List of all relevant changes made in the manuscript and point-by-point response to the reviews (open discussion)

Title: Modelling of snow processes in catchment hydrology by means of downscaled WRF meteorological data fields (discussion paper)
Author(s): K. Förster et al.
MS No.: hess-2014-76
MS Type: Research Article
Iteration: Major Revision

We would like to thank three anonymous referees and Dr. Bettina Schaefli for their effort to review and edit our manuscript. We believe that the comments and suggestions helped us to improve our manuscript in several ways. The following describes all major changes in the manuscript and includes the answers to the reviewer comments in the open discussion. Please note that in order to best possibly address the reviewer's comments, we have restructured the manuscript. Some detailed replies provided online in the open discussion might therefore refer to issues that are not part of the updated version of the paper anymore. This document includes the reviewer's comments (blue), our detailed answers in the open discussion (black) and the respective implications for the manuscript (red).

Based on the comments of the reviewers, we have intensively discussed the revision of our manuscript. Since the main criticism pertains to the combined use meteorological observations and simulations in one model run (which we agree does not really allow to assess model performance using WRF data), we have decided to only compare model runs driven by downscaled WRF data to model runs driven by meteorological observations in the new version of the manuscript. Following the reviewer's suggestions, the revised version of our manuscript provides more detailed information with respect to model calibration including a brief description of model parameters. To best possibly address all reviewer's comments, most parts of the manuscript have been updated including title, abstract, introduction, results, and the summary section. Please find below a list of changes that have been made to the discussion paper.

Major changes

- The aim of our study is now stated more precisely and addresses the following research questions:
    1. To what extend do downscaled meteorological simulations enhance model performance?
    2. Does increasing snow model complexity using these data increase model performance?
- Following anonymous referee #2's suggestion, we have included the study site in the title and have reformulated the title to better reflect the content of our study: New *title*: "Effect of meteorological forcing and snow model complexity on hydrological simulations in the Sieber catchment (Harz Mountains, Germany)"
- *Abstract*: The abstract was updated in accordance to the edited content of the manuscript.

- *Introduction*: The introduction was also updated with respect to the reformulated research question. Most parts of the literature review were not changed.
- *Data and methods*:
    - *Study area*: minor changes according to the reviewer's comments.
    - *Selected winter seasons*: The problem of computation time is discussed in order to better explain the limited number of considered winter seasons. Moreover, we have updated the section *Dynamical downscaling using WRF* in order to provide more information about the computation time required for one single WRF run in the current model setup – the high computational demands are the major cause for the selection of only two winter seasons.
    - *Snow models*: Reviewer #3 suggests adding specific information about the snow models (e.g., separation of rain and snow). Therefore, our revisions include information about adjustable parameters of each model. We briefly outline how rain and snow are separated.
    - *Hydrological modelling*: As suggested by referee #1 and #3, we have extended the model description with respect to model calibration. The calibration part was moved to a separate subsection (*2.6 Model calibration*).
    - *Model calibration*: We included this additional section with a revised description of model calibration following the recommendations of referee #1 and #3. The calibration procedure with respect to the Sieber catchment using PANTA RHEI is now explained in much more detail including the number of altered parameters and performance measures. With reference to the brief description of model parameters in the *Snow models* section, we now provide the parameter set for each snow model. Performance measures are also given (including Nash Sutcliffe as suggested by referee #1).
- *Results and discussion*: As recommended by the reviewers, the revised manuscript should clearly distinguish between model runs driven by meteorological observations and WRF-driven model runs and avoid the combined use of meteorological observation and simulations (reviewer #2, #1). We have followed these suggestions with the respective impacts on the structure of the *Results and discussion* section. This approach also corresponds to the "comparative study" suggested by reviewer #2.
    - *Snowmelt simulations using observed meteorological data*: In this section, we show results that were obtained using the temperature-index model with observed meteorological time series. Results are displayed for the point as well as the catchment scale.
    - *Snowmelt simulations using downscaled WRF data*: This section includes an evaluation of downscaled WRF time series. This comparison of downscaled and observed meteorological time series corresponds to the *Meteorological fields* section of the discussion paper. In the revised version of the manuscript more specific information is given for precipitation simulations. In addition, we briefly explain the model performance with respect to humidity, wind speed, and radiation as has been suggested by reviewer #3. Since all relevant meteorological variables are available from WRF, this model configuration is not restricted to an application of the temperature index approach but enables the application of energy balance methods. In

contrast to the discussion paper all results are based on downscaled meteorological variables only.
    - As the comparison of different sublimation simulations could be misinterpreted as part of the validation (see comments from reviewer #1 and #3), we have removed the respective paragraphs from the manuscript.
- The *Summary and conclusions* section now reflects all above-mentioned revisions.

- Tables:
    - Table 2 added: This table provides a summary of the calibration procedure of the snow models including the parameters that have been altered and their respective values (as recommended by all reviewers).
    - Table 3 added: This table includes the results of the calibration procedure. Nash-Sutcliffe model efficiencies for all snowmelt simulations for the point and the catchment scale are provided (as recommended by all reviewers).
    - Table 4 added: As suggested by reviewer #1, Nash-Sutcliffe and other criteria for model performance with respect to downscaled precipitation are now given for each station.
    - Table 5 added: The model performance for all other relevant variables is now included, a modification suggested by reviewer #3.
- Figures:
    - *Map of the study area* (Fig. 1): The station network was added indicating the observed variables for each station (as suggested by reviewers #1 and #2, respectively).
    - Figure added (Fig. 4 of the new manuscript): SWE and snow depth are illustrated for the point-scale simulations using observed meteorological time series as model input (as suggested by all reviewers).
    - Figure added (Fig. 5 of the new manuscript): Melt runoff is illustrated for the point-scale simulations using observed meteorological time series as model input ("comparative study" as proposed by reviewer #2).
    - Figure added (Fig. 6 of the new manuscript): Streamflow simulations for the catchment scale are illustrated using observed meteorological time series as model input ("comparative study" as proposed by reviewer #2).
    - Figure added (Fig. 9 of the new manuscript): Comparison of modelled SWE and observed snow depth at the point scale using downscaled meteorological time series as model input (as suggested by all reviewers).
    - Revision of melt runoff and streamflow simulation plots (Fig. 10-13 of the new manuscript). Now the four plots include the results of all snow models at both scales consistently showing results based on downscaled precipitation only.
    - Figure 10 of the discussion paper removed: The Taylor plots were removed in order to keep the manuscript more concise. Nash-Sutcliffe model efficiency values are provided instead (Fig. 10-13), as suggested by reviewer #1.
    - Figure 11 of the discussion paper removed: As described, the model comparison with respect to sublimation is not considered anymore (see changes to the *Results and discussion* section).

