

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

This paper compares the efficiency of a Convection-Permitting Model (CPM : ICON-CLM 2.6.4 at 3km) and its driving Regional Climate Model (RCM : ICON-CLM 2.6.4 at 11km) to simulate some climate variables and discharge when forcing WaSim distributed hydrological model. The study is conducted over a cluster of catchments of small and medium size in central eastern Germany. The evaluated climate variables are the one relevant for hydrological modelling such as surface temperature, relative humidity, wind speed, global radiation and precipitation. The CPM exhibits no clear added-value for the simulation of all the climate variables. For precipitation, especially extreme precipitation, the increase in model resolution switches from a negative bias with the RCM to a strong positive bias and sometimes unrealistic values with the CPM. This behavior leads to an overestimation of discharge over the studied catchments even if the strongest recorded flood of the period remains underestimated.

Due to the scarcity of hydroclimatic studies using high resolution regional climate models, this paper is welcomed in the community, even if it consists of a simple case study using only one CPM-RCM couple and one hydrological model on one particular cluster of catchments. The study adds further scientific content in the CPM bibliography by assessing a CPM potential benefits and understanding their transferability to the hydrology.

The article is interesting, well organized and globally well written. However, some important information is lacking and some important aspects need to be addressed to improve either the clarity of the message and the consistency of the paper to make it suitable for publication in HESS.

1 - According to the paper title, the aim of the study is the assessment of the added value of the CPM compared to the RCM on hydroclimatic simulations, meaning climate and related hydrological simulations. The evaluation of global radiation, wind speed and relative humidity, even if bringing interest, are too detailed for the real purpose of the paper. The evaluation results presented for those variables are discussed separately and are not connected in any way to the final output that is the hydrological simulations.

If keeping this level of details (annual cycles of global radiation, diurnal cycles of wind speed, relative humidity) I would expect to see :

- a) A quick comparison of Potential Evapotranspiration (PE) obtained from simulations and observations to assess the impact of these variables on hydrological model forcing. PE could be directly accessible from the hydrological model output. If not, you could compute it with the Penman-Monteith equation from the climate data. PE annual and diurnal cycles

would be at least as interesting and relevant for your study as diurnal cycles of wind speed and annual cycles of global radiation.

The monthly mean bias of hourly total evapotranspiration (ET), as calculated by the hydrological model driven with ICON3km and ICON11km in respect to when driven with observational data (with precipitation data from RADOLAN3km and RADOLAN11km respectively), is shown in Fig. 1. During summer, the biases of the climate models differ most strongly. In fact, the identified reduction of the negative summer bias in temperature and global radiation by ICON3km may contribute to the (slightly) higher ET estimates in summer. We will expand the paper by analyses of simulated evapotranspiration.

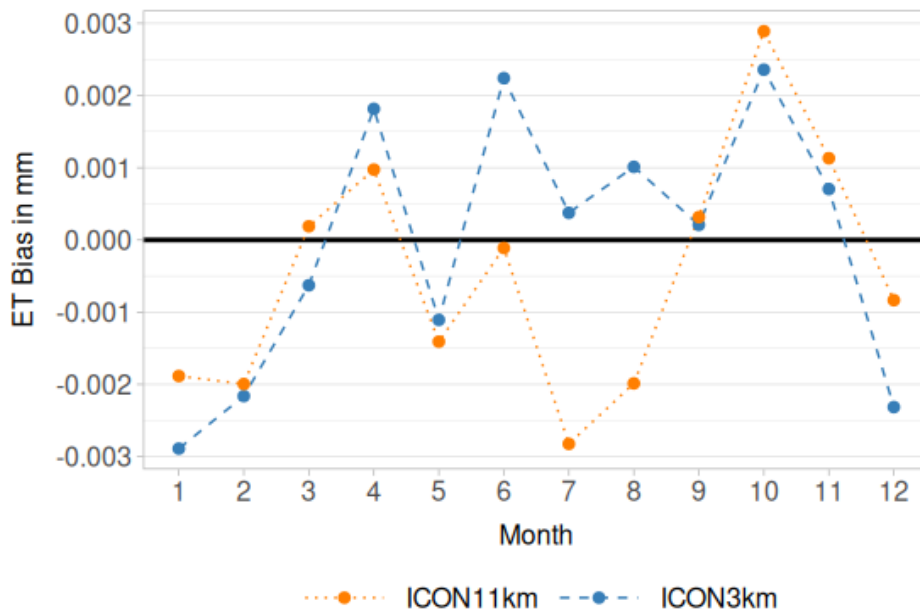


Fig. 1: Monthly mean bias of hourly total evapotranspiration, as calculated by the hydrological model driven with ICON3km and ICON11km for the period 2006 to 2014

- b) If possible, the impact on these variables on snowmelt in the model, and their contribution to discharge biases.

Over the nine-year simulation period, the hydrological model driven with meteorological data from the climate models overestimates the catchment spatial average nine-year snow melt sum (1538 mm for ICON11km compared to 1103 mm for RADOLAN11km, and 1571 mm for ICON3km compared to 1113 mm for RADOLAN3km). This is reflected in a positive discharge bias in early spring. The focus of this paper is however on convective summer storms, where CPRCMs and RCMs are likely to show the greatest differences.

