Dear Dr. Carlson,

Please find below the reviewer comments (black) as well as our responses and modifications to the manuscript (blue) below.

Best regards,
Emily Collier & Thomas Mölg

Reviewer 1: Benjamin Poschlod

General comments

The manuscript by Emily Collier & Thomas Mölg gives a comprehensive overview of a high-resolution 30-year climatological data set over Bavaria. The climate simulations were produced by the WRF model in 1.5 km resolution, nested in a 7.5-km-resolution domain and driven by ERA5 boundary conditions. The authors evaluate the model performance for air temperature, relative humidity, winds, surface pressure, precipitation, and land surface temperature for a 12-month period where they compare simulated values to observational data. Additionally, the effect of the application of nudging is assessed. Generally, the manuscript is well-written, and the figures support the presentation of the data set and its evaluation. In particular, the authors' handling of errors in the data set (e.g. sub-surface temperature in single glacier pixels) and explanation of deviations/biases (e.g. urban heat islands, connection between overestimated air temperature and overestimated radiation) are very valuable features of the data description.

The data are easily accessible and valuable for further application with focus on impact-related studies. Though, the total size of the 30-year daily-resolution data set (~450 GB) may not be easy to handle for users, who are new to the application of high-resolution climate data. On the other hand, users from the field of climate science would be interested in even higher temporal resolution, especially regarding the precipitation data. In sum, I consider the manuscript and the data appropriate for the publication within ESSD, although I recommend minor revisions based on the following remarks.

Thank you for your favorable assessment of our manuscript. Although the total dataset size is ~450 GB, the 3D variables that are most likely to be used for impacts assessments (e.g., near-surface air temperature, humidity and precipitation) amount to a more manageable 57 GB. With regards to the provision of higher temporal resolution precipitation data, please see our response to Michael Warscher for more details.

Specific comments

L1: Title: the data set is described as "convection-resolving". Though, within the whole manuscript, no convective events have been evaluated. Furthermore, the data set is provided in daily resolution, which is why short convective events cannot be investigated properly. Hence, I would suggest replacing "convection-resolving" by "high resolution".

We used the term "convection resolving" to describe the dataset following convention for atmospheric simulations with grid spacings below ~4-km. We did not mean to imply that we analyze convective events, however we agree that the use of this term could be misleading, especially to a wider audience, and therefore changed the title as suggested.

L80 / Table 1: The Kain-Fritsch cumulus scheme is applied for the 7.5 km domain, but not for the 1.5 km domain. According to that, not only deep, but also shallow convection is explicitly resolved in the 1.5 km domain? I would suggest clarifying this in the text.

Yes, as no additional parameterization is employed, deep and shallow convection are explicitly represented in the 1.5 km domain. We added to Sect. 2.1 “As no cumulus parameterization was employed in D2, both deep and shallow convection are assumed to be explicitly resolved.

L143: “For the distributed trend analysis, we did not apply a field significance test (e.g., Wilks, 2016) due to the small sample size.” – Does the “distributed trend analysis” refer to the results in L241 – 246 and Fig. 9? If yes – can you explain why is the sample size too small? If you test the trend at all 351x351 locations, the p-value should be adjusted for statistical tests at many locations (following e.g. Wilks 2016). Moreover, the reference (Wilks, 2016) is missing in your Reference section. Please also clarify, which test or method you used to detect trends.
Here we were referring to the sample size of years. Please note that based on the suggestion of the other reviewer, we removed the trends analysis from the manuscript in favor of expanding the model evaluation. Please see our response to this reviewer for more details.

L180: Has the observational precipitation data from DWD been corrected for undercatch? Especially in (pre-)alpine regions, this plays a major role, in particular for solid precipitation. I would recommend to briefly discuss this source of uncertainty.
We did not correct precipitation for undercatch and have added this information in Section 2.2. We also added to Section 3.1: “The MD is positive at the majority of stations, indicating that WRF generally underestimates observed precipitation. The underestimate is likely greater than reported here, since the observations were not corrected for wind-induced undercatch.”

L385: Figure 4 gives a good overview of the biases averaged for all locations. Though, the spatial distribution of biases would be of high interest as well. As the manuscript is already quite long and contains many figures, I would suggest creating such bias maps and moving these additional figures to a supplementary file.
We added some spatially distributed bias analysis as part of the expanded model evaluation.

Technical corrections
L156: 273.16 unit is missing
We changed to “exceeded the melting point.”

Figures 3,4,5,7,8: Temperature unit is “C” instead of “C”. Figure 6: Here the unit is missing in the figure (and given in the caption instead) Figure 9: Here you use “K” –> Please unify.
We changed the units to degrees Celsius throughout the paper and corrected the figure labels and captions.

Reviewer 2: Michael Warscher

In their manuscript “BAYWRF: a convection-resolving, present-day climatological atmospheric dataset for Bavaria”, the authors present a new high-resolution RCM simulation using WRF and ERA5 reanalysis data as boundary condition. They evaluate the performance for the target region of Bavaria using station observations.

General Comments
The manuscript is very well written and of high technical and scientific quality. It fits very well in the scope of ESSD. However, I have several issues, questions, and suggestions which at

large could lead to major revisions. However, I understand that the manuscript is mainly an overview of the presented data and thereby, the amount and detail of the analyses and following content has to be limited at certain points. I would gladly leave the decision to the editors on how much of and at what detail level my suggestions should be addressed. The dataset they produced is generally very valuable for the scientific community, as well as for many users in different sectors.

The authors have chosen a single year as specific validation period. In addition, they point out in L. 101 that it is not an average year in terms of seasonal climatology (record heatwave in 2018). The chosen year might therefore not be a representative period for the RCM performance in other years. However, an extension of the evaluation period seems just limited by a missing run using the NO_NUDGE configuration. As the whole exercise is a historic / present-day reanalysis driven simulation effort, the NUDGE setup was run for the whole 30-year period anyways and would be available for additional validation years. I would highly recommend to add at least one additional year of validation (the more the better) to strengthen the results under different conditions. The validation could potentially even be done for the whole 30-years (just being limited by available observations and – by now - the missing NO_NUDGE for more than one year). I am also quite sure that an extension of the analysis would not limited by available observation data. If no additional simulations (NO_NUDGE) can be performed, the authors might think of some additional validation using the 30-year NUDGE run only.

While I really acknowledge the direct comparison to station data, the study would highly benefit from a comparison to gridded observation data sets such as REG- NIE (1 km, https://www.dwd.de/DE/leistungen/regnie/regnie.html) or HYRAS (5 km, https://www.dwd.de/DE/leistungen/hyras/hyras.html). Besides the correct representation of single stations, the real benefit of such a computationally expensive high- resolution simulation might or should be – besides the reproduction of observed station data - the resolving of spatial distributions.
To address this comment, we replaced the trend analysis with a new section evaluating the performance of BAYWRF over the whole simulation period compared with all available station data from the DWD datasets TT_TU_MN009, RF_TU_MN009, and R1_MN008 (*T*, *RH* and *PREC*) at daily timescales. We note that the *PREC* dataset is only available after 1 September 1995. We therefore also added a brief pattern correlation analysis between simulated monthly total precipitation and the suggested REGNIE dataset for the full simulation period.

We note that we have added the following information after posting this response in the interactive discussion:
Added section: 3.2 Evaluation of BAYWRF
"Averaged over the full simulation period, BAYWRF shows a similar magnitude of agreement with station T and RH data at daily timescales as found at sub-diurnal timescales for the single evaluation year (Fig. 8; cf. Fig. 3). For T, the MD has lower and upper quartiles of -0.3 and 0.4°C, respectively, while the values of R2 uniformly exceed 0.92. For RH, the MD has lower and upper quartiles of 0.4 and 4.4% while the respective values for R2 are 0.57 and 0.65. For PREC, the upper and lower quartiles of MDs considering days with observed precipitation are -0.1 and 0.1 mm while for R2 the values are 0.41 and 0.47. A similar number and magnitude of wet false events are simulated (twenty times less than the sample size of observed events). Spatially, BAYWRF exhibits a positive bias in T and a negative bias in RH in the interior of Bavaria, and the converse anomalies in the pre-alpine and alpine areas in the south and along the eastern border of the region (Fig. 8a, c). The mean R2 values for RH show a clear meridional gradient (Fig. 8d), which suggests that the model has

some difficulty capturing processes governing near-surface moisture fluctuations in the southern part of Bavaria. Nonetheless, the highest correlation coefficients for observed precipitation events are found in this region (Fig. 8f). In addition, considering monthly precipitation sums, the centered pattern correlation between REGNIE and BAYWRF ranges from a lower quartile of 0.64 to an upper quartile of 0.82. Therefore, the characteristics of precipitation variability in time and space are captured by the dataset."

The authors point out, that they use a convection-permitting resolution of 1.5 km. However, the topic of simulating convective precipitation is not referred to again in the manuscript. This is still a very important and relevant topic, and the presented data would be ideally suited to look into this. Several questions arise and could easily be tackled. E.g. the authors show an underestimation of precipitation at some point. Could this be explained by the 1.5 km still being too coarse to resolve all or enough convective events? Are the results of the KF parameterization in the 7.5 km (D1) similar or totally different? Are sub-daily precipitation dynamics captured? Some of these questions could quite easily be investigated by comparing your results of D1 (convection parameterized) and D2 (convection resolved) to gridded precipitation products such as REGNIE and maybe even to station data.

This leads to another question regarding the resolution. While I do not question the validity and satisfying performance of the presented simulation, I would be very glad to see more about the added value of such a high resolution. This could be done by a comparison of the performance between the results of Domain D1 and Domain D2. There are no analyses in the manuscript that try to address this important question.
We agree that the question of convective characteristics is an interesting one, yet do not evaluate sub-diurnal precipitation or the added value in D2 in this manuscript for several reasons. do not evaluate sub-diurnal precipitation or the added value in D2 in this manuscript for several reasons. First, we made this dataset for (dendroclimatological) impact studies, which require kilometer-scale resolution but only daily (e.g., Dietrich et al., 2019) or even monthly temporal resolution. Given the intended application of the dataset for impact studies, we have also only made data from D2 available through the OSF repository, since users interested in climate data at ~ O(10 km) grid spacing are likely to use the ERA5 data directly.

Second, we note that there is already some consensus in the literature as to the added value of kilometer-scale grid spacing, in particular for precipitation (e.g., Ban et al., 2014; Mölg and Kaser, 2011; Prein et al., 2015) and examination of this scientific question (data interpretation) would appear to be outside of the scope of ESSD.

Finally, the full two-hourly time series of precipitation alone amounts to nearly 50 GB of data. As our project already exceeds the desired size for (free) data storage on the OSF, uploading more data to the repository is problematic – as would be presenting and evaluating data that are not available for download.

Another important issue regarding the trend analysis can be found in the specific comments.

Specific Comments

L. 12: I suggest to remove the reference to the project here (and at other positions in the manuscript) and state the project name solely in the acknowledgement section.
We removed the reference to the project from the abstract and the last paragraph of the introduction. Please see below for our reply about removing references elsewhere in the manuscript.

L. 32 – 38: I see that the linkage to dendroclimatological studies refers to the research project, but in my opinion, this is not needed here. You don't show any further results regarding this topic, and the general effort and method of dynamical downscaling does not really need to be justified or explained within this manuscript.
While we agree that the introduction could be re-formulated to focus only on climate downscaling, we feel the project description and dendroecological discussion in the introduction provides an important context for why we generated these data as well as some configuration choices (e.g., the temporal resolution of the dataset). While we have removed two references to our specific project in the manuscript, we would prefer to keep some mention of this information in the introduction.

L. 49 – 59: The same as the comments above: this is in general interesting information, but not within this manuscript. The paragraph should be shortened, maybe only keep the last sentence: "High-temporal. . ."
Please see above.

L. 66 – 68: Please remove the sentence: "These data. . ." for the reason stated above.
We removed this sentence.

L. 69: While your statement here is certainly true, I would prefer a more moderate phrasing, e.g., "These data has the potential to find. . ." instead of "These data will also find. . .".
We made this change.