Reply to Anonymous Referee #1

We thank anonymous reviewer #1 for her/his detailed comments on the manuscript. The comments really helped us to improve the manuscript.

General comments:

"Different snow models are driven by a set of parameters delivered by the WRF model. The investigation area is located in the Hartz Mountains (Germany). The general aim of the paper, to introduce a model chain which is independent of surface measurements, is an important research topic. The authors clearly describe the deficits of sometimes sparsely distributed point measurements with respect to spatially distributed models.
Hence, they are using WRF fields for driving land surface models. This would be a favorable approach but because of deficits within the precipitation fields they are substituted by measurements in the course of the paper. This definitely limits the significance of the whole paper and stands in contradiction to title, abstract and introduction. Beside of this the calibration strategy remains unclear and is sometimes inexplicable. It is said that the snow models are calibrated by using WRF fields instead of measurements but precipitation measurements are used for driving the model afterwards. This is hard to understand. Finally, the availability of validation data is extremely limited and the data seems to be inappropriate for checking the quality of different key results of the snow models. If these deficits can be eliminated the paper could be worth for publication."

Reply in the open discussion:
We agree that the substitution of simulated precipitation by observations limits the application of the approach that has been presented. Our intention was to present that the meteorological fields, which were derived using downscaled WRF analysis data, are generally suitable datasets to drive snowmelt models. It was possible to derive reliable values of precipitation depth at the seasonal scale. However, at smaller time scales simulated precipitation intensities are less reliable. This finding holds especially for the period 25 Mar 2006 until 09 Apr 2006, during which a rain on snow event occurred (Fig. 4). Therefore, we decided to substitute simulated precipitation with observations. In this case, the simulations of runoff at both scales coincide well when compared to observations. Hence, it could be shown that the other variables, which are also relevant to snowmelt, are suitable to drive snowmelt models at both scales.
Alternatively, we will show that using simulated precipitation does not yield good performance, by showing results based on simulated precipitation. This approach proves that 1) the precipitation simulations are not as accurate as desired and 2) the other variables including temperature, humidity, wind speed, shortwave radiation, and longwave radiation are very useful in order to serve as input for snowmelt models.
Using observed precipitation could be used to improve the representation of precipitation.
These findings restrict the applicability of this approach. However, in many cases all these variables are less widely available than precipitation. Hence, a substitute might be relevant in regions where only precipitation is available and some of the other variables are missing. This is a common situation. As a prospect for future research, a combined product of precipitation observed at stations, and simulated precipitation fields could also be useful to improve the spatial and temporal representation of precipitation at the catchment scale.

We regret that the explanations regarding to the calibration of the models have led to misunderstandings. As we will explain in the specific comments sections, we will clarify the text with respect to the calibration procedure.
Unfortunately, snow water equivalent observations are not available at the Torfhaus meteorological station. We will provide a figure including both observed snow depth and snow water equivalent simulations in order to improve the validation of the models.

Implications for the manuscript:
Following the reviewers suggestions we no longer use a combination of observed precipitation and other simulated meteorological variables as input for the snow models in the updated version of the manuscript. Instead, we separately evaluate and discuss the snow model results achievable with observed and simulated meteorological input to fully assess the value of simulated meteorological forcing as a substitute for observations. We also follow the reviewer comments and provide detailed information on model calibration in the revised version of our manuscript. We have included an additional section providing a more detailed description of model calibration for PANTA RHEI and the snow models. Moreover, we added plots illustrating modelled SWE and observed snow depth. We thank reviewer #1 for these valuable suggestions, they really strengthened our manuscript!

<u>Specific comments:</u>

Comment #1:
"Line 9: Can you proof this assumption by data or by a citation. Does it matter at all?"

Reply in the open discussion:
If you refer to page 4067, line 9, we agree that it does not matter at all. A comparison between climatological means of temperature and precipitation among stations in the Harz and the Alps could underline this statement. However, as it is not particularly relevant for this study, we will delete this sentence.

Implications for the manuscript:
The sentence has been deleted – thank you for this suggestion!

Comment #2:
"P4068 Line 1 pp.: What is the database for the mentioned values? How many stations were available? Where are these stations located? What is the configuration of the stations?"

Reply in the open discussion:
All values refer to monthly means as well as long-term recordings of Braunlage automatic weather station. We will briefly describe this and we will indicate the location of this station in the map (Fig. 1).

Implications for the manuscript:
Done – thanks!

Comment #3:
"P4068 Line10 pp.: What are the reasons for choosing these two winter seasons? Why have you chosen seasons which are significantly different from the average? Would it make sense to analyze a season which is close to the average as a benchmark?"

Reply in the open discussion:
The basic idea behind choosing two different winter seasons is based on the concept of applying a differential split sample test, which is described at the end of this section. The question you have raised is interesting and it would be worth pursuing this issue. As explained in Sect. 2.2, observations with high temporal resolution were only available for the last decade. But this would also raise the question which winter seasons during the last decade are closest to average? We therefore decided to select two different winter seasons in order to test the models for a wide range of possible meteorological conditions.

Implications for the manuscript:
Reviewer #1 is right; the explanations with respect to the chosen winter seasons have been revised. We added some information about the computation time, which still limits dynamical downscaling over longer periods of time. We have added references for this ongoing discussion to the revised version of the manuscript.

Comment #4:
"P4074 Line 3: On the basis of which criteria?"

Reply in the open discussion:
The subdivision of sub-catchments was carried out using digital elevation data, as it is a straightforward approach. In a subsequent step, hydrological response units were derived using land use and soil information. This section will be completed with respect to these explanations.

