Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-154-AC3, 2016

Geoscientific
Model Development
Discussions
Open Access

Interactive comment on "Reinitialised versus continuous regional climate simulations using ALARO-0 coupled to the land surface model SURFEX" *by* Julie Berckmans et al.

Julie Berckmans et al.

julie.berckmans@meteo.be

Received and published: 4 November 2016

[1,2]JulieBerckmans [1,2]OlivierGiot [1,3]RozemienDe Troch [1,3]RafiqHamdi [2]ReinhartCeulemans [1,3]PietTermonia

[1]Royal Meteorological Institute, Brussels, Belgium [2]Centre of Excellence PLECO (Plant and Vegetation Ecology), Department of Biology, University of Antwerp, Antwerp, Belgium [3]Department of Physics and Astronomy, Ghent University, Ghent, Belgium

Julie Berckmans (julie.berckmans@meteo.be)

11

November 4, 2016

1 Reply to the Editor

Dear Editor,

we have prepared a majorily improved version of the manuscript by incorporating all suggestions and critical comments raised by both reviewers. Please find attached our response to the referees' comments on our above mentioned manuscript, titled "Reinitialised versus continuous regional climate simulations using ALARO-0 coupled to the land surface model SURFEX". Below mentioned you will find our detailed responses to all the reviewers' comments and suggestions (put in italics and red). We have also explained where and how they were incorporated in the revised manuscipt.

2 Reply to Anonymous Referee #1

We would like to thank the Anonymous Referee #1 for the encouraging and constructive comments, which have improved the manuscript. Below is a list of modifications that we have implemented based on your major and specific comments. The supplemented "manuscript_figures.zip" contains Figures 1 and 2, together with the marked-up manuscript showing the changes made in the new version.

Major comments:

(1) I was disappointed on the level of detail describing all the different models referred to in this manuscript: ALARO-0, SURFEX, ISBA, TEB and ALADIN and found that if I wanted a reasonable understanding of what the key parameterizations were or the runtime options used or the hierarchy of these models I would need to consult several other manuscripts. . .an exhausting exercise when there are so many models referred to here. Furthermore, I found it confusing to follow how the model is run. In RCMs like WRF, reanalysis (or GCM) data is used to update the BCs on a 6 hourly interval, which is necessary when running climate simulations (rather than short-term forecasts) to avoid drift. Therefore from the beginning I was confused by what the authors mean by "continuous" and "reinitialized" for an RCM when running a 10 year simulation. What variables are reinitialized, how are they reinitialized (at the boundaries or across the domain) and at what frequency? This information was not clearly articulated, making it difficult for someone to reproduce the experimental design.

The level of detail for the different models has been increased in the revised manuscript for all models described. The setup of the experiment has been clarified, using a schematic diagram (Fig. 2 in revised manuscript), showing the different downscaling simulations. The meaning of continuous and reinitialised is better defined and supported by the diagram. The wording "initial conditions" and "lateral boundary conditions" has been used in its correct context, to get a proper understanding of the differences between the downscaling simulations. The new version (Page 4 Lines 26-28) reads:
"The zonal and meridional wind components, atmospheric temperature, specific humidity, surface pressure and surface components were provided every 6 hrs as lateral boundary conditions and interpolated hourly. They were introduced as initial

conditions accross the domain."

(2) Although there are 3 objectives detailed in the introduction, I could see two possible aims of the manuscript: (a) Trying to show that forecast skill is improved when a new land surface scheme is added (CRDX vs. CON) – which is documented in Hamdi et al. 2014 (b) Trying to show that forecast skill is improved when reinitializing daily (FS vs. CON) If the aim is to present the benefits of the daily reinitialization, then perhaps excluding the CRDX results would improve the focus of the manuscript.

The aim of the manuscript was to present the improved simulation skill of the daily reinitialisation. It is a good suggestion to focus on this objective. We have therefore removed the first part on the comparison with the EURO-CORDEX simulations from the revised manuscript. The latter will be used in a separate study.

(3) There are different spin-up periods for FS (3 months) and CON (1 year). No explanation is provided for why the set up is different between the experiments. In particular, if the same land surface scheme is used, then best practice would be to have the same spin-up period for the land-surface state variables (soil moisture and soil temperature). If FS and CON have different spin-up lengths then how can the authors be sure that the differences in simulation skill are due to the "continuous/reinitialized" configuration and not the spin-up? This would actually make me advocate that the whole experiment needs to be run again with same configuration (e.g. Bcs, spin-up length, IC) so that differences in skill between "continuous" or "reinitialized" runtime modes can be fairly evaluated. At the moment I don't think this is really possible.