L. 78: Could you please give some more information on how the WRF configuration was chosen? This should then also be added to the manuscript. You state that the setup is based on Collier et al. (2019) but the study seems to be located in completely different climate and terrain conditions (East Africa). It is widely shown in literature that the performance of the chosen configuration strongly depends on the region. I understand that it is not feasible to perform a full configuration optimization ensemble, but some more information on this issue should be added.
The configuration is based on all of our previous convection resolving modelling efforts, including some mesoscale and LES simulations for the European Alps that are as-of-yet unpublished and therefore cannot be cited. The configuration was not specifically optimized for the Bavaria domain, given temporal and computational constraints. To clarify, we amended this sentence to: "The physics and dynamics options used in the simulations are based on several recent convection-permitting applications of WRF by the authors (e.g., Collier et al., 2019) but were not specifically optimized for these domains due to the computational expense of the simulations."

L. 122: It would be very interesting to see sub-daily results also for precipitation from such a simulation. By permitting convective events, this could potentially be one of the strong points of such a high-resolution simulation.
Please see our response to the general comment on this issue.

L. 123: What about all the cases where modeled precipitation > 0 but the observed precipitation = 0? These cases should somehow be analyzed too and not be neglected in the performance analysis.
Excellent point. If these events are included in the evaluation in Fig. 3, the bias evaluation is artificially improved, because WRF underestimates the magnitude of observed precipitation events and the magnitude of simulated "false" events is generally small (for example, the mean and median magnitudes are ~ 0.6 and 0.1 mm /day in WRF_NUDGE). We therefore

left Figure 3 as is but added the following sentence to Section 3.1: “In addition to underestimating observed daily precipitation events (total sample size of 35,791 for all stations and record lengths), the simulations also produce false daily precipitation events, the vast majority of which are very small in magnitude (the median value in both WRF simulations is less than 0.1 mm/day). Considering wetter days (precipitation exceeding 1 mm/day; Ban et al., 2014), the number of false events is more than ten times smaller than the number of observed events (sample sizes of 3,096 and 2,249 in WRF_NO_NUDGE and WRF_NUDGE, respectively).”

L. 128: Why did you choose two hourly WRF output? I see that the output somehow has to be confined, but hourly values would also be very valuable! Do you still have these available? I think it is fine for the manuscript to keep two hourly results, but at least for the main surface variables, it would be very useful to have hourly values as well. If they are available upon request, you could add this information to the data availability section.
The history write frequency for the WRF simulations was set to two-hourly, so hourly data were not stored and are not available. This choice was made as a compromise between temporal resolution and storage requirements, and because the dendroclimatological part of the project only requires daily resolution. At two-hourly temporal resolution, the unprocessed model output already amounts to 55 TB of storage, which represented a huge logistical challenge to store, analyze, and make available through the public repository.

We added a sentence to emphasize this point in the last paragraph of Section 2.1: “We selected this write frequency as a compromise between high-temporal resolution and the logistical challenges of storing, analyzing, and disseminating the data.

L. 147 - 157: I highly appreciate the very well investigated and documented error handling here!
Thank you!

L. 160: See comment above: why did you choose two-hourly output? Was it just to save storage or is there another reason?
Please see our reply above.

L. 169 - 170: It is very valuable that you try to compare the results to other studies (here to the work by Warscher et al. 2019), but the numbers are not really comparable here (different investigation area, nudging strategy, stations, terrain, etc.). I would either keep your statement and add an explanation regarding the differences in the analyses or remove the statement or phrase it differently (“similar but lower” is quite inexplicit).
Our intention was to contextualize our results with previous literature without going into great detail, since method comparison is outside of the scope of ESSD and the exact biases depend on many factors. We rephrased this sentence to “These values are comparable to previous high-resolution applications of WRF over Bavaria (Warscher et al., 2019)” and hope this change addresses your concern.

L. 180 - 181: To me, this is a strong hint that the used resolution is still not high enough to correctly simulate absolute convective precipitation amounts. That’s one reason why it would be so valuable to analyze more than one year of data and to compare results between D1 and D2.
Given the resolution requirements for explicitly simulating the evolution of individual clouds, it is likely that some influence of moist convection is not being captured at 1.5-km grid spacing. We added to this paragraph: “The reported seasonal and mean biases in daily precipitation are consistent with a potential underestimate of deep convection and convective

precipitation at 1.5-km grid spacing. Although simulated mean precipitation shows a weak grid dependency below a spacing of ~ 4 km (Langhans et al., 2012), sub-kilometer spatial resolution is required to explicitly resolve the evolution and characteristics of clouds (e.g., Bryan et al., 2003; Craig and Dörnbrack, 2008; Prein et al., 2015).

L. 213: You clearly show that the grid-nudged run is performing better than the "free" simulation, which again leads to the question of the benefits of the simulation. This result indicates that WRF adds biases compared to the ERA5 forcing simulation when not grid-nudged to them. The DWD stations you used for your validation might even have been assimilated in ERA5 which again questions to some point the added value of the simulation.
Simulation drift is a well-known issue with regional climate simulations, especially when larger domains are used, as in our simulations (e.g., Prein et al., 2015 and references therein). We contacted DWD to inquire which datasets may have been used for data assimilation in ERA5 but unfortunately have not received a reply. Nonetheless, we think it is a logical consequence (and not a particular drawback of the model or simulations) that nudging towards a dataset that assimilates some observational data produces results that agree more closely with the same, or other, observations. The benefit of the simulations is the dynamically consistent (and in D2 where no nudging occurs, physically consistent) representation of local climate at the kilometer scale.

L. 215 - 216 Remove "(the temporal resolution of data available in BAYWRF)". This is not important here.
We made this change.

L. 228: Three typos: add spaces after "WRF:"
We corrected these typos.

Sect. 3.5: Trend analysis: it is quite obvious that the trends are reproduced by the simulation when it is forced by a reanalysis product such as ERA5 (and grid-nudging is used). You could think about removing the whole section, as I do not see a value in this information. If the trends would not have been reproduced, it would be an argument that something goes wrong, but – the other way round - these results are not proving a good performance of WRF (as stated in the paragraph). You simply see the overall dynamics of the forcing (which includes assimilations of historic observations and therefore reproduces historic trends).
We removed the trend section in favour of expanded model evaluation, as suggested above.

L. 241 - 246: The paragraph falls a bit short compared to the other ones. The spatial distribution of trends could be more elaborated (if a trend section is kept). The fine scale spatial differences of trends is in the end the information that is produced by the RCM simulations (see the statements regarding trends and reanalysis above).
We removed this paragraph.

L. 261 - 263: The statement regarding grid-nudging may be true, but I do not see it as a success, as the forcing obviously includes assimilated observations (see comments above).
We intended this statement to report a result rather than a success. We attempted to clarify by re-phrasing: "Comparison of simulations for the period of September 2017 to August 2018 with and without grid-analysis nudging against extensive meteorological measurements across Bavaria showed that nudging decreased the mean deviations and increased the coefficient of determinations at the majority of sites for nearly all evaluated atmospheric variables, in particular precipitation. This approach was therefore adopted for generating the full BAYWRF dataset."

Fig. 7 c) and d): If I understand it right, the values in the legend should be reversed (wrong sign).

The difference in night-time land-surface temperature was computed as MODIS minus WRF. Somewhat unconventionally, here orange colors delineate negative values, i.e. where WRF is warmer. We realize the color choice is somewhat confusing and added an additional sentence to the figure caption: “Note that the orange and red colours in panels c and d shade areas where WRF is warmer than MODIS (MODIS minus WRF is negative) and vice versa for blues.”

References

Ban, N., Schmidli, J. and Schär, C.: Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations, J. Geophys. Res., 119(13), 7889–7907 [online] Available from: http://onlinelibrary.wiley.com.proxy.library.uu.nl/doi/10.1002/2014JD021478/full, 2014.

Dietrich, H., Wolf, T., Kawohl, T., Wehberg, J., Kändler, G., Mette, T., Röder, A. and Böhner, J.: Temporal and spatial high-resolution climate data from 1961 to 2100 for the German National Forest Inventory (NFI), Ann. For. Sci., doi:10.1007/s13595-018-0788-5, 2019.

Mölg, T. and Kaser, G.: A new approach to resolving climate-cryosphere relations: Downscaling climate dynamics to glacier-scale mass and energy balance without statistical scale linking, J. Geophys. Res. Atmos., 116(16), doi:10.1029/2011JD015669, 2011.

Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., Keller, M., Tölle, 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., doi:10.1002/2014RG000475, 2015.

BAYWRF: a high-resolution present-day climatological atmospheric dataset for Bavaria

Emily Collier[1] and Thomas Mölg[1]

[1]Climate System Research Group, Institute of Geography, Friedrich-Alexander University Erlangen-Nürnberg (FAU), Erlangen, Germany

Correspondence to: Emily Collier (emily.collier@fau.de)

Abstract.

Climate impact assessments require information about climate change at regional and ideally local scales. In dendroecological studies, this information has traditionally been obtained using statistical methods, which preclude the linkage of local climate changes to large-scale drivers in a process-based way. As part of recent efforts to investigate the impact of climate change on forest ecosystems in Bavaria, Germany, we developed a high-resolution atmospheric modelling dataset, BAYWRF, for this region over the thirty-year period of September 1987 to August 2018. The atmospheric model employed in this study, WRF, was configured with two nested domains of 7.5- and 1.5-km grid spacing, centred over Bavaria and forced at the outer lateral boundaries by ERA5 reanalysis data. Using an extensive network of observational data, we evaluate: (i) the impact of using grid-analysis nudging for a single-year simulation of the period of September 2017 to August 2018; and (ii) the full BAYWRF dataset generated using nudging. The evaluation shows that the model represents variability in near-surface meteorological conditions generally well, although there are both seasonal and spatial biases in the dataset that interested users should take into account. BAYWRF provides a unique and valuable tool for investigating climate change in Bavaria with high interdisciplinary relevance. Data from the finest resolution WRF domain are available for download at daily temporal resolution from a public repository at the Open Science Framework (Collier, 2020; https://www.doi.org/10.17605/OSF.IO/AQ58B).

1 Introduction

The forcing of climate change in modern times is clearly of global nature, and many important scientific problems can be understood at the global scale as well (e.g., Held and Soden, 2006). Climate impact assessments, however, must also understand the effects at regional and even local scales in order to develop appropriate adaptation and mitigation measures. Although local phenomena such as glaciers, lakes, vegetation patterns, or stream flow show a strong dependence on the large-scale climate dynamics, these proxies experience further variability when the large-scale signal is transferred to their

Deleted: convection-resolving,

Deleted: within the BayTreeNet project,

Deleted: the

Deleted: of Bavaria

Deleted: Based on a shorter evaluation period of September 2017 to August 2018, we evaluate two aspects of the simulations: (i) we assess model biases compared with an extensive

Deleted: at both two-hourly and daily mean temporal resolutions, and (ii) we investigate the influence of using grid

Deleted:

Deleted: . The

Deleted: well, with a clear improvement when nudging is used

Deleted: cold

Deleted: warm

Deleted: winter and summer, respectively. We also present a brief overview of the full dataset, which will provide…

Deleted:

location (e.g., Mölg et al., 2014). In order to contextualize local changes, there is a need to link local climate to the large-scale climate, ideally in a process-based way.

In dendroclimatological studies, the traditional approach is to compute a calibration function between local or regional tree-ring parameters and climatic variables. Typically, such a statistical relationship would try to utilize local station data (which are generally sparse), gridded observations (which tend to be coarse resolution), or indices of large-scale climate dynamics (which describe coupled atmosphere-ocean modes) as the climatic influence (e.g., Hochreuther et al., 2016). Besides known problems like stationarity (e.g., Frías et al., 2006), statistical approaches also limit the possibilities to explain the influences at the various scales on a process-resolving level. Dynamical downscaling with a full numerical atmospheric model provides a physical answer (Giorgi and Mearns, 1991), yet the disadvantage is the high computational cost. Hence, dynamical downscaling at near-kilometer resolution has traditionally been performed on a case-study basis for weather events (e.g., Gohm et al., 2008). Multi-decadal simulations, on the other hand, were typically limited to resolutions of tens of kilometers (e.g., Di Luca et al., 2016). With the progress of computational resources, dynamical downscaling is becoming a candidate for climate impact studies that require local-scale information, and the first decadal simulations at ~1-km resolution are now available (e.g., Collier et al., 2018). From the resultant model output, impact studies could utilize information about local meteorological conditions at high-spatial and high-temporal resolution, and over long, climatologically relevant temporal periods. Moreover, the physically consistent output would enable to generate the said process understanding of influences across the various climatic scales.