Implications for the manuscript:
The respective explanations have been added, thank you!

Comment #5:
"P4074 Line 4: What information have you used for doing this?"

Reply in the open discussion:
The local water supply company provides seasonal flow rates, which have also been observed using gauging stations. We will include this information in the updated manuscript and provide the respective reference.

Implications for the manuscript:
The respective information and reference has been included.

Comment #6:
"P 4074 Line 6pp: Which parameters where calibrated? Which calibration scheme was used, what are the quality criteria's? What was the quality of the model after calibration?

Why have you calibrated the model by using met-stations instead of using WRF fields, which are used for the successive model runs? What was calibrated in which snow model? Why have you used WRF fields here and no meteorological stations? Please show why a calibration scheme which id based on two input data sets (meteorological stations for the hydrological part/ WRF for the snow model part) is consistent and applicable."

Reply in the open discussion:
The model was calibrated manually by maximizing the Nash-Sutcliffe model efficiency. We will add the model efficiency for both calibration steps. Mainly the parameters of the soil, the runoff concentration as well as the routing model were altered. In addition the temperature-index model was calibrated for the long-term simulations. We will briefly describe this. We think that the description of every single parameter would be too detailed for the scope of this study. However, we will provide the number of adjusted parameters.
You ask for proof of the consistency and applicability of the calibration procedure, since it is based on two input datasets. We cannot prove this in a more rigorous way, e.g., by using a cross-validation. However, the results of the passively coupled atmospheric and hydrological model are still good. Since we have also considered a validation period for each calibration step, we expect the model to be valid.

Implications for the manuscript:
To address the reviewer comment, the revised manuscript includes an additional section describing model calibration for both the hydrological model and the snow models. We explain that the model calibration was carried out using downscaled meteorological fields and observed precipitation. This combined dataset is used for calibration only in order to avoid misleading parameterizations due to uncertainties in modelled precipitation. Moreover, the model efficiency is provided for each of the models. Thanks, these suggestions have really improved the manuscript!

Comment #7:
"P4074 Line 21: Please show the network."

Reply in the open discussion:
The network will be added to the map (Fig. 1).

Implications for the manuscript:
The network has been added to the map (Fig. 1).

Comment #8:
"P4074 Line 24: Why have you chosen such a generalized illustration of precipitation? If 19 stations area available a spatially distributed illustration would be possible by e.g. showing the Nash Sutcliff coefficient for any station. This would also allow a more precise estimation of the quality of precipitation within the study area. Moreover, it would be also necessary to see if WRF is able to calculate the correct phase of precipitation."

Reply in the open discussion:
We could provide a table including the precipitation depth for both the entire winter season and the snowmelt event in March/April 2006 for all stations. We could also present Nash-Sutcliffe model efficiencies for different time steps and both periods. However, we think that this information would go too far within the context of this paper. Such a table could be enclosed as an appendix.

Implications for the manuscript:
We have followed reviewer #1's suggestion and have added a table providing information about the model performance with respect to precipitation including model efficiencies for each station.

Comment #9:
"P4075 Line 12: The usage of Nash Sutcliff would be again more meaningful than r2."

Reply in the open discussion:
We will also provide this information for the temperature simulations.

Implications for the manuscript:
The suggestion of reviewer #1 has been followed and the respective section has been updated!

Comment #10:
"P4075 Line 16pp: This is critical and leads to a negative evaluation of the whole paper. As you said precipitation is a key parameter with respect to snow cover modelling. You mention the problem of the areal representativeness of point measurements in the introduction and mentioning that this is the reason for using models like WRF. But right now you are going in the opposite direction because you are argue that point scale precipitation is better than WRF."

Reply in the open discussion:
As outlined in the general comments section, we will provide results based on WRF precipitation. These results will clearly underline that is not necessarily possible to simulate single events reliably. By substituting simulated with observed precipitation we show that in this case, point scale information is better than the model output because the snowmelt models perform well using observed precipitation (we will carry out a comparative study as suggested by Anonymous Referee #2). Alternatively, this configuration underlines the applicability of temperature, humidity, radiation, and wind speed. We believe that these results are worth being reported, regardless of the fact that it somewhat constitutes a failure story. As recommended by Andréassian et al. (2010, p. 855), "hydrologists should be encouraged to dedicate part of their publication efforts to reporting their mistakes or what can be called negative results." However, there is still a significant contribution to hydrological modelling in a positive sense, since temperature, humidity, radiation and wind speed can be downscaled with sufficient accuracy for snowmelt modelling at both scales. These results may be of interest to the scientific community since these observations are sometimes hardly available.

Implications for the manuscript:
As explained in the open discussion, the revised manuscript has been substantially updated with respect to reviewer #1's comments. We have decided to compare simulations that involve either observed or downscaled meteorological time series only, rather than showing results based on combined observed and simulated meteorological input. We believe that this modification of the manuscript has really improved our paper. This new approach enables us to really assess the value of downscaled WRF data in hydrological modelling and also corresponds to the comparative study, which was suggested by reviewer #2. Thank you very much for these valuable suggestions!

Comment #11:
"P4076 Line 1pp: The results are hard to interpret. First of all you are discussing the results for the calibration period by using no quantitative measures. Moreover, it is not clear if you have used WRF for the calibration of the snow models (as it was mentioned before) or WRF and measured precipitation which would be more meaningful as you are driving the models in this configuration."

Reply in the open discussion:
We will add further quantitative measures including Nash-Sutcliffe model efficiencies. We agree that the current explanations with respect to calibration and the precipitation input are ambiguous and unclear. In all cases, we used observed precipitation for the calibration as well as the validation. We will clarify all statements that refer to the calibration of the snowmelt models.

Implications for the manuscript:
As suggested by the reviewer, model calibration is now explained in much more detail (please also refer to our reply to comment #6). Observed precipitation is now only involved for model calibration. The evaluated and discussed results in the updated version of the manuscript are now either based on observed meteorological time series or WRF data.

