Point-by-point response to the reviewers

J. J. Gómez-Navarro et al.

May 15, 2018

Anonymous Reviewer #1:

The paper is well written and addresses a relevant scientific question by describing a promising bias correction method, based on quantile mapping (QM) conditioned to regions with similar temporal variability. It is in general well-structured and represents a substantial contribution to the modelling and impacts community.

We appreciate the positive view the reviewer expresses about this version of the manuscript. We have tried to address below the concerns he/she poses about the manuscript. We hope the new version improves the deficiencies pointed out by the reviewer.

Still there are some explanations missing to be able to understand the whole methodology and these explanations may probably answer some of my specific comments. In particular, the regions/clusters are obtained for observations and model independently, I do not understand how the bias correction is trained and applied for each grid box, since the regions are different for each dataset and a grid box may belong to different clusters in both datasets. Thus, how are the calibrated corrections obtained for a region? Which correction is applied to a grid box that belong to different regions in the model and observations?

As correctly guessed by the reviewer, this is a misunderstanding possibly motivated by insufficient or inaccurate explanations of the details of the methodology. We have carefully edited the manuscript to emphasise these details and clarify how the regions are defined.
In summary, the regions to apply the correction are those defined using the WRF-CESM simulation alone. These regions, i.e. the ones shown in the third column of Fig. 2, are the ones that condition the QM correction. As such, there is no ambiguity in the selection of to which region a given grid point belongs. The reason for the application of this approach is that the aim of the regionalisation is precisely to group regions that (miss)behave similarly, so that we correct them in a similar fashion. The alternative approach, i.e. defining regions based on the observations, would naturally lead to regions that although behave similarly in reality, contain grid points which are in principle affected by biases of different nature in the simulation, which is precisely what the condition of QM to regions tries to avoid.
One could ask, if this misunderstanding comes from including in the discussion the regions obtained in the other datasets, why did include them? We did so because, as stated in the introduction, the aim of this manuscript it twofold. Although the regions in these other datasets are not relevant for the application of bias correction, comparing the regions obtained in WRF-CESM and WRF-ERA, and those within the observations serves for validation purposes. It allows us to draw conclusions regarding the confidence we can put on the ability of these simulations to reproduce the precipitation paterns in the complex area of Switzerland.

1) If the cluster classification of the raw model data is used, this classification is based on biased data.

This is indeed the case. The regionalisation is part of the process to remove bias (actually, it is prior to it), therefore it necessarily has to operate on biased data. The regionalisation aims at identifying regions whose precipitation variability is similar, so that we can apply bias correction to such regions coherently. The rationale is that since grid points within a region behave similarly, it makes sense to remove their biases with the same transfer function, because they are arguably affected by the misrepresentation of common physical mechanisms.
A critic of this approach could be that biases might be so prominent that they affect the regionalisation itself, leading to unrealistic/unphysical regions. Fortunately this is not problematic in this case, if this is what the reviewer is arguing. Demonstrating how this is not the case is indeed what the inclusion of the WRF-ERA simulation in this analysis pursues. Figures 2 and 3 compare the results

obtained with the regionalisation of WRF-CESM and WRF-ERA. It turns out that, although the former is affected by larger biases (note that WRF-ERA is able to nicely reproduce the annual cycle, albeit with a consistent wet bias), and the temporal correlation between both runs is negligible (for being CESM a free run), the regions obtained using both datasets present strong similarities. We emphasise this important result in the manuscript:

> In summary, the regions identified in both simulations are similar and resemble the orographical barrier imposed by the Alps. This similarity demonstrates that the spatial structure of precipitation regimes are largely independent on the driving dataset.

2) If only the classification for the train/test of the QM based on observations is used, how would be the method applied in a changing climate in which the grid boxes could move to another cluster?

We believe this question is related to the misinterpretation of the methodology addressed in previous answers to this reviewer. The regionalisation is actually based on simulated data, as we have tried to clarify above, so it is not bounded to the existence of observational records, which obviously do not exist under climate change conditions.
However, the approach could be extended to consider this scenario. Under climate change conditions it is in principle perfectly possible to repeat the regionalisation for the biased climate change projection. Once the regions are found, the transfer functions can be obtained for such regions using simulated and observed data for a control period, and then apply the correction to the projected precipitation.
Still, although technically possible, more tricky is to answer to what extent this approach would lead to physical or meaningful results. In the opinion of the authors, the limitations and uncertainties that such a methodology could pose are in line with the raised concerns (now more extensively discussed in the manuscript after the suggestions of this reviewer) about the application of bias correction techniques in a climate change context.

3) Can one relate those "objective" clusters to e.g. hydrological catchments relevant for impact studies?

We have tried to relate through the discussion of the results in Section 4.1 the main features of the regions in relation to the most prominent physical characteristics of this area, as they lead to the most clear agreements between the regions found across datasets. Although a more detailed analysis of the precipitation regions, and how they can be related to actual hydrological catchments, could be in principle carried out in this context, it is in our opinion beyond the scope of this manuscript. We believe we should limit ourselves here to the development of the proposed methodology and the validation of two simulations, rather than tackling as well the analysis of the spatio-temporal variability of precipitation and how it relates to hydrological features of Switzerland.

4) How does the different number of grid boxes in each cluster affect the results? The authors may include the number of grid boxes per region in Fig.2.

This is an interesting suggestion. We have created a new table (Table 2) with the number of grid points per region. In principle, having fewer grid points leads to transfer functions not so efficiently estimated by the finite sample. However, as the sample consists of pooling all grid points that belong to the same region for the whole period, the number of data points that contribute to the estimation of the quantile-quantile curve that is responsible for the correction is large, i.e. #days within a month in the period 1979-2005 $\times$ #grid points per region. For instance, 48600 pairs of numbers populate the smallest region (Region #6 in WRF-CESM in Winter, with 60 grid points), and this number of much larger in general. We have noted this in the main text (beginning of Section 4.1)

A further concern is if the authors checked differences/improvements with respect to standard QM (without conditioning to regions). Some discussion about this would be appreciated.

We have included now a comparison with a simpler method. In particular, we introduce and discuss the results by Felder et al. (2018), being a study which applies a bias-corrected version of the simulations we present here, but carried out with a simpler method. The authors briefly evaluate the model (that study focuses on impacts, and includes several types of models), and the published figures in that reference demonstrate the modest improvement achieved with basic quantile mapping. The study by Felder et al. (2018), which is co-authored by several researchers in the present study, was under review at the time of submitting this manuscript, but it is now published and accessible online, and it can be regarded as the main motivation to develop the new methodology we present here. Therefore, we have included various references to this work in the discussion of the results through the new version of the manuscript.

Here I list some specific comments and typos, giving the page and line numbers.
Specific comments:
P1 L6 "minimise disturbances to the physical consistency" -> not clear, please rephrase or elaborate.

We have rephrased this sentence, as marked in the document that highlight the differences. In the following, we do not discuss in detail the changes, but we invite the editor/reviewer to check such document to evaluate if the changes satisfy the reviewer queries.

P1 L16 which variables? So far only precipitation was mentioned (also in the title). If the clustering depends on the variable, why does the method preserve the physical consistency among variables more than the standard QM?

We restrict the sentence to precipitation.

P2 L3 The authors may consider citing the newer analysis including EURO-CORDEX data: Rajczak, J. and C. Schar (2017), Projections of future precipitation extremes over Europe: a multi-model assessment of climate simulations | J. Geophys. Res. Atmos., doi:10.1002/2017JD027176.

This reference has been included and commented.

P3 In the review of bias correction methods, the authors may consider the following paper, with some similarities from a technical point of view, where the bias correction is conditioned to circulation types: Wetterhall, F., Pappenberger, F., He, Y., Freer, J., and Cloke, H. L.: Conditioning model output statistics of regional climate model precipitation on circulation patterns, Nonlin. Processes Geophys., 19, 623-633, https://doi.org/10.5194/npg-19-623-2012, 2012.

This reference has been included and commented.

P3 L15 After this paragraph I suggest to include a sentence mentioning the implications in the climate change context, something like "As a consequence, the climate change signal might be unrealistically modified", as stated e.g. by:
Casanueva, A., Bedia, J., Herrera, S., Fernández J. and Gutiérrez J.M. Di- rect and component-wise bias correction of multi-variate climate indices: the per- centile adjustment function diagnostic tool. Climatic Change (2018) 147: 411. https://doi.org/10.1007/s10584-018-2167-5 Teng J, Potter NJ, Chiew FHS, Zhang L, Wang B, Vaze J, Evans JP (2015) How does bias correction of regional climate model precipitation affect modelled runoff? Hydrol Earth Syst Sci 19(2):711–728. https://doi.org/10.5194/hess-19-711-2015

Such sentence and references have been included.

P3 L16-20 The authors may consider the above paper (Rajczak and Schar 2017) to update that summary of previous works.

This reference has been included.

P3 L20-21 what do the authors mean with "similar"? different model version? Parameterizations?

We refer to what the authors show in this publication: same model configuration, just different spatial resolution. We have edited the text accordingly.

P3 L16-25 I would suggest to move the entire paragraph before the previous one, in which bias correction is introduced, since it reads better after line 3 and here it is again about previous studies in which bias correction is not applied. Also the final lines of the paragraph (23-25) are more or less repeating what it is already said in P2 L34.

We agree with this suggestion, so we have swapped the paragraphs.

P5 L24 Is there a reason for using 27 years instead of e.g. 30?

The reason is availability of global data to drive the RCM simulations. The ERA-Interim and CESM runs span different periods of time, but for comparison and validation purposes we focus the analysis on the overlap of both datasets. The ERA-Interim period starts in 1979, therefore this provides the lower bound. The CESM simulation runs up to 2005 (starting in 1850), which provides the upper boundary. We have carefully rephrased this in section 2.5 to explain this detail. The WRF-ERA simulation is actually longer, but we did not include the 2006-2013 period for consistency with WRF-CESM.

P7 L7 I suggest to add "smooths out the transfer functions prior to the correction".

We modified the text accordingly.

P7 L9 Until now it is not clear which is the analysis domain, the title says Alpine region, simulations are performed for the Alpine region but observations are available for Switzerland. Please consider to mention Switzerland explicitly in the experimental design and title.

We acknowledge that the difference between the area that is downscaled (the entire Alpine area) and the one we analyse in detail due to available observations (Switzerland) was not clear enough in the first version of the manuscript. Therefore, we have emphasised this difference in the new version in sections 1, 2.4 and 6.
Regarding the title, we believe that including "Switzerland" in it seems to artificially limit the scope of our analysis, as both the simulations we present here, as well as the bias correction techniques, are not limited to this country. Therefore, we have opted for *"A new region-aware bias correction method for simulated precipitation in areas of complex orography"*

P7 L16 As mentioned before, the authors mention the preservation of physical consistency. My question now is how coherent is the method in a multivariate case? I guess a different division in clusters would be performed for each variable. Can the authors comment something on this?

We have adjusted the sentence to limit the scope of what physical consistency is meant here. As we are dealing with a single variable, breaking the physical consistency would imply in this context that we correct the precipitation in a way that it breaks down the spatial and temporal structure of variability of this single variable. For instance, corrections could smooth out differences in precipitation in opposite sides of a mountain, therefore destroying part of the added value of the high-resolution simulation, or even breaking conservation laws (of mass, in this case) implicit in the simulation.

P7 L21 "varies per season" why seasonally? In section 3 it is said that the method is applied for each month, thus one expect to have different clusters at the monthly scale.

The reviewer is right, and this is a detail that was not very clearly explained in the first version of the manuscript. The correction is indeed applied to each month separately to efficiently account for the annual cycle. Therefore, in principle the regionalisation, which is a prerequisite for obtaining the transfer functions that conform the correction, should be carried out on a monthly basis as well. However this has some drawbacks. First, the computational cost of the clustering is relatively high: 3×12 regionalisations should be carried out, each of which including a previous EOF analysis and the clustering of the resulting temporal series of principal components. Second, it leads to 36 maps that we should show and analyse. But eventually, the details that these 36 maps provide are limited, as there exist great redundancy across the annual cycle, because months that belong to the same season behave similarly in terms of precipitation. Having this into account, we decided to apply a simplification that consists of performing the regionalisation on a seasonal, rather than a monthly basis, and then using the same regions for the three months within each season. This simplification does not impact the outcome of the correction, while it optimises the computational cost and reduces by a factor of 3 number of maps needed to show and have into consideration to discuss the similarities and differences among datasets and through the annual cycle that is supported by Fig. 2. Therefore, we have added a whole new paragraph to Section 3:

> We note that the application of this methodology implies a previous regionalisation of the series for each month separately, which in general involves notable computational cost. Further, months belonging to the same season behave similarly, so that the resulting regions are hardly distinguishable and the analysis presents some level of redundancy. For these reasons, we propose a simplified form of the methodology, which we apply hereafter, and consists of carrying out the regionalisation on a seasonal basis. Once identified, these regions can be regarded as representative and common for the three months within each season, so that the final correction can be applied on a monthly basis.

P8 L13 The authors mention several times the insufficient effective resolution of the observations, what about the effective resolution of the simulations? The authors should include it in the discussion as well.

Perhaps "insufficient" is not an appropriate term here, as it suggest an absolute measure. In this part of the manuscript we compare the results obtained from the WRF simulations with the observational product, and try to attribute the differences to physical mechanisms. The difference in the spatial resolution is an obvious candidate, as the effective resolution of the gridded product acknowledged by its authors is about one order of magnitude lower than the one implemented in both simulations. The "effective resolution" in RCM simulations is difficult to stablish, but it is generally accepted to be between 2 and 4 times larger than the spatial resolution, depending on the variable (Pielke Sr, 2013). Therefore, the spatial resolution of the gridded product is lower than the one in the simulations, and precludes it from capturing the finer orographic features of the domain of study, especially over mountain tips, which in turn can be resolved to a generally larger extent by the simulations presented in the manuscript. Therefore, a word that suggest the comparison of relative measures is perhaps more adequate. We have rephrased this sentence and included the comment about effective resolution of RCMs accordingly.

