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.

We thank the reviewer for these comments, for identifying this gap and for the interesting direction of resolution effects on the representation of non-linear behaviour. We have done several additional analyses in response, and propose to make several adjustments to the paper, and give more explanations and answers below. We hope that these answers and adjustments will help to improve our paper, and would make it acceptable for publication.

Disentangling the effects of bias of individual meteorological variables on the flood generation is particularly challenging since many of the key processes in hydrological modelling are affected by the whole range of meteorological input variables. We want to give this more thorough consideration, on the one hand, by giving an overview of the input requirements of the individual modules of WaSiM and, on the other hand, by expanding our focus beyond discharge, conducting analyses on evapotranspiration and soil moisture. A look into threshold processes will allow to merge and extend the gained insights.

The modules in WaSiM for high-resolution hydrological simulation in the temperate zone are for processes of evapotranspiration, snow accumulation and snow melt, interception of snow and precipitation, as well as an unsatured-zone model and a groundwater model. Potential evapotranspiration is calculated after Penman-Monteith, thereby showing dependency on the meteorological input variables of air temperature, global radiation, relative humidity and wind speed. Interception processes are described by an energy balance approach, considering the complete suite of meteorological input variables, just as is the case for the simulation of snow accumulation and snow melt. The unsaturated zone model is built on the Richards-equation and draws on the aforementioned modules, as well as on the groundwater model.

The soil moisture bias in the hydrological model driven with ICON11km is found to be negative compared to when driven with RADOLAN11km (see Fig. 1 below), which is likely a reflection of the underestimation of precipitation intensity by ICON11km. Looking at the model driven with ICON3km, the sign of the bias changes for moderate to high soil moisture estimates given the relative overestimation of rainfall intensities by ICON3km compared to RADOLAN3km. We will include this new QQ-plot in the revised paper. Higher soil moisture in the hydrological model driven by ICON3km is likely to lead to higher runoff in case of a moderate to heavy rainfall event. Soil moisture is further shaped by total evapotranspiration but becomes the limiting factor for the latter in summer.

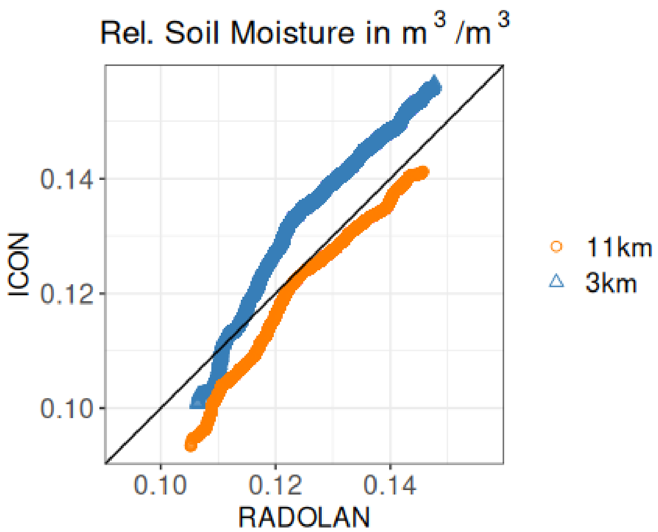

Fig. 1: QQ-plots for relative hourly soil moisture, as calculated by the hydrological model driven with ICON11km and RADOLAN11km, respectively ICON3km and RADOLAN3km for the period 2006 to 2014

In fact, a change in sign of bias is also visible for evapotranspiration estimates in the summer months, when comparing ICON11km and ICON3km (see Fig. 2 below). Besides higher water availability, a reduction of the negative summer bias in temperature and global radiation by ICON3km may contribute to the (slightly) higher ET estimates.

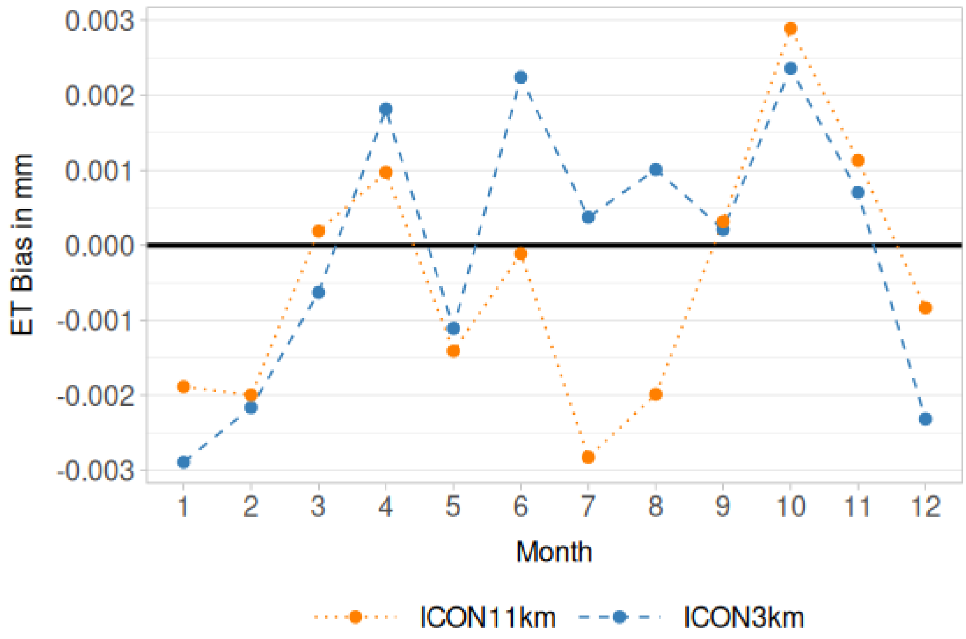

Fig. 2: Monthly mean bias of hourly total evapotranspiration, as calculated by the hydrological model driven with ICON3km and ICON11km for the period 2006 to 2014

We further looked into threshold processes by analysing the hydrological response of runoff volume (Q_{tot}) to the sum of rainfall volume (T_{tot}) and antecedent rainfall (AT). The choice of considered meteorological factors was based on Ross et al. (2021), who found non-linear behaviour in the flood generation to be most commonly present when using the T_{tot} + AR threshold. We employed a modified version of the HydRun tool (Tang & Carey, 2017; with modifications to the runoff peak search algorithm) for the rainfall-runoff-attribution and chose an antecedent rainfall window of 8 hours. We discarded the months of November to May to avoid the effects of snow melt. Fig. 3 shows the relation between runoff and the sum of total rainfall volume and antecedent rainfall for simulations with ICON11km, ICON3km, RADOLAN11km and RADOLAN3km. The segmented regression is highly sensitive to rainfall-runoff events of high return period, such as predominantly recorded from the ICON model simulations. As such, no firm conclusions can be drawn from Fig. 3 on the existence or position of a break point in the relation between runoff and rainfall.

