Author's response (AR), changes marked red

Response to reviewer 1

Reviewer #1 general comments:

RC: But the bias corrected 50km simulation achieves, from my point of view, at least as good results as the uncorrected 3km run (figure 10, 11, 13, 14). This raises the question, whether a computationally expensive downscaling to 3km is necessary for a statistical consideration of floods or if bias correction of coarse data is sufficient? Based on the presented results, I would suggest that the improvements (if existing) of a downscaling to 3km do not justify its additional costs. I would recommend to put this question as central statement of the paper and thus, a major revision is needed.

AR: We agree that results of the model chain using the coarser RCM together with error correction procedure look good. However, as the uncorrected RCM results show, seasonality is wrong in the coarser models, which indicates a lack in capturing the main atmospheric mechanisms for flood generation. From the point of applying a climate impact model chain, bias correction should either (1) only correct (small) biases, i.e. systematic errors, and not compensate errors in process description in order to prevent from the 'model is right for the wrong reasons' case (Klemes 1986) or (2) make use of process-informed approaches (e.g. Maraun et al., 2017) that are currently discussed in the climate modelling communities but are far from being established. Moreover, as it was recently shown by Blöschl et al. (2017), shifts in the seasonality are the only consistent large-scale climate change signal regarding floods identified so far. Our bias correction is largely compensating the improper representation of seasonality in the coarser models, however this is done in a statistical manner and we cannot exclude that we are still doing the 'right for the wrong reasons'. On the other hand, convection-permitting models gain from high resolutions and numerically resolved deep convective processes. This represents a fundamental change in the modelling technique which can have a substantial impact on projected climate change effects. For instance, Kendon et al. (2014) found significant increases in summertime precipitation in convection-permitting climate simulations in UK while the coarser resolved counterpart does not show any significant change. Ban et al. (2015) and Berthou et al. (2018) found similar results for short-term extreme precipitation events in the Alpine region and in the Mediterranean. However, since such simulations are relatively new, their benefits and shortcomings in climate applications are largely unknown, especially their potential in flood-modelling has not been explored. Nevertheless in the conclusion, we will reduce the strength of the requirement to use a 3 km model, since it also hinges on the partly wrong coarser models as its driving data and biases propagate along the downscaling chain (e.g. Addor et al., 2016). But the first results are promising. We will also include a statement in the conclusions that so far the coarser models could be used for climate impact studies in larger catchments for rough estimations, but they should not be taken for granted regarding local/regional flood change. We agree that so far, there is a trade-off in the additional costs of a 3 km simulation and the postulated (small scale) process description as long as the physical representation of such small scale processes can be substituted by statistical ones. We will include this trade-off question into the introduction, discussion of the simulation runs and conclusions. We will therefore extend the synthesis chapter by a comprehensive discussion. It was not the focus of the study; that would also include cost-benefit analyses, also monetarily. We believe/hope that computational infrastructure and efficiency will further improve to reduce these costs.

We have decided to add most of the discussion points (and references) into the chapters and have rewritten the conclusions. Additional analyses particularly focused on the representation of event rainfall in the RCMs and the influence of the bias correction. We tried to understand the good performance of CCLM 3km in small catchments on the one hand and the bad performance after bias correction. One figure has been added into chapter 4, and three figures have been added into chapter 6.2. Also the abstract has been changed corresponding to the new conclusions. The following references have been added.

Addor, N., Rohrer, M., Furrer, R. and Seibert, J.: Propagation of biases in climate models from the synoptic to the regional scale: Implications for bias adjustment, J. Geophys. Res.-Atmospheres, 121(5), 2075–2089, doi:10.1002/2015JD024040, 2016.

Ban, N., Schmidli, J. and Schaer, C.: Heavy precipitation in a changing climate: Does short-term summer precipitation increase faster?, Geophys. Res. Lett., 42(4), 1165–1172, doi:10.1002/2014GL062588, 2015.

Berthou, S., Kendon, E. J., Chan, S. C., Ban, N., Leutwyler, D., Schär, C. and Fosser, G.: Pan-European climate at convection-permitting scale: a model intercomparison study, Clim. Dyn., doi:10.1007/s00382-018-4114-6, 2018.

Blöschl et al.: Changing climate shifts timing of European floods. Science, 357 (2017), 6351; 588 – 590, 2017.

Kendon, E. J., Roberts, N. M., Fowler, H. J., Roberts, M. J., Chan, S. C. and Senior, C. A.: Heavier summer downpours with climate change revealed by weather forecast resolution model, Nat. Clim. Change, 4(7), 570–576, doi:10.1038/NCLIMATE2258, 2014.

Klemeš, V.: Operational testing of hydrological simulation models. Hydrological Sciences Journal, 31(1), 13-24, 1986.

Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J. M., Hagemann, S., Richter, I., Soares, P. M. M., Hall, A. and Mearns, L. O.: Towards process-informed bias correction of climate change simulations, Nat. Clim. Change, 7(11), 664–773, doi:10.1038/nclimate3418, 2017.

RC: […] floods are only well represented in CCLM and not in WRF (figure 13, 14). This highlights the relevance of an adjusted RCM for each research area and should be mentioned and discussed more prominently.

AR: We agree, CCLM and WRF show different behaviour depending on resolution and research area which asks for a bias correction. We will include a more rigorous analysis of the effects and the applicability of our bias correction technique in flood modelling attempts. Especially, since the bias correction method does not affect the frequency of precipitation it is rather unclear at the current stage how the statistical correction of precipitation intensities affect flood events that rely on a correct representation of the precipitation sequence and their occurrence in a climatological sense. See above

Reviewer #1 specific comments:

RC: page 1, line 16: I would not say "ensemble" in this context, since the simulations are not really used as an ensemble. "Model chain" would be more appropriate.

AR: We agree and will change this. Done

RC: page 1, line 21: I would use the term "coupling time step" to avoid confusion with the model time step.

AR: We agree and will change this. Done

RC: page 6, line 9-11: The example is difficult to understand and should be rewritten. In general, the method of the bias correction should be described in more detail.

AR: Since the new error correction method has been published recently, we didn't include a comprehensive description. But we will do this for a better understanding. Done

RC: page 7, line 1: Why does KAMPUS use a temperature threshold to calculate the snow accumulation out of precipitation, instead of using directly the simulated snow from the RCM? CCLM 4.8_clm17

AR: For consistency with the calibration we will stick to this simple model. This is widely used in flood forecasting to avoid additional uncertainties introduced by the use of highly variable climate variables from weather/climate models and therefore increase robustness of the models. Same with evapotranspiration.

RC: [..] By correcting the cold bias in the CCLM results, this overestimated snow accumulation may be reduced, potentially explaining the improved seasonality in the bias corrected 50km simulation.

AR: Yes, this is captured in the bias correction by correcting air temperature. However, the shift in runoff seasonality (overestimation of runoff in spring and underestimation in summer) is the consequence of the (same) shift in precipitation. We will add the precipitation distribution from the different RCMs over a year (monthly basis) in the suppl. material. Done

RC: page 8, figure 3: What are the red areas?

AR: The figure will be redrawn. Layout is misunderstanding. The blue colour is used to denote nested catchments within the larger ones (in red). Done

RC: page 10, figure 4: Why are you showing the average January precipitation amounts during night to highlight the added value of increasing model resolution? This is not the time frame in which I would expect the highest benefit from high resolution simulations (especially convection-permitting), but rather for summer (afternoon) precipitation.

AR: Figure 4 is only an example. We decided to remove it, since it does not contain any additional information needed for the explanation in the text. Done

RC: page10, figure 5: The figure shows that the added value of an increased resolution is mainly caused by an improved diurnal cycle of precipitation. I would recommend to mention this more prominent, since this is very important for a realistic description of floods in smaller catchments.

AR: We agree. This will be mentioned together with a better explanation of the SDM method.

RC: page 11, line 7: Please add a reference for NSE.

AR: Will be done. Done

RC: page 12, figure 6: calibration and validation results should be drawn in different colors. In this way, it's difficult to assess the quality of the validation results.

AR: Will be done. Done

RC: page 24, conclusions: see above the general comments

AR: A comprehensive discussion section will be included (see above). Done

Response to reviewer 2

Reviewer #2 general comments:

RC: Pg. 13, lines 8-11: In Figure 7 the improvement in using CCLM at 0.03° is very clear for the smallest catchment. But for the catchment with 119 km2, 0.11° obtains the best agreement, and even 0.70° seems closer to observations than 0.03° (except for the longest return period). In other catchments, coarser resolutions are also closer to observations. Stating that "simulations with coarser RCM data already yield reasonable results" is somewhat insufficient. Because increasing the resolution (to convective permitting) seems to degrade the flood frequency simulation in some cases (e.g. Voistsberg/U. Kanaish except for the highest return period, or Fluttendorf/Gnasb. compared to 0.11° or even 0.70°; Tillmitsch/Lassnitz and Leibniz./Sulm at intermediate return periods). This requires a more careful discussion. The presented results imply the need for a priori knowledge of the best resolution for flood frequency simulation in each catchment.

AR: We will add a comprehensive discussion with reduced strength regarding the need of CPS and the benefit of the bias correction. Improvement of CPS is not evident in every case regarding flood frequency, but it is evident for seasonality (see response to general comments of reviewer #1). Done, see above.

RC: The improvement is even less clear when using WRF, which is given in supplementary material. On this matter, the choice of presenting WRF results as supplementary material is not clear to me, and I have some concerns about it. It is stated in the abstract that the manuscript is discussing two RCMs (and no further distinction is made between them is made). This is again repeated in the last paragraph of the introduction. The fact that the added value of convective-permitting resolution in WRF is lower, and often non-existent for both flood frequency and seasonality seems like a main result, given the aim of the proposed investigation.

AR: We focused on the CCLM results for explaining the evaluation procedure in order to avoid a too long paper. Also, during the study starting in 2013 we had problems with the WRF simulation. For a long time it was not clear if we receive results that can be interpreted like the CCLM results (crash, bug, etc., see specific response to reviewer #3, p.10). We will add a deeper discussion with possible reasons why the WRF 3km more or less fails in representing floods (processes, nesting, etc.). Results of both model types are compared within a deeper discussion showing no systematic behaviour.

RC: Inspection of Figures 13 and 14 also raises major questions about the value of convective permitting resolution and bias-correction. The best resolution and whether bias correction improves the results seems to vary significantly between different catchments.

Then in the conclusions it is stated that: "Flood frequency and seasonality is represented well in all catchments. However, the 3km grid size is essential for catchments smaller than 200 km2. This seems like an overstatement. For Fluttendorf (119km2), using CCLM uncorrected at 0.11° is better than 0.03° for simulation of flood frequency; and in the corrected case the essential nature of 3km is not clear at all. For WRF, 0.44° and 0.70° are better for uncorrected case considering flood frequency over Fluttendorf. For the corrected case, the essential nature of 3km for flood frequency is not evident. Seasonality in WRF is very often degraded by 0.03° resolution, for both corrected and uncorrected, including in the catchments with <200km2.

It is also stated in the conclusions that "in the larger catchments, the 12.5 km and 50 km resolution already yield satisfying results regards flood statistics". Concerning the flood frequency, the results are not "often not already satisfying", the problem here is that increasing to 0.03° degrades the results. Hence we need a priori knowledge of whether we should use convective permitting or not. For Seasonality, CCLM does seem to be improved by using 0.03°, but not WRF, which is also problematic. The abstract also reflects these unclear statements of added value, when compared to the results.

AR: This is true, there is no systematic behaviour. Any differences are further amplified by the non-linearity in the flood generation process (particularly above return periods of 5-8 years, small differences in precipitation can induce large differences in flood peaks). We will include this into the discussion. On the other hand, this shows the importance of a test of the models against historical data before applying in impact analyses.

Reviewer #2 minor comments:

RC: Pg. 1, Line 10: "an in increase in regional climate model". Instead of repeating "regional climate model", it could be replaced by RCM or just model.

AR: OK.

RC: Pg. 1, Line 10: "Increase in regional climate model resolution and in particular, at the convection permitting scale, will lead to a better representation of the spatial and temporal characteristics of heavy precipitation at small and medium scales". This sentence is technically correct, but it's not very clear. It could be re-written. Increasing the resolution will lead to a better representation of small scales. But if we are using a coarser resolution the small scales are not represented (they are not explicitly simulated). Of course, this will depend on what is meant by "small and medium scales", which is not entirely clear. Perhaps quantify these. Notice that throughout the text "small and medium scales" is also used rather loosely. For example, in Pg. 3 line 13 it is (30km2 to 1000 km2), but in pg. 8 line 17 it is (75kms to 200km2), while <1100 km2 is referred to as large.

AR: The first sentence will be re-written. We take out the term "small and medium scales" here, in order to avoid the mismatch of terms from modelling techniques ("processes on resolved/unresolved scales") and the size of the investigated catchments ("small and medium"). The new version of the sentence now is: "Regional climate model (RCM) evaluations and inter-comparisons have shown that an Increase in regional climate model resolution and in particular, at the convection permitting scale, will lead to a better representation of the spatial and temporal characteristics of heavy precipitation." Further on, the terms "small and medium scales" will be clarified and used consistently through the text.

RC: Pg. 24, line 10: "Moreover, catchments with an area less than 100 km2 require a 1-hour time step due to the short response times": this is based on one single case (Schwanberg)? Perhaps it should be stated that "Moreover, the catchment with an area …". It this generalizable?

AR: We tested the influence of the time step by using the 3hr sums of the CCLM 3km and comparing to the 1 hr results. There is a decrease of flood peaks, but the main decrease in performance in the small catchment Schwanberg is due to the error correction. We will discuss the reasons more in detail on the basis of a better explanation of the SDM method (see response to reviewer #1). Done

RC: Figure 14 the green circle in the last panel (bottom right) is not visible.

AR: It is behind the red-filled square. The symbol for the observations will be made larger. Done

Response to reviewer 3

Reviewer #3 general comments:

RC: [..] but it raises more questions than it answers and as such I think the authors need some more nuanced discussion and also to tone down the conclusions a bit. [..]

AR: We will include a comprehensive discussion about the different results and reduce the strength of the conclusion (see response to reviewer #1). We agree, that there are still some open questions which we will address (some questions "OQ" are marked later in the text).

RC: [..] Also the sample of two catchments is too small to make any generalizable conclusions. Maybe the authors could add a few more small catchments to the study to help bolster their case. [..]

AR: Of course, the extent of the test area in south eastern Austria cannot claim any generalized conclusions for say, the European scale. It is an area, where a spatially distributed model could be calibrated on a very small scale sub-catchment basis. The evaluation catchments were carefully selected in order to be representative in the area. As stated in the text, despite the small overall test area extent, the variety of climatic, topographic, geologic and pedologic properties is high. The catchments were selected to represent these different properties. There are several gauges in neighbouring or nested catchments available and the results there are consistent with the selected catchment for each region. **OQ:** Perhaps this small scale variability of catchment properties (which lead to different model parameters – response times, non-linearity!) is one reason for the unsystematic results. Also, the quality of some temporal characteristics of the catchment-accumulated precipitation (frequency, duration, intensity) that is simulated by CCLM and WRF in their various resolutions has not been investigated yet. We will address these issues (see also response to reviewer #1). Particularly the latter was analysed in a comprehensive way.

<u>RC:</u> Also there is the fact is that these results appear to be highly model dependent. The authors offer some explanation by way of the fact that CCLM is well tuned and widely used over this region while WRF is not. However, recent studies show comparable performance by WRF over the Greater Alpine Region and Europe more generally (Awan et al., 2015; Knist et al., 2018; Kunstmann et al. 2018). Rather than a hand waving generalization about model family rather an more detailed description of which processes the simulations reproduce correctly and why might be more informative.

<u>AR:</u> Yes, we agree. The generalization is too short-sighted. It has been demonstrated multiple times, that CCLM and WRF show similar performance indices for precipitation on coarser resolutions (12.5 km, 50 km), e.g. Kotlarski et al. (2014), Smiatek et al. (2016). However, when it comes to convection-permitting simulations, a judgement of the model performance becomes difficult, because such systematic model-inter-comparison studies do not exist and one has to rely on published evaluation studies, that vary in the model domain and/or in the length of the simulation period. For instance, Knist et al. (2018) conducted and compared a pan-European WRF simulation with 3 km horizontal grid spacing with data records from ground based stations, however, the majority of the stations covers Germany and only a few of them are located in the Alpine region. On the other hand, Ban et al. (2014) evaluated a CCLM simulation with 2.2 km grid spacing, but only with station data from Switzerland, which raises questions about the comparability with the results of Knist et al. (2018). Nonetheless, both studies have in common, that the convection permitting simulations capture the frequencies of heavy and extreme hourly precipitation better than their coarser resolved counterparts and, that extreme events are more overestimated in mountainous regions. We agree, that our CCLM and WRF simulations are ideal to fill the gap of missing comparable convection permitting simulations; however, the focus of this paper lies on driving a flood model with RCMs and to investigate the effects of the RCMs' resolutions and a bias correction on the representation of floods. For the time being, we will put more effort on understanding the effects of the bias correction (see response to reviewer #1) and provide understating of the sources of RCM biases only as far as it is necessary. A thorough CCLM/WRF inter-comparison study is out of the scope of this paper. Such systematic model-inter-comparison studies are subject to the Flag Ship Pilot Study (FPS) "Convective phenomena at high resolution over Europe and the Mediterranean" of the Coordinated Downscaling Experiment (CORDEX) of the World Climate Research Programme (WCRP) that has started in 2016 and to which CCLM and WRF simulations of the Wegener Center contribute to.

Ban, N., Schmidli, J. and Schaer, C.: Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations, J. Geophys. Res.-Atmospheres, 119(13), doi:10.1002/2014JD021478, 2014.

Knist, S., Goergen, K. and Simmer, C.: Evaluation and projected changes of precipitation statistics in convection-permitting WRF climate simulations over Central Europe, Clim. Dyn., doi:10.1007/s00382-018-4147-x, 2018.

Kotlarski, S., Keuler, K., Christensen, O. B., Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., van Meijgaard, E., Nikulin, G., Schär, C., Teichmann, C., Vautard, R., Warrach-Sagi, K. and Wulfmeyer, V.: Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble, Geosci. Model Dev., 7(4), 1297–1333, doi:10.5194/gmd-7-1297-2014, 2014.

Smiatek, G., Kunstmann, H. and Senatore, A.: EURO-CORDEX regional climate model analysis for the Greater Alpine Region: Performance and expected future change: CLIMATE CHANGE IN THE GAR AREA, J. Geophys. Res. Atmospheres, 121(13), 7710–7728, doi:10.1002/2015JD024727, 2016.