P8 L15-35 The authors should motivate better the correlation analysis in Fig.3. I do not see the point of this analysis, especially since the clusters are built in a way that the differences among clusters are maximized. Moreover, the clusters are different in each dataset, so there is not a clear correspondence. This lack of correspondence is only mentioned and resolved in Fig.4.

The aim of the correlation analysis is to provide a quantitative assessment of the level of similarity/dissimilarity among regions. Note that the regionalisation step will always produce a number of regions, but how different these regions really are is something that can not be answered by looking at Fig. 2 alone. The correlation analysis shown in Fig. 3 tries to overcome this caveat by providing numbers that allow to better judge objectively the coherence of the regions.
We have rephrased this introductory paragraph to motivate better this analysis.

P9 L31 "averaged over Switzerland" Given the differences in the annual cycle among the regions, the authors may consider doing this analysis per cluster, based on the observations or the WRF-ERA classifications.

Although we believe the reviewer makes here a valuable suggestion, we have been carefully considering how to account for it. Finally we have decided to leave the figure as is for a number of reasons, including:

- It is not clear which regions we should use. Figure 2 shows how each dataset leads to three possibilities, and they are all in principle equally valid for such an analysis. Should we consider and show all possible variations?
- For each dataset, the regions vary through the annual cycle. Should we change the way the average is calculated through the annual cycle? This would complicate the way the results have to be read.
- Considering all possibilities and showning them, we would end up with a very complex figure with tens of bars which would make the reading and interpretation of the figure difficult.
- The aim of this figure is to illustrate, in general terms, how precipitation varies through the annual cycle across datasets. The point of the figure is to show the consistent overestimation of precipitation in WRF-ERA, as well as the seasonality isues of WRF-CESM. For a spatially disaggregated version, that illustrates the different behaviour across the domain, the reader can get further insight in Figure 6.
- We believe it is important achieve a certain level of consistency with similar studies that facilitates their inter-comparability. In this regard, Figures 5 and 6 are, in their current shape, easily comparable to similar results in other references suggested by the reviewers, as indeed we do in the new version of the manuscript (e.g. Torma Csaba et al.; Fantini et al.). Changing this would dificult or make imposible such a comparison.

P10 L1-8 The underestimation of precipitation in the Ticcino during autumn is worth to mention. Can the authors give a reason for this?

We have discussed this issue through personal communication with other researchers working with WRF in the same target region,. They have also found a negative precipitation bias over Ticino in summer, and they traced it to a negative moisture bias in the lowest layers. However, to thoroughly test this, we would need to step by step check if WRF represents all processes detailed below correctly, in a comprehensive analysis of this bias and its causes which is beyond the scope of the manuscript. Still, we have included a brief discussion of the possible causes in the main text.

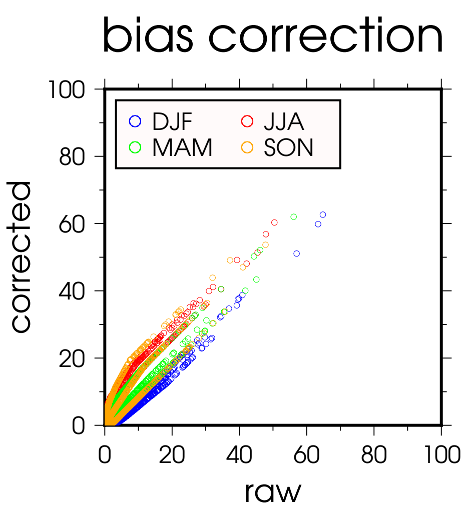
> Isotta et al. (2014) show that in the region of Ticino up to 70% of the yearly precipitation accumulation is due to the top 25% of the wet days, so it is sensible to assume that the bias stems from high to extreme precipitation events. In Ticino these heavy precipitation events are driven by the transport of moist and potentially unstable (moist neutral stratification) air masses against the Alps from the south (Martius et al., 2006; Froidevaux and Martius, 2016). Locally, the vertical shear between south-easterly flow near the surface and southerly to southwesterly above 850 hPa leads to moisture convergence and repeated formation convective cells (Panziera et al., 2015). On an even more local scale, strong vertical shear can result in small-scale circulation that results in local precipitation maxima (Houze et al., 2001). Therefore if the RCM fails to capture any of these local and highly driven by the orography processes properly, it will result in an underestimation of the precipitation.

P10 L12 The authors should also explicitly mention in the methods how the precipitation frequency is adjusted by this method (relevant for the interpretation of Fig.7). Standard QM is able to correct for a higher frequency of wet days in the model, but the opposite problem (here shown in Fig. 7, winter) could be corrected by applying the frequency adaptation, otherwise an overestimation of the wet day frequency is found in the corrected data. See : Themeßl, M.J., Gobiet, A. & Hein- rich, G. Empirical-statistical downscaling and error correction of regional climate mod- els and its impact on the climate change signal. Climatic Change (2012) 112: 449. https://doi.org/10.1007/s10584-011-0224-4

We do not use frequency adaptation techniques. This indeed leads to the wet bias in the corrected precipitation in winter, as pointed out by the reviewer. We now explicitly acknowledge this limitation and point out how it could be a suitable solution to this bias.

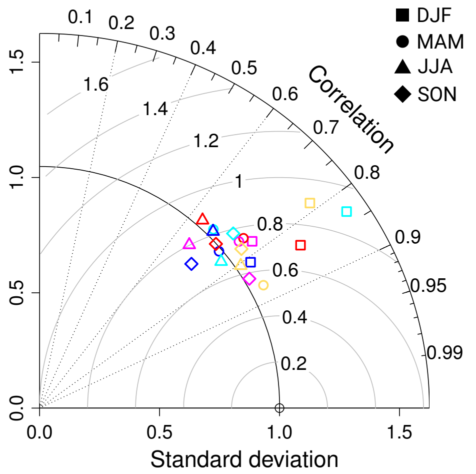
P12 L4 Why are the temporal correlations lower in autumn? This may be related to the way the corrections are trained and applied.

This is motivated by the variability of biases within this season, and is related to the cancellation of biases in the intermediate seasons further discussed in response to reviewer #2 in the context of Fig. 7. The figure below shows the dispersion map of the raw versus corrected series in a single grid point where correlation is low (i.e. a grid point within one of the red spots in Fig. 8). For each season, the points line up around three different curves, which are the transfer functions for each of the months within the season. The large spread Autumn (orange) is evident, especially when compared to the homogeneity of the three hardly distinguishable curves for winter (blue) and summer (red). This spread deteriorates the linear relationship between raw and corrected data, and therefore reduces the correlation. Although we do not show this figure in the manuscript for the sake of brevity, we have introduced a short explanation of this effect in the main text.



P22 Fig.4 The decimal dots are missing in the labels of the Taylor diagram. And more important than that, it is completely unclear to me what is shown by the angular scale (azimuthal angle). I *would expect to have represented there correlation values but that legend must be something else. Please explain in the caption how this should be interpreted.*

This is a standard Taylor diagram, where the angular scale represents correlation. The missing points (and labels!) were not missing in our original manuscript. Unfortunately some technical issue in the conversion to produce final uploaded document seems to have removed some of the information in the figure. We include below the figure as it was supposed to be included in the manuscript.



We believe this figure does not lead to the issues raised by the reviewer, and certainly we will be more careful in the final submission.

Technical corrections:

All technical corrections have been implemented as suggested by the reviewer.

Anonymous Reviewer #2:

This paper presents a bias correction method for regional climate simulations over the Alps at very high resolution. A observational database for the region is used for the validation, and ERAinterim and GCM-CESM forcing fields are used to WRF modelling work. To my opinion, it shows enough aspects to novelty and adequate analysis and understanding of the obtained results. I suggest it to be considered for publication, once the questions and requested item can be properly answered or at least taken into account in some way

We thank the reviewer for the time devoted to carefully read the manuscript, the positive vision expressed about it, and the constructive comments that will certainly improve the final version. We have tried to answer point by point all his/her comments below.

1. Missing references. It is always the case that not all the relevant references are included when a work is presented. Here I find some that I consider that are essential to be included, not only for the introductory aspects, but also for the methods and results description. Let me indicate them to the authors for them to be considered a properly used throughout the text

The new version of the manuscripts includes many more references, including most of those suggested by both reviewers, and that clearly allow to better contextualize this piece of work in the existing literature.

2. Apart from the pure bibliography missing items, there are some aspects that could be more deeply described by the authors. One of them should be to compare the proposed bias correction method with other similar ones, if there are some, to see more clearly differences and similarities with others already proposed. I am sure the quantile mapping procedures have been used before, if one goes to those references. Therefore, I recommend the ongoing work by Nikulin and others in the frame of EuroCORDEX activities, named BCIP. Take a look at this abstract at EGU2015: Nikulin, G., Bosshard, T., Yang, W., Bärring, L., Wilcke, R., Vrac, M., ... & Fernández, J. (2015, April). Bias Correction Intercomparison Project (BCIP): an introduction and the first results. In EGU General Assembly Conference Abstracts (Vol. 17). In a more general sense, perhaps a mention to this recommendation by CORDEX community could be made. take a look at http://cordex.org/data-access/bias-adjusted-rcm-data/, and from there, to a IPCC work focused on this topic: See Breakout Group 3bis: Bias Correction (pp. 21-23) in IPCC, 2015: Workshop Report of the Intergovernmental Panel on Climate Change Workshop on Regional Climate Projections and their Use in Impacts and Risk Analysis Studies [Stocker, T.F., D. Qin, G.-K. Plattner, and M. Tignor (eds.)]. IPCC Working Group I Technical Support Unit, University of Bern, Bern, Switzerland, pp. 171. (https://www.ipcc.ch/pdf/supporting-material/RPW_WorkshopReport.pdf). I can imagine that authors do not want to go too far on this aspect, but I think that some more comments, to have this work inside the wider context, should be made. Even a mention to some developed software for this kind of analysis could be included, such as Bedia, J., Iturbide, M., Herrera, S., Manzanas, R., & Gutiérrez, J. (2017). downscaleR: an R Package for Bias Correction and Statistical Downscaling. R Package Version 2.0-3.

We have included a discussion of the issues raised by the reviewer about on-going debate about the use of bias correction, including the mentioned references in the new version of the manuscript.

3. I am not sure if the authors have a comment about the fact that this bias correction method has been applied to a region with a very deep orography, and to precipitation field. Which could be the potential to apply it to other regions with smoother orography, and/or to other variables?

We have added a new paragraph in the conclusions to discuss how this method can be exported to other regions/variables. We reproduce it below:

> We note that the rationale of the developed methodology is to divide a large domain into smaller subregions according to the behaviour of the target variable. We have applied it here to daily precipitation in Switzerland for being a variable strongly affected by complex orographical details that lead to strong horizontal gradients. With more generality, spatial regionalisation is an efficient method to break down complexity in areas and variables whose behaviour strongly varies through the domain. Still, the bias correction applied separately to subregions can be in principle adapted to other cases with simpler topography, or other variables with lower horizontal gradients. The only practical difference is that in this case the regionalisation will naturally lead to a lower number of subregions which are necessary to obtains clusters with coherent features.

Specific comments

1. It has been indirectly mentioned on the general comments section, but here I want to comment if explicitly: I miss a mention to the EuroCORDEX/MedCORDEX activities, that have used plenty of simulations at high resolutions (0.11) over Europe, and several studies with not a single RCM as here, but an ensemble of them, that have analyzed, also forced with ERAinterim fields, how precipitation is described. I do not mean a full comparison with other RCMs, but at least some mention and comparison with them, to see more clearly if WRF-RCM is similar to the state-of-the-art RCMs modelling alpine precipitation for current climate conditions.

We have added plenty of references and explicit mentions to EURO-CORDEX and MED-CORDEX activities in the newer version.

2. And also related to this point, I miss some comparison of your figure 5, for example, with figure 2 of Torma et al., 2015 or Fantini et al., 2016, figure 5, not only for RCMs, but also for observational datasets, I am not sure if they are totally consistent. Or for your figure 6 and 7, and their corresponding figures.

This is an important pitfall that has been corrected in the new version. We have enlarged the discussion of the results in Section 4.2 with explicit mention to the ones in the corresponding figures in the two sources pointed out by the reviewer. However, we have not included a discussion comparing the results about daily PDFs in Fig. 7. The reason is that the daily PDFs in those references include all seasons, and they are built to emphasise the different results across spatial resolutions. Therefore they are somewhat different figures, and it is difficult to stablish a fair and meaningful comparison.

3. I have a concern about the domain of study chosen here. On figure 1, D4 subdomain seems to be the one used for the analysis, but then figures with the political borders of Switzerland seem to be used. This relatively artificial borders could add some non-physical or modelling aspects to the analysis, and specially when obtaining the subregions from the clustering procedure. Which is the opinion of the authors about this aspect?.

This confusion between the simulated and analysed domains (the Alpine region vs. Switzerland) has also been pointed out by reviewer #1. We believe we were not clear enough in the former version in the description of the dataset and the methodology. Therefore we have clarified this in the newer version of the manuscript. The reason for using Switzerland, i.e. a political boundary, instead of a natural one, is that the observational product we used is limited to this domain. This imposes a unavoidable bottleneck of the validation. Certainly there are observations beyond the borders of Switzerland, but we believe that they do not contain the high density, even distribution, and quality-tested of the observations blended to create the gridded product developed by Meteoswiss.

4. Another point I would like to hear from the authors is about the very high resolution used for the WRF D4 domain (page 5, line 14): 2km. Which one is the real advantage here of using such resolution compared with the even-very-high 6km one?. It seems that no much mention or usefulness is made by the authors to this resolution, by far much larger than the mentioned 0.11 "high resolution" EuroCORDEX standard values these days. It is also a tricky aspect, since the comparison and bias correction method is made against the roughly 20km observational dataset information, and so some statements are made through the text related to this resolution differences. A more complete study should perhaps include at least some other resolution from the WRF model to a better understanding of the resolution topic?.