Comment #12:
"P4077 Line11pp: You are validating your models mainly on the basis of snow melt. It would be good to know which kind of Lysimeter you have used. Moreover, it is well known that the melt rates measured by Lysimeters can be significantly biased. Hence, additional parameters would be needed for estimating the quality of the models (e.g. snow depth, SWE)"

Reply in the open discussion:
The snowmelt lysimeter located at Torfhaus is an unenclosed type with a small rim above the collector, which covers an area of 2m^2. A tipping bucket is installed below the collector in order to continuously record melt rates. We will complete this information about the lysimeter in Sect. 2.1. We agree that lysimeters may be biased, e.g., by lateral flow. However, we carefully checked the data by comparing total runoff depth with precipitation measurements and the timing of the melt rates. In our opinion, the lysimeter seems to provide reliable melt rates. Moreover, the evaluation against melt rates enables a comparison of the models across scales when considering stream flow observations at the catchment scale.

Since snow water equivalent measurements are not available, we will provide a figure including snow water equivalent simulations and the corresponding snow depth observations at Torfhaus.

Implications for the manuscript:
The explanations with respect to the lysimeter were added. We included two plots illustrating modelled SWE and observed snow depth in order to provide more information about model performance.

Comment #13:
"P4078 Line 3: How? Which parameters of which models? What are the input parameters?"

Reply in the open discussion:
As explained earlier, we will clarify the text with respect to precipitation input. Moreover, we will briefly explain which parameters were adjusted (mainly concerning the calibration of forest effects).

Implications for the manuscript:
All adjustable parameters for each snow model are now briefly described in the *Snow models* section. Moreover, we have added a table that includes the parameter sets of each snow model after the calibration procedure. Thanks for pointing us in this direction!

Comment #14:
"P4078 Line 7pp: I don't see your calibration strategy. How have you altered the parameters? How have you defined parameter ranges?"

Reply in the open discussion:
The parameters, which will be added to the text as stated in the previous comment reply, were also altered manually, as were the parameters for the hydrological model of the Sieber catchment.

Implications for the manuscript:
The information above was added to the manuscript.

Comment #15:
"P4078 Line 29pp: This is problematic because in the abstract you are talking about an approach able to simulate snow pack and snow melt processes on the catchment scale by using WRF fields. First of all you have excluded the precipitation field and now you are not able to give a measure for the accuracy of the snow pack evolution."

Reply in the open discussion:
We agree that snowmelt alone does not necessarily provide a comprehensive measure of snow pack evaluation. Since we will add snow depth observations, which also enable a model evaluation throughout the entire winter season, this statement will be proven in more comprehensive way.

Implications for the manuscript:
The manuscript has been revised with respect to both issues. Downscaled precipitation is now used as model input and the model results (SWE) are also compared with snow depth observations.

Comment #16:
"P4079 Line 13pp: I don't think that this kind of evaluation is adequate. The model should be quantitatively validated by available and meaningful parameters. At least everything is guessing in here. It is probably that the inclusion of the canopy stands in ESCIMO improves the whole model as it was shown by Warscher (2013) but they have used a consistent model package and consistent validation data for showing the effect of different model components on the quality of the results. Here we have a highly calibrated model and an improvement can be due to the inclusion of the named effects or we have a typical case for being right for the wrong reasons."

Reply in the open discussion:
We do not agree with the reviewer's statement "everything is guessing in here". All snowmelt models were quantitatively validated including RMSE, correlation and a comparison of standard deviations. In contrast to exclusively using a single measure, this enables gaining insight into the question why models do not perfectly match the observed time series, as is always the case. The correlation gives some information about the mismatch in phase, whereas the RMSE provides information about the bias of the model (even though the central pattern RMSE is somewhat limited with respect to that).
In our opinion, both Taylor plots adequately represent the model performance for single models and among all models between both scales. However, we could provide a table with all performance measures including an additional column that includes Nash-Sutcliffe model efficiencies.
The comparison of water vapour mass flux simulations (Fig. 11) is not intended to be a validation of models because adequate measurements are not available. We do not state that any of the models perform better with respect to runoff as a result of these processes representations. We merely pointed out that the inclusion of canopy processes is generally more reliable and may be relevant for other sites.
The concept behind this comparison is that processes may change due to scaling, which may also result in different representations of specific processes. We think that the different sign of water vapour mass flux is worth noting because the different representations of processes may lead to opposing results with respect to this component of the water balance. In other regions, this difference may also affect the water balance simulation to a higher degree. In conclusion, Fig. 11 is neither meant for model validation nor for showing the effect of different model components on the quality of results (as described by Warscher et al., 2013). We will clarify that this comparison is intended to emphasize the different behaviour of the tested models for plausibility only.

Implications for the manuscript:
As explained in the open discussion, it was not our intention to validate models with respect to sublimation. This is not possible because adequate observations are not available. As the comparison of different sublimation simulations could be

misinterpreted as part of the validation, we have removed the respective paragraphs from the manuscript.

References:
Andréassian, V., Perrin, C., Parent, E., and Bárdossy, A.: The Court of Miracles of Hydrology: can failure stories contribute to hydrological science?, Hydrolog. Sci. J., 55, 6, 849-856, http://dx.doi.org/10.1080/02626667.2010.506050, 2010.
Warscher, M., Strasser, U., Kraller, G., Marke, T., Franz, H., and Kunstmann, H.: Performance of complex snow cover descriptions in a distributed hydrological model system: a case study for the high Alpine terrain of the Berchtesgaden Alps, Water Resour. Res., 49, 2619–2637, 2013.

Reply to Anonymous Referee #2

We would like to thank Anonymous Referee #2 for a detailed review of the manuscript. The comments will help us to improve our manuscript. Please find below our detailed response.

General comments:

Comment #1:
"The paper deals with the value of WRF downscaled meteorological fields for driving snow modules. Ground measurements usually limit performances of hydrologic model in mountain catchments, where atmospheric forcings may vary within restricted horizontal distances due to topographic effects. Several published contributions describe the value of downscaled meteo data for driving energy-balance/temperature-index snowmelt models. Here, the innovative contribution is related to the use of four independent snowmelt models, as it is indicated within the introduction. This should help in making results model-independent, if one is investigating the value of downscaled inputs for modeling snow processes.

Overall, the topic is interesting and meets the requirements of the journal. Methods are clearly explained with synthetic sentences.

