Reply on RC3

Joel Fiddes et al.

Author comment on "TopoCLIM: rapid topography-based downscaling of regional climate model output in complex terrain v1.1" by Joel Fiddes et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-60-AC3, 2021

REVIEWER #3

We thank the reviewer for his thoughtful comments which have undoubtedly helped to improve this work. We also apologize for the somewhat delayed response. We provide here a point-by-point response where reviewer comments are in bold type and author responses in non-bold. Changed or inserted text is given in italics.

Code:

The provided methods could be highly useful for many people requiring downscaled climate model data for running impact models. Unfortunately, the source code in its current form is in my opinion not yet usable for the community.

First, there is no manual. The GitHub repository only contains a very short Readme explaining how to download the model and run the examples, but no information whatsoever on how to apply the methods for one's own purposes. Also, the code is quite messy. Most importantly, the parameters of the methods should not be set directly in the source files (parameters and code should always be separated (not only) for reusability and reproducibility reasons). Ideally, the code should be provided in the form of a Python package e.g. with individual functions for each method which accept their respective parameters as arguments. This would allow to run the methods from Python directly instead of having to manually execute a number of Python scripts in a particular sequence. Finally, even getting the example to run is not straightforward with the provided instructions. Running it seems to require an installation of R including the packages ncdf4 and qmap (which I found out by trial and error), but I still could not get it to run. It crashes when trying to run the R script qmap_hour_plots_daily_12.R: Error in cordex_dates_cp : object 'cordex_dates_cp' not found subprocess.CalledProcessError: Command '['Rscript', '../rsrc/qmap_hour_plots_daily_12.R', '../examples/', '_1D', '_1D.csv', '9.809', '46.83', '../examples//cordex']' returned non-zero exit status 1.

I would encourage the authors to rework the code and to provide a proper manual (or extended Readme) containing at least a description of the individual

methods and how to run them, their parameters, as well as the required prerequisites apart from the Python libraries (R + packages, cdo etc.) and the supported Python versions.

We realise that the code was not only poorly documented but focused on the analysis presented in the paper. We have now done the following to make the code more clear, simple and generic to user needs, including:

- Resolved dependency issue, all dependencies (except R) are now handled by conda package manager (dependency related issue was responsible for crash the reviewer experienced)
- Reduced dependencies to make the install smoother.
- Reworked code (removed all paper specific routines (eval plots etc) which are not generic.
- Simplified ESGF helper routines so they are more generic for application to other CORDEX domains and removed inscript parameters in these helper scripts.
- Documented functions using docstrings (in script) .
- Reworked and extended the README.
- Added Python version information.

Paper:

General comments:

The paper is concise and well-written. My only major remark is that in several parts of the paper it is not immediately clear which parts are "hardcoded" in the TopoCLIM methods and which are only exemplarily used for the particular snow modeling application. E.g., does the scheme currently only allow to use CORDEX and ERA5 data or is it already possible to use other data sets (as mentioned in the conclusions)? It would be very helpful to the reader if this would be made more specific (see also my individual remarks below).

We hope we have addressed this adequately in the various specific comments below. As a very general response the basic principle of generating a pseudo-observation using TopoSCALE and using this to downscale climate data to produce forcings suitable for driving an impact model are of course (in principle) generalisable to other climate model datasets and other reanalysis. However the devil is in the details, as some of the current code base is data-management and I/O which of course is specific to the used datasets and data structures. Nonetheless, CMIP6 could for example be accessed and preprocessed without much additional effort due to a similar data structure to CORDEX. However, this step has not been yet made. We would point out that CORDEX and ERA5 permit global studies to be conducted and therefore of wide interest, as is. In the future a generalisation of the methods to accommodate other datasets (with a plugin structure) would be worthwhile.

Remembering the several schemes/data sets and their acronyms (CLIM, T-MET, T-CLIM, QM, QM_MONTH, …) can get quite challenging; maybe consider adding a table listing all of the acronyms and their meaning.

We have added the data set acronyms to Figure 1 which hopefully adds clarity and concentrates this overview in a single Figure. We hope this makes an additional table unnecessary.

Both "snow height" and "snow depth" are used throughout the paper, ideally this should be consistent.

Changed to snow depth throughout.

Specific comments:

Title: is the first dot in "v.1.0" intentional or should it be v1.0?

Edited to v1.0.

Abstract: references should be avoided in the abstract.

Removed.

Section 2.1:

When reading this, it is not really clear if the procedure refers to a single CORDEX grid point or if the method considers all grid points within a specified region.