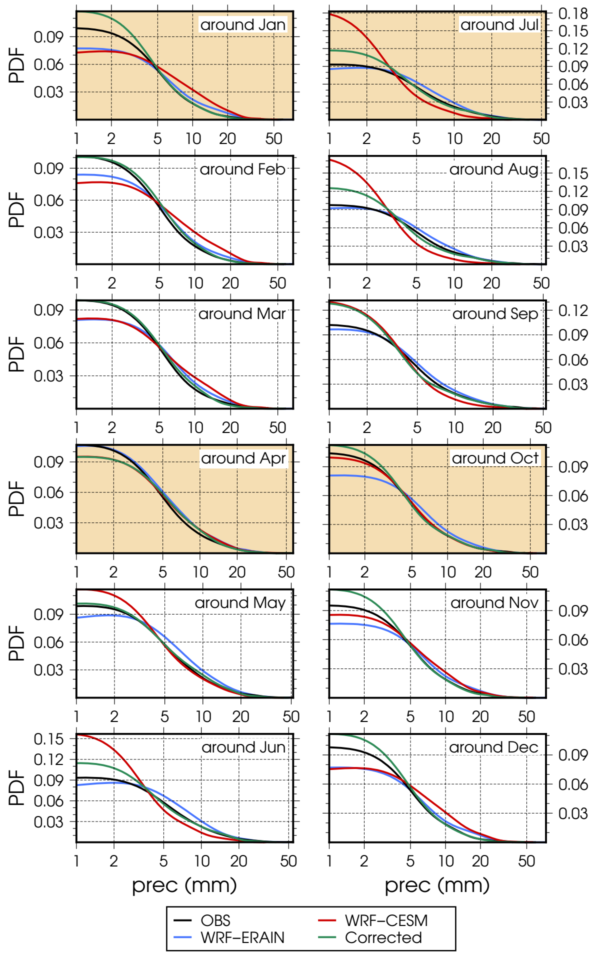
We have tried to motivate the added value of the high resolution. In particular, the fact that we can avoid the use of parametrisations of convective processes, and we provide references that back the added value of such simulations. The difference between 2 km and 6 km can be substantial. For instance, in response to reviewer #1 we have discussed the effective resolution. If we use the factor 3-4 mentioned in some references (e.g. Pielke Sr, 2013), 6 km of spatial resolution would have an effective resolution clearly above 10 km, which can be argued that it is not sufficient to account for all convective processes. This can be hardly put in doubt with a resolution of 2 km. More precisely, (Gómez-Navarro et al., 2015) investigated the particular issue of the skill as a function of spatial resolution, and found that there is indeed a large gain in switching from 6 km to 2 km. Unfortunately the latter study is based on the performance of surface wind, not precipitation. With this context, it seems reasonable to carry out a study of the added value of the model performance as a function of spatial resolution, using the precipitation produced within D3 or even D2 of this simulation. Unfortunately such an analysis can not be carried out with the present simulation. The reason is that this run was carried out with all domains nested two-way, as described in section 2.4. This implies that the precipitation as simulated by each coarser domain is replaced by the one within the innermost domain in the overlap region, i.e. the precipitation recorded for D3 inside the region span by D4 is actually a spatial smoothed version of the latter. Therefore it is does not correspond to the actual precipitation as resolved by a 6 km configuration, but an improved version that accounts for phenomena explicitly resolved within D4. This effectively precludes the use of this data for the fair evaluation of the model performance as a function of spatial resolution suggested by the reviewer. At this stage, a proper evaluation of this issue would require re-running great part of the simulations, which would involve a prohibitive computational cost.

5. I understand that the forcing GCM is always an open question, but the usage of just one instead of, at least, a couple of them, does not limit a little bit the representativity of the GCM-forced RCM analysis?

Certainly. It is always better to target at an ensemble, as such an approach allows to better characterise GCM-specific biases. This is indeed what we aim to some extent with the inclusion of the simulation driven by ERA-Interim in the analysis. However, computational cost is a bottleneck in this study. Carrying out a single realisation with a single GCM costed thousand of hours in one of the most powerful supercomputer available, CSCS. It is completely unaffordable for us the repetition of this simulation driven by alternative GCMs to produce a proper ensemble. We hope that this limitation is overcome in future studies, but unfortunately we are currently limited by this.
Still, we have added a paragraph in section 2.5 to discuss this issue.

6. The result shown in pages 10-11 that related intermediate seasons with cancellation artifacts sounds reasonable, but perhaps a more specific analysis could be made, with moving seasons, to see if more clear picture of that can be obtained. Because on the other hand, this result could be found non-intuitive, as one can think that precisely those transition seasons are more difficult to be properly captured. Which are the thoughts of the authors about it?.

The figure below shows the result of the calculation suggested by the reviewer. It shows PDFs of daily precipitation within "moving seasons". There are 12 panels, each one obtained considering as the window the given month, the previous and the former. The coloured panels highlight the standard seasons shown in Fig. 7 in the manuscript. The compensation of errors in intermediate seasons becomes apparent in WRF-CESM, as this simulation shows opposite biases in the previous and following seasons. We have briefly discussed this results in the manuscript, although we believe that the inclusion of the figure is not necessary.



7. Page 11, line 22. The bias corrected result over the frequency distribution that changes from underestimation to overestimation in winter looks a little bit peculiar. Could this result be a little bit further explained?

This issues has been raised by reviewer #1. As discussed by Themeßl et al. (2011), this effect occurs when models tend to underestimate the dry-day frequency (which is a rather infrequent feature of some RCMs, as most of them exhibit the oposite behaviour, i.e. drizzling-effect), as all these days become mapped onto a precipitation day, leading to a wet bias. This could be further corrected using frequency adaptation techniques, although we have not considered such techniques here. A brief discussion of this aspect has been included in the manuscript.

Technical corrections

1. When describing the experimental design (page 5, line 25) I do not understand those 6-day chunks and 12h spinup periods. I thought that a whole year or even two or more where needed for the soil moisture to be adapted. Could this aspect be explained a little bit more? I understand that more details can be found in Gomez-Navarro et al., 2015, but perhaps here it is too little what is said. It is the same about D1-D2-D3-D4 subdomains and nesting aspects.

We carry out the simulation in so-called reforecast mode. This approach is not new, but a well-settled methodology to conduct RCM simulations that splits the simulations into small tranches. As explained in the cited reference (Gómez-Navarro et al. 2015), *"The method consists of splitting a long simulation into shorter simulation periods of 1 to a few days, running each period separately and finally merging them. This method effectively minimises the impact of the boundaries, transforming the problem into a mostly initial-value problem. The reforecast method is regularly applied (Jiménez and Dudhia, 2012; García-Díez et al., 2013; Menendez et al., 2014, among others), and the increased skill of this method compared to continuous runs has been reported (Lo et al., 2008)."*. In a nutshell, splitting the simulation allows to bind the RCM to the driving dataset, and it can be regarded as a form of nudging. Further, this strategy has computational advantages: several simulations can be run simultaneously, which naturally leads to the parallelization of the problem.
Of course there are drawbacks. As pointed out by the reviewer, the short spinup period does not allow the soil moisture to reach an actual equilibrium with the atmosphere, which in opinion of

the authors reduces the land-atmosphere coupling (still, this coupling does not disappears, as it is borrowed from both the soil and atmosphere initial conditions used to run both submodels within each tranche). Although we believe this can bias certain applications of the simulations, for instance in the study of severe droughts and certain type of floodings, this approach does not impose a fundamental problem in general, as the successful validation of the simulations carried out in this same study demonstrates.
Regarding the domains, we clearly state their setup in Section 2.4 and even show them explicitly in Fig. 1: "Horizontally, we use four two-way nested domains with grid sizes of 54, 18, 6 and 2 km, respectively (top map in Fig. 1)"
We have added more details and a brief exposition of these arguments in the Section 2.5 in new version of the manuscript.

2. Close to this point, I do not also understand why nudging is applied to ERAinterim forced simulation, but not to the ESM one.

This is not arbitrary, but there is a rationale behind this choice. We developed it in the first version of the manuscript:

> The rationale behind this choice is that avoiding nudging gives the model more freedom to develop a more precise representation of the physical processes at regional scales (due to the higher resolution), and thus is potentially able to better correct systematic biases of the ESM, which, e.g., simulate a too strong zonal circulation (Bracegirdle et al., 2013)."

References

Fantini, A., Raffaele, F., Torma, C., Bacer, S., Coppola, E., Giorgi, F., Ahrens, B., Dubois, C., Sanchez, E., and Verdecchia, M.: Assessment of multiple daily precipitation statistics in ERA-Interim driven Med-CORDEX and EURO-CORDEX experiments against high resolution observations, pp. 1–24, https://doi.org/10.1007/s00382-016-3453-4.

Felder, G., Gómez-Navarro, J. J., Zischg, A., Raible, C. C., Röthlisberger, V., Bozhinova, D., Martius, O., and Weingartner, R.: From global circulation to local flood loss: Coupling models across the scales, Science of the Total Environment, 635, 1225–1239, https://doi.org/10.1016/j.scitotenv.2018.04.170, 2018.

Froidevaux, P. and Martius, O.: Exceptional integrated vapour transport toward orography: an important precursor to severe floods in Switzerland: Integrated Vapour Transport and Floods in Switzerland, 142, 1997–2012, https://doi.org/10.1002/qj.2793, 2016.

Gómez-Navarro, J. J., Raible, C. C., and Dierer, S.: Sensitivity of the WRF model to PBL parametrisations and nesting techniques: evaluation of wind storms over complex terrain, Geoscientific Model Development, 8, 3349–3363, https://doi.org/10.5194/gmd-8-3349-2015, 2015.

Houze, R. A., James, C. N., and Medina, S.: Radar observations of precipitation and airflow on the Mediterranean side of the Alps: Autumn 1998 and 1999, 127, 2537–2558, https://doi.org/10.1002/qj.49712757804, 2001.

Isotta, F. A., Frei, C., Weilguni, V., Perčec Tadić, M., Lassègues, P., Rudolf, B., Pavan, V., Cacciamani, C., Antolini, G., Ratto, S. M., Munari, M., Micheletti, S., Bonati, V., Lussana, C., Ronchi, C., Panettieri, E., Marigo, G., and Vertačnik, G.: The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data: CLIMATE OF DAILY PRECIPITATION IN THE ALPS, 34, 1657–1675, https://doi.org/10.1002/joc.3794, 2014.

Martius, O., Zenklusen, E., Schwierz, C., and Davies, H. C.: Episodes of alpine heavy precipitation with an overlying elongated stratospheric intrusion: a climatology, 26, 1149–1164, https://doi.org/10.1002/joc.1295, URL `http://doi.wiley.com/10.1002/joc.1295`, 2006.

Panziera, L., James, C. N., and Germann, U.: Mesoscale organization and structure of orographic precipitation producing flash floods in the Lago Maggiore region: Orographic Convection in the Lago Maggiore Area, 141, 224–248, https://doi.org/10.1002/qj.2351, 2015.

Pielke Sr, R. A.: Mesoscale meteorological modeling, vol. 98, Academic press, 2013.

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

Torma Csaba, Giorgi Filippo, and Coppola Erika: Added value of regional climate modeling over areas characterized by complex terrain—Precipitation over the Alps, 120, 3957–3972, https://doi.org/10.1002/2014JD022781.

A new region-aware bias correction method for simulated precipitation in ~~the Alpine region~~areas of complex orography

Juan José Gómez-Navarro[1,2,3], Christoph C. Raible[1,2], Denica Bozhinova[1,2], Olivia Martius[2,4], Juan Andrés García Valero[3,5], and Juan Pedro Montávez[3]

[1]Climate and Environmental Physics, University of Bern, Bern, Switzerland
[2]Oeschger Centre for Climate Change Research, Bern, Switzerland
[3]Department of Physics, University of Murcia, Murcia, Spain
[4]Institute of Geography, University of Bern, Bern, Switzerland
[5]AEMET, Agencia Estatal de Meteorología, Spain

Correspondence: Juan José Gómez-Navarro (jjgomeznavarro@um.es)

Abstract. Regional climate modelling is used to ~~better capture~~ simulate the hydrological cycle, which is fundamental for climate impact investigations. However, the output of these models is affected by biases that hamper its direct use in impact modelling. Here, we present ~~and evaluate the performance of~~ two high-resolution (2 km) climate simulations of precipitation in the Alpine region, evaluate their performance over Switzerland, and develop a new bias correction technique for precipitation suitable for complex topography. The latter is based on quantile mapping, which is applied separately across a number of non-overlapping regions defined through cluster analysis. This technique allows removing prominent biases while it aims at minimising the disturbances to the physical consistency ~~of the simulation~~inherent in all statistical corrections of simulated data.

The simulations span the period 1979-2005 and are carried out with the Weather Research and Forecasting model (WRF), driven by the reanalysis ERA-Interim (hereafter WRF-ERA), and the Community Earth System Model (hereafter WRF-CESM). The simulated precipitation is in both cases validated against observations in Switzerland. In a first step, ~~Switzerland~~ the area is classified into regions of similar temporal variability of precipitation. Similar spatial patterns emerge in all datasets, with a clear Northwest-Southeast separation following the main orographic features of this region. The daily evolution and the annual cycle of precipitation in WRF-ERA closely reproduces the observations. ~~This is in contrast to WRF-CESM, which~~ Conversely, WRF-CESM shows a different seasonality with peak precipitation in Winter and not in Summer as in the observations or in WRF-ERA. The application of the new bias correction technique minimises systematic biases in the WRF-CESM simulation, and substantially improves the seasonality, while the temporal and physical consistency ~~among simulated variables is~~ of simulated precipitation is greatly preserved.