2 - Some important results of the paper are discussed but never presented in tables or figures. On the contrary some others are, in my opinion, secondary if the analysis proposed in the previous comment is not produced. These intermediate results take up space at the detriment of the hydrological analysis which is the main point of the study.

- a) I308-311 and 426-428 : You present and discuss results concerning discharges above the 99.5th percentile but these results are never shown in tables or figures. Yet, one of the main reasons for CPM development and their use is to enhance impact studies for extreme events such as floods. Here, the results about extreme discharges are really important to show, either under the form of CDF comparison, QQ plots or even a table comparing biases for particular quantiles of discharge.

Boxplots of the 99.5th percentile of hourly discharge computed by the hydrological model driven with ICON3km, ICON11km, RADOLAN3km and RADOLAN11km respectively are displayed in Fig. 2 and will be added to the appendix of the revised paper.

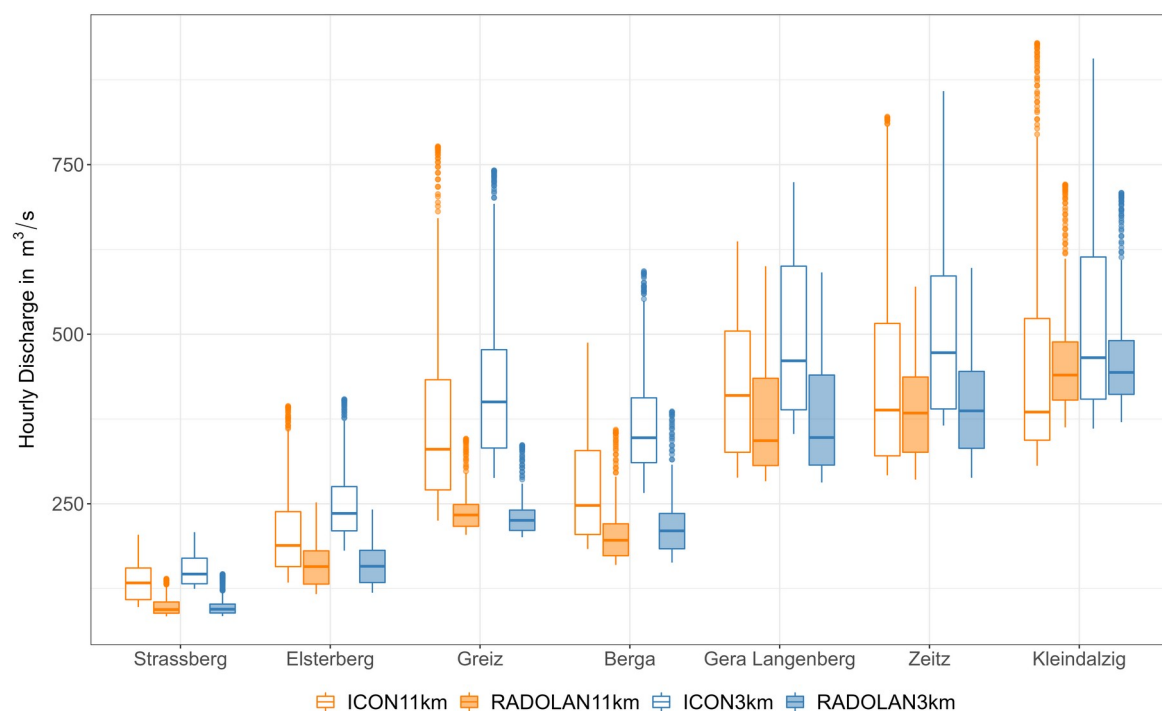
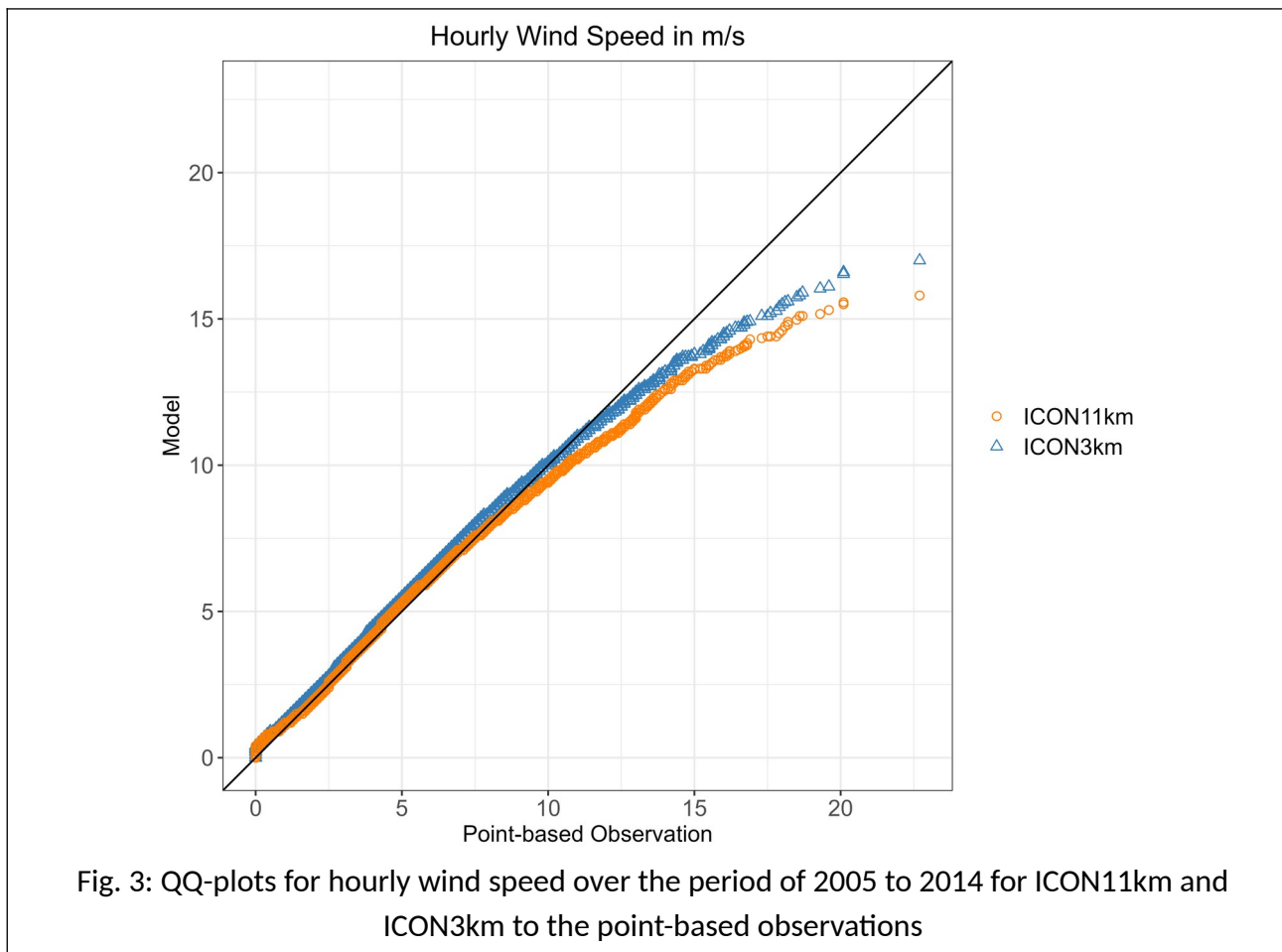