This is a good point. Based on a general consensus within the climate community, we have used one-year spin-up for the continuous simulation. When making FS simulations, we generally run each year in parallel to reduce computing time. Therefore, we apply the smallest possible spin-up time which still provides an equilibrium state. Past tests with different initialisation times and periods for a domain covering Belgium and the neighbouring countries showed that starting in March and keeping a spin-up time of three months is reasonable for the model to come into an equilibrium state (see Fig. 1).

Fig. 1 represents the deep soil moisture simulated by the three dynamical downscaling approaches DRI, FS, and CON, for the period of January 1990 to January 1991. The initial conditions for the surface restarted every day within DRI. For CON and FS, the initial conditions for the surface were initialised once at the beginning of each run. For FS the initialisation of the surface in March 1990 resulted in a shorter time period to reach equilibrium than the initialisation in January or February 1990. When starting in March 1990, the FS lines were close in June and thereafter. This means that the variability within FS is smaller than the variability between FS and CON. However, when starting in January, we would have to spin-up the model for 6 months until July to reach equilibrium. A similar test was done for our study domain. Fig. 2 shows the deep soil moisture simulated by the two dynamical downscaling approaches CON and FS, for the period of March 1991 to August 1991. FS SHORTSPINUP shows the deep soil moisture when starting the spin up in March 1991, while FS LONGSPINUP shows the deep soil moisture when starting the spin up the previous year in March 1990. The variability within FS is smaller than the variability between FS and CON after 3 months, in June. Both results show that the different spin-up periods for CON and FS do not impact the simulation skill of the two downscaling simulations, whereas the different configurations do impact the simulation skill.
The new version (Page 5 Lines 21-22):
"Although CON required one year spin-up time, 3 months were sufficient for the FS deep soil moisture to reach equilibrium state, when starting in March (not shown)."

(4) I wasn't convinced at all by the analysis on the land-atmosphere feedback. Perhaps it would be good to consider the coupling metrics in detailed in Lorenz et al. (2015) that

are suitable for the fully coupled simulations. It would be more convincing to calculate the coupling metrics for each experiment independently and then evaluate the difference between FS and CON to examine changes in coupling. However, is this analysis relevant here, given that the differences between CON and FS is the frequency that the lateral BCs are updated, not the land surface state as understood from Page 4/5: "The soil variables evolved freely after the first initialisation and were never corrected or nudged in the course of the simulation."

We agree with the comment of the reviewer; consequently we removed the correlation analysis from the revised manuscript. In the revised manuscript we now concentrate the analysis on the mean state of the atmosphere and the surface, but we do not investigate the feedback. However, we still evaluate the diurnal cycle of the surface energy fluxes, as they are impacted by a different response of the surface to the continuous vs. reinitialised atmosphere. This has been done in section 4 in the revised manuscript, where we validated the spatial distribution of the Bowen Ratio and the diurnal cycle of the surface energy fluxes. The coupling metrics as in Lorenz et al. (2015) might be interesting for a future separate study.

(5) I am a bit concerned about the limited number of sites used to evaluate the experiments against FLUXNET. This is likely due to the choice of simulation period (1991-2000) where FLUXNET data coverage is limited. It sounds like the authors were already aware of this limitation too. Perhaps a more recent simulation period would resolve this issue when more FLUXNET sites are available for a more rigorous validation. Alternatively, the authors could consider the LandFLUX (Mueller et al., 2013) or GLEAM (Miralles et al., 2013) datasets to validate the surface fluxes more comprehensively.

The simulations used for this study, driven by ERA-Interim, were done in parallel to the model simulations driven by the global climate model ARPEGE CMIP5 (1976-2005). The period of 1991-2000 was chosen to have an overlapping period for comparison.

At the onset of the present study we had not decided to investigate the surface fluxes. This decision came later on, to look more into detail at the surface processes to explain the differences between the downscaling approaches. In particular, the FLUXNET database was selected for the model validation as our research group is familiar with this network. Besides, one of the co-authors is PI of a FLUXNET station (not used in this study).

(6) The surface fluxes in the FS and CON configurations are also only evaluated in a second set of shorter simulations (3 months) rather than the original 10 year simulations. It would make better sense to evaluate the fluxes and land-atmosphere feedback in the 10 year simulations given that the purpose of the manuscript is to evaluate the simulation skill of long simulations. Unfortunately this provides the reader with the impression that the experimental design was either poorly designed or that a random bunch of simulations with different set-ups were cobbled together to evaluate the different runtime modes.

We agree with the reviewer's comment. In the revised manuscript the fluxes were calculated for 10 years of summer, to make this consistent with the overall setup of the experiment. As the observations from FLUXNET only provide data from 1996 onwards, the overlapping period with the model is 5 years. Therefore, the Bowen Ratio was presented in the revised manuscript for 5 years, as well as the results for the stations. Additionally, the Bowen ratio was presented for 10 years, to indicate that the 5-year period is still robust for the 10-year period, and can be used for the model validation of the fluxes.