1 Introduction

Producing reliable climate information is fundamental to address many of the currently open research questions about climate change (IPCC, 2013). Many of these questions pertain the future evolution of hydrological variables, as they are especially important for potentially impacting society. An important source of uncertainty in current climate projections originates from the inability to resolve all relevant processes of the hydrological cycle, e.g. convection, which affect in particular statements about extreme events of hydrological variables (IPCC-SREX, 2012). For instance, Rajczak et al. (2013) used simulations from the ENSEMBLE project to conclude that in the Alpine region some simulations project an intensification of heavy precipitation events during fall, albeit this result is clearly model-dependent, and therefore burdened by large uncertainties (Rajczak et al., 2013). . More recently, Rajczak Jan and Schär Christoph (2017) updated this results using an large ensemble of 100 Regional Climate Model (RCM) simulations from both ENSEMBLES and EURO-CORDEX. These authors indicate that newer simulations exhibit no clear agreement on the projection of a reduction in summer precipitation and rainy days, and point out to the use of different convection parametrizations as one of the main sources of this uncertainty. In this regard, Giorgi et al. (2016) have shown how convective precipitation is indeed a fundamental mechanism that modulates the response of precipitation in the Alpine region to climate change.

To gain insights in the hydrological cycle, different sources of information are available, namely observations and model simulations. Particularly important for this study are gridded observational products (e.g. Haylock et al., 2008; MeteoSwiss, 2016), as their spatial homogeneity becomes particularly useful in the validation of climate models (Gómez-Navarro et al., 2012). Simulation of the climate is performed with a wide variety of models ranging from simple box models to state-of-the-art comprehensive Earth System Models (ESM) (e.g. Hurrell et al., 2013; Lehner et al., 2015). These models are used in, e.g., process understanding as well as in simulating past, present, and future climate conditions. Observations and simulations offer complementary viewpoints to climate variability. The cornerstone of climate simulations is their internal physical consistency, which emerges from the underlying set of physical equations that are solved internally as part of the simulation. However, internal variability, the counterpart of natural variability in the model world, precludes the simulation from following the actual path of climate, which indeed can be seen as a single random realization of such variability. As a compromise between models and observations, reanalysis products combine the physical consistency of climate simulations with the assimilation of observations, therefore blending physical consistency with a temporal evolution that mimics the actual past evolution of climate (e.g. Dee et al., 2011). Both ESMs and reanalysis products are useful in different contexts, and the choice of using one over the other depends ultimately on the question being addressed.

Regardless of the type of simulation being employed, a bottleneck is the spatial resolution. Global reanalysis products or simulations with state-of-the-art ESMs, e.g. in Climate Model Intercomparison Project (CMIP5) (Taylor et al., 2012; Wang et al., 2014), have a spatial resolution of 50 to 200 km (Dee et al., 2011; Rienecker et al., 2011; Taylor et al., 2012; Lehner et al., 2015). Although this spatial resolution is sufficient to explicitly simulate the physical processes that dominate the large-scale atmospheric dynamics, it cannot resolve the sub-grid physical processes that are important for the hydrological cycle, e.g., microphysics and convective processes, and therefore have to be parametrized, thereby being an important source of uncertainty in current climate projections (Rajczak Jan and Schär Christoph, 2017). This is especially problematic for the accurate simulation of the climate in areas of complex topography, such as the Alps (e.g. Rajczak et al., 2013; Gómez-Navarro et al., 2015) (Rajczak et al., 2013; Torma et al., 2015; Giorgi et al., 2016; Rajczak Jan and Schär Christoph, 2017, among others), and in variables where for which the interaction with terrain is very important, such as precipitation and wind (Montesarchio et al., 2014) (Montesarchio et al., 2014; Gómez-Navarro et al., 2015).

One way to overcome these problems is to increase the spatial resolution enabling the explicit simulation of a wider range of physical phenomena over the area of interest with the help of a Regional Climate Model (RCM)RCM. This so-called dynamical downscaling approach allows to simulate the climate over a limited-area domain according to the initial and boundary conditions prescribed by either a ESM or a reanalysis product (Jacob et al., 2013; Rajczak et al., 2013; Kotlarski et al., 2014, among others) (Jacob et al., 2013; Rajczak et al., 2013; Kotlarski et al., 2014; Torma et al., 2015; Fantini et al., 2016; Giorgi et al., 2016, among others) . The use of RCMs has proven to be a very valuable tool to downscale global datasets in the Alpine region, and indeed it has been the target area of various studies under the umbrella of large coordinate projects such as ENSEMBLES and more recently EURO-CORDEX and MED-CORDEX (e.g. Torma et al., 2015; Casanueva et al., 2016; Giorgi et al., 2016). For wind, Gómez-Navarro et al. (2015) proved that a change in spatial resolution from 6 km to 2 km has a great impact in the ability of the simulation to reproduce the observed surface wind. Regarding hydrological variables, several studies within the frame of EURO-CORDEX have recently evaluated the added value of increasing the RCM resolution from 0.44º to 0.11º in the spatial patterns and daily variability of precipitation (Torma et al., 2015; Casanueva et al., 2016; Fantini et al., 2016; Giorgi et al., 2016) . At even higher spatial resolution, Ban et al. (2014) showed that an increase in horizontal resolution from 12 km to 2.2 km leads to a noticeably increased ability of the same model configuration to simulate the observed frequency of heavy hourly precipitation events. This improvement with increasing resolution has been confirmed using a different RCM in a similar area of study (Montesarchio et al., 2014). The reason for this improvement is that convective precipitation is explicitly simulated, which otherwise has to be parametrized ; being a major source of model uncertainties (Awan et al., 2011).

So far, regional simulations performed with different RCMs over complex terrain with resolutions from 2 to 25 km have been analyzed. Rajczak et al. (2013) used 10 RCM simulations for the Alpine region in the context of the ENSEMBLES project, where the horizontal resolution was set to 25 km. The conclusions drawn in the former study were validated and updated using a 100-member ensemble which includes the former runs plus the newer EURO-CORDEX simulations, in which the spatial resolution is set to 12 km (Rajczak Jan and Schär Christoph, 2017). A number of recent studies have further improved the spatial resolution. Montesarchio et al. (2014) conducted a simulation with the COSMO-CLM for the period 1979-2000 driven by ERA-40 reanalysis at a spatial resolution of about 8 km. This simulation allows for a satisfactory representation of temperature and precipitation, and clearly outperforms a simulation run with the same model setup but at a coarser resolution of 25 km. Ban et al. (2014) carried out a similar simulation also with COSMO-CLM for the 10-year period 1998-2007 driven with ERA-Interim with an increased resolution of 2.2 km, therefore being able to explicitly simulate convection processes.

Still, noticeable and systematic biases remain that can be attributed to either limited process understanding, insufficient resolution, or biases introduced by the driving dataset (Themeßl et al., 2011). To overcome this, statistical post-processing of RCM output is used to remove known systematic biases (Gudmundsson et al., 2012; Teutschbein and Seibert, 2012; Maraun, 2016). The underlying idea is to apply a statistical transformation to the simulated model output so that the distribution of modelled data resembles the observed one. There are a variety of correction methods, which can be broadly classified into distribution derived transformations, parametric transformations and nonparametric transformations (Gudmundsson et al., 2012). Various studies have reviewed the possibilities, with an overall emphasis on hydrological variables, and quantile mapping has emerged as a nonparametric method that slightly outperforms other approaches, at least in areas of complex topography (Themeßl et al., 2011; Gudmundsson et al., 2012; Teutschbein and Seibert, 2012). However, the Different versions of these techniques have been tested in the recent literature, and even software packages have been specifically developed and made publicly available, e.g. downscaleR (https://github.com/SantanderMetGroup/downscaleR). Casanueva et al. (2016) applied three different methodologies to correct daily precipitation within the EURO-CORDEX ensemble and found that the improvements introduced by the correction depends on the model, region and details of the methodology, concluding that there is no single optimal approach. Dosio Alessandro (2016) used the same RCM ensemble to produce an ensemble of bias-corrected projections of climate change based on a number of climate indices from the Expert Team on Climate Change. The authors conclude that results depend on the index, season and region of interest. In particular, percentile-based indices are barely affected by bias adjustment, whereas absolute-threshold indices are very sensitive to the techniques. Further, some refinements to these techniques have been proposed. Wetterhall et al. (2012) proposed to correct the model output differently for each day, conditioned to several types of circulation patterns. Argüeso et al. (2013) introduced a variant of quantile mapping that is not corrected against gridded observations, but station data. This allows to overcome an emerging problem in very high-resolution simulations, namely that they produce fewer rain days than gridded observations, which is an assumption most bias correction techniques are based on. Felder et al. (2018) applied a preliminarily bias-corrected version of the dataset of simulated precipitation we thoughtfully present here as part of a larger study aimed at the simulation of impacts of extreme events with a compressive model chain. It this study, the authors apply and briefly evaluate a simple bias correction method, where some limitations of the technique, imposed by the complexity of the Alpine region and the high resolution of the data set, stand out. Indeed, the latter study motivated some of the improvements to the bias correction we introduce and analyse in the present study.

Despite the abundant literature on the suitability and added value of these techniques, the use of bias correction is currently still intensely debated. Maraun (2016) argues that it is difficult to establish the actual performance of these techniques in climate simulations, and Maraun et al. (2017) demonstrates how statistical corrections cannot overcome fundamental deficiencies in climate models, pointing out that new process-informed methods should be developed.

So far, regional simulations performed with different RCMs over complex terrain with resolutions from 2 to 25 km have been analyzed and bias corrected. Rajczak et al. (2013) used 10 RCM simulations for the Alpine region in the context of the ENSEMBLES project, where the horizontal resolution was set to 25 km. More recently, Montesarchio et al. (2014) conducted a simulation with the COSMO-CLM for the period 1979-2000 driven by ERA-40 reanalysis at a spatial resolution of about 8 km. This simulation allows for a satisfactory representation of temperature and precipitation, and clearly outperforms a similar simulation run with a coarser resolution of 25 km. Ban et al. (2014) carried out a similar simulation also with COSMO-CLM for the 10-year period 1998-2007 driven with ERA-Interim with an increased resolution of 2.2 km. The setting of the model allowed convection to be explicitly simulated. Ban et al. (2014) were able to show that convection permitting resolutions lead to more accurate representation of These limitations have implications in studies addressing climate change and impacts, as the climate change signal can be unrealistically yet unwittingly modified (see discussion in Teng et al., 2015; Casanueva et al., 2018). These concerns are acknowledged and summarised in a report from the IPCC (Stocker et al., 2015). Among other recommendations, this report advices to identify and try to understand most prominent model deficiencies prior applying any bias corrections, as well as always proving the raw uncorrected data along with a clear description of the frequency of hourly extreme precipitation events compared to resolution where convection is parametrizedmethodology applied to remove biases. In this direction, a new initiative associated to the CORDEX experiment called Bias Correction Intercomparison Project (BCIP, Nikulin et al., 2015) has been created and aims to "i) quantify what level of uncertainties bias adjustment introduces to workflow of climate information, ii) advance bias-adjustment technique and iii) provide the best practice on use of the bias-adjusted climate simulations".

Here, we tackle some of the problems discussed by Maraun et al. (2017)presenting a new , and demonstrated in practice in the low performance of a preliminary bias correction dataset of precipitation in the Aare catchment by Felder et al. (2018). We describe an improved approach based on the combination of dynamical downscaling to a very high resolution that explicitly considers a greater number of physical processes at regional scale, followed by a quantile mapping correction applied separately to regions which are defined according to their different precipitation regimes. Thus, the aim of this study is twofold. First, we present and assess the performance of describe two high-resolution climate simulations (2 km horizontal resolution) for the Alpine region in the period 1979-2005, and assess their performance over Switzerland with the emphasis put on the ability of the model to reproduce precipitation. These simulations supersedes existing studies (Ban et al., 2014; Montesarchio et al., 2014) in terms of length (27 years) and spatial resolution (2 km). The RCM is driven by two different datasets: the reanalysis ERA-Interim (Dee et al., 2011) and a transient simulation of an ESM (Lehner et al., 2015). The comparison of both datasets allows the characterization of errors and their attribution to biases in the driving conditions, therefore fulfilling recommendations by the IPCC for the AR6 (Stocker et al., 2015), while it enables the identification of robust features, therefore increasing which increases the reliability of both simulations. Second, the new process-informed bias correction technique for precipitation is introduced and applied to the simulation driven by the ESM, which shall be regarded as the preferred dataset for impact assessments. Thereby we can evaluate improvements with respect to previous results obtained with more simple bias correction techniques that do not explicitly account for complex topography (Felder et al., 2018).

2 Data, model and experimental design

2.1 Gridded observational dataset

This study relies on an observational dataset to evaluate and bias-correct the precipitation in our model simulations. We use the gridded product RhiresD, developed by MeteoSwiss (2016). This product is based on daily precipitation totals as recorded by a network of rain-gauge stations of MeteoSwiss. It uses quality checked observations to ensure maximum effective resolution and accuracy. The observations undergo an interpolation to fill a homogeneous 1 by 1 km grid with an effective resolution of 15 to 20 km. To directly compare the observations to the simulations, we bi-linearly interpolated the observations to 2 km. Although this dataset is considered as generally reliable, it may underestimate precipitation in high altitudes due to the data sparsity (e.g., Messmer et al., 2017). More generally, observational products contain uncertainty whose magnitude can be sometimes comparable to model errors (Gómez-Navarro et al., 2012). Still, in this study we do not explicitly consider this uncertainty, and instead assume that these observations represent the true precipitation without errors.

2.2 Global Reanalysis: ERA-Interim

The reanalysis ERA-Interim (Dee et al., 2011) is used to provide boundary conditions for one of the RCM simulations. ERA-Interim is a reanalysis product released by the European Centre for Medium Range Weather Forecast, and is generated running the IFS model at a spectral resolution of T255 and 60 vertical levels while it assimilates observational data. The assimilation technique is the 4-D variational analysis that digest a number of observations of the actual state of the atmosphere (Dee et al., 2011). While the reanalysis covers the period from 1979 to today, a shorter period spanning 1979–2005 is downscaled. The reanalysis data used has a 6-hourly temporal resolution and a spatial resolution of $0.75° \times 0.75°$.

