RCM L. 1980-2010 N0	Simulation carried out over the period considered in the figures, with the ALADIN RCM, and the learning period 1980-2010
RCM L. 1980-2010 N50	Simulation carried out over the period considered in the figures, with the ALADIN RCM, and the learning period 1980-2010, using the spatial and altitudinal RCM grid points neighbour selection technique ($N = 50$)
RCM L. 1980-2010 no corr	Same as RCM L. 1980-2010 N0, but without performing the last quantile mapping for rain and snow

Table 3. Mean values and scores of the ADAMONT-adjusted RCM L. 1980-2010 simulation compared to SAFRAN over the period 1980-2010 for each massif of the French Alps (massif numbers # indicated in Fig. 3) and for the Northern and Southern Alps, at 1200 m and 2100 m elevation : mean annual temperature (T, in K) and precipitation (P, in kg m^{-2} yr^{-1}), mean winter (DJFMAM) snow depth (SD, in m), mean annual bias of T and P, mean winter bias of SD, annual root mean square error (RMSE) of T and P, winter RMSE of SD, and annual correlation of T and P.

#	Massif	Altitude	Mean value			Mean bias			RMSE			Correlation	
			T	P	SD	T	P	SD	T	P	SD	T	P
	Northern Alps	1200 m	280.1	991	0.32	-0.04	-217	-0.04	0.40	643	0.11	0.99	0.92
		2100 m	275.8	675	1.25	0.03	-294	-0.03	0.50	804	0.16	0.96	0.91
1	Chablais	1200 m	279.5	1247	0.40	-0.05	-233	-0.04	0.56	1010	0.16	0.97	0.56
		2100 m	275.5	845	1.54	0.07	-313	-0.04	0.55	1222	0.27	0.95	0.52
2	Aravis	1200 m	279.8	1205	0.42	-0.02	-282	-0.05	0.47	1021	0.17	0.98	0.88
		2100 m	275.7	814	1.65	0.07	-389	-0.03	0.56	1310	0.29	0.95	0.88
3	Mont Blanc	1200 m	279.7	1104	0.35	-0.06	-232	-0.04	0.55	981	0.12	0.97	0.58
		2100 m	275.6	854	1.44	0.04	-367	-0.10	0.51	1316	0.29	0.97	0.59
4	Bauges	1200 m	279.7	1177	0.44	-0.02	-273	-0.04	0.44	948	0.17	0.98	0.90
		2100 m	275.6	751	1.65	0.07	-408	0.01	0.56	1099	0.31	0.95	0.90
5	Beaufortin	1200 m	280.1	921	0.40	-0.02	-195	-0.02	0.45	786	0.14	0.98	0.79
		2100 m	275.6	653	1.36	0.05	-291	-0.05	0.53	974	0.20	0.96	0.78
6	Haute Tarentaise	1200 m	280.3	727	0.33	-0.04	-177	-0.05	0.65	686	0.16	0.97	0.75
		2100 m	275.4	509	1.01	0.00	-199	-0.06	0.62	789	0.25	0.97	0.74
7	Chartreuse	1200 m	280.0	1225	0.37	-0.02	-303	-0.04	0.51	1070	0.21	0.97	0.87
		2100 m	276.1	761	1.57	0.07	-409	0.06	0.75	1307	0.30	0.89	0.84
8	Belledonne	1200 m	280.1	1112	0.34	-0.05	-229	-0.03	0.48	917	0.16	0.98	0.89
		2100 m	275.9	771	1.45	0.03	-314	0.05	0.66	1175	0.26	0.91	0.88
9	Maurienne	1200 m	280.4	854	0.33	-0.04	-184	-0.01	0.48	767	0.15	0.99	0.84
		2100 m	275.8	548	1.10	0.03	-241	-0.02	0.55	868	0.21	0.95	0.85
10	Vanoise	1200 m	280.4	771	0.31	-0.03	-129	-0.02	0.53	694	0.11	0.98	0.82
		2100 m	275.6	549	1.00	0.00	-186	-0.04	0.54	833	0.20	0.96	0.81
11	Haute Maurienne	1200 m	280.7	642	0.15	-0.05	-147	-0.04	0.59	693	0.10	0.97	0.87
		2100 m	275.5	487	0.61	-0.03	-185	-0.08	0.48	858	0.19	0.98	0.84
12	Grandes Rousses	1200 m	280.4	907	0.26	-0.06	-200	-0.03	0.65	902	0.14	0.96	0.83
		2100 m	276.0	591	1.06	0.00	-244	-0.02	0.76	998	0.26	0.88	0.85
13	Vercors	1200 m	280.2	1032	0.20	-0.02	-228	-0.04	0.50	768	0.13	0.97	0.89

Table 3 continues on the next page

Table 3 continued

#	Massif	Altitude	Mean value			Mean bias			RMSE			Correlation	
			T	P	SD	T	P	SD	T	P	SD	T	P
		2100 m	276.2	686	1.21	0.05	-308	-0.03	0.73	971	0.25	0.88	0.85
14	Oisans	1200 m	280.5	947	0.19	-0.03	-223	-0.05	0.49	903	0.11	0.98	0.84
		2100 m	276.2	629	0.91	0.01	-264	-0.07	0.65	1038	0.28	0.93	0.85
	Southern Alps	1200 m	281.2	775	0.10	0.03	-150	-0.02	0.49	530	0.05	0.98	0.93
		2100 m	276.4	546	0.63	0.02	-194	-0.04	0.47	646	0.15	0.98	0.93
15	Thabor	2100 m	275.9	452	0.70	0.00	-220	-0.03	0.61	868	0.20	0.96	0.87
16	Pelvoux	1200 m	280.9	733	0.20	0.00	-146	-0.01	0.75	676	0.08	0.94	0.93
		2100 m	276.2	533	0.92	0.02	-204	0.00	0.67	878	0.24	0.94	0.92
17	Queyras	1200 m	281.1	568	0.10	0.01	-138	-0.03	0.56	641	0.07	0.98	0.83
		2100 m	276.0	426	0.46	0.02	-163	-0.04	0.54	770	0.18	0.98	0.82
18	Dévoluy	1200 m	280.6	935	0.10	-0.02	-171	-0.03	0.50	784	0.08	0.97	0.86
		2100 m	276.4	633	0.77	0.05	-186	-0.02	0.66	919	0.24	0.94	0.84
19	Champsaur	1200 m	280.8	823	0.13	-0.02	-180	-0.02	0.57	705	0.08	0.98	0.90
		2100 m	276.4	580	0.74	0.01	-217	-0.04	0.57	880	0.24	0.97	0.88
20	Parpaillon	1200 m	281.1	629	0.13	0.02	-145	-0.02	0.60	644	0.07	0.97	0.87
		2100 m	276.4	467	0.54	0.02	-179	-0.03	0.52	736	0.17	0.99	0.87
21	Ubaye	1200 m	281.2	682	0.06	0.04	-132	-0.01	0.82	580	0.05	0.92	0.89
		2100 m	276.6	525	0.43	0.03	-179	-0.06	0.58	705	0.18	0.98	0.89
22	Alpes Azur	1200 m	281.7	877	0.05	0.10	-119	-0.02	0.66	728	0.08	0.94	0.78
		2100 m	277.0	590	0.53	0.03	-180	-0.10	0.48	854	0.22	0.98	0.78
23	Mercantour	1200 m	282.3	952	0.05	0.09	-168	-0.03	0.71	974	0.07	0.94	0.68
		2100 m	276.9	707	0.56	-0.02	-223	-0.06	0.61	1133	0.27	0.96	0.69