The management of forests is a classical impact study where adaptation and mitigation measures meet the heterogeneous effects of climate change at local scales (e.g., Lindner et al., 2014). With this background, the project BayTreeNet was started recently under the umbrella of the interdisciplinary climatological research network Bayklif (https://www.bayklif.de; last accessed 1 March 2020), and aims to investigate the response of forest ecosystems to current and future climate dynamics across different growth areas in Bavaria, Germany. The project comprises a network of 10 measurement sites where meteorological and dendroecological data will be monitored and used both for research and for public and educational outreach, which are currently in the process of being established. High-temporal (approximately daily) and high-spatial resolution data is a key component of dendroecological impact studies, since the physiological behavior of trees, their structural properties and functional wood anatomy, as well as other important parameters such as wood density and mortality risk are not only influenced by seasonal averages, but also by short-term extreme events and weather anomalies (e.g., Bräuning et al., 2016).

Previous regional climate simulations including Bavaria over continuous multi-decadal periods were performed with model resolutions as high as 5-7 km and up to the year 2009 (e.g., Berg et al., 2013; Warscher et al., 2019). However, to the best of our knowledge, such datasets at the kilometer scale and up to the near present do not yet exist, despite previous research

Deleted: . The locations were selected carefully to account for ecological and elevational variety in the study region, and the sites…

Deleted: installed

highlighting the importance of convection-permitting resolution in this region (Fosser et al., 2014). We address this data gap by performing simulations with an atmospheric model, configured with convection-permitting spatial resolution in a nested domain over Bavaria, for the recent climatological period of 1987 to 2018. These data have the potential to find multidisciplinary interest among researchers assessing ecological and human dependencies on the climate for scientific and practical questions.

Deleted: These data were generated as part of the aforementioned BayTreeNet project in order to assess the dendroecological consequences of local climate change on forests in the study region. These data will also…

2 Methods

2.1 Atmospheric model

The atmospheric simulations were performed using the advanced research version of the Weather Research & Forecasting (WRF) model v. 4.1 (Skamarock and Klemp, 2008) configured with two one-way-nested domains of 7.5- and 1.5-km grid spacing situated over Bavaria (Fig. 1), hereafter referred to as D1 and D2. Terrain data were taken from NASA Shuttle Radar Topographic Mission data re-sampled to 1-km and 500-m grids (Jarvis et al., 2008; https://cgiarcsi.community/data/srtm-90m-digital-elevation-database-v4-1; last accessed 24 May 2020) for D1 and D2, respectively, while land-use data was updated based on the European Space Agency Climate Change Initiative Land Cover data at 300-m spatial resolution (http://maps.elie.ucl.ac.be/CCI/viewer/download.php; last accessed 18 April 2018). The physics and dynamics options used in the simulations are based on several recent convection-permitting applications of WRF by the authors (e.g., Collier et al., 2019) but were not specifically optimized for these domains due to the computational expense of the simulations. The options are summarized in Table 1. As no cumulus parameterization was employed in D2, both deep and shallow convection are assumed to be explicitly resolved. We note that no additional urban physics were enabled beyond the default parameterization used by the Noah family of land surface models (Liu et al., 2006) and land-use sub-tiling was not enabled.

Deleted: 4.1 (Skamarock and Klemp, 2008)

Deleted: (Collier et al., 2019)

Deleted: and are summarized in Table 1.

Forcing data at the lateral boundary of D1 and bottom boundaries of both domains was taken from the ERA5 reanalysis (Copernicus Climate Change Service (C3S), 2017) at three-hourly temporal resolution. The 30-year simulation was divided into 30 annual simulations that were run continuously from 15 August of year *n-1* to 31 August of year *n*. The first 16 days of each simulation were discarded as spin-up time, retaining data from 1 September of year *n-1* onwards. Atmospheric carbon dioxide (CO_2) was updated in WRF for each simulation year using annually and globally averaged concentrations at the surface from the National Oceanic and Atmospheric Administration Earth System Research Laboratory (Tans and Keeling, 2019). Each simulation employed the CO_2 concentration of year *n*, which ranged from 351 to 407 ppm between 1988 and 2018. All other parameters and bottom boundary conditions (e.g., vegetation and land use) were held constant for all simulations. Therefore, they do not capture the impact of known land-use changes over the study period (e.g., Fuchs et al., 2013).

Each run required 12 days of wall-time with 320 processors on the Meggie compute cluster at the Erlangen Regional Computing Center, for a total of 2.86 million core hours. The model was compiled using intel 17.0 compilers and run using distributed-memory parallelization. Model output was written at two-hourly intervals, amounting to more than 55 TB of data, in addition to ~30 TB of pre-processing and input files. We selected this write frequency as a compromise between high-temporal resolution and the logistical challenges of storing, analyzing, and disseminating the data.

2.2 Evaluation of Forcing Strategy

For the period of 00 UTC 1 September 2017 to 00 UTC 1 September 2018, we compared two simulations with different forcing approaches: one excluding and one including grid-analysis nudging to constrain drift in the large-scale circulation (e.g., Bowden et al., 2013). This period was selected due to the higher availability of observational data closer to present day and because the summer of 2018 contained a record heatwave with drought conditions (Beyer, 2018), permitting evaluation of an extreme event. We refer to these simulations as WRF_NO_NUDGE and WRF_NUDGE, respectively. For the WRF_NUDGE simulation, analysis nudging was applied in D1 outside of the planetary boundary layer and above the lowest 10 model levels using the default strength (3.0×10^{-4}) for temperature and winds and reduced strength (5.0×10^{-5}) for the water vapor mixing ratio (e.g., Otte et al., 2012), consistent with a previous decadal application of WRF (Collier et al., 2018). Given the computational expense of each annual simulation, we did not attempt to optimize the nudging coefficients for our study area and instead evaluate simply whether nudging in this form improves the simulated atmospheric variables or not.

2.3 Observational Data

For model evaluation, we used data from the German Weather Service (DWD) Climate Data Center for all stations in Bavaria with hourly temporal resolution available, which provide good spatial coverage of our study area (Table 2; Fig. 2). To evaluate the forcing approach, we compared the following near-surface atmospheric variables at the highest temporal resolution available in the simulations, which is two-hourly: air temperature and relative humidity at 2 m (*T* and *RH*), zonal and meridional wind components at 10 m (*U* and *V*), and surface pressure (*PS*). In addition, we compared with daily total precipitation (*PR*). In our comparison with observations, we excluded measurement sites where the observed terrain height differed from the modelled value by more than 100 m (similar to e.g., Vionnet et al., 2019), corresponding to four sites in total for all variables except for *PS* (three) and *PR* (nine). After this exclusion, the average difference between modelled and observed terrain height at all stations is within ± 8 m for each dataset. We also excluded any days with missing observational data when computing daily statistics. We note that observed precipitation was not corrected for undercatch. We did not evaluate radiation variables, as only sunshine hours are available from the DWD in sufficiently large sample sizes. However, for understanding temperature biases in WRF during summer 2018, we used incoming shortwave radiation from the DWD Climate Data Center dataset entitled "*Hourly station observations of solar incoming (total/diffuse) and longwave downward radiation for Germany*" (Table 2). In total, there were four sites with both incoming shortwave (*SW*) and *T* data available in

Deleted: ¶

Deleted: Model

Moved (insertion) [1]

Deleted: detailed

Deleted: selected the period of 00 UTC 1 September 2017 to 00 UTC 1 September 2018, as data availability is highest closest to present day and the summer of 2018 contained a record heatwave with drought conditions (Beyer, 2018). Neither extensive physics optimization nor a longer evaluation period was possible due to the computational expense of the simulations. For the evaluation period, we compared two test simulations (see Sect. 2.3) with…

Deleted: We

Deleted:), and

Bavaria between 1 June and 31 August 2018 whose elevation was represented within ±100 m in D2: Nürnberg (id 3668), Weihenstephan-Dürnast (5404), Würzburg (5705), and Fürstenzell (5856).

For statistical analysis, we computed the mean deviation (MD), mean absolute deviation (MAD), and the coefficient of determination (R^2) between station data and data from the closest grid point in D2 without spatial interpolation at two-hourly and, for precipitation, at daily temporal frequency. The MD, also referred to here as the model bias, and the MAD were computed from observation minus model data. For precipitation, only daily totals were evaluated, and the MD and MAD were computed considering only days with non-zero observed precipitation.

Finally, we also compared night-time land surface temperature (LST) from the MODIS MYD11A1 dataset (Table 2) at 1-km spatial and daily temporal resolution with simulated skin temperature in D2 for the period of 1 June to 31 August 2018. The night view time ranged from 1.2 to 2.8 hours in local solar time, with a domain and time averaged value of 2.2 hours. As WRF data were only available at two-hourly timesteps, we averaged 00 and 02 UTC (01 and 03 local time) data from D2 for comparison with MODIS. In our comparison, we excluded nights when MODIS had more than 50% missing data over D2, leaving a sample size of 52.

For evaluating the full simulation, we performed a similar analysis with the aforementioned station datasets for *T*, *RH* and *PREC* (Table 2), however we averaged and summed the data to daily timescales for comparison with BAYWRF. In addition to comparing with individual stations, we also compared monthly total precipitation in BAYWRF with the gridded dataset REGNIE from the DWD CDC, which is based on interpolated station data and available at 1-km spatial resolution (e.g., Rauthe et al., 2013). For the comparison, REGNIE data were regridded to the WRF grid using patch interpolation and the ESMF regridding toolbox in NCL (https://www.ncl.ucar.edu/Document/Functions/ESMF/ESMF_regrid.shtml; last accessed 10 September 2020) and the centered pattern correlation between the two datasets was computed.

2.4 Numerical issue in BAYWRF

We note that unphysically large sub-surface temperatures were simulated at a number of glacierized grid points, primarily during the months of July to September. Considering all of D2, the daily average number of affected grid cells was 24, compared with 294 glacierized and 122,500 total cells. The maximum number of affected grid points was 274 on 31 August 2017, corresponding to 0.2% of D2. In addition, over the climatological simulation, only one grid point in Bavaria was affected (J = 71, I = 285; 47.4952°N, 13.6039°E). Surface temperature remained physical, since it is limited at the melting point over glacier surfaces, and soil moisture was unaffected, since it is specified to be fully saturated in glacierized grid cells. No other land-use categories were affected, and adjacent grid points were also unaffected, as the land surface model operates as a column model with no lateral communication. To preclude usage of these data, sub-surface temperature was set to missing where it exceeded the melting point at glacierized grid points in BAYWRF. More information about this

Deleted: 2.3 Forcing Strategy¶
For the evaluation period, we compared two simulations with different forcing approaches, one excluding and one including grid analysis nudging to constrain drift in the large-scale circulation (e.g., Bowden et al., 2013). We refer to these simulations as WRF_NO_NUDGE and WRF_NUDGE in Section 3.1,

Moved up [1]: For the WRF_NUDGE simulation, analysis nudging was applied in D1 outside of the planetary boundary layer and above the lowest 10 model levels using the default strength (3.0×10^{-4}) for temperature and winds and reduced strength (5.0×10^{-5}) for the water vapor mixing ratio (e.g., Otte et al., 2012), consistent with a previous decadal application of WRF (Collier et al., 2018). Given the computational expense of each annual simulation, we did not attempt to optimize the nudging coefficients for our study area and instead evaluate simply whether nudging in this form improves the simulated atmospheric variables or not.¶

Deleted: ¶
2.4 Climatological Analysis¶
To briefly evaluate the full climatological simulation, we compared simulated and observed monthly mean *T* from the DWD dataset '*MO_TT_MN004*' (Table 2), with a sample size of 26 stations that remained after filtering for height differences exceeding 100 m, the presence of missing data, and stations located in grid cells classified as urban (see Sect. 3.1). For the distributed trend analysis, we did not apply a field significance test (e.g., Wilks, 2016) due to the small sample size. Future users should rigorously evaluate biases for the variables, time periods, and resolutions relevant to their particular applications.¶

Deleted: 273.16

numerical issue is available on the model's GitHub repository (https://github.com/wrf-model/WRF/issues/1185; last accessed 24 May 2020).