2.3 Global model simulation: CESM

The second dataset which provides boundary conditioned of the RCM simulations is obtained from a seamless transient simulation with the Community Earth System Model (CESM, 1.0.1 release; Hurrell et al., 2013). This model is a state-of-the-art fully-coupled Earth System Model developed by the National Center for Atmospheric Research run at a resolution of about 1° in all physical model components (atmosphere, ocean, land and sea ice) (CCSM; Gent et al., 2011) and the carbon cycle module. The latter interactively calculates CO_2 concentrations and exchange these between the model components. Further details for the particular setting are presented in Lehner et al. (2015).

The transient simulation spans the entire last millennium from AD 850 to 2099, but for this study we focus on the period 1979 to 2005. The simulation is initialized from a 500-yr control simulation under perpetual AD 850 conditions. The transient external forcing is obtained from the Paleo Model Intercomparison Project 3 (PMIP3) protocols (Schmidt et al., 2011). It consists of Total Solar Irradiance (TSI), volcanic and anthropogenic aerosols, land use change, and greenhouse gases. TSI forcing deviates from the PMIP3 protocol, as the amplitude between the Maunder Minimum (1640-1715) and today is doubled. Note further that CO_2 concentrations obtained by the carbon cycle module are radiatively inactive. Instead, observed/reconstructed CO_2 concentrations (according to the PMIP3 protocol) are applied in the radiation schemes of the physical model components. Beyond AD 2005 the external forcing is obtained from the Representative Concentration Pathways RCP8.5, which corresponds to a radiative forcing of approximately 8.5 W m^{-2} in the year 2100. Further details on the simulation are summarized in Lehner et al. (2015) and analyses of this simulation ~~is~~ are presented elsewhere (Keller et al., 2015; PAGES 2k-PMIP3 group, 2015; Camenisch et al., 2016; Chikamoto et al., 2016).

2.4 The regional climate model WRF

The dynamical downscaling of the reanalysis data and the CESM simulation is performed with the Weather Research and Forecasting Model (WRF, version 3.5; Skamarock et al., 2008). This non-hydrostatic model uses a Eulerian mass-coordinate solver. The setting follows the one discussed in Gómez-Navarro et al. (2015): It is vertically discretized by a terrain-following eta-coordinate system with 40 levels. Horizontally, we use four two-way nested domains with grid sizes of 54, 18, 6 and 2 km, respectively (top map in Fig. 1). Although the innermost domain of the simulation spans the Alpine region almost entirely, the analysis hereafter is based on the area covered by RhiresD, which is limited to the interior of Switzerland (bottom map in Fig.

1). The physical parametrizations include the micro-physics WRF single-moment six-class scheme (Hong and Lim, 2006), the Kain-Fritsch scheme for cumulus parametrization (Kain, 2004), which is implemented only in the two outermost domains. In the inner most domain the convection parametrization is disabled as at this resolution the model is convection-permitting. The planetary boundary layer is parametrized by a modified version of the fully non-local scheme developed at Yonsei University (hereafter YSU) (Hong et al., 2006), which accounts for unresolved orography (Jiménez and Dudhia, 2012). The radiation is treated by the Rapid and accurate Radiative Transfer Model (RRTM) (Mlawer et al., 1997) and the short-wave radiation scheme by (Dudhia, 1989). Finally, land processes are simulated by the Noah land soil model Chen and Dudhia (2001).

2.5 Experimental design: downscaling ERA-Interim and CESM

Two ~~regional~~ RCM simulations for the European Alps are ~~performed:~~ conducted for the same period 1979-2005. This period is chosen for being the overlap between the ERA-Interim and the CESM simulation. First, the ERA-Interim reanalysis dataset is dynamically downscaled with WRF ~~for the period 1979 to 2005~~ (hereinafter referred as WRF-ERA). ~~Therefore, ERA-Interim is split in 6-day chunks and~~ The simulation is run in so-called reforecast mode. This consists of dividing the full period into small tranches of 6 days with a spin-up period of 12 hours~~is used.~~. This approach allows to efficiently parallelize the problem, although it has the drawback of reducing the coupling between the land and the atmosphere. This can, in turn, introduce biases in the simulation of phenomena where the feedback between both systems is of prominent relevance, e.g. severe drought or certain type of floodings. Still, it does not impose a bottleneck of the model performance in terms of its ability to simulate surface wind, as shown by Gómez-Navarro et al. (2015), or in precipitation, as demonstrated here. Further, analysis nudging of wind, temperature and humidity above the PBL is allowed within the regional model domain, as this setting proved to outperform other configurations for this domain and model setup (Gómez-Navarro et al., 2015).

Secondly, ~~the same~~ this period of the CESM simulation is dynamically downscaled (hereinafter referred as WRF-CESM). For this simulation, the WRF setup is almost identical to the one of WRF-ERA in order to facilitate comparison between the simulations and to be able to analyze the influence of different driving datasets. Still, one important difference exists: the absence of analysis nudging. The rationale behind this choice is that avoiding nudging gives the model more freedom to develop a more precise representation of the physical processes at regional scales (due to the higher resolution), and thus is potentially able to better correct systematic biases of the ESM, which, e.g., simulate a too strong zonal circulation (Bracegirdle et al., 2013).

The comparison between WRF-ERA and WRF-CESM allows the identification of biases attributable to the driving conditions for the RCM, as described below. In this regard, it would be desirable to repeat the latter simulation using different Global Climate Models. Unfortunately the high resolution used in the RCM configuration demands a high computational cost that currently precludes the repetition of the experiment to produce an ensemble.

3 Bias correction technique

Although dynamical downscaling should improve coarsely resolved datasets, biases from either the driving dataset or the regional model still remain, as shown in the next section. In a previous study, Felder et al. (2018) used a bias-corrected version of the precipitation in WRF-CESM. The results (see Figs. 4 and 5 in Felder et al., 2018) demonstrate a modest performance of quantile mapping, and motivate further improvements to the methodology. Therefore, we developed a new bias correction technique, which combines a cluster analysis-based selection of regions with similar variability and quantile mapping for these regions. This technique is applied to each month separately which is justified as biases can be related to processes which undergo a strong seasonal cycle. This separation into regions of similar variability and through the annual cycle explicitly acknowledges that errors can be due to different physical processes, and therefore allows more physically coherent corrections.

In the first step, regions of similar variability are defined according to an objective criterion. In doing so, an Empirical Orthogonal Functions (EOF) analysis is applied to the precipitation series in order to obtain a rank-reduced phase space where the search of distances necessary in the subsequent cluster analysis is facilitated. We retain 7 leading EOFs, as they account for more than 80% of the total variance in the original datasets, while drastically reduces the computational cost. Then, a hierarchical clustering approach identifies regions of similar precipitation variability in the rank-reduced EOF space according to the Ward algorithm (Ward, 1963). To minimise the inherent subjectivity in the choice of the number of clusters to retain, we use a method based on the spectra of distances after every merge. To find the number of cluster centroids, the Euclidian distances between the centroids needs need to show a noticeable gap in the dendrogram that is built as part of the clustering procedure (not shown). A complementary criterion consists of aiming at retaining a low number of cluster centroids (and thus regions) so that a large number of grid points per centroid is available, which will improve the estimation of the transfer function in the quantile mapping step. The resulting cluster centroids are then used as initial seeds for a k-means clustering, which allows for fine-rearranging of grid points across regions (as one drawback of the hierarchical clustering is that a grid point once attributed to a specific cluster centroid will belong to it despite the fact that it might be more meaningfully attached to another cluster centroid in the end). Note that this regionalisation is not only a preliminary step of the bias correction procedure, but it is also used as an analysis technique to investigate the variability of precipitation over Switzerland and how consistent it is through various datasets.

In the second step, quantile mapping is applied separately to each of the regions identified within the first step. This non-parametric method corrects the empirical cumulative distribution function (ECDF) of the simulated precipitation with the observation (Themeßl et al., 2011; Rajczak et al., 2016). Assume that the climate model daily time series is $X_{\text{model}}(t,x,y)$ with t the time and x,y the location. To obtain a corrected time series $X_{\text{corr}}(t,x,y)$ the following rule is used:

$$X_{\text{corr}}(t,x,y) = \left(\text{ECDF}_{\text{obs}}(t,x,y)\right)^{-1}\left(\text{ECDF}_{\text{model}}(t,x,y)\right)$$

with ECDF^{-1} indicating the inverse ECDF, i.e., a quantile. Therefore, it can be seen as a transfer function between the ECDFs of the simulation and the observations. The quantile interval is set to 1, so quantiles corresponding to percentiles from 1st to the 99th are corrected. The transfer function is obtained for each region independently by pooling all grid points that belong to it (therefore a larger number of grid points per cluster facilitates the estimation of such function, as outlined above). Finally, the correction is applied to the daily series of precipitation in every grid point, with a transfer function that is common to all elements within the same region, but varies across the various regions defined by the cluster analysis. A small drawback of the separation into regions is that they lead to artificial and abrupt boundaries across the domain that would leave a fingerprint in the corrected data. To minimise these artificial boundaries, we perform a spatial smoothing in the obtained quantiles with a radius of 4 km, which smooths out the transfer functions prior to the correction, effectively removing ~~statistical artifacts.~~ such artifacts. Note that this scheme can lead to wet biases after the correction when the dry-day frequency is underestimated by the model, which then become systematically mapped onto precipitation days. These biases can be further removed with frequency adaptation techniques (Themeßl et al., 2012), although we do not consider them in our scheme, which can be related to wet biases in the corrected precipitation in Winter (see discussion below).

It is important to note the rationale for the separation into regions. Quantile mapping can be in principle used either for each grid point separately or on the entire domain, here Switzerland. Both options have advantages and disadvantages. Using an average transfer function over a large heterogeneous region may lead to problems when it contains positive and negative biases that can cancel each other and disable any correction. This problem disappears applying a correction to each grid point separately, but it has the disadvantage that the potential gain of a highly resolved physical consistent estimate of the climate obtained by the regional model is destroyed. These caveats contribute to the on-going discussion on the suitability of bias correction techniques and the necessity of more physical-based methods (Maraun et al., 2017). In this sense, the new bias correction technique based on objective regionalization presents a compromise between these two extremes, as regions with ~~coherent~~ similar precipitation behavior are corrected ~~similarly preserving physical consistency in these regions~~ coherent and jointly, thus preserving a great part of the physical self-consistency of this variable for each region dictated by the RCM, but still avoiding the cancellation of positive and negative biases.

We note that the application of this methodology implies a previous regionalisation of the series for each month separately, which in general involves notable computational cost. Further, months belonging to the same season behave similarly, so that the resulting regions are hardly distinguishable and the analysis presents some level of redundancy. For these reasons, we propose a simplified form of the methodology, which we apply hereafter, and consists of carrying out the regionalisation on a seasonal basis. Once identified, these regions can be regarded as representative and common for the three months within each season, so that the final correction can be applied on a monthly basis.

4 Evaluation of the simulations

4.1 Regions of common variability and time behavior

Using the cluster analysis introduced in Sec. 3, the number of regions with common variability (clusters) slightly varies per season and dataset (Table ~~2~~1). Their spatial distribution is depicted in Fig 2, where different colors represent grid points belonging to each region~~.~~, and the number of grid points within the Swiss domain that belong each region is shown in Table **??**. Note that in the smallest region the number of grid points is 60, which implies that 48600 pairs of numbers (i.e. 27 years $\times$ 30 days per year $\times$ 60 points per day) are used to obtain the transfer function that effectively carries out the correction in the less favourable case. This ensures that such function is efficiently estimated from the sample in all regions and cases. The number of clusters obtained is similar in all cases, and a clear Northwest-Southeast pattern emerges concurrently with the main orographic features over Switzerland (see bottom of Fig. 1). The resemblance between the regions obtained for both WRF simulations is remarkable. In all cases, a large region that includes the plains in the center of Switzerland, but also the Valais and Engadin valleys, stands out. Further, the southern part of the country, South of the Alps also emerges as a distinct region, although in some cases it is further sub-divided (see SON in the WRF-ERA simulation). The Alps themselves are another cluster in most of the seasons and datasets. The orographic pattern is explicit, with a cluster encompassing the mountains tops, in Winter in both simulations, and Spring in the WRF-CESM case. Such strong differentiation as a function of terrain height is not so explicit in other seasons. Still, it should be noted that differences in the sub-regions beyond North and South of the Alps are not so robust, and might be attributed to the subjective component in the choice of number of regions. The similarity between the regions in both simulations indicates that the precipitation regimes across Switzerland are mostly imposed by the RCM, being robust regarding the boundaries that impose the temporal evolution of the simulation. This is a non-trivial finding, as the CESM simulation is affected by acknowledged biases compared to ERA-Interim, and thus the output of the regionalization might shed to very different results. Instead, and although such biases leave a strong footprint in the amount and location of the simulated precipitation (further discussed below), the CESM boundary conditions lead to a spatial distribution of precipitation variability that, once dynamically downscaled, is greatly consistent with ERA-Interim.

Larger differences appear however when comparing the regions obtained with both simulations to the observations. As in the case of the simulations, two main superclusters stand out covering both sides of the Alps through the annual cycle, with some seasonal differences (the Northwest-Southeast pattern is less dominant in Autumn and Winter). The presence of the Alps and its orographic footprint is less obvious, and the regions are defined with clear boundaries. There are a number of reasons that help to explain such differences. The most prominent is the different resolution. OBS has an effective resolution of about 20 km (see 2.1), whereas both simulations reach 2 km in the innermost domain (although the regionalization has been obtained with a coarser resolution version of the data of 6 km due to computational constrains). Note that the effective resolution of the simulations is coarser than 2 km, as it is between 2 and 4 times the one implemented in the simulation (Pielke Sr, 2013) . The coarser resolution in the gridded product of observations contributes to the smoothing of the regions and therefore ~~a~~ to their clearer definition. The absence of strong orographic features (mountain tops, valleys, etc.) that can be recognized in Fig. 2 for the gridded observations might be attributable to the combined effect of ~~insufficient~~ coarser effective resolution plus the fact that there are fewer observations in the high mountain regions. This is an important limiting factor in gridded products for precipitation in complex topography areas.

