Dear Professor Zehe, dear reviewers,

thank you for your valuable comments. Based on these comments we could improve the manuscript and revised the following main points:

1. Specification of the used variogram (as requested by Erwin Zehe).

2. Estimation of the interpolation errors based on the Jacknife method (as requested by Erwin Zehe).

3. Comparison of evapotranspiration estimates from mHM and MODIS data.

4. Relation of land surface and hydro-climatic conditions with model performance of the 222 basins.

5. An additional section in the introduction discussion the parametrization of large scale models as well as a revision of the section discussing the objectives of the study.

6. A reformulation of section 4.1.

7. Discussing of additional studies in the results and discussion section.

For further details please find below the point-by-point answers and the marked-up manuscript version.

Kind regards
Matthias Zink, Rohini Kumar, Matthias Cuntz, and Luis Samaniego

1 Anonymous Referee #1

1.1 Overall:

This paper provides an excellent and useful product, and the complexities and methods used to derive this product. There are some minor organizational and grammatical errors, but overall I think this paper is a good contribution to broad-scale hydrologic modeling and analysis.
Thank you for offering such encouraging and detailed suggestions for improving the structure of the paper. We appreciate your efforts to improve the structure and the clarity of the paper. We revised the introduction as suggested by the reviewer. We added additional analyses for addressing the questions: 1) How the assumption of a static land cover before 1990 impacts the model results, and 2) What are the relative characteristics of the 222 catchments and how do they relate to model performance. In the following, we present the referee's comments as well as our point-by-point response to all of them.
The paper could benefit from some additional attention to organization, specifically in the introductory sections. The authors limit themselves in the stated aim of the paper in the introduction, and go on to state other aims later in the paper. The aim as stated in the introduction is to derive a consistent set of national-scale hydrologic data for Germany at high spatial and temporal resolutions. If this was the extent of the paper, I would recommend that this be

re-submitted as a methods paper; however, the authors go on to append additional aims/goals in the body of the paper which go beyond this, such as: Page 5, line 22: to derived consistent model parameters to perform nationwide simulations of water fluxes and states. Page 7, line 8: to analyze the temporal dynamics of soil moisture Page 13, line 21: spatio-temporal differences of uncertainties caused by the 100 ensemble parameter sets As it is written, this paper reads as an aggregation of papers instead of one cohesive contribution. This could be easily fixed by restating the aim in the introduction to include all the parametric uncertainty analyses that are presented in the paper. Please gather and reassess the purpose and scope of the entire papers contribution to the field of hydrologic science, and state this in the introduction.

We thank the reviewer for his/her assessment of the manuscript. We revised the manuscript based on your suggestion and reorganize the (methodological) aims in the introduction.

There are also several spots in the paper which could be improved by directly assessing the limitations of the data or analyses. An example of where this is done well in the paper is page 7, lines 7-8: A direct comparison between observed and simulated soil moisture may therefore be misleading due to differences in spatial representativeness and sampling depth.

We address the limitations of data and analyses where we identified them in the revised manuscript, e.g., the limitations of an observational driven, simulated hydrologic dataset.

1.2 Specific comments

Page 1, line 24: Formatting of the citation does not match others.
We edited the format of the citation.
Page 2, line 6: State limitations of using observational data.
Thank you for pointing this out. We mention this limitations in the revised manuscript as "First, due to a limited amount of observed variables modeling approaches like the estimation of potential evapotranspiration have to be adopted to the available data. In consequence temperature based methods may be preferred to more physically based radiation approaches. Second, the interpolation of point observations induces uncertainties depending on the applied interpolation method. Further, small-scale, convective precipitation events may not be caught by gauging networks and lead to an underestimation in precipitation."
Page 2, line 9: add contiguous or continental United States.
Done.
Page 2, line 13-14: grammar. ... who stated a need for higher-resolution spatial data and models...
Thank you, done.
Page 3, line 11: add entirely, only catchment entirely covered by German territory
We changed that accordingly.
Page 3, line 17: grammar, average discharge of the seven catchments ranges
Thank you, done.
Page 4, line 1-2: How does this assumption of static land cover before 1990 impact results?

The impact of changing land cover is remarkably high near urban areas, since most of the changes there happened between 1950 and 2010. The effects of these changes are, however, low at the model resolution of $4\times4\text{km}^2$ as the table underneath shows. This table shows the differences between the hydrological state and fluxes between two scenarios. In the first scenario the land cover is fixed to the state of 1990. In the second scenario we fix the land cover to the conditions observed in 2006. The model time period is 1951 to 2010 and the domain is Germany. The comparison is based on daily values of the respective flux or state. Mean relative biases of less than 5% indicate that the assumption of static land cover has an impact on the modeled fluxes and states which is low compared to the effect of the parametric uncertainty. Changes apart from urbanization will have low effects on the modeled hydrological variables because mHM works with three land cover classes, i.e., sealed (mostly urban), forest and a mixed class. We restricted our study to only 3 land cover scenes because the well established CORINE data are only available for the years 1996, 2000, and 2006.

The table shows the mean and standard deviation between 2 land cover scenarios.

variable	bias	rel. bias	correlation	RMSE
	[mm d^{-1}]	[%]	[-]	[mm d^{-1}]
evapotranspiration	0±0.02	0.22±1.46	1.0±0	0.01±0.03
soil moisture	0±0	-0.02±0.67	1.0±0	0±0
generated runoff	0±0.02	-0.57±4.91	1.0±0.03	0.03±0.07
recharge	0±0.01	0.43±2.79	1.0±0	0±0.01

Page 4, line 2: What are your aggregation/resampling methods?
We remapped the data using a nearest neighbor approach. We address this question in the revised manuscript.
Page 4, line 4: Remove information after gauging station
Done.
Page 4, line 7: What are the relative characteristics of the 222 catchments? Size? Is there a map?
The location of the catchments and their size can be retrieved from Figure 4 in the manuscript. For a better insight we relate basin specific characteristics to the model performance (see Figure 1 below). This figure is included and discussed in the revised manuscript. Furthermore, a table containing all the relevant information (location, mean elevation, mean slope, mean precipitation, etc.) will be published as supplement to this manuscript.
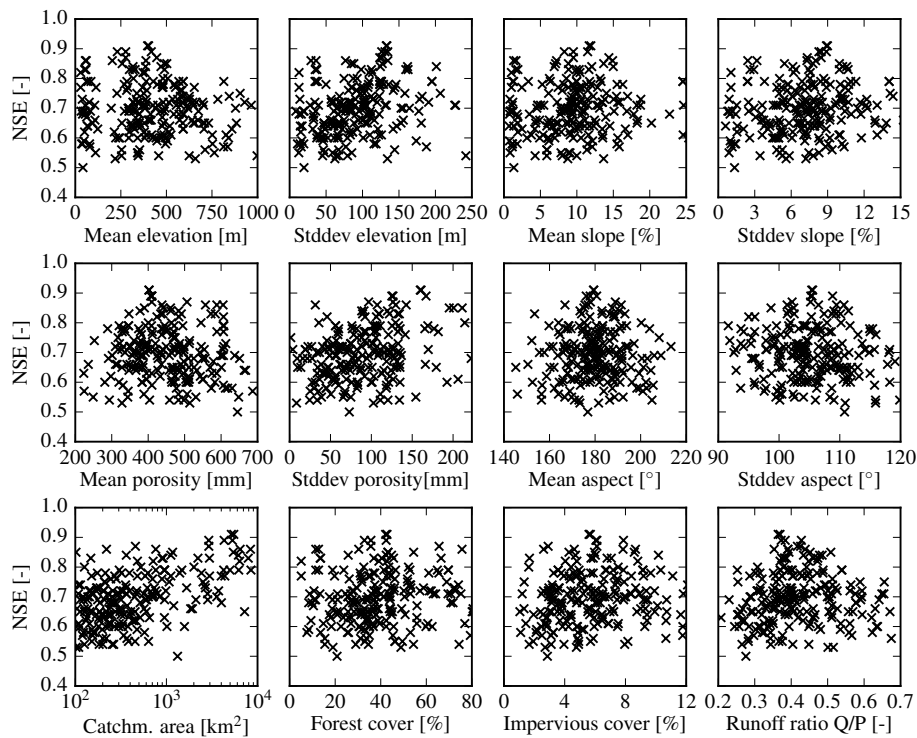

Figure 1: Relation between land surface and hydro-climatic conditions and model performance for the 222 river basins. The mean and standard deviation (stddev) of a characteristic for the single basins are based on the morphological input data at the 100×100 m^2 resolution. standard deviation.