3 Results & Discussion

3.1 Evaluation of forcing approach

Deleted: Model

Averaged over the evaluation year, both WRF simulations capture the magnitude and variability of sub-diurnal near-surface meteorological conditions at most sites well (Fig. 3; Table 3). The interquartile range (IQR; range between upper and lower quartile) of MDs is one order of magnitude smaller than the observed standard deviation for all variables. As expected, variability is best captured for *T* and *PS*, with R^2 values that uniformly exceed 0.87 and 0.96, respectively. Those of *RH* have a larger range but a lower quartile above ~0.55. Compared with these variables, the model shows less skill in simulating sub-diurnal variability in winds, with lower quartiles of R^2 for *U* and *V* of approximately 0.39 and 0.27, respectively.

Shifting to daily timescales, both simulations represent variability in daily total *PR* surprisingly well, with the upper quartile of MDs below ~1.25 mm and lower quartiles of R^2 exceeding 0.18 and 0.33, depending on the simulation. The MD is positive at the majority of stations, indicating that WRF generally underestimates observed precipitation. The underestimate is likely greater than reported here, since the observations were not corrected for wind-induced undercatch. In addition to underestimating observed daily precipitation events (total sample size of 35,791 for all stations and record lengths), the simulations also produce false daily precipitation events, the vast majority of which are very small in magnitude (the median value in both WRF simulations is less than 0.1 mm/day). Considering wetter days (precipitation exceeding 1 mm/day; Ban et al., 2014), the number of false events is more than ten times smaller than the number of observed events (sample sizes of 3,096 and 2,249 in WRF_NO_NUDGE and WRF_NUDGE, respectively).

Deleted: IQRs

Previous studies evaluating WRF over this region have reported Root Mean Square Deviations (RMSD). For direct comparison, the mean RMSD in WRF_NUDGE for two-hourly *T* and *RH* is 2.67°C and 13.7%, respectively, and for daily total precipitation is 5.27 mm. These values are comparable to previous high-resolution applications of WRF over Bavaria (Warscher et al., 2019).

Deleted: K

Deleted: similar

Deleted: but lower than previously reported biases in

Deleted: at 5-km grid spacing

Deleted: for the period of 2001—2009

Examination of model biases on a monthly basis reveals further insights into the model performance (Fig. 4). The amplitude of the annual cycle is overpredicted in WRF, indicating that the good average agreement in *T* results from compensating biases: there is a cold bias in WRF in winter, a well-known issue with the model over snow-covered surfaces (e.g., Tomasi et al., 2017), and a warm bias in summer (Fig. 4a). The latter bias results in an underprediction of *RH* during this season (Fig. 4b), suggesting that WRF represents absolute humidity more accurately. The summer temperature bias is also more sustained than the winter one, resulting in the long tails (heads) in the distribution of MDs of *T* (*RH*) in Fig. 3. There is also a general

underprediction of near-surface winds from fall to early winter, as exemplified by the results for *U* in Fig. 4c and the slight positive skewness of the distribution of MDs for both *U* and *V* in Fig. 3, consistent with overly stable atmospheric conditions resulting from the cold bias. Finally, the model tends to overestimate precipitation in early spring and underestimate it in summer and fall. The reported seasonal and mean biases in daily precipitation are consistent with a potential underestimate of deep convection and convective precipitation at 1.5-km grid spacing. Although simulated mean precipitation shows a weak grid dependency below a spacing of ~ 4 km (Langhans et al., 2012), sub-kilometer spatial resolution is required to explicitly resolve the evolution and characteristics of clouds (e.g., Bryan et al., 2003; Craig and Dörnbrack, 2008; Prein et al., 2015).

Figure 5 shows a representative timeseries of *T* and *SW* for the station in Nürnberg (3668) in June 2018. The timeseries illustrates that the positive temperature bias in summer 2018 results from two distinct contributions. First, there is an overestimation of daytime maximum *T*, coinciding with an overestimation of *SW*. This relationship is observed both at Nürnberg and at the other three stations for which both datasets are available (Fig. 6a; cf. Sect. 2.2). The overestimation suggests there is an underestimation of either daytime cloudiness or its impact on incoming *SW* at the surface, likely stemming from the microphysics parameterization. Ban et al. (2014) identified similar processes underlying a warm bias in summer in a convection-permitting decadal simulation over central Europe. Second, there is an overestimation of night-time minimum *T*, suggesting that land-surface processes may play a role. Of the 101 stations with *T* measurements available, the dominant land-use categories of the grid cells containing stations are: *'Urban'* (10 sites); *'Dryland Cropland and Pasture'* (4 sites); *'Grassland'* (72 sites); *'Deciduous Broadleaf Forest'* (1 sites); *'Evergreen Needleleaf Forest'* (11 sites); and, *'Mixed Forest'* (3 sites). The overestimation of night-time *T* is greatest at stations located in grid cells classified as urban (Fig. 6b), consistent with a previous evaluation of WRF with the Noah-MP LSM for urban and rural stations in summer (Salamanca et al., 2018). The bias amplification in urban grid cells may reflect an incorrect classification of the underlying land surface in WRF, as only the München-Stadt station (id 3379) is listed as an urban station on the DWD's list for computing heat island effects. It may also result from an overestimation of heat storage when a mosaic approach is not used, and therefore the entire grid cell is treated as urban (Daniel Fenner, personal communication). The potential role of the land-surface specification or properties is reinforced by the comparison with MODIS data (Fig. 7), which shows the largest warm biases over grid cells classified as urban or croplands while biases are smaller in forested areas. There is also a cold bias along the foothills and at higher elevations in the Alps. The biases are slightly smaller in WRF_NUDGE than in WRF_NO_NUDGE, consistent with the station-based assessment.

In addition to factors internal to WRF, we note that the driving reanalysis data may also contribute to the warm bias, at least at some locations. From the available observations, 60 stations have both valid *T* data between June and August 2018 and a modelled elevation in ERA5 that is within ±100 m of reality. Averaged over the summer months and all stations, ERA5 has

a mean warm bias of 0.37°C. At 25 of the sites, a warm bias exceeding 0.5°C is present, with an average value over these sites of 0.92°C.

The inclusion of grid-analysis nudging leads to a small but nearly uniform improvement in agreement between observed and simulated variables. The distribution of MDs is closer to zero for all variables except *U* and *PS*, while those of MADs are closer for all variables (cf. Fig. 3 and Table 3). R^2 values are also uniformly higher when nudging is used, and the lowest lower-quartile value is 0.3 in WRF_NUDGE compared with only 0.18 in WRF_NO_NUDGE. Nudging produces a particularly noticeable improvement in simulated precipitation, halving the MD and nearly doubling the R^2 values (cf. Fig. 3, Fig. 4 and Table 3). Its usage also reduces the magnitude of the seasonal temperature biases and the number of extreme occurrences of the warm bias in summer (cf. Fig. 4 and Fig. 6). Considering daily timescales, the agreement of WRF_NUDGE with the observations is similar or even improved (Table 4): the mean MD is largely unaffected, but the average MAD decreases and average R^2 increases. Based on these improvements, grid-analysis nudging was adopted for the climatological simulations.

Deleted: 3.2 Impact of Grid Analysis Nudging¶

Deleted:

Deleted: (the temporal resolution of data available in BAYWRF),

Deleted:

3.2 Evaluation of BAYWRF

Deleted: 3 Climatological Simulations¶

Figure 8a compares simulated and observed annual mean *T* between 1988 and 2018 at the 33 stations with data available after filtering (cf. Sect. 2

Deleted: .4). WRF captures the variability in annual *T* well, although the data and its spread are slightly underestimated during the late 1990s and 2000s. The observations have a statistically significant trend…

Deleted: 0.28 K/decade (p = 0.04) while the model data do not (0.2 K/decade, p = 0.2) although this statistical analysis is very sensitive with such a small sample size. However, inspection of *T* trends by month reveals that WRF overestimates the slight cooling trends observed between January and March (statistically insignificant in both data sources; not shown), due to the winter cold bias and/or potential errors in snow depth and extent, which may contribute to an underestimate of the annual warming trend. Trends are otherwise well captured, including the three months with statistically significant observed trends: April (OBS: 0.83; WRF:0.75 K/decade); June (OBS: 0.75; WRF:0.71 K/decade); and, November (OBS: 0.72; WRF:0.71 K/decade). WRF similarly captures the variability of summer mean (June, July, August; JJA) temperature (Fig. 8b), although the warm bias at the station locations is again apparent. The observed trend is 0.49 K/decade compared with 0.48 in WRF ($p<0.05$). Furthermore, the model captures recent extreme summers in Western and Central Europe in 2003, 2015 and 2018, which in combination with drought have had severe impacts on economies, public health (e.g., Muthers et al., 2017), and primary productivity (e.g., Ciais et al., 2005). ¶

Averaged over the full simulation period, BAYWRF shows a similar magnitude of agreement with station *T* and *RH* data at daily timescales as found at sub-diurnal timescales for the single evaluation year (Fig. 8; cf. Fig. 3). For *T*, the MD has lower and upper quartiles of -0.3 and 0.4°C, respectively, while the values of R^2 uniformly exceed 0.92. For *RH,* the MD has lower and upper quartiles of 0.4 and 4.4% while the respective values for R^2 are 0.57 and 0.65. For *PREC*, the upper and lower quartiles of MDs considering days with observed precipitation are -0.1 and 0.1 mm while for R^2 the values are 0.41 and 0.47. A similar number and magnitude of wet false events are simulated (twenty times less than the sample size of observed events). Spatially, BAYWRF exhibits a positive bias in *T* and a negative bias in *RH* in the interior of Bavaria, and the converse anomalies in the pre-alpine and alpine areas in the south and along the eastern border of the region (Fig. 8a, c). The mean R^2 values for *RH* show a clear meridional gradient (Fig. 8d), which suggests that the model has some difficulty capturing processes governing near-surface moisture fluctuations in the southern part of Bavaria. Nonetheless, the highest correlation coefficients for observed precipitation events are found in this region (Fig. 8f). In addition, considering monthly precipitation sums, the centered pattern correlation between REGNIE and BAYWRF ranges from a lower quartile of 0.64 to an upper quartile of 0.82. Therefore, the characteristics of precipitation variability in time and space are captured by the dataset.

For BAYWRF, we note that in addition to the potential factors contributing to temperature biases discussed in Section 3.1, evaluation of the climatological simulation is also affected by discontinuities in station location and instrumentation. One example is Nürnberg (id 3668), which moved on 4 December 1995 from (49.4947ºN, 11.0806ºE) to (49.5030ºN, 11.0549ºE). The older station position is shifted one grid cell to the south and one grid cell to the west compared with its current location, corresponding to a shift in land use from urban (old position) to grasslands (new). Any discontinuities in location and

Deleted: Here

Deleted: .

underlying surface type are not captured since the most recent station positions are used for extracting meteorological data from D2. This potential source of discrepancies should be taken into consideration for climatological analyses (e.g., comparing observed and simulated trends).

4 Data Availability

Data from BAYWRF are available for download on the Open Science Framework (OSF; Collier, 2020; https://www.doi.org/10.17605/OSF.IO/AQ58B). Due to the size of the simulations, we have only provided daily mean data from the finest WRF domain (D2; 1.5-km grid spacing) after cropping close to the extent of Bavaria and removing vertical levels above ~ 200 hPa, amounting to 450 GB in total. Data are divided into three- and four-dimensional fields by year and month, with respective file sizes of ~150 MB and 1.1 GB. For the four-dimensional data, perturbation and base-state atmospheric pressure (WRF variables P and PB) and geopotential (PH and PHB) were combined to generate full model fields, while perturbation potential temperature (T) was converted to atmospheric temperature.