Fig. 2: Boxplots of the 99.5h percentiles of hourly discharge (period of 2006 to 2014) computed by the WaSiM hydrological model driven with meteorological data from ICON11km and ICON3km, as well as with adjusted radar data of respective equal resolution (RADOLAN11km and RADOLAN3km) for catchments of the main stem of the Weiße Elster river within the study area. Boxplots are built according to McGill et al. (1978).

- b) Too much attention is brought to wind speed frequency evaluation, for really tiny frequency biases that probably do not affect discharge modelling at all. In my opinion, seeing intensity biases on the whole distributions would be more informative. Figure 4 should be moved to supportive information and potentially replaced by a QQ plot of the type shown in figure 6.

QQ-plots for hourly wind speed over the period of 2005 to 2014 for ICON11km and ICON3km to the point-based observations are shown in Fig. 3. We will add the figure to the appendix of the manuscript.



- c) I221 - 225 : The main finding in Figure 6 (and later Figure 8) that stands out is the significant overestimation of precipitation for ICON3km, with unrealistic values ($> 100\text{mm}$ in 1 hour). Either here or in the discussion, these surprising results are not highlighted, nor discussed enough. What are the synoptic scale conditions leading to the occurrence of these values ? Where do they occur ? Are they numerical errors ? Are these findings in line with previous versions of ICON3km ?

We will conduct event analyses with a focus on the registered very high rainfall intensities to detect among others if they group together and to determine their seasonality.

- Furthermore, the median of precipitation above the 99.5th percentile is not representative of extreme values. An option is to conduct the same computation over a different quantiles (25th percentile, median, 90th percentile, 95th percentile and 99th percentile) of wet hourly rainfall ($> 0.1\text{mm/h}$ or $> 1\text{mm/h}$) distribution. The biases of ICON3km seem to exceed 200% for the highest quantiles of wet hourly precipitation.

The 99.5th percentile is by definition a reflection of the extreme values and we present this analysis therefore to discuss the heavy precipitation simulations.

3 - Figure 1 is a map showing the study area and its contrasted topography. In the rest of the text, the biases of climate variables are presented either aggregated on the whole area or without

distinction of geographical location. As a case study over a limited area, it would be very interesting to use the distributed nature of the data to visualize certain results in a map format. For example, a map of extreme precipitation biases for the two climate models, even in their native resolution, would be of great help to understand the spatial distribution of these biases and their transfer role in hydrological biases. It could as well help to understand the spatial patterns of hourly extreme rainfall values ($> 100\text{mm/h}$, visible in figure 6). Are these values simulated on high elevation areas (potentially related to biases of the orography forcing) or randomly over the study area (numerical errors) ?