Page 4, line 18: Authors state that this spatial resolution is appropriate without additional reasoning or citation. Please provide one or both.
We orientate the choice for a spatial resolution to the density of the precipitation stations. Two main arguments were considered: 1) The spatial resolution should be lower than the mean lowest distances between existing stations, 2) The chosen resolution shouldn't be so low that the interpolated meteorological variable is

mainly an artifact of the interpolation method (e.g., elevation driven external drift). Therefore, we argument that half of the minimum distances, i.e., 3 km, is a reasonable choice. Because of model specific reasons, we decided for the closest even number to 3 km, i.e., 4 km. We revised the manuscript to make this connection between spatial model resolution and average minimum distance between precipitation stations more clear.

Page 5, line 3: Change precipitation to rainfall.

Done.

Page 5, line 6: On average, it is 1.8 m deep in Germany. Given the previous sentence this is confusing. What is the it in this sentence?

We clarified this point in the revised manuscript.

Page 5, line 11: since Germany is a part of Europe, various river basins across Europe (including Germany), and the USA. . .

Done.

Page 5, line 14: remove the before porosity.

Done.

Page 5, line 22: This goal is not included in the introductory aim. Page 7, line 8-9: this aim is not included in the introduction.

Thank you. We address both points in revised the manuscript as stated above.

Page 8, line 1-2: mean and standard deviation symbols defined in line 1, just use symbols in following sentence.

Done.

Page 8, line 25: Choose basins or catchments and be consistent. I would recommend Basins, which would require going back and changing this throughout the paper.

We follow your recommendation and changed the manuscript accordingly.

Page 8-9:This section could benefit from more organization, such as the use of a more explicit introductory paragraph. Presentation jumps from analysis of results overall to specific basins, which is difficult to follow.

We rewrote and reorganized this section.

Page 9, line 26: remove is at the end of the sentence.

Done.

Page 10, lines 16-24: You discuss the energy- and water-limited conditions, which could also be added to the figure. It also appears that at this point the data switch from under representation to over representation (within the uncertainty bounds). Please discuss this.

Since Figure 4 has already a high information content, we dare to add further information as, e.g., the separation to energy and water limited areas, by introducing another line or coloring. Nevertheless, we elaborated the caption of the figure to address this point. We discuss the issue of under- and overrepresentation in the text now. Thank you for pointing this out.

Page 11, line 5: Your data groupings get confusing here. In the figure daily and monthly (I think?) data are grouped by color in seasons. In the table you report monthly and daily values. Be explicitly clear about what is being reported here. Maybe have a different symbol for monthly and daily data to differentiate (if they are both in the scatterplots. . . still not clear.).

The scatterplot only shows daily data. We adopted the figure caption accordingly.

Page 11, line 6: "The results of the scatter plot... indicate..." this phrasing is awkward. Consider "The scatterplots shown in Figure 5 indicate..."

Thank you for your suggestion. We reformulated the sentence.

Page 11, lines 17-23: Add "limitations" before Hargreaves-Samani approach to improve clarity of the paragraph.

Done.

Page 11, line 24-page 12 line 2: Does land cover type play a role in the ability to interpolate point to grid data? Are some land cover types likely to be more spatially heterogenous with respect to ET and soil moisture? Could this be incorporated into an uncertainty analysis?

For the above mentioned analysis no spatial variabilities of morphological information are considered at all. The hydrological model is run on the point scale, i.e., a single grid cell ($100{\times}100$m^2), for this analysis. The relation of land cover and modeled evapotranspiration at the resolution of $4{\times}4$ km^2 is shown in Figure 8 and shortly discussed in section 4.5.

Page 12, line 11: These features (Central Uplands, Alps) are not on the map.

We changed the text in way that the location can be identified on the map in Figure 9.

Tables: Capitalize all headers, to be consistent (Table 1 no headers are capitalized, Table 2 some are, some arent.

We capitalized all table headers.

Table 1. Header for Major German Basins

We added the header in the revised manuscript.

Table 2. Describe RMSE, BIAS and in a footnote, and what [-] means in a footnote or caption. Station- name should be two words.

Thank you for pointing out that the abbreviations are not explained at all. We added the descriptions in the captions of both tables.

Figures:

Figure 1: What do colors represent? Please describe in figure caption.

The colors are only to better distinguish the different catchments. We added an explanation to the caption.

Figure 4: Using the same color scale is a little misleading. Include the locations of the eddy covariance stations on the map.

The location of the eddy covariance stations are depicted in Figure 1. We have chosen the same colors for the Budyko plot and the map plot to give the reader the opportunity to see wether model performances are clustering for a particular climatic regime and/or geographical location. For the maps on the right side, we are using the inverse color bar. We intend to ease up the comparison of catchments showing a relatively good performance (green) with relative low NSE ranges (green) and vice versa.

Figure 5: Consider changing symbols so that this figure is readable in black and white.

We changed the illustration of the different seasons to four different marker symbols.

Figure 6: Why are these four stations selected?

First, of all to represent the major mHM land cover classes (forest and mixed) are represented. And second, because they have at least continuous 3 year long time series without big data gaps. Further the four station are spread over the three regions where eddy covariance observations are available. We added this reasoning to the figure caption.

Figure 7: The Central Uplands and Alps (referenced on page 12, line 11) are not explicitly shown on this map.

We changed the text such that the location can be identified on the map in Figure 9.

2 Anonymous Referee #2

The authors provide a description of a publicly available dataset that they have developed for Germany. Their product will be useful for the scientific community. Aside from a few problematic oversights, the paper is generally well-written, with appropriate figures and references. In my opinion the paper will be suitable for publication after a minor revision.

Thank you for your helpful comments which are highly appreciated by us. The manuscript benefited from your suggested analyses and literature. We added the mentioned references to the revised manuscript and discuss them. Further, we address the questions: 1) Could the model performance of the 222 catchments be explained by any land surface or hydro-meteorological conditions?, and 2) How does the model estimate of ET compares with a remotely-sensed product? by additional analyses. In the following, we present the referee's comments as well as our point-by-point response to all of them.

2.1 Major

A major oversight of this paper is the lack of referencing a relevant paper that provides a similar dataset, at least in scope. The dataset of Newman et al. (2015) is also a 100-sample ensemble and needs to be cited here. The similarities and differences of the authors dataset with that of Newman et al. (2015) should be noted.

We discuss the difference in among these datasets in the introduction and the conclusions now. The mentioned references were added.

It is surprising that ET would have less uncertainty than streamflow since the latter is a more direct measurement. The authors only evaluate ET at 7 locations, while discharge is evaluated at over 200. It seems inconsistent to suggest that uncertainty across these two observations could be readily compared. Additional discussion is warranted here, including the scale mismatch between a 4×4 km^2 grid cell and a point observation.

The uncertainty of evapotranspiration and generated runoff is compared on the grid cell level, e.g., Figures 8 and 9. This comparison does not consider any observations. This analysis is based on the ensemble spread of the simulations at the 4×4 km^2 resolution. The model and model parameters are beforehand evaluated at point scale, i.e., 100×100 m^2, with observations at eddy covariance stations and at the 4×4 km^2 with discharge observations among others. For these evaluations we do not compare uncertainties of different variables as they would not be "readily comparable" because of the scale mismatch as you mentioned. So we fully agree with you in the argument that uncertainties from hydrologic variables at different resolutions are not comparable.

Further, the authors should comment more directly on why they did not evaluate the spatial patterns of their model against remotely-sensed ET and consider doing this evaluation.