5 Conclusions

We presented a climatological kilometer-scale simulation with the atmospheric model WRF over Bavaria for the period of September 1987 to August 2018. Comparison of simulations for the period of September 2017 to August 2018 with and without grid-analysis nudging against extensive meteorological measurements across Bavaria showed that nudging decreased the mean deviations and increased the coefficient of determinations at the majority of sites for nearly all evaluated atmospheric variables, in particular precipitation. This approach was therefore adopted for generating the full BAYWRF dataset. In general, BAYWRF represents the variability of near-surface meteorological conditions well, albeit with both seasonal and spatial biases that are explored briefly here. Future users of this dataset are encouraged to rigorously evaluate biases for the variables and time periods relevant to their particular study areas and applications. BAYWRF provides a useful database for linking large-scale climate, as represented by the ERA5 reanalysis, to mesoscale climate over Germany, to local conditions in Bavaria, in a physically based way. The data are intended for dendroecological research applications but would also provide a valuable tool for investigations of the climate dependence of economic, societal, ecological, and agricultural processes in Bavaria.

6 Author contributions

EC performed the simulations, analyzed the data and wrote the manuscript. TM developed the study concept and wrote the manuscript.

Deleted: *T*

Deleted: ¶

Deleted: Spatially distributed trends in *T* are strongest and significant over the largest in area during JJA (Fig. 9a; other seasons not shown), ranging from ~0.3 to 0.7 K/decade over Bavaria. Trends in precipitable water are likewise uniformly positive over the study region, ranging from ~0.2 to 0.3 mm/decade. The trends of both fields also have a positive gradient between southwestern and northeastern Bavaria. These results agree qualitatively and quantitively with previous studies (e.g., Alshawaf et al., 2017). ¶

Deleted: (OSF; Collier, 2020; https://www.doi.org/10.17605/OSF.IO/AQ58B)…

Deleted: Subsets of D1 or sub-diurnal data can be made available upon request. …

Deleted: , convection-permitting

Deleted: For

Deleted: evaluation

Deleted: , we compared the simulations

Deleted: and evaluated the impact of using grid-analysis nudging. We found that the model reproduced variability in near-surface

Deleted: the model reproduced variability in near-surface meteorological conditions well, although seasonal temperature biases were present. Grid analysis …

Deleted: correlations between simulations and observations

Deleted: with a particularly noticeable improvement for simulated daily precipitation.

7 Competing interests

The authors declare that they have no conflict of interest.

8 Acknowledgements

This project is sponsored by the Bavarian State Ministry of Science and the Arts in the context of the Bavarian Climate Research Network (bayklif). We gratefully acknowledge the compute resources and support provided by the Erlangen Regional Computing Center (RRZE) and we thank Thomas Zeiser for his assistance with the timely completion of the simulations.

References

Alshawaf, F., Balidakis, K., Dick, G., Heise, S. and Wickert, J.: Estimating trends in atmospheric water vapor and temperature time series over Germany, Atmos. Meas. Tech., doi:10.5194/amt-10-3117-2017, 2017.

Berg, P., Wagner, S., Kunstmann, H. and Schädler, G.: High resolution regional climate model simulations for Germany: Part I-validation, Clim. Dyn., doi:10.1007/s00382-012-1508-8, 2013.

Beyer, M.: Hitzewelle Sommer 2018 - Einordnung und Ausblick, [online] Available from: https://www.dwd.de/DE/wetter/thema_des_tages/2018/8/6.html (Accessed 1 August 2019), 2018.

Bowden, J. H., Nolte, C. G. and Otte, T. L.: Simulating the impact of the large-scale circulation on the 2-m temperature and precipitation climatology, Clim. Dyn. [online] Available from: http://link.springer.com/article/10.1007/s00382-012-1440-y, 2013.

Bräuning, A., De Ridder, M., Zafirov, N., García-González, I., Dimitrov, D. P. and Gärtner, H.: TREE-RING FEATURES: INDICATORS of EXTREME EVENT IMPACTS, IAWA J., doi:10.1163/22941932-20160131, 2016.

Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogée, J., Allard, V., Aubinet, M., Buchmann, N., Bernhofer, C., Carrara, A., Chevallier, F., De Noblet, N., Friend, A. D., Friedlingstein, P., Grünwald, T., Heinesch, B., Keronen, P., Knohl, A., Krinner, G., Loustau, D., Manca, G., Matteucci, G., Miglietta, F., Ourcival, J. M., Papale, D., Pilegaard, K., Rambal, S., Seufert, G., Soussana, J. F., Sanz, M. J., Schulze, E. D., Vesala, T. and Valentini, R.: Europe-wide reduction in primary productivity caused by the heat and drought in 2003, Nature, doi:10.1038/nature03972, 2005.

Collier, E.: BAYWRF, [online] Available from: https://www.doi.org/10.17605/OSF.IO/AQ58B, 2020.

Collier, E., Mölg, T. and Sauter, T.: Recent atmospheric variability at Kibo summit, Kilimanjaro, and its relation to climate mode activity, J. Clim., 31(10), 3875–3891, doi:10.1175/JCLI-D-17-0551.1, 2018.

Collier, E., Sauter, T., Mölg, T. and Hardy, D.: The influence of tropical cyclones on circulation, moisture transport, and snow accumulation at Kilimanjaro during the 2006 - 2007 season, J. Geophys. Res. Atmos., 2019.

Copernicus Climate Change Service (C3S): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global

climate., Copernicus Clim. Chang. Serv. Clim. Data Store [online] Available from: https://cds.climate.copernicus.eu/cdsapp#!/home (Accessed 16 June 2019), 2017.

Esty, W. W. and Banfield, J. D.: The Box-Percentile Plot, J. Stat. Softw., doi:10.18637/jss.v008.i17, 2003.

Fosser, G., Khodayar, S. and Berg, P.: Benefit of convection permitting climate model simulations in the representation of convective precipitation, Clim. Dyn., doi:10.1007/s00382-014-2242-1, 2014.

Frías, M. D., Zorita, E., Fernández, J. and Rodríguez-Puebla, C.: Testing statistical downscaling methods in simulated climates, Geophys. Res. Lett., doi:10.1029/2006GL027453, 2006.

Fuchs, R., Herold, M., Verburg, P. H. and Clevers, J. G. P. W.: A high-resolution and harmonized model approach for reconstructing and analysing historic land changes in Europe, Biogeosciences, doi:10.5194/bg-10-1543-2013, 2013.

Giorgi, F. and Mearns, L. O.: Approaches to the simulation of regional climate change: A review, Rev. Geophys., doi:10.1029/90RG02636, 1991.

Gohm, A., Mayr, G. J., Fix, A. and Giez, A.: On the onset of bora and the formation of rotors and jumps near a mountain gap, Q. J. R. Meteorol. Soc., doi:10.1002/qj.206, 2008.

Held, I. M. and Soden, B. J.: Robust responses of the hydrological cycle to global warming, J. Clim., doi:10.1175/JCLI3990.1, 2006.

Hochreuther, P., Wernicke, J., Grießinger, J., Mölg, T., Zhu, H., Wang, L. and Bräuning, A.: Influence of the Indian Ocean Dipole on tree-ring δ18O of monsoonal Southeast Tibet, Clim. Change, doi:10.1007/s10584-016-1663-8, 2016.

Hong, S. Y., Noh, Y. and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, Mon. Weather Rev. [online] Available from: http://journals.ametsoc.org/doi/abs/10.1175/MWR3199.1, 2006.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A. and Collins, W. D.: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models, J. Geophys. Res., 113(D13), D13103, doi:10.1029/2008JD009944, 2008.

Jarvis, A., Reuter, H. I., Nelson, A. and Guevara, E.: Hole-filled SRTM for the globe Version 4, available from CGIAR-CSI SRTM 90m Database (http//srtm. csi. cgiar. org), 2008.

Jiménez, P. A., Dudhia, J., González-Rouco, J. F., Navarro, J., Montávez, J. P. and García-Bustamante, E.: A Revised Scheme for the WRF Surface Layer Formulation, Mon. Weather Rev., 140(3), 898–918 [online] Available from: http://journals.ametsoc.org/doi/abs/10.1175/MWR-D-11-00056.1, 2012.

Kain, J. S.: The Kain–Fritsch Convective Parameterization: An Update, J. Appl. Meteorol., 43(1), 170–181, 2004.

Lindner, M., Fitzgerald, J. B., Zimmermann, N. E., Reyer, C., Delzon, S., van der Maaten, E., Schelhaas, M. J., Lasch, P., Eggers, J., van der Maaten-Theunissen, M., Suckow, F., Psomas, A., Poulter, B. and Hanewinkel, M.: Climate change and European forests: What do we know, what are the uncertainties, and what are the implications for forest management?, J. Environ. Manage., doi:10.1016/j.jenvman.2014.07.030, 2014.

Liu, Y., Chen, F., Warner, T. and Basara, J.: Verification of a mesoscale data-assimilation and forecasting system for the Oklahoma City area during the joint urban 2003 field project, J. Appl. Meteorol. Climatol., doi:10.1175/JAM2383.1, 2006.

Di Luca, A., Argüeso, D., Evans, J. P., De Elía, R. and Laprise, R.: Quantifying the overall added value of dynamical downscaling and the contribution from different spatial scales, J. Geophys. Res., doi:10.1002/2015JD024009, 2016.

Mölg, T., Maussion, F. and Scherer, D.: Mid-latitude westerlies as a driver of glacier variability in monsoonal High Asia, Nat. Clim. Chang. [online] Available from: http://www.nature.com/nclimate/journal/v4/n1/full/nclimate2055.html, 2014.

Morrison, H., Thompson, G. and Tatarskii, V.: Impact of Cloud Microphysics on the Development of Trailing Stratiform Precipitation in a Simulated Squall Line: Comparison of One- and Two-Moment Schemes, Mon. Weather Rev., 137(3), 991–1007, doi:10.1175/2008MWR2556.1, 2009.

Muthers, S., Laschewski, G. and Matzarakis, A.: The summers 2003 and 2015 in South-West Germany: Heat waves and heat-related mortality in the context of climate change, Atmosphere (Basel)., doi:10.3390/atmos8110224, 2017.

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M. and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, J. Geophys. Res., 116(D12) [online] Available from: http://www.agu.org/pubs/crossref/2011/2010JD015139.shtml, 2011.

Otte, T. L., Nolte, C. G., Otte, M. J. and Bowden, J. H.: Does nudging squelch the extremes in regional climate modeling?, J. Clim., 25(20), 7046–7066, 2012.

Salamanca, F., Zhang, Y., Barlage, M., Chen, F., Mahalov, A. and Miao, S.: Evaluation of the WRF-Urban Modeling System Coupled to Noah and Noah-MP Land Surface Models Over a Semiarid Urban Environment, J. Geophys. Res. Atmos., doi:10.1002/2018JD028377, 2018.

Skamarock, W. C. and Klemp, J. B.: ScienceDirect - Journal of Computational Physics : A time-split nonhydrostatic atmospheric model for weather research and forecasting applications, J. Comput. Phys., 227, 3465–3485, 2008.