The assessment of the manuscript is highly conditioned by the fact that downscaled precipitation is not involved. This invalidates title, abstract and introduction. It even affects results, since it is not possible to draw conclusions about the usefulness of WRF atmospheric forcings for modeling snow accumulation and snowmelt. Your approach simulates well both discharge and melt runoff, but it is still dependent on in-situ data. Results do not provide useful information if one wants to applied WRF outputs for hydrologic purposes, thus avoiding observations. These limitations must be overcome if you are going to maintain the same targets for your paper. Results obtained using downscaled precipitation must be shown in accordance to the title."

Reply in the open discussion:
We agree that involving modelled precipitation for snowmelt simulations would be more appropriate with respect to title, abstract and introduction. Using simulated precipitation would indeed give an insight into the applicability of downscaled data. Thus, we will revise our manuscript with respect to this issue.

Implications for the manuscript:
The revised manuscript has been updated according to reviewer #2's comment. We are now comparing simulations that involve either observed or downscaled meteorological time series, rather than showing results based on combined observed and simulated meteorological input. This approach enables us to fully assess the value of downscaled WRF data in hydrological modelling and has also been suggested by reviewer #1. Thanks for this valuable suggestion!

Comment #2:
"An idea could be an additional comparative study between the performances you got at the point scale, driving snowmelt models by WRF fields (including precipitation, even if

the simulation is not good) and the results obtained using in situ measurements (you stated that temperature and precipitation are recorded: these seem the key factors and they are certainly enough for the temperature-index approach and maybe for the model Walter et al. 2005.). This will give an indication about the loss of accuracy due to downscaled forcings when compared to reliable in situ data."

Reply in the open discussion:
We wish to thank Anonymous Referee #2 for this suggestion. We believe that such a comparative study would be very helpful in order to improve our manuscript. We will take the suggestion into account and we will compare temperature-index simulations for both cases (observed and simulated meteorological input).

Implications for the manuscript:
The *Results and discussion* section has been revised according to this suggestion (also see reply to comment #1).

Comment #3:
"Again, catchment scale simulations should use downscaled precipitation. Here, a comparison could be performed for understanding whether this approach outperforms the results provided by a spatial distribution of observed temperature and precipitation. You spent several sentences on the issue of "area representativeness of point observations" within the introduction. At the end, it is not clear if WRF data help for hydrologic modeling when compared against standard inputs (ie ground-based measurements). Methods for distributing point values (eg kriging) exist. For instance, using the simple degree-day approach, it could be implemented a spatial distribution of temperature with altitude (a constant lapse rate is usually adopted). Then, you can run the degree-day model coupled to PANTA RHEI."

Reply in the open discussion:
For the catchment scale simulations we will also present results that rely on downscaled precipitation. According to your previous suggestion, which focuses on the point scale simulations, a similar comparative study for the catchment scale will be carried out and presented. Therefore, two configurations of PANTA RHEI / temperature-index input will be taken into account: i) standard input including observed precipitation and temperature assuming a constant lapse rate and ii) downscaled temperature and precipitation.
This comparison also emphasises that it was not possible to derive precipitation fields with high accuracy at the considered spatial and temporal scales.

Implications for the manuscript:
We have followed reviewer #2's suggestion and have extended the comparative study towards the catchment scale using PANTA RHEI. The manuscript has been updated accordingly.

Comment #4:
"It is not clear whether you are discussing the value of the combined use of WRF and ground-data or the value of WRF meteo data alone. In the first case, you should state that your study uses both data sources from the beginning, and the results must be

discussed with this focus. Even in this case, additional analysis should be provided. For example, it could be possible to use precipitation and temperature recordings (which seem available at several stations around the catchment) and downscaled radiation, wind speed etc., if such data are not measured."

Reply in the open discussion:
We think that your previously mentioned suggestions will also help us in the process to improve our discussion. Since we will show results based on both observed and downscaled precipitation, we will focus on the value of all downscaled meteorological variables, even though precipitation is not as accurate. As explained in our earlier comment (15 May 2014), we suggest using observed precipitation if available.

Implications for the manuscript:
According to the revised *Results and discussion* section, we have also updated the *Summary and conclusions* section with respect to the changed structure. Following the comments of all reviewers we are now comparing simulations that involve either observed or downscaled meteorological time series and are not showing results using a combined dataset.

Specific comments

Comment #1:
"If you are not going to use WRF precipitation the title should be changed. This must be claimed also within the abstract. In the title, the sentence "modelling of snow processes in catchment hydrology" seems very general. Actually you are considering only one (and particular) case study. Probably the title should indicate this."

Reply in the open discussion:
As explained in our reply to your general comments, we will consider downscaled precipitation in our revised manuscript. Hence, we think that it is not necessary to change our title with respect to precipitation. However, we could indicate the study area through providing a subtitle: "A numerical experiment for the Sieber catchment (Harz Mountains, Germany)"

Implications for the manuscript:
Following reviewer #2's suggestion, we have changed the title including the study area: "Effect of meteorological forcing and snow model complexity on hydrological simulations in the Sieber catchment (Harz Mountains, Germany)"

Comment #2:
"P 4066 Line 5: I do not think this is always true. Please state that the problem of spatial resolution is mainly related to complex topographies of mountain regions."

Reply in the open discussion:
Indeed, this statement holds especially in the case of mountainous terrain. We will rephrase our statement accordingly.

Implications for the manuscript:
The sentence has been rephrased according to reviewer #2's suggestion.

Comment #3:
"Study area" section: It is not clear the whole number of stations available around the catchment and what kind of data are provided (only temperature and precipitation?). This is useful since you might run the models using ground observations and see what happens in comparison with WRF inputs. This might be an interesting contribution by your paper."

Reply in the open discussion:
As suggested by Anonymous Referee #1, we will add the station network to the map depicted in Fig. 1. We will also indicate the variables measured at each station.

Implications for the manuscript:
The manuscript has been updated following reviewer #2's advice – thank you!

Comment #4:
"Selected winter season" section: I agree with referee 1 that two particular winter seasons" are not enough for drawing conclusions. Please consider the possibility of involving a third hydrologic cycle."