Thank you for mentioning this point. We added a comparison of the ensemble mean of modeled evapotranspiration with MODIS evapotranspiration. We elaborated the results of this comparison in the revised manuscript and added Figure 2 to the manuscript).
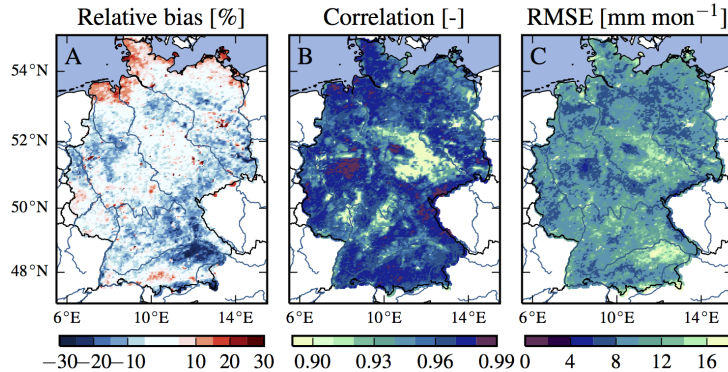

Figure 2: Comparison of monthly estimates of evapotranspiration from mHM and MODIS in the period 2001-2010. The ensemble is represented by the ensemble mean of 100 evapotranspiration estimates. The comparison is based on three statistical assessments: A) relative bias, B) Pearson correlation coefficient, and C) root mean squared error (RMSE). The respective units are given in brackets.

The validation watersheds range in size by nearly two orders of magnitude. If the model spatial resolution is the same for all, the authors should comment and hypothesize whether they see higher model performance in larger basins. Does performance increase monotonically with basin size?

We comment on this issue in the manuscript as "However, a tendency to perform better in large catchments basins is observed." We also note that there is no clear (monotonic) relationship between basin area and NSE as can be observed from the Figure 3 below.

In Figure 4, climatic regime does not appear to be a good predictor of model performance, with some of the highest NSE scores distributed throughout the range of conditions. The authors should comment on what, if anything, will best predict model performance, to guide a potential user of the dataset.

Fortunately, we did not find any meteorological or morphological characteristics which explained why model performance is different for different catchments. This makes us confident that the retrieved parameter sets are representative for various climatic and physiographic conditions. We performed an analysis for identifying relations between land surface and hydro-climatic characteristics and model performance. Figure 4 was added in the revised manuscript and the findings are shortly discussed in section 4.2 of the revised manuscript.
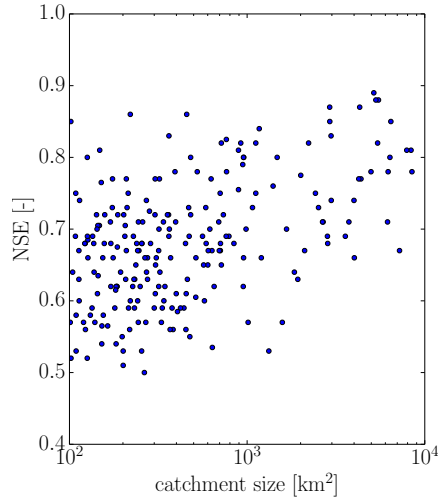
Figure 3: Relation of model performance and catchment area for the 222 basins.
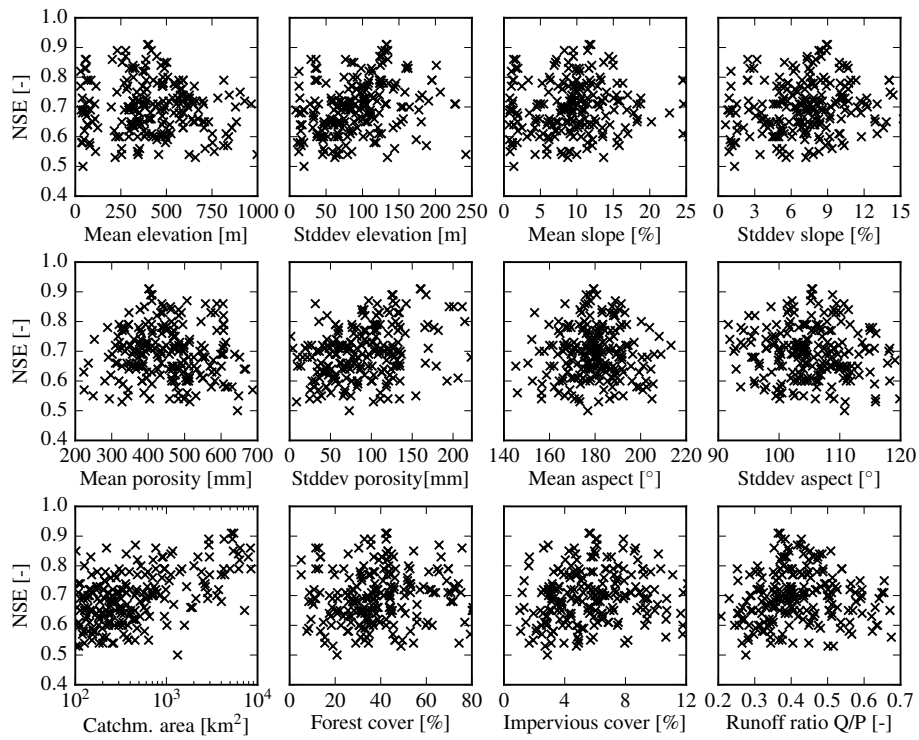

Figure 4: Relation between land surface and hydro-climatic conditions and model performance for the 222 river basins. The mean and standard deviation (stddev) of a characteristic for the single basins are based on the morphological input data at the $100{\times}100$ m^2 resolution. standard deviation.

10

2.2 Minor

P1L24: Grammar: have a footprints

Changed.

P1L24: 827 stations worldwide perhaps more apt to say "less than 1,000 locations worldwid", since there are other observational sources beyond fluxnet.

Changed.

P2L1: replace "reanalysis data" with "reanalysis products" and make this change elsewhere

Done.

P2L9: Maurer et al (2002) and Livneh et al. (2015) also cover a significant area in Canada (i.e. not just US, MX, and China).

Thanks for pointing out this fact. We changed the text "' accordingly.

P9: Here and elsewhere the use of the plural form of the word "performance" as "performances" is grammatically incorrect. Please correct this.

Changed.

References:

Newman, A. J., M. P. Clark, J. Craig, B. Nijssen, A. W. Wood, E. D. Gutmann, N. Mizukami, L. Brekke, and J. R. Arnold (2015). An observationally based gridded ensemble of precipitation and temperature data for the contiguous USA. J. Hydrometeorology, doi:10.1175/JHM-D-15-0026.1.

Livneh, B., Bohn, T. J., Pierce, D. W., Munoz-Arriola, F., Nijssen, B., Vose, R., & Brekke, L. (2015). A spatially comprehensive, hydrometeorological data set for Mexico, the US, and Southern Canada 19502013. Scientific data, 2.

3   Anonymous Referee #3

In the submitted manuscript "A High-Resolution Dataset of Water Fluxes and States for Germany Accounting for Parametric Uncertainty" Zink et al. present a new approach to calibrate a distributed model (the mesoscale Hydrological Mode mHM) across all basins across Germany on a 4×4 km^2 resolution. They use a 2 step calibration procedure, during which they firstly calibrate 7 major basins individually, and, secondly use a subset of the calibrated parameter samples with sufficient performance (NSE$\geq$0.65) at all 7 basins to apply them over the remaining catchments over Germany. Using split-sample tests and auxiliary information (AET, soil moisture, recharge) they evaluate the model and the combined parameter set concerning its general performance and uncertainty. Overall, the approach is well chosen and the provided results make sense.
However, the manuscript needs serious improvements before it can be considered for publication in HESS. Most of the points of criticism are related to the need for more rigorousness.

We would like to thank the reviewer for his/her valuable comments. We highly appreciate them. We think the manuscript improved significantly by addressing these comments. Based on the comments of the reviewer we revised the introduction and the results and discussion section. We discuss additional references and strengthened the rigorousness of the manuscript and the scientific analyses therein. In the following, we present the referee's comments as well as our point-by-point response to all of them.