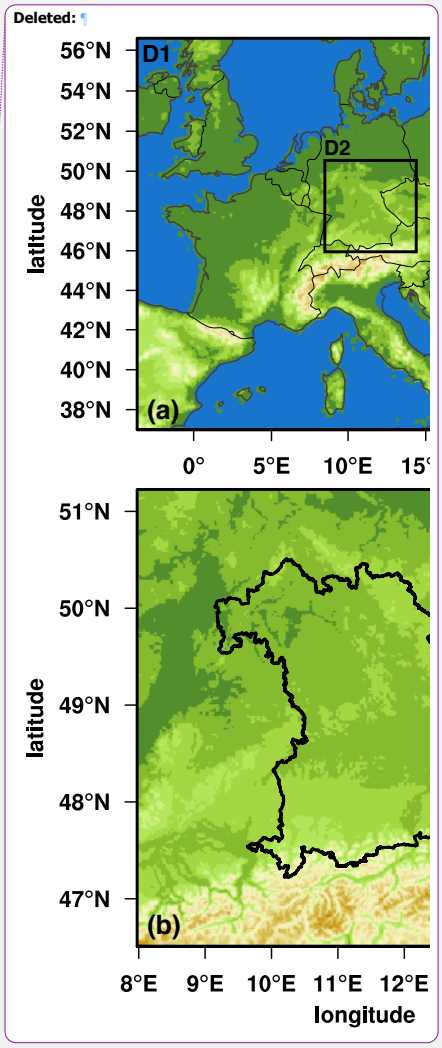
Tans, P. and Keeling, R.: Trends in Atmospheric Carbon Dioxide, [online] Available from: https://www.esrl.noaa.gov/gmd/ccgg/trends/data.html (Accessed 1 August 2019), 2019.

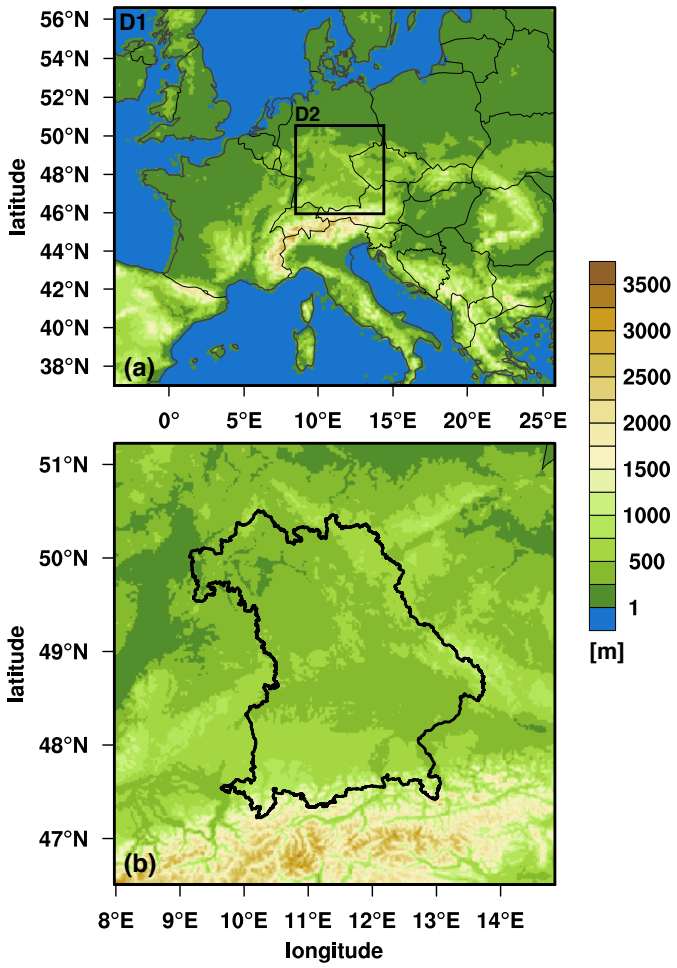
Tomasi, E., Giovannini, L., Zardi, D. and de Franceschi, M.: Optimization of Noah and Noah_MP WRF land surface schemes in snow-melting conditions over complex terrain, Mon. Weather Rev., doi:10.1175/MWR-D-16-0408.1, 2017.

Vionnet, V., Six, D., Auger, L., Dumont, M., Lafaysse, M., Quéno, L., Réveillet, M., Dombrowski-Etchevers, I., Thibert, E. and Vincent, C.: Sub-kilometer Precipitation Datasets for Snowpack and Glacier Modeling in Alpine Terrain, Front. Earth Sci., doi:10.3389/feart.2019.00182, 2019.

Warscher, M., Wagner, S., Marke, T., Laux, P., Smiatek, G., Strasser, U. and Kunstmann, H.: A 5 km Resolution Regional Climate Simulation for Central Europe: Performance in High Mountain Areas and Seasonal, Regional and Elevation-Dependent Variations, Atmosphere (Basel)., 10(11), 682, 2019.

Deleted: ¶





515

Figure 1: Extent and modelled topographic height in WRF D1 (a) and D2 (b). The extent of D2 and of Bavaria are delineated in black in the top and bottom panels, respectively.

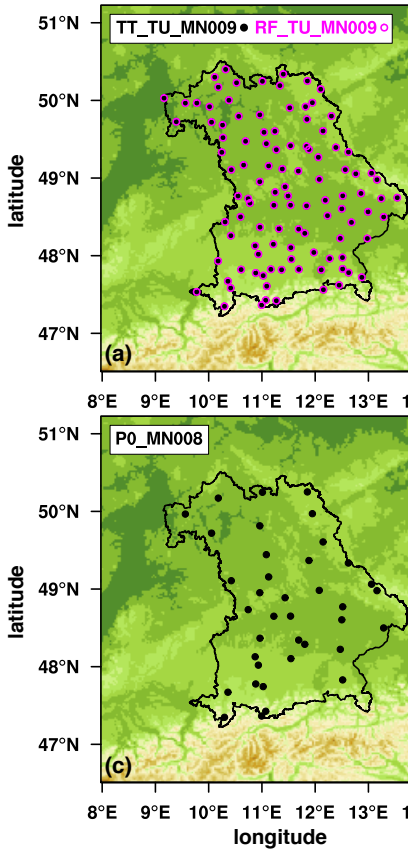
Formatted: Justified

Deleted: [m]

Deleted: ¶

¶

¶



Formatted: Font: Not Bold

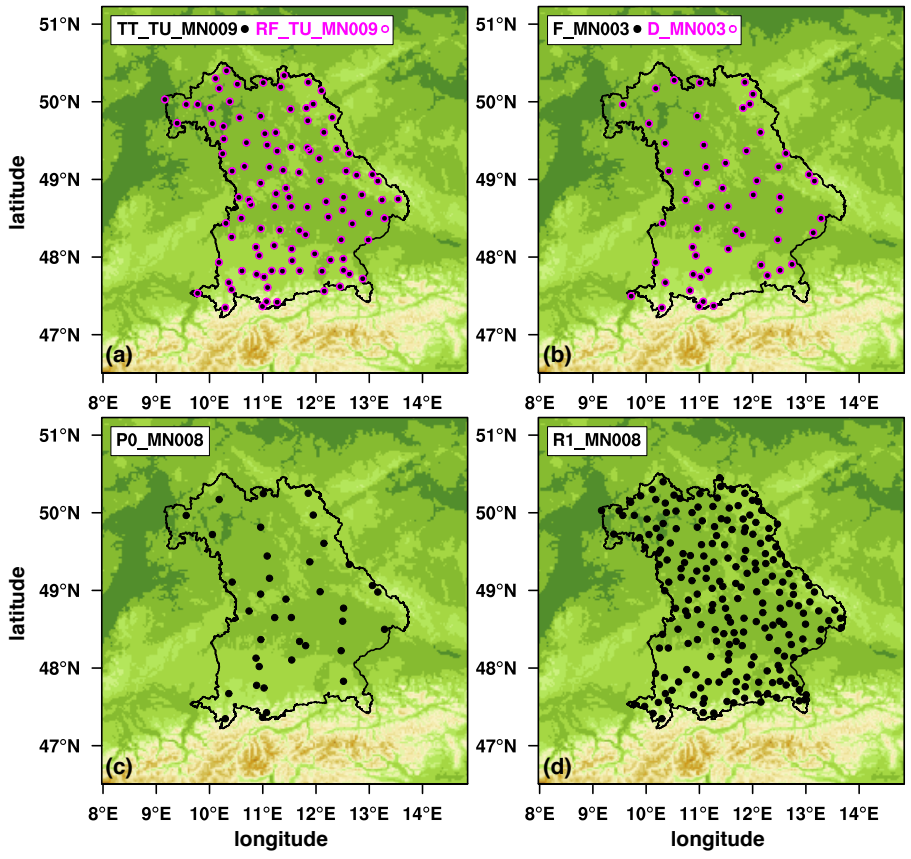
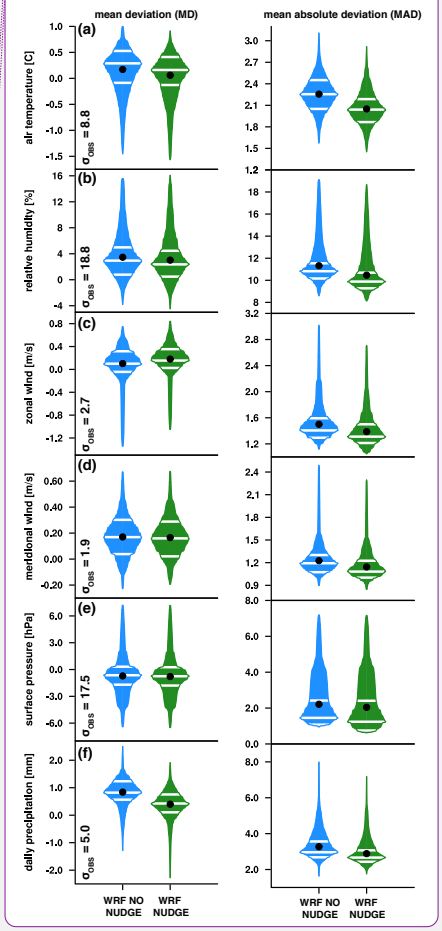


Figure 2: The location of the stations used for model evaluation during the most recent simulation year (September 2017 to August 2018) for each dataset listed in Table 2. Datasets labelled in black are shown by filled black circles, while datasets labelled in pink are shown by open pink circles, illustrating that locations for measurements of air temperature and humidity (a; TT_TU_MN009 & RF_TU_MN009) and of wind speed and direction (b; F_MN003 & D_MN003) were the same. The locations for measurements of surface pressure (P0_MN008) and of precipitation (R1_MN008) are shown in panels c and d, respectively.

Formatted: Caption, Left

Deleted: ¶



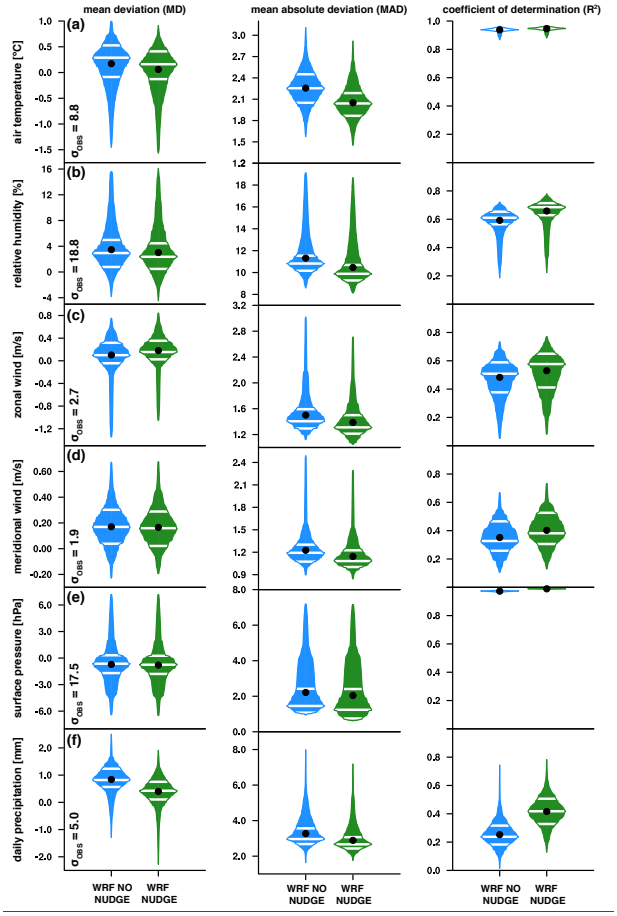
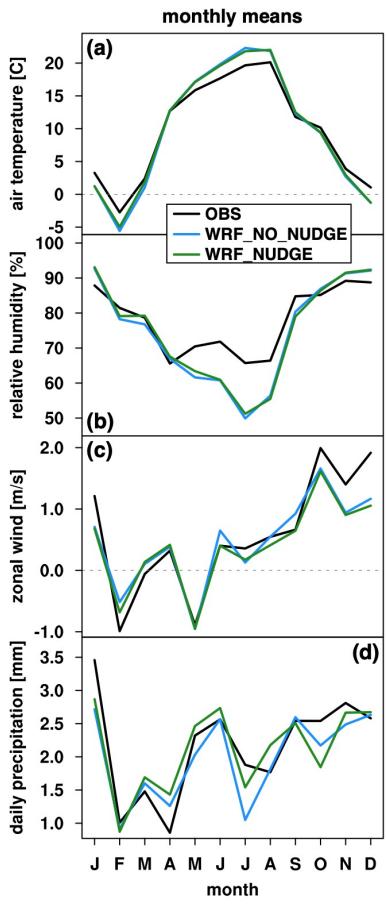