This map could be included in the main body or in the supportive information depending on the relevance and of the article available space.

We employ a two-nest approach, in which the regional climate model ICON-CLM is run over the EURO-CORDEX domain at 11 km resolution driven by ECMWF-ERA5, and ICON-CLM in convection-permitting setup at 3 km resolution is embedded over the Central European (CEU) domain. With domains of such size and without nudging, high-intensity rainfall events are unlikely to be simulated at the precise location they have been observed. As such, a pixel-based computation of bias and its map-representation are not advised.

We will conduct event-based analyses and look at the most extreme storms simulated, mapping the climate model results side-by-side.

4 - The methodology section should be complemented by :

a) A small paragraph presenting the hydro-climatic conditions of the study area (climate type, hydrological regime...)

Thank you for the comment. We will implement it.

b) A new paragraph presenting hydrological model calibration and validation periods and choices. We understand later on in the results (I 269-270 and I275-276) that the authors chose to calibrate WaSim on the June 2013 flood and validate it over a winter flood, but no periods or dates are specified. This methodology remains therefore unclear and arises some questions:

- The analysis of the paper focuses on the whole range of discharges, but not particularly on extremes (except for one particular flood). Why have you chosen to calibrate the hydrological model only on extreme events?

The hydrological model was calibrated on the 2013 flood, as well as on the calendar year of 2012. We looked at weekly discharge sums, whereby the weeks of the year were defined according to ISO 8601.

- What are the exact period dates for calibration and validation?

The periods for calibration are:

- the 2013 flood event: 2013-05-31 00:00 to 2013-06-06 23:00
- the 2012 calendar year: 2012-01-02 00:00 to 2012-12-30 23:00

The periods for validation are:

- the 2011 flood event: 2011-01-07 00:00 to 2011-01-24 23:00
- the 2017 calendar year: 2007-12-31 00:00 to 2008-12-28 23:00

- Can certain hydrological processes, such as those governing long-term soil moisture, be excluded from calibration and therefore poorly simulated?

Long-term processes should receive consideration in calibration, therefore we have also calibrated our model on a complete calendar year.

- A classical modelling approach is to perform a split-sample test to calibrate and validate your hydrological model (KLEMEŠ, 1986). Given the short period of data, another recommended approach is to calibrate over the entire length of data (Arsenault et al., 2018; Shen et al., 2022). Have you tried one or the other approach? If so, it would be interesting to see a summary of the calibration/validation results. If not, could you justify your calibration choices?

We calibrated the hydrological model on the most severe summer flood observed in the catchment, as convective summer storms are particularly interesting in the discourse on the added value of CPRCMs for hydrological impact modelling. In fact, we showed that biases of ICON3km and ICON11km deviate most strongly during the summer months, likely since deep convection is parametrised in ICON11km and explicitly resolved in ICON3km.

We furthermore calibrated the model on the calendar year of 2012, a year of comparatively low monthly precipitation anomalies.

The paragraph should clarify this aspect of methodology, by justifying the authors' choices and answering the question above as best as possible.

Specific comments

- In my opinion, the passive form is overused and makes the text sometimes difficult to read. The expression "was/were found to" is too frequent (lines 154, 170, 175, 188, 192, 193, 204, 211, 250, 258, 260, 274, 276, 283, 297, 306, 332, 341, 346, 358, 366, 369, 378, 392, 396, 403) and could be changed in many cases by a more active form. For example : "ICON3km was found to overestimate" could be replaced by "ICON3km tends to...", "ICON3km shows an overestimation" or "ICON3km overestimates...", or "We notice an overestimation..."

Thank you for the suggestion. The phrasing has been chosen deliberately to underline that the conclusions were drawn from the one simulation run we did. The active form (e.g. "ICON3km overestimates") entails the risk of assumed generalisation, such as that ICON-CLM per se strongly overestimates rainfall intensities. We will revise the text for improved readability.

- Table 1 : The table 1 in Introduction listing existing hydroclimatic studies using CPM is greatly appreciated, but needs to be completed with the most recent work up to date (Dale and Shelton, 2025; Xie et al., 2025, maybe others...).