- The introduction is too short and does not provide a proper view on the research gaps of the approaches and methods applied in this study (for instance calibration and model evolution approaches). It appears to be series of vaguely related short paragraphs - a more robust story line is needed.
  We added a paragraph discussing the calibration of hydrologic models for large spatial domains to the introduction.

- The methods are incomplete, partially referring to previous research, partially omitting parts of the analysis that later appears in the results section. On the other hand some information is irrelevant. Very important information, for instance introducing the model parameters that are calibrated, is completely missing. Up to the end of the manuscript it is not clear, which parameters were calibrated, which ranges were used and there was no discussion of their physical meaning.
  For giving deeper insight to the model parameterization we rewrote the model description part (section 3.1) which made it hopefully better understandable. Further, we added tables of the effective model parameters in the revised manuscript. A deeper insight to the model and model parameterization is however out of the scope of this study. We refer to Samaniego et al. 2010 and Kumar et al. 2103 (also mentioned in the manuscript) for further details.
  Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. Water Resources Research, 46(5), W05523.
  Kumar, R., Samaniego, L., & Attinger, S. (2013). Implications of dis-

tributed hydrologic model parameterization on water fluxes at multiple scales and locations. Water Resources Research, 49(1), 360379.

- There is generally too little referencing of other studies. In particular in the Results and Discussion section, there are some interpretations that are hardly supported by the results and almost no comparison to the research of others.
We addressed the need for more references in the discussion of the results by adding comparisons to similar studies where appropriate. Further, we relaxed or deleted some of the interpretations which are not supported by the data.

- In general there is a lack of self-criticism. There are many obvious and hidden assumptions in the approach and the authors should spend significantly more effort discussing them.
We added discussions of limitations of chosen approaches and assumptions at places in the manuscript which could be identified by us, and which were pointed out by you (see point-by-point answers).

For all these reasons, which are elaborated in more detail in the commented pdf, I recommend major revisions. I am convinced that the approach and the results are novel and reasonable but the authors have to show this in a rigorous scientific way.

Introduction: more structure needed, storyline incomplete too general, mixed up with results
We reorganized and rewrote major parts of the introduction.
P1L8: please explain acronym
Done.
P2L1: reanalysis data: What type of data?
We added some examples of potential reanalysis data.
P2L5: observational data: Please be more specific on Scale and type of data
Thanks for the comment. We specified what we mean with observational data in the revised manuscript.
P2L17-21: You mention observational uncertainty and then you decide to only consider parameter uncertainty. Please establish link between these different types of uncertainty.
In this paragraph of the introduction we gave an overview on all possible sources of predictive uncertainty in hydrologic modeling but surely we can not pursue all aspects of uncertainty within a single paper. Therefore, we aim to analyze other sources of uncertainty in separate studies, e.g., Baroni et al. 2016. A discussion about the links between the different sources of uncertainty is added to the revised manuscript.
Baroni, G., Zink, M., Kumar, R., Samaniego, L., and Attinger, S.: On the effect of the uncertainty in soil properties on the simulated hydrological state and fluxes at different spatio-temporal scales, Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-657, in review, 2016.
P2L27-28: This is already results - don't mention here
We deleted the respective sentence.

P3L26: hydrogeological vector map: please clarify: is this a hydro geological map?

Yes it is a hydrogeological map. We have provided a reference to this map for further details. We also revised wording to "hydrogeological map".

P4L12: On average? What does this mean when referring to the number of stations?

The number of meteorological stations is varying over time. New stations are established while others are disassembled. We provide the average number of stations of the modeled time period of 1951-2010.

P4L19: Why are you not using a more physically based method like Penman Monteith?

We use the well established Hargreaves-Samani approach in this study because it has the best support with observational data. As mentioned in the paper we use about 570 climate stations over Germany for providing input to the Hargreaves-Samani method. In contrast radiation observations are sparsly conducted within Germany. Right now approximately 80 global radiation measurement stations exist in Germany and still longwave radiation information are missing. Therefore, we can not estimate PET based on the Penman-Monteith approach. Moreover, several studies showed that PET estimates of regionalized Hargreaves-Samani approaches are close to those of Penman-Montheith estimates. Herein we are using a regionalized Hargreaves-Samani approach which is based on the aspect of the respective grid cell.

further reading:

Almorox, J., Quej, V. H., & Mart, P. (2015). Global performance ranking of temperature-based approaches for evapotranspiration estimation considering Köppen climate classes. Journal of Hydrology, 528, 514522. doi:2015.06.057

Droogers, P., & Allen, R. G. (2002). Estimating reference evapotranspiration under inaccurate data conditions. Irrigation and Drainage Systems, 16(1), 3345. doi:10.1023/A:1015508322413

Temesgen, B., Eching, S., Davidoff, B., & Frame, K. (2005). Comparison of Some Reference Evapotranspiration Equations for California. Journal of Irrigation and Drainage Engineering, 131(1), 7384. doi:(ASCE)0733-9437(2005)131:1(73)

P4L23: REGNIE: Why didn't you use this data as direct input for the model?

First, The German Meteorological Service was working at the development of the REGNIE data set in parallel to us. So after we finished the establishment of our interpolation routines in 2011 the REGNIE product was released. Second, we intended to use daily updated station data from the German Meteorological Service for running hydrological simulations on an operational basis. We could realize this aim in 2014 (www.ufz.de/droughtmonitor). And third, we publish our precipitation data set herein to address the need of investigating and analyzing input data uncertainties. Since both interpolation approaches are based on different methodologies we consider the publication of an alternative gridded precipitation product as added value for future research activities.

The mesoscale Hydrologic Model mHM: parameter estimation not clear

We now elaborate more on the estimation of parameters within mHM in the revised manuscript. A detailed description of the Multiscale Parameter Regionalization technique is out of the scope of this study since it was already published in Samaniego et al. 2010 and Kumar et al. 2013. We refer to those papers for getting deeper insight to the parameterization of mHM.

P5L19-20 Is the sub grid variability also up scaled by distribution functions or is it finally One effective value derived by sub grid information?

The effective parameter is an effective value which was derived by sub grid information. We clarified this in the manuscript.

P5L22: How many calibration parameters do you have?

We use 67 *global* or *transfer* parameters which were calibrated. We mention this fact in the revised manuscript. We add an overview of these parameters and their ranges to the supplementary material.

P5L25: It is not clear how the different parameter sets derived from the 7 basins are put together to be used at the remaining basins.

We transfer the *global* parameters which were inferred by calibration from one catchment to another (receiver) basins. mHM allows for this flexibility because the *global* parameters are time-invariant and location-independent. These parameters are then used for the hydrologic simulation in each of the receiver catchments.

P5L26: Mention studies that used similar approaches for parameter estimation and model evaluation such as

Choi, H. T. and Beven, K.: Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of TOPMODEL within the GLUE framework, J. Hydrol., 332(34), 316336, doi:10.1016/j.jhydrol.2006.07.012, 2007.

Hartmann, A., Gleeson, T., Rosolem, R., Pianosi, F., Wada, Y. and Wagener, T.: A large-scale simulation model to assess karstic groundwater recharge over Europe and the Mediterranean, Geosci. Model Dev., 8(6), 17291746, doi:10.5194/gmd-8-1729-2015, 2015.

and there are surely more if you take a closer look

We discuss the below mentioned approaches for model evaluation and parameter estimation in the introduction of the revised manuscript.

Hartmann, A., Gleeson, T., Rosolem, R., Pianosi, F., Wada, Y. and Wagener, T.: A large-scale simulation model to assess karstic groundwater recharge over Europe and the Mediterranean, Geosci. Model Dev., 8(6), 17291746, doi:10.5194/gmd-8-1729-2015, 2015.

Perrin, C., Andrassian, V., Rojas Serna, C., Mathevet, T., & Le Moine, N. (2008). Discrete parameterization of hydrological models: Evaluating the use of parameter sets libraries over 900 catchments. Water Resources Research, 44(8), W08447.