The rationale of regionalization consists of finding groups of grid points where ~~differences in precipitation variability between elements of such region region are minimized~~precipitation variability within such region is coherent, whereas differences between different regions are maximized. ~~To analyze this in more detail ,~~ The discussion so far has focussed on a qualitative description of the outcome of the regionalisation, without analyzing in detail to what extent these regions can be regarded as different (the dendrograms used to stablish the number of regions are not shown, for instance). Therefore we analyze next in a quantitative fashion the coherence of the regions ~~is further evaluated~~ through correlation analysis. For this, the daily precipitation series in each grid point is grouped for each region and averaged to obtain regional series. Then the cross-correlation between all series is calculated for each dataset and season, and shown with a color scale in Fig. 3.

Note that there is no one-to-one correspondence between the regions in different datasets and seasons, so the labeling (1 to 6) of this figures has to be carefully read from Fig. 2. Correlations of daily regional-averaged precipitation are generally large, above 0.7 in many cases and never negative. This indicates that, despite the complex orography of the regions under study, precipitation evolves very coherently across Switzerland. Still, there are noticeable exceptions that appears as bands with more greenish and reddish colours. In Winter, region 4 in the observations, 3 in WRF-ERA and 2 in WRF-CESM exhibit the lowest correlations, reaching 0.2 in certain combinations of regions. Comparing with Fig. 2, these regions are located south of the Alps, and largely correspond to southern Switzerland, which stand out as regions with a remarkable, different behaviour. Similarly, in Spring the regions most strongly detached to the behaviour of the rest are 4 and 5 in the observations, 4 and 5 in WRF-ERA and 2 and 5 in WRF-CESM, which again correspond to the same Southeastern part of the country (see Fig. 2). In Summer, the Northwest-Southeast separation is still apparent and similar in both simulations (region 5 in both simulations, which corresponds to Ticino, is the most clearly decoupled), while such differentiation, although qualitatively similar, is not so strong in the observations, which exhibits correlations of up to 0.6 with region 1 in the Northeast. Finally, in Autumn the number of regions in both simulations is different (6 and 4) in WRF-ERA and WRF-CESM, respectively. However, the correlations in the bottom row in Fig. 3 show that this apparently different regionalisation can be understood in the same terms of Northwest-Southeast separation, as regions 4, 5 and 6 in WRF-ERA are the counterpart of region 2 in WRF-CESM, and the three formers behave collectively as the latter in terms of separation with respect the rest of the domain. The observations also reproduce this pattern in Autumn, although less clear, as correlations between regions are never below 0.4.

The skill of the WRF-ERA regarding its ability to reproduce the temporal evolution of observed precipitation in the period 1979-2005 is explored through a Taylor diagram that compares this dataset to the observations. Note that in this case the comparison with WRF-CESM is not meaningful due to the lack of assimilation of observations in the CESM simulation, therefore we skipped that dataset in the following analysis. The skill is assessed for each regional series, separately. This generates an inconsistency that complicates the calculation, as the number and shape of regions are different for the observations and WRF-ERA (see first and second columns in Fig. 2). We solve this by using the same regions to obtain the regional series in both datasets, which correspond to the ones obtained with WRF-ERA (second column in Fig. 2). The assessment of the skill is shown in Fig. 4, which depicts the results for each season (symbols) and region (colors). Daily correlations between WRF-ERA and OBS range between 0.6 and 0.9 in all cases, with an average of 0.78 (0.74 for Summer and 0.83 for Winter, respectively). This supports the lack of systematic errors attributable to driving conditions. Differences also ~~appears~~ appear in the ability of the simulation to mimic the temporal variability of precipitation. Region 1, which represents fairly consistently the central plains of Switzerland in all seasons, is where the agreement between the simulation and ~~the~~ observations is best, with a ratio of standard deviations close to one. In the rest of regions, the model overestimates the variance about 20% compared to the observations. Part of this bias can be explained in terms of the systematic overestimation of precipitation through the annual cycle in the WRF-ERA simulation described in the next section. However, a striking feature is the severe overestimation of simulated precipitation in region 4 in Winter, which corresponds to a cluster that is only identified in the simulation, and spans the highest mountains in the Alps (see Fig. 2). As argued above, the observations in such locations are generally less reliable and are more strongly affected by extrapolation artifacts (due to data sparsity), and therefore a plausible explanation for this outlier is the underestimation of actual precipitation and its variance in the observational product.

In summary, the regions identified in both simulations are similar and resemble the orographical barrier imposed by the Alps. This similarity demonstrates that the spatial structure of precipitation regimes are largely independent on the driving dataset. ~~The~~ This spatial structure is ~~well~~ similarly reproduced in the observations, although boundaries are more sharply defined and correlations among regions are slightly larger (see for example the lack of correlations below 0.4 in Summer, or 0.3 in Autumn). The more pronounced differentiation of regional characteristics in the simulations compared to the observations might be explained by the effectively coarser resolution of the observational gridded product of precipitation. Moreover, the Taylor diagram demonstrates the acceptable performance of the WRF-ERA simulation as a plausible surrogate of the evolution of precipitation in Switzerland during the ERA-Interim period.

4.2 Climatology and annual cycle

In this section we compare the downscaled precipitation driven by ERA-Interim and CESM to observations to identify systematic model deficiencies leading to biases of the downscaled precipitation (Figs. 5 and 6). Figure 5 shows the precipitation averaged over Switzerland separately for each month, thereby emphasising the annual cycle, whereas the Figure 6 presents the maps of accumulated precipitation for each season (by columns) and dataset (columns 1 to 3).

The seasonality of precipitation is well reproduced by the WRF-ERA simulation (see blue bars in Fig. 5, as well as first and second columns in Fig. 6), showing a peak in the Summer months June to August and the driest months in Winter. However, the WRF-ERA simulation generally overestimates precipitation throughout the year, in particular during December and January, which can be linked to the overestimation of precipitation variability identified in the previous section. This overestimation is especially noticeable in the highest locations around the Alps, but given the, in principle, larger uncertainties in the observations of precipitation in these locations, it is hard to judge to what extent this difference is directly attributable to just model deficiencies. In this regard, it is worth to note that there is a high agreement between WRF-ERA and OBS at low altitudes and valleys. ~~Thus, the~~ Despite the general wet bias, the model underestimates precipitation in Ticino in Autumn. Isotta et al. (2014) show that in the region of Ticino up to 70% of the yearly precipitation accumulation is due to the top 25% of the wet days, so it is sensible to assume that the bias stems from high to extreme precipitation events. In Ticino these heavy precipitation events are driven by the transport of moist and potentially unstable (moist neutral stratification) air masses against the Alps from the south (Martius et al., 2006; Froidevaux and Martius, 2016). Locally, the vertical shear between south-easterly flow near the surface and southerly to southwesterly above 850 hPa leads to moisture convergence and repeated formation convective cells (Panziera et al., 2015). On an even more local scale, strong vertical shear can result in small-scale circulation that results in local precipitation maxima (Houze et al., 2001). Therefore if the RCM fails to capture any of these local and highly driven by the orography processes properly, it will result in an underestimation of the precipitation. The simulation is able to capture great part of the complex spatial structure of the climatology of precipitation which is induced by the complex topography (Fig. 6). The spatial correlation between the simulated and observed patterns (Fig. 6) lies between

0.78 (in Winter) and 0.84 (in Summer). These results can be compared to those obtained with an ensemble of RCM simulations driven by ERA-Interim within the EURO-CORDEX and MED-CORDEX projects. Fig. 2 in Fantini et al. (2016) is similar to Fig. 6 here, although the model resolution and observational gridded product used to validate the models are different. Further, Fig. 5 in Fantini et al. (2016) shows similar annual cycle as Fig. 5 here, but the Alps domain they consider is considerably larger, including western France, great part of Austria and the northern half of Italy. The comparison of these figures shows strong agreements, e.g. the simulations reproduce an orographical pattern with the highest precipitation over the Alps, they consistently overestimates precipitation, and they closely follow the annual cycle with the respective observational product. However a remarkable difference is that the annual cycle in the Alps domain in Fantini et al. (2016) presents a bi-modal curve without the unique and clear summer maximum we find for Switzerland and is consistent between WRF-ERA and the observations. Since the observational products are both of high quality and similar characteristics, this discrepancy is attributable to the disparity between the domains both studies consider.

As expected, the performance of the simulation when WRF is driven by CESM is lower (see red bars in Fig. 5, and first and third columns in Fig. 6). WRF-CESM shows strong deviations in the seasonal cycle with a maximum of precipitation in the extended Winter season from November to March and a strong underestimation of precipitation in Summer (Fig. 5). Strikingly, this behaviour is reverse to the observations, which show a peak in the Summer months from June to August and less precipitation in Winter. The spatial disaggregation of these biases are further explored in the seasonal precipitation patterns in Fig. 6. WRF-CESM strongly overestimates precipitation at high altitudes in Winter beyond the problems already stated regarding WRF-ERA. Further, it severely underestimates Summer precipitation (spatial average of 429.94 mm in the observations vs. 195.76 in WRF-CESM, respectively), without a clear footprint of orography in this bias. The spatial correlations between the simulated (WRF-CESM) and observed patterns, although lower than in WRF-ERA, are still fairly high, ranging from 0.55 (in Autumn) to 0.78 (in Summer). Again, this correlation is due to the strong influence of orography. This further emphasises how the spatial distribution of precipitation regimes are, to a great extent, imposed by the RCM setup alone, whereas the ability of the simulation to reproduce the annual cycle is largely governed by the driving conditions provided externally through the boundaries. The performance of WRF-CESM can be compared to ESM-driven simulations within the EURO-CORDEX and MED-CORDEX ensembles. Figs. 3 and 4 in Torma et al. (2015) show the averaged winter and summer precipitation in the observations and the ensemble mean, and provide results consistent with the discussion about the influence of orography on precipitation presented above. Fig. 2 in Torma et al. (2015) shows the annual cycle for the same Alps domain employed by Fantini et al. (2016). The ensemble mean of ESM-driven simulations does reproduce the bi-modal annual cycle present in the observations for this domain, and the overestimation of precipitation is similar to the one obtained with these models are driven by ERA-Interim (Fantini et al., 2016). Therefore the seasonality biases of WRF-CESM seem not to be a general problem across ESM-driven simulations, but rather an issue specific to this ESM.

An important outcome of these simulations is the potential application to the study of extreme events. This type of study demands the disaggregation of precipitation into shorter periods than monthly averages. Although the daily correlation between WRF-ERA and OBS was shown in the Taylor diagram in Fig. 4, the ability of WRF to reproduce daily precipitation has not been explicitly analysed so far. Therefore, we evaluate model biases at daily scale by showing the Probability Density Function

(PDF) of daily precipitation averaged over Switzerland for each season (Fig. 7). The overestimation of Winter precipitation in the WRF-CESM simulation stands out as an underestimation of the frequency of days with precipitation below 5 mm, i.e. the so-called "drizzling-effect", and its counterpart in the higher frequency of precipitation above 10 mm. WRF-ERA behaves similar to WRF-CESM, although the magnitude of this bias is lower. In Summer, the WRF-ERA simulation is able to mimic the distribution of precipitation. The WRF-CESM simulation exhibits a distorted PDF of daily precipitation in Summer, as the frequency of days with precipitation below 3 mm is strongly overestimated. This leads to the severe underestimation of precipitation apparent in Fig. 6. The comparison with the simulation driven by ERA-Interim~~shows~~ , as well as the aforementioned results within the EURO-CORDEX ensemble (Fantini et al., 2016), show that this systematic error becomes attributable to biases in the boundary conditions provided by the CESM model. In the intermediate seasons of Spring and Autumn, both simulations exhibit an mixed behaviour, and their skill is remarkably good in Spring. ~~Although~~ Indeed, WRF-CESM allegedly outperforms WRF-ERA in Autumn~~, this can be explained as~~ . However the latter is not a demonstration of model performance, but an error cancellation artifact, as can be shown evaluating the performance through moving seasons (not shown). The behaviour of biases during this season are a combination of the ones in Summer and Winter, which are opposite and therefore tend to cancel out when pooled to obtain the PDF.

5 Bias correction of the WRF-CESM simulation

From the results described so far, three important conclusions can be drawn:

- WRF-ERA mimics many important features of the observed spatio-temporal distribution of precipitation, even at daily scale and through the annual cycle.
- The spatial structure of precipitation variability is strongly affected by orographic features, and is prescribed by the RCM. This leads to consistency between WRF-ERA and WRF-CESM, and together with the first point, supports the reliability of the latter simulation.
- The temporal evolution is driven by the boundary conditions, and in particular the WRF-CESM presents important systematic biases through the annual cycle that can not be removed with dynamical downscaling alone.

These conclusions together suggest that although the output of the WRF-CESM is a valuable resource with potential applications, it might be desirable to post-process this dataset in a way that systematic biases are ameliorated. Therefore, the new bias correction method ~~binning~~ binding cluster analysis and quantile mapping (Sec. 3) is applied to the WRF-CESM simulation.

The results of the bias correction method are presented in the Figs. 5 to 7 showing the desired improvements: the mean precipitation fields agree better with the observations, so that the annual cycle is corrected in a way that closely follows the observed values (green bars in Fig. 5). In particular, the strong overestimation (underestimation) in Winter (Summer) has been removed to a large extent. It is worth to note the clear improvements in the ability of the bias-corrected dataset to mimic the annual cycle compared to the results obtained with a simpler method that does not account for the spatial heterogeneity (Fig. 5 in Felder et al., 2018), as well as in the spatial patterns of precipitation (Fig. 4 in Felder et al., 2018). The bias correction also improves the intensity of precipitation and preserve its spatial structure (compare second and fourth columns in Fig. 6). This is important, as according to the results above, this structure is in agreement with the more reliable WRF-ERA simulation. However, it does not improve the spatial correlation with the observations, which ranges between 0.54 (in Autumn) and 0.78 (in Summer). Interestingly, an improvement is also found on a daily scale (green curve in Fig. 7). The underestimation of the frequency of days with very low precipitation in Winter is corrected, leading although it leads to a slight overestimation. This effect occurs when models tend to underestimate the dry-day frequency, as all days become mapped onto a precipitation day, producing a wet bias. This could be further corrected using frequency adaptation techniques (Themeßl et al., 2012), although we have not considered such techniques here. Above 5 mm the precipitation PDF is remarkably well captured. Similarly, in Summer the bias correction improves the PDF, although does not completely remove the overestimation (underestimation) of the frequency of dry (wet) days; above 4 mm the simulated PDF is barely indistinguishable from the observed one. Again, intermediate seasons exhibit a mixed behavior. In Autumn, the PDF of bias-corrected WRF-CESM simulation is apparently worse than the uncorrected WRF-CESM simulation. This reinforces the argument developed above regarding the apparent skill of the simulation in this season due to error cancellation.