We will include these evaluations in the revised paper.

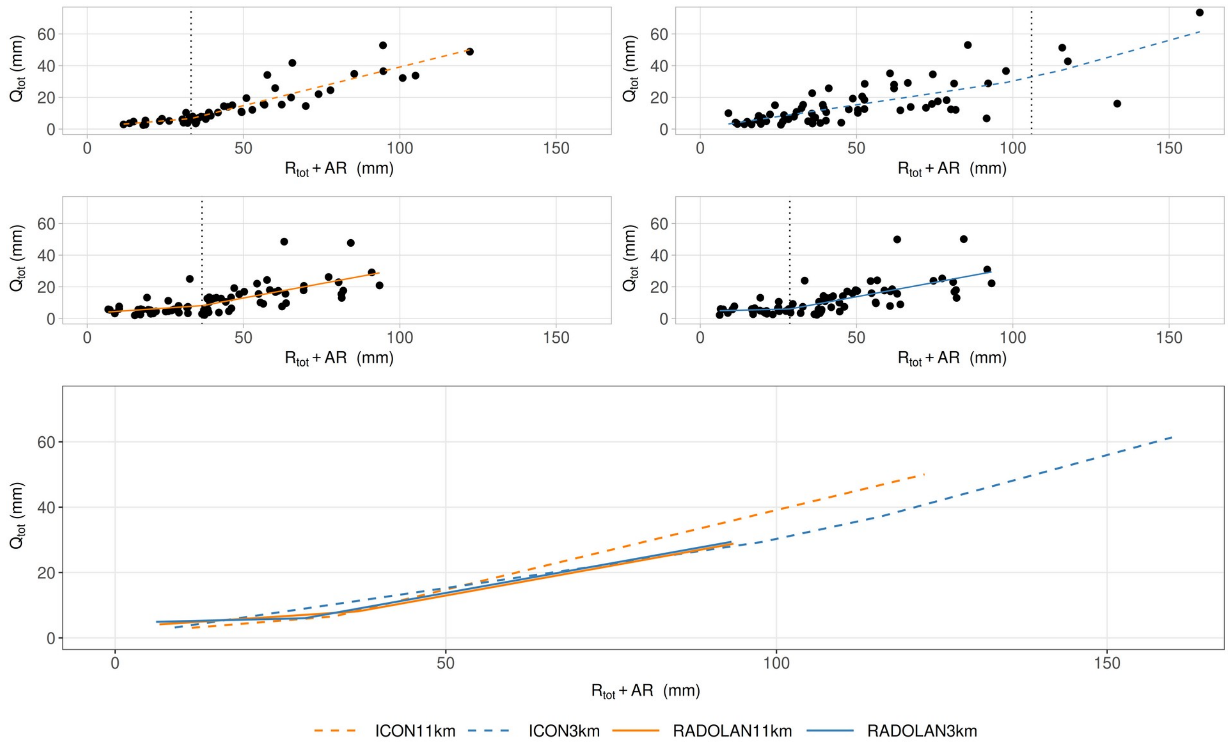

Fig. 3: Upper half: Scatter plots of total runoff (Q_{tot}, min. peak threshold for runoff events: 0.03 mm/h) to the sum of total rainfall (R_{tot}) and 8-h antecedent rainfall (AR) over the catchment of Kleindalzig and those upstream, fitted with a piecewise linear regression model. The thresholds are indicated with a dotted vertical line. Lower half: Combined piecewise linear regression model results of Q_{tot} to R_{tot} + AR

This also implies a more comprehensive evaluation of the used hydrological model to accurately represent these processes within the calibration and validation procedure.

We calibrated the model on the flood of June 2013, which is the largest flood observed in the catchment during the study period. At Zeitz, peak discharges were measured of an estimated return period of 100 years (LHW, 2014). The model was further calibrated for its long-term statistics on the discharge measurements of the year 2012. The year of 2012 was chosen as it shows comparatively low monthly precipitation anomalies. We looked at weekly discharge sums, whereby the weeks of the year were defined according to ISO 8601.

The second largest flood, which occurred in January 2011 as a result of snowmelt, was used for validation despite having a different genesis, since the short study period did not capture another significant summer flood. The return period of the measured peak discharge at the gauge Zeitz was of 10 years and at the gauge Kleindalzig of > 25 years (LHW, 2011). The model was further validated on the calendar year of 2008, another year of low monthly precipitation anomalies.

The calibration results on the summer flood of 2013 (see Fig. 4 below) are considered satisfactory, except for the headwater catchment of Weida. However, the model did not capture the steep peak (flashiness) of the rainfall-runoff response and the responsiveness of the baseflow. The validation results underline these conclusions and show an overestimation of direct discharge from snowmelt, as well as too high and retarded flood peaks. We will add this explanation in the revision of the paper.

Looking at a whole calendar year, the weekly discharge estimates show a positive bias in the direct discharge, interflow and baseflow simulations.