Reply in the open discussion:
We are in the process of preparing a third winter season.

Implications for the manuscript:
We actually carried out simulations for a third winter season (see, Fig. 1 of this document). The results are in agreement to the winter seasons that have already been considered. However, this additional winter season does neither bring any further insight nor does it fulfil the requirements of long-term simulations, as it would be highly desirable (most existing studies that consider LAM / RCM data with high spatial resolution consider a few months only, which can be related to computational demands). Thus, we decided to omit this third winter season, not least because of the additional plots that would be necessary.
Moreover, we have added some discussion and references that justifies the chosen period of time.

Comment #5:
"Snowmelt models" section: please, spend some times in describing what parameters you calibrated, perhaps by inserting a summary table. Are they calibrated using melt rates from the lysimeter?
Could be possible to better explain how you designed accumulation and melting for the temperature-index model? Did you consider refreezing? Are you using only one degree-day factor for the entire basin? Did you consider the additional energy input by rain-on-snow?"

Reply in the open discussion:
We will briefly describe the parameters that have been altered in the calibration process of each snowmelt model. Refreezing and different energy sources were neglected and we provided land use dependent degree-day factors. Calibrations were carried out using lysimeter data. This information will also be added.

Implications for the manuscript:
Following reviewer #2's advice, the snow models section now includes more details about adjustable parameters and we explain that we only consider the most basic version of the temperature-index model, which neglects the above-mentioned processes. The calibration was carried out using lysimeter data, which is also explained in the revised manuscript. Following your suggestion, we have added a summary table.

Comment #6:
"P 4074 line 9: how did you calibrate snowmelt models for catchment scale simulations? What did you calibrate?"

Reply in the open discussion:
Each snowmelt model was calibrated through altering the parameter most relevant for canopy effects. The description of this approach will be added to the text.

Implications for the manuscript:
The descriptions described above have been added to the manuscript.

Comment #7:
"P 4077 line 20: please remove comma after "concluded"."

Implications for the manuscript:
Done.

Comment #8:
"P 4081 line 19: the fact that considering snow processes instead of "no snow" you are improving runoff simulations is not a finding. It would have been a problem if it happened the opposite. On the contrary, you should discuss how much you improve the "no snow" simulation and if it justifies the use of a more complex hydrologic model. Anyway, your sentence does not seem scientifically relevant for the goals of your manuscript."

Reply in the open discussion:
We will delete this sentence.

Implications for the manuscript:
Done.

Comment #9:
"P 4082 line 3: to do what? you are combining data sources. Please explain why and where the presented approach could be applied."

Reply in the open discussion:
We will rewrite these concluding remarks according to our first comment and the updated simulations. Using observed precipitation in combination with downscaled data could be seen as an alternative for regions where e.g. only a few precipitation observations are available and all other meteorological data are missing.

Implications for the manuscript:
As explained earlier, we have revised the *Summary and conclusions* section according to the restructured *Results and discussion* section, which now includes the comparative study suggested by reviewer #1 and #2. In contrast to our reply during the open discussion we do not consider combined WRF / observed precipitation datasets in the revised manuscript. We consider this a major change in the revised manuscript that has really improved the quality of our study. We thankfully acknowledge this valuable input from the reviewers!

Comment #10:
"Fig. 5: the plot is not very clear due to the high frequency. Please consider to enlarge the x-axis or restrict the temporal window you are showing. Otherwise you may split it into two time frames."

Reply in the open discussion:
The aspect ratio of the axes aspect ratio will be improved accordingly.

Implications for the manuscript:
Done.

Comment #11:
Fig. 8: why the name "snowmelt simultation" as in fig. 7? Are they streamflows at the closure section? Here, the name of the figure should be "catchment discharges considering snowmelt", or something like that."

Reply in the open discussion:
We will change the figure captions to "Stream flow simulations considering different snowmelt models".

Implications for the manuscript:
We have followed reviewer #2's suggestion and have modified the manuscript accordingly.

Reply to Anonymous Referee #3

We would like to thank Anonymous Referee #3 for a detailed review of the manuscript. The comments and suggestions helped us to improve our manuscript. Please find our detailed response below.

General comments:

"Overall the paper deals with an interesting question, because data availability is often the bottleneck for modeling.
I generally agree with the other referees, in particular it is essential to mention that WRF precipitation is not used to run the models in the abstract. Moreover there should be more specific information given, e.g. about how the calibration of the models was done. Also it was not clear if the measured precipitation is corrected for systematic errors like undercatch.
Overall I suggest to focus revision on the calibration of the models and giving more specific information about them, e.g. if they use different thresholds to divide between rain and snow. This makes results hard to compare. Also the degree-day method should be explained better regarding the use of a fixed or variable degree-day factor. Finally, if there are snow height measurements available, they could be used to evaluate snow cover development for the point scale modeling.
The specific comments may help to provide missing information and if these deficits can be eliminated the paper meets the requirements of the journal."

Reply in the open discussion:
In the abstract we will mention that we also consider observed precipitation for snowmelt simulations. As explained earlier, we will also involve WRF precipitation in order to more clearly and consistently show the applicability of downscaled meteorological fields.
We will also rewrite the text with respect to calibration. In this context, the different ways in which the models separate rain and snow will also be explained briefly. More emphasis will also be put on how the degree-day factor has been considered. Since snow depth observations are available to some extent, a comparison with modelled SWE will be included.

Implications for the manuscript:
The revised manuscript shows results based on downscaled precipitation. We followed reviewer #3's advice and added information with regard to the parameters of the snow models in the *Snow model* section including a more detailed description of the temperature-index model. As suggested, we briefly explain how snow and rain are separated. Furthermore, it is now indicated thoroughly in the captions whether precipitation is corrected or not. SWE simulations are now plotted against snow depth observations.
As suggested, we focused our revision on model calibration and added a new section *Model calibration*, which includes a description of model calibration for the hydrological model and the snow models. The parameters of the calibrated snow models are listed using an additional table. Thank you very much for these valuable suggestions!

Specific comments:

Comment #1:
"P4067: Line 8pp: Regarding which climatological parameters?"

Reply in the open discussion:
This statement refers to mean annual precipitation depth and temperature. As explained in our response to Anonymous Referee #1, we will delete this statement since it is not relevant.