(7) Due to the writing style, I found the paper hard to read in many places. The structure also requires refinement, as there are many instances where information is provided in the wrong section that would be more useful in another.

The structure of the revised text has been improved. The revised text now fits better within each of the sections.

Specific comments:

Abstract

Here I got the impression that the manuscript was about evaluating the updated land surface scheme rather than the different running modes. Please revise the abstract to appropriately reflect the aim and scope of the manuscript and the key results.

The Abstract has been revised as suggested.
The new version (Page 1 Lines 4-9) reads:
"We evaluated the dependence of the simulation potential on the running mode of the ALARO model coupled to the land surface model SURFEX, driven by the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Re-Analysis (ERA-Interim) data. Three types of downscaling simulations were carried out for a 10-year period covering 1991 to 2000, over a Western European domain at 20 km horizontal resolution: ..."

Please define all acronyms.

This has been done in the revised manuscript. All acronyms have been properly and clearly defined.

It is perhaps not necessary to mention the ALADIN modeling system here to avoid overwhelming the reader with acronyms.

We appreciate the reviewer's suggestion. The acronym has not been mentioned anymore in the revised Abstract.

Sentence starting "This contribution . . ." perhaps better to say "We evaluate the dependence of simulation skill on the running mode (continuous or reinitialized) of the ALARO-0 model."

We thank the reviewer for the suggestion; the sentence has been reworded in the revised manuscript.
The new versions (Page 1 Lines 4-5):
"We evaluated the dependence of the simulation potential on the running mode of the ALARO model coupled to the land surface model SURFEX, and driven by the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Re-Analysis (ERA-Interim) data."

Sentence starting with: "The results show that the introduction of SURFEX..." Could be revised to something like: "The results show that the SURFEX land surface scheme improves the simulation of 2 m temperature but has a negligible impact of the simulation skill of daily precipitation totals."

We thank the reviewer for the suggestion; the sentence has been reworded in the revised manuscript.
The new versions (Page 1 Lines 10-12):
" The results showed that the daily reinitialisation of the atmosphere improved the simulation of the 2 m temperature for all seasons. It revealed a neutral impact on the daily precipitation totals during winter, but the results were improved for the summer when the surface was kept continuous."

Introduction

The narrative introduces the reader to global climate modeling, numerical weather prediction, regional climate modeling, downscaling, limited area modeling. However there is insufficient detail on their differences particularly on the frequency that Bcs are updated, what is meant by 'continuous' or how the 'reinitialization' is done (at the boundaries or across the domain). This needs simplifying, and can perhaps be resolved by limiting to just a few terms that are explicitly relevant to the study.

We added more detail to the revised manuscript as suggested. The explanation on the initial and the lateral boundary conditions has been significantly improved. We added a figure/schematic diagram (Figure 2 of the revised manuscript) explaining the different downscaling approaches and their simulation periods, spin-up times, frequency of initialisation and update of lateral boundary conditions etc.

It is never defined explicitly here or in the methods what "frequent reinitializations" and "continuous simulations" means. I got the impression that climate simulations were run using an RCM where the BCs were updated once a month (in CON) or daily (in FS) when most state of the art RCMs would be updating the BCs on a more frequent basis.

In the revised manuscript this has been explained better now, in combination with a new figure (Figure 2 in the revised manuscript). The LBCs where updated every 6 hours, but the update frequency of the initial conditions was different. For CON there was only one single initialisation, but with monthly updates of the SSTs; for DRI and FS there was a daily reinitialisation. This has been rephrased and clarified in the revised manuscript.
The new version (Page 4 Lines 26-28) reads:
""The zonal and meridional wind components, atmospheric temperature, specific humidity, surface pressure and surface components were provided every 6 hrs as lateral boundary conditions and interpolated hourly. They were introduced as initial conditions accross the domain."
And Page 4-5 Lines 34-2:

" To evaluate the sensitivity of the model to the update frequency of the initial conditions, three types of downscaling approaches were conducted with ALARO version 0 coupled to SURFEX version 5."
And further on in the description of the downscaling approaches.

Page 2 Line 6: Please check for spelling errors!

All spelling errors have been corrected in the revised manuscript as requested.

Page 2 Line 14: "The model used in this study is the ALARO-0 model configuration of the ALADIN system." This won't mean much to those who have never used this model configuration. Perhaps this information is best in the methods where you describe the model and can then elaborate on the specific details of the ALARO-0 model configuration.

We thank the reviewer for this suggestion. This is moved to the Model section 2.1 and the model configuration has been described more in-depth to get a good idea on the specific details of the model.
The new version (Page 3 Lines 17-18) reads: "The regional climate model used in this study is the ALARO model version 0, a configuration of the Aire Limitée Adaptation Dynamique Développement International (ALADIN) model with improved physical parameterisations (Gerard et al., 2009)."