_RC:_ *What would help greatly would be a more in-depth look at the bias correction method, discussion of the effects of the different nesting strategies, the addition of more small catchments and more nuanced discussion and conclusion sections. […]*

AR: A more comprehensive description of the bias-correction method and how it alters precipitation events that cause floods in our catchments will be given (see response to reviewer #1). Concerning nesting: CCLM 0.44° and CCLM 0.11° are directly driven by ERA-Interim (single nesting), while WRF 0.11° is nested into WRF 0.44° (double nesting). Note, in all 0.44°/0.11° domains sea surface temperature (which has a major impact because about 50% of the European model domain is covered by ocean) is deducted from ERA-Interim. Since the 0.44° domains are only slightly larger than the 0.11° domains and hence model internal variability can only slightly introduce deviations from ERA-Interim within this narrow boundary zone (area of 0.44° domain minus 0.11° domain), the effect of these two nesting strategies is expected to be minor. These nesting issues will be shortly addressed in the discussion. The reasons why the CCLM and WRF performs in the way shown in the paper was additionally analysed. See above.

_RC:_ *[..] Also the conclusion section should be rewritten with a more nuanced interpretation of the results. Is convection permitting modeling really needed if, after bias adjustment, results are no better or only modestly improved compared to coarser resolution simulations? Should multi-model, multi-realization, ensembles be employed or rather one highly tuned simulation? For present climate this might be sufficient but such tuning has well demonstrated shortcomings at climate time scales. Clearly there are substantial challenges remaining before these types of simulations can reliably be used for impacts models. At present the authors fail to acknowledge this and I think somewhat overstate their results. Also, the authors claim that recommendations can be made but then fail to deliver on this promise. What general recommendations, if any, can be made based on this study? Does convection permitting modeling only provide added value over particular areas, for particular cases, particular time scales and particular cases/phenomena? The results here certainly seem to point towards such a limited use or at the very least a need to balance expectations with current capabilities.*

AR: We agree that "*there are substantial challenges remaining before these types of simulations can reliably be used for impacts models*". We will acknowledge these. We were a bit too optimistic in the conclusion because we didn't expect the good CCLM 3km results (uncorrected); also when looking at event scale. Large events were simulated plausible (magnitude and dynamics), particularly events induced by large scale frontal systems. **OQ**: difficulties seem to occur under weak synoptic forcing. This can be partly explained with model-internal-variability as it was published first by Kida et al. (1991). Since our nesting strategy does not make use of any nudging technique, the interior grid cells of the RCMs' domains decouple from the synoptic situation as it prescribed by ERA-Interim. This decoupling effect is largest when synoptic forcing is weak and even small deviations in mesoscale dynamics (that are responsible for moisture supply) can have a large impact on the resultant precipitation on a given location. A paper from the CORDEX-FPS community that points out this issue is currently under review (Coppola et al.). This event type issue was discussed in the paper by Fig. 9 and 12. Of course, this raises questions about the usage of bias correction methods that are not aware of temporal and/or spatial displacements of single events as well as future long-term climate projections (ensemble recommended, testing against historical flood data necessary, etc.). We will acknowledge this and will try to give some concluding remarks. Done

In general, we will add a detailed discussion section, either within the synthesis section or by a separate section, where we will include all the discussion text hidden in the different chapters and extended by the very valuable issues addressed by the reviewers.

Coppola, E., S. Sobolowski, E. Pichelli, F. Raffaele, B. Ahrens , I. Anders, N. Ban, S. Bastin, M. Belda, D. Belusic, A. Caldas-Alvarez, R. M. Cardoso, S. Davolio, A. Dobler, J. Fernandez, L. Fita Borrell, Q. Fumiere, F. Giorgi, K. Goergen, I. Güttler, T. Halenka, D. Heinzeller, Ø. Hodnebrog, D. Jacob, S. Kartsios, E. Katragkou, E. Kendon, S. Khodayar, H. Kunstmann, S. Knist, A. Lavín-Gullón, P. Lind, T. Lorenz, D. Maraun, L. Marelle, E. van Meijgaard, J. Milovac, G. Myhre, H.-J. Panitz, M. Piazza, M. Raffa, T. Raub, B. Rockel, C. Schär, K. Sieck, P. M. M. Soares, S. Somot, L. Srnec, P. Stocchi, M. H. Tölle, H. Truhetz, R. Vautard, H. de Vries, K. Warrach-Sagi, A first-of-its-kind multi-model convection permitting ensemble for investigating convective phenomena over Europe and the Mediterranean, Climate Dynamics, under review.

Kida, H., Koide, T., Sasaki, H. and Chiba, M.: A New Approach for Coupling a Limited Area Model to a Gcm for Regional Climate Simulations, J. Meteorol. Soc. Jpn., 69(6), 723–728, 1991.

Reviewer #3 specific comments:

P3L7-8: The community is well beyond "first attempts". WRF Hydro (a fully coupled distributed hydrological model within WRF) is far enough in its development to be the core model for the United States' National Water Model. https://ral.ucar.edu/projects/wrf_hydro/overview. Other such systems in operation are TerrSysMP which features a 3-D groundwater model coupled to COSMO-CLM (e.g. Keune et al., 2017). Note that I do not make any on their reliability over climate time scales (i.e. simulations around a decade or more).

_AR:_ We agree, that the idea to couple hydrological models with RCMs is not new. However, our hydrological model KAMPUS is – thanks to the calibration process that is described in this paper - operationally used for flood-forecasts in the Styrian region (see Fig. 1). Driving KAMPUS with RCM-output to test its applicability for climate impact studies in future activities was a dedicated goal of the underlying research project (CHC-FloodS, see Acknowledgements) that funded this paper. Using a fully coupled modelling system, like WRF-Hydro, has many advantages concerning physical consistency etc., but it also has the drawback, that it is limited to the usage of one RCM (WRF in the case of WRF Hydro). In climate impact application ensembles of RCMs need to be used, if uncertainty in projected climate changes should not be underestimated. Hence, in our case it was important to build up CCLM-KAMPUS and WRF-KAMPUS modelling chains and in this aspect this is the first attempt. We will correct this in the paper. _By the way:_ TerrSysMP is a groundwater model that makes use of a surface runoff module that is also part of WRF Hydro, KAMPUS is a specified flood model calibrated to our catchments.

P3L13: Please be clear that when you write "coupled" you mean limited one-way coupling (I actually wouldn't call this coupling at all) wherein there is no feedback between the hydrological model and the atmospheric model and the atmospheric model only passes temperature and precipitation.

_AR:_ We will clarify that: It is a sequential coupling with no feedback. Done

P4 L1-7: What are the potential ranges of observational uncertainty? In addition to sensor errors and under catch there is also uncertainty resulting from interpolating to a grid from point based station data. How are these taken into account?

AR: The hydrological model calibration aims to remove systematic input errors. Of course, other interpolation methods would yield different model parameters. But accuracy depends mainly on the number of stations and available additional information, not so much on the interpolation method (see work of U. Haberlandt et al.). To analyse this would be beyond of this paper. We used all available station data in the area with the dense network of daily recording rain gauges as additional information. Snow under catch is accounted for by a model parameter (snow correction factor) which is calibrated.

P4L8: The version of WRF used here is quite old (almost 8 years!) and many of the issues related to this version have been corrected. In fact WRF is now 6 full versions more advanced as of this writing. How might this have affected to the results?

AR: WRF 3.3.1 is the main model version that has been extensively used in the regional climate modelling initiative EURO-CORDEX. The 0.44° and 0.11° simulations have been conducted by Klaus Görgen and contributed to EURO-CORDEX. In our 0.03° simulation we did not want to divert too much from the EURO-CORDEX version for comparability reasons. We agree that the model was improved multiple times since then. One of the major improvements was developed based on our 3.3.1 simulation: we found a bug in the treatment of lateral boundary conditions in the original 3.3.1 version that introduced unphysical artefacts and caused unforeseeable model crashes after simulating two years. The fixed code entered the official release in version 3.7 (see http://www2.mmm.ucar.edu/wrf/users/wrfv3.7/updates-3.7.html topic: "improved specified bdy for long simulations"). In our 3.3.1 version this bug is already fixed. We agree that there might be other changes in the code that would significantly change the results. We will acknowledge this in the conclusion section accordingly.

P6-Error Correction: More details on the bias adjustment are needed given that it is being pitched as a "novel" approach. How does it perform relative to other approaches? What are its limitations and/or tradeoffs? What are the implications of univariate approach to bias adjustments when the two variables corrected, temperature and precipitation, are related to each other? Also more explanation of the issues/limitations behind current approaches is needed. Maraun et al (2017) have an excellent overview of the current state of bias adjustment shortcomings and placing the SDM approach among these would be helpful to readers.

AR: A more detailed description of the bias correction and its effects on the flood modelling will be given (see response to reviewers #1 and #2). Done, particularly the influence on flood modelling is analysed by additional figures and text.

P8L29: "relatively".

AR: Done

P12L1: Specify which figure/panel you are referring to.

AR: OK

P13L10: I don't believe it has been demonstrated that CPM is "absolutely necessary". I would suggest either making a stronger argument with stronger supporting evidence or modify this claim.

AR: OK, see response to general comments above and previous reviewers.

P17L23: The authors write that “performance using 3km data is still best” but it is hard to discern this from the figures. Figure 11 show 24 seasons in total and of those the 3km is closest to observations in only 11 of these. In the other seasons either the 0.11 or 0.44 degrees simulation is closer to observations or the performance across resolutions is equal.

AR: Indeed, results of all the different RCMs after bias correction are similar (deviations are - with some exceptions - only between 1 or 2 events), which can be interpreted as accurate.

P16L10: Performance after bias adjustment is degraded over the smallest catchment. Yet this is precisely the type of application (i.e. small catchments) where the authors argue we see the greatest added value of convection permitting modeling. Here it appears that the two techniques, high-resolution modeling and bias adjustment, are not working in concert but in opposition.

AR: This is true and we will address this. We analysed the effect of the 3hr aggregation (see response above – p.5, bottom)

P20L30-31: This is in direct contradiction to earlier, and later statements, that CP scales are “absolutely necessary”. I would say rather that it is clear that there is still quite some work to do before these models can be reliably used for these sorts of applications.

AR: We agree.

P24L10: See previous comment. I do not think the authors have presented evidence sufficient to make this statement.

AR: We agree. Essential is too strong. But the results using CCLM before bias correction are really promising (and surprisingly good at the same time)

P25L17-18: Clearly CCLM has higher performance than WRF in this study. However, the authors never discussed the nesting strategy (see table 2), which is different for each model system. Specifically, WRF goes through an additional intermediate nest, a step that will certainly have an impact. How then are then are the WRF and CCLM simulations directly comparable?

AR: Both nesting strategies (see above) are frequently used in regional climate modelling frameworks. From the climate modelling point of view, both strategies are equally binned. The goal in our study was to investigate how well statistical properties of floods can be simulated by physically based modelling chains in usual climate modelling frameworks. Having this in mind, the nesting strategy becomes only relevant if unphysical perturbations are introduced via one or the other nesting technique, which is not the case. Nevertheless, we will make more statements about the nesting strategy.

P25L4: What “recommendations”, specifically, can be made?

AR: We agree, due to the unsystematic results and the small test domain general recommendations are hard to give, we will add some specific recommendations - or better remarks - regarding the results of the paper, open questions and possible solutions (e.g., CCLM 3km results promising, but trade-off with computational costs, seasonality important – indication for an accurate representation of atmospheric processes, error correction degrades results –

should error correction compensate large errors?, test with historical data necessary). We have re-written the conclusions.

P25L5: The modeling systems used here are not "coupled" they are used in a model chain

AR: Sequential coupling. We will clarify this. Done

Figure 1. It is almost impossible to make out the catchment boundaries and initials in this busy figure. I suggest moving the catchment labeling to the larger figure 3.

Figure 3. Place catchment labels here in bold. Also bold lines around the catchments themselves so that the six catchments under investigation are clearly delineated.

AR: Fig. 1 and Fig 3 will be re-structured. OK, labels will be moved to Fig. 3. Done

Figure 4. What region is shown here?

AR: Fig. 4 will be removed (see response to reviewer #1, p10) Done

Figure 5. Remove the empty panel in the lower right corner.

AR: OK.

Figure 7. It is very hard to distinguish between the blue and black circles. Also including red and green is not colorblind friendly. I suggest a different color scale that has greater separation. This comment applies to Figures 7-12 and 14.

AR: Colours depend also on the printing. We will check colours before final print.

Convection-permitting regional climate simulations for representing floods in small and medium sized catchments in the Eastern Alps

Christian Reszler[1], Matthew Blaise Switanek[2], Heimo Truhetz[2]

[1]JR-AquaConSol GmbH, Steyrergasse 21, 8010 Graz, Austria
[2]Wegener Center for Climate and Global Change, University of Graz, 8010 Graz, Austria

Correspondence to: Christian Reszler. E-mail: christian.reszler@jr-aquaconsol.at

Abstract. Small scale floods are a consequence of high precipitation rates in small areas that can occur along frontal activity and convective storms. This situation is expected to become more severe due to a warming climate, when single precipitation events resulting from deep convection become more intense ("Super Clausius-Clapeyron effect"). Regional climate model (RCM) evaluations and inter-comparisons have shown that there is evidence that an increase in regional climate modelRCM resolution and in particular, at the convection permitting scale, will lead to a better representation of the spatial and temporal characteristics of heavy precipitation at small and medium scales. In this paper, the benefit of grid size reduction and bias correction in climate models are evaluated in their ability to properly represent flood generation in small and medium sized catchments. The climate models are sequentially coupled with a distributed hydrological model. The study area is the Eastern Alps, where small scale storms often occur along with heterogeneous rainfall distributions leading to a very local flash flood generation. The work is carried out in a small multi-model (ensemble) framework using two different RCMs (CCLM and WRF) in different grid sizes. Bias correction is performed by the use of the novel Scaled Distribution Mapping (SDM, similar to usual quantile mapping) method. The results show, that in the investigated RCM ensemble, no clear added value of the usage of convection permitting RCMs for the purpose of flood modelling can be found. This is based on the fact that flood events are the consequence of an interplay between the total precipitation amount per event and the temporal distribution of rainfall intensities on a sub-daily scale. The RCM ensemble either lacks on one and/or the other. In the small catchment (< 100 km²), a favourable superposition of the errors leads to seemingly good CCLM 3km results both for flood statistics and seasonal occurrence. This is however, not systematic across the catchments. that for small catchments (< 200 km²) a resolution of 3 km is essential to accurately simulate the magnitude of flood events. Flood frequency and seasonality are both represented well in all catchments. In the larger catchments resolutions of 12.5 km and 50 km already yield statistically satisfying results, but poor results regarding seasonality. Also, due to the short response times in the small sub-catchments a time step of 1 hour is required. In all setups a bias still exists in precipitation and temperature, which sometimes leads to unrealistic hydrological conditions The applied bias correction only corrects total event rainfall amounts in an attempt to reduce systematic errors on seasonal basis. It does not account for errors in the temporal dynamics and deteriorates the results in the small catchment. Therefore it cannot be recommended for flood modelling. demonstrating the

necessity of bias correction. The results show the added value of reducing grid size and bias correction in climate models that can be used to model flood mechanisms in small and medium sized catchments.

1 Introduction

Floods in small and medium sized catchments are often triggered by atmospheric processes on small scales, i.e., small scale frontal systems (Schemm et al., 2016) and convective storms. In the Austrian Alpine area, these types of small scale storms cause millions of Euros in damage every year. This situation is expected to become more severe as a result of a warming climate and the Clausius-Clapeyron relationship. Single precipitation events are expected to become more intense (e.g., Allen and Ingram, 2002; Trenberth et al., 2003; Allan and Soden, 2008; Gobiet et al., 2013), and recent investigations have shown increases in deep convective precipitation can exceed the Clausius-Clapeyron relationship (known as the Super Clausius-Clapeyron Scaling effect, e.g., Lenderink and Van Meijgaard, 2009; Berg et al., 2013; Wang et al., 2017; Lenderink et al., 2017).

Regional climate models (RCMs) are valuable tools for studying climate change effects on water resources. They are employed to generate climate simulations at scales below a 50 km horizontal resolution, like in the EU-FP7 project ENSEMBLES (Hewitt and Griggs, 2004) or the North American Regional Climate Change Assessment Program (NARCCAP) (Mearns et al., 2009). RCMs operating with 0.11° (~12 km) grid spacing became the standard in Europe as a result of EURO-CORDEX (www.euro-cordex.net) (Jacob et al., 20132014), which is the on-going European branch of the global Coordinated Regional Downscaling Experiment (CORDEX) (Giorgi et al., 2009) of the World Climate Research Programme (WCRP). Prein et al. (2016) investigated the added value in precipitation in the EURO-CORDEX RCMs. They demonstrated that as model resolution increased, atmospheric processes such as extreme precipitation are more realistically represented, especially in regions of complex terrain (e.g., the Alpine region). Nissen and Ulbrich (2017) focused on the representation of heavy precipitation events in the EURO-CORDEX ensemble. They found that the frequency and size of heavy precipitation events are predicted to increase over most of Europe with increasing greenhouse gas concentrations. Moreover, the most severe events were detected to be in the projection period.

With improvements in numerical weather prediction (NWP) and computing technology, RCM grid spacing can now be further reduced to allow convection permitting climate simulations (CPCSs). Two major advantages of CPCSs benefit from two major advantages with respect to concerning precipitation extremes are: (1) deep moist convection, which is the most important process in the majority of extreme precipitation events, are physically resolved by the RCM and (2) the representation of orography and surface fields is improved. Multiple studies have already demonstrated the added value of convection-permitting models (CPMs, Prein et al., 2015) in capturing extreme precipitation (e.g., Chan et al. 2013; Chan et al., 2014; Meredith et al., 2015; Chan et al., 2017; Zittis et al., 2017) and their frequency of occurrence (Ban et al. 2014; Knist et al. 2018). However, there are only a few future projections that use CPCSs, like (e.g., Prein et al., (2017), ; Ban et al., (2015), ; Kendon et al., (2014), and Knist et al. (2018). Although processes are better represented in CPCSs, It it should be noted that local biases using CPCSs doesare not necessarily being reduced biases. Their bandwidths are large and (spatial and temporal) correlation coefficients are poor, when they are compared to highly resolved observation data (e.g. Prein et al., 2013; Ban et al. 2014; Knist et al. 2018). Especially, Ban et al. (2014) and Knist et al. (2018) found in common, that their models (CCLM and WRF) increasingly overestimate extreme events in mountainous regions. This makes bias correction techniques indispensable, even if deep convection becomes resolved by RCMs. Also, additional computational costs are high which can limit an application particular for climate change studies in decision making.

