Review of

Do convection-permitting regional climate models have added value for hydroclimatic simulations? A test case over small and medium-sized catchments in Germany

By Oakley Wagner, Verena Maleska, and Laurens M. Bouwer

General Comments

The presented paper uses the convection permitting regional climate model (CPRCM) ICON-CLM 2.6.4 with the spatial resolution of 3 km together with its driving model with parametrised convection at 11 km resolution for a comparison with measured climate data and as input into a hydrological model in small to medium sized catchments in Eastern Central Germany. Particularly the first part, the comparison with measured climate data, is well and clearly written and provide valuable information about the performance of the models. As the authors stated, it is important to compile more and more example data of CPRCMs to illustrate their possible added value compared to their coarser sister-models (scientific data). This is particularly true regarding the focus on the variables that are crucial for hydrological implications (water budget and floods). Regarding the hydrological modelling part, however, I doubt that the corresponding chapters are ready for publication without revisions (see specific comments). In its current state, the modelling work does not provide additional findings about the usage of these RCMs for climate impact studies. In particular, there are no conclusions given regarding the impact of the different model input variables on the model output, in this case also the representation of floods. Of course, the huge overestimation of precipitation intensities in the CPRCM will play the main role. In this respect, I miss conclusions about the impact of such an overestimation on the flood representation, particularly the non-linear behaviour of the flood generation (threshold processes) compared to the calibrated case and gauge data. This also implies a more comprehensive evaluation of the used hydrological model to accurately represent these processes within the calibration and validation procedure.

Specific comments

P. 6 / L. 88: Are the stations used here implemented in the RADOLAN scheme? If so, are the station values preserved after regionalisation in RADOLAN? Please give a short clarification.

P. 6 / L. 102: Are the climate models run on hourly time step? Is this the effective temporal resolution? Please add a few words or point to the reference.

P 7 / L. 122: Please give the reason to choose the Sturges' rule.

P 7 / L. 136: How do the THIESSEN interpolated rainfall compare to the RADOLAN product? See first comment above. Please add a short comment.

P. 8 / L. 145: The bandwidth is hard to read. I assume it is ±0.08 to ±0.76 K. Please clarify.

P. 8 / L. 154 and further lines and plots (e.g., Fig. 2) in the manuscript: please remove "error" in "monthly mean bias error", this is redundant and may mislead.

P. 10 / Fig. 3: For clarity, I would recommend to make 4 plots for the 4 seasons out of this plot.

P. 11 / L. 200ff: Please rewrite the sentence to clarify and add the general measurement uncertainties (or add references).

P 12 / after Fig. 5: Please add a short chapter of the calculated potential evapotranspiration (I assume, ET0 by the PENMAN-MONTEITH formula), since this is a relevant input variable and includes the mentioned variables. Please also give a short comment about the partial influence of the variables (probably add the ET0 formula). Systematic errors may be relevant in the context of error propagation through the model chain, particularly regarding water budget and soil moisture.

P. 13 / L. 225: Please discuss shortly this (huge) overestimation of the 3km model. Are there many hourly intensities in the 3km model in the range of 100 to 150 mm/hour? Are these single intensities or embedded into longer rainfall events? What does this mean for flood representation in small catchments (flash floods)?

P. 14 / Fig. 7: Please give the threshold for the 99.5 % percentile (app. 2 mm/hour?).

P 15 / L. 246 ff: Is there a flood season? If so, please add the information, in which season mainly the (large) floods occur (seasonality of floods)? Do snow melt induced floods play a role? If so, please refer to the temperature data evaluation.

P. 16 / Fig. 9: Is there a reason to use this kind of presentation (anomalies)? In my view, the monthly values can also be explicative.

P. 16 / L. 269 fff: Please add which precipitation input data where used for calibration, the THIESSEN interpolated or the RADOLAN? How were the hydrological model parameters chosen, which were fixed, and what were the main parameters calibrated? Please add the graphs of the two events for the catchments and quote the (estimated) return periods? How is the performance at smaller events? Is there a non-linear shift in runoff generation from smaller to larger events? Can this be identified in the measured gauge data?

P. 17 / L. 280: In my opinion, this assumption can only be made, if the model adequately represents the main runoff generation processes (water budget, flood generation - non-linearity) and error propagation can be quantified. Therefore, please show the calibration and validation results in a more comprehensive way (see above).