Page 2 Line 17: Interaction appears to be used twice here. Please correct.

Corrected in the revised manuscript as suggested.

Page 2 Line 19: Please provide the reference evaluating ALARO-0 with ISBA for continuous simulations.

The reference to Giot et al. (2016) had already been given in the original manuscript, but in the revised manuscript we moved the location of the reference so that it is more clear.
The new version (Page 4 Lines 3-4) reads:
"In addition, this setup has been validated for continuous climate simulations and is now contributing to the EURO-CORDEX project (Giot et al., 2016; Jacob et al., 2014)."

Page 2 Line 21: "has been implemented in the ALARO-0 version." Seems like the version number is missing at the end of the sentence.

The version number is zero; this has been corrected and explicitly stated in the revised manuscript in Section 2. In the Introduction, ALARO is mentioned without version number.
The new version (Page 3 Lines 17-18) reads:
"The regional climate model used in this study is the ALARO model version 0, a configuration of the Aire Limitée Adaptation Dynamique Développement International (ALADIN) model with improved physical parameterisations (Gerard et al., 2009)."

Page 2 Line 22: "the introduction of SURFEX with ALARO-0 has shown neutral to positive. . ." This phrasing is used a couple of times, perhaps its best to be more specific; which variables show no sensitivity, which ones are sensitive and what is the sign and magnitude?

In the revised manuscript we have included a more detailed description of the variables that show sensitivity as well as the sign, without mentioning the magnitude.
The new version (Page 4 Lines 6-9) reads:
"With respect to NWP applications, the introduction of SURFEXv5 within ALARO-0 has shown neutral effects on the winter 2 m temperature and on the vertical profile of the wind speed. However, it has shown positive effects on the summer 2 m temperature,

2 m relative humidity, and resulted in improved precipitation scores compared to the previously used ISBA model (Hamdi et al., 2014)."

Page 2 Line 24: "Therefore the evaluation of SURFEX within ALARO-0 is highly demanding for regional climate simulations" Hamdi et al. 2014 already evaluates SURFEX within ALARO-0 so perhaps the authors need to be specific here by saying that Hamdi et al. evaluate SURFEX within ALARO-0 for NWP but this manuscript will evaluate the same model environment for longer simulations.

In the revised manuscript we added a sentence as suggested.
The new version (Page 4 Lines 9-10):
"Next to the validation of this setup for NWP, the implementation of SURFEXv5 within ALARO-0 is highly demanding for long-term climate simulations."

Page 2 Line 27: "The second objective is to evaluate the continuous setup with an upper air daily reinitialized setup, where the surface is simulated continuously." I think this is where a lot of the confusion on terminology and model runtime configuration stems from. It would help if you can articulate what variables are continuous for each experiment, what variables are reinitialized and the frequency to which this is done. Could also add that information to Table 1.

The requested information has been added in Figure 2 of the revised manuscript. The initial and lateral boundary conditions for the atmosphere and surface (as stated earlier) are either initialised once at the beginning or reinitialised daily or reinitialised daily for the atmosphere and only initialised once at the beginning for the surface. For the continuous simulation, the SSTs are still reinitialised monthly.

Page 2 Line 29: "Therefore one expects to see improvements for the second method over continental parts of the domain, but not so much in coastal areas." Why not in the

coastal areas?

When applying a continuous simulation of the regional climate, we made a similar simulation as is commonly done by the climate community, using monthly SST updates. However, when we applied the daily reinitialisation of the atmosphere, the SSTs were updated daily as well. We expect a different behaviour of the model with different SSTs update frequency. As we did not test the sensitivity of the model to daily or monthly updates of SSTs, we are not able to make any statements on the reasons for different behaviour of the model over the coastal areas. This is outside the scope of our study. Therefore, we removed this statement.

Page 3 Line 2: Update to "Therefore the diurnal cycle of soil moisture was analyzed at particular locations in this study."

This analysis has been removed from the revised manuscript.

Page 3 Line 3: This paragraph doesn't quite fit in with the previous narrative; please provide an explanation on why the focus is on the summer season (i.e. when the soil moisture limitation on evapotranspiration is greatest) for those less familiar with the land-atmosphere coupling literature.

As requested we have provided more background information in the revised manuscript on why we have done this only for summer.
The new version (Page 2 Lines 31-33) reads:
"The surface interacts with the climate through the soil moisture and soil temperature, by influencing the surface energy budget (Giorgi and Mearns, 1999). The soil moisture controls the partitioning of the incoming energy into a latent and sensible heat flux. The soil moisture limitation on the evapotranspiration is largest during the summer (Seneviratne et al., 2010)."