Implications for the manuscript:
The respective statement has been deleted.

Comment #2:
"P4067: Line 21pp: What is the range in altitude in the Sieber catchment? Are rain gauges representing the topography of the catchment?"

Reply in the open discussion:
The altitudinal range of the Sieber catchment covers 340 m to 920 m, which will be included in the improved text. The rain gauges represent the general topography of the catchment. We will provide a map including all stations, displaying that the surrounding stations represent the altitudinal range of the catchment.

Implications for the manuscript:
The manuscript has been updated with a description and a map as described above.

Comment #3:
"P4068: Line 3pp: Is there a reference? Or derived through own data analysis?"

Reply in the open discussion:
This value is derived through own analysis. Mean daily values of discharge are available from 1930 onward, in the form of mean daily values. Rather than to limit these to 2008, we will extend the time period to 2013.

Implications for the manuscript:
The value has been changed accordingly.

Comment #4:
"P4068: Line 10pp: Why are just 2 years considered in the study? How well are the meteorological values modeled for the other seasons?"

Reply in the open discussion:
We agree that a long-term evaluation would be of great value. Depending on the computer hardware, WRF applications run 3 up to 20 days in order to simulate the meteorological fields of one single winter season. The 3 day estimate refers to high performance computing with 24 cores whereas the 20 days would be necessary to run the model on a computer with 4 cores. We will add the information regarding computing

time in order to show the limitations of this approach. Unfortunately, due to required computing power it is not possible to downscale a greater number of winter seasons under the present conditions.
We configured the model only for the winter seasons mentioned in the text. Hence, a continuous run is not available and the study focuses on melt events. We tested the applicability of downscaled meteorological fields for two winter seasons in order to show the principle applicability of this approach. Instead of using long-term simulations we focused on running different snowmelt models at the point and the catchment scale to compensate the use of short periods of time to evaluate the applicability of downscaled meteorological fields.
We see the WRF application of longer time series as an opportunity for future research, a point of view we will add to the concluding remarks.
However, we carried out simulation runs for a third winter season, as suggested. The point scale results are depicted in Fig. 1. In contrast to the Taylor plots in the manuscript, the herein displayed results are based on modelled precipitation. Since we will revise our manuscript to include modelled precipitation input, as announced, the updated manuscript will be considerably more extensive than the current version. Hence, we believe that providing all additional plots including a discussion for the additional winter season would go beyond the scope of this study, since the evaluation of the additional winter season does not provide further insight. The already discussed results, based on two winter seasons, represent sufficiently different conditions to prove the main research question.
We see this as a realistic compromise between conciseness and completeness.

Implications for the manuscript:
As explained earlier (reply to comment #4 of reviewer #2), we actually carried out simulations for a third winter season (see, Fig. 1 of this document). The results are in agreement to the winter seasons that have already been considered in the submitted manuscript.
However, these simulations do not bring any further insight nor does adding a third winter season fulfil the requirements of long-term simulations, as it would be highly desirable (most existing studies that consider LAM / RCM data with high spatial resolution consider a few months only, which can be related to computational demands). Thus, we decided to omit this third winter season, not least because of the additional plots that would be necessary.
Moreover, we have added some discussion and references that explain and justify the chosen period of time.

Comment #5:
"P4071: Line 26 pp: Please mention more about the used degree day method. Is the degree-day factor changing over the season or is a fixed factor used over the whole period?"

Reply in the open discussion:
We use time-independent degree-day values in order to consider the basic temperature-index method. However, we provide a lookup table for PANTA RHEI simulations including various land use classes for catchment scale applications. This will be outlined more precisely.

Implications for the manuscript:
Information on the applied temperature-index method has been added as suggested by reviewer #3.

Comment #6:
"P4074: Line 6 pp: Please clarify more how the calibration was done. Why did you calibrate with meteorological data from 1971 to 2000 but precipitation from 2002 to 2008? Why to use measured meteorological data if the model is driven by WRF data?"

Reply in the open discussion:
We carried out two calibration steps. For the first calibration period, only daily meteorological time series are available. This first calibration procedure accounts for the calibration of the water balance components of the hydrological model (e.g., soil model, evapotranspiration). The subsequent shorter period enables the calibration of flood peaks since hourly precipitation time series were considered. We will describe the calibration procedure in a more detailed manner.

Implications for the manuscript:
An additional section has been added to the revised manuscript in order to provide more detailed information on the calibration procedure for both the hydrological model and the snow models. Thank you for this suggestion!

Comment #7:
"P4074: Line 8pp: Why don't you use another year (representing more average conditions) for calibration instead of one included in the study? Especially since you are just modeling 2 years."

Reply in the open discussion:
The basic idea was to make optimal use of the short periods of time. Therefore, we decided to use the first winter season for calibration while the second winter season is seen as validation period. Since both seasons differ in terms of meteorological conditions, this modelling experiment can be seen as a differential split sample test. We will clarify this.

Implications for the manuscript:
In the revised manuscript we explain that we intended to consider different meteorological conditions at still reasonable computational costs (see reply to comment #4). Therefore, we have selected two different winter seasons that reflect different meteorological conditions.

Comment #8:
"P4074: Line 17 p: It would be interesting to have a plot of the other meteorological time series as well."

Reply in the open discussion:
It is possible to provide such plots for humidity, wind speed and radiation components. We are considering providing an additional plot, including these variables or at least performance measures.

Implications for the manuscript:
In order to keep the paper concise we provide a summary table instead of plots. This table includes performance measures for the other meteorological variables. Thank you for pointing this out!

Comment #9:
„P4074: Line 20pp: FIGURE 4: Is the measured data compensated for errors like wind error? How does data from different stations look like? Are there some stations representing the modeled data?"

Reply in the open discussion:
The observed precipitation time series are not corrected with respect to systematic errors. We will add this information.

Implications for the manuscript:
The respective information has been added to the revised manuscript.

Comment #10:
„P4076: Line 1pp: FIGURE 6: Is snow height measured at point scale? At the beginning of the event SWE values range from approx. 275 to 375 mm for the different models. So not only runoff for this event is interesting, also if the whole season is represented correctly."