P. 17 / Chapter 3.2.2: The model was calibrated to one single event and validated to another single event. Please show, that the model accurately captures the long term water balance. Please add the observation (or the calibration - is RADOLAN 3km closest to the calibration?) results into Fig. 10 and discuss the differences. Particularly in the context of long term model runs, in a climate change framework occurring systematic errors (uncorrected precipitation and potential evaporation) may lead to an accumulation of errors. In this respect, please discuss also possible biases in the ET0 and the impact on soil moisture simulation. Please also show what flood event peaks are “produced” by the extraordinary high rainfall intensities. Probably give short flood peak statistics (annual maxima or POT - peaks over a certain threshold).

P. 17 / Fig. 10 and further lines and plots: please remove “routed”, just use “discharge”.

P. 18 / Fig. 11: Please add the catchment sizes of the four examples. Also, the relative differences (in %) would be good to mention and to compare to the corresponding (relative) differences in the catchment precipitation.

P. 18 / L. 307: Please clarify “full range of ICON11km meteorological data” or rewrite.

P. 19 / L. 13: Please repeat for clarification that this is the calibration event (e.g., in brackets).

P. 19 & 20 / Fig. 12: I believe this chart requires further discussion. Again: Are the RADOLAN 3km and 11km driven simulations close to the calibration and, can they be seen as reference simulation? Please add a short comment. If so, it seems to me, that the calibration focused on the representation of the flood peak. In my view, with the corresponding graphs, this could be discussed in detail in chapter 3.2.1. Also, the performance regarding the annual and seasonal water balance (discharge volumes) together with the soil moisture and storage simulation should be analysed. The smaller preceding events are significantly overestimated with the RADOLAN input, so one can assume, that the antecedent conditions at the start of the main event are overestimated as well. In this respect, it should be analysed, if the different event sizes and event types are generally represented well by the calibrated model(s) (frequency).

Please add, how the initial conditions in May 2013 were chosen. Were they obtained by the continuous simulation? The high discharge in the RADOLAN driven simulation implies also high soil moisture and storage fillings in the simulation and furthermore, lower antecedent losses and a higher initial discharge at the main event. The observed discharge indicates, that the smaller rainfall events recorded in the RADOLAN data did not lead to a discharge rise, which could be related to a lower soil moisture status in May than simulated.

With the climate model input the simulation results preceding the event are closer to the observation. It would be interesting, how the results would look like if the same antecedent conditions were used for all model inputs. This would help to clarify the impact of the different rainfall input at the particular event and also the role of the initial catchment conditions, i.e. soil moisture. This would lead to the question, if the systematic errors in precipitation and possibly also in ET0, may be more relevant and distort the interpretation of single flood events.

Generally, the underestimation of the flood by the ICON 3km driven model is a rather surprising result, because this model setup considerably overestimates the discharge in the mean and in all presented percentiles in all catchments (Fig. 10). The initial baseflow-to-peak rise, however, seems to be similar to the RADOLAN driven models (or even larger). In any case, these discrepancies have - in my view - to be further addressed.

P. 21 / Chpt. 4: In my view, the discussion needs revision after further analyses have been performed.

P. 21f / Chpt. 4.1, 4.2, 4.3, and 4.4: Please add the implications on hydrology, e.g., snow melt, ET (water budget).

P. 22 / Chpt. 4.5: Please add a paragraph about the impact of the huge precipitation overestimation (refer, e.g., to depth-duration frequency curve (Fig. 8)) on the hydrological model output (non-linearity). What does this mean for flood peaks in terms of shifting of return periods (please also refer to literature). What are the implications for future predictions when such biases occur?

P. 22 / L: 379: ICON3km overestimates the intensity of the highest quantiles. Apparently, it fails at the event. Can this be explained? See above.

P. 23 / Chpt. 4.6: Hydrological simulations: it would be interesting, what event peaks were simulated with these extraordinary intensities (higher than 2013, see Fig. 6). Please provide some conclusions regarding flood statistics. In general, please discuss shortly the usability or advantages of the application of the complex model with lots of input variables and model parameters that are usually difficult to measure (e.g. soil hydraulic conductivity) as well as the large computation time (numerical solution of Richard's Equation), when such high uncertainties in precipitation occur. Can it be recommended for larger catchments, sensitivity analyses or ensemble modelling (e.g., sensitivity of the hydraulic conductivities)?

P. 24 / Chpt. 5 Conclusion: Would you recommend to perform more local studies or to set the focus on larger scale studies? The point-by-point comparison of the meteorological variables, which is a rather strict test, can also be carried out on larger scale. Please add a concluding comment about the applicability of the model chain for future predictions, when such high precipitation biases occur. Would bias correction makes sense?