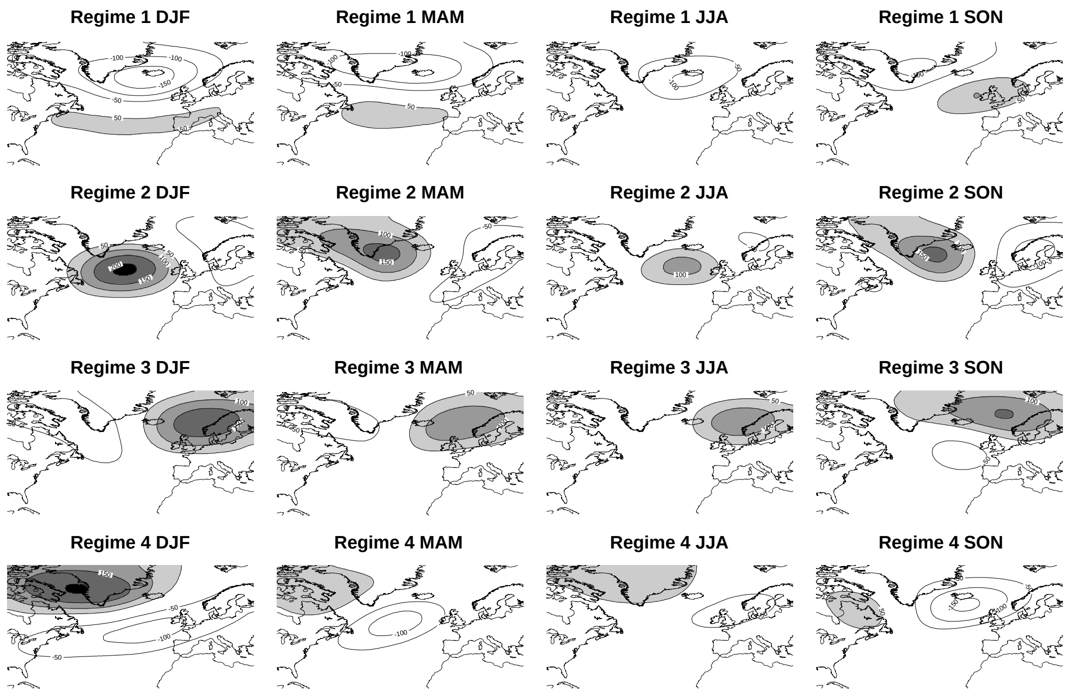

Figure 1. Clusters of each weather regime for the different seasons (winter: DJF, spring: MAM, summer: JJA, autumn: SON) used in this study: mean geopotential height at 500 hPa (m) in ERA-40 over the period 1958-2001. The seasonal climatological mean was removed. Isohypses are represented every 50 m and the zero isohypse is not represented. For readability, positive values are shaded progressively.