Hinging on the scale of the driving data, climate change impact studies have often focussed on water balance in relatively large catchments (e.g., Fowler et al., 2007). Regarding floods, numerous studies were performed and pointed out the high uncertainties in the GCM-RCM-hydrological model chain (e.g., Hennegriff et al., 2006; Dankers et al, 2007; Hanel and Buishand, 2010). Maraun et al. (2010) provided a comprehensive review on the requirements of hydrological models and their fulfilment via RCMs. They define the requirements in a correct representation of (1) intensities, (2) temporal variability, (3) spatial variability, and (4) consistency between different local-scale variables. Köplin et al. (2014) used future climate change scenarios from the ENSEMBLES project to analyse the seasonality and magnitude of floods in Switzerland. They found that the simulated change in flood seasonality is a function of the change in flow regime type. Magnitudes of both mean annual floods and maximum floods (in a 22-year period) are expected to increase in the future because of changes in flood-generating processes and scaled extreme precipitation. Using the new EURO-CORDEX models Alfieri et al. (2015) assessed projected changes in flood hazard in Europe based on the RCP8.5 scenario and the hydrological LISFLOOD model. Their results indicate that the change in frequency of discharge extremes is likely to have a larger impact on the overall flood hazard as compared to the change in their magnitude. On average, in Europe, flood peaks with return periods above 100 years are projected to double in frequency within 3 decades. In an effort to sequentially couple convection permitting RCMs with a hydrological model, first attempts have been made. For example, Kay et al. (2015) use results of 1.5 km RCM nested in a 12 km RCM driven by European-Reanalysis boundary conditions to drive a gridded hydrological model. However, they found that the 1.5 km RCM generally performs worse than the 12 km RCM for simulating river flows in 32 example catchments.

In this study, two regional climate models (CCLM, WRF) with different grid spacing (~50 km, ~12.5 km, and ~3 km) are sequentially coupled (one way) with a hydrological model for representing floods on small and medium spatial scales (30 km² to 1000 km²). An improved bias correction technique (Switanek et al., 2017) is used to minimize error propagation throughout the modelling chain. The study area is located in South-Eastern Austria (Styria), where local flash floods are the predominant flood type (e.g., Merz and Blöschl, 2003). The spatially distributed hydrological model KAMPUS (Blöschl et al., 2008) is used, which is in operational use for flood forecasting in Austria in small to medium sized scales (Blöschl et al., 2008; Ruch et al., 2012). The added value of the highly resolved convection permitting RCM setup (~3 km grid spacing) is evaluated in the period 1989-2010 by quantitative and qualitative criteria regarding flood generation.

2 Study area and observation data

The study area is located in South-Eastern Austria, at the border of the Eastern Alps (Fig. 1). Meteorological data of all available stations in the region were acquired from the Hydrographic Service of the provincial government of Styria and the Austrian Central Department of Meteorology and Geodynamics (ZAMG). Figure 1 shows the distribution of the stations between the period 2000 to 2009, which corresponds to the calibration period of the hydrological model. Data coverage has improved through the years by installing new stations. Historically in Austria, network of stations with daily data ("Ombrometer") is much denser than the network of stations with high temporal recording (e.g., every 15 minutes or hourly). In the bottom right plot the development of the station availability in Southern Styria is shown. At the beginning of 2000 the number of stations with high temporal resolution significantly increased, whereas the number of stations with daily data was high since the beginning of the study period in 1989.

Interpolated fields of precipitation and air temperature are generated on an hourly basis. Stations with daily data are incorporated into the interpolation procedure to benefit from the dense network as follows (Reszler et al., 2006): First, daily data are interpolated on the model grid (1 km). Then, hourly data are interpolated on the same grid and the daily sum of the cells is calculated and scaled to the daily grid. Spatial distribution of daily precipitation is expected to be accurate even in the years before 2000, which is important for an accurate representation of the general water balance. However, due to the high spatial variability of precipitation in the region, hourly fields before 2000 contain more uncertainty. In contrast, uncertainty in interpolated hourly air temperature is generally much lower. The data were interpolated by a regression with station altitude and an interpolation of the residuals on the 1 km working grid. As an interpolation method for both variables the "Inverse Squared Distance" method was used. The interpolated fields for model calibration serve also as a reference data set for the RCM evaluation.

Runoff data for a high number of stream gauges are available at an hourly time step. These gauges are all used for model calibration (black triangles in Fig. 1, data provided by the Hydrographic Service of Styria). Representative gauges were selected in this study (labelled triangles with corresponding catchment boundaries, Table 1) in order to cover a wide range of catchment sizes (75 km² to 1100 km²) and different characteristics of soils and geology. There are more gauges used in the Western part, because the catchments vary largely by slope, climate, geology and soil type. This leads to differences in flood response and occurrence. For example, catchments of the gauges Schwanberg (S) and Voitsberg (V) reach relative high altitudes up to 2100 m a.s.l. at the Koralpe massif and are therefore expected to show significant influences of snow in winter and spring. Geology is crystalline (predominating gneiss and schist) with deep weathering zone (Flügel and Neubauer, 1984; BMLFUW, 2007) which implies significant storage capacities. Areas at the foot of the Koralpe consist mainly of tertiary sediments with low storage capacities (BMLFUW, 2007). Runoff at the corresponding gauges (e.g., gauge Gündorf – Gü) shows a relatively rapid response to rainfall and low baseflow (Ruch et al., 2012). The gauges Tillmitsch (T) covers the whole Lassnitz branch which flows into the Sulm which is gauged at the catchment outlet in Leibnitz (L).

In the Eastern part (so-called Grabenland creeks) only one gauge (Fluttendorf – Fl) is selected because of the relatively homogeneous climate, geology and soils. The runoff record extends over the whole simulation period and data are assumed very reliable according to the data provider Hydrographic Service. Geology and soils mainly consist of tertiary material. Influence of continental climate is increasing towards the east with values of annual precipitation in the order of annual evapotranspiration: Mean annual precipitation (MAP) is 700 mm in the East, whereas in the Western part, MAP ranges from 1100 mm at the foot to 1500 at the high altitudes.

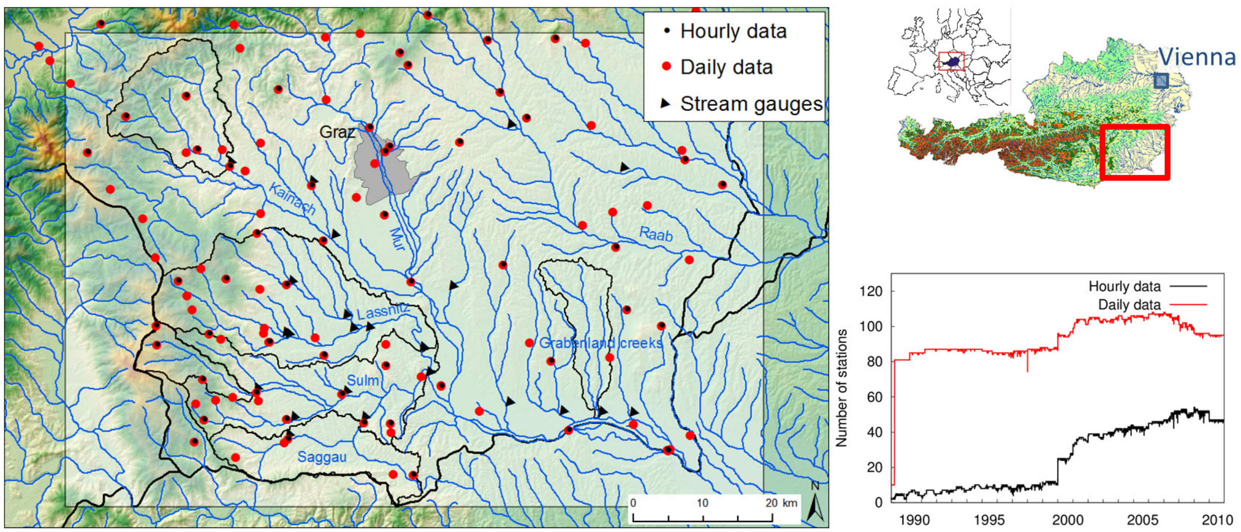

Figure 1: Study area and station distribution (left) as well as availability of high resolution meteorological data (bottom right). Polygons are the catchment boundaries of the stream gauges for evaluation in this study (Table 1, Fig. 3).

3 Method

3.1 Regional climate models

The RCMs we employ are the non-hydrostatic COnsortium for Small-scale Modeling (COSMO) model in CLimate Mode (COSMO-CLM or CCLM) (Böhm et al., 2006; Rockel et al., 2009) version 4.8 clm 17 and the Advanced Research version of the Weather Research and Forecasting model (WRF/ARW) (Skamarock et al., 2007) version 3.3.1. Both models are driven by the re-analysis dataset ERA-Interim (Dee et al., 2011) and cover the period 1989 to 2010. The models' innermost domain, the Greater Alpine region (GAR) with 3 km grid spacing, is reached via intermediate pan-European ~~(EURO-CORDEX)~~ domains (without nudging) with 12.5 km grid spacing for CCLM and 50 km and 12.5 km grid spacing for WRF. By doing so, we mimic a typical setup as it is used in regional climate modelling applications and we do not run the risk of underestimating ~~uncertainty~~ internal variability in our investigations. The simulations of the pan-European domains have contributed to the EURO-CORDEX initiative and have been evaluated in several studies, e.g. Katragkou et al. (2015), Kotlarski et al. (2014) and Prein et al. (2016). The model configurations for the convection-permitting (3 km grid spacing) simulations in the GAR are based on experiences from previous sensitivity experiments (Suklitsch et al. 2011; Awan et al. 2011; Prein et al. 2013; Prein et al. 2015). Our RCMs differ from their coarser resolved counterparts (EURO-CORDEX) insofar that the parametrization for deep-convection, the Tiedke scheme (Tiedke, 1989) in CCLM and the Kain-Fritsch scheme (Kain, 2004) in WRF, has been turned off in the GAR. Overview of the model domains and simulations used are given in Fig.2 and Table 2, respectively.

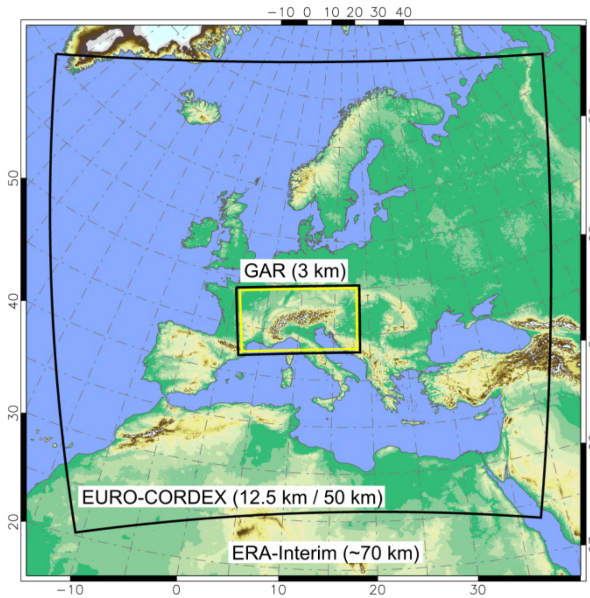

Figure 2: RCM domains. ERA-Interim is dynamically downscaled with CCLM and WRF from its initial resolution of ~70 km to ~~3~~ 3 km in the Greater Alpine Region (GAR) by making use of an intermediate model domain, the EURO-CORDEX domain with 12.5 km and 50 km grid spacing.

3.2 Error correction

The novel method Scaled distribution mapping (SDM) is used to bias correct the model precipitation and temperature data time series (Switanek et al., 2017). SDM is a parametric method, but it is nearly identical to that of quantile mapping (QM) when correcting the historical period. However, for a "future" period (or any period outside of calibration), the method scales the observed distribution by the relative (for precipitation) or absolute (for temperature) distances between the "future" and historical modelled CDFs. The commonly used bias correction method of ~~quantile mapping~~QM (Wood et al., 2004; Piani et al., 2010; Themeßl et al., 2011; Teutschbein and Seibert, 2013) assumes that error correction functions can be treated as stationary from one time period to another. This assumption is responsible for altering the projected climate change signal. For example, a projected mean increase of precipitation of 20% can be inflated to be 30%, while extremes can be altered even more dramatically. However, Maraun (2012), Teutschbein and Seibert (2013), Maurer and Pierce (2014) and Switanek et al. (2017) showed this assumption of a stationary error correction function to be invalid, and as a result, the altering of the raw model projected changes to precipitation and temperature was found to be unjustified. In addition, quantile mapping was found to overestimate values of low precipitation and underestimate high precipitation (Maraun, 2013). SDM, in contrast, does not rely on a stationary error correction function, but rather attempts to best preserve the raw model projected changes across the entire distribution. However, the over (under) estimation of low (high) precipitation intensities remains. Bias correction was performed on RCM precipitation and temperature data independently for each grid cell and calendar month. It was implemented on a 3-hourly window to more accurately capture the observed diurnal cycle.

3.3 Hydrological model

The spatially distributed model KAMPUS (Blöschl et al., 2008) is used, which is in operational use for flood forecasting in Austria. It contains conceptual models for snow melt, soil moisture accounting and flow routing. The snow model is based on the degree-day approach which calculates snow melt depending on the air temperature. For snow accumulation precipitation is split into snow and rainfall by a lower and an upper threshold temperature with a linear transition. Depending on the actual soil moisture, rainfall and snowmelt are non-linearly partitioned into a component that increases soil moisture and a component that contributes to runoff, dQ. Soil moisture can only be depleted by evapotranspiration. Runoff routing on a raster cell (hillslope) is represented by an upper zone and two lower zones, which are formulated as linear reservoirs. dQ is the input into the upper zone. The zone has three outlets: (i) outflow with a low storage coefficient (k_1) that represents interflow, (ii) percolation to the lower reservoirs (saturated zone), and (iii), when a defined threshold, L_1, is exceeded, outflow with a very low storage coefficient (k_0) representing surface or near surface runoff. The percolation rate into the two lower zones is separated into two components by a factor. Outflow of the lower zones is defined as groundwater flow and deep groundwater flow, respectively. A bypass flow, dQ_{by}, routes rainfall and snow melt directly into the lower storages (macro-pore flow). Model structure is described in detail in Blöschl et al. (2008). In this work, the original vertical structure is extended by a module for infiltration excess. At very high intensities ($I > I_{crit}$) parameters of soil storage are reduced, and bypass and deep percolation is set to zero. Values for I_{crit} and the reduction of infiltration parameters are obtained by calibration.

Total runoff on a grid cell is calculated as the sum of the outflows from all zones. It is then aggregated to sub-catchments and convoluted by a linear storage cascade which represents runoff routing in the stream network within each of the sub-catchments. Routing in the river reaches which connect model nodes is formulated by a cascade of linear reservoirs (Reszler et al., 2008b). By a step-wise linear formulation, this model allows for incorporating non-linear effects in flood rooting, such as flood wave acceleration at high water levels and flood retention at flood plains. For the latter, board-full discharge and existing 2D hydrodynamic studies have been provided by the Hydrographic Service of Styria for calibrating the corresponding parameters. This is particularly important for a plausible representation of flood peak attenuation at very large floods. Since the hydrological model is also driven by simulated, often biased, precipitation input, flood peaks may be simulated which exceed observations.

The model domain extends over all of Southern Styria (grey shaded window in Fig. 1). The Western part has previously been calibrated (Ruch et al., 2012), as it is implemented for operational flood warning by the provincial government of Styria. The

Eastern part was extended in the current study. The model has a sub-catchment structure with 96 catchments and 152 internal model nodes (Fig. 3). The model is driven by precipitation and air temperature with an hourly temporal resolution and a 1 km gridded spatial resolution. No further climate variables are required; the potential evaporation is represented by the modified Blaney-Criddle method (Schrödter et al., 1985), which only requires air temperature as input.

The method of extending the model to the Eastern domain followed the strategy outlined by Reszler et al. (2006, 2008a). This approach contains several steps for parameter identification based on the Dominant Processes Concept (e.g., Grayson and Blöschl, 2002), and proposes the usage of auxiliary information and data (e.g., field surveys, snow depths, hydrogeological data) and the stratification into different event types (convective, advective and snow melt events). Spatially distributed information is incorporated in a GIS framework, but the resulting hydrotope structure is manually fine-tuned. The following hydrotope types were chosen (compare to Reszler et al., 2006): urban areas, low density urban areas, steep slopes open, steep slopes forest, flat agricultural areas with porous aquifer, saturation areas and karstic areas. Hydrotope structure and parameter values are chosen in consistency with the existing model in Western Styria, where in some catchments (e.g., at the foothills of the Koralpe massif) the physiographic situation is similar.

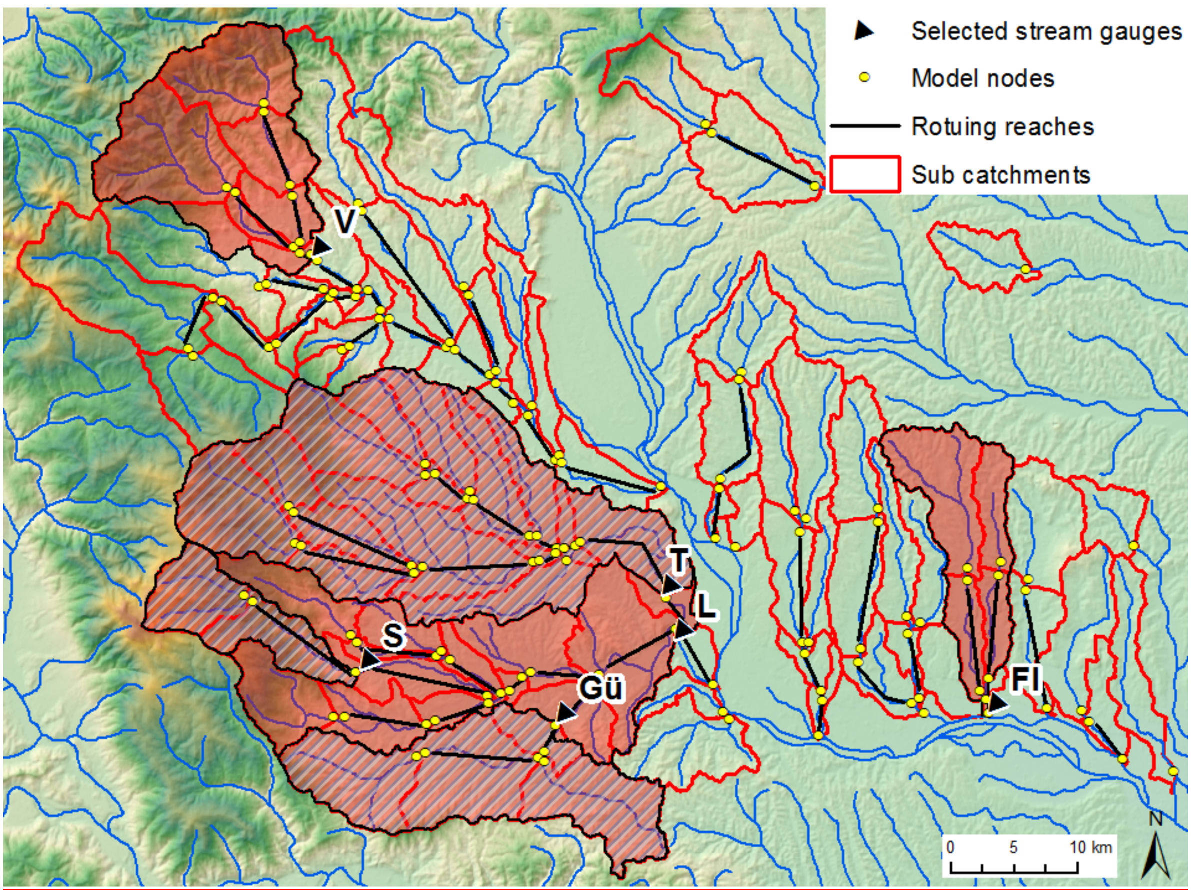

Figure 3: Spatial model structure (sub-catchments, nodes, routing reaches), available gauges for calibration and catchments of stream gauges for evaluation highlighted (nested catchments ~~in blue~~are shaded). Evaluation gauges ~~labels~~ see ~~Fig. 1~~Tab. 1.