Reply in the open discussion:
We wholeheartedly agree with this. As explained in our reply to Anonymous Referee #1, we will provide a plot including the modelled SWE for all models including observed snow depth.

Implications for the manuscript:
A plot illustrating the modelled SWE for all models including observed snow depth has been added to the manuscript.

Comment #11:
„P4077: Line 19pp: Is that a good measure for model performance?"

Reply in the open discussion:
In response to this very valid question: it is indeed not a good performance measure. However, this information gains insight to the question if snowmelt processes are generally relevant for the selected sites. This comparison is intended to show the plausibility of results.

Implications for the manuscript:
As the comparison of different sublimation simulations could be misinterpreted as part of the validation, we have removed the respective paragraphs from the manuscript.

Comment #12:
„P4077: Line 23p: Why weren't more winters used then? If comparing different models in performance, than a comparison or ranking should be possible."

Reply in the open discussion:
As previously explained, carrying out further evaluations is limited by computation time. You are right to say that using longer time series would enable a comparison of model performance. We consider your comment a possibility for future research. However, our primary aim was to show that the downscaled meteorological data are suitable to drive snowmelt models at different scales. We agree it is not necessarily possible to compare the models using only data of two winter seasons. We see the application of more than one model as a surrogate for long time-series applied to one model in order to prove our hypothesis.

Implications for the manuscript:
As explained in our reply to comment #4, computational costs still limit LAM applications at high spatial resolution. We now explain this issue in the revised text and show that comparable studies struggle with the same limitations leading to a consideration of even shorter periods of time that merely range from single events to several months. Of course we agree that a model inter-comparison study would benefit from long-term datasets.

Comment #13:
„P4077: It would be interesting to also see whole winter seasons for point observations."

Reply in the open discussion:
We will add a plot for the winter season in its entirety including both observed snow depth and modelled SWE.

Implications for the manuscript:
Done.

Comment #14:
„P4078: Line 26pp: Not all models used the same thresholds to separate rain and snow? How are parameters for the models set? Do comparable parameters differ for the different models?"

Reply in the open discussion:
The parameterization of this separation differs. The temperature-index model and the modified Walter model rely on a simple temperature threshold, the Utah Energy Balance model includes two thresholds considering mixed precipitation and ESCIMO is parameterized using a wet-bulb temperature threshold.

Implications for the manuscript:
The parameter settings for separating rain and snow are now provided for each model.

Comment #15:
„P4079: Line 23: It is not clear, how snow and rain are separated.“

Reply in the open discussion:
We will add a brief description of these parameters in order to clarify this statement.

Implications for the manuscript:
The *Snow models* section now includes a brief description about the rain and snow separation for each model. Thank you for pointing this out!

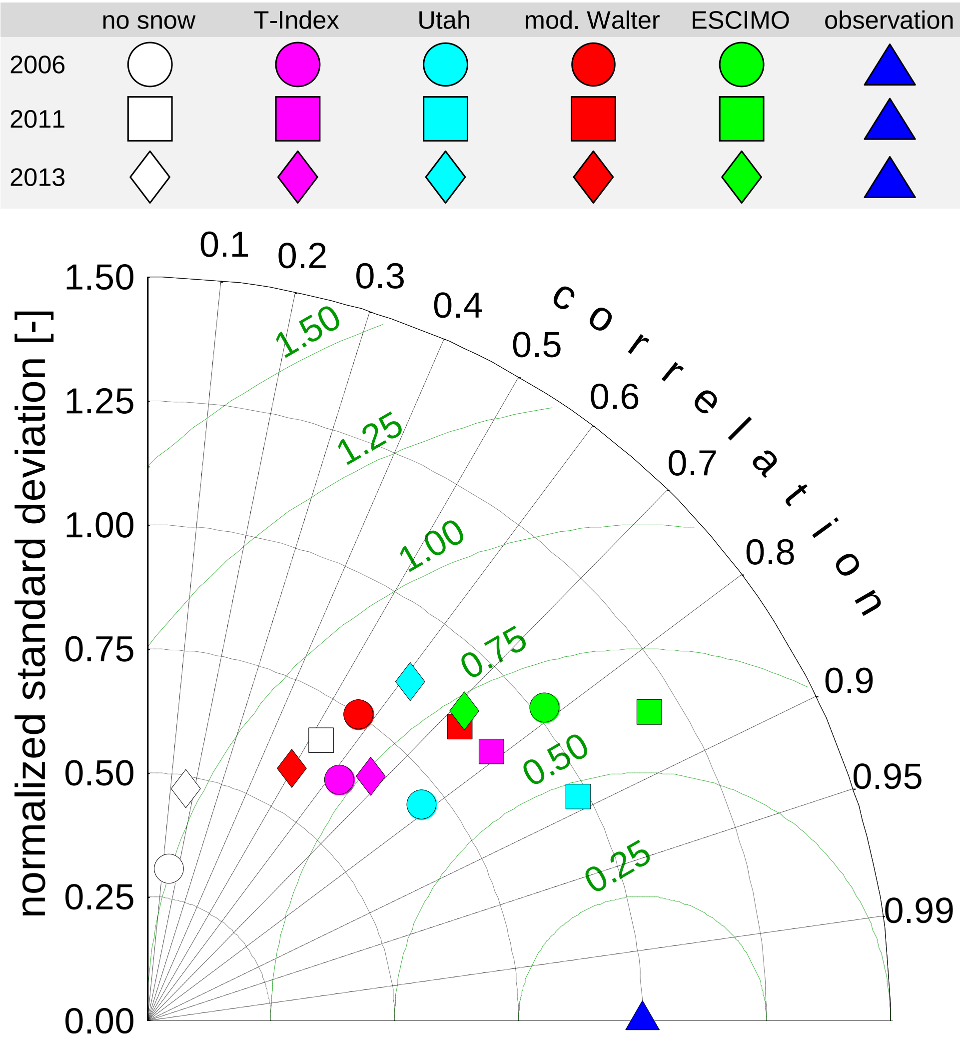

Fig. 1. Taylor plot including point scale simulations for three winter seasons. All simulation runs are based on modelled WRF precipitation.