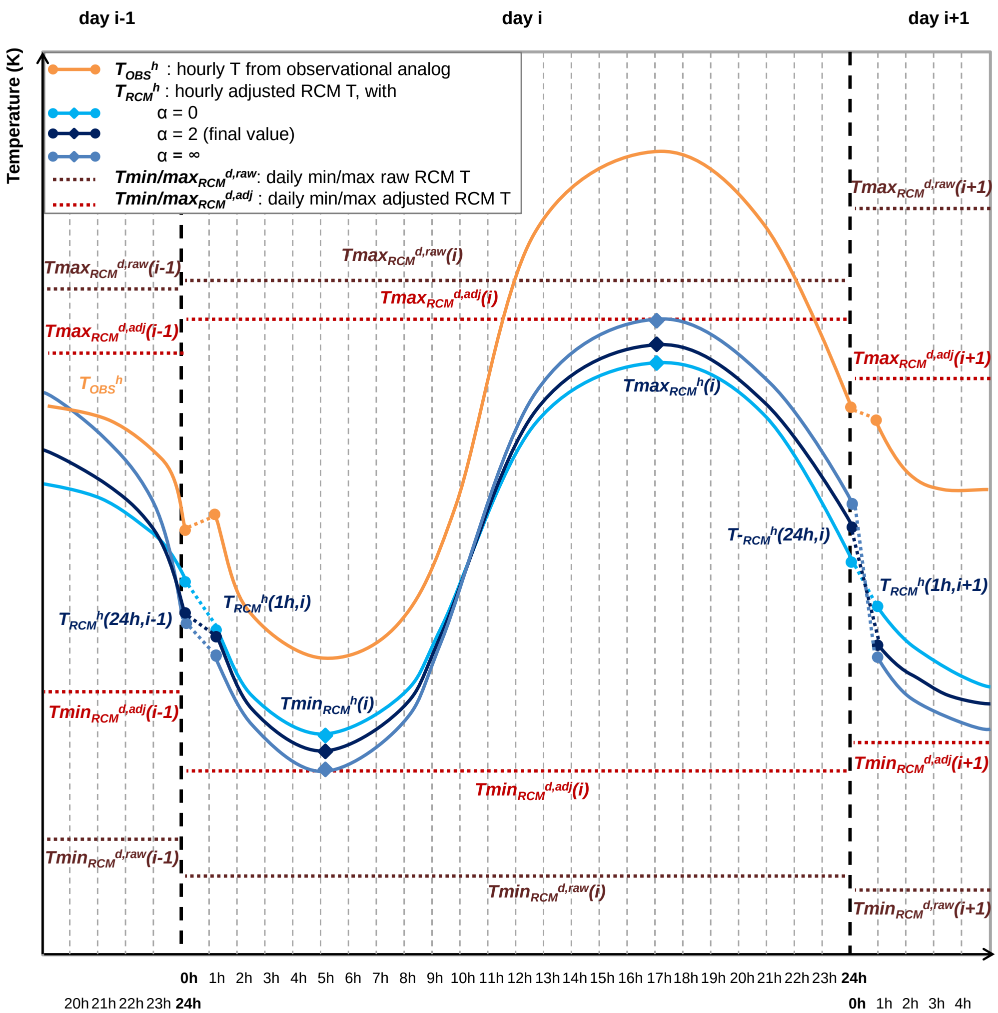

Figure 2. Illustration of the different parameters taken into account in the disaggregation of RCM temperature from a daily integration period into an hourly time step. $T^h_{RCM}(1h, i)$ and $T^h_{RCM}(24h, i-1)$ are the hourly adjusted RCM temperature values at the first time step of day i and at the last time step of the day before (i-1), $Tmin^h_{RCM}(i)$ and $Tmax^h_{RCM}(i)$ are the hourly minimum and maximum adjusted RCM temperature values respectively, and $Tmin^{d,adj}_{RCM}(i)$ and $Tmax^{d,adj}_{RCM}(i)$ are the daily minimum and maximum adjusted RCM temperature values respectively. α is a parameter which can be tuned to give more importance to the minimisation of differences between daily and hourly RCM minima and maxima. Hourly adjusted RCM temperature time series for values of α of zero, 2 and infinite are shown. T^h_{OBS} corresponds to the hourly series of the chosen daily analogue, and $Tmin^{d,raw}_{RCM}(i)$ and $Tmax^{d,raw}_{RCM}(i)$ are the daily raw minimum and maximum RCM temperature values (before adjustment).

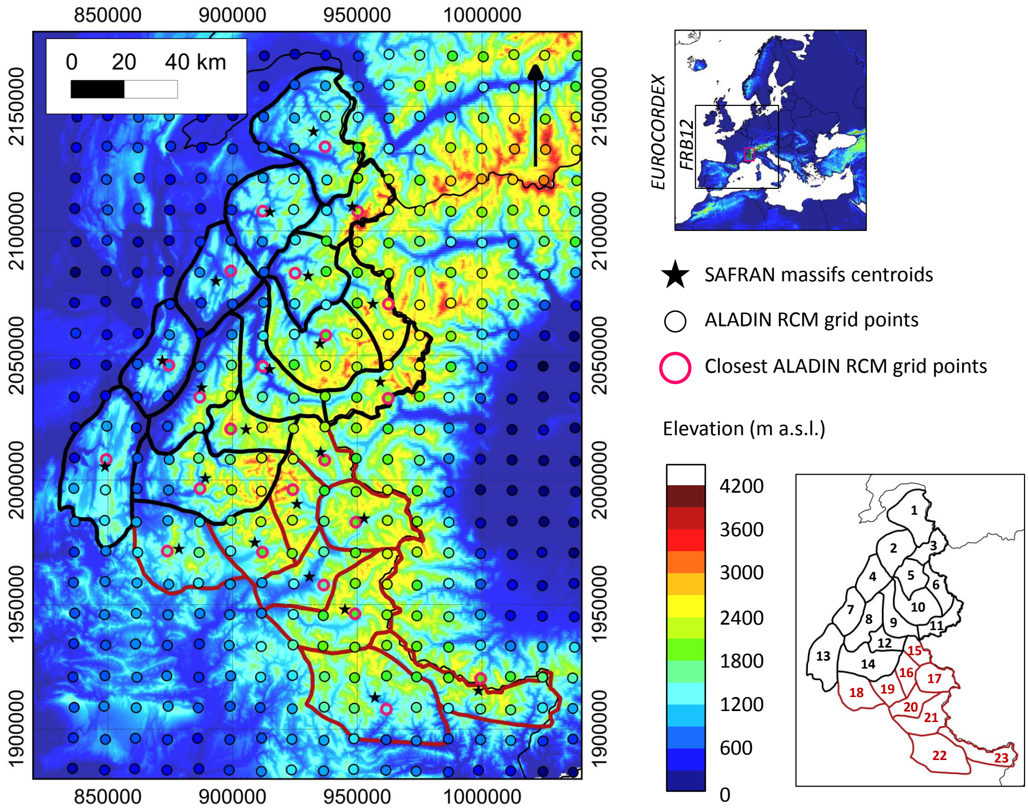

Figure 3. Description of geographical configuration of the SAFRAN reanalysis and the ALADIN RCM. The top right panel illustrates the spatial domains covered by the simulation (FRB12) and by EURO-CORDEX, and the location of the study area is indicated by the pink box. In the main panel, SAFRAN massifs are delimited by the black contours for the Northern Alps and by the burgundy contours for the Southern Alps, and their centre points are indicated by the black stars. ALADIN grid points are represented by dots, with pink contour for the grid points closest to each SAFRAN massif centre point. Surface elevation in France is from the 50m-DEM from the Institut National d'Information Géographique et Forestière (IGN) and outside France from GTOPO30 (resolution of 30 arc seconds $\approx$ 1 km). The elevation of ALADIN grid points is indicated by the color palette (in m above sea level (a.s.l.)). The bottom right panel indicates the location of each massif used in Table 3. Projection is in Lambert II étendu (L2E).