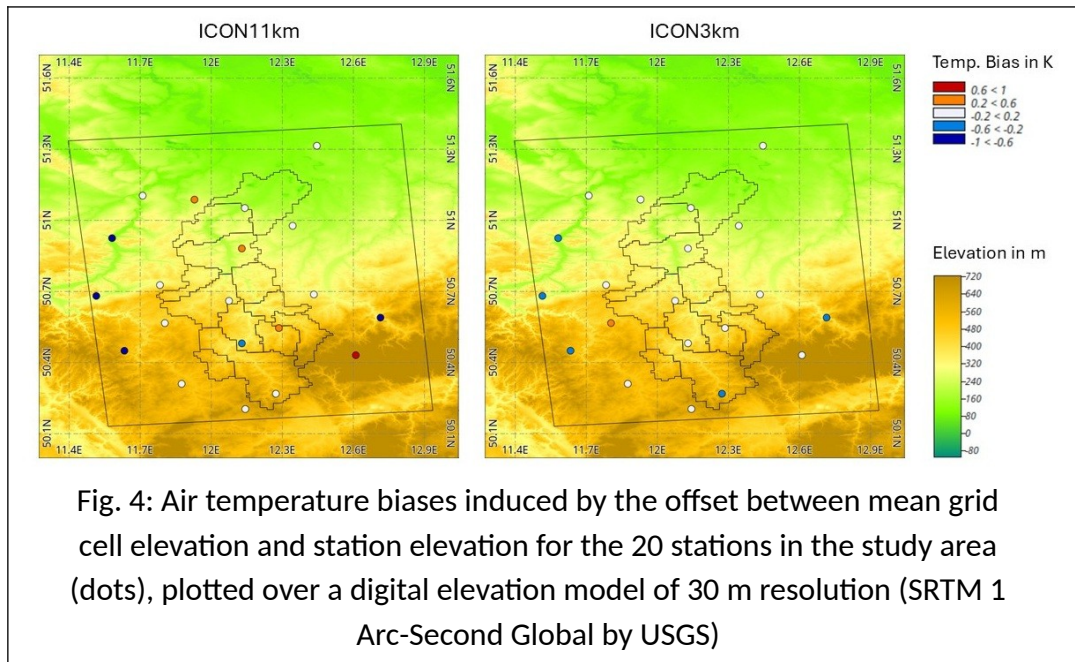
Thank you very much for pointing this out. We will update the literature review.

- l47: This sentence is misleading. The part of the sentence “come to offer substantial added value for flood simulation” implies that a consensus has been reached about the added-value of CPM on discharge modelling, which is not the case yet, thus justifying the interest of your study. The listed studies in Table 1 focus on various aspects of hydrological modelling, from low flows to floods, covering very different regions and climates and using diverse CPM and are too few to conclude. I would recommend the authors to take into account this comment and change this part of the sentence.

We will rewrite the respective sentence taking the comment into account.

- l81 : To compare point-based climate data to stations, are you applying a correction for the temperature vertical profile depending on grid mean elevation and station elevation ?

Thank you for the comment. We did not apply a correction for the temperature vertical profile. We have followed this up and taken a look at the effect that the mismatch of the station elevations and the elevations of the associated climate model grid cells has on the temperature estimations. As a digital elevation model, we used SRTM 1 Arc-Second Global provided by USGS at 30 m resolution. We upscaled in a two-step-process, once to 1 km keeping the native grid orientation, and then to 3 km, resp. 11 km while regriding to the ICON3km, resp. ICON11km grid. We employed the environmental lapse rate (0.65 °C/ 100 m) and found elevation-induced biases in the range of ± 1 K (see Fig. 4 below). A set of stations, particularly in the west of the study area, are located in river valleys and therefore at lower elevation than the climate model cells. These deviations in elevation appear as a cold bias by the climate models. Other stations are located at hill tops and therefore at higher elevation than the climate model cells, which appears as a warm bias by the climate models. We will discuss this in the revised paper.



- I111: If I understand this part of methodology well, precipitation observations have been upscaled to the climate model resolution, only with an aggregation of the initial grid and no regridding ? No interpolation has been done ?

The RADOLAN product has been retrieved at 1 km resolution. It has been upscaled and regridded to the ICON3km and ICON11km grid (becoming RADOLAN3km and RADOLAN11km).

- I113-115 : This part of the sentence has to be rephrased to improve clarity.

We will do that.

- I125 : Can you justify the differences of duration increments between ICON3km and ICON11km ?

Sliding a short moving window over an hourly time series of 10 years to extract the precipitation sums within each window position is a computationally intensive endeavour. For the 11-km-resolution data, we worked with window sizes of 2 h, 4 h to 24 h by 2-hour-increments. For the higher resolution data (3 km), we increased the increment size from 2 hours to 4 hours to stay within a reasonable computation time.

- In section 2.4 or 3.1, please explicitly detail the extent of the evaluation domain. Is it the rectangle displayed in figure 1.b or the mask of the catchments ?

Thank you for identifying this lack of clarity. The evaluation domain is the whole study area, i.e. the rectangle displayed in figure 1.b in the manuscript. We will clarify this in the revised paper.

- I146-147 and 148-149 : These results (temperature estimates and frequency distribution) are not linked to any figure or table. Please show them.

An overview of the temperature estimates is given by the boxplots in Fig. 5 below, whereas the frequency distributions are shown in Fig. 6.