It applies to however many grid points are in a given domain. In fact the spatial model is defined by the TopoSCALE procedure which could generate forcings for a single or multiple points on the Earth's surface. These are associated with the closest CORDEX grid centroid on a nearest neighbour basis for the quantile mapping step. This extends the same procedure described in Fiddes et al. 2015 to climate data. We have tried to clarify this in the text as:

"and additionally generalising this to a map product using the subgrid scheme, TopoSUB (Fiddes & Gruber 2012), which efficiently spatialises 1D model results (multiple subgrid samples per CORDEX gridbox) to a map domain according to important dimensions of land surface heterogeneity. In this way, hillslope resolution (100 m) map results are generated with a laptop-feasible number of subgrid simulations (in this case 100) per large scale CORDEX gridbox. "

Section 2.2:

It would be helpful to list exactly which preprocessing steps are performed with the CORDEX data instead of providing only some examples. This could be done in the text and/or integrated in Fig. 1 (i.e., listing all the steps instead of the generic "Preprocess CORDEX" block).

We actually don't do more than is described so this was a slightly misleading use of "such as". We have edited for clarity:

"All preprocessing of raw CORDEX data (Figure 1): concatenating NetCDF time series, extracting region of interest and regridding from rotated pole projections, is accomplished using standard tools from the Climate Data Operators (CDO) suite."

We haven't, however, added this to Fig 1 as we feel that this extra detail clutters what is meant to be a high-level overview - to make the conceptual flow of the scheme clear. We think that giving this detail in the text is sufficient.

"The CDO tools are incorporated into ..." is a little ambiguous – I assume the tools aren ot directly integrated into the package but have to be pre-installed and are called from within the package?

Yes this is correct, we have edited the text to reflect this as:

"The CDO tools are called from the preprocessing module of TopoCLIM and not used as standalone command line tools."

Section 2.4:

Again, here it is not clear if the two time periods (1980-1995 and 1996-2006) are fixed in the method or if these are only used for this particular study.

These are just used in this study. Clarified as:

"It should be noted that these periods are constraints imposed by the datasets used in this study and can be changed in other applications of the method."

Section 2.5: Please consider adding some more detail about the used disaggregation functions for the different variables (as these can likely have considerable impact on the impact modeling results).

We have added the methods used to Table 1 as this seems to make most sense. To accommodate this extra column we have removed the CF standard name column, as this was somewhat redundant (we still give the CF long name).

"An adapted version of the Melodist package" - adapted in which way?

This was written somewhat inaccurately since we used the melodist package as is. We have edited the sentence to simply:

The "Melodist" package is used for this purpose \citep{Forster2016-xx}.

However not all variables are covered by Melodist (air pressure and incoming longwave) so we augmented Melodist with our own methods, as now described in the text:

"Melodist does not provide methods for air pressure or longwave radiation, these are handled with the following procedure: (1) taking advantage of the relationship between incoming longwave radiation (ILWR) and air temperature (TA):

ILWR = ϵ σ * TA^4*

where σ is the Stefan-Boltzmann constant (5.67×10−8 W/m2·K4) we diagnosed the daily all sky emissivity (ϵ). We then used ϵ as a daily scaling factor to convert disaggregated TA into ILWR. This procedure therefore assumes a constant ϵ at sub daily timestep (which of course will not normally be true) yet ensures that ILWR scales correctly with TA. Therefore higher TA lead to higher ILWR values and vice-versa. Air pressure is simply linearly interpolated to the sub-daily timestep."

Section 2.6:

Again here it is not clear if this is part of the TopoCLIM method or only part of the example application for using TopoCLIM-generated data in an impact model. Since I assume the latter is the case, this section (along with 2.7) should probably be moved e.g. to section 4?

We have moved both Section 2.6 and 2.7 to Section 4 and agree it makes more sense like this to separate out core methods and study setup.

"Typical setups use ..." - perhaps add some more details about a typical setup

(region size, resolution).

To give a concrete example we have added: *"for example to produce the results given in Figure 7."*

Section 3.2:

What does "globally" (L172) mean in this context?

It means that while the high resolution dataset would of course be advantageous, the fact that 11 km CORDEX is not available globally (as compared to the 44 km product) means that it does not fit the design spec of what we try to do in developing and testing methods that can be used anywhere on Earth (=globally in this context).

Section 5.1:

Last paragraph: the agreement seems to be good for only some of the years in the mentioned periods. However, individual years should probably generally not be compared to observations for both the historical and the scenario period.