3.4 Evaluation measures

In order to combine quantitative and qualitative evaluation of the different model simulations, the following measures are chosen:

- Catchment size as an indicator for general attenuation affects
- Frequency of floods, i.e. maximum annual floods (MAF)
- Seasonality of floods
- Other variables, such as soil moisture (simulated by the hydrological model)
- Event-based analyses (performance at particular events, event/weather types)

Catchment size is implicitly incorporated by the selection of the gauges with a wide range of catchment areas from small to medium scale (75 to 200 km²) as well as the larger catchments of the gauges downstream (< 1100 km²) (Table 1). In the evaluation plots in this paper the size is identified by the letters "L" for large, "M" for medium and "S" for small. In addition, differences between the catchments in runoff generation and response times are evaluated by different model parameters obtained by the calibration.

Frequency of floods are analysed by typical statistics of maximum annual flood peaks using the following "plotting position" (Weibull)

$$RP = \frac{(N+1)}{m} \qquad (1)$$

RP is the (empirical) return period, *N* the number of values (years) and *m* the ranking (1 for the maximum and N for the minimum flood).

The seasonality of floods gives first insights into the main hydrological drivers for flood occurrence (Parajka et al., 2010). It is the result of the relative influences of soil moisture, evaporation and snow processes and varies considerably in space. In their event type analyses, Merz and Blöschl (2003) used the seasonality of maximum annual flood (MAF) peaks as an indicator describing the timing of floods. Here, seasonality is first, analysed simply by counting MAF peaks in the four seasons December, January, February (DJF), March, April, May (MAM) June, July, August (JJA), and September, October, November (SON). Second, in order to illustrate seasonality for different simulation runs in the small multi-model ensemble, circular statistics are performed. For each event the date of occurrence of the MAF is transposed to an angle by

$$\alpha_i = D_i \frac{2\pi}{365} \qquad i = 1, \dots n \qquad (2)$$

where D_i denotes the day of the year ($D_i = 1$ for Jan. 1st, $D_i = 365$ for Dec. 31st). This angle is averaged by the following equations

$$Y = \frac{\sum_{i=1}^{n} \sin(\alpha_i)}{n}$$

$$X = \frac{\sum_{i=1}^{n} \cos(\alpha_i)}{n} \qquad (3\text{-}6)$$

$$r = \sqrt{X^2 + Y^2}$$

$$\theta_r = \arctan\left(\frac{Y}{X}\right)$$

where X and Y are the rectangular coordinates of the mean angle Θ_r, and r is the mean vector length, which is a measure of strengths of the seasonality ($r = 1$ if all events occur at the same date). Note that the final resulting mean angle depends on the quadrant of the calculated mean angle.

Using a hydrological model for an evaluation of climate model results also enables the incorporation of other hydrological quantities, which give indications about the performance of the climate model regarding the hydrological conditions. Soil moisture is an important variable to be analysed in terms of non-linearity and threshold processes in flood generation (e.g., Penna et al., 2011). It is continuously calculated by the hydrological model, and hence, can be used as a comparison between the different simulation runs.

At last, mainly using the 3 km convection permitting RCM results, runoff simulations at characteristic events are checked for their realistic event evolution and the plausibility of the corresponding atmospheric and hydrological conditions.

4 Added value in RCMs due to increased resolution

In order to demonstrate added value due to a reduction in the model grid spacing, we derived averaged precipitation fields of the models and the observational data ~~in time~~ and calculate the spatial correlation coefficient between them. ~~Prior to bias correction, it is necessary to know whether increasing the model resolution results in added value. Figure 4 shows an example of the spatial variability, on average, of precipitation for observations and the uncorrected WRF RCM. As the resolution of WRF increases, the model more closely simulates the observed precipitation patterns. The closeness of these fields can be evaluated using correlation coefficients, which~~ Figure 4 illustrates the resultant correlation coefficients for all models, months and hours of the day. ~~The correlation coefficients between the observations and each of the three WRF resolutions in the example of Fig. 4 correspond to the upper left coloured squares in each of the bottom row's subplots in Fig. 5.~~ Higher correlations for both models, illustrated by the warmer colours, are more clearly observed towards the left side of Fig. 4, the side where highest model resolutions are depicted. This shows that the RCMs improve, on average, in their ability to simulate precipitation fields across space as the resolution of the model increases.

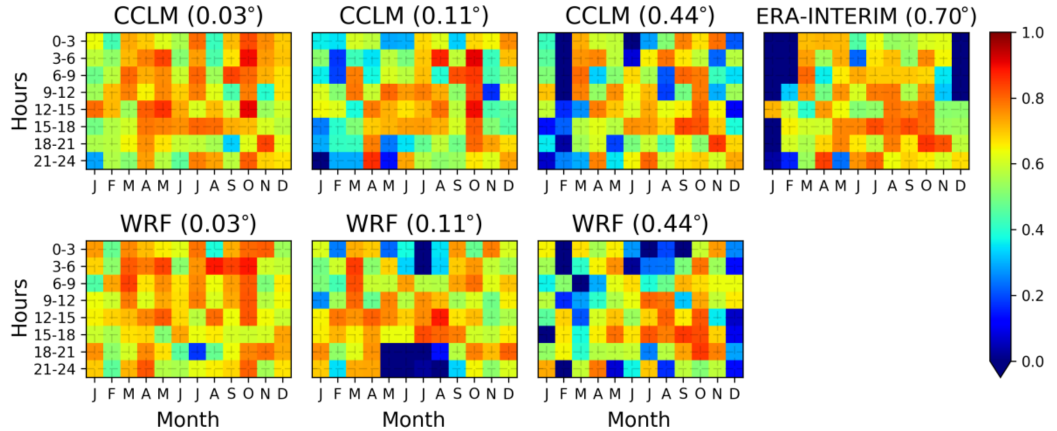

Figure 4: Added value of using higher model resolution. The colour bar corresponds to the correlation coefficients between the observed and the modelled spatial fields of averaged precipitation. The x-axes and the y-axes show the months of the year and the hours of the day, respectively.

Added value is also seen on catchment averaged quantities. Generally, the convection-permitting models increase precipitation intensities from heavy (>90th percentile) precipitation events in all catchments. In the case of CCLM, this results in added value (together with some overestimation), since the coarser resolved counterpart CCLM 0.11° largely underestimates (in the range of -16% to -26% across the catchments) precipitation intensities on average (see Figure 5). In contrast, WRF does not show such strong underestimations in the 0.11° simulation and WRF 0.03° gives an overestimation, because of its linkage to the coarser resolved 0.11° simulation that enables error propagation (Addor et al., 2016). The reasons for the enhancement of intensities in mountainous regions may be a result of the higher resolved orography and is in agreement with previous evaluation studies (Knist et al., 2018; Prein et al., 2013; Ban et al., 2014; Langhans et al., 2013). Note, WRF 0.11° is generally in a better agreement with the observations than CCLM 0.11° (Figure 5), although both simulations are of comparable performance on the European domain (e.g., Katragkou et al. 2015; Kotlarski et al. 2014) and add value in mountainous regions compared to their 0.44° counterparts (Prein et al., 2016).

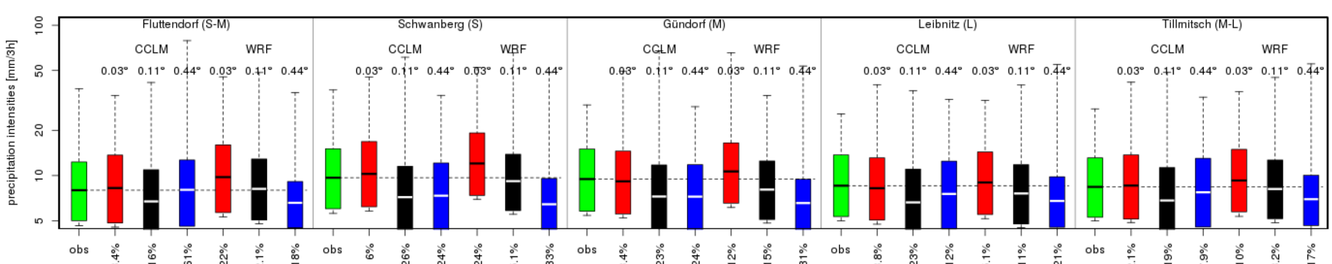

Figure 5: Catchment averaged heavy (>90th percentile) precipitation intensities of observations (green) and CCLM and WRF with 0.03° (3 km; red), 0.11° (12.5 km; black), and 0.44° (50 km; blue) grid spacing. Y-axis has a logarithmic scale. Coloured boxes indicate the 10th – 90th inter-quantile range, horizontal markers in the boxes denote mean values, whiskers refer to maximum values. Relative biases of mean values are given along the x-axis.

5 Hydrological model calibration and validation

The hydrological model was calibrated, for each sub-catchment, against runoff data of all available stream gauges in the period 2000-2009 (Fig. 1). Calibration results in Western Styria are available, and the found parameters in catchments with similar soil and geological properties serve as a priori values for the catchments in the extended part. The historical data in the current study (1989-1999) are used for model validation. This allows also a validation of the existing model; these data were provided for the current study and had not been used for model calibration. Quantitative metrics such as the commonly used Nash-Sutcliffe-Efficiency (NSE, Nash and Sutcliffe 1970), the BIAS based on mean runoff values, and the Root Mean Square Error (RMSE) are used to measure model calibration. In Table 3 the results for the selected gauges are listed. As it is often the case, NSE is lowest in the smaller catchments, e.g., Schwanberg and Fluttendorf with 0.77 and 0.78 in the calibration period, respectively. In the validation period the NSE falls below 0.7 in these two catchments. The historical period also includes phases with poor data availability (see Fig. 1), which is also the reason for the drop of the NSE value in the validation period at the Gündorf gauge.

Examples of hydrographs in the calibration and validation period are attached in supplementary material (Fig. S1 and Fig. S2). In addition to flood peaks, runoff generation and rainfall response is represented very well. Differences in the shape of hydrographs are also accurately simulated. For example, the Schwanberg gauge shows short peaks due to short concentration times in the small catchment but at the same time high baseflow. The latter indicates high fraction of slowly draining flow component (groundwater) from long term storage. On the other hand, in the medium sized catchment Gündorf, short peaks also indicate short response times, but baseflow is significantly lower. This difference can be attributed to the different geologic conditions in the area. In the Schwanberg catchment the significant subsurface storage can be attributed to a deep weathering zone overlaying schists and gneiss, and geology in the Gündorf catchment consists mainly of tertiary material (silt, loam) with very low storage capacities. In the larger catchments flood peaks are smooth lasting over several hours which shows the attenuation effects (Tillmitsch, Leibnitz). The resulting model parameter values representing time scales of runoff response show the flashy character in the catchments: The time constant of the fast flow component, i.e. surface runoff in open steep slopes (k_0), at hillslope scale is in the order of simulation time step of 1 hour, routing within the sub-catchments (30 - 50 km²) is in the same order, and travel time within most river reaches connecting the nodes is 2-3 hours.

For this study, representation of flood frequency is important. Model simulated maximum annual floods for the entire study period (calibration and validation period combined, 1989-2010, 22 years) are compared with observed flood peaks in Fig. 6 (flood frequency plot). Although the MAF distribution was not explicitly subject to calibration, and the data availability was relatively poor in the period 1990-2000, the model accurately simulates observed flood statistics at the selected gauges. The largest flood is simulated well at all gauges, while simulation results at the smaller events are reasonable. Both in the calibration and validation period, deviations at significant events are analysed in terms of probable errors in input (precipitation), model structure, model parameters and/or runoff data. At the exceptional events threshold processes are operative, which are accurately simulated. For example, at the largest flood at the Schwanberg gauge in August 2005 (Figure

6, plot above left, extrapolated RP would be more than 100 years) was a very local event (see following Fig. 9 lowest panel) and the interpolated rainfall is assumed to be relative uncertain. In order to simulate the observed flood peak parameters would be needed which are not plausible and decrease model performance at other large events. Also inundation occurred during the event in the Schwanberg town, which likely led to uncertainties in the observed peak runoff data. At the Fluttendorf/Gnasbach gauge, flooding occurred at the two events in 2009 (see Figure 15), and retention by inundation in flood plains was calibrated successfully in the flood routing model. At the Voitsberg gauge the two largest floods were slightly underestimated. The largest flood occurred in the October 1993, within the validation period, which was underestimated in the simulation. The largest flood in the simulation is the 2009 flood which was represented very well (see supplementary material). Data quality used to be poor in 1993 in the high altitude catchment in the North-Western part. Station density in this part today is still lower than for example, in the South-Western part (see Fig. 1). Same situation can be stated for the Tillmitsch gauge. At this gauge, four medium event peaks (from 92 to 117 m³/s) are the MAFs in the years 1993 – 1996, and the simulated flood peaks at the corresponding plotting positions were slightly underestimated.

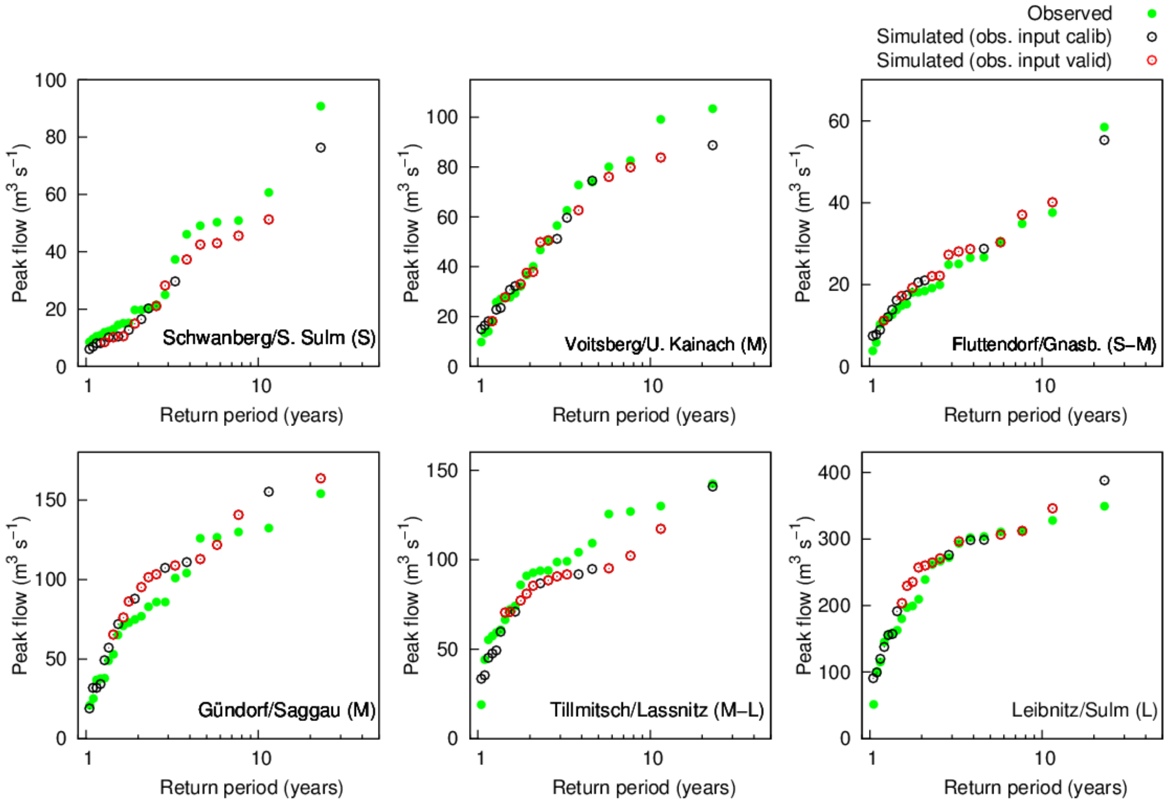

Figure 6: Simulated and observed maximum annual flood peaks vs. empirical return periods (Eq. (1), flood frequency plots) of the selected gauges in the period 1989-2010. The peaks in the validation period are marked with red colour.

6 Evaluation of simulation results using RCM data as input

With the calibrated hydrological model, simulations are performed using the results of the RCMs as input. In the following sections 6.1 and 6.2 evaluation results are discussed in detail for the runoff simulations using the CCLM results. The same procedure has been applied using the WRF results (provided in supplementary material). In a synthesis step (section 6.3), all the results of the small multi-model ensemble are summarized and compared for formulating final conclusions.

6.1 Uncorrected RCM data

Flood frequency plots for the selected gauges using uncorrected CCLM data (and the ERA-Interim data) as input compared to the observations are shown in Fig. 7. The figure illustrates the improvement of the results using the ~~hourly~~ 3 km CCLM data, particularly for the smallest catchment Schwanberg/S. Sulm with a size of app. 75 km². ~~This indicates that at~~ In this ~~spatial scale~~specific region, the convection permitting simulation ~~is~~ seems to be ~~absolutely~~ a ~~necessary~~ necessity in order to accurately represent the magnitude of floods. In some larger catchments the simulations with the coarser RCM data already yield reasonable results. As the plotting positions suggest, the statistical properties, mean and standard deviation, decrease with increasing grid size. The skewness does not decrease; in Voitsberg and Schwanberg the CCLM 0.11° simulation yields a high skewness. At the latter only three events are simulated with peak flows above 20 m³/s, whereas the observation reaches a maximum of 90 m³/s.

Most of the RCM settings show negative biases regarding MAF peaks; however, some are significantly positively biased, e.g., Voitsberg/Kainach in the North-Western Alpine part ~~(see Fig. 4).~~ At the Fluttendorf gauge (upper right sub-plot) the 0.44° data lead to a maximum flood that significantly exceeds the observations; this peak is the consequence of overestimated heavy rainfall intensities (app. 300 mm in 18 hours), which was simulated in August 2005 during the “Alpenhochwasser.” For comparison, the observed MAF at the Fluttendorf gauge occurred during an extreme precipitation event in June 2009 (Figure 15 in the following section).