Figure 3: Box-percentile plots (Esty and Banfield, 2003) of mean deviation (MD), mean absolute deviation (MAD), and coefficient of determination (R^2) between observations and the two WRF simulations, WRF_NO_NUDGE (blue) and WRF_NUDGE (green), for 2-m (a) air temperature and (b) relative humidity, 10-m (c) zonal and (d) meridional winds, (e) surface pressure and (f) precipitation. The statistics for all variables except for precipitation were computed from two-hourly instantaneous values, while those for precipitation were computed using daily totals. The shape of the plots shows the distribution of data over their range of values, white lines delineate 25th, 50th and 75th percentiles, and a black dot indicates the mean. The observed standard deviation (σ_{obs}) for each variable is provided in the left column.

Formatted: Justified

Deleted: ¶



Formatted: Font: Not Bold

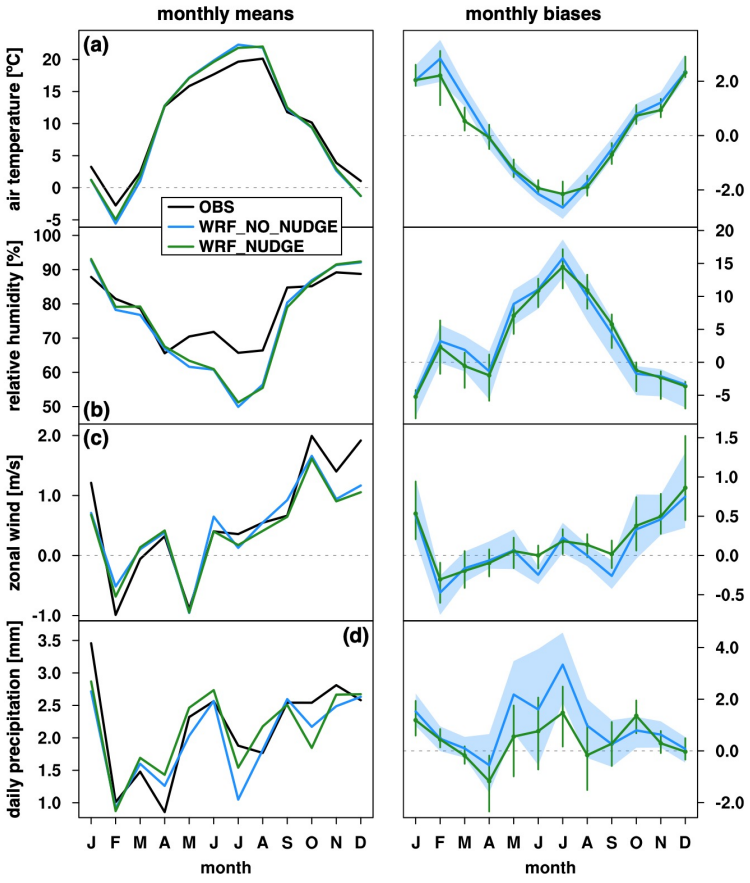
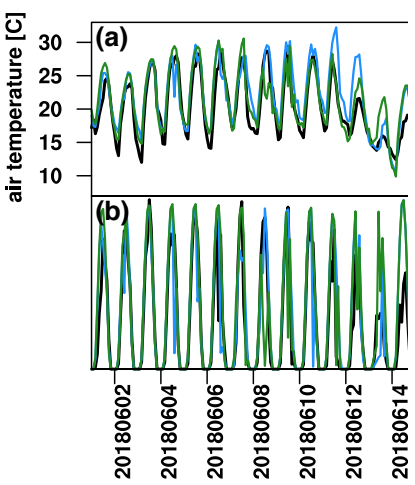


Figure 4: Timeseries of monthly mean 2-m (a) air temperature and (b) relative humidity, (c) 10-m zonal winds, and (d) daily total precipitation (left column) between September 2017 and August 2018. Observational, WRF_NO_NUDGE and WRF_NUDGE data are shown in black, blue and green, respectively. Timeseries of monthly mean biases of the same variables (right column). The mean bias over all stations is shown for each simulation using the same colour assignment, while the lower and upper quartile of the station biases is shown as a blue polygon and green bars for WRF_NO_NUDGE and WRF_NUDGE data, respectively.

Deleted: ¶



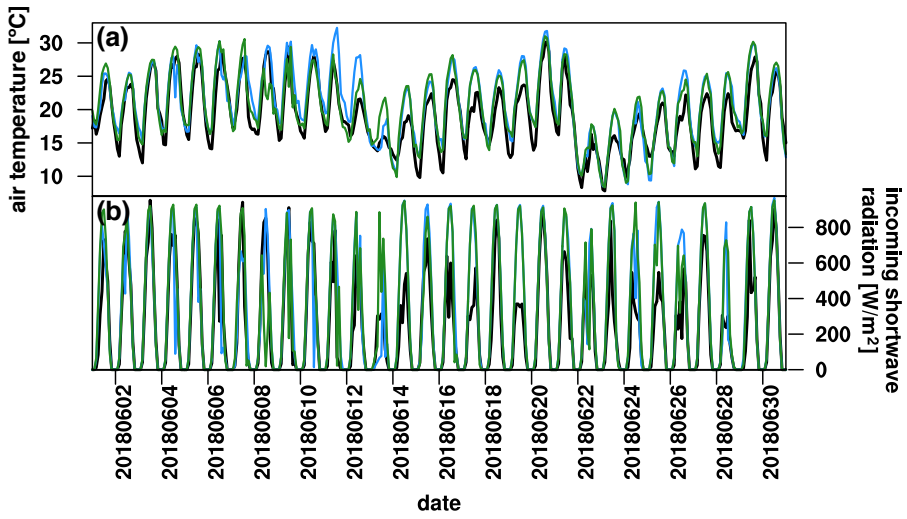
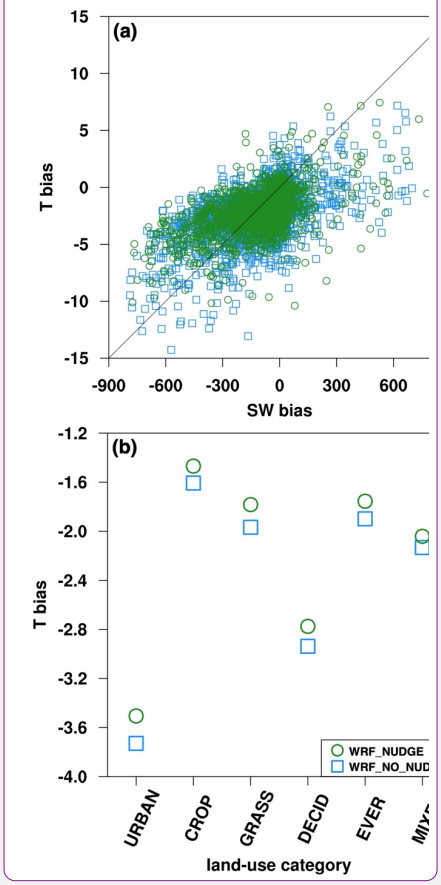


Figure 5: Timeseries of (a) 2-m air temperature and (b) incoming shortwave radiation at the station in Nürnberg (id 3668) from 1 June to 1 July 2018. Observational, WRF_NO_NUDGE and WRF_NUDGE data are shown in black, blue and green, respectively.

Formatted: Caption, Left

Deleted:



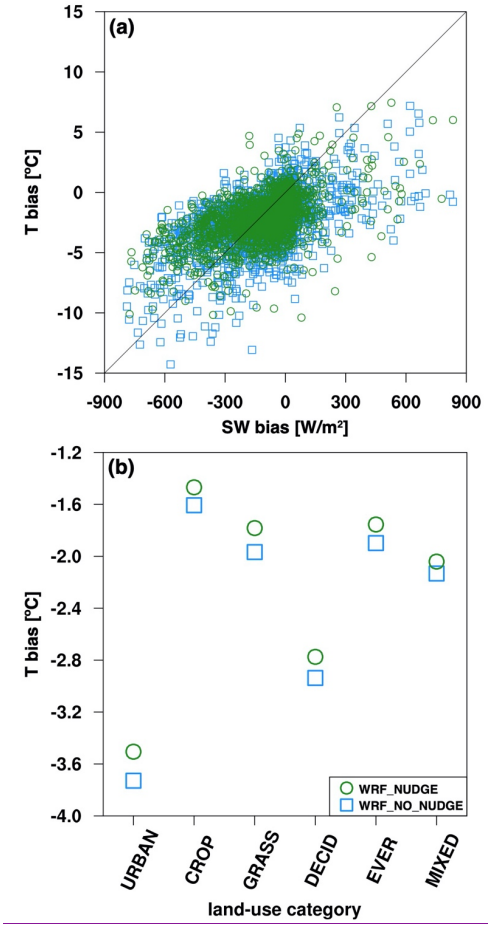


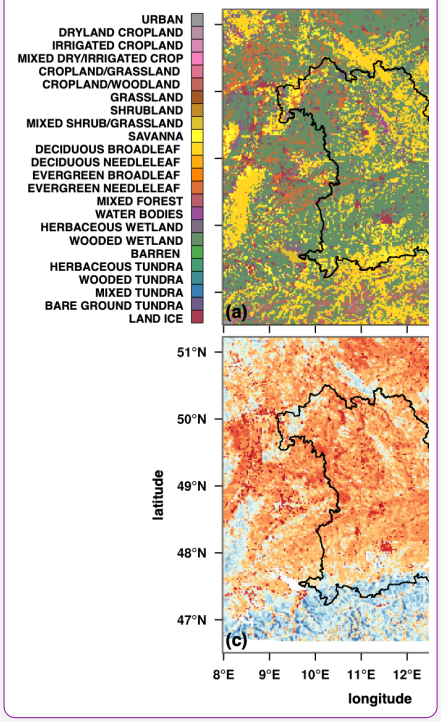
Figure 6: Scatter plots of (a) air temperature bias vs. incoming shortwave radiation bias and (b) air temperature bias vs. land-use category in closest grid cell to station. The category abbreviations from left to right describe: 'Urban and Built-Up Land' (10 sites); 'Dryland Cropland and Pasture' (4 sites); 'Grassland' (72 sites); 'Deciduous Broadleaf Forest' (1 sites); 'Evergreen Needleleaf Forest' (11 sites); and, 'Mixed Forest' (3 sites). For both panels, data from WRF_NO_NUDGE and WRF_NUDGE are displayed as blue square and green circle markers, respectively.