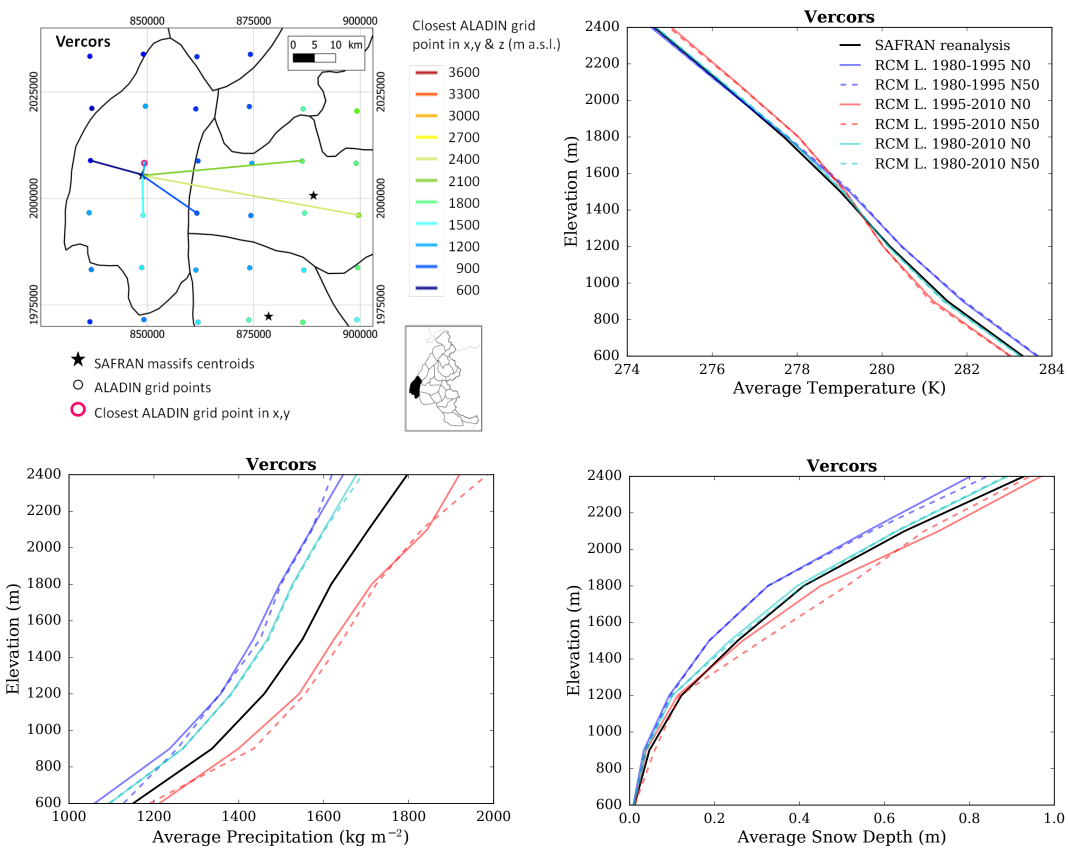

Figure 4. (top left) Location of the Vercors massif, with ALADIN RCM grid points chosen as the closest in x, y ($N = 0$, pink contour) and in x, y and z (using $N \neq 0$). Coloured lines link each SAFRAN massif centre point with the corresponding grid point in ALADIN for the different elevations considered (600-2400~~in~~ m above sea level (a.s.l.)). In this case the 1500 m and 1800 m lines are similar. (top right) Mean temperature for each elevation band over the evaluation period in each adjusted RCM simulation (different learning periods and 2 grid points neighbour selection methods) and in SAFRAN (1980-2010). (bottom left) Mean precipitation for each elevation band over the evaluation period. (bottom right) Mean snow depth (using Crocus in this case) for each elevation band over the evaluation period.

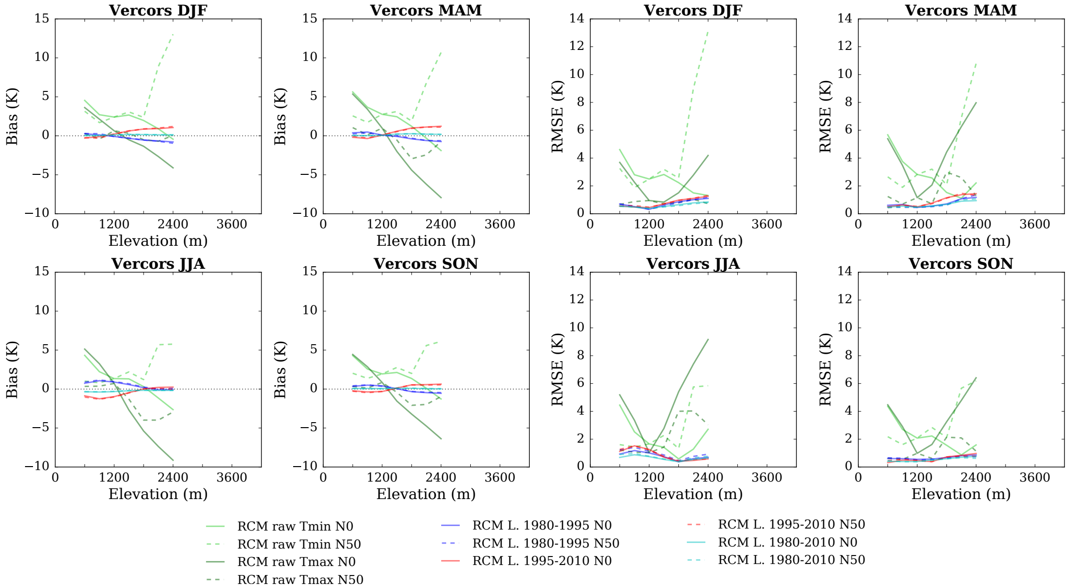

Figure 5. Temperature mean bias and root mean square error (RMSE) of each raw and adjusted RCM simulation compared to the SAFRAN reanalysis over the evaluation period for the Vercors massif as a function of elevation. Scores computed for the raw RCM simulations concern minimum and maximum daily temperatures.

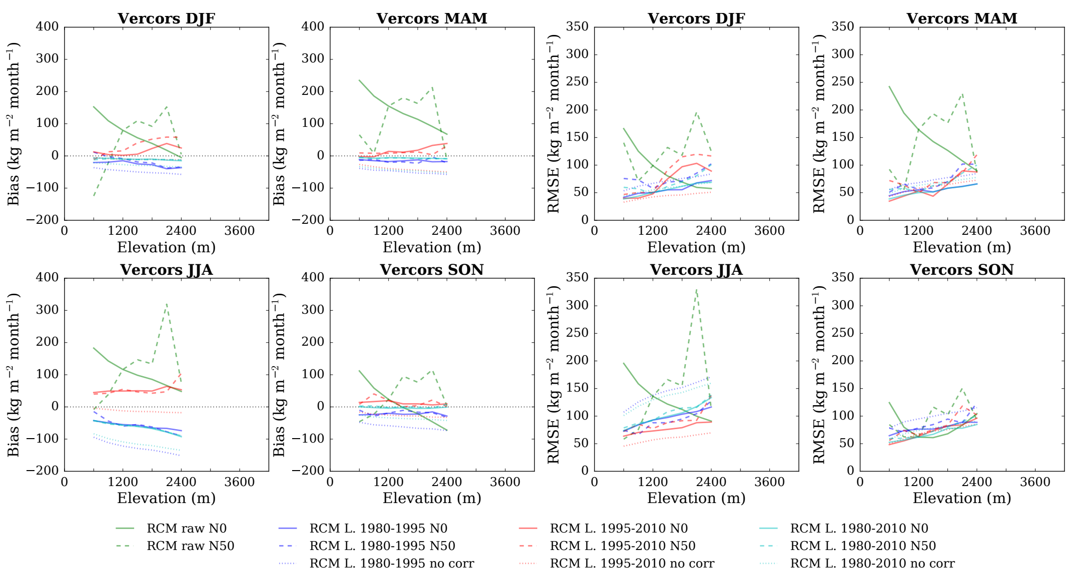

Figure 6. Precipitation mean bias and root mean square error (RMSE) of each raw and adjusted RCM simulation compared to the SAFRAN reanalysis over the evaluation period for the Vercors massif as a function of elevation.