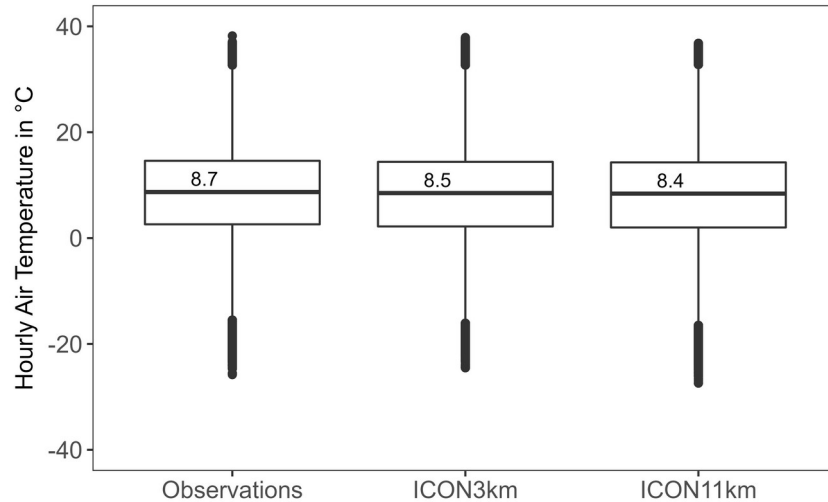


Fig. 5: Boxplots of hourly air temperature for the observations, ICON3km and ICON11km for the period of 2005 to 2014. The boxplots are built according to McGill et al. (1978).

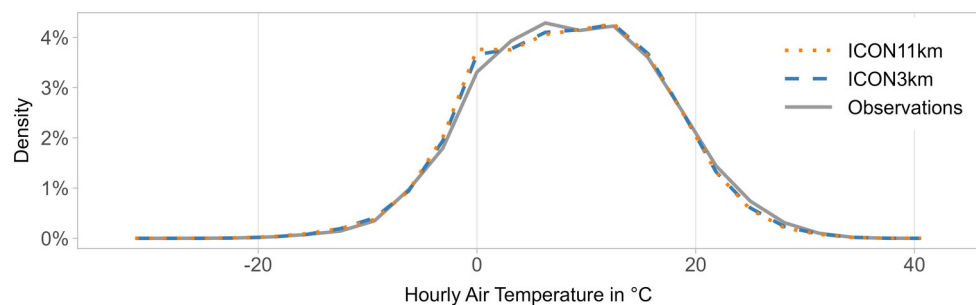


Fig. 6: Frequency polygons for hourly air temperature as observed, and calculated by ICON3km and ICON11km for the period of 2005 to 2014

- I154 : Could you be more specific about the sign of the error (negative) ?

Monthly mean bias of hourly air temperature by ICON11km and ICON3km was either zero or negative for all months of the year. We will add this to the text in the paper.

- I171-173 : This sentence should be rephrased to clarify the message : the passive voice and all the commas make it difficult to read.

We suggest the following rephrasing: „A similar picture can be drawn for the intensity estimates of global radiation. The climate models show a positive bias in the intensity of moderate daily mean global radiation. Extremes however were underestimated.“

- I176 : Can you convert this $2.5\text{J}/\text{cm}^2$ bias reduction in percent ? It is hard to realize if it is a big improvement or negligible.

A unit of absolute bias of daily mean global radiation translates to a higher relative bias on a cloudy day than on a sunny day. We therefore abstain from computing relative biases of daily mean global radiation. For

reference, station-averaged daily mean global radiation in July was of 77.6 J/cm^2 for the study period of 2005 to 2014. We will add this number to the paper.

- I194-195 : This sentence is hardly understandable. I propose this modification : “The largest difference in the average monthly errors between the two models over the year is only 1.3%, ...”. A figure could help visualize this aspect.

We will adopt the sentence as follows: „The largest absolute difference in the monthly means of hourly errors between the two models occurs in September and October and is of only 1.3 % relative humidity.“ Fig. 7 below supports the statement and will be added to the appendix.