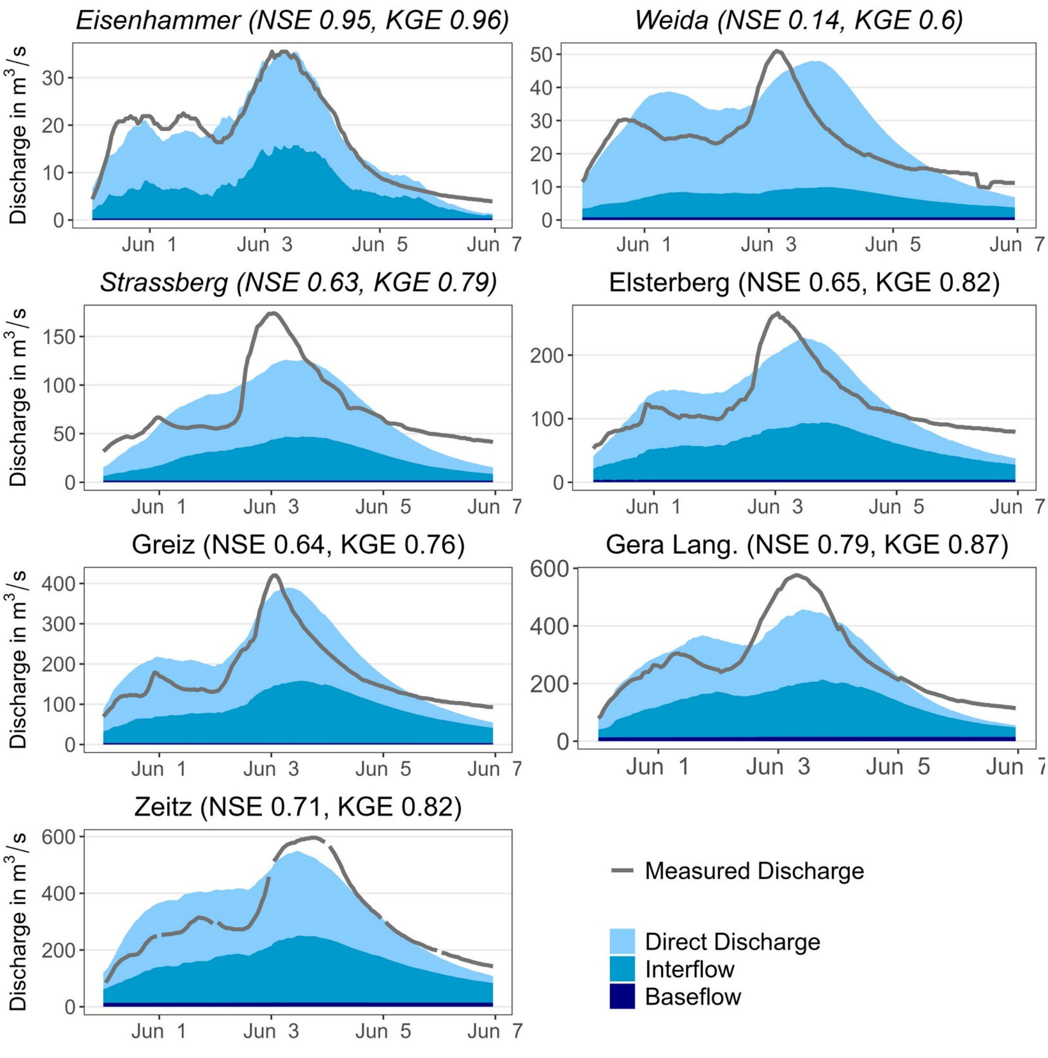

Fig. 4: Calibration results for the flood of 2013

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.

We used the processed RADOLAN data set offered via the ReKIS portal, which covers the German federal states of Saxony, Saxony-Anhalt and Thuringia (Körner, 2022). The native RADOLAN product builds the foundation but has further been gap-filled using the ground stations of the German Weather Service (DWD). Data from DWD's precipitation gauges is freely available and has been used for calibrating and validating our hydrological model. The same station network was employed by DWD in the adjustment of the radar measurements in the RADOLAN scheme (Winterrath et al., 2012). The choice of adjustment procedure is determined based on a performance test with a few control stations (see Winterrath et al., 2012). It is unknown whether any, respectively which of these control stations correspond precisely to the stations used in our study, but in principle the station values should be preserved. We will clarify this in the revision of the paper.

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.

The analysed climate model data was outputted by the ICON-CLM model in an hourly resolution and provided to users in the frame of the NUKLEUS (Actionable Local Climate Information for Germany) project.

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

Sturges' rule has established itself as the first choice in finding the optimal number of bins in a histogram (Scott, 2009). It should however indeed be noted that for large datasets it is prone to oversmoothing (Scott, 2009).

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

Fig. 5 shows the QQ-plots for catchment spatial averages of the Thiessen-interpolated hourly rainfall estimates from RADOLAN3km, resp. RADOLAN11km to the catchment spatial averages of the Thiessen-interpolated hourly rain gauge measurements, as computed in a preprocessing step by the hydrological model WaSiM.

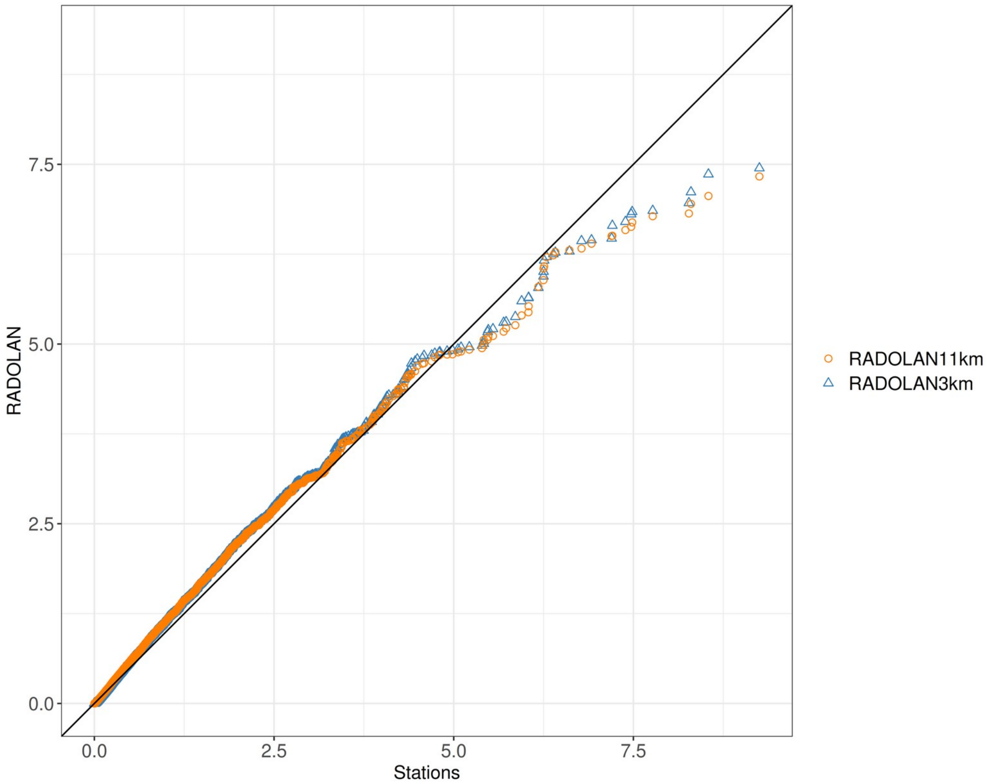

Fig. 5: QQ-plots for catchment spatial averages of the Thiessen-interpolated hourly rainfall estimates from RADOLAN3km, resp. RADOLAN11km to the catchment spatial averages of the Thiessen-interpolated hourly rain gauge measurements for the period 2005 to 2014

The Thiessen-interpolated RADOLAN products derive most strongly from the Thiessen-interpolated station measurements for moderate to high intensity rainfalls. We looked exemplarily at the effect this has on the simulation of the largest flood in the catchment during the study period (Fig. 6 below) by adding the calibration results into Fig. 12 of the original submission. The deviations of the discharge simulations driven with RADOLAN data from those driven with precipitation station data are found to be minor. We will include these explanations in the revised paper and add the corresponding figures to the appendix.