As the proposed bias correction employs a non-linear transformation on a daily basis, which is based on a transfer function that differs for each month within the annual cycle, it does not simply scale precipitation, but modifies it in a different mannerevery daycomplex manner. Such modification slightly changes the temporal evolution of precipitation at every grid point. This is an undesired side effect, as the temporal co-evolution of all simulated variables is bounded by the equations being solved by the model, and therefore modifications to this evolution may underscore the most valuable aspects of the dynamical downscaling: its physical consistency (Maraun, 2016). This effect is unavoidable, but it it depends on factors such as the magnitude of the biases, their location within the precipitation distribution, or their variability through the annual cycle, and should be ideally kept to a minimum. We demonstrate how the applied bias correction has only slightly affected the temporal evolution in Fig. 8, which shows the daily correlation separately by seasons to avoid the overestimation of correlation due to the annual cycle. The point-wise correlation between the raw and corrected simulation is well above 0.8 in all seasons across the domain, and lower than 0.9 in Autumn in just few quasi-random locations. The lower correlation in this season is motivated by the larger variability of the nature of biases within this season, which drives a large spread between the transfer functions for the three months, and therefore reduces the linear relationship between raw and corrected series (not shown). There is no obvious indication in these maps of geographical influences, e.g. orographic, longitudinal, etc. that might point out systematical errors attributable to a misrepresentation of physical processes at regional scales.

6 Conclusions

This study presents the performance and biases of two high-resolution climate simulations, and introduces a new bias correction technique that reduces systematic biases based on the regionalisation of precipitation. The simulations span the recent past 1979-2005 over the Alpine area, and entire Alpine region, although we limit the analysis and bias correction of the simulation to the area of Switzerland due to the limited spatial coverage of the observational product we use as reference. Both simulations are carried out with a RCM driven by two global datasets, an ESM (CESM) and a reanalysis product (ERA-Interim). The bias correction is based on quantile mapping, but it is separately applied to different regions of common variability, which are identified by objective cluster analyses.

The comparison between simulations and observations shows that regions of common variability agree between the two simulations and to a great extent with the observations. Still, the observed regions of common variability lack of many fine details found in the simulations due to the coarser effective resolution RhiresD data and potentially the sparse data network at high altitudes. Besides the regional classification, further agreements and differences between the simulations and observations are found. The WRF-ERA simulation is able to simulate the seasonal cycle but consistently overestimates precipitation by about 20%. The day-to-day variability is captured by the WRF-ERA simulation with rather high positive correlation, but the simulated variability is again larger than in the observations. At least for Winter, overestimation of simulated variance is related to a potential underestimation of observed precipitation due to the sparsity of observations in high mountains. The biases of the WRF-CESM simulation are expected to be larger as the driving CESM data do not incorporate observations. The WRF-CESM simulation is not able to simulate the seasonal cycle correctly with a strong overestimation (underestimation) of Winter (Summer) precipitation.

To correct for these systematic biases a new bias correction technique is applied to the WRF-CESM simulation. The separation in regions of common variability by the cluster analysis acknowledges the fact that biases in different regions and seasons are produced by different physical mechanisms, and minimises the risk of error cancellation. ~~The latter~~ This method clearly improves simpler approaches that do not account for this heterogeneity, and is an issue when ~~applied quantile mapping~~ quantile mapping is applied to larger regions like the entire Switzerland (Felder et al., 2018). The spatial structure of bias corrected precipitation is preserved compared to the original WRF-CESM, but the seasonality is corrected in a way that nearly mimics the observations. This improvement is also found when analysing the daily scale. This means that the temporal evolution of the simulation, which emerges from the physical consistency of the simulation, is greatly preserved, as the daily temporal correlation between the raw and corrected versions of the WRF-CESM simulation is above 0.9 in most cases, except for few quasi-random grid-points in Autumn.

~~The~~ We note that the rationale of the developed methodology is to divide a large domain into smaller subregions according to the behaviour of the target variable. We have applied it here to daily precipitation in Switzerland for being a variable strongly affected by complex orographical details that lead to strong horizontal gradients. With more generality, spatial regionalisation is an efficient method to break down complexity in areas and variables whose behaviour strongly varies through the domain. Still, the bias correction applied separately to subregions can be in principle adapted to other cases with simpler topography, or other variables with lower horizontal gradients. The only practical difference is that in this case the regionalisation will naturally lead to a lower number of subregions which are necessary to obtains clusters with coherent features.

Finally, the applicability of the three datasets, i.e. WRF-ERA, the raw WRF-CESM, or the corrected version of WRF-CESM, depends on the nature of the question to be addressed. For applications where a match with the actual observed climate is needed, the ERA-Interim driven simulations is suitable. However, there are research questions ~~where~~ for which a simulation driven by an ESM, such as WRF-CESM, is necessary. This is for example the case for climate change projections, but also climate simulations of past conditions, or studies of extreme situations in long simulations (Felder et al., 2018) or sensitivity studies (Messmer et al., 2015). Finally, the use of corrected variables is advisable only when an accurate simulation of the magnitude of the variable under consideration is critical for the application. An example is the use of output of climate simulation as input in hydrological modelling (Camici S. et al., 2014; Felder et al., 2018), as the magnitude of rainfall in a given location, and not only its large-scale structure or temporal consistency, is crucial for an realistic simulation of river discharge.

Code availability. All code used through this manuscript is open source. WRF is a community model that can be downloaded from its webpage (http://www2.mmm.ucar.edu/wrf/users). The code to perform the regionalisation, as well as the Taylor diagram, is based on R and Bash scripts, whereas quantile mapping and PDF estimation is implemented with Fortran 90. The source code of these tools is available upon request. Simple calculations carried out at each grid point, e.g. means, correlations, etc. have been performed with CDO (https://code.mpimet.mpg.de/projects/cdo). The figures have been prepared with GMT (http://gmt.soest.hawaii.edu)

Data availability. The CESM simulation was carried out at the University of Bern, and is available once approved by the original authors. The ERA-Interim dataset can be downloaded from the ECMWF webpage, although it requires previous registration. The two datasets produced, WRF-ERA and WRF-CESM consists of hourly output of a number of variables, and therefore occupies several Terabytes and is not freely accessible. Still, it can be accessed upon request to the authors of this manuscript.

Author contributions. JJGN coordinated the work, carried out the WRF-ERA simulation and the calculations of this manuscript. CR contributed in the design of the simulations and their analysis. DB carried out the WRF-CESM simulation. OM helped in the design of the simulations and the discussion of the results. JAGV provided the code to carry out the regionalisation and helped in its analysis. JPMG provided ideas for new approaches in the analysis of the simulations that have been integrated in the final manuscript. The manuscript has been written by JJGN and CR, and all authors have contributed reviewing the text.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work is supported by the Mobiliar Lab and the Swiss National Science Foundation (grant pleistoCEP (Nr. 200020_159563)). The CESM and WRF simulations were performed on the super computing architecture of the Swiss National Supercomputing Centre (CSCS). JJGN acknowledges the CARM for the funding provided through the Seneca Foundation (project 20022/SF/16). Special thanks are due to Flavio Lehner who performed the seamless transient simulation with CESM.

References

Argüeso, D., Evans, J. P., and Fita, L.: Precipitation bias correction of very high resolution regional climate models, 17, 4379–4388, https://doi.org/10.5194/hess-17-4379-2013, 2013.

Awan, N. K., Truhetz, H., and Gobiet, A.: Parameterization-induced error characteristics of MM5 and WRF operated in climate mode over the Alpine region: an ensemble-based analysis, Journal of Climate, 24, 3107–3123, https://doi.org/10.1175/2011JCLI3674.1, 2011.

Ban, N., Schmidli, J., and Schär, C.: Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations, Journal of Geophysical Research: Atmospheres, 119, 7889–7907, https://doi.org/10.1002/2014JD021478, 2014.

Bracegirdle, T. J., Shuckburgh, E., Sallee, J.-B., Wang, Z., Meijers, A. J. S., Bruneau, N., Phillips, T., and Wilcox, L. J.: Assessment of surface winds over the Atlantic, Indian, and Pacific Ocean sectors of the Southern Ocean in CMIP5 models: historical bias, forcing response, and state dependence, Journal of Geophysical Research-Atmospheres, 118, 547–562, https://doi.org/10.1002/jgrd.50153, 2013.

Camenisch, C., Keller, K. M., Salvisberg, M., Amann, B., Bauch, M., Blumer, S., Brázdil, R., Brönnimann, S., Büntgen, U., Campbell, B. M. S., Fernández-Donado, L., Fleitmann, D., Glaser, R., González-Rouco, F., Grosjean, M., Hoffmann, R. C., Huhtamaa, H., Joos, F., Kiss, A., Kotyza, O., Lehner, F., Luterbacher, J., Maughan, N., Neukom, R., Novy, T., Pribyl, K., Raible, C. C., Riemann, D., Schuh, M., Slavin, P., Werner, J. P., and Wetter, O.: The 1430s: A cold period of extraordinary internal climate variability during the early Spörer Minimum with social and economic impacts in north-western and central Europe, Clim. Past, 12, 2107–2126, https://doi.org/10.5194/cp-12-2107-2016, 2016.

Camici S., Brocca L., Melone F., and Moramarco T.: Impact of climate change on flood frequency using different climate models and downscaling approaches, Journal of Hydrologic Engineering, 19, 04014 002, https://doi.org/10.1061/(ASCE)HE.1943-5584.0000959, 2014.

Casanueva, A., Kotlarski, S., Herrera, S., Fernández, J., Gutiérrez, J. M., Boberg, F., Colette, A., Christensen, O. B., Goergen, K., Jacob, D., Keuler, K., Nikulin, G., Teichmann, C., and Vautard, R.: Daily precipitation statistics in a EURO-CORDEX RCM ensemble: added value of raw and bias-corrected high-resolution simulations, 47, 719–737, https://doi.org/10.1007/s00382-015-2865-x, 2016.

Casanueva, A., Bedia, J., Herrera, S., Fernández, J., and Gutiérrez, J. M.: Direct and component-wise bias correction of multi-variate climate indices: the percentile adjustment function diagnostic tool, 147, 411–425, https://doi.org/10.1007/s10584-018-2167-5, 2018.

Chen, F. and Dudhia, J.: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity, Monthly Weather Review, 129, 569–585, https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2, 2001.

Chikamoto, M. O., Timmermann, A., Yoshimori, M., Lehner, F., Laurian, A., Abe-Ouchi, A., Mouchet, A., Joos, F., Raible, C. C., and Cobb, K. M.: Intensification of tropical Pacific biological productivity due to volcanic eruptions, Geophysical Research Letters, 43, 2015GL067 359, https://doi.org/10.1002/2015GL067359, 2016.

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., Kållberg, 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, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Dosio Alessandro: Projections of climate change indices of temperature and precipitation from an ensemble of bias-adjusted high-resolution EURO-CORDEX regional climate models, 121, 5488–5511, https://doi.org/10.1002/2015JD024411, 2016.

Dudhia, J.: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model, Journal of the Atmospheric Sciences, 46, 3077–3107, https://doi.org/10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2, 1989.

Fantini, A., Raffaele, F., Torma, C., Bacer, S., Coppola, E., Giorgi, F., Ahrens, B., Dubois, C., Sanchez, E., and Verdecchia, M.: Assessment of multiple daily precipitation statistics in ERA-Interim driven Med-CORDEX and EURO-CORDEX experiments against high resolution observations, pp. 1–24, https://doi.org/10.1007/s00382-016-3453-4, 2016.

Felder, G., Gómez-Navarro, J. J., Zischg, A., Raible, C. C., Röthlisberger, V., Bozhinova, D., Martius, O., and Weingartner, R.: From global circulation to local flood loss: Coupling models across the scales, Science of the Total Environment, 635, 1225–1239, https://doi.org/10.1016/j.scitotenv.2018.04.170, 2018.

Froidevaux, P. and Martius, O.: Exceptional integrated vapour transport toward orography: an important precursor to severe floods in Switzerland: Integrated Vapour Transport and Floods in Switzerland, 142, 1997–2012, https://doi.org/10.1002/qj.2793, 2016.

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., Lawrence, D. M., Neale, R. B., Rasch, P. J., Vertenstein, M., Worley, P. H., Yang, Z.-L., and Zhang, M.: The Community Climate System Model Version 4, Journal of Climate, 24, 4973–4991, https://doi.org/10.1175/2011JCLI4083.1, 2011.

Giorgi, F., Torma, C., Coppola, E., Ban, N., Schär, C., and Somot, S.: Enhanced summer convective rainfall at Alpine high elevations in response to climate warming, 9, 584–589, https://doi.org/10.1038/ngeo2761, 2016.

Gómez-Navarro, J. J., Montávez, J. P., Jerez, S., Jiménez-Guerrero, P., and Zorita, E.: What is the role of the observational dataset in the evaluation and scoring of climate models?, Geophysical Research Letters, 39, L24 701–L24 701, https://doi.org/10.1029/2012GL054206, 2012.

Gómez-Navarro, J. J., Raible, C. C., and Dierer, S.: Sensitivity of the WRF model to PBL parametrisations and nesting techniques: evaluation of wind storms over complex terrain, Geoscientific Model Development, 8, 3349–3363, https://doi.org/10.5194/gmd-8-3349-2015, 2015.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E., and Engen-Skaugen, T.: Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods, Hydrology and Earth System Sciences, 16, 3383–3390, https://doi.org/10.5194/hess-16-3383-2012, 2012.