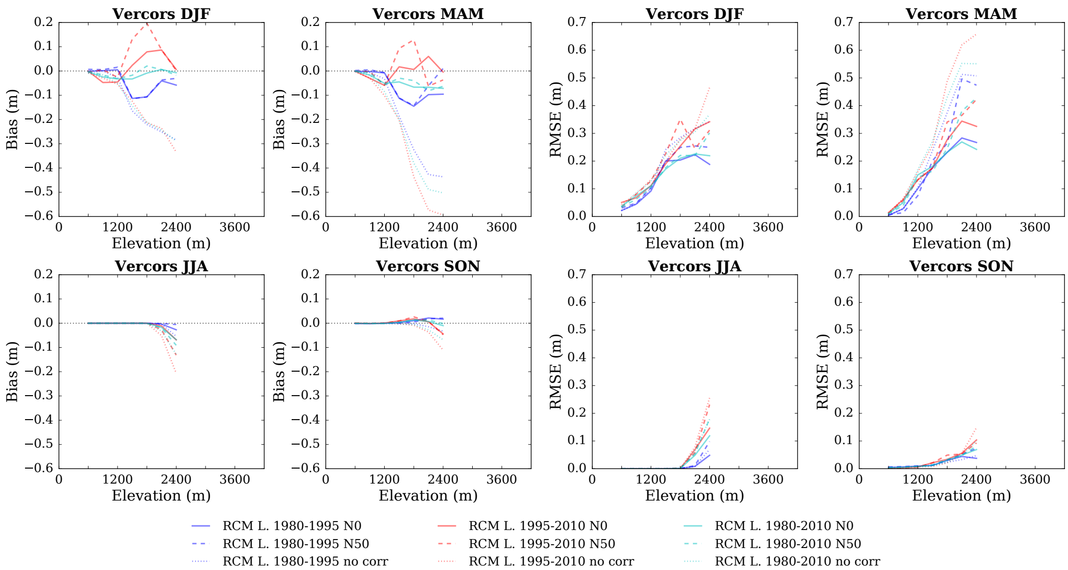

Figure 7. Snow depth mean bias and root mean square error (RMSE) of each adjusted RCM simulation (used as input of Crocus) compared to the SAFRAN/Crocus reanalysis over the evaluation period for the Vercors massif as a function of elevation.

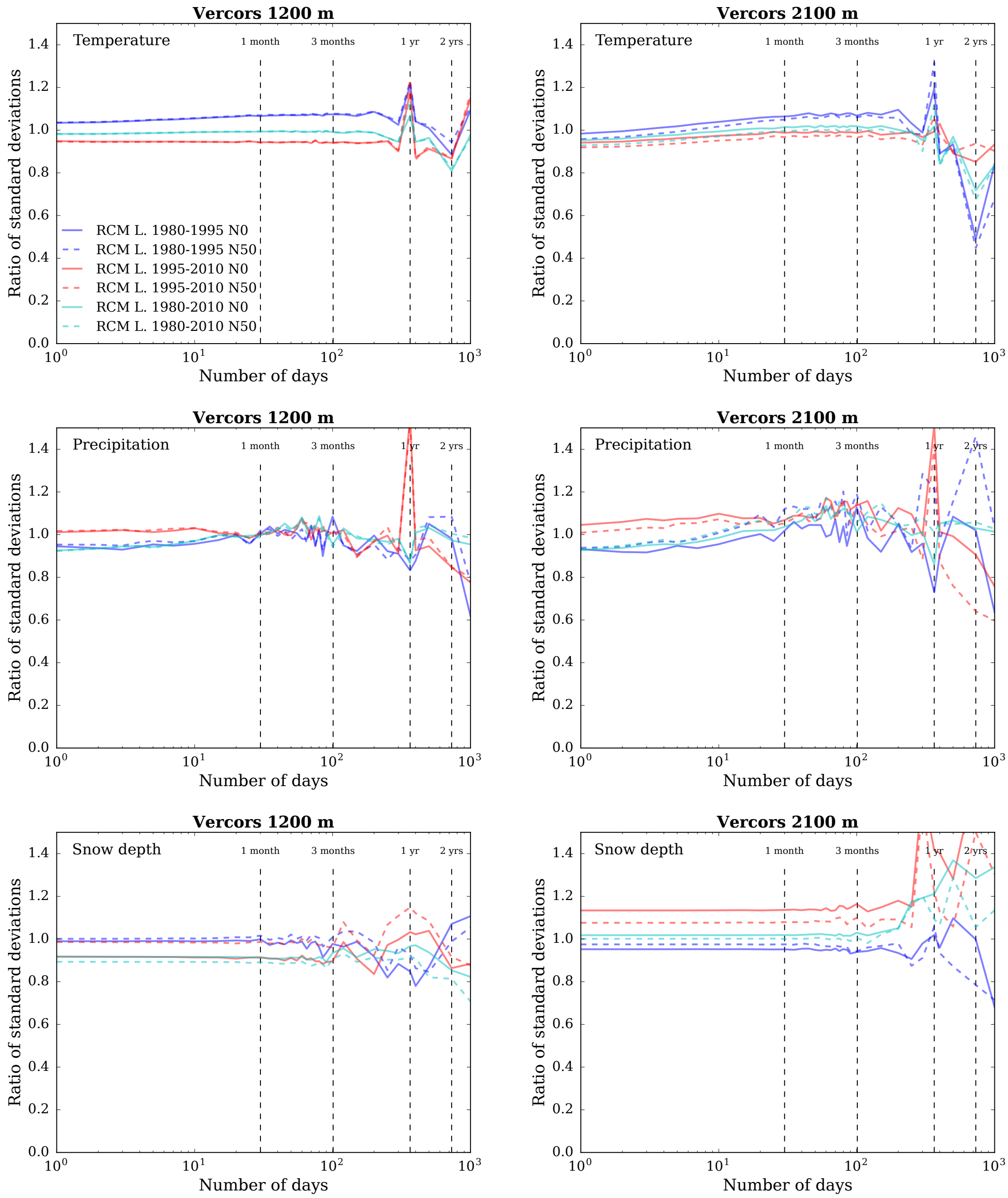

Figure 8. Ratio of standard deviations between the SAFRAN reanalysis and adjusted RCM temperature, precipitation and snow depth (using Crocus in this case) as a function of the integration window over the evaluation period, for the Vercors massif at 1200 m a.s.l. and 2100 m a.s.l..