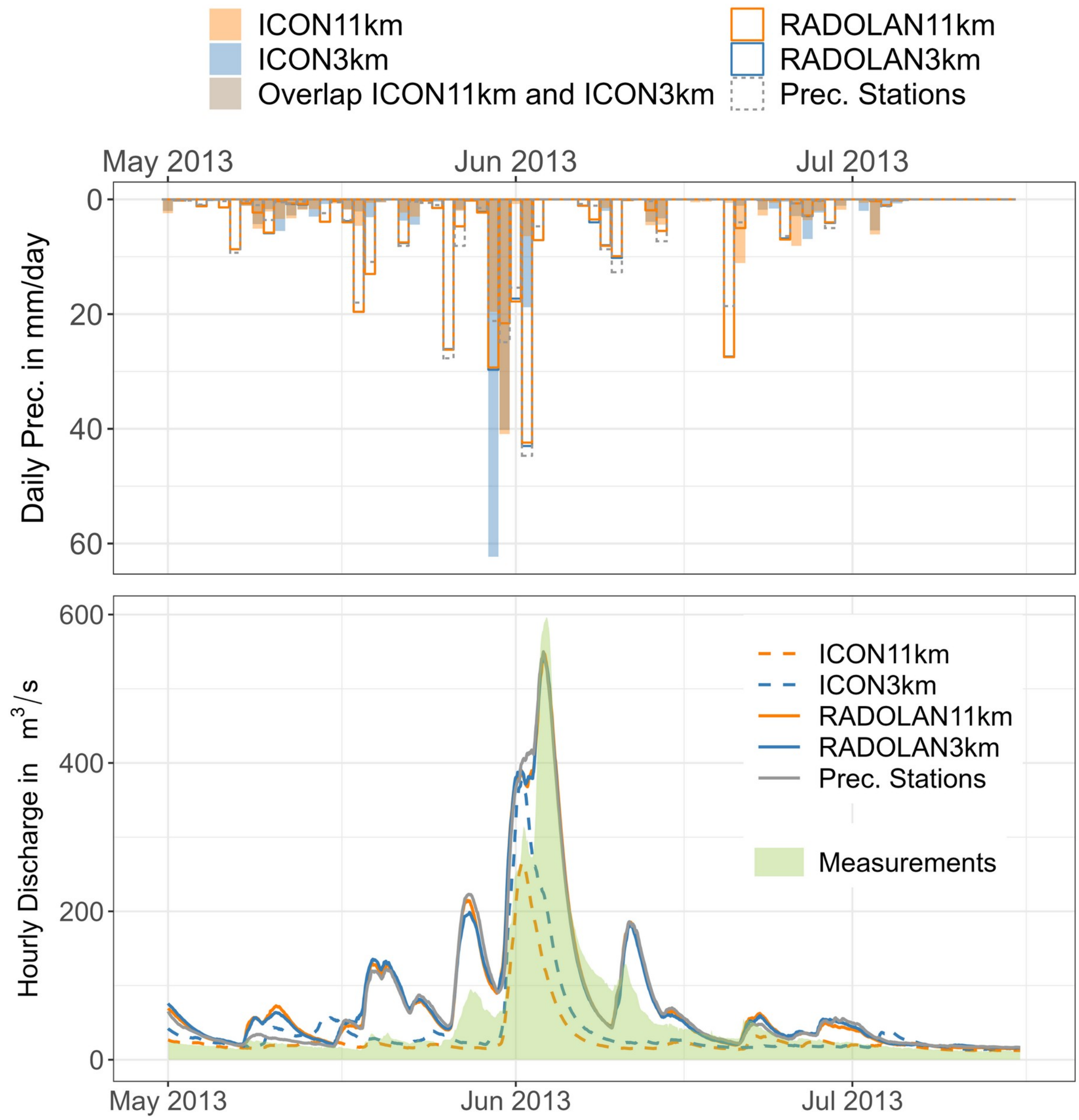

Fig. 6: Daily spatial average precipitation estimates over the catchment of Zeitz and those upstream for the period of the 2013 flood for ICON11km, ICON3km, RADOLAN11km, RADOLAN3km and the interpolated precipitation station measurements, Bottom: the resulting hydrographs (using hourly data) for the catchment of Zeitz together with the discharge measurements

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

Yes, that's right. We will clarify this in the revised paper.

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.

Thank you for pointing this out. We will change it.

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

Thank you. We will follow the suggestion.

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

> The German Weather Service commonly measures wind speed with the Ultrasonic Anemometer 2D, which has an accuracy of ± 0.1 m/s for wind speed below 5 m/s and of ± 1.5% for wind speeds higher than that (METEK, n.d.). We will clarify this in the manuscript.

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.

> We will clarify that potential evapotranspiration was calculated by the Penman-Monteith formula. This approach has established itself as the standard in hydrological modelling (Ndulue & Ranjan, 2021). A range of studies have analysed how uncertainties in meteorological variables shape the estimation of current and future potential evaporation and evapotranspiration using the Penman-Monteith method (e.g. Meyer et al., 1989; Kay & Davies, 2008; Lai et al., 2022). We will include this information in the revised paper. It shall be noted that while evapotranspiration greatly influences the long-term water balance, however for heavy rainfall events, as are the focus of this paper, evapotranspiration is of secondary importance given that soil saturation is quickly reached.

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)?

> We will take a closer look at the individual rainfall events and their associated runoff events, as identified using the HydRun tool (Tang & Carey, 2017). A particular focus will be placed on extreme intensity rainfall estimates, their occurrence and effects on runoff generation.

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

> The 99.5% percentile of hourly precipitation by RADOLAN3km is at 2.8 mm/h and by RADOLAN11km at 2.7 mm/h.

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.

> In the study region, floods occur primarily in summer as a result of high-intensity convective storms, as well as in winter due to prolonged rainfall, sometimes amplified by rapid snowmelt. CPRCMs are expected to offer little improvement for the representation of winter floods, as these are governed by large-scale synoptic weather systems (Strandberg & Lind, 2021), but CPRCMs have great potential for the simulation of local convective summer storms. The focus of our paper is therefore on summer convective rainfall events.

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.