Henriksen, H. J., Troldborg, L., Nyegaard, P., Sonnenborg, T. O., Refsgaard, J. C., & Madsen, B. (2003). Methodology for construction, calibration and validation of a national hydrological model for Denmark. Journal of Hydrology, 280(14), 5271.

Troy, T. J., Wood, E. F., & Sheffield, J. (2008). An efficient calibration method for continental-scale land surface modeling. Water Resources Research, 44(9), W09411.

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Mocko, D. (2012). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. Journal of Geophysical Research: Atmospheres, 117(D3), D03109.

Hostetler, S. W., & Alder, J. R. (2016). Implementation and evaluation of a monthly water balance model over the US on an 800 m grid. Water

Resources Research, 52, 120.

Mitchell, K. E. (2004). The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. Journal of Geophysical Research, 109(D7), D07S90.

Zhang, X.-J., Tang, Q., Pan, M., & Tang, Y. (2014). A Long-Term Land Surface Hydrologic Fluxes and States Dataset for China. Journal of Hydrometeorology, 15(5), 20672084.

Zhu, C., & Lettenmaier, D. P. (2007). Long-Term Climate and Derived Surface Hydrology and Energy Flux Data for Mexico: 19252004. Journal of Climate, 20(9), 19361946.

P6L13: Is this number large enough to find the best parameter sets?

As the results in Figure 1 shows this number iterations is sufficient to obtain reasonable performances. We have to admit the dynamically dimensioned search algorithm will not find optimal parameter values. This algorithm is design to find sufficient objective function values in a reasonable amount of time. Consequently another algorithm, e.g., the Shuffled Complex Evolution algorithm, needs to be applied for identifying the optimum of the objective function. For the herein proposed purpose the choice for DDS is reasonable because the aim is to identify reasonable parameter sets, rather than the best ones, for a set of 7 big catchments in a reasonable amount of time. The results of the model calibration are shown in Figure 2 as white boxes in the upper left corner. With exception of the Saale river basin all catchments reveal sufficient discharge estimations (median NSE$\geq$0.85, overall mean NSE=0.89).

P7L1: What justifies this? Other studies discarded time periods during the energy balance which could not be closed (e.g. Miralles, D. G., De Jeu, R. A. M., Gash, J. H., Holmes, T. R. H. and Dolman, A. J.: Magnitude and variability of land evaporation and its components at the global scale, Hydrol. Earth Syst. Sci., 15(3), 967981, doi:10.5194/hess-15-967-2011, 2011.")

The energy balance is not closed on the majority of the eddy flux towers worldwide due to a variety of reasons (e.g., Stoy et al. 2007, Foken 2008, Leuning 2012). There is an extensive literature on how to correct the observed fluxes (e.g., Twine et al. 2000, Wilson et al. 2002, Boldocchi 2003, Stoy et al. 2007, Allen 2008, Hendricks Franssen et al. 2010, Mauder et al. 2010, 2013, Foken et al. 2011, Kessomkiat et al. 2013, Charuchittipan et al. 2014, Ingwersen et al. 2015) ranging from correcting mostly latent heat to correcting mostly sensible heat. Two prominent arguments, which show immediately why latent heat should be corrected as well, are 1. meso-scale circulations that remove energy horizontally, i.e., in a movement perpendicular to the tower observations (e.g., Stoy et al. 2007) and 2. dampening of the water vapour signal in the tubing of the so-called closed path analysers and hence loss of high-frequency contributions especially for latent heat (e.g., Leuning 2012). We use a conservative correction, which is similar to preserving the observed Bowen ratio.

further reading:

Allen, R. G. (2008), Quality assessment of weather data and micrometeological flux - Impacts on evapotranspiration calculation, Journal of Agricultural Meteorology, 64(4), 191204.

Baldocchi, D. D. (2003), Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future, Global Change Biology, 9(4), 479492.

Charuchittipan, D., W. Babel, M. Mauder, J.-P. Leps, and T. Foken (2014), Extension of the averaging time in Eddy-covariance measurements and its effect on the energy balance closure, Boundary-Layer Meteorology, 152(3), 303327, doi:10.1007/s10546-014-9922-6.

Foken, T., M. Aubinet, J. J. Finnigan, M. Y. Leclerc, M. Mauder, and K. T. Paw U (2011), Results of a panel discussion about the energy balance closure correction for trace gases, Bulletin of the American Meteorological Society, 92(4), ES13ES18, doi:10.1175/2011BAMS3130.1.

Hendricks Franssen, H. J., R. Stckli, I. Lehner, E. Rotenberg, and S. I. Seneviratne (2010), Energy balance closure of eddy-covariance data: A multisite analysis for European FLUXNET stations, Agricultural and Forest Meteorology, 150(12), 15531567, doi:10.1016/j.agrformet.2010.08.005.

Ingwersen, J., K. Imukova, P. Hgy, and T. Streck (2015), On the use of the post-closure methods uncertainty band to evaluate the performance of land surface models against eddy covariance flux data, Biogeosciences, 12, 23112326, doi:10.5194/bg-12-2311-2015.

Kessomkiat, W., H.-J. H. Franssen, A. Graf, and H. Vereecken (2013), Estimating random errors of eddy covariance data: An extended two-tower approach, Agricultural and Forest Meteorology, 171-172, 203219, doi: 10.1016/j.agrformet.2012.11.019.

Mauder, M., R. L. Desjardins, E. Pattey, and D. Worth (2010), An attempt to close the daytime surface energy balance using spatially-averaged flux measurements, Boundary-Layer Meteorology, 136(2), 175191, doi:10.1007/s10546-010-9497-9.

Mauder, M., M. Cuntz, C. Dre, A. Graf, C. Rebmann, H.-P. Schmid, M. Schmidt, and R. Steinbrecher (2013), A strategy for quality and uncertainty assessment of long-term eddy-covariance measurements, Agricultural and Forest Meteorology, 169, 122135, doi:10.1016/j.agrformet.2012.09.006.

Stoy, P. C., S. Palmroth, A. C. Oishi, M. B. S. Siqueira, J.-Y. Juang, K. A. Novick, E. J. Ward, G. G. Katul, and R. Oren (2007), Are ecosystem carbon inputs and outputs coupled at short time scales? A case study from adjacent pine and hardwood forests using impulse-response analysis, Plant, Cell & Environment, 30(6), 700710, doi:10.1111/j.1365-3040.2007.01655.x.

Twine, T. E., W. P. Kustas, J. M. Norman, D. R. Cook, P. R. Houser, T. P. Meyers, J. H. Prueger, P. J. Starks, and M. L. Wesely (2000), Correcting eddy-covariance flux underestimates over a grassland, Agricultural and Forest Meteorology, 103(3), 279300.

Wilson, K. et al. (2002), Energy balance closure at FLUXNET sites, Agricultural and Forest Meteorology, 113(1-4), 223243.

P8L20: These are no results - delete or move to methods section

Thank you for pointing this out. We moved the sentence to the methods section.

P8L26: Fig 2 and text do not fit well together..,

We will rewrite the corresponding section in the revised manuscript.

P8l29: How is it possible that some of the grey whiskers fall below 0.65?

The parameter selection procedure is applied to model performances on daily basis in the validation period (see section 3.2). Thus, the grey boxes in the upper right panel of Figure 2 show the resulting performances after parameter selection. All of the whiskers are exceeding an NSE of 0.65.

P9L1: There are large drops for Mulde, Neckar and Danube - how can the average drop b. only 6%?

For clarification we provide the numbers below. As can be seen the average drop is 5.64%.

medians	Mulde	Ems	Neckar	Saale	Main	Weser	Danube	average
NSE on-site	0.80	0.82	0.90	0.69	0.92	0.91	0.84	
NSE ensemble	0.69	0.78	0.79	0.72	0.86	0.91	0.83	
difference	-0.12	-0.04	-0.10	0.03	-0.06	0.00	-0.01	-0.04
normalization *100 [%]	-16.76	-5.44	-13.03	4.44	-7.14	0.00	-1.57	-5.64

P9L3: What about Danube and Main? For those two the ranges change significantly

We revised the text to reflect your comment.