Haylock, M. R., Hofstra, N., Klein Tank, a. M. G., Klok, E. J., Jones, P. D., and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, Journal of Geophysical Research, 113, D20 119–D20 119, https://doi.org/10.1029/2008JD010201, http://doi.wiley.com/10.1029/2008JD010201, 2008.

Hong, S. and Lim, J.: The WRF single-moment 6-class microphysics scheme (WSM6), Journal of Korean Meteorology Society, 42, 129–151, 2006.

Hong, S.-Y., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, Monthly Weather Review, 134, 2318–2341, https://doi.org/10.1175/MWR3199.1, 2006.

Houze, R. A., James, C. N., and Medina, S.: Radar observations of precipitation and airflow on the Mediterranean side of the Alps: Autumn 1998 and 1999, 127, 2537–2558, https://doi.org/10.1002/qj.49712757804, 2001.

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., Lamarque, J.-F., Large, W. G., Lawrence, D., Lindsay, K., Lipscomb, W. H., Long, M. C., Mahowald, N., Marsh, D. R., Neale, R. B., Rasch, P., Vavrus, S., Vertenstein, M., Bader, D., Collins, W. D., Hack, J. J., Kiehl, J., and Marshall, S.: The Community Earth System Model: A framework for collaborative research, Bulletin of the American Meteorological Society, 94, 1339–1360, https://doi.org/10.1175/BAMS-D-12-00121.1, 2013.

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, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/CBO9781107415324, 2013.

IPCC-SREX: Managing the risks of extreme events and disasters to advance climate change adaption: special report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 582, 2012.

Isotta, F. A., Frei, C., Weilguni, V., Perčec Tadić, M., Lassègues, P., Rudolf, B., Pavan, V., Cacciamani, C., Antolini, G., Ratto, S. M., Munari, M., Micheletti, S., Bonati, V., Lussana, C., Ronchi, C., Panettieri, E., Marigo, G., and Vertačnik, G.: The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data: CLIMATE OF DAILY PRECIPITATION IN THE ALPS, 34, 1657–1675, https://doi.org/10.1002/joc.3794, 2014.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, 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, https://doi.org/10.1007/s10113-013-0499-2, 2013.

Jiménez, P. A. and Dudhia, J.: Improving the representation of resolved and unresolved topographic effects on surface wind in the WRF model, Journal of Applied Meteorology and Climatology, 51, 300–316, https://doi.org/10.1175/JAMC-D-11-084.1, 2012.

Kain, J. S.: The Kain–Fritsch convective parameterization: An update, Journal of Applied Meteorology, 43, 170–181, https://doi.org/10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2, 2004.

Keller, K. M., Joos, F., Lehner, F., and Raible, C. C.: Detecting changes in marine responses to ENSO from 850 to 2100 C.E.: Insights from the ocean carbon cycle, Geophysical Research Letters, 42, 2014GL062 398, https://doi.org/10.1002/2014GL062398, http://onlinelibrary.wiley.com/doi/10.1002/2014GL062398/abstract, 2015.

Kotlarski, S., Keuler, K., Christensen, O. B., Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., van Meijgaard, E., Nikulin, G., Schär, 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, https://doi.org/10.5194/gmd-7-1297-2014, 2014.

Lehner, F., Joos, F., Raible, C. C., Mignot, J., Born, A., Keller, K. M., and Stocker, T. F.: Climate and carbon cycle dynamics in a CESM simulation from 850 to 2100 CE, Earth System Dynamics, 6, 411–434, https://doi.org/10.5194/esd-6-411-2015, 2015.

Maraun, D.: Bias correcting climate change simulations – a critical review, Current Climate Change Reports, 2, 211–220, https://doi.org/10.1007/s40641-016-0050-x, 2016.

Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J. M., Hagemann, S., Richter, I., Soares, P. M. M., Hall, A., and Mearns, L. O.: Towards process-informed bias correction of climate change simulations, Nature Climate Change, 7, nclimate3418, https://doi.org/10.1038/nclimate3418, 2017.

Martius, O., Zenklusen, E., Schwierz, C., and Davies, H. C.: Episodes of alpine heavy precipitation with an overlying elongated stratospheric intrusion: a climatology, 26, 1149–1164, https://doi.org/10.1002/joc.1295, http://doi.wiley.com/10.1002/joc.1295, 2006.

Messmer, M., Gómez-Navarro, J. J., and Raible, C. C.: Climatology of Vb cyclones, physical mechanisms and their impact on extreme precipitation over Central Europe, Earth System Dynamics, 6, 541–553, https://doi.org/10.5194/esd-6-541-2015, 2015.

Messmer, M., Gómez-Navarro, J. J., and Raible, C. C.: Sensitivity experiments on the response of Vb cyclones to sea surface temperature and soil moisture changes, Earth System Dynamics, 8, 477–493, https://doi.org/10.5194/esd-8-477-2017, 2017.

MeteoSwiss: Daily precipitation (final analysis): RhiresD, http://www.meteoswiss.admin.ch/content/dam/meteoswiss/de/service-und-publikationen/produkt/raeumliche-daten-niederschlag/doc/ProdDoc_RhiresD.pdf, 2016.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A.: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, Journal of Geophysical Research: Atmospheres, 102, 16 663–16 682, https://doi.org/10.1029/97JD00237, 1997.

Montesarchio, M., Zollo, A. L., Bucchignani, E., Mercogliano, P., and Castellari, S.: Performance evaluation of high-resolution regional climate simulations in the Alpine space and analysis of extreme events, Journal of Geophysical Research: Atmospheres, 119, 2013JD021 105, https://doi.org/10.1002/2013JD021105, 2014.

Nikulin, G., Bosshard, T., Yang, W., Bärring, L., Wilcke, R., Vrac, M., Vautard, R., Noel, T., Gutiérrez, J. M., Herrera, S., Fernández, J., Haugen, J. E., Benestad, R., Landgren, O. A., Grillakis, M., Ioannis, T., Koutroulis, A., Dosio, A., Ferrone, A., , and Switanek, M.: Bias Correction Intercomparison Project (BCIP): an introduction and the first results, vol. 17, Europen Geosciences Union, abstract number 2250, 2015.

PAGES 2k-PMIP3 group: Continental-scale temperature variability in PMIP3 simulations and PAGES 2k regional temperature reconstructions over the past millennium, Clim. Past, 11, 1673–1699, https://doi.org/10.5194/cp-11-1673-2015, http://www.clim-past.net/11/1673/2015/, 2015.

Panziera, L., James, C. N., and Germann, U.: Mesoscale organization and structure of orographic precipitation producing flash floods in the Lago Maggiore region: Orographic Convection in the Lago Maggiore Area, 141, 224–248, https://doi.org/10.1002/qj.2351, 2015.

Pielke Sr, R. A.: Mesoscale meteorological modeling, vol. 98, Academic press, 2013.

Rajczak, J., Pall, P., and Schär, C.: Projections of extreme precipitation events in regional climate simulations for Europe and the Alpine region, Journal of Geophysical Research: Atmospheres, 118, 3610–3626, https://doi.org/10.1002/jgrd.50297, 2013.

Rajczak, J., Kotlarski, S., and Schär, C.: Does quantile mapping of simulated precipitation correct for biases in transition probabilities and spell-lengths?, Journal of Climate, 29, 1605–1615, https://doi.org/10.1175/JCLI-D-15-0162.1, 2016.

Rajczak Jan and Schär Christoph: Projections of Future Precipitation Extremes Over Europe: A Multimodel Assessment of Climate Simulations, 122, 10,773–10,800, https://doi.org/10.1002/2017JD027176, 2017.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, Journal of Climate, 24, 3624–3648, https://doi.org/10.1175/JCLI-D-11-00015.1, 2011.

Schmidt, G. A., Jungclaus, J. H., Ammann, C. M., Bard, E., Braconnot, P., Crowley, T. J., Delaygue, G., Joos, F., Krivova, N. A., Muscheler, R., Otto-Bliesner, B. L., Pongratz, J., Shindell, D. T., Solanki, S. K., Steinhilber, F., and Vieira, L. E. A.: Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0), Geoscientific Model Development, 4, 33–45, 2011.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W., and Powers, J. G.: A description of the advanced research WRF version 3, Tech. Rep. TN–475+STR, National Center for Atmospheric Research, 2008.

Stocker, T. F., Dahe, Q., Plattner, G.-K., and Tignor, M.: IPCC Workshop on Regional Climate Projections and their Use in Impacts and Risk Analysis Studies, chap. Breakout Group 3bis: Bias Correction, pp. 21–23, IPCC Working Group I Technical Support Unit, University of Bern, Falkenplatz 16, 3012 Bern, Switzerland, 2015.

Taylor, K. E., Stouffer, R. J., and Meehl, G. a.: An overview of CMIP5 and the experiment design, Bulletin of the American Meteorological Society, 93, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1, 2012.

Teng, J., Potter, N. J., Chiew, F. H. S., Zhang, L., Wang, B., Vaze, J., and Evans, J. P.: How does bias correction of regional climate model precipitation affect modelled runoff?, 19, 711–728, https://doi.org/10.5194/hess-19-711-2015, 2015.

Teutschbein, C. and Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, Journal of Hydrology, 456-457, 12–29, https://doi.org/10.1016/j.jhydrol.2012.05.052, 2012.

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

Themeßl, M. J., Gobiet, A., and Heinrich, G.: Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal, 112, 449–468, https://doi.org/10.1007/s10584-011-0224-4, 2012.

Torma, C., Giorgi, F., and Coppola, E.: Added value of regional climate modeling over areas characterized by complex terrain—Precipitation over the Alps, 120, 3957–3972, https://doi.org/10.1002/2014JD022781, 2015.

Wang, C., Zhang, L., Lee, S.-K., Wu, L., and Mechoso, C. R.: A global perspective on CMIP5 climate model biases, Nature Climate Change, 4, 201–205, https://doi.org/10.1038/nclimate2118, 2014.

Ward, J. H.: Hierarchical grouping to optimize an objective function, Journal of the American Statistical Association, 58, 236–244, https://doi.org/10.1080/01621459.1963.10500845, 1963.

Wetterhall, F., Pappenberger, F., He, Y., Freer, J., and Cloke, H. L.: Conditioning model output statistics of regional climate model precipitation on circulation patterns, 19, 623–633, https://doi.org/10.5194/npg-19-623-2012, 2012.

Table 1. Number of regions obtained after the cluster analysis of daily precipitation. The shape of such regions is shown in Fig. 2. The number of EOFs retained is kept to 7 in all cases, which corresponds to a explained variance above 80% in all cases.

height	OBS	WRF-CESM WRF-ERA	WRF-CESM
DJF	5	6 4	6
MAM	5	6 5	6
JJA	5	5	5
SON	5	6	4

Table 2. Number of grid points that belong to each of the regions shown in Fig. 2. Only grid points within the Swiss domain, i.e. those not missing values in OBS, are counted. Note that in some cases the number of regions is lower than 6, therefore we indicate it with ah dash.

	6 DJF			MAM			JJA			SON		
	OBS	WE	WC	OBS	WE	WC	OBS	WE	WC	OBS	WE	WC
heightReg. 1	1233	1830	1719	1017	1956	1800	897	1746	1293	1116	1812	2193
Reg. 2	1203	954	579	837	471	438	846	579	618	945	492	693
Reg. 3	738	564	372	825	708	678	822	777	786	786	747	606
Reg. 4	375	435	708	771	426	294	735	327	630	471	291	291
Reg. 5	234	–	345	333	222	246	483	354	456	465	183	–
Reg. 6	–	–	60	–	–	327	–	–	–	–	258	–

Number of regions obtained after the cluster analysis of daily precipitation. The number of EOFs retained is kept to 7 in all cases, which corresponds to a variance explained above 80% in all cases.

Figure 1. Top: configuration of the four nested domains used in both the WRF-ERA and WRF-CESM simulations. Bottom: detail of the actual orography implemented in the 2-km resolution simulation over Switzerland.

[redacted]

Figure 2. Regions obtained from the cluster analysis described in Sec. 3. The maps correspond to the 12 possible combinations, 3 for each dataset (OBS, WRF-ERA and WRF-CESM) and 4 for each season. Note that the colors are set arbitrarily as a label within the algorithm, so no one-to-one correspondence is implied between regions of the same colour in different maps.

[redacted]

Figure 3. Temporal cross-correlation matrices between all regional series. The calculation, as the definition of regions, is carried out independently for each dataset and season. The order of matrices is from region 1 (bottom-left) to region 6 (top-right), and the spatial distribution of the regions is shown in Fig. 2. Note that all matrices are symmetric with 1 across the diagonal.

[redacted]

Figure 4. Taylor diagram showing the temporal correlation and ratio of standard deviation between the regional series in the WRF-ERA simulation and the observations across all 4 seasons. For obtaining the regional series, the regions defined for WRF-ERA are used in both datasets. Different symbols denote the result for each season, whereas the colours correspond to the different regions according to the legend and spatial structure shown in middle column in Fig. 2.

Figure 5. Seasonal cycle of monthly precipitation over Switzerland in the observations (black), the WRF-ERA simulation (blue), the WRF-CESM simulation (red), and bias-corrected WRF-CESM simulation (green).

Figure 6. Mean seasonal accumulated precipitation over Switzerland across seasons (different rows) in the gridded observations (first column), in the WRF-ERA simulation (second column), in the WRF-CESM simulation (third column) and the bias-corrected WRF-CESM simulation (forth column).

Figure 7. Estimated PDFs of daily precipitation averaged over Switzerland. Each panel depicts the result for a season, and different colors are representative of the results for different datasets according to the choice in Fig. 5. Note the logarithmic scale in the x axis, which precludes the area below all curves being equal.

Figure 8. Correlation maps between the daily series of precipitation in the raw WRF-CESM simulation and the output of the bias corrected. The analysis is carried out separately by seasons to minimize the effect of the annual cycle on correlation.