We looked at the precipitation anomalies to the long-term mean to differentiate whether the climate models are able to depict months with (1) unusual wet conditions and (2) unusual dry conditions. We identified a difference in performance for the two cases, with ICON3km overestimating the first, but in contrast proving able to capture even deep negative precipitation anomalies.

For reference, the monthly precipitation sums averaged over the study area are shown in Fig. 7 for the ICON models and the RADOLAN data. In contrast, the precipitation anomalies of the ICON models (Fig. 9 in the paper) were computed only for cells overlaying rain gauges and expressed in relation to the long-term measurements of the latter.

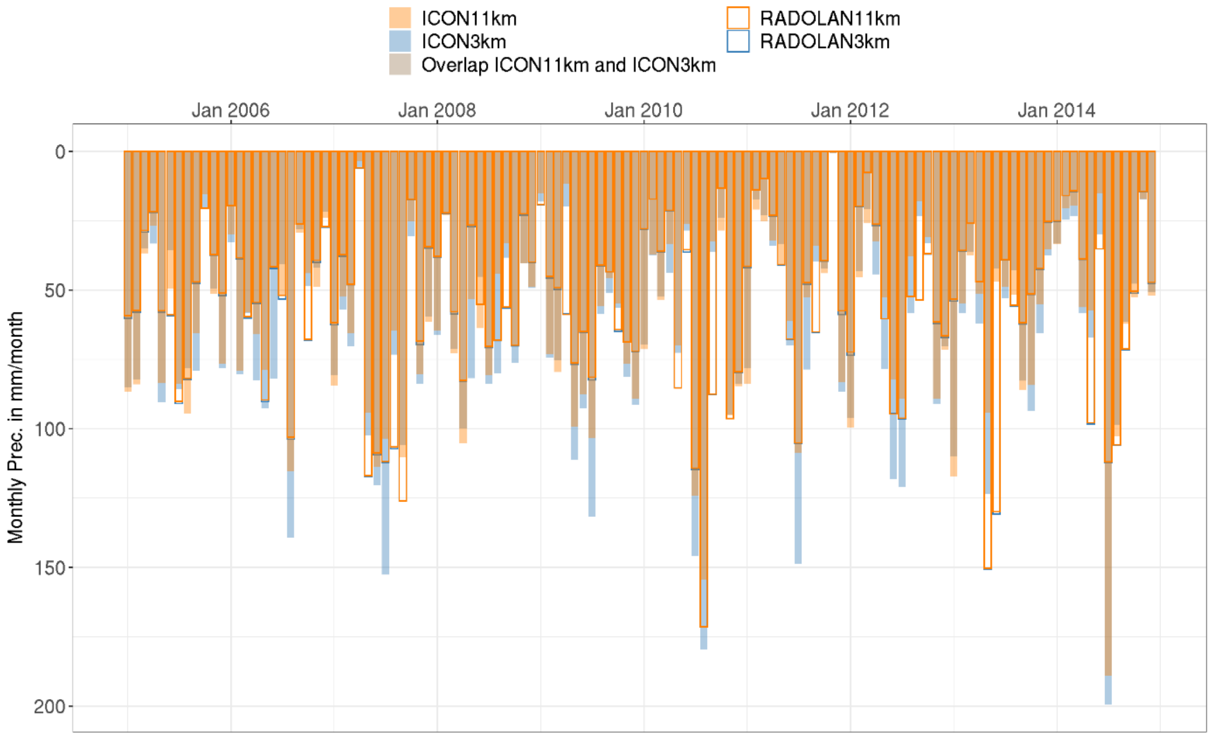

Fig. 7: Monthly precipitation sums averaged over the study area for ICON11km, ICON3km, RADOLAN11km, RADOLAN3km for the period of 2005 to 2014

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).

We used the distributed physically-based hydrological model WaSiM, which we ran with Thiessen-interpolated meteorological station data for calibration. Four main parameters were used for calibration, namely (1) the single reservoir recession constant for surface runoff, (2) the single reservoir recession constant for interflow, (3) the drainage density for interflow and (4) the fraction of surface runoff on snow melt. More information on the hydrological model is given by Schulla (2021).

A discussion of the calibration and validation results is given as an answer to the general comments.

Fig. 8 shows a scatter plot of total runoff (Q_{tot}, estimated from the measured discharge) to the sum of thiessen-interpolated measured total rainfall (R_{tot}) and 8-h antecedent rainfall (AR) over the catchment of Zeitz and those upstream. The data is widely scattered and doesn't allow to draw conclusions on a non-linear behaviour of flood generation.

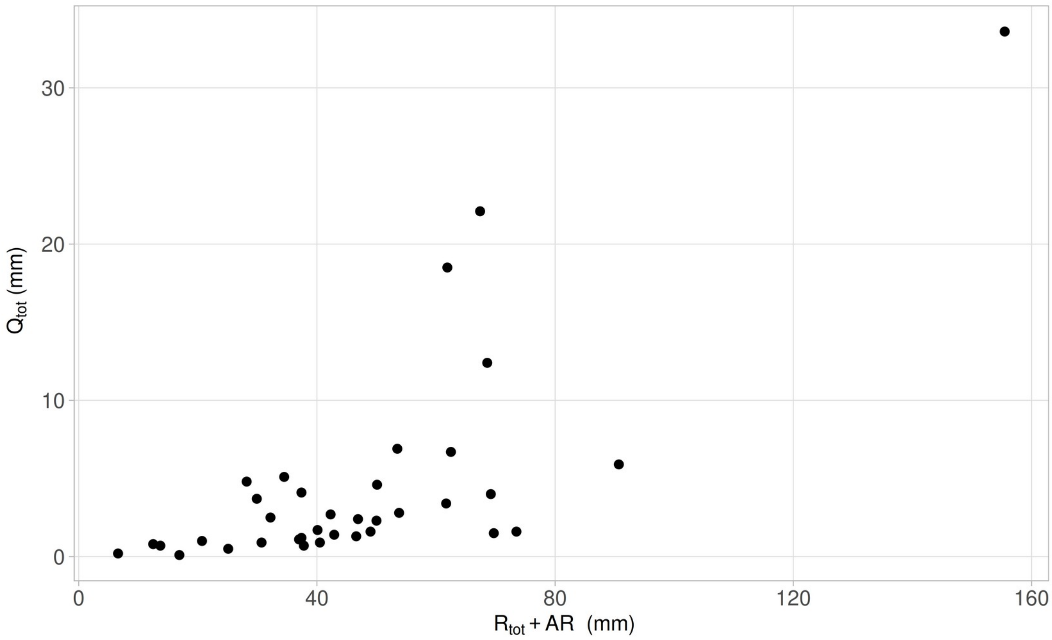

Fig. 8: Scatter plot of total runoff (Q_{tot}, estimated from the measured discharge with a peak threshold for runoff events of 0.015 mm/h) to the sum of thiessen-interpolated measured total rainfall (R_{tot}) and 8-h antecedent rainfall (AR) over the catchment of Zeitz and those upstream

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).