The seasonal occurrence (winter: DJF, spring: MAM, summer: JJA and autumn: SON) of the simulated MAFs is analysed in Fig. 8. The improvements are evident when reducing grid size; the simulation with the uncorrected 3 km CCLM data represents the observed seasonality very well. The figure further shows that both the CCLM with 0.44° (~50 km) and 0.11° (~12.5 km) grid sizes yield a shift of the flood season from summer (JJA) to spring (MAM) in all catchments - except Schwanberg. Also, in the catchments in Western Styria (Kainach, Sulm, Saggau, Lassnitz) numbers of MAF in autumn are underestimated in all CCLM settings. In autumn, frequently occurring low pressure systems in the Mediterranean or east of the Alpine region induce heavy rainfall which can often lead to large floods (Seibert et al., 2007). The simulations indicate that this is underrepresented in all CCLM data. This shows the value of the use of seasonality for an evaluation of an accurate representation of the main flood generating mechanisms. Flood statistics in the mentioned cases yielded reasonable results, but this criterion alone could be misleading. In the catchment in Eastern Styria (Gnasbach) the relatively uniform distribution is captured well (upper right sub-plot). Both for flood frequency and seasonality, using the WRF data the results are worse than using the CCLM data (shown in supplementary material, Fig. S3 and Fig. S4, and discussed later in section 6.3).

Simulated soil moisture on monthly basis (attached in supplementary material, Fig. S5 above) shows annual dynamics that are similar to the seasonality of MAF. Also, the improvements using the 3 km CCLM (0.03°) compared to the coarser resolution are evident, and ERA-Interim is closest to the reference. Within ERA-Interim the observed situation is represented, however, the coarse resolution also leads to a bias. Underestimation in summer is significant, particularly in the case of the 0.44° (~50 km) and 0.11° (~12.5 km) grid sizes. In this season heavy storms occur with often convective character or double events. The corresponding flood magnitude is (non-linearly) controlled by the antecedent soil moisture, which is the consequence of the meteorological and hydrological history prior to the flood events. The same is true for the autumn (SON in Figure 8), when the soil moisture is underestimated and often floods occur as a consequence of Mediterranean low pressure systems in combination with wet soils due to reduced evapotranspiration.

As the first results show, the CCLM 3 km setting yields a clear benefit regarding magnitude and frequency of large floods particularly in small catchments. As stated above, the largest floods of the simulations are not necessarily aligned in time with observations. Figure 9 shows two simulation periods for the gauge Schwanberg/S. Sulm (75 km²). In the panel above the period August and September 1996 is plotted, when the largest flood was simulated with the 3 km CCLM input. There are several small rainfall events in the observation but no large flood occurred during this period. The panel below shows the largest flood in the record which occurred in August 2005 ("Alpenhochwasser"). This flood was completely missed using the CCLM input. As it happens, the size and the month of occurrence of the two simulated floods is the same. Also, in 1996 the temporal rainfall distribution simulated by the climate model, showing a very high one hour block embedded into a slight pre- and post-rainfall, leads to a plausible shape of the hydrograph. This example indicates that also for small catchments, large floods are "produced" which leads to a rather good statistical representation of maximum annual flood peaks (see Fig. 6), but an event-by-event comparison partly fails, because of the large RCM domain: the 3 km CCLM simulation is driven by a 12.5 km CCLM simulation covering the European domain that is in turn driven by re-analysis data (ERA-Interim) without nudging. Due to the internal model variability, the 12.5 km simulation partly deviates from its driving data (ERA-Interim) and even synoptically forced events (like the 2005 flood) may not be correctly represented in space and time in the RCM. This decoupling-effect is well known in Regional Climate Modelling and Numerical Weather Prediction and was first published by Kida et al. (1991). Along the modelling chain, This affects also the convection-permitting 3 km simulation is affected by decoupling for two times: (1) via the 12.5 km domain that is partly decoupled on the synoptic scale and (2) via its own internal variability, where so that single thunderstorms (with under weak synoptic forcing) may occur at different places or/and at different time as in the observations ("double penalty" problem). This limits the applicability of event-by-event comparisons and emphasises statistical evaluation approaches.

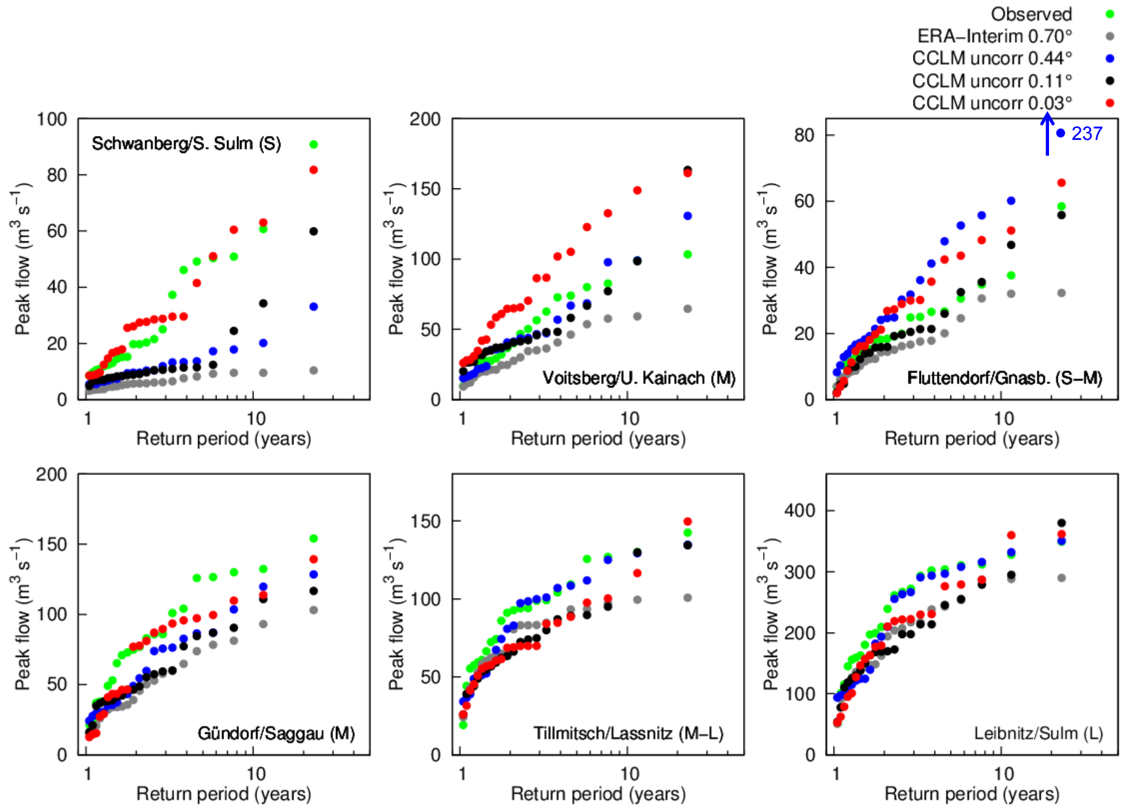

Figure 7: Simulated maximum annual flood peaks using raw CCLM data as input and observed maximum annual flood peaks vs. empirical return periods (Eq. (1), flood frequency plots) of the selected gauges in the period 1989-2010.

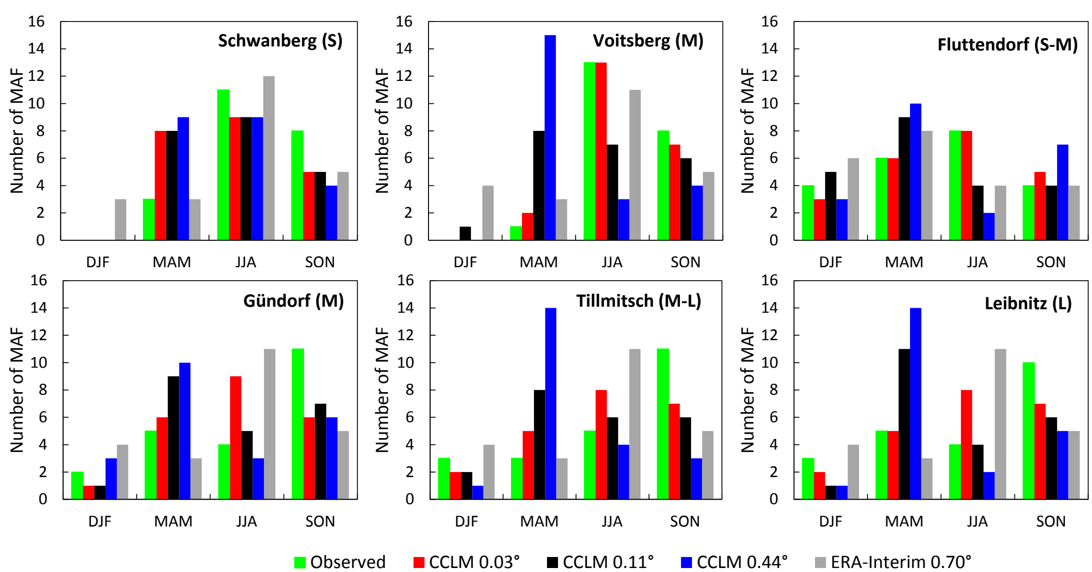

Figure 8: Number of maximum annual floods in the four seasons (seasonality) from the simulation using raw CCLM data as input compared to the observation at the selected gauges in the period 1989-2010.

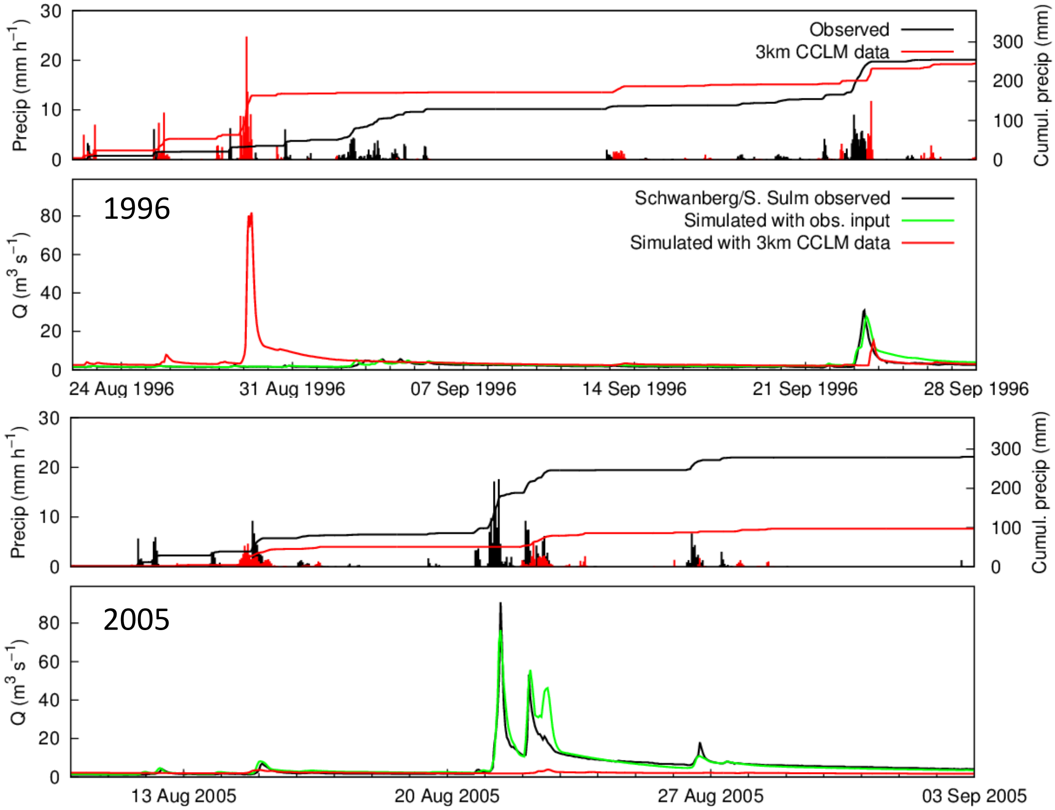

Figure 9: Example of two events (Aug. 1996, above and Aug. 2005 below, in each case plotted with catchment precipitation above the runoff) simulated with raw CCLM 3 km data compared to the simulation with observed input and observed runoff data for the Schwanberg gauge.

6.2 Bias corrected RCM data

In the same way as for the raw RCM data, the hydrological model is driven using bias corrected data. After bias correction results of flood statistics using CCLM (Figure 13) are improved, except for the smallest catchment Schwanberg. Here, particularly the results deteriorate compared to the run using the uncorrected data (Fig. 6). ~~This is due to the 3 hour time step after the bias correction. The short response time in the 75 km² catchment requires a one hour time step to accurately represent runoff response to short intensive rainfall events. These events are characterized by a steep rise of the hydrograph and short peaks and can induce significant floods at this spatial scale. The aggregated three hour time step chosen in the bias correction has two reasons. First, a relatively smooth transition of the diurnal cycle is required, where fitting distributions to one hourly data is too noisy. Second, all the models should be compared under a common framework (some models were at a 1 hour temporal resolution while others were at a 3 hour resolution). Then, differences or improvements between models can be attributed to the spatial scale without the influence of varying temporal time steps.~~

This can be explained by an interference of the temporal distribution of precipitation intensities during the flood generating rainfall events and the bias correction that simply ignores such temporal relationships. Figure 10 shows the precipitation intensities that contributed to the maximum annual flood events in Schwanberg simulated by CCLM 3km, before and after bias correction. Each event is limited to a duration of two days before the maximum peak flow is reached. Figure 10a demonstrates the work of the bias correction that removes severe under (over) estimation of low (high) intensities in the CCLM 3km data, but leaves the total amount of precipitation of these events largely unaffected so that a median under catchment of -15% to -16% remains (Figure 10b). The success of CCLM 3km in capturing the flood events (Figure 13) lies in the precipitation amount that is accumulated over a shorter time period prior to the flood events. Figure 10c shows the averaged relative bias of accumulated precipitation as a function of the accumulation time prior to the event. On average, CCLM 3km increasingly overestimates the accumulated precipitation as the accumulation time is shortened. The lack of total precipitation is compensated by the temporal evolution that gives about 20% more precipitation within a time range of 24 h before the flood event. The bias correction removes these (compensating) overestimated intensities and the positioning of the peak flows (Figure 13) rapidly drops. Note, the reason why single intensities are not properly corrected is based on the fact that the bias correction is independently applied on each grid cell. The remaining deviations from the observations (Figure 10a) result from the aggregation of single grid cells to areas that cover the entire catchment.

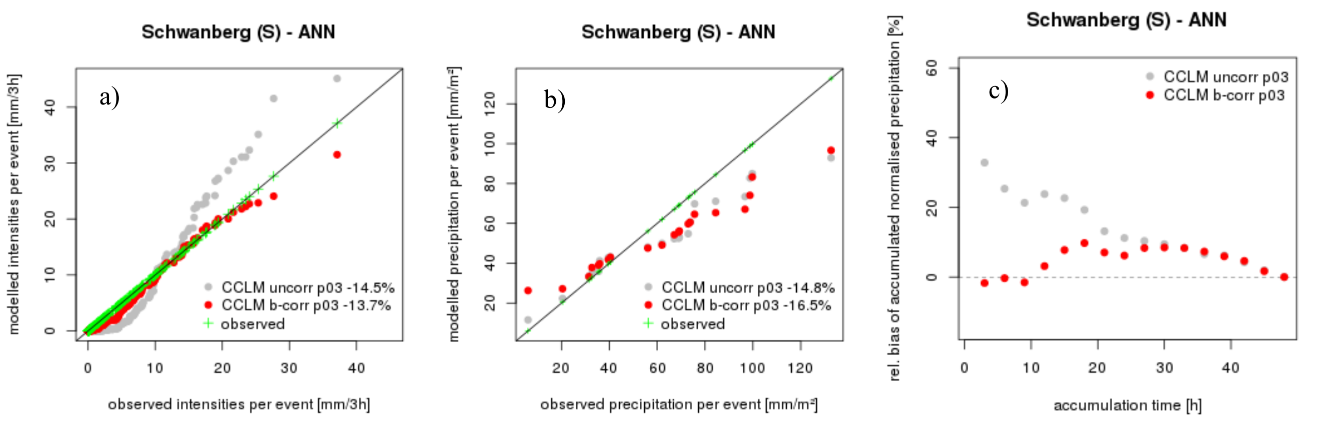

Figure 10: Comparison between modelled (CCLM 3km), bias corrected, and observed precipitation characteristics tied to the 22 MAF events in the catchment Schwanberg. The left panel shows the single precipitation intensities. The middle panel depicts the total precipitation per event (defined as the 2 day period before the maximum peak flow). The right panel shows the (event averaged) relative bias of the accumulated precipitation amount (normalised by the events' total precipitation) prior to the flood events as a function of its accumulation time. Numbers in the legends give the relative median bias of the plotted data.

In contrast, the positioning of WRF 3km peak flows in Schwanberg lies above the observations and the bias correction leads to a deterioration (Fig. S3). In this case, WRF 3km overestimates precipitation intensities across the flood events and the bias correction changes this (due to the aggregation of single grid cells to catchments) into an underestimation (Figure 11a). This leads to an overestimation (underestimation) of event-related precipitation amounts (Figure 11b) for the uncorrected

(corrected) data. In WRF 3km the temporal distribution of the intensities is in a much better agreement with the observations than in CCLM 3km (compare Figure 11c and Figure 10c). However, since the total amount is overestimated, the peak flows are higher. The bias correction furtherly deteriorates the temporal distribution of the intensities that lie closer to the flood event and together with the underestimation of the total amount this gives a rapid drop in the positioning of the peak flows (Fig. S3). Note, this good representation of the temporal distribution in WRF 3km is a catchment specific feature.

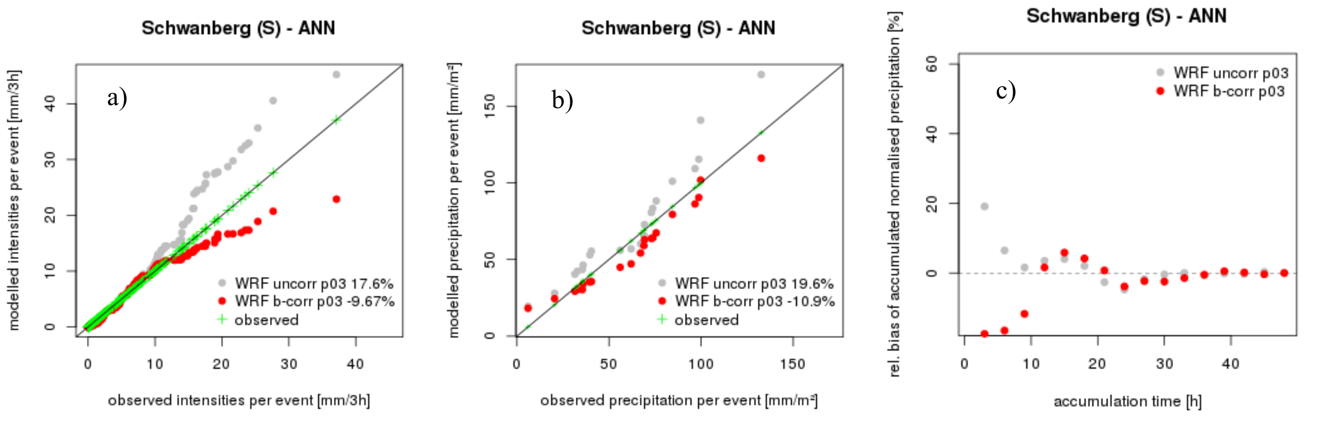

Figure 11: Same as Figure 10, but for WRF 3km.