P9L9: Compensate for errors in the model structure: Provide some references to such cases.

We added a reference which is analyzing this problem (Clark and Vrugt 2006).

Clark, M. P., & Vrugt, J. a. (2006). Unraveling uncertainties in hydrologic model calibration: Addressing the problem of compensatory parameters. Geophysical Research Letters, 33(6), L06406. doi:10.1029/2005GL025604

P10L16: Should be mentioned in discussion

The manuscript does not have a separate discussion section. Therefore we included those discussion in the "Results and Discussion" section.

P10L24: 0.1: of what? NSE?

Yes. We revised the manuscript.

P11L17-20: So why not using a more physical approach?

Because an approach like Penman-Monteith (PM) is based on observations which are usually sparsely available as we elaborated beforehand. Thus, estimating evapotranspiration based on PM would imply to apply reanalysis products which introduce another degree of uncertainties because these data are partly relying on model estimations. The intention of this study was to use observational forcing data.

P12L27-P13L3: Doesn't this rather belong to the study site description?

These analyses are based on the gridded precipitation and potential evapotranspiration product which were develop in this study. For this reason we think these analyses are appropriately placed in the result part.

P13L6-9: Can you quantify the strength of the relation between AET/Q uncertainty and its explanatory variables?

Thank you for this comment. We quantified the strength of the connections between uncertainty patterns of the evapotranspiration and generated runoff with porosity and dryness index using the Spearman rank correlation (see table below). We elaborated on these results in the revised manuscript.

Spearman rank corr.	Porosity	Dryness index
Evapotranspiration uncertainty	0.58	0.28
Generated runoff uncertainty	0.32	0.92

P13L9-10: You cannot state this without a proper sensitivity analysis

We reformulated this sentence.

P13L12-13: see my previous comment and provide numbers

Done.

P13L15-17: See previous comments. Right now, the data does not support such strong statements

We removed this sentence.

P13L30-31: Why? Please clarify.

We elaborated on that in the revised manuscript.

P14L9-14: Without mentioning or explaining the model parameters and visualizing that relationship between the snow and the soil parameters this statement is not supported be the analysis.

We revised the manuscript accordingly and deleted statements which are not supported by the data and parameters.

P14L25-26: You cannot state this without discussing actual values of the parameters. An acceptable NSE does not mean that the related model parameters are sensitive.

We kindly disagree with the interpretation of the reviewer. We are not aiming identifying parameter sensitivities in this study. Our aim is to find reasonable parameter sets on the basis of observed discharge. As we demonstrate within this study the chosen method can yield reasonably good model performance for discharge evaluation and is able to capture the spatio-temporal variability of *ET* data.

Figure 2: Shouldn't this be filled white?

These boxes are filled white. The impression that they are grey may arise because of the narrow boxes. We assume that potential readers of the plot will assess the rationale behind the plot and interpret it in the right way as you did.

Figure 4: panel D?

Thank you. Changed.

Figure 6: observations hardly visible - please improve

We improved the plot.

Figure 9: The information on the range of uncertainties is provided by the grey area enveloping the median. I don't think the normalized range adds significantly more information to that - delete?

We argue that the normalization of the ranges is needed for the comparison of uncertainties among the hydrologic variables. Also interpreting the uncertainty behavior through the course of a year is more difficult without proper normalization. The uncertainty in evapotranspiration, for example, does not significantly change over the course of a year. Such a behavior would be difficult to observe without the normalized ranges. For that reason we prefer to stay with the figure as it is.

4 The marked-up manuscript version

A High-Resolution Dataset of Water Fluxes and States for Germany Accounting for Parametric Uncertainty

Matthias Zink[1], Rohini Kumar[1], Matthias Cuntz[1,2], and Luis Samaniego[1]

[1]Helmholtz Centre for Environmental Research - UFZ, Department Computational Hydrosystems, Permoserstrasse 15, 04318 Leipzig, Germany

[2]INRA, Université de Lorraine, UMR1137 Ecologie et Ecophysiologie Forestières, Champenoux, France

Correspondence to: Matthias Zink (matthias.zink@ufz.de)

Abstract. Long term, high-resolution data about hydrologic fluxes and states are needed for many hydrological applications. Because continuous large-scale observations of such variables are not feasible, hydrologic or land surface models are applied to derive them. This study aims to analyze and provide a consistent high-resolution dataset of land surface variables over Germany, accounting for uncertainties caused by equifinal model parameters. The mesoscale Hydrological Model (mHM) is employed

5   to derive an ensemble (100 members) of evapotranspiration, groundwater recharge, soil moisture and ~~generated runoff~~ runoff generated at high spatial and temporal resolutions (4 km and daily, respectively) for the period 1951-2010. The model is cross-evaluated against the observed ~~runoff~~ daily streamflow in 222 ~~catchments~~basins, which are not used for model calibration. The mean (standard deviation) of the ensemble median ~~NSE~~ Nash-Sutcliffe Efficiency estimated for these ~~catchments~~ basins is 0.68 (0.09) for daily ~~discharge~~ streamflow simulations. The modeled evapotranspiration and soil moisture reasonably represent the

10   observations from eddy covariance stations. Our analysis indicates the lowest parametric uncertainty for evapotranspiration, and the largest is observed for groundwater recharge. The uncertainty of the hydrologic variables varies over the course of a year, with the exception of evapotranspiration, which remains almost constant. This study emphasizes the role of accounting for the parametric uncertainty in model-derived hydrological datasets.

1 Introduction

Consistent, long-term data of meteorological and hydrological variables at a high spatial resolution are needed for many applications, including i) impact assessment studies, such as for drought, flood or climate change analysis (Sheffield and Wood, 2007; Huang et al., 2010; Samaniego et al., 2013; Kumar et al., 2016; Zink et al., 2016), and ii) studies that need spatially and temporally continuous, observation-based datasets, e.g., for downscaling or disaggregating climate model outputs (Wood et al., 2004; Thober et al., 2014) or for establishing Ensemble Streamflow Prediction (Day, 1985) and reverse Ensemble Streamflow Prediction ~~(Wood and Lettenmaier, 2008) approaches~~ approaches (Wood and Lettenmaier, 2008) .

Continuous observations of hydrologic fluxes and states are economically and logistically not feasible on regional to national scales (Vereecken et al., 2008). In-situ soil moisture observations, for example, are scarcely available. These point-scale observations are ~~usually only~~ representative for a small control volume of a few cm^3. Evapotranspiration measurements at eddy covariance stations have ~~a~~ footprints of tens to hundreds of meters but they are available at ~~only 827~~ less than 1000 stations worldwide (~~http://fluxnet.ornl.gov, April 2016~~FLUXNET (2007)).

Alternatives include remote sensing or reanalysis ~~data~~products such as NCEP-CFSR (Saha et al., 2010) or ERA-INTERIM (Dee et al., 2011) . Hydrologic products derived from remote sensing are broadly available, but they do not consider the conservation of mass, i.e., the closure of the water balance. Moreover, these products are not spatially and temporally continuous due to reliance on cloud-free conditions (Mu et al., 2007; Liu et al., 2012). Reanalysis ~~data, alternatively,~~products, in contrast, ~~provide continuous data but they~~ have coarse spatial resolutions of at most 1/4° (Dee et al., 2016), which is not suitable for regional scale applications.

Hydrologic models driven by ~~observational data~~ground-based meteorological observations are the prime alternative to derive spatially and temporally consistent water fluxes and states ~~for~~at large spatial domains.

For example, Maurer et al. (2002); Zhu and Lettenmaier (2007); Livneh et al. (2013); and Zhang et al. (2014) provided model-based datasets on a national scale. These data are based on the Variable Infiltration Capacity (VIC) model (Liang et al., 1994) and have, at most, a spatial resolution of 1/16° and cover the contiguous United States, Mexico~~and China.~~, China, and parts of Canada. Livneh et al. (2015); Newman et al. (2015a) ; and Newman et al. (2015b) provide data on the same domain with a focus on meteorological data. A set of four models was used in the NLDAS project to assess the water balance components over the contiguous United States (Mitchell, 2004; Xia et al., 2012b, a) . Studies by Nijssen et al. (2001); Fan and van den Dool (2004); Berg et al. (2005); and Sheffield et al. (2006) focus on the global domain. The spatial resolution of these global data sets is at most 1/2°, and many of these studies focus on meteorological forcings rather than hydrologic variables.