- Regarding the calibration and validation results, please refer to the answers given to the general comments above.
- Regarding the comparison of the discharge simulations by the hydrological model driven with RADOLAN3km and the interpolated rainfall gauge measurements, please refer to the answers given to the comment to P 7 / L. 136.
- Regarding the biases in ET0 and the impact on soil moisture simulation, please refer to the answers given to the general comments.
- We will conduct analyses on the simulated extreme intensity rainfall estimates, looking at their occurrence and effects on runoff generation under application of flood peak statistics.

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

Thank you for the comment. We will implement it.

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.

The catchment of Berga has an area of 195 km², Gera Langenberg of 303 km², Zeitz of 295 km² and Kleindalzig of 425 km². We will clarify this in the manuscript.

We computed the mean monthly sums of catchment precipitation and discharge estimates, and deduced the relative biases of ICON3km and ICON11km to RADOLAN3km and RADOLAN11km respectively (Fig. 9 below). In the paper we show how both ICON3km and ICON11km overestimate mean yearly precipitation sums. These findings are in line with the results of the mean monthly sums shown below. Only during the peak of summer do ICON3km and ICON11km show biases of markedly different extent, with higher positive biases by the CPRCM (ICON3km). The overestimation of summer precipitation translates into the discharge simulations. A direct comparison of the relative biases is hindered by the difference in reference and scale of the absolute values from which they have been derived.

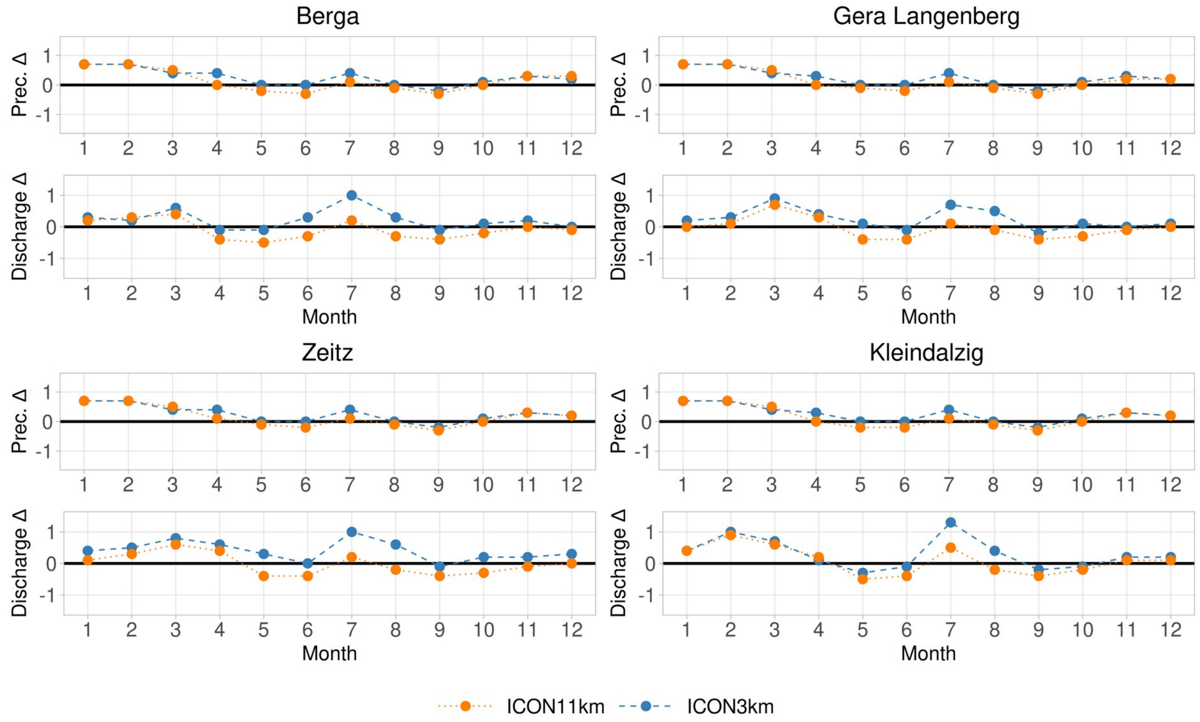

Fig. 9: Relative deviations of average monthly sums of catchment precipitation and discharge estimates for ICON3km and ICON11km to RADOLAN3km and RADOLAN11km for the four most downstream catchments on the main stem of the Weiße Elster river within the study area over the period of 2006 to 2014

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

> We will rewrite this sentence to: „Driving the hydrological model with ICON11km meteorological data was found to lead to an underestimation of median hourly discharge, but results in an overestimation when looking only at the 99.5th percentile of hourly discharge for six out of the seven catchments on the main stem."

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

> We will add this clarification in the manuscript.

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.

> This has been discussed previously as an answer to the comment to P 7 / L. 136.

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.

> Soil moisture has been studied and discussed as a response to the general comments.

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.

Fig. 10 expands the hyetograph-hydrograph-plot of the 2013 flood (Fig. 12 in the manuscript) by the soil moisture time series, as gained from continuous hydrological simulation. ICON11km climate input data leads to too low soil moisture, a finding in keeping with Fig. 1 of this document. As a consequence, the flood wave is greatly buffered by the soils, leading to a rise in soil moisture and a comparatively low flood peak. Driven with RADOLAN3km and RADOLAN11km, the soil moisture results suggest that the storm rained off over saturated soils; the peak rainfall does not lead to an increase of soil moisture, but to a pronounced runoff peak. While on average, soils in the hydrological model driven with ICON3km showed similar initial soil moisture values, however the rainfall likely fell over catchments with not yet fully exhausted storage capacities. As a consequence, average soil moisture rose further and the flood peak was attenuated. ICON3km was run over the vast Central European (CEU) domain, forced three-hourly at its boundaries by ECMWF-ERA5 reanalysis data. While the general events and patterns match, the climate model, as is widely known, is unlikely to perfectly match the actual geographic position of a convective storm and spatial offsets are to be expected.

We will replace Fig. 12 in the manuscript by the extended version including the soil moisture time series (Fig. 10 below) and discuss the results.