Also, in the small catchments, the aggregation to 3hr sums has an influence on the performance. We tested it by using the 3hr sums of the CCLM 3km and comparing to the 1 hr results (not shown). There is a decrease of flood peaks, but the main decrease in performance in the small catchment Schwanberg is due to the error correction explained above.

In some cases bias correction leads to an over-compensation of the flood peaks, particularly in the case of the ERA-Interim data. For instance in Gündorf, flood event related precipitation intensities and amounts are largely underestimated in ERA-Interim by more than -30% on average (median) (Figure 12ab), but the precipitation amount within an time range of ~3/4 day before the flood event is over estimated by ~25% on average (Figure 12c). However, this overestimation is too small and the peak flows of the corresponding flood events (Figure 13) stay below the observations. The bias correction overcorrects the catchment averaged intensities that are larger than ~7 mm/3h (Figure 12a) and leaves smaller intensities undercorrected (as an effect of catchment aggregation) which however yields to a well representation of the precipitation amounts (Figure 12b). The overcorrection of higher intensities leads to a further increase of the accumulated precipitation amount 3/4 day prior to the flood events and the corresponding positioning of the peak flows (Figure 13) lie above the observations in general.

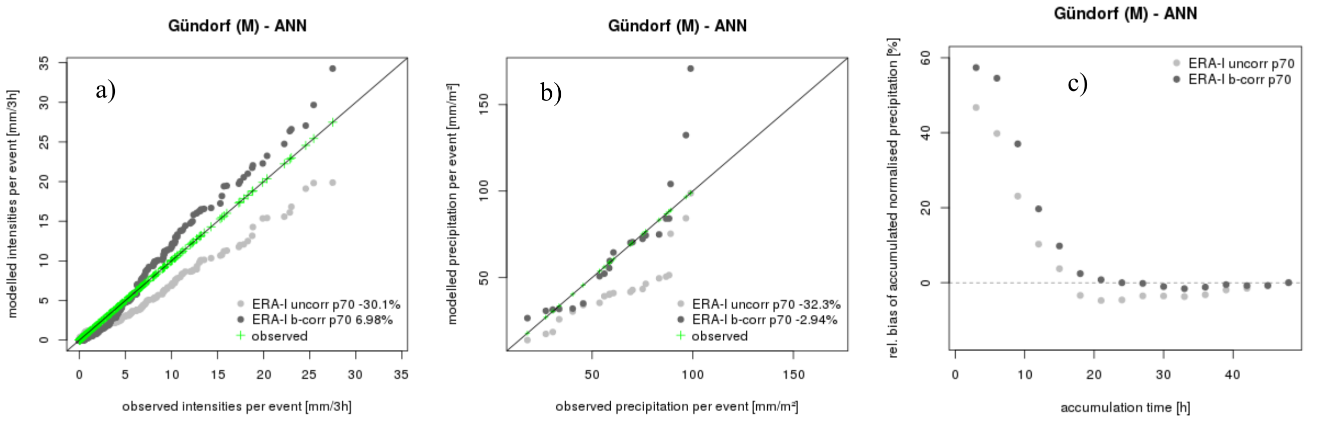

Figure 12: Same as Figure 10, but for ERA-Interim in the catchment Gündorf.

From a return period of 6-10 years the flood simulations are very sensitive to overestimations (e.g., gauges Voitsberg and Gündorf in Figure 13) and underestimations (see Fig. 7) of the simulated rainfall, which is due to the non-linearity in the rainfall-runoff process (e.g., Komma et al., 2007; Rogger et al., 2012). This threshold is consistent with usual concepts in hydrology, such as the concept of the GRADEX method (e.g., Merz et al., 1999). At this size of floods the soils have been saturated by a high amount of precipitation and 100% of the subsequent rainfall comes to runoff. This is vital to take into account when it comes to correct high rainfall intensities within the bias correction procedure.

Seasonal occurrence is improved for all CCLM settings after bias correction (Figure 14). In particular, the shift from summer to spring using the raw 0.11° and 0.44° data is removed. Again, the 3 km data yield results closest to the observed distribution. Again, results using the bias corrected WRF data as input are incorporated into discussion in the synthesis step in the following section.

As for the seasonality, the seasonal shift in the simulated soil moisture is removed after bias correction, but the underestimation in summer and autumn cannot be entirely compensated (see supplementary material, Fig. S5 below). This can be attributed to the fact that the modelled events are different in size, shape, and overall structure to those of observations. The SDM methodology is performed independently for each grid cell, and as a result is not imposing the structure of typical broad-scale observed weather events. Therefore, even though the distributions of bias corrected precipitation align to observations at individual grid cells, the average precipitation amounts across multiple grid cells can differ from observations. ERA-Interim results now lie exactly on the observation. However, for the MAF performance using ERA-Interim data is not sufficient (compare Figure 13). This shows that using observed atmospheric conditions with large grid size (~70 km) is able to reproduce mean monthly hydrological conditions, but fails in flood event representation on this scale. Out of the CCLM data, performance using the 3 km data is still best, and underestimation using the 0.11° and 0.44° data in summer is still evident in all catchments.

For an event based illustration of the effect of bias correction two events in 2009 at the Fluttendorf/Gnasbach gauge were chosen using the 3 km CCLM data as input (Figure 15). The first event in June is the largest in the series and the second event in August is the second largest in the series. Synoptic forcing is different between the two events: the first event is controlled by a persistent upper-air cut-off low that is located over the Balkan region and brings warm and moist air towards the Eastern Alpine region from the East (Godina and Müller, 2009). This led to floods in the whole southern Styrian region~~he first event is controlled by large scale frontal processes, leading to floods in the whole southern Styria region~~, whereas the second flood is mainly driven by convective processes and concentrated on the eastern part. For the first event, the model with the uncorrected 3 km CCLM data simulates an event with the same order of magnitude, but slightly different timing, as the observation. After the bias correction flood peak is decreased due to a general reduction of precipitation in the bias correction in this period. A reduction of rainfall in this period is resulting from the bias correction as a consequence of the overestimation of the MAFs by raw CCLM data (compare Fig. 7, upper right sub-plot). However, after bias correction, this is still the largest flood peak in the series (see Figure 13, upper right sub-plot). The second event is completely missed by the simulation run with the raw climate model data. No significant rainfall is simulated in the RCM and hence, bias correction is totally ineffective. It is clear, that at such missed events there is no possibility to correct raw RCM data by any statistical bias correction method. Bias correction is not able to compensate general uncertainties in representing convective situations. Note that bias corrected intensities in upper panel are aggregated three hour sums.

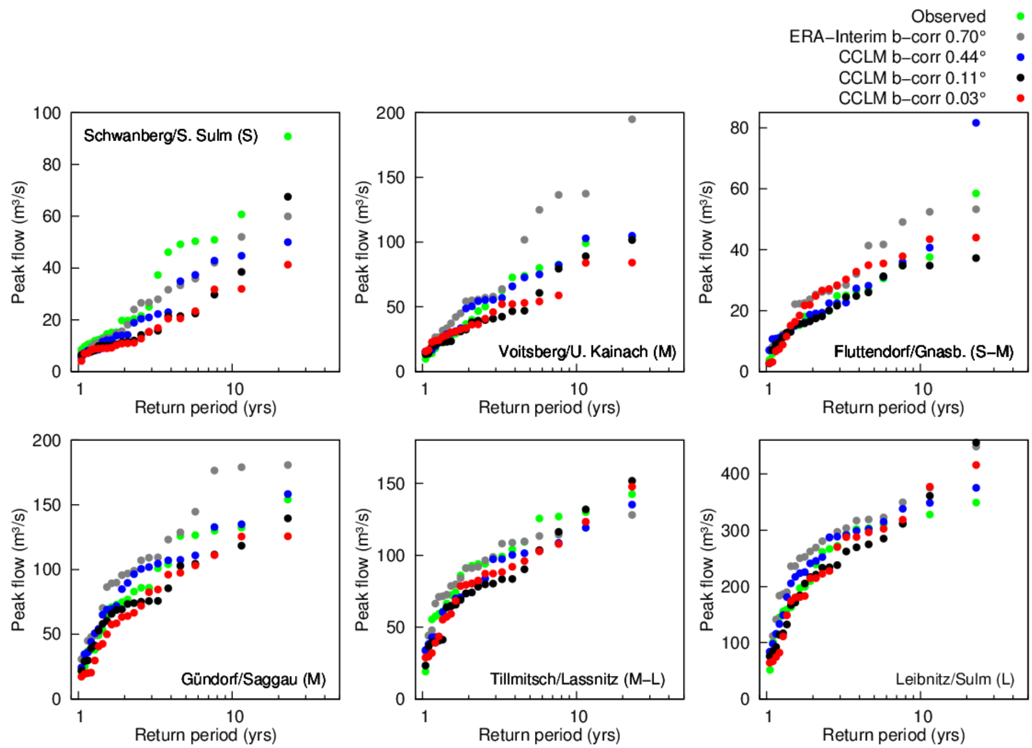

Figure 13: Simulated maximum annual flood peaks using bias corrected CCLM data as input and observed maximum annual flood peaks vs. empirical return periods (Eq. (1), flood frequency plots) of the selected gauges in the period 1989-2010.

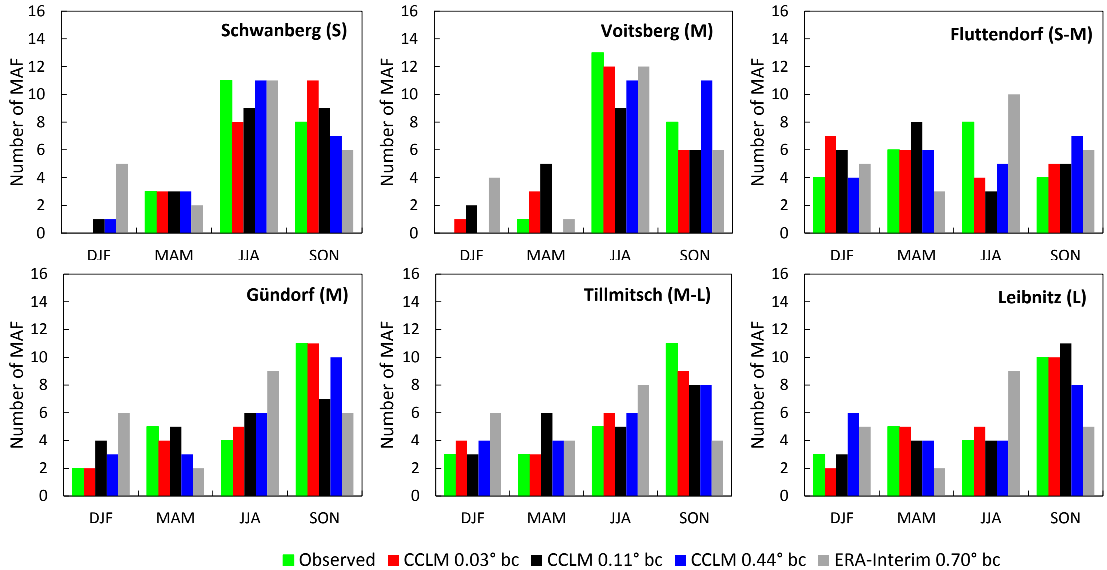

Figure 14: Number of maximum annual floods in the four seasons (seasonality) from the simulation using bias corrected CCLM data as input compared to the observation at the selected gauges in the period 1989-2010.

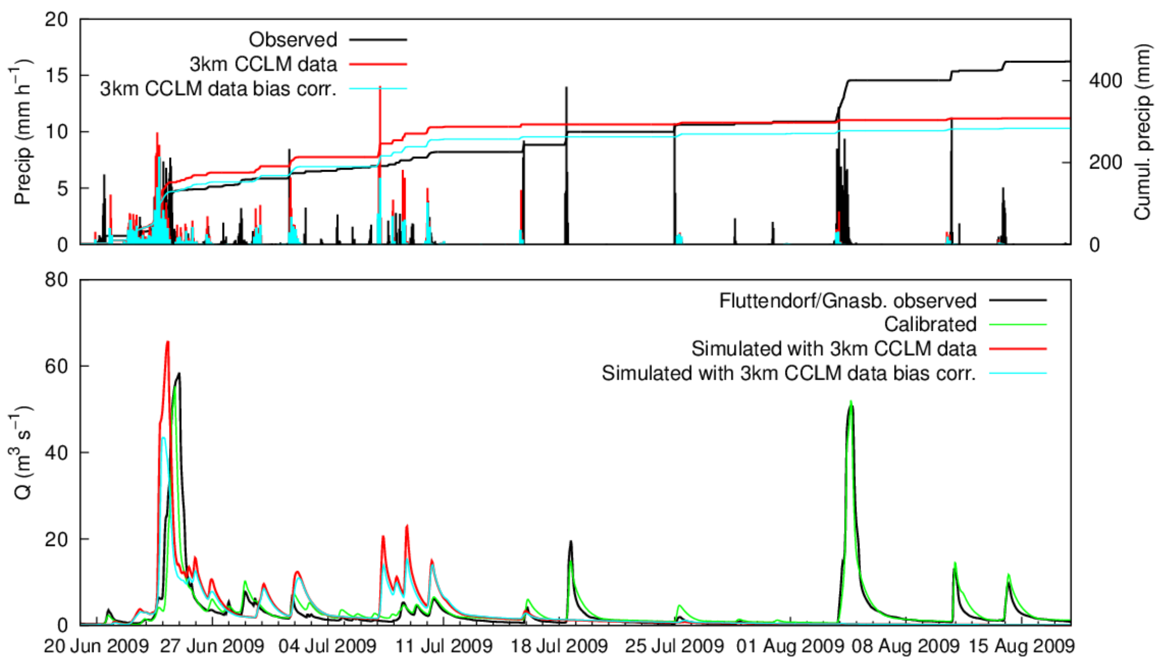

Figure 15: Runoff simulated with the uncorrected (1h rainfall sums) and bias corrected (3h rainfall sums) 3km CCLM data for the period with the largest floods at the gauge Fluttendorf/Gnasbach. Above: catchment precipitation.

6.3 Synthesis

The statistical measures of mean, standard deviation and skewness, for the 22 years sample of maximum annual floods resulting from the 14 different variants are illustrated in Figure 16Fig. 16. The mean (left plot column) and the standard deviation (middle plot column) are related to the catchment area in order to compare these measures between the gauges. Results using the ERA-Interim data are plotted in the centre, and the results using the different RCM settings with decreasing grid size are plotted towards the left (CCLM) and the right (WRF). The values with raw RCM data as input are plotted as black points; the values with bias corrected RCM data as an input are plotted as red points. The observed measures are indicated with a thin horizontal line for each gauge. The figure first clearly shows the decrease of mean specific runoff peaks and – in connection to this – the specific standard deviation with the catchment sizes (S to L from above) for all variants. This is the mainly the consequence of a decrease of mean areal precipitation for large rainfall intensities and short durations (e.g., Hershfield, 1961; Lorenz and Skoda, 2000) but also of attenuation effects through flood routing. As discussed in the previous section, in most of the CCLM data driven simulations the statistical properties are improved reducing the grid size (black points) and further improved after bias correction (red points). For the larger catchments Tillmitsch and Leibnitz, the differences between the model variants are small, which, again, indicates the good performance of the coarser RCMs regarding general flood statistics (particularly CCLM). This improvement is not always the case for the WRF driven runs. Particularly large biases from the uncorrected run are either not compensated (e.g., WRF 0.44° for Schwanberg) or even over-compensated after bias correction (e.g., WRF 0.03° for Schwanberg and Voitsberg). The 3 km WRF produces in some periods unrealistic high rainfall intensities over several time steps, which leads to exceptional high flood peaks in the simulation. Examples are the very high values for the skewness (right plot column) at Gündorf and Voitsberg gauge. This high skewness could sometimes not be compensated after bias correction, e.g., Voitsberg gauge. At this point, we have to mention that the complex interplay between single precipitation intensities and their contribution to the total rainfall amount per flood event via their temporal distribution is not correctly represented by any of the investigated RCM or ERA-Interim – even if resultant flood events show proper statistics. This specifically holds for the convection-permitting simulations. Moreover, the bias correction method is not able to correct displacements in this interplay per construction.

In order to summarize performance of the small multi-model ensemble regarding seasonality, the following Figure 17 shows the results applying Eq. (2) and Eq. (3-6) on the simulated MAFs using the different RCM data, raw (above) and after bias correction (below). The observation is plotted with a green filled square. As discussed in section 6.1, the results illustrate again the improvement of the seasonality using the 3 km CCLM data (full red squares) compared to the simulations with the coarser CCLM data for all gauges. For example, the highest concentration of timing, i.e. length of vector, of floods in a year in Voitsberg is represented well by the raw 3 km CCLM (upper middle sub-plot). However, this supreme result of CCLM 3km is the result of compensating errors: the complex interplay between single precipitation intensities and their temporal distribution during flood generating rainfall events is not correctly represented (section 6.2). Either the total precipitation amount is properly captured but the temporal distribution is failed or vice versa. This also holds for the other RCM simulations, including WRF 3km, and ERA-Interim. In addition, the bias correction method is not able to correct displacements in this complex interplay per construction and hence.

Using the coarser RCMs, both the timing and strength of seasonality of MAFs deviate significantly from the observations in all catchments. Moreover, the scatter between the different settings is large. However, to some extent all CCLM settings represent the weak seasonality in the Eastern part (Fluttendorf catchment, upper right sub-plot). As for the flood statistics, most of the WRF settings (white filled squares) yield results, which are significantly worse than the CCLM settings. The convection-permitting WRF 3 km setting does not provide any improvements compared to the coarser resolutions. Timing of MAFs tends to be concentrated in May/June for all catchments, whereas flood events occur mainly from July to September. This indicates that more or less all WRF settings fail in representing the general mechanisms for flood generation in this area and at this scale. Mostly, discrepancies can be compensated by the bias correction in the CCLM case, but not for the WRF case. In some catchments using the WRF 3 km settings the results are even worse after bias correction. For example, at the Fluttendorf gauge (upper right sub-plot in Figure 17Fig. 17 below) the concentration of timing shifts from the beginning of May (with a low strength) to February (with a relatively high strength), a month when flood generation is also influenced by snow melt processes.

CCLM, due to its long-term application in operational weather forecasts in Europe (e.g. Deutscher Wetterdienst, DWD) and in the Alpine region (e.g. MeteoSwiss) also with convection permitting resolutions, is much more properly adapted in terms of parameterisations used and their internal tunings to the Alpine domain than WRF. This can help to explain the generally poor results using the WRF 3km data, and this might induce disturbances that pile up to larger biases seen in WRF.