The resolution of the above mentioned model-derived datasets are coarse according to Wood et al. (2011), who stated ~~that a need exist to have higher spatially resolved~~ a need for higher-resolution data and models for purposes like flood and drought forecasting. Moreover, Bierkens et al. (2015) ~~states~~state that water resources or river basin managers will favor highly resolved data at resolutions of 1-5 km.

The application of observational derived model products, however, also has some limitations. First, due to a limited amount of observed variables modeling approaches, like the estimation of potential evapotranspiration (PET), have to be adopted to

the available data. In consequence temperature based PET methods may be preferred to more physically based approaches (e.g., radiation based). Second, the interpolation of point observations induces uncertainties depending on the applied interpolation method. Further, small-scale, convective precipitation events may not be caught by gauging networks and lead to an underestimation in precipitation.

5   Furthermore, hydrological models are subject to different sources of uncertainty, i.e., input, model structural and parametric uncertainty (Beven, 1993). ~~These uncertainties~~ All of the afore-mentioned uncertainties propagate to the model results and can superpose each other (Zappa et al., 2011) . The overall uncertainty of hydrological models is therefore summarized as predictive uncertainty. Uncertainties are often not considered when deriving hydrological or hydro-meteorological datasets (e.g., Huang et al., 2010; Livneh et al., 2013; Zhang et al., 2014). In consequence, predictive uncertainties are often not

10 addressed but may have substantial implications on subsequent studies, as shown by Samaniego et al. (2013). Herein, we will focus on the predictive uncertainties caused by equifinal parameter sets.

    The specification of model parameters which are valid beyond catchment boundaries poses another challenge in the application of hydrologic models over large domains. Large scale hydrologic model studies apply either parameters originating from a single catchment (Henriksen et al., 2003) , filter behavioral parameters from predefined sets (Perrin et al., 2008; Hartmann et al., 2015) ,

15 extrapolate or regionalize parameters or hydrological variables from observed to unknown locations (Zhu and Lettenmaier, 2007; Troy et al use an uncalibrated model (Mitchell, 2004; Hostetler and Alder, 2016) . A methodology considering the calibration in individual basins for creating a set of regionalized parameters which are later on filtered for behavioral solutions in all considered basins could be an alternative approach. Such an approach combines all of the afore mentioned strategies.

    The aim of this study is to derive a model based, consistent set of national-scale hydrological data for Germany ~~at high~~

20 ~~spatial and temporal resolutions. The mesoscale Hydrologic model (mHM; Samaniego et al. (2010); Kumar et al. (2013b))~~ ~~is implemented over Germany to provide gridded fields of evapotranspiration, soil moisture, groundwater recharge, and~~ ~~grid-cell-generated runoff on a daily basis for~~ within the period 1951-2010. We address the ~~issue of predictive uncertainties~~ ~~in deriving these datasets by considering an ensemble of equifinal parameter sets. The 100 ensemble members of the model~~ ~~parameters are selected based on a compromise solution and show satisfactory performances in seven major German river~~

25 ~~basins for daily streamflow simulations. We cross-evaluated these parameter sets in 222 other catchments that have not been~~ ~~used for parameter inference. Additionally, the model simulations are evaluated by using evapotranspiration and soil moisture~~ ~~observations from seven eddy covariance stations.~~

    ~~We also provide the daily forcing datasets, including precipitation, temperature, and potential evapotranspiration. We address~~ ~~the~~ need for highly resolved data by conducting observation-driven hydrological simulations at a spatial resolution of 4×4 km^2

30 (1/25°). Daily fields of evapotranspiration, soil moisture, groundwater recharge, and grid-cell-generated runoff as well as precipitation, temperatures, and potential evapotranspiration are made freely available. To our knowledge, such a consistent and long-term dataset for Germany has not been freely available until now. The ~~internal consistency among variables is ensured~~ ~~by applying the observation-driven hydrological model (mHM) over the entire domain of Germany. The set of ensemble~~ ~~simulations allowed us to investigate~~ dataset accounts for predictive uncertainties by considering a set of equifinal parameters.

35 An parameter estimation approach for deriving a set of 100 parameters on the national scale is developed. We further aim to as-

sess and evaluate the spatio-temporal ~~propagation of uncertainties between different model compartments. Further, we provide an analysis of potential causes for the distinct spatial distribution of uncertainties~~ distribution of the simulated hydrological states and fluxes as well as their uncertainties using multiple validation variables at different scales. Finally, the parametric uncertainties are analyzed regarding their explanatory variables for the simulated fluxes and their propagation between different

5   model compartments.

2   Study Domain and Datasets

The study is conducted on the territory of Germany, which covers an area of approximately 357,000 km^2 (Figure 1). The region, located in Central Europe, is mainly characterized by a humid climate but nonetheless has north-to-south and east-to-west climatic gradients. The topography varies from low-altitude, flat areas in the north (North German Plain) over mid-altitude

10   mountains in Central Germany (Central Uplands) to the high altitude Alpine Foothills and the Alps in the south. Whereas the northwestern part of Germany is still under maritime influence, the eastern part has a more continental climate that is characterized by colder winters and less precipitation.

The assessment of water fluxes and states is restricted to the national borders of Germany because meteorological data and land surface characteristics are available in this domain. Thus, only ~~catchments~~ basins entirely covered by German territory

15   are used to derive parameters for the hydrological model. These seven major ~~catchments~~ basins are depicted in Figure 1. These basins represent the topographic and hydro-climatic gradient within Germany (see Table 1). They range in size from 6,000 km^2 to 48,000 km^2 and are characterized by mean elevations ranging from 60 m.a.s.l. (Ems ~~catchment~~basin) to 560 m.a.s.l. (Danube ~~catchment~~basin). All ~~catchments~~ basins have a comparable degree of urbanization ranging between 6% and 10%. A remarkably low amount of forest is observed in the Ems ~~catchment~~basin, where agriculture and pasture are the dominant land use.

20   Due to different climatic regimes the ~~averages discharge~~ average streamflow of the seven ~~catchments range~~ basins ranges from 161 mm a^{-1} to 469 mm a^{-1}. The low-lying Ems reaches a remarkably high discharge due to maritime influence, whereas the Saale river is characterized by the lowest ~~discharge~~streamflow. The runoff coefficient of the Saale differs significantly from the other ~~catchments~~basins, which originates from the high degree of anthropogenic influence within this basin. Three of the ten largest dams in Germany are located there (Bleiloch - 215 Mio. m^3, the Hohenwarte - 182 Mio. m^3 and the Rappbode

25   reservoir - 109 Mio. m^3). Furthermore, open pit mining has a large influence on the water budget of this ~~catchment~~basin.

2.1   Land Surface Properties

The land surface characteristics required by the hydrologic model include a 50 m digital elevation model (DEM) acquired from the Federal Agency for Cartography and Geodesy (Federal Agency for Cartography and Geodesy (BKG), 2010), a digitized soil map at a scale of 1:1,000,000 (Federal Institute for Geosciences and Natural Resources (BGR), 1998), and a hydrogeological

30   ~~vector~~ map at a scale of 1:200,000 (Federal Institute for Geosciences and Natural Resources (BGR), 2009). The soil map contains information on soil textural properties, such as the sand and clay contents of different soil horizons. The soils are classified into 72 soil types and have an average depth of 1.8 m. The hydrogeological map comprises 23 classes and gives

4

information about saturated hydraulic conductivities and karstic areas. Based on the DEM, additional information, such as the slope, aspect, flow direction and flow accumulation, are inferred. Land cover information is derived from CORINE land cover scenes of the years 1990, 2000, and 2006 (European Environmental Agency (EEA), 2009). The period prior to 1990 is assumed to be static and is represented by the scene of 1990. All data sets are remapped to a common spatial resolution of 100×100 m^2

5   using a nearest neighbor approach.

