Response to reviewer # 1

This study presents the results from a coupled regional climate model HIRHAM with MIKE SHE hydrological model, which also includes SWET land surface model. The coupling between the models are only done over the Skjern catchment , which is an interesting feature of this new tool. With this new tool, a series of real data numerical experiments with coupled and uncoupled models are presented to explore the influence of coupling frequency and internal variability of the atmospheric model. The study is quite interesting, and the multiple simulations along with comparisons to observations are comprehensive. However, at the same time, there are several shortcomings in the current version of the paper also, which needs to be addressed before being suitable for publication.

In general, the manuscript is poorly written with lot of grammatical mistakes and not well organized.

We will have an English native speaker to help improving the English language and the organization of the manuscript.

The authors conclude that the coupled simulations give poor results because the coupled model is not tuned or calibrated. However, there are no results presented in the current manuscript that supports their conclusion.

I understand the comment as a call for; 1) proof that the coupled setup provides poorer results than the uncoupled HIRHAM <u>due to the lack of combined calibration</u> and *not* 2) a call for proof that the coupled setup is actually poorer than the uncoupled and the reply here is written accordingly. Also because the proof for the latter (2) is evident in figures 3, 5 and 8.

The question of the need for a coupled calibration is centred on either providing the right answer for the wrong reason or providing the wrong answer for the right reason. As stated, both models are refined, tuned or calibrated (phrasing depends on community; climate/hydrology) to reproduce observations and any change in forcing data (surface scheme and climate input data), constituting significant elements of each model, are likely to worsen the results over the area in question.

One way of indirectly supporting this statement would be to actually show these forcings to have a significant impact on the model outcome. In the present paper this is especially seen in figure 5 where the RMSE values are both higher and include more variability for coupled runs. Another figure showing the influence of model forcing is seen in Butts et al. (submitted) where distributed evapotranspiration output for a one-week period is seen for 1) MIKE SHE forced with observations, 2) MIKE SHE forced with HIRHAM input (one-way with no feedback), 3) MIKE SHE forced with HIRHAM input (two-way including feedback) and 4) for HIRHAM alone. From this figure the influence of the coupling is evident as i), MIKE SHE produces higher evapotranspiration with observation input as compared to using HIRHAM input and, ii) the feedback between models significantly influences the coupled setup outputs as the two-way coupled evapotranspiration is higher than one-way coupled (no feedback from MIKE SHE).

We are a bit cautious on expanding the overall volume of the paper, but in the revised version we will elaborate on the issue of dynamics and coupled calibration.

A way of directly supporting the statement of the need for coupled calibration would be to include just that; coupled calibration. However, this is beyond the reach of the present study. Also, the question of coupled calibration is used in the present paper as a general comment for the coupling of systems calibrated and tuned individually to provide physical sense on their own in terms of energy- and water balance closure (providing an answer – right or wrong – for the right reason). We could add a short reference to coupled ocean-atmosphere studies where the experience is very similar.

It could also be potentially influenced by their limited area coupling. In their approach, they couple the two models over a small catchment only, which is less than 0.1% of the total atmospheric domain. So, does it create a very different patch of land surface over Skjern catchment compared to the adjacent cells, where HIRHAM uses its own land surface model? If there are significant differences in soil temperature due to difference in partitioning of surface energy fluxes by the new model, it can generate local circulations, which can influence the simulated variables. This needs to be discussed.

In spite of the coupling only covering 0.02% of the RCM domain our results show that it nonetheless has significant impacts on the land surface variables. This can e.g. be seen from figure 5, where the cells outside the coupled domain (domain) show much less difference between the RCM (TUV) and the coupled model (CV) as compared to the cells within the Skjern catchment (e.g. domain 1)" This is an interesting result in itself and documents that the local partitioning of surface energy fluxes within the Skjern catchment has a significant impact on the land surface variables and that those are not only determined by boundary effects (advection). We will include a discussion on the issue of local circulations deriving from the now area-limited forcing of the coupled area.

1. Pg. 3007, Line 10: The authors here can definitely not say that it is due to the calibration of the models alone. There could be many other reasons. This needs to be addressed. Again, we are cautious on expanding the paper volume but we would be happy to rephrase here into something along the lines of: "It is discussed whether this may be caused by the lack of calibration of the coupled model" in both abstract and discussion.