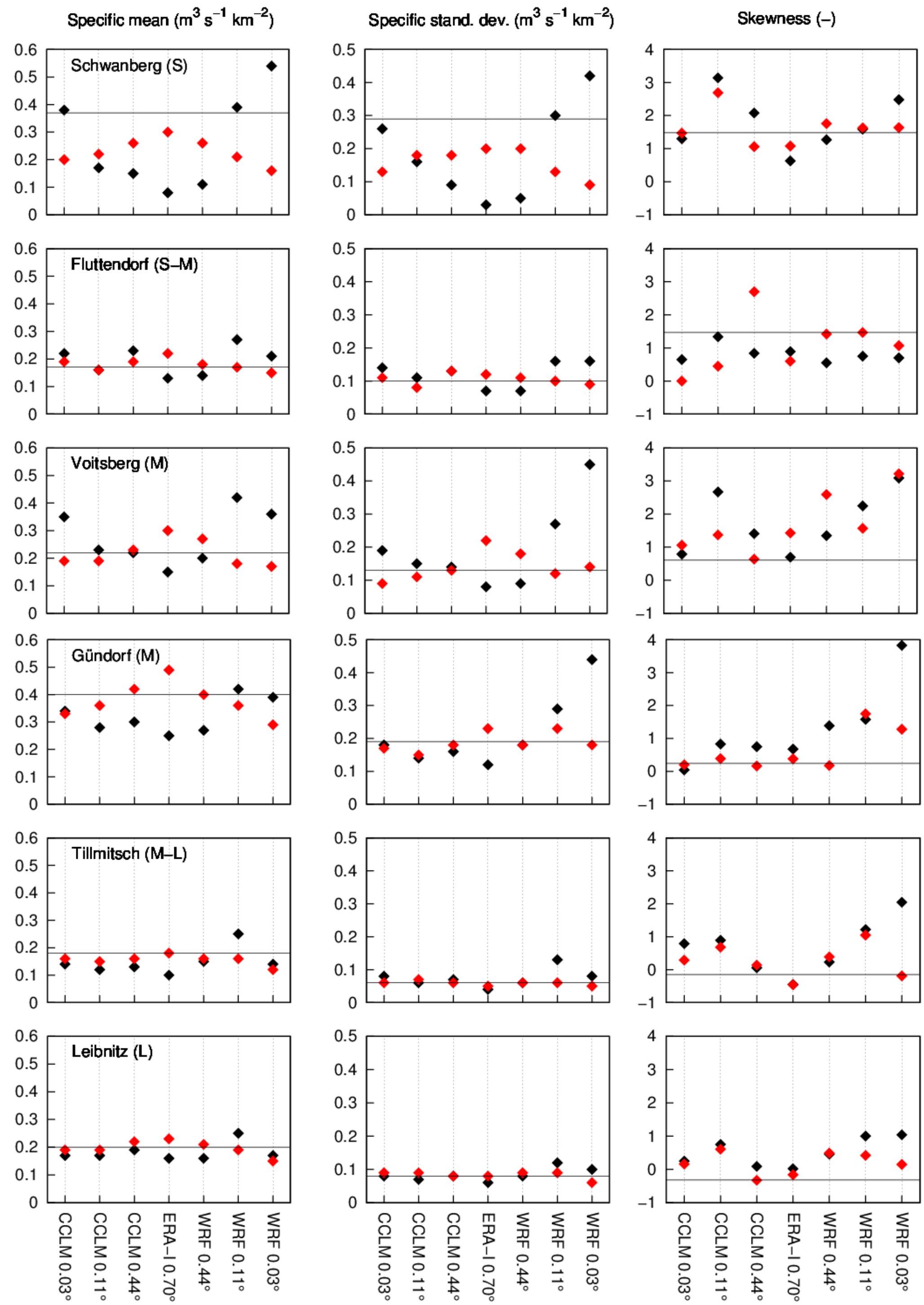

Figure 16: Statistical measures of maximum annual flood peak distribution evolving the different model runs. Column 1: specific mean, column 2: specific standard deviation, column 3: skewness. Black: raw RCM data; red: bias corrected RCM data as input. Black horizontal line denotes the values from the observed flood peak series.

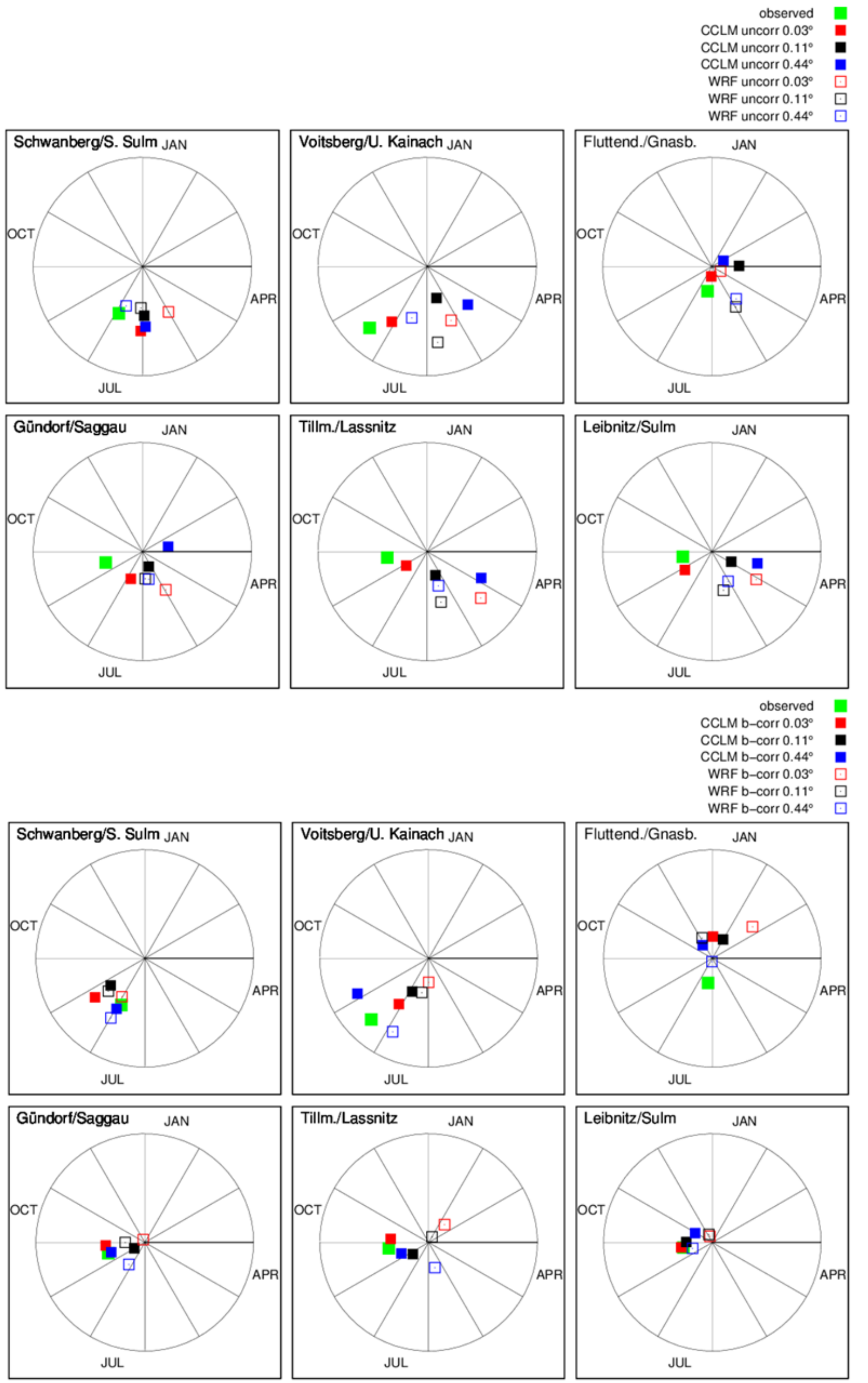

Figure 17: Results of seasonality (circular statistics of the maximum annual floods) evolving the different model runs. Above: raw RCM data as input; below: bias corrected RCM data as input. The distance from the centre is the mean vector r and a measure for the seasonality strength, i.e. concentration of timing.

7 Conclusions

This study implemented regional climate models sequentially coupled with a spatially distributed hydrological model to be used for enhanced flood modelling on small and medium spatial scales (up to app. 1000 km^2) in the Eastern Alps. The work is carried out in a small multi-model (ensemble) framework using two different RCMs (CCLM and WRF) in different grid sizes, ~50 km, ~12.5 km including two runs at convection permitting scale (~3 km). Additionally, a novel bias correction method (i.e. a modified version of quantile mapping) is applied to minimize error propagation throughout the modelling chain. Together with the driving ERA-Interim data (grid size ~70 km) the ensemble contains 14 model variants.

Evaluations using observed data in a historical period (1989-2010) ~~clearly~~ showed that in the investigated RCM ensemble, no clear added value of the usage of convection permitting RCMs for the purpose of flood modelling can be found, although CCLM 3 km outperforms in most flood statistics. This is based on the fact that flood events are the consequence of an interplay between the total precipitation amount per event and the temporal distribution of rainfall intensities on a sub-daily scale. The investigated RCM ensemble either lacks on one and/or the other. The seemingly good CCLM 3 km results in the small catchment lie on an overestimation of the intensities and underestimation of the total rainfall amount. This superposition is not systematic across the catchments. ~~the benefit of the convection-permitting simulations for all catchments, particularly in the CCLM case. Flood frequency and seasonality is represented well in all catchments.~~ From a statistical perspective, a~~A~~ll RCMs with all resolutions are able to produce precipitation rates that may cause floods in the study area. ~~However, the 3 km grid size is essential for catchments smaller than 200 km^2~~. ~~Moreover,~~ In catchments with an area less than 100 km^2 ~~require~~ a 1 hour time step due to the short response times is favourable but the influence is small. In the larger catchments, the 12.5 km and 50 km resolutions already yield satisfying results regarding flood statistics. However, with the coarser grid size the seasonality of floods, i.e. date of occurrence in a year, is not accurately represented. This indicates that some main flood generation mechanisms are not captured with the coarser models. CCLM 3km improves the seasonality of the maximum annual floods; however, in the light of the discrepancies mentioned above, the reason for this is not clear so far. ~~which calls into question the reliability of coarser RCM model settings for flood event representation in climate change studies at this spatial scale.~~ An accurate representation of seasonality is important also in the light of recent findings by Blöschl et al. (2017) that shifts in the seasonality are the only consistent large-scale climate change signal regarding floods identified so far.

The bias correction method Scaled Distribution Mapping (SDM) is able to systematically reduce biases on a seasonal basis. SDM improves results in magnitude and seasonality of maximum annual floods in all settings except for the small catchment~~s~~ (< 100 km^2), which has to do with the intensity-rainfall amount interplay mentioned above. ~~is due to the 3 hour time step used~~. The procedure corrects the rainfall amount but cannot correct the temporal dynamics. Also, ~~However, when single events and the corresponding atmospheric conditions are analysed, this systematic behaviour is not evident. Due~~due to the internal model variability, the RCM simulations partly decouple from their driving data and both synoptically forced and convective events may occur at different places or/and at different time as in the observations. Hence, in a usual climate modelling framework, i.e. long simulation periods and large RCM domain without nudging, an event-by-event analysis is not possible. Since the bias correction does not account for this effect and since it does not account for the number of sequential precipitation events (persistence), it might fail for single events and in weather type related approaches. Single events with very large biases – as seen using the WRF results – are over-compensated, i.e. an over-estimation is turned into an underestimation and vice versa. This affects the simulated flood peaks particularly for the higher return periods. The results further showed that the bias correction method is not able to compensate deviations in the hydrological conditions, particularly in summer. This has implications on flood generation at summer storms, which are frequent in the study area, and highlights the need for further research regarding modifying rainfall events in this season within the bias correction.

With respect to climate change applications of convection permitting simulations for flood representation we can conclude that, despite the seemingly good results in the CCLM 3 km setting, attention has to be paid and the test of the results against historical data is of utmost importance. On the other hand, deep-convection parameterisations in coarser resolved standard RCMs have shown to be a source of "deep" uncertainty. For instance, Kendon et al. (2014) found significant increases in summertime precipitation in convection-permitting climate simulations in UK while the coarser resolved counterpart does not show any significant change. Ban et al. (2015) and Berthou et al. (2018) found similar results for short-term extreme precipitation events in the Alpine region and in the Mediterranean. In order to circumvent possibly misguided but far reaching climate change adaptation strategies, either convection permitting RCMs or proper statistical "convection emulators" (that are currently discussed in the climate modelling communities) should be used. Coarser models could still be used in larger catchments for rough estimations, but they should not be taken for granted regarding local/regional flood change. Also, there is a trade-off in the additional costs of a 3 km simulation and the postulated (small scale) process description as long as the physical representation of such small scale processes can be substituted by statistical ones. Regarding bias correction, the temporal dynamics of the rainfall have to be analysed; only if RCM errors are found to be systematic an application of a current error correction method can be recommended.

~~The study results point out the general requirements of the hydrologic modeller on RCM simulations for flood representation to be used in climate change impact studies. Recommendations can be derived to further bridge the scale gap between climatology and hydrology in small and medium sized catchments. The coupled model enables estimation of climate change impacts on floods in small and medium sized catchments and can be therefore a valuable tool in water resources planning and management.~~

Acknowledgments

The study was funded by the Austrian Climate Research Programme (ACRP 6th call, Proj.No. KR13AC6K11102 - CHC-FloodS) by the Austrian Climate and Energy Funds (KLIEN). Regional climate model output on the EURO-CORDEX domain was provided by Klaus Keuler (BTU Cottbus, Germany) and Klaus Görgen (Institute of Bio- and Geosciences, Agrosphere, Research Centre Jülich). The WRF simulations from K. Goergen used in this study were conducted at the

Centre de Recherche Public -- Gabriel Lippmann, now: Luxembourg Institute of Science and Technology, under grant FNR C09/SR/16 (CLIMPACT) from the Luxembourg National Research Fund. The convection-permitting simulations were conducted in the course of the project NHCM-2 (nhcm-2.uni-graz.at), funded by the Austrian Science Fund (FWF, project no P24758-N29). We also thank Andreas F. Prein (National Center for Atmospheric Research, U.S.). He conducted the convection-permitting CCLM simulation and helped with fruitful comments. Computational resources were provided by the Jülich Supercomputing Centre (JSC) and by the Vienna Scientific Cluster (VSC). Additionally, we gratefully acknowledge the European Centre for Medium-Range Weather Forecasts (ECMWF) for providing ERA-Interim data. Hydro-meteorological data were provided by the Hydrographic Service of the province of Styria and the Central Institute for Meteorology and Geodynamics (ZAMG).

References

Addor, N., Rohrer, M., Furrer, R. and Seibert, J.: Propagation of biases in climate models from the synoptic to the regional scale: Implications for bias adjustment, J. Geophys. Res.-Atmospheres, 121(5), 2075–2089, doi:10.1002/2015JD024040, 2016.

Alfieri, L., Burek, P., Feyen, L., and Forzieri, G.: Global warming increases the frequency of river floods in Europe. Hydrol. and Earth Sys. Sci., 19(5), 2247-2260, doi: 10.5194/hess-19-2247-2015, 2015.

Allen, M. R. and Ingram, W. J.: Constraints on future changes in climate and the hydrologic cycle, 419 (6903), 224–232, doi: 10.1038/nature01092, 2002.

Allan, R. P., and Soden, B. J.: Atmospheric warming and the amplification of precipitation extremes. Science, 321(5895), 1481-1484, doi: 10.1126/science.1160787, 2008.

Awan, N. K., Truhetz, H. and Gobiet, A.: Parameterization-Induced Error Characteristics of MM5 and WRF Operated in Climate Mode over the Alpine Region: An Ensemble-Based Analysis, J. Clim., 24(12), 3107-3123, doi: 10.1175/2011JCLI3674.1, 2011.

Ban, N., Schmidli, J. and Schaer, C.: Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations, J. Geophys. Res.-Atmospheres, 119(13), doi:10.1002/2014JD021478, 2014.

Ban, N., Schmidli, J. and Schaer, C.: Heavy precipitation in a changing climate: Does short-term summer precipitation increase faster?, Geophys. Res. Lett., 42(4), 1165–1172, doi:10.1002/2014GL062588, 2015.

Berg, P., Moseley, C., and Haerter, J. O.: Strong increase in convective precipitation in response to higher temperatures. Nat. Geosci., 6, 181-185, doi: 10.1038/ngeo1731, 2013.

Berthou, S., Kendon, E. J., Chan, S. C., Ban, N., Leutwyler, D., Schär, C. and Fosser, G.: Pan-European climate at convection-permitting scale: a model intercomparison study, Clim. Dyn., doi:10.1007/s00382-018-4114-6, 2018.

Blöschl, G., Reszler, C. and Komma, J.: A spatially distributed flash flood forecasting model. Environmental Modelling and Software, 23, 4; 464-478, doi: 10.1016/j.envsoft.2007.06.010, 2008.

Blöschl et al.: Changing climate shifts timing of European floods. Science, 357 (2017), 6351; 588 – 590, 2017.

BMLFUW: Hydrologic Atlas Austria, Chapter 6.2: Hydrogeology. Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft (BMLFUW); Sektion Wasser - Abt. Wasserhaushalt, 2007.

Böhm, U., Kücken, M., Ahrens, W., Block, A., Hauffe, D., Keuler, K., Rockel, B., and Will, A.: CLM - The Climate Version of LM: Brief Description and Long-Term Applications, COSMO Newsletter, 6, 225-235, available via http://cosmo-odel.cscs.ch/content/model/documentation/newsLetters/newsLetter06/, 2006.

Chan, S. C., Kendon, E. J., Fowler, H. J., Blenkinsop, S., Ferro, C. A., and Stephenson, D. B.: Does increasing the spatial resolution of a regional climate model improve the simulated daily precipitation?. Clim. Dyn., 41(5-6), 1475-1495, doi: 10.1007/s00382-012-1568-9, 2013.

Chan, S. C., Kendon, E. J., Fowler, H. J., Blenkinsop, S., Roberts, N. M., and Ferro, C. A.: The value of high-resolution met office regional climate models in the simulation of multihourly precipitation extremes. J. Clim., 27(16), 6155-6174, doi: 10.1175/JCLI-D-13-00723.1, 2014.

Chan, S. C., Kendon, E. J., Roberts, N. M., Fowler, H. J., and Blenkinsop, S.: The characteristics of summer sub-hourly rainfall over the southern UK in a high-resolution convective permitting model. Environmental Research Letters, 11(9), 094024, 2016.

Cheng, C. S., Li, G., Li, Q., and Auld, H.: A Synoptic Typing Approach to Simulate Daily Rainfall and Extremes in Ontario, Canada: Potential for Climate Change Projections, J. Appl. Met. Clim., 49, 845 – 866, doi: 10.1175/2010JAMC2016.1, 2010.