The location and shape of the major ~~catchments~~ basins (Figure 1) are derived via an automated delineation, which is based on gauging station ~~information~~ and terrain information (flow accumulation and flow direction). ~~Discharge~~ Streamflow data are provided by the European Water Archive (EWA) (2011) and the Global Runoff Data Centre (GRDC) (2011). The results of the delineation are approved via comparison with the CCM River and Catchment Database (European Commission - Joint

10 Research center (JRC), 2007; Vogt et al., 2007). In addition to the seven major ~~catchments~~ basins (as described above), the model is set up in 222 additional, smaller ~~catchments~~ basins to cross-validate the model performance.

2.2   Meteorological Forcings

The hydrologic model is forced with daily fields of precipitation and minimum, maximum, and average temperature. They are derived from local observations operated by the national weather service (Deutscher Wetterdienst (DWD), 2015). The

15 station network comprises, on average, 3,800 rain gauges and 570 climate stations per year (period: 1951-2010), which have an average minimum distance of 6 km and 14 km between neighboring stations, respectively.

These local observations are interpolated on a regular grid of 4×4 km^2 using external drift Kriging. The terrain elevation (DEM) is used as the external drift, and the Kriging weights are based on a theoretical variogram. The variogram is estimated for all of Germany by fitting to an empirical variogram (see appendix A1). To avoid discontinuities in the interpolated meteo-

20 rological forcings and consecutively in the hydrologic simulation, an estimation of multiple variograms for different climatic zones or distinct morphological regions has been rejected. The spatial resolution of 4×4 km^2 is seen as appropriate, considering the aforementioned station network density ~~.~~ of precipitation observations. The quality of the interpolation is assessed by the Jacknife method (leave one out strategy) which leads to a mean relative bias of 0.64% for all precipitation stations (see appendix A2). Subsequently, daily fields of potential evapotranspiration ~~were~~ are estimated with the Hargreaves-Samani

25 method (Hargreaves and Samani, 1985) using interpolated temperatures (average, minimum, and maximum).

The interpolation of the precipitation is evaluated with gridded precipitation data (REGNIE) provided by the German Meteorological Service (Deutscher Wetterdienst (DWD) (2013); Rauthe et al. (2013)). The REGNIE data are based on the same observations and have a spatial resolution of 1 km. They are derived by applying a multiple linear regression approach, which accounts for daily atmospheric conditions and terrain properties, such as elevation, slope, and aspect (Rauthe et al., 2013). Af-

30 ter remapping the REGNIE data to the aforementioned 4×4 km^2 grid by bilinear interpolation, a satisfactory correspondence between the interpolation and the REGNIE precipitation data is found (see Samaniego et al. (2013)). The spatially averaged bias of the daily fields is 0 with a standard deviation of 0.11 mm d^{-1} within the period 1951-2010.

3 Methodology

3.1 The mesoscale Hydrologic Model mHM

mHM (www.ufz.de/mhm) is a distributed hydrologic model that accounts for the following main processes: snow accumulation and melting, evapotranspiration, canopy interception, soil water infiltration and storage, percolation, and runoff generation. These processes are conceptualized as water fluxes between internal model states similar to existing models, such as HBV (Bergrström, 1976) or VIC (Liang et al., 1994). Snow accumulation and melting processes are based on the improved degree-day method, which accounts for increased snow melting during intense ~~precipitation~~ rainfall events (Hundecha and Bárdossy, 2004). A three-layer discretization is used to account for the processes that represent the root-zone soil moisture dynamics. The two upper layers end in 0.05 m nad 0.25 m, and the lowest layer is spatially variable in depth depending on the soil map. On average, ~~it~~ the lowest layer is 1.8 m deep in Germany. The evapotranspiration from soil layers is estimated as a fraction of the potential evapotranspiration depending on the soil moisture stress and the fraction of vegetation roots present in each layer. The runoff generation in mHM is formalized as the sum of the direct runoff, slow and fast interflow, and baseflow components. The runoff generated at every grid cell is routed to the outlet using the Muskingum-Cunge algorithm. For a detailed model description, interested readers may refer to Samaniego et al. (2010) and Kumar et al. (2013b). To date the model has been successfully applied to various river basins across ~~Germany, Europe, and~~ Europe (including Germany), the USA (Kumar et al., 2010; Samaniego et al., 2013; Kumar et al., 2013a; Thober et al., 2015; Rakovec et al., 2016; Zink et al., 2016), and worldwide (Samaniego et al., 2016) .

A feature that is unique to mHM is its technique for estimating effective model parameters: Multiscale Parameter Regionalization (MPR, Samaniego et al. (2010); Kumar et al. (2013b)). Its basic concept is to estimate parameters (e.g., ~~the~~ soil porosity) based on ~~terrain~~ physiographic properties (e.g., sand and clay content) and transfer functions (e.g., pedotransfer functions). These transfer functions depend on ~~transfer~~ *transfer* or *global* parameters (e.g., factors of the pedotransfer functions) that are time-invariant and location-independent~~and are used for~~ . For the domain of Germany 68 *global* parameters were purpose to an automated calibration (described in section 3.2). ~~The parameter estimation is performed on~~ An overview of the *global* parameters and the resulting effective model parameters can be found in the supplemental material.

This regionalization of model parameters is conducted at the high-resolution land surface property input, e.g., 100×100 m^2~~,~~ ~~and~~ . In a second step these parameters are subsequently upscaled to the user-specified resolution of the hydrologic simulations, e.g., 4×4 km^2~~. Thus, mHM explicitly accounts~~ , by applying parameter-specific upscaling rules (Samaniego et al., 2010) . This procedure yields in effective parameter values (e.g., soil porosity) which are used for the simulation of hydrological processes (e.g., soil water retention). Thus, the effective parameters account for the sub-grid ~~variability~~ variabilities of land surface properties, such as terrain or soil information.

3.2 Derivation of Representative Parameter Sets

One of the goals of this study is to derive consistent model parameters to perform nationwide simulations of water fluxes and states. A two-step parameter selection procedure was used for this purpose. In the first step, we estimate ~~multiple parameter sets~~ 100 sets of *global* parameters via calibration in each of the seven inner German river basins (Figure 1) independently.

In the next step, we transfer these calibrated parameter sets to the remaining basins. The parameter sets exceeding a Nash-Sutcliffe model efficiency of 0.65 (NSE $\geq$ 0.65) in all seven basins during the evaluation period (1965-1999) are retained. This parameter selection procedure ensures that the resulting ensemble parameter sets do not exhibit spatial discontinuities at ~~catchment~~ basin boundaries.

The calibration is performed using the dynamically dimensioned search (DDS) algorithm (Tolson and Shoemaker (2007)). The objective function for calibration consists of an equally weighted power law function for the NSE (Nash and Sutcliffe, 1970) of the ~~discharge~~ streamflow and the logarithm of the ~~discharge~~ streamflow to consider high and low flows within the objective function. A compromise programming technique (Duckstein, 1984) using a power law with an exponent $p = 6$ is used to estimate the multi-objective function (Φ). This technique ensures equal improvement of the different measures ϕ_i during a multi-objective calibration. The overall objective function Φ is given as

$$\Phi = \left(\sum_{i=1}^{2} w_i^p \phi_i^p \right)^{\frac{1}{p}} \text{ with } \sum w_i = 1 \tag{1}$$

with

$$\phi_1 = \text{NSE}(Q) \quad = 1 - \frac{\sum_{t=1}^{T}(\widehat{Q_t} - Q_t)^2}{\sum_{t=1}^{T}(Q_t - \overline{Q})^2} \tag{2}$$

$$\phi_2 = \text{NSE}(\ln Q) = 1 - \frac{\sum_{t=1}^{T}(\ln \widehat{Q_t} - \ln Q_t)^2}{\sum_{t=1}^{T}(\ln Q_t - \overline{\ln Q})^2} \tag{3}$$

where w_i is the weight ($w_1 = w_2 = 0.5$) for a particular measure ϕ_i, $\widehat{Q_t}$ and Q_t denote the modeled and observed ~~discharge~~