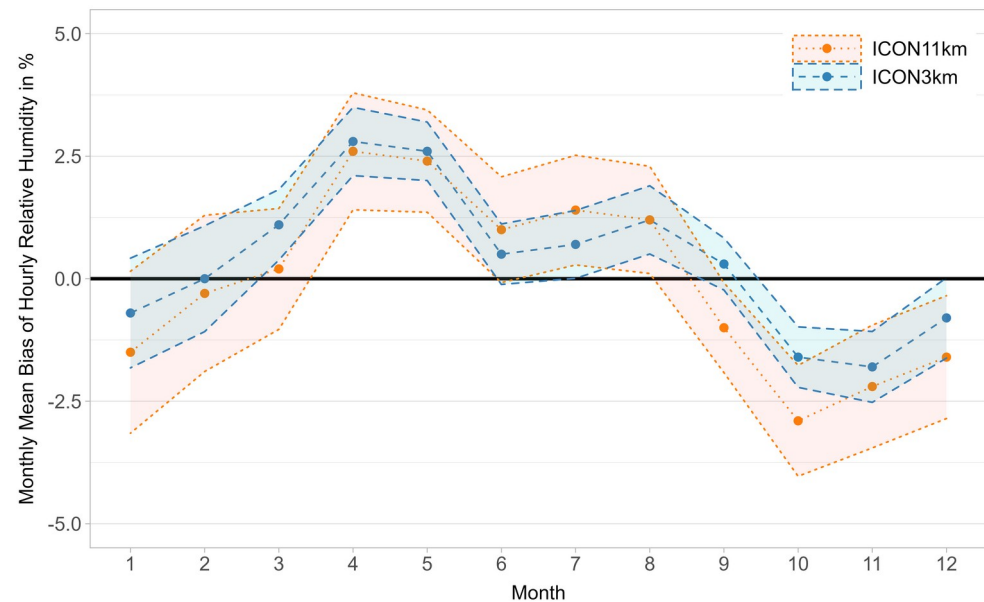


Fig. 7: Monthly mean bias of hourly relative humidity of ICON3km and ICON11km during the period 2005 to 2014, as well as the 95%-confidence intervals from the station means

- I201-202 : What is the measurement uncertainty for relative humidity ?

The measurement uncertainty for relative humidity is stated in line 186 of the paper. The employed sensor HMP45D comes with a measurement uncertainty of $\pm 2 \%$ (Kyrouac & Theisen, 2017).

- L237 : change “In keeping is” by “We notice”

Here we juxtapose the conclusions drawn from the depth-duration-frequency curves (with durations of up to 25 hours) to the yearly precipitation totals. The overestimation of high precipitation intensities and their frequencies by ICON3km is reflected both in the depth-duration-frequency curves and the yearly precipitation totals. As such, the results are in keeping but cannot be derived from each other, wherefore we chose to introduce the yearly precipitation totals using „in keeping“.

- L246: A presentation and an analysis of the diurnal cycle of precipitation would be interesting, considering the task is done for global radiation and wind speed. In addition, it could help to understand the representation of summer convective precipitation by the model and the biases shown for July in Fig S4.. Depending on the length of the paper, this could fit in Supportive Information.

We studied the diurnal variation of median hourly precipitation during wet days and did not find a clear signal in the climate model data. In fact, Ban et al. (2021) who show the diurnal cycles for over 20 simulations at 3 km resolution and 12 km resolution respectively over Switzerland, France and Italy, also did not see a noticeable diurnal cycle of mean precipitation over France in summer for some of the climate models. Their ensemble has a large intermodal spread, but in the mean does however show an improvement by the CPRCMs.

- L254 - 256 : This sentence should be rephrased. Here is a suggestion : “Additionally, the finer resolution of ICON3km improves the delineation of heavy rainfall events, preventing runoff generation and routing outside of the concerned catchments, as it occurs in RCM”

Thank you for the suggestion, which however does not reflect what the initial sentence intended to convey. We can suggest the following alternative phrasing: „Thereby they offer great advantage for hydrological impact modelling. In fact, coarse climate model cells overlapping the catchment boundaries bear the risk of having their precipitation estimates risen by heavy rainfall that occurs just outside the catchment. As these climate model cells stretch into the catchment, their high precipitation estimates will be attributed to the catchment, when in fact the storm occurred slightly outside of the catchment.“

- I259 : I suggest to cut the sentence after ICON3km and to start after it with : “This is also reflected...”

We will follow the suggestion.

- I276-277 : Could you share an example of this behavior ?

The validation results are a representative example of the model's ability to simulate snow melt floods. Similar behaviour can be seen for other snow melt induced floods in the time series. Flood peaks are computed as too high and retarded. The hydrographs are depicted as too flashy, while the baseflow is underestimated. We validated our model on a snow melt flood, as no other summer flood of comparable return period is present in the time series. The focus of the paper is however on heavy convective rainfall events.

- I278 : “were too high” → “underestimates”

We can adjust the wording and substitute „the model results were too high (...)“ with „the models **overestimate** (...)“

- I291 : The sentence of fig 10 legend is too long. Could you separate it into different parts ?

Our boxplots are built according to McGill et al. (1978). We will indicate this in the methodology section and leave out the specifications in the figure captions.

- I304 : What do you call the “full range of ICON11km meteorological data” ?

We will rewrite the corresponding sentence as: „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.“

- I306 : Why are you considering only catchments relative to the main stem ?

For all catchments in the study area, the climate models show a similar annual course of bias. Given the similarities, we opted for a 2x2-plot-matrix and limited visualisation to the four most downstream catchments on the main stem. It should be noted that bias on upstream catchments is reflected in downstream catchments.

- I314 : In methodology, it is stated that the calibration was performed on the July 2013 flood. How did you calibrate the hydrological model on other catchments if discharge measurements are not available ?

We calibrated the models using the discharge measurements at Eisenhammer, Weida, Strassberg, Elsterberg, Greiz, Gera Langenberg and Zeitz. The catchments of Kleindalzig, Berga, Dröda, Pöhl and Mylau are ungauged. However, with the sole exception of Kleindalzig, each of these catchments feeds into gauged downstream catchments and their effects are thereby seen in the downstream discharge measurements.

- I336-340 : This discussion is interesting. This behavior could have been checked easily in your study by looking if the temperature biases are more important on Tmin than Tmax.