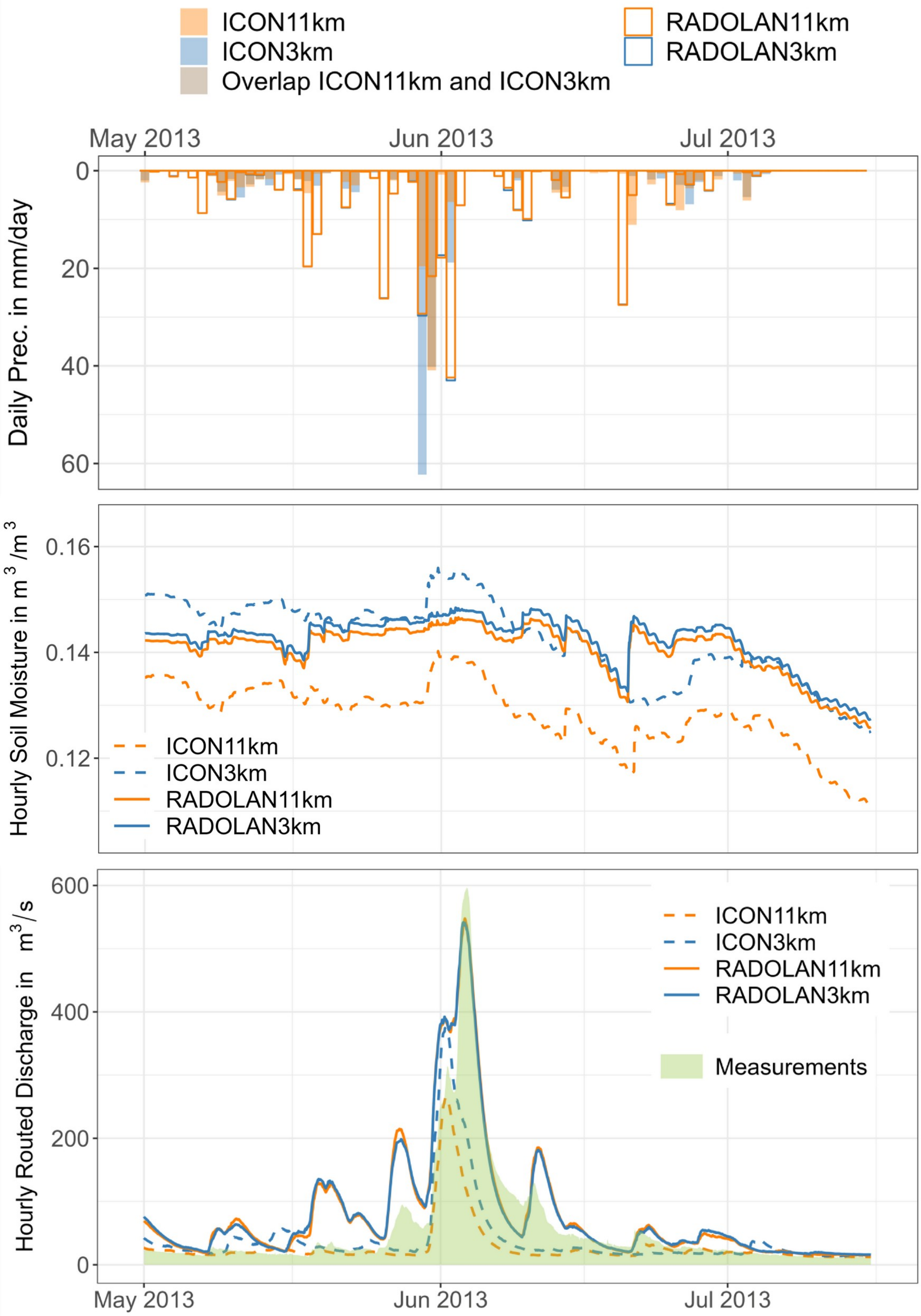

Fig. 10: Top: Daily spatial average precipitation estimates over the catchment of Zeitz and those upstream for the period of the 2013 flood for ICON11km, ICON3km, RADOLAN11km and RADOLAN3km, Centre: the respective hourly relative soil moisture spatial averages, Bottom: the resulting hydrographs (using hourly data) for the catchment of Zeitz together with the discharge measurements

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

> We will revise and extend the discussion with the new analyses.

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).

> In the revised manuscript, we will discuss the results on the implications on hydrology presented above as a reply to the general comments.

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?

> In the planned analyses of individual rainfall events and their associated runoff events, we will also look into flood statistics and potential shifts in return periods by varying modelling resolution. However, the short time series of 9 years only captured a limited number of floods, posing a strong limitation to this endeavour.

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

> In the case study of the 2013 flood event, ICON3km showed an overestimation of peak rainfall intensity, but the flood peak was underestimated. As outlined as a response to the comment to P. 19 & 20 / Fig. 12, the simulated rainfall cell was likely shifted in space and rained off over soils which were not yet full saturated, while in the RADOLAN3km-driven simulation, the rainfall hit saturated soils and was converted integrally into runoff.

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.

> Thank you for pointing us in this direction. We will add some comments in the revised paper on the flood statistics, under consideration of the literature, given that the time series used here is too short to estimate high return periods.

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)?

> The added value of process-based fully distributed hydrological models compared to conceptual lumped models for discharge simulations driven with convection-permitting regional climate models is a field of ongoing research. Poncet et al. (2024) used hourly convection-permitting regional climate model data to drive the process-based distributed hydrological model CREST and the conceptual lumped model GR5H. They found an

improvement in the simulation of flood intensity and frequency using the CPRCM compared to a RCM. They further studied the number of peaks over threshold and found CREST to overestimate, while GR5H offered better agreement with observations. These results show that conceptual lumped models remain valid, but more research is needed to draw firm conclusion, since deterministically better process representation, as in process-based models, goes in hand with reduced bias. A major bottleneck of process-based fully distributed models is however their high computational cost. We will add some of these points in the revision.

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.

Climate model data of high spatial and temporal resolution is needed for local hydrological impact studies in small catchments of mountainous or highly urban character (e.g. Schaller et al., 2020; Tamm et al., 2023). Here CPRCMs have the potential to provide a real improvement. It is therefore recommended to direct research to complex study sites where coarser RCMs with parametrised convection fall short. We will add text along these lines to the paper.

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?

High precipitation bias poses a strong limitation to the applicability of the model chain, however bias correction of climate model data of high spatial and temporal resolution, as needed for hydrological impact modelling, comes with major challenges (e.g. Haerter et al., 2011; Addor & Seibert, 2014). Bias correction furthermore relies on the assumption of stationary bias, making it questionable for climate projections (Huang, Krysanova & Hattermann, 2014). A correction of the climate model data is outside the scope of our paper.