Dankers, R., Christensen, O. B., Feyen, L., Kalas, M., and de Roo, A.: Evaluation of very high resolution climate model data for simulating flood hazards in the Upper Danube Basin, J. .Hydrol., 347, 319–331, doi: 10.1016/j.jhydrol.2007.09.055, 2007.

Dee, D. P., Uppala, S. M., Simmons A. J., et al.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc., 137(656), 553-597, doi: 10.1002/qj.828, 2011.

Flügel H. W., and Neubauer, F.: Geologische Karte der Steiermark 1:200.000, Geologische Bundesanstalt Wien, 1984.

Fowler, H. J., Blenkinsop, S., and Tebaldi C.: Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling, Int. J. Climatol., 27, 1547–1578, doi: 10.1002/joc.1556, 2007.

Giorgi, F., Jones, C., and Asrar, G. R.: Addressing climate information needs at the regional level: the CORDEX framework, WMO Bulletin, 58(3), 175-183, 2009.

Gobiet, A., Kotlarski, S., Beniston, M., Heinrich, G., Rajczak, J., and Stoffel, M.: 21st Century Climate Change in the European Alps – A Review, Science of the Total Environment, online first, ISSN 0048-9697, http://dx.doi.org/10.1016/j.scitotenv.2013.07.050, 2013.

Godina, R. and Müller, G.: Das Hochwasser in Österreich vom 22. bis 30 Juni, 2009 – Beschreibung der hydrologischen Situation (in German), Report, pp 21, Austrian Federal Ministry of Agriculture, Forestry, Environment and Water Management (BMLFUW), Abt. VII/3., Vienna, Austria, 2009.

Grayson, R., and Blöschl, G.: Summary of pattern comparison and concluding remarks, chapter 14. In Spatial Patterns in Catchment Hydrology: Observations and Modelling, Grayson R, Blöschl G (eds). Cambridge University Press: Cambridge, UK; 355 – 367, 2000.

Hanel, M., and Buishand, T.A.: On the value of hourly precipitation extremes in regional climate model simulations. J. Hydrol., 393, 265-273, doi: 10.1016/j.jhydrol.2010.08.024, 2010.

Hennegriff, W., Kolokotronis, V., Weber, H., and Bartels, H.: Klimawandel und Hochwasser -Erkenntnisse und Anpassungsstrategien beim Hochwasserschutz. KA – Abwasser, Abfall 2006 (53) Nr. 8, 2006.

Hershfield, D. M.: Estimating the probable maximum precipitation. J. Hydraul, Div., Am. Soc. Civ. Eng., 87 (Hy5), 99-116, 1961.

Hewitt, C. D., and Griggs, D. J.: Ensembles-Based Predictions of Climate Changes and Their Impacts (ENSEMBLES), Eos Trans. AGU, 85(52), 566, doi: 10.1029/2004EO520005, 2004.

Jacob, D., Petersen, J., Eggert, B. et al.: EURO-CORDEX: new high-resolution climate change projections for European impact research. Reg. Environ. Change, online first, doi: 10.1007/s10113-013-0499-2, 2013.

Katragkou, E., Garcia-Diez, M., Vautard, R., Sobolowski, S., Zanis, P., Alexandri, G., Cardoso, R. M., Colette, A., Fernandez, J., Gobiet, A., Goergen, K., Karacostas, T., Knist, S., Mayer, S., Soares, P. M. M., Pytharoulis, I., Tegoulias, I., Tsikerdekis, A., and Jacob, D.: Regional climate hindcast simulations within EURO-CORDEX: evaluation of a WRF multi-physics ensemble, Geosci. Model Dev., 8(3), 603-618, doi: 10.5194/gmd-8-603-2015, 2015.

Kain, J. S.: The Kain–Fritsch convective parameterization: An update. J. Appl. Meteor., 43, 170–181, 2004.

Kay, A. L., Rudd, A. C., Davies, H. N., Kendon, E. J., and Jones, R. G.: Use of very high resolution climate model data for hydrological modelling: baseline performance and future flood changes. Climatic change, 133(2), 193-208, doi: 10.1007/s10584-015-1455-6, 2015.

Kendon, E. J., Roberts, N. M., Fowler, H. J., Roberts, M. J., Chan, S. C. and Senior, C. A.: Heavier summer downpours with climate change revealed by weather forecast resolution model, Nat. Clim. Change, 4(7), 570–576, doi:10.1038/NCLIMATE2258, 2014.

Kendon, E. J., Ban, N., Roberts, N. M., Fowler, H. J., Roberts, M. J., Chan, S. C., Evans, J.P., Fosser, G. and Wilkinson, J. M.: Do convection-permitting regional climate models improve projections of future precipitation change?. Bulletin of the American Meteorological Society, 98(1), 79-93, doi: 10.1175/BAMS-D-15-0004.1, 2017.

Kida, H., Koide, T., Sasaki, H. and Chiba, M.: A New Approach for Coupling a Limited Area Model to a Gcm for Regional Climate Simulations, J. Meteorol. Soc. Jpn., 69(6), 723–728, 1991.

Knist, S., Goergen, K., and Simmer, C.: Evaluation and projected changes of precipitation statistics in convection-permitting WRF climate simulations over Central Europe. Clim. Dyn., doi: 10.1007/s00382-018-4147-x, 2018.

Komma, J., Reszler, C., Blöschl, G., and Haiden, T.: Ensemble prediction of floods - catchment non-linearity and forecast probabilities. Nat. Hazards Earth Sys. Sci., 7, 431-444, 2007.

Köplin, N., Schädler, B., Viviroli, D., and Weingartner, R.: Seasonality and magnitude of floods in Switzerland under future climate change. Hydrol. Proc., 28(4), 2567-2578, doi: 10.1002/hyp.9757, 2014.

Kotlarski, S., Keuler, K., Christensen, O. B., Colette, A., Deque, M., Gobiet, A., Goergen, K., Jacob, D., Luethi, D., van Meijgaard, E., Nikulin, G., Schaer, C., Teichmann, C., Vautard, R., Warrach-Sagi, K., and Wulfmeyer, V.: Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble, Geosci. Model Dev., 7(4), 1297-1333, doi: 10.5194/gmd-7-1297-2014, 2014.

Langhans, W., Schmidli, J., Fuhrer, O., Bieri, S. and Schär, C.: Long-Term Simulations of Thermally Driven Flows and Orographic Convection at Convection-Parameterizing and Cloud-Resolving Resolutions, Journal of Applied Meteorology and Climatology, 52(6), 1490–1510, doi:10.1175/JAMC-D-12-0167.1, 2013.

Lenderink, G., and van Meijgaard E.: Unexpected rise in extreme precipitation caused by a shift in rain type?, Nat. Geosci., 2 (6), 373, doi:10.1038/ngeo524, 2009.

Lenderink, G., Barbero, R., Loriaux, J. M., and Fowler, H. J.: Super-Clausius-Clapeyron Scaling of Extreme Hourly Convective Precipitation and Its Relation to Large-Scale Atmospheric Conditions, J. Clim., 30(15), 6037-6052, doi: 10.1175/JCLI-D-16-0808.1, 2017.

Lorenz, P., Skoda, G.: Bemessungsniederschläge kurzer Dauerstufen mit inadäquaten Daten. Mitteilungsblatt des Hydrographischen Dienstes in Österreich, Nr. 80, Wien, 2000.

Maraun, D.: Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums. Geophysical Research Letters, 39(Journal Article), L06706–L06706, doi: 10.1029/2012GL051210, 2012.

Maraun, D.: Bias Correction, Quantile Mapping, and Downscaling: Revisiting the Inflation Issue, J.Clim., 26(6), 2137–2143, doi:10.1175/JCLI-D-12-00821.1, 2013.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themessl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I.:

Precipitation Downscaling Under Climate Change: Recent Developments to Bridge the Gap between Dynamical Models and the End User, Rev. Geophys., 48, RG3003, doi: 10.1029/2009RG000314, 2010.

Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J. M., Hagemann, S., Richter, I., Soares, P. M. M., Hall, A. and Mearns, L. O.: Towards process-informed bias correction of climate change simulations, Nat. Clim. Change, 7(11), 664–773, doi:10.1038/nclimate3418, 2017.

Maurer, E. P. and Pierce, D. W.: Bias correction can modify climate model simulated precipitation changes without adverse effect on the ensemble mean, Hydrology and Earth System Sciences, 18(3), 915–925, doi:10.5194/hess-18-915-2014, 2014.

Mearns, L. O., Gutowski, W. J., Jones, R., et al.: A Regional Climate Change Assessment Program for North America, Eos Trans. AGU, 90(39), 311-312, doi: 10.1029/2009EO360002, 2009.

Meredith, E. P., Maraun, D., Semenov, V. A., and Park, W.: Evidence for added value of convection -permitting studying changes in extreme precipitation. Journal of Geophysical Research: Atmospheres, 120(24), 12500-12513, 2015.

Merz R., and Blöschl, G.: A process typology of regional floods. Water Resour. Res., 39 (12), article number 1340, 2003.

Merz, R., Blöschl, G., und Piock-Ellena, U.: Zur Anwendbarkeit des Gradex-Verfahrens in Österreich (Applicability of the Gradex-Method in Austria). Österreichische Wasser- und Abfallwirtschaft, 51, (11/12), pp. 291-30, 1999.

Nash, J. E., and Sutcliffe, J. V.: River flow forecasting through conceptual models part I—A discussion of principles. Journal of hydrology, 10(3), 282-290. 1970.

Nissen, K. M., and Ulbrich, U.: Increasing frequencies and changing characteristics of heavy precipitation events threatening infrastructure in Europe under climate change. Nat. Hazards Earth Sys. Sci., 17(7), 1177-1190, doi: 10.5194/nhess-17-1177-2017, 2017.

Parajka, J., Kohnová, S., Bálint, G., Barbuc, M., Borga, M., Claps, P., Cheval, S., Dumitrescu, A., Gaume, E., Hlavčová, K., Merz, R., Pfaundler, M., Stancalie, G., Szolgay, J., and Blöschl, G.: Seasonal characteristics of flood regimes across the Alpine–Carpathian range. J. Hydrol., 394(1), 78-89. doi.org/10.1016/j.jhydrol.2010.05.015, 2010.

Penna, D., Tromp-van Meerveld, H. J., Gobbi, A., Borga, M., and Dalla Fontana, G.: The influence of soil moisture on threshold runoff generation processes in an alpine headwater catchment. Hydrol. Earth Sys. Sci., 15(3), 689, doi: 10.5194/hess-15-689-2011, 2011.

Piani, D., and Haerter, J.: Two-dimensional bias correction of temperature and precipitation copulas in climate models. Geophys. Res. Lett., 39, L20401, doi: 10.1029/2012GL053839, 2012.

Prein, A. F., Gobiet, A., Truhetz, H., Keuler, K., Goergen, K., Teichmann, C., Maule, C. F., van Meijgaard, E., Deque, M., Nikulin, G., Vautard, R., Colette, A., Kjellström, E., and Jacob, D.: Precipitation in the EURO-CORDEX 0.11° and 0.44° simulations: high resolution, high benefits? Clim. Dyn., 46(1), 383-412, doi: 10.1007/s00382-015-2589-y, 2016.

Prein, A. F., Gobiet, A., Suklitsch, M., Truhetz, H., Awan, N. K., Keuler, K., and Georgievski, G.: Added value of convection permitting seasonal simulations, Clim. Dyn., 41(9-10), 2655-2677, doi: 10.1007/s00382-013-1744-6, 2013.

Prein, A. F., Rasmussen, R. M., Ikeda, K., Liu, C., Clark, M. P., and Holland, G. J.: The future intensification of hourly precipitation extremes, Nature Climate Change, 7(1), 48-+, doi: 10.1038/NCLIMATE3168, 2017.

Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., Keller, M., Toelle, M., Gutjahr, O., Feser, F., Brisson, E., Kollet, S., Schmidli, J., van Lipzig, N. P. M., and Leung, R.:, A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges, Rev. Geophys., 53(2), 323-361, doi: 10.1002/2014RG000475, 2015.

Reszler, C., Komma, J., Blöschl, G., and Gutknecht, D.: Ein Ansatz zur Identifikation flächendetaillierter Abflussmodelle für die Hochwasservorhersage. Hydrologie und Wasserbewirtschaftung, 50, 5; 220 – 232, 2006.

Reszler, C., Komma, J., Blöschl, G., and Gutknecht, D.: Dominante Prozesse und Ereignistypen zur Plausibilisierung flächendetaillierter Niederschlag-Abflussmodelle. Hydrologie und Wasserbewirtschaftung, 52 (3), 120-131, 2008a.

Reszler, C., Blöschl, G., and Komma, J.: Identifying runoff routing parameters for operational flood forecasting in small to medium sized catchments. Hydrol. Sci. J., 53 (1), 112-129, doi: 10.1623/hysj.53.1.112, 2008b.

Rockel, B., Will, A., Hense, A.: Special issue on regional climate modelling with COSMO-CLM (CCLM). Meteorol. Z., (17), 477–485, 2008.

Rogger, M., Pirkl, H., Viglione, A., Komma, J., Kohl, B., Kirnbauer, R., Merz, R., and Blöschl, G.: Step changes in the flood frequency curve: Process controls, Water Resour. Res., 48, W05544, doi:10.1029/2011WR011187, 2012.

Ruch, C., Reszler, C., Rock, G., and Plieschnegger, M.: Flash Flood Warning System – One Model, Two Methods. Geophysical Research Abstracts, Vol. 14, EGU2012- 9437, 2012.

Schemm, S., Nisi, L., Martinov, A., Leuenberger, D., and Martius, O.: On the link between cold fronts and hail in Switzerland, Atmospheric Science Letters, 17(5), 315-325, doi: 10.1002/asl.660,2016.

Schrödter, H.: Verdunstung – Anwendungsorientierte Meßverfahren und Bestimmungsmethoden. Springer Verlag, 186 p., 1985.

Seibert, P., Frank, A., and Formayer, H.: Synoptic and regional patterns of heavy precipitation in Austria, Theor. Appl. Climatol., 87(1-4), 139-153, doi: 10.1007/s00704-006-0198-8, 2007.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D., Wang, W., and Powers, J.: A description of the advanced research WRF version 2, Tech. Rep., NCAR, National Center for Atmospheric Research, Boulder, Colorado, USA, 2007.

Suklitsch, M., Gobiet, A., Truhetz, H., Awan, N. K., Göttel, H., and Jacob, D.: Error characteristics of high resolution regional climate models over the Alpine area, Clim. Dyn., 37(1-2), 377-390, doi: 10.1007/s00382-010-0848-5, 2011.

Switanek, M. B., Troch, P. A., Castro, C. L., Leuprecht, A., Chang, H.-I., Mukherjee, R., and Demaria, E. M. C.: Scaled distribution mapping: a bias correction method that preserves raw climate model projected changes, Hydrol. Earth Syst. Sci., doi:10.5194/hess-21-2649-2017, 2017.

Teutschbein, C., and Seibert, J.: Is bias correction of regional climate model (RCM) simulations possible for nonstationary conditions?, Hydrol. Earth Syst. Sci., 17, 5061–5077, doi: 10.5194/hess-17-5061-2013, 2013.

Themeßl M., Gobiet, A., and Leuprecht, A.: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models, International Journal of Climatology, 31 (10), 1530–1544, doi: 10.1002/joc.2168, 2011.

Trenberth, K. E., Dai, A., Rasmussen, R. M., and Parsons, D. B.: The changing character of precipitation, Bull. Am. Met. Soc., 84 (9), 1205–1217, doi: 10.1175/BAMS-84-9-1205, 2003.

Tiedtke, M.: A comprehensive mass flux scheme for cumulus parameterization in large–scale models. Mon. Wea. Rev., 117, 1779–1800, 1989.

Wang, G., Wang, D., Trenberth, K. E., Erfanian, A., Yu, M., Bosilovich, M. G., and Parr, D. T.: The peak structure and future changes of the relationships between extreme precipitation and temperature, Nature Climate Change, 7(4), 268-+, doi: 10.1038/NCLIMATE3239, 2017.

Wood, A. W., Leung, L. R., Sridhar, V., and Lettenmaier, D. P.: Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs, Climatic Change, 62, 189–216, doi: 10.1023/B:CLIM.0000013685.99609.9e, 2004.

Zittis, G., Bruggeman, A., Camera, C., Hadjinicolaou, P., and Lelieveld, J.: The added value of convection permitting simulations of extreme precipitation events over the eastern Mediterranean. Atmospheric Research, 191, 20–33, doi: 10.1016/j.atmosres.2017.03.002, 2017.

Tables

Table 1 Stream gauges used for evaluation (Fig. 1).

Abbrev. in Fig. 1	Gauge	River	Area (km²)	Data since
V	Voitsberg	Upper Kainach	211	1966
S	Schwanberg	Schwarze Sulm	75	1951
L	Leibnitz	Sulm	1103	1951
Gü	Gündorf	Saggau	200	1982
T	Tillmitsch	Lassnitz	480	1961
Fl	Fluttendorf	Gnasbach	119	1968

Table 2 RCMs and their settings.

Model	Grid size	Time step	nested in	conducted by
ERA-Interim	0.7° (~70 km)	6 h	-	ECMWF
CCLM 4.8 clm 17	0.44° (~50 km)	1 h	ERA-Interim	WEGC (CORDEX)
CCLM 4.8 clm 17	0.11° (~12.5 km)	3 h	ERA-Interim	BTU Cottbus (CORDEX)
CCLM 4.8 clm 17	0.03° (~3 km)	1 h	CCLM 0.11°	WEGC (NHCM-2)
WRF 3.3.1	0.44° (~50 km)	3 h	ERA-Interim	CRP-GL (CORDEX)
WRF 3.3.1	0.11° (~12.5 km)	3 h	WRF 0.44°	CRP-GL (CORDEX)
WRF 3.3.1	0.03° (~3 km)	1 h	WRF 0.11°	WEGC (NHCM-2)

Table 3 Model efficiency at the selected gauges in the calibration and historical (validation) period.

Gauge	Area (km²)	Calibration period 2000-2009			Historical (validation) period 1989-1999		
		BIAS (m³/s)	NSE (-)	RMSE (m³/s)	BIAS (m³/s)	NSE (-)	RMSE (m³/s)
Leibnitz/Sulm	1103	0.85	0.88	5.86	-0.28	0.83	8.15
Tillmitsch/Lassnitz	480	0.23	0.86	2.45	-0.35	0.82	3.42
Gündorf/Saggau	200	0.28	0.84	1.93	0.39	0.56	3.28
Schwanberg S. Sulm	75	0.20	0.78	0.66	0.22	0.65	0.88
Voitsberg/U. Kainach*	211	0.15	0.83	1.26	-0.13	0.85	1.33
Fluttendorf/Gnasbach	119	0.00	0.77	0.75	-0.03	0.67	0.91

* Continuous runoff data since 1996 (only four years in historical period), but historical maximum annual flood peaks available (hydrographic year book)