Deleted: [°C]

Deleted: [W m^{-2}]

Deleted: [°C]

Deleted: ¶



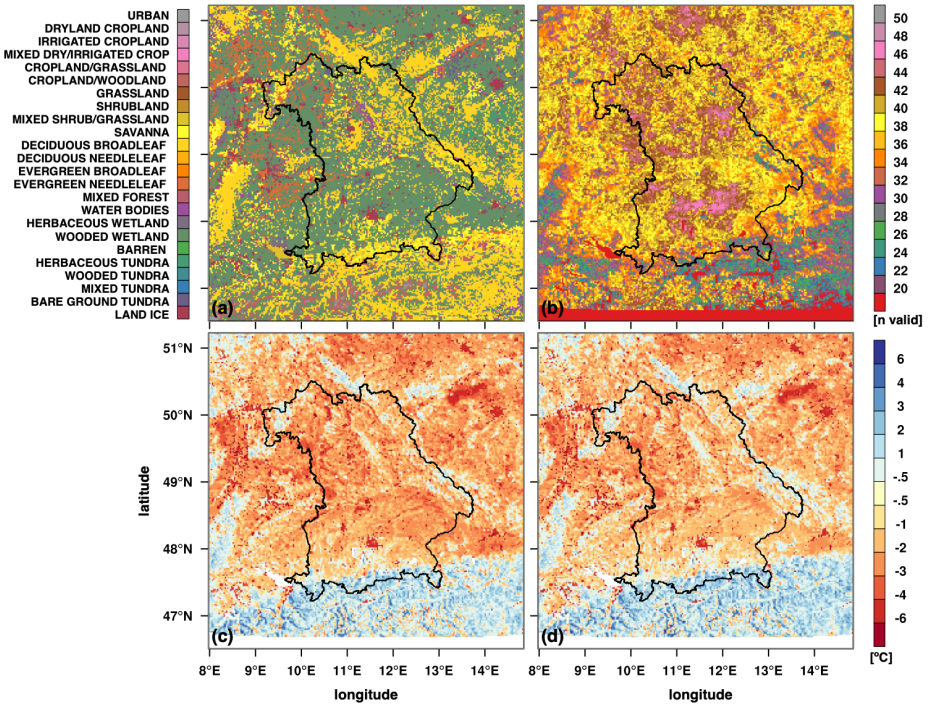
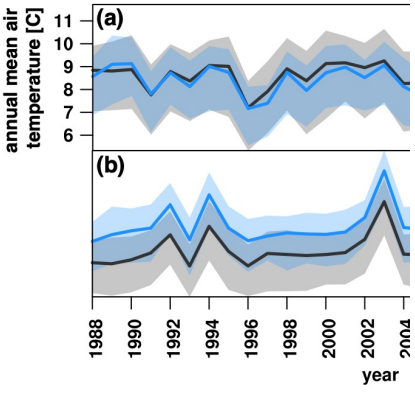


Figure 7: (a) Land-use classification in D2. (b) Number of timesteps with valid night-time LST data in the MODIS MYD11A1 dataset between 1 June and 31 August 2018 out of a maximum of 52 with less than 50% missing data in D2. The average difference in observed and simulated LST for (c) WRF_NO_NUDGE and (d) WRF_NUDGE. Note that the orange and red colours in panels c and d shade areas where WRF is warmer than MODIS (MODIS minus WRF is negative) and vice versa for blues.

Formatted: Caption, Left

Deleted: ¶



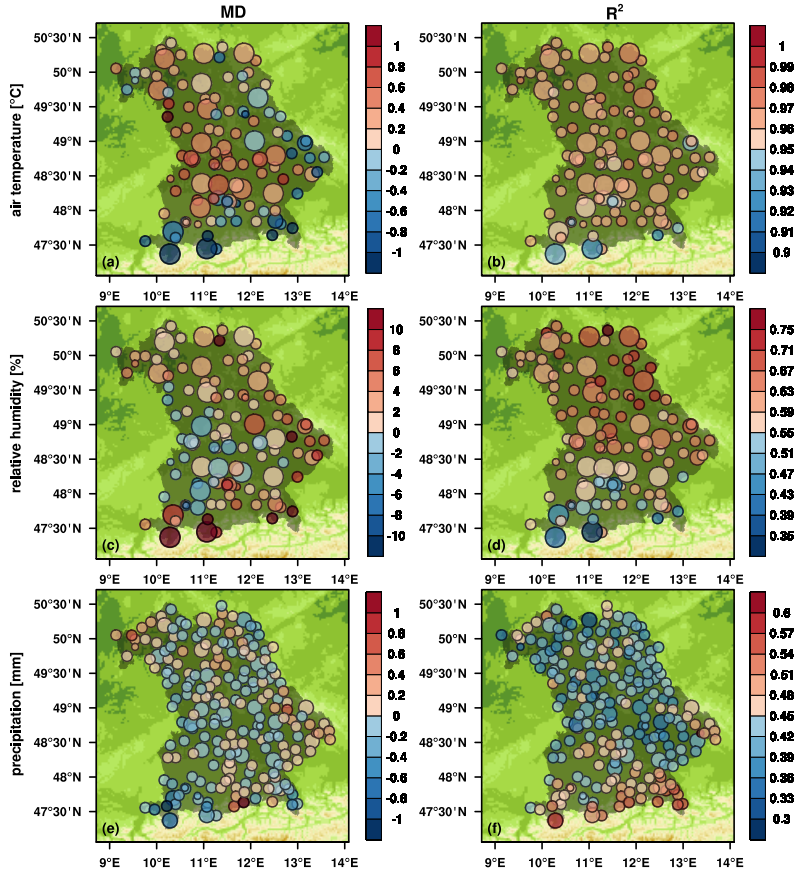


Figure 8: Spatial maps of mean MD (left column) and R^2 (right) at all stations with valid data between September 1987 and August 2018 for daily mean (a, b) *T* and (c, d) *RH*, and for daily total (e, f) *PREC*. The four marker sizes group the percentage of the total timesteps (11,323 days) for which data were available at each station into the four quartiles. The largest marker size, which delineates records with more than 75% valid data points, is therefore not available for *PREC*, as this dataset begins on 1 September 1995.

Formatted: Caption

Deleted: **Figure 8:** Comparison of (a) annual mean and (b) JJA mean 2-m air temperature from 1988 to 2018 averaged over all observations (black curve) and the corresponding grid points in WRF (blue curve). The shaded polygons delineate the range of values. To maximize data availability, annual means were computed from September of year *n-1* to August of year *n*.¶

... [1]

Table 1: A summary of the WRF configuration used for the simulations.

Table 1: WRF configuration		
Domain configuration		
Horizontal grid spacing	7.5 & 1.5 km (D1–2)	
Grid dimensions	351x301, 351x351	
Time step	45 & 9 s	
Vertical levels	60	
Model top pressure	10 hPa	
Model physics		
Radiation	RRTMG	(Iacono et al., 2008)
Microphysics	Morrison	(Morrison et al., 2009)
Cumulus	Kain-Fritsch (none in D2)	(Kain, 2004)
Planetary boundary layer	Yonsei State University	(Hong et al., 2006)
Atmospheric surface layer	Monin Obukhov	(Jiménez et al., 2012)
Land surface	Noah-MP	(Niu et al., 2011)
Dynamics		
Top boundary condition	Rayleigh damping	
Diffusion	Calculated in physical space	

Deleted: 1000

Deleted: ¶
¶

Table 2: A summary of data used for model evaluation. Rows highlighted in grey provide information about observational data from the DWD CDC Data Portal, whose measurement locations for the evaluation for the 2017 to 2018 period are shown in Figure 2.

Dataset Name	Variable [unit]	Temporal Resolution	Total Stations in Bavaria 2017-2018 (1987-2018)	Version	Access URL	Last Accessed	Dataset Description
TT_TU_MN009	2-m air temperature [°C]	Hourly	106 (120)	v19.3	https://cdc.dwd.de/portal	10 Sep 2020	https://cdc.dwd.de/sdi/pid/TT_TU_MN009/DESCRIPTION_TT_TU_MN009_en.pdf
RF_TU_MN009	2-m relative humidity [%]	Hourly	106 (120)	v19.3	https://cdc.dwd.de/portal		https://cdc.dwd.de/sdi/pid/RF_TU_MN009/DESCRIPTION_RF_TU_MN009_en.pdf
F_MN003	10-m wind speed [m/s]	Hourly	57	v19.3	https://cdc.dwd.de/portal		https://cdc.dwd.de/sdi/pid/F_MN003/DESCRIPTION_F_MN003_en.pdf
D_MN003	10-m wind direction [deg]	Hourly	57	v19.3	https://cdc.dwd.de/portal		https://cdc.dwd.de/sdi/pid/D_MN003/DESCRIPTION_D_MN003_en.pdf
P0_MN008	surface pressure [hPa]	Hourly	38	v19.3	https://cdc.dwd.de/portal		https://cdc.dwd.de/sdi/pid/P0_MN008/DESCRIPTION_P0_MN008_en.pdf
R1_MN008	precipitation [mm]	Hourly	213 (219)	v19.3	https://cdc.dwd.de/portal		https://cdc.dwd.de/sdi/pid/R1_MN008/DESCRIPTION_R1_MN008_en.pdf
Hourly station observations of solar incoming (total/diffuse) and longwave downward radiation for Germany	Incoming longwave and shortwave radiation [J/cm2]	Hourly	10	recent	https://cdc.dwd.de/portal		https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/hourly/solar//DESCRIPTION_obsgermany_climate_hourly_solar_en.pdf
REGNIE	precipitation [mm]	Monthly sum	--	recent	https://opendata.dwd.de/climate_environment/CDC/grids_germany/monthly/regnie/		https://opendata.dwd.de/climate_environment/CDC/grids_germany/monthly/regnie/DESCRIPTION_gridsgermany_monthly_regnie_en.pdf
MODIS MYD11A1	land surface temperature [K]	Daily	--	v006	https://lpdaacsvc.cr.usgs.gov/appeears		https://lpdaac.usgs.gov/products/myd11a1v006/

Deleted:

Dataset Name	Variable [unit]	Temporal Resolution	Total Stations in Bavaria
TT_TU_MN009	2-m air temperature [°C]	Hourly	106
RF_TU_MN009	2-m relative humidity [%]	Hourly	106
F_MN003	10-m wind speed [m/s]	Hourly	57
D_MN003	10-m wind direction [deg]	Hourly	57
P0_MN008	surface pressure [hPa]	Hourly	38
R1_MN008	precipitation [mm]	Hourly	213
Hourly station observations of solar incoming (total/diffuse) and longwave downward radiation for Germany	Incoming longwave and shortwave radiation [J/cm2]	Hourly	10
MO_TT_MN004	2-m air temperature [°C]	Monthly	244
MODIS MYD11A1	land surface temperature [K]	Daily	--

¶

Table 3: A summary of the statistical evaluation of the WRF_NO_NUDGE (italics) and WRF_NUDGE (bold italics) simulations, considering the whole evaluation period of 1 September 2017 to 1 September 2018. The table presents the mean deviation (MD), the mean absolute deviation (MAD) and the coefficient of determination (R^2) for two-hourly 2-m air temperature (*T*) and relative humidity *RH*), 10-m zonal wind (*U*) and meridional wind (*V*), surface pressure (*PS*), and daily total precipitation (*PR*). All computations are made from observations minus model data.

Variable	MD	MAD	R2
T (WRF_NO_NUDGE)	0.2	2.3	0.94
T (WRF_NUDGE)	**0.1**	**2.0**	**0.95**
RH	3.5	11.3	0.59
RH	**3.0**	**10.5**	**0.66**
U	0.1	1.5	0.48
U	**0.2**	**1.4**	**0.53**
V	0.2	1.2	0.35
V	**0.2**	**1.1**	**0.40**
PS	-0.7	2.2	0.97
PS	**-0.8**	**2.0**	**0.99**
PR	0.8	3.3	0.25
PR	**0.4**	**2.9**	**0.42**

Table 4: Same as Table 3 but for daily mean variables in WRF_NUDGE only.

Variable	MD	MAD	R2
T	**0.1**	**1.7**	**0.97**
RH	**3.0**	**8.4**	**0.71**
U	**0.2**	**0.9**	**0.72**
V	**0.2**	**0.6**	**0.64**
PS	**-0.8**	**2.0**	**0.99**

Formatted: Not Highlight

Page 20: [1] Deleted Revised 16/09/2020 15:19:00