We have followed this up by defining Tmin as temperature measurements below the 0.5th percentile and Tmax as temperature measurements above the 99.5th percentile. For the corresponding hourly time steps, we identified higher bias by the climate models for Tmin (ICON3km: 3.3 ± 4.2 K, ICON11km: 3.0 ± 4.3 K) than for Tmax (ICON3km: -1.2 ± 2.2 K, ICON11km: -1.3 ± 2.1 K). However, the biases on Tmin are positive and therefore not a reflection of the relatively strong negative biases we saw for winter when looking at the monthly means of hourly biases across the whole distribution. We attributed these

relatively high negative monthly mean biases to higher albedo and shorter roughness lengths on cropland and pasture in winter (Lind et al., 2020). However, extremely cold days are likely to be windstill and the presence of snow is not a given. As such, we are likely to see these effects on the monthly means of hourly biases and not on Tmin.

- I342-344 : The source of summer temperature improvement can be checked in your study by looking at the improvement of diurnal cycle of precipitation. Have you taken a look at it ?

We looked at the diurnal cycle of precipitation. Please refer to our answer to the comment to L246.

- I352-353 : Wrong citation yea. It is Keller et al, 2016.

Thank you for pointing this out.

- Fig 4 and lines 366-367 in discussion : The underestimation of extremes is not clear. There is a very slight underrepresentation of very light winds (can we consider them extremes ?) but no visible results on strong winds. To state this, you should represent and analyze a QQ-plot of ICON3km and ICON11km against observations, as advised in Major comment 2.b, and change these lines.

The QQ-plots for hourly wind speed for ICON11km and ICON3km to the point-based observations are presented in Fig. 3 above. The climate models show an underestimation of the intensity of high wind speeds. The figure will be added to the appendix of the revised paper.

- I380-381 : This sentence should be rephrased for clarity.

We suggest alternatively: „These results agree with the consensus from literature (Lucas-Picher et al., 2021), e.g. with Ban et al. (2021) and Adinolfi et al. (2021). In their study over the southern United Kingdom, Kendon et al. (2012) showed an RCM of 12 km resolution to compute heavy rainfall as not intense enough, but as too persistent and widespread.“

- I403: I advice the authors to rephrase the sentence to something like : “ICON3km does not improve the simulation of monthly precipitation, exhibiting a negative bias as ICON11km”

ICON3km exhibits a **positive** bias, just as does ICON11km. Here we discuss negative monthly precipitation anomalies, i.e. months during which the observed precipitation sums were below the long-term mean. We will clarify what we mean with negative monthly precipitation anomalies.

- I414-415 and 425-426: You have to be careful in your statement here. The different studies are not analyzing the same aspects of hydrology. Some did for the whole discharge range and others only for floods. Please modify these statements accordingly.

Thank you for pointing this out. We will refine our statement.

- I436-437 : Complete the sentence here to remind that the study is a case study and the results are valid over these particular catchments with the specific used climate models.

We will clarify this in the manuscript.

- Conclusion : a sentence should be added to contrast between the results and what is written line 47. Your study is not in line with recent studies and brings an interesting result : CPM does not systematically perform better than RCM for hydrological simulation because of some important biases on climate variables (here extreme precipitation). However, these results are consistent with some aspects of Xie et al. (2025) study.

We will follow this recommendation.

- I447 : You should split the sentence at the comma and start again with “The study is conducted exemplarily...”

Thank you, we will do so.

- I461 : From which study are you concluding of a “particular potential” ? This paper cannot conclude on a potential of ICON-CLM given the strong biases and the absence of added-value of using this CPM for hydrological simulations. I suggest you end your conclusion on an open note on the efforts put in place to correct some biases of the CPM (bias-correction, development, better understanding on extreme precipitation biases of ICON3km) to enhance CPM simulation and impact studies.

We will extend our concluding remarks taking these comments into account.

References

- Ban, N., Caillaud, C., Coppola, E., Pichelli, E., Sobolowski, S., Adinolfi, M., Ahrens, B., Alias, A., Anders, I., and Bastin, S.: The first multi-model ensemble of regional climate simulations at kilometer-scale resolution, part I: evaluation of precipitation, *Clim Dyn*, 57, 275–302, <https://doi.org/10.1007/s00382-021-05708-w>, 2021.
- Kyröuac, J. and Theisen, A.: Biases of the MET Temperature and Relative Humidity Sensor (HMP45) Report, 2017.
- Lind, P., Belušić, D., Christensen, O. B., Dobler, A., Kjellström, E., Landgren, O., Lindstedt, D., Matte, D., Pedersen, R. A., and Toivonen, E.: Benefits and added value of convection-permitting climate modeling over Fenno-Scandinavia, *Clim Dyn*, 55, 1893–1912, <https://doi.org/10.1007/s00382-020-05359-3>, 2020.
- McGill, R., Tukey, J. W., and Larsen, W. A.: Variations of Box Plots, *Am. Stat.*, 32, 12, <https://doi.org/10.2307/2683468>, 1978.