We have added the following text to Section 5.1:

Available snow depth observations from WFJ show good agreement with both the historical period and first decade of the RCP runs in terms of lying within the model ensemble. It should be stressed that we do not attempt a quantitative comparison here with individual years as CORDEX variability variability(and climate models more generally) is not expected to be perfectly synchronised with observed variability at such temporal resolutions.

Section 5.2:

If I understand correctly, DOY is used here as "days since September 1"? If so, this is really confusing, since the term DOY has a very specific meaning (with 1 being January 1). I suggest to either use the actual DOY in Fig. 6 or use another term instead of DOY.

True, we have changed both axis and caption to *day of water year (DOWY)*

Section 5.3:

The section title is confusing, since the previous results also already included TopoCLIMresults and climate change impacts on the Alpine snow cover (albeit on the point scale).

We agree this was somewhat confusing - to address this we have changed the section titles to be more precise:

5.1 Evaluation at the Weissfluhjoch station

5.2 Evaluation for the IMIS network

5.3 Climate change impacts on Alpine snow cover across Switzerland

"by coupling TopoCLIM with the TopoSUB spatial framework" - and with FSM in between, correct?

Yes correct, we clarified the sentence to:

"As an example application of the full model pipeline, the results in Figure 7 were generated by feeding model results (TopoCLIM/FSM) to the TopoSUB spatial framework to generate transiently modelled snow depth maps at 100 m resolution."

Fig. 1:

As stated above, this figure contains some very specific terms and dates (CORDEX,

ERA5, 1980-2100, 1980-2020, 90 m, 1980-2100). Since the figure should be a general overview of the TopoCLIM method I would remove all terms and dates which are not "hardcoded" in TopoCLIM.

We prefer the detail given as it gives an overview of date ranges used in the study, which we think is important for the high level overview that Figure 1 aims to achieve. However, the point is well taken and we have edited the caption to make it clear that these are indicative for this study:

"Figure 1: Schematic overview of the main TopoCLIM processes and experimental setup used in this study."

There is a typo in "Qantlle map".

Corrected

Fig. 2:

"with CORDEX grid overlaid" - please specify which grid (44 km?).

Edited to: *"with CORDEX EUR-44 grid (44 km) overlaid"*

Fig. 6:

The term TSCALE appears only in this figure. Probably this should be T-MET?

Corrected to T-MET in legend.

Fig. 7:

The figure has a very poor resolution.

We have improved the resolution, this also addresses plot artefact noticed by reviewer 2.

Figure caption: "Example TopoCLIM climate change maps" - maybe rephrase, since the maps are not a direct TopoCLIM result but a combination of TopoCLIM, FSM and TopoSUB.

Edited to: *"Example TopoCLIM forced climate change maps:"*

Table 1:

The table is very useful, but as far as I see it is not referenced from the paper. Consider adding a reference e.g. at the beginning of the results section.

We added the reference to the section describing the climate data, L.178-79 now reads:

 "A full description of CORDEX variables is given in Table 1 and model chains used in Table 2."

What do the timesteps of 3 and 6 hours mean?

This is a legacy mistake from previous work. Timestep has been corrected to "daily" in all cases.

Minor remarks:

P3 L65: Add reference to ERA5 here (currently in section 3.3) if this is the first Mention.

Moved reference as suggested.

P3 L69: Either write "a subgrid clustering scheme" or "the subgrid clustering scheme TopoSUB".

Corrected to the second suggestion

P4 L118: Add a colon at the end of the line.

Done.

P5 L130 and P8 L228: replace QMAP_MONTH by QM_MONTH as in the rest of the paper (assuming this does not actually refer to a different variant).

Corrected for consistency (as addressed also by Reviewer 2).

P5 L145: Remove the second "model".

Done.

P6 L173: Capitalize "python".

Done.

P7 L202: Adding "(cm)" is probably not necessary.

Removed.

P7 L207: Replace 2031-60 and 2070-100 with 2031-2060 and 2070-2100.

Done.

P8 L219: The term "QM_QM" appears only here and in the caption of Fig. 3, I assume this is what is referred to as only "QM" in the rest of the paper?

Corrected for consistency (as addressed also by Reviewer 2)

P8 L225: Typo ("erro").

Corrected.

P10 L283: "25 km" -> not 30 km (as stated in section 3.3)?

Corrected to 0.25 degrees throughout.

P10 L293: The acronyms (TA, ILWR, ISWR) appear here for the first and only time; please write the full variable names.

Done as suggested.