Methods:

Are all simulations run with the same ERA-Interim BCs? It would be useful to add this information to Table 1 and manuscript text.

Yes, this information has been added to Figure 2 of the revised manuscript and in the revised text of the manuscript.
The new versions (Page 4 Lines 24-25) reads:
"The regional climate model was driven by initial and lateral boundary conditions provided by the ERA-Interim reanalysis, available at a horizontal resolution of ca. 79 km."

What data is used for the 'reinitializations' in FS?

The ERA-Interim reanalysis data have been used to provide the reinitial conditions for the zonal and meridional wind components, atmospheric temperature, specific humidity, surface pressure and surface components. This information has been added to the revised text (see earlier).

I found most of the narrative of Section 2.1 to go between describing the model/s and information that would be better placed in Section 2.2 Experimental Design – needs revising

This is a valid comment. The structure of Section 2 has now been revised and improved. The revised Section 2.1 is purely on the model descriptions, and revised Section 2.2 is on the experimental setup of the study.

Page 3 Line 14: These sentences would be more suitable in Section 2.2 Experimental Design

Thank you for this suggestion. The sentences have been moved to the revised Section on the Experimental Design.

Page 3 Line 17: Another model is introduced but not used in this study. Please remove to simplify the narrative.

We agree that this introduction does not make any sense. Accordingly we have removed this reference from the revised manuscript.

Page 3 Lines 14-22: There is no detail here on the microphysics, cumulus convection scheme or planetary boundary layer scheme. The reader is not provided with sufficient information on what the ALARO-0 configuration of the ALADIN modeling system actually means. This information is necessary in my opinion, particularly for someone not familiar with the model but interested in what was tested.

More information on the technical details of the model has been added to the revised manuscript as requested.
The new version (Page 3 Lines 24-27) reads:
"The new physical parameterisation within the ALARO-0 model was specifically designed to be run at convection-permitting scales, with a particular focus on an improved convection and cloud scheme, developed by Gerard and Geleyn (2005) and further improved by Gerard (2007) and Gerard et al. (2009). "

Page 3 Line 20: Please change "following-terrain" to "terrain-following"

This has been changed on Page 3 Line 21.

Page 3 Lines 23-29: Based on what is written here I got the impression that SURFEX was basically ISBA with tiling and coupled to TEB. If the intention is to conduct a *comparison between SURFEX and ISBA then the model descriptions need to be much more explicit on what the differences are between these models.*

The model descriptions have been made more explicit in the revised manuscript. We have clarified what the differences are about between SURFEX and ISBA. SURFEX is providing tiles for new surface types, uses ECOCLIMAP as input for the land cover and is externalised.
The new version (Page 4 Lines 16-21) reads:
"The initial parameterisation ISBA for the nature tile was conserved, and parameterisations for the other surface tiles were added, such as the Town Energy Balance scheme (TEB, Masson, 2000) for the town tile. TEB uses a canopy approach with three urban energy budgets for the layers roof, wall and road. The ISBA and TEB scheme were combined, together with parameterisation schemes for inland water and oceans, and externalised, based on the algorithm of Best et al. (2004). Each tile is divided in different patches, according to the tile type. These patches correspond to the plant functional types described in ECOCLIMAP (Masson et al., 2003)."

Page 3 Line 30: 'ECOCLIMAP' Please define all acronyms the first time they are introduced

ECOCLIMAP is not an acronym, but the name of a database.

Page 4 Lines 5-10: I gather that this is an explanation of the what variables are exchanged between the land surface model and the atmospheric model. The wording could be revised to make this easier to understand.

We have removed this from the revised manuscript, as it does not provide added value to the content of the manuscript. These details can be found in related literature.

Page 4 Lines 11-19: If the model is not run on the EURO-CORDEX domain then why provide detail on it? Also: "The present study was done in the framework of another project" please tell the reader what this project is.

This statement has been deleted from the revised version as we did not have any comparison with EURO-CORDEX output anymore.

In Table 1 a CRDX experiment is listed but not defined in Section 2.2; which one is it?

This experiment is no longer part of the manuscript anymore and has been removed.

Page 4 Line 22: "It started at 00UTC on 01 January 1990, and ran continuously until 00UTC on 1 January 2000. The first year was treated as a spin-up year, and the analysis period covers the 10-year period 01 January 1991 to 31 December 2000." Please revise the inconsistency here as two different end dates are mentioned.

Done, the references to dates and times have been made consistent in the revised manuscript.
The new version (Page 4 Lines 31-32) reads:
"The analysis covered a 10-year period from 00UTC on 01 January 1991 to 00UTC on 01 January 2001."
And (Page 5 Lines 4-5) reads:
"The model was simulated from 00UTC on 01 January 1990, and ran continuously until 00UTC on 01 January 2001."