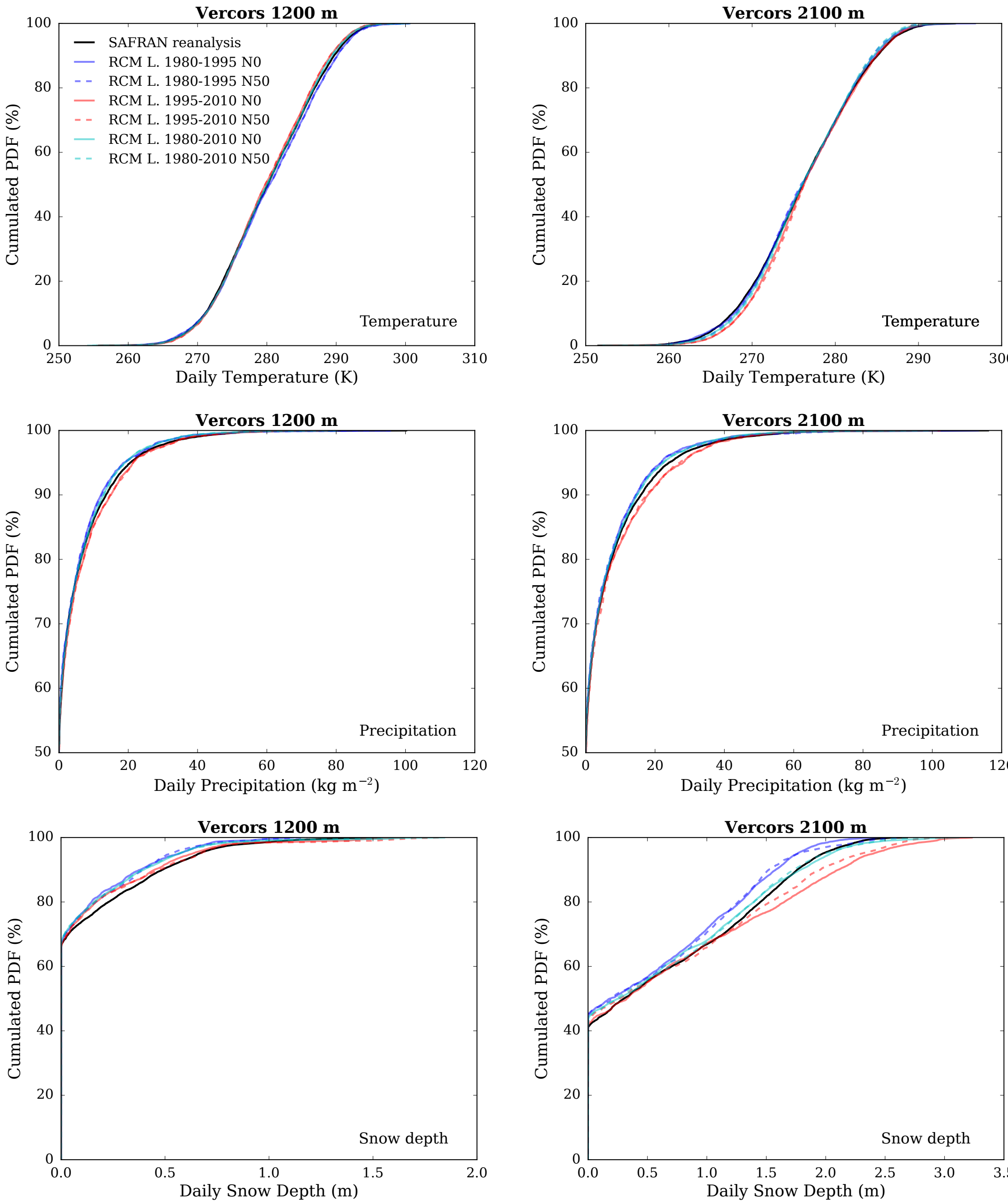

Figure 9. Cumulated probability density function (PDF) of daily temperature, precipitation and snow depth (using Crocus in this case) in each adjusted RCM simulation and in the SAFRAN reanalysis (1980-2010) over the evaluation period, for the Vercors massif at 1200 m a.s.l. and 2100 m a.s.l..

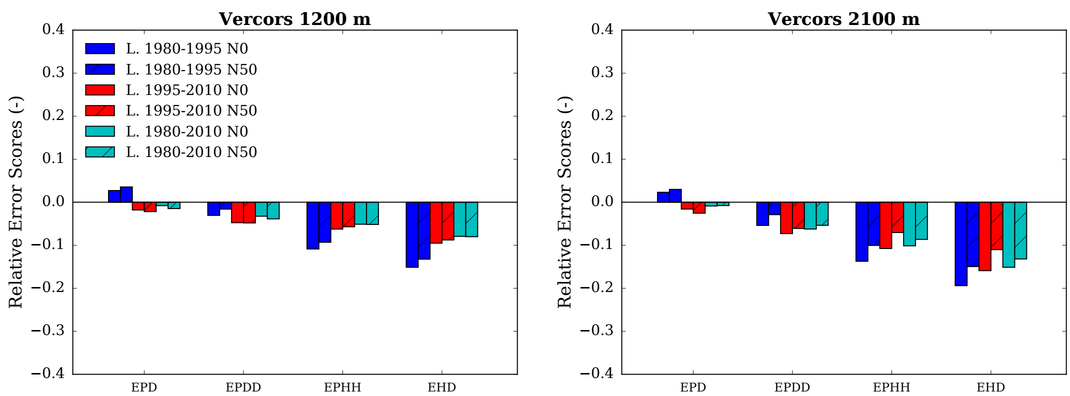

Figure 10. Scores for the duration and persistence of precipitation events in each adjusted RCM simulation compared to the SAFRAN reanalysis over the evaluation period, for the Vercors massif at 1200 m a.s.l. and 2100 m a.s.l.. EPD = relative error on the probability of a dry day, EPDD = relative error on the probability of a dry day following a dry day, EPHH = relative error on the probability of a wet day following a wet day, EHD = relative error on the mean duration of wet periods.