Given that a number of studies have indeed identified added value in the use of CPRCMs (e.g. Rudd et al., 2020; Kay et al., 2022; Poncet et al. 2024), an ensemble approach would be recommended. Coordinated research efforts are needed and have gained momentum in recent years (Lucas-Picher et al., 2021). We will add a concluding remark on this.

References

Addor, N. and Seibert, J.: Bias correction for hydrological impact studies – beyond the daily perspective, Hydrol. Process., 28, 4823–4828, https://doi.org/10.1002/hyp.10238, 2014.

Haerter, J. O., Hagemann, S., Moseley, C., and Piani, C.: Climate model bias correction and the role of timescales, Hydrol. Earth Syst. Sci., 15, 1065–1079, https://doi.org/10.5194/hess-15-1065-2011, 2011.

Huang, S., Krysanova, V., and Hattermann, F. F.: Does bias correction increase reliability of flood projections under climate change? A case study of large rivers in Germany, Intl Journal of Climatology, 34, 3780–3800, https://doi.org/10.1002/joc.3945, 2014.

Kay, A.: Differences in hydrological impacts using regional climate model and nested convection-permitting model data, Climatic Change, 173, https://doi.org/10.1007/s10584-022-03405-z, 2022.

Kay, A. L. and Davies, H. N.: Calculating potential evaporation from climate model data: A source of uncertainty for hydrological climate change impacts, J. Hydrol., 358, 221–239, https://doi.org/10.1016/j.jhydrol.2008.06.005, 2008.

Körner, P.: Radarbasierte Niederschlagsdaten 2001 - 2020: Radarbasierte Niederschlagsdaten im Zeitraum 2001-2020 Beschaffung, Aufbereitung und Bereitstellung im ReKIS, https://publikationen.sachsen.de/bdb/artikel/39692/documents/60999, 2022.

Lai, C., Chen, X., Zhong, R., and Wang, Z.: Implication of climate variable selections on the uncertainty of reference crop evapotranspiration projections propagated from climate variables projections under climate change, Agric. Water Manag., 259, 107273, https://doi.org/10.1016/j.agwat.2021.107273, 2022.

LHW (Landesbetrieb für Hochwasserschutz und Wasserwirtschaft Sachsen-Anhalt): Bericht über das Hochwasser Januar 2011, https://klimawandel.geo.uni-halle.de/sites/default/files/abschlussbericht_2011.pdf, 2011.

LHW (Landesbetrieb für Hochwasserschutz und Wasserwirtschaft Sachsen-Anhalt): Bericht über das Hochwasser im Juni 2013 in Sachsen-Anhalt: Entstehung, Ablauf, Management und statistische Einordnung, last access: https://lhw.sachsen-anhalt.de/fileadmin/Bibliothek/Politik_und_Verwaltung/Landesbetriebe/LHW/neu_PDF/4.0/SB_Hochwasserschutz/Hochwasserbericht_2013.pdf, 2014.

Lucas-Picher, P., Argüeso, D., Brisson, E., Tramblay, Y., Berg, P., Lemonsu, A., Kotlarski, S., and Caillaud, C.: Convection-permitting modeling with regional climate models: Latest developments and next steps, WIREs Climate Change, 12, https://doi.org/10.1002/wcc.731, 2021.

METEK: 2D Ultrasonic Anemometer, https://www.th-friedrichs.de/en/products/wind/ultrasonic-anemometer/2d-ultrasonic-anemometer/, last access: 30 October 2023.

Meyer, S. J., Hubbard, K. G., and Wilhite, D. A.: Estimating potential evapotranspiration: the effect of random and systematic errors, Agric For Meteorol., 46, 285–296, https://doi.org/10.1016/0168-1923(89)90032-4, 1989.

Ndulue, E. and Ranjan, R. S.: Performance of the FAO Penman-Monteith equation under limiting conditions and fourteen reference evapotranspiration models in southern Manitoba, Theor Appl Climatol, 143, 1285–1298, https://doi.org/10.1007/s00704-020-03505-9, 2021.

Poncet, N., Lucas-Picher, P., Tramblay, Y., Thirel, G., Vergara, H., Gourley, J., and Alias, A.: Does a convection-permitting regional climate model bring new perspectives on the projection of Mediterranean floods?, Nat. Hazards Earth Syst. Sci., 24, 1163–1183, 2024.

Ross, C. A., Ali, G. A., Spence, C., and Courchesne, F.: Evaluating the Ubiquity of Thresholds in Rainfall-Runoff Response Across Contrasting Environments, Water Resour. Res., 57, https://doi.org/10.1029/2020WR027498, 2021.
Rudd, A. C., Kay, A. L., Wells, S. C., Aldridge, T., Cole, S. J., Kendon, E. J., and Stewart, E. J.: Investigating potential future changes in surface water flooding hazard and impact, Hydrol. Process., 34, 139–149, https://doi.org/10.1002/hyp.13572, 2020.
Schaller, N., Sillmann, J., Müller, M., Haarsma, R., Hazeleger, W., Hegdahl, T. J., Kelder, T., van den Oord, G., Weerts, A., and Whan, K.: The role of spatial and temporal model resolution in a flood event storyline approach in western Norway, Weather Clim. Extrem., 29, 100259, https://doi.org/10.1016/j.wace.2020.100259, 2020.
Schulla, J.: Model Description WaSiM: Water balance Simulation Model, 2021.
Scott, D. W.: Sturges' rule, WIREs Computational Stats, 1, 303–306, https://doi.org/10.1002/wics.35, 2009.
Strandberg, G. and Lind, P.: The importance of horizontal model resolution on simulated precipitation in Europe – from global to regional models, Weather Clim. Dynam., 2, 181–204, https://doi.org/10.5194/wcd-2-181-2021, 2021.
Tamm, O., Kokkonen, T., Warsta, L., Dubovik, M., and Koivusalo, H.: Modelling urban stormwater management changes using SWMM and convection-permitting climate simulations in cold areas, J. Hydrol., 622, 129656, https://doi.org/10.1016/j.jhydrol.2023.129656, 2023.
Tang, W. and Carey, S. K.: HydRun: A MATLAB toolbox for rainfall–runoff analysis, Hydrol. Process., 31, 2670–2682, https://doi.org/10.1002/hyp.11185, 2017.
Winterrath, T., Rosenow, W., and Weigl, E.: On the DWD quantitative precipitation analysis and nowcasting system for real-time application in German flood risk management, Weather Radar and Hydrology, 323-329, 2012.