Why are simulations run for 1991-2000 when a more recent period would enable comparison to more FLUXNET sites?

This time period was selected because a similar experiment was done with forcing from the global climate model ARPEGE CMIP5 version. This forcing covers the period of 1976-2005 and we selected an overlapping period between the two simulations.

Why does CON have a spin-up of 1 year starting in January 1990 at 00UTC and why does FS have a 3 month spin-up with simulations starting in March each year at 12UTC? Usually the spin-up and start date should be the same unless the focus of the study is on how long a spin-up is necessary to maximize simulation skill which is not the aim of this paper!

The 1-year spin-up time is commonly done in climate simulations in the continuous mode. Considering the FS simulations, we showed that the ideal setup of the initialisation period is three months starting in March. This relatively short spin-up period is enough to reach a equilibrium state (shown earlier in this Reply at point (3)). If we had selected a start in January, the model would have needed a longer spin-up time and the computing time of the simulations would significantly increase. This has been rephrased and clarified in the revised manuscript.
The new version (Page 5 Lines 21-22):
"Although CON required one year spin-up time, 3 months were sufficient for the FS deep soil moisture to reach equilibrium state, when starting in March (not shown)."

Why are atmospheric variables saved at a 3 hourly interval and land surface variables at daily? It would make sense to save them all at the same interval.

ALARO and SURFEX are two different models, with different output files. The surface output files are much larger than the atmospheric output files. Therefore, we had to limit the surface output files because of storage limitations at our institute. This has been rephrased in the revised manuscript.
The new version (Page 5 Line 25) reads:
" The model output at every 3 hrs was used for the model evaluation. "
And (Page 5 Lines 32-34) reads:

"As land-surface processes play an important role primarily during summer, the model output was stored at every hour for the summer period of June-July-August (JJA) during the 10-year period."

Page 4 Line 33: "Each daily simulation extended up to 60 hours of which the first 36 hours were treated as spin-up" This seems to be contradicting the previous explanation where there is a 3 month spin-up for 1 year simulations! Please revise the description of how each experiment was run as it is not clear at the moment and limits the reader's ability to reproduce the experiment.

The original wording was indeed confusing. Therefore, we used the wording "atmospheric spin-up" in the revised manuscript, which is typically in the order of 24 hours, and the "surface spin-up", which is typically in the order of a few months to one year.
The new version (Page 4 Lines 29-31) reads:
"For the sake of a good understanding, the following description makes a distinction between atmospheric spin-up time, typically of a few days, and surface spin-up time, typically of a few months to one year."

Page 5 Line 3: "A third simulation was applied for both CON and FS" Technically third and fourth simulations?

This has been removed, as these are technically not different simulations. For the three downscaling approaches, the variables were stored hourly in addition for the summer.
And (Page 5 Lines 32-34) reads:
"As land-surface processes play an important role primarily during summer, the model output was stored at every hour for the summer period of June-July-August (JJA) during the 10-year period."

Page 5 the extra CON and FS simulations saving hourly output, why wasn't this done *from the start rather than doing additional shorter simulations*

Initially we only intended to investigate atmospheric parameters. We noticed the large changes in summer for 2 m temperature between the different simulations and decided to further investigate surface fluxes. Therefore, we needed hourly output to look at the diurnal cycle. The explanation of the output has been reworded accordingly in the revised text (see earlier).

Page 5 Line 6: "The first simulation. . ." Is this referring to the CRDX experiment?

This sentence has been removed from the revised manuscript.

Page 5 Line 13: This is the first time PRUDENCE is mentioned – define acronym

Ok, we have defined the project acronym in the revised manuscript as requested.
The new version (Page 5 Lines 28-29) reads:
"project "Prediction of Regional scenarios and Uncertainties for Defining European Climate change risks and Effects" (PRUDENCE) (Christensen et al., 2007)"

Page 5 Line 16: "relaxation zone was excluded" It looks like Figures 3, 5 and 7 need yo be cropped to exclude the relaxation zone.

The relaxation zone has been excluded from the plots in the revised manuscript. See Figure 3, Figure 4 and Figure 6 in the revised manuscript.

Observational data:

Perhaps better to combine the model description, experimental design and observational data under one Methods, Models and Datasets section

Ok, we have followed the reviewer's suggestion and we have combined Model

description, Experimental design and observational data in the revised manuscript.

Page 5 Line 21: replace 'sum' with 'total'

Done; has been replaced in the revised manuscript.
The new version (Page 6 Line 8) reads:
"daily precipitation total."

Page 5 Line 26: Usually one interpolates to the coarser resolution. i.e. interpolate the model data to the E-OBS resolution.