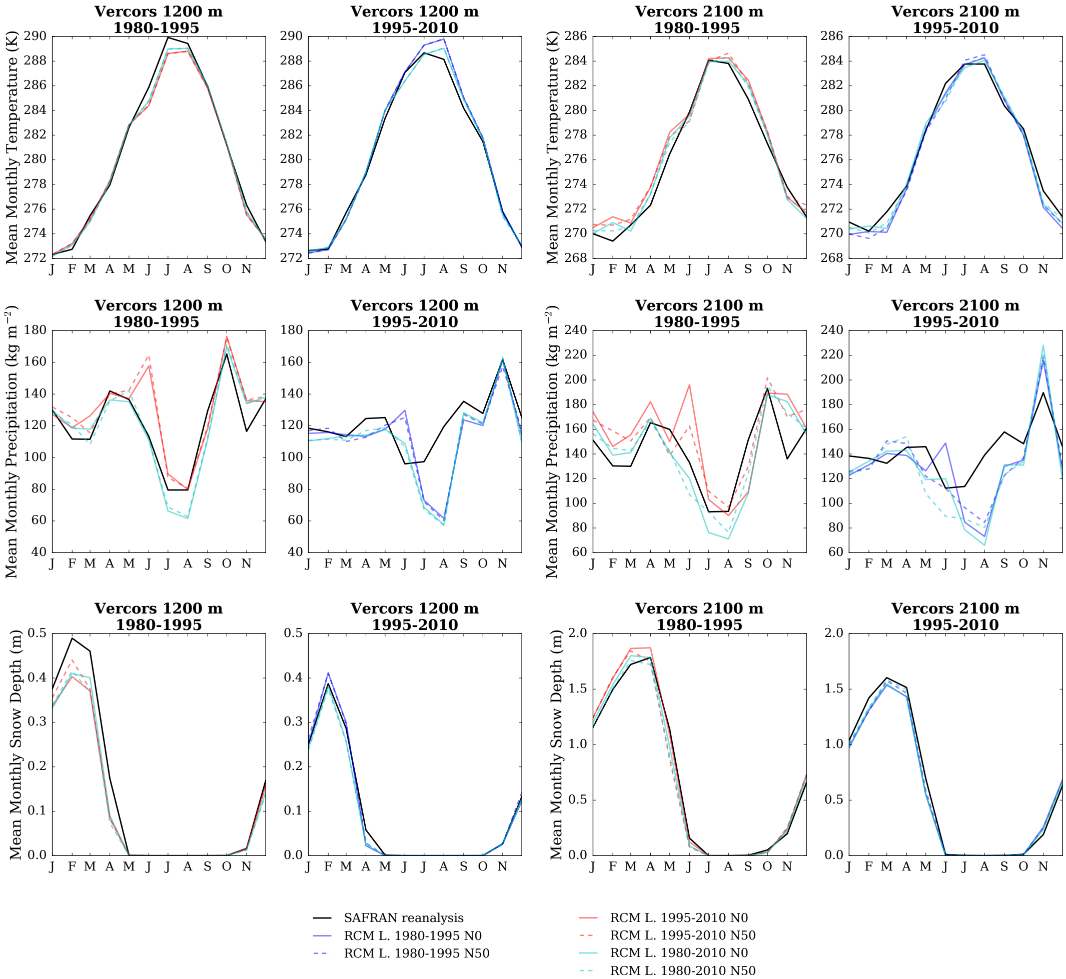

Figure 11. Mean annual cycle of temperature, precipitation and snow depth (using Crocus in this case) in each adjusted RCM simulation and in the SAFRAN reanalysis over the period 1980-1995 and 1995-2010, for the Vercors massif at 1200 m a.s.l. and 2100 m a.s.l.. Letters on the x-axis correspond to the different months of the calendar (J = January, F = February, etc.).

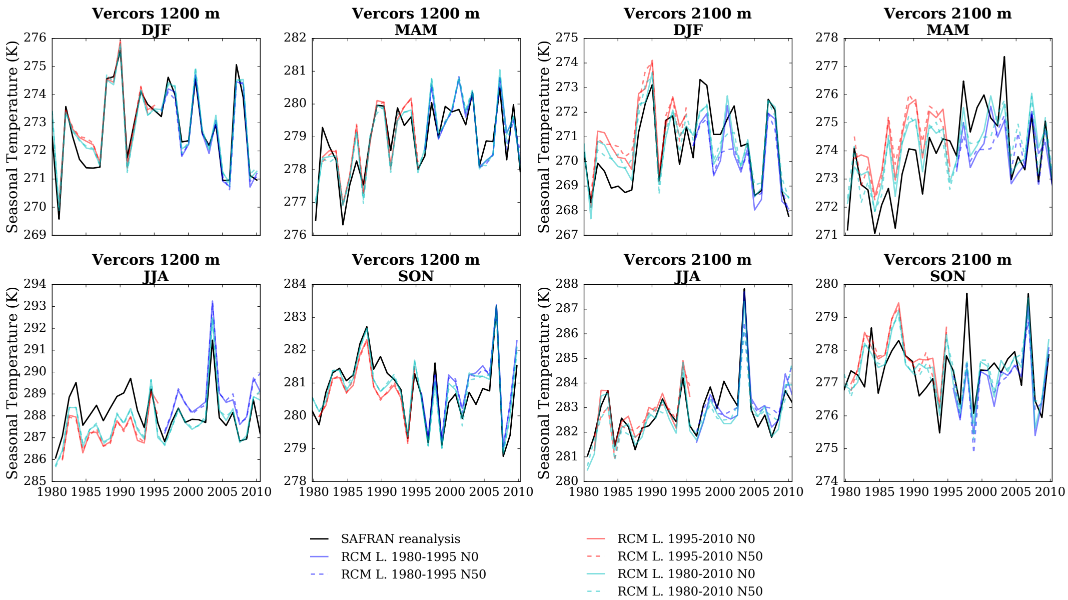

Figure 12. Seasonal average time series of temperature from 1980 to 2010 in each adjusted RCM simulation and in the SAFRAN reanalysis, for the Vercors massif at 1200 m a.s.l. and 2100 m a.s.l..

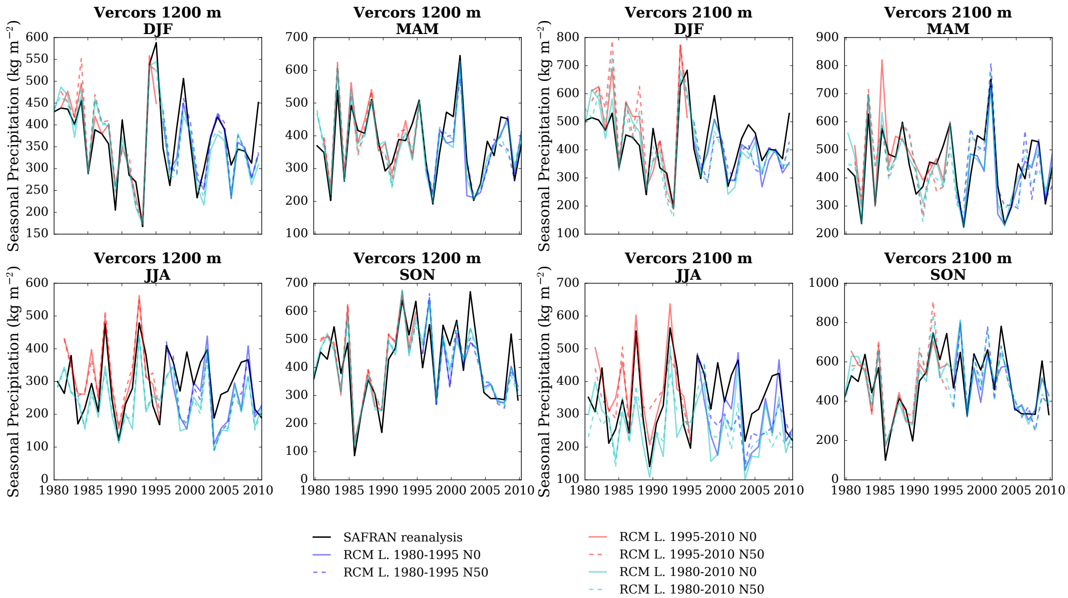

Figure 13. Seasonal average time series of precipitation from 1980 to 2010 in each adjusted RCM simulation and in the SAFRAN reanalysis, for the Vercors massif at 1200 m a.s.l. and 2100 m a.s.l..