2. Pg. 3007, Line 18: Change "ranges" to "spread". Ok.

3. Pg. 3009: The introduction is poorly written, the authors discuss about future global climate predictions and importance of uncertainty in climate models, and then at the end, present what they do in this study. A more focused discussion on the scientific questions they want to answer with this new tool and the motivation behind this work, would strengthen this section. We agree that this section sets of very broadly, something which is done to frame the study into the context of highlighting the need for further knowledge and investigation of the atmosphere – land surface process interrelations. We however agree to shorten the broader perspectives and more clearly emphasize the specific motivation for the study.

4. Pg.3009, Line 6-16: These works were mostly related to short term simulations using mesoscale atmospheric models, not climate simulations. The spatio-temporal time scales of these studies compared to the preceding introduction are different. This needs to be rephrased. We agree and will rephrase.

5. Pg. 3009, Line 26-29: For example, read York et al. 2002, Jiang et al .(2009), Anyah et al. (2008). C926

These are some very interesting papers. Thank you. Since York et al. (2002) uses a single atmospheric model grid, Jiang et al. (2009) uses a simple non-3D SIMGM groundwater model and Anyah et al. (2008) (as described in Miguez-Macho et al. 2007) uses Darcy's law for the horizontal groundwater fluxes we believe the statement still holds true whereas we would be happy to modify from:

"To our knowledge, no studies have been reported on long term simulations (more than a few days) with couplings between a regional climate model and a 3-D groundwater–surface water hydrological model using catchments larger than a single regional climate model grid point."

To:

"To our knowledge, no studies have been reported on long term simulations (more than a few days) with couplings between a distributed regional climate model and a full 3-D groundwater–surface water hydrological model using catchments larger than a single regional climate model grid point."

Also, we will add these suggested papers in the introduction literature review.

6. Pg. 3011, Line 19: Clarify "the undercatch corrected precipitation". Good suggestion.

7. Pg. 3012, Line 9: Does it mean that the fluxes measured over forested area was used for agricultural site? Explain the rationale behind it. Missing data alone does not justify this approach. This approach is well documented and approved in peer reviewed papers for these specific data (Ringgaard et al 2011). We however clearly acknowledge the pitfalls of this approach and will make a short comment on this in the paper.

8. Pg. 3013, Line 20: Is this calibration done for every particular year? The calibration is described in another paper (submitted – as described in the reference) where sensitivity analysis and inverse modelling is applied for a one-year period with subsequent validation for another period.

9. Pg. 3015, Line 11: Clarify "safety regulations". So, the coupling is based on reading and writing of output files? Yes.

10. Pg. 3015-3017: The description of the different simulations requires a Table with two sub-categories: "coupled and uncoupled simulations" followed by experiment name and description. Otherwise, it becomes too difficult for the readers to follow which experiment is which, and it is very annoying. The evaluations are performed in terms of RMSE and MAE. The results could be interpreted better by using MSE and examining the contribution of model bias, variance of the simulated variables and the correlations. And, the use of Taylor diagrams would be even more appropriate to present the results for comparison of different variables with multiple simulations, coupled or uncoupled. Adding a table for a general simulation overview is a good idea that we could

try to fit in without much added volume. Regarding performance measures we have used MAE as this is more intuitive to understand the actual absolute differences and the choice between RMSE and MSE is simply a matter of reducing the plotted numbers for a good and balanced overview. Again, due to volume concerns we are hesitant to add Taylor diagrams as an additional figure as we believe the present figures still provide important information used in the discussion regarding model performance, data transfer frequency and variability.

11. Pg. 3020 - 3021: See above comments. Answered above.

12. Pg. 3023 and Pg. 3030, Line 10: This needs to be rephrased. See above comments. We can broaden this statement based on our answer above.

13. All figures have a very small font size which is not readable. Good observation – we will revise.

References

- Butts, M., Drews, M., Larsen, M.A.D., Lerer, S., Rasmussen, S.H., Groos, J., Overgaard, J., Refsgaard, J.C., Christensen, O.B. and Christensen, J.H. (submitted): Embedding complex hydrology in the regional climate system – dynamic coupling across different modelling domains.
- Larsen, M. A. D., P. Thejll, J. H. Christensen, J. C. Refsgaard, and K. H. Jensen (2013a) On the role of domain size and resolution in the simulations with the HIRHAM region climate model, Clim. Dynam., 40, 2903–2918, doi:10.1007/s00382-012-1513-y.
- Maxwell, R.M., Lundquist, J.K., Mirocha, J.D., Smith, S.G., Woordward, C.S. and Tompson, A.F.B. (2011) Development of a Coupled Groundwater–Atmosphere Model. Mon Weather Rev, Vol 139, 96-116, doi:10.1175/2010MWR3392.1.