In line with the reviewer's suggestion we regridded the model to the E-OBS resolution and we recalculated the differences at the coarser resolution. We have clarified this in the revised manuscript.
The new version (Page 6 Lines 12-14) reads:
"In order to validate the model data, the ALARO-0 data at 20 km horizontal resolution were bilinearly interpolated towards E-OBS at 25 km horizontal resolution and replotted to our study domain."

Page 5 Line 27: replace 'implied' with 'applied'

Done, this has been replaced in the revised manuscript.
The new version (Page 6 Lines 15) reads:
"applied"

Page 5 Line 30: replace 'exchanges' with 'exchange'

Done; has been replace in the revised manuscript.
The new version (Page 6 Lines 18) reads:
"exchange"

Page 6 Line 1: remove the sentence starting "The technique. . ."

Done; the sentence has been removed from the revised manuscript as requested.

Page 6 Line 4: replace 'with regard to the' with 'against these'

This sentence is rephrased.
The new version (Page 6 Lines 20-21) reads:
"No gap-filling has been done and the comparison to the model output was only done at hours when no gaps occurred."

Page 6 Line 5: revise sentence starting "The model resolution is quite low..."

The sentence has been removed in the revised manuscript, as it is straightforward that model resolution is lower than station resolution.

In Section 5.3 a justification is provided on why the two FLUXNET sites were selected. It would be more appropriate to put that in Section 3.2.

This has been clarified in Section 2.3 of the revised manuscript.
The new version (Page 6 Lines 23-25) reads:
"Two FLUXNET stations were selected, that provided data already during this period and where the model grid cell represented more than 50% of the corresponding land cover, to show energy fluxes that were representative for the particular land cover."

Results – more general rather than line by line:

The values in Table 2 and 3 are referred to more often than Figures 2 and 3. There

are also several instances where it is not clear which results are being referenced. In particular, reporting the percentage change between experiments was quite confusing given that these values are not presented in either the Tables or Figures. This meant that I had to spend a lot of time checking where the values were coming from, or calculating the percentage changes myself. This could be resolved by adding detail in the manuscript text on the values shown in the figures. If the percentage change between experiments is quoted then please include these values in the Table or replace with something along the lines of: "CON has a larger bias of X relative to FS which has a bias of Y". This will make it easier for a reader to match the narrative to what is presented in Figures and Tables.

We thank the reviewer for these suggestions. The text has been rephrased and improved. In the revised manuscript we refer both to the Figures and Tables. The percentage change between simulations (%) were indeed very confusing; consequently we have removed them from the revised manuscript. Only the exact numbers are used in the revised manuscript, as they are represented in the Tables.

There is a tendency to use words such as 'large', 'slight', 'excessive' or 'improved' in the narrative. Please be specific and insert detail. For example, "the bias improves" could be replaced with "the temperature bias decreases by X in experiment Y". There are instances where this is done, but it has not been applied consistently. However doing so will make it easier for the reader to understand the results that are presented.

Thank you for this suggestion. In the revised manuscript we have avoided the use of these words, and we predominantly use the exact numbers.
One example (Page 7 Lines 14-15) reads:
"This resulted in a smaller bias for Eastern Europe of -0.3 C and 0.0 C for DRI relative to CON which had a bias of -1.1 C and -0.5 C for winter and summer respectively."

Why are the temperature biases presented in degC but the precipitation biases presented as the percentage change? It would be preferable to use one approach consistently throughout the manuscript. In particular for the precipitation results, the % bias is often large when the observed precipitation is small – this is perhaps an instance where using the bias (MODEL minus OBS) would be more useful where small precipitation values inflate the value of the % bias.

This approach is generally used in climate model validations, because this domain covers different climatic regimes. By doing so, we can better compare regions. This has not explicitly been clarified in the manuscript.

Page 7 last sentence: I don't agree with this, why would SSTs only be influential in winter?

We have removed statements like this, as discussed earlier. As we did not test the sensitivity of the model simulations to the different SSTs update frequency, we are not able to relate directly the differences in winter precipitation to differences in SSTs update frequency. This is outside the scope of this study, and thus has not been explicitly clarified in the revised manuscript.

There is a tendency to start sentences with "Similarly to Experiment X, Experiment Y . . ." please use either: "Similar to Experiment X, Experiment Y. . ." or just start with "Experiment Y . . ."

Done; this wording has been improved in the revised manuscript.
One example (Page 9 Line 20) reads:
" Similar to Mid-Europe, ..."

Sentences such as: "FS (CON) overestimated (underestimated) the summer 2 m tem-

perature" are really hard to read and understand. It is actually easier to read: "FS overestimated and CON underestimated the summer 2 m temperature". Please revise all instances of this.

Done; this has been revised at all instances in the revised manuscript.
The new version (Page 9 Lines 15-16) reads:
"However, FS overestimated the summer 2 m temperature and CON and DRI underestimated the summer 2 m temperature."

