

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

**Anonymous Referee #1**

Received and published: 22 August 2016

Referee Comment on “Reinitialised versus continuous regional climate simulations using ALARO-0 coupled to the land surface model SURFEX” by Berckmans et al. Geosci. Model Dev. Discuss. doi:10.5194/gmd-2016-154

This study investigates the impact of different modeling techniques within the ALARO-0 Numerical Weather Prediction (NWP) model. These techniques differ in terms of whether the atmospheric boundary conditions (BCs) and sea surface temperatures (SSTs) are not updated during the simulation (continuous) or updated on a daily interval (reinitialized). I do believe that it is good to document model updates, particularly those concerning the best practice for initialization and running of forecast models. Therefore, I do feel that this manuscript fits within the scope of GMD.

C1

There are however some major concerns over the experimental design and the explanation of the results that require significant amendment before this manuscript is suitable for publication.

(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.

(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.

(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

C2

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.

(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.”

(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.

(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 dif-

C3

ferent runtime modes.

(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.

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.

- Please define all acronyms.

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

- 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.”

- 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.”

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 re-

C4

solved by limiting to just a few terms that are explicitly relevant to the study.

- 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.

- Page 2 Line 6: Please check for spelling errors!

- 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.

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

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

- 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.

- 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?

- 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.

C5

- 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.

- 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?

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

- 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.

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.

- What data is used for the ‘reinitializations’ in FS?

- 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

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

- Page 3 Line 17: Another model is introduced but not used in this study. Please

C6

remove to simplify the narrative.

- 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.
- Page 3 Line 20: Please change “following-terrain” to “terrain-following”
- 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.
- Page 3 Line 30: ‘ECOCLIMAP’ Please define all acronyms the first time they are introduced
- 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.
- 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.
- In Table 1 a CRDX experiment is listed but not defined in Section 2.2; which one is it?
- 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.
- Why are simulations run for 1991-2000 when a more recent period would enable

C7

comparison to more FLUXNET sites?

- 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!
- 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.
- 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.
- Page 5 Line 3: “A third simulation was applied for both CON and FS” Technically third and fourth simulations?
- 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?
- Page 5 Line 6: “The first simulation. . .” Is this referring to the CRDX experiment?
- Page 5 Line 13: This is the first time PRUDENCE is mentioned – define acronym
- Page 5 Line 16: “relaxation zone was excluded” It looks like Figures 3, 5 and 7 need to be cropped to exclude the relaxation zone.

Observational data:

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

C8

- Page 5 Line 21: replace 'sum' with 'total'
- Page 5 Line 26: Usually one interpolates to the coarser resolution. i.e. interpolate the model data to the E-OBS resolution.
- Page 5 Line 27: replace 'implied' with 'applied'
- Page 5 Line 30: replace 'exchanges' with 'exchange'
- Page 6 Line 1: remove the sentence starting "The technique..."
- Page 6 Line 4: replace 'with regard to the' with 'against these'
- Page 6 Line 5: revise sentence starting "The model resolution is quite low..."
- 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.

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.
- 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

C9

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.

- 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.
- Page 7 last sentence: I don't agree with this, why would SSTs only be influential in winter?
- 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..."
- Sentences such as: "FS (CON) overestimated (underestimated) the summer 2 m temperature" 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.

Section 5.1 and 5.2:

- Replace: "We assume two-way interactions" with "There are two-way interactions"
- At no point do we know the depth of the soil layers in the land surface scheme. This would be useful to know.
- 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

C10

be applied to the data.

- 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.

Conclusions:

- Page 11 Line 31: This definition of 'continuous' and 'reinitialized' simulation should be defined 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.

Tables and Figures

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

- Table 3 – it would be easier if these values were presented as mm day<sup>-1</sup> 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.

- 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

- 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

C11

exclude the relaxation zone.

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

- 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. . .

References:

- Lorenz et al. 2015: Intraseasonal versus interannual measures of land-atmosphere coupling strength in a global climate model: GLACE-1 versus GLACE-CMIP5 experiments in ACCESS1.3b. *J. Hydrometeor.*, 16, 2276-2295.

- Miralles et al. 2011: Global land-surface evaporation estimated from satellite-based observations. *Hydrol. Earth Syst. Sci.*, 15, 453–469.

- Mueller et al. 2013: Benchmark products for land evapotranspiration: LandFlux-EVAL multi-dataset synthesis. *Hydrol. Earth Syst. Sci.*, 17, 3707-3720.

---

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-154, 2016.

C12