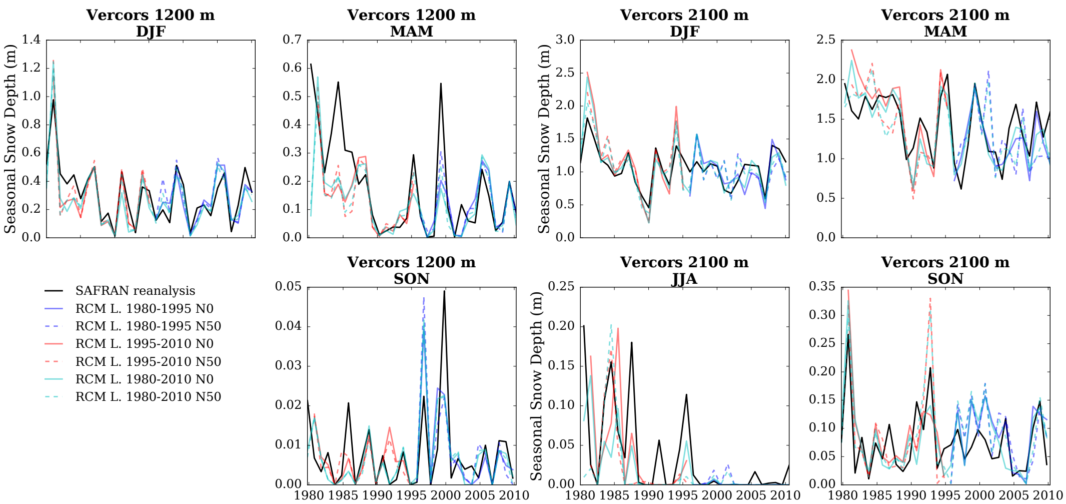

Figure 14. Seasonal average time series of snow depth from 1980 to 2010 in each adjusted RCM simulation (used as input for Crocus) and in the SAFRAN/Crocus reanalysis, for the Vercors massif at 1200 m a.s.l. and 2100 m a.s.l..

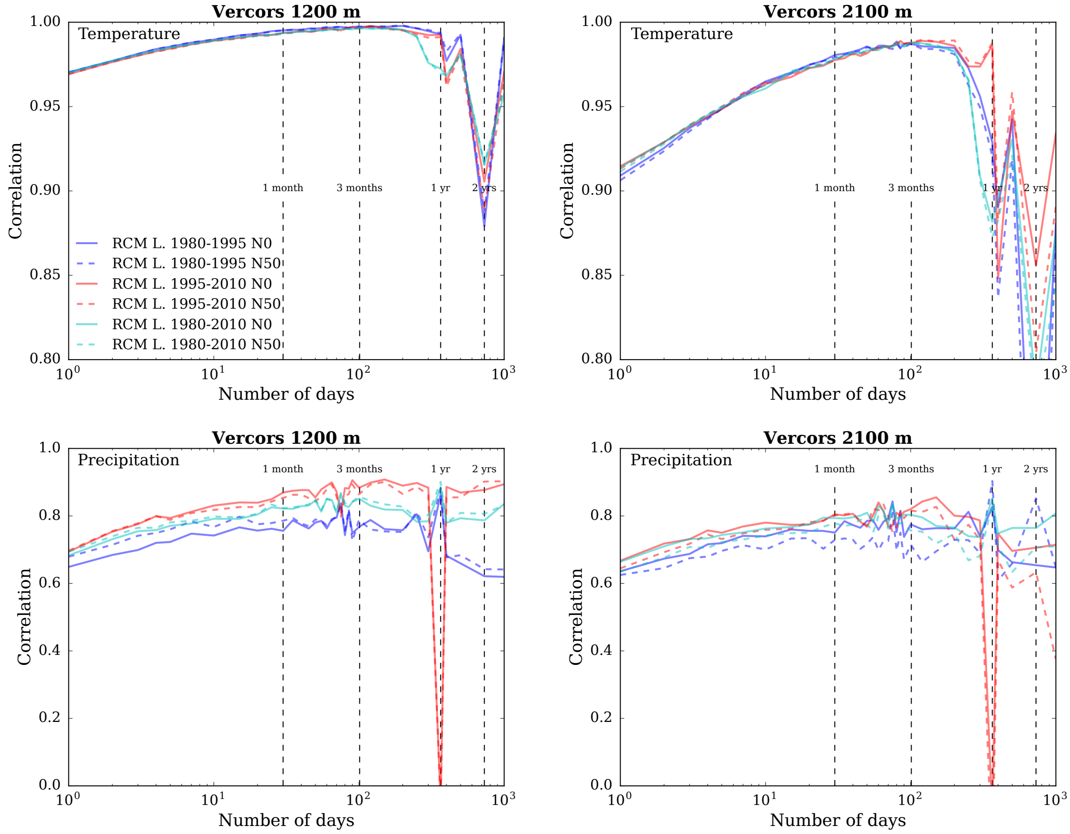

Figure 15. Correlation between the SAFRAN reanalysis and adjusted RCM temperature and precipitation as a function of the integration window over the evaluation period, for the Vercors massif at 1200 m a.s.l. and 2100 m a.s.l..

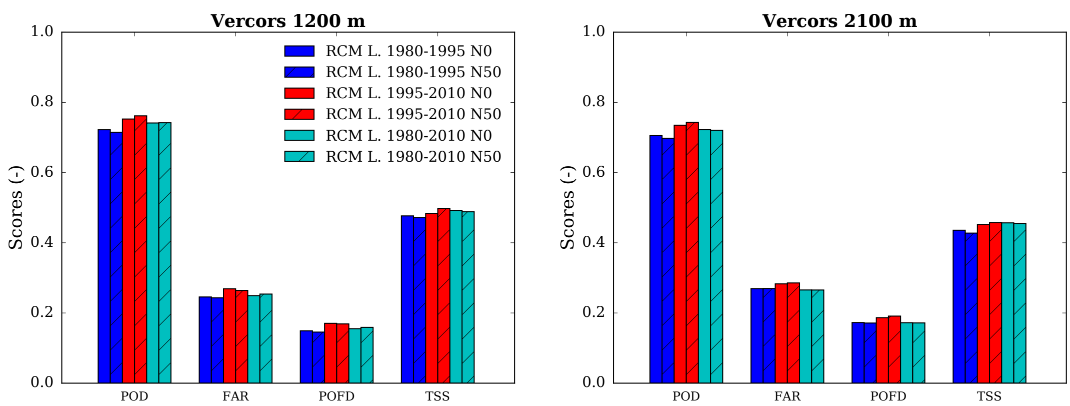

Figure 16. Scores for the detection of precipitation events in each adjusted RCM simulation compared to the SAFRAN reanalysis over the evaluation period, for the Vercors massif at 1200 m a.s.l. and 2100 m a.s.l.. POD = probability of detection, FAR = false alarm rate, POFD = probability of false detection, TSS = true skill score.