Section 5.1 and 5.2:

Replace: "We assume two-way interactions" with "There are two-way interactions"

This sentence and the related analysis have been removed from the revised manuscript.

At no point do we know the depth of the soil layers in the land surface scheme. This would be useful to know.

Yes indeed, this will be used for a separate, future study where we will analyse the correlations in more detail. We have removed this statement from the revised manuscript.

It is not clear to me why the land-atmosphere feedback is evaluated by calculating the soil moisture temperature (ST-T) correlation using (FS minus CON / CON). It would be better to calculate the SM-T correlation for FS, the SM-T correlation for CON and then the difference between these two estimates. I think this analysis needs to be redone. Lorenz et al. (2015) provides a good description of different coupling metrics that could be applied to the data.

The soil-moisture-temperature correlation will be content for a separate, future study,

as the current study primarily focuses on the simulation potential of different downscaling experiments instead of looking at correlations between variables. This has been clarified in the revised manuscript.
The new version (Page 12 Lines 25-27) reads:
"A more in-depth analysis on the interaction between 2 m temperature, precipitation, and surface energy fluxes can reveal soil-moisture-temperature coupling (Jaeger et al., 2009), but this lies outside the scope of this study."

Here I also think that the spin-up length will have some influence on evaluating the land-atmosphere feedback. If CON has a longer spin-up than FS of the soil moisture and soil temperature then the results are surely already biased as the land surface state fields will be more resolved in the simulations with the longer spin-up.

We have shown that three months of spin-up is reasonable for a good equilibrium. Accordingly we concluded that the different spin-up periods did not influence the land-atmosphere feedback. However, the land-atmosphere feedback has no longer been evaluated in the revised manuscript.

Conclusions:

Page 11 Line 31: This definition of 'continuous' and 'reinitialized' simulation should be defined much earlier in the manuscript!

Done. The definition has been provided much earlier in the manuscript.

Page 11 Line 7: "The differences in 2 m temperature and precipitation between the downscaling setups during summer are demonstrated by an interaction with the soil moisture." I don't agree with this because it's not actually calculated for the 10 year simulations where the temperature and precipitation differences are evaluated.

We agree with the reviewer's comment. As the surface fluxes have been calculated for 10 years in the revised manuscript, we can make more reliable statements on this. This has not explicitly been clarified in the Conclusions section.

Tables and Figures

Table 1 – perhaps remove the CON-FS line or separate into two.

Table 1 has been removed from the revised manuscript.

Table 3 – it would be easier if these values were presented as mm day-1 rather than the relative bias because some values are very high but might only be so because these are regions where the precipitation is very low: e.g. MD regions DJF FS experiment.

As a matter of fact the difference in MD for DJF is high. We have decided to maintain this relative bias in the revised manuscript as explained earlier. See Figure 3 in the revised manuscript.

Figure 2 – it would be good to include statistical significance as done in Figure 7; perhaps update the labels in the top left hand corner of panels c to h with CRDX –E-OBS

Done; the statistical significance has been included in the revised manuscript. See Figure 3 and 4 in the revised manuscript.

Figure 3 – please put all panels in the same units it makes it easier to compare panels c to h with a and b. Is there missing data over Africa where there is a weird white triangle shape at the bottom of all panels? It is obvious in this figure that there are boundary affects for the domain and that the figures have not been cropped to exclude the relaxation zone.

Different units are used for the absolute precipitation presented by E-OBS and the precipitatiob bias of the model with E-OBS. No precipitation has been measured over this triangle in Africa. This also appears in the EURO-CORDEX output. The relaxation zone has now been excluded from the revised manuscript as rightly requested by the reviewer. See Figure 3, 4 and 6 in the revised manuscript.

Figure 4 – if the authors choose to keep the CRDX simulations then they should be included here and other figures. It would also be handy if dashed lines were added to each panel to delineate the seasonal breaks referred to in the manuscript text.

Figure 4 has been removed from the revised manuscript.

Figure 5 – obvious boundary affect in all panels. Why is the absolute difference used for temperature but the relative difference used for precipitation and soil moisture? It would be easier if they all presented in the same way.

Figure 5 has been removed from the revised manuscript.

Figure 7 – This needs to be redone to show SM-T and SM-P correlations for each experiment separately and then their difference.

The suggested analysis will be done in a separate, future study.

Figure 8 – It would be easier if FLUXNET, CON and FS were all on the same panel to directly compare differences. QS looks very different between the observations and CON and FS, are the authors certain that they are comparing like for like? QS looks like a flat line in panel c – check if there is a plotting error but this may just be because QS is very small relative to the axis scale or that it is not measured. . .

This is a good suggestion. In the revised manuscript separate panels are shown for

the different variables. See Figure 7 in the revised manuscript.

3 Manuscript version with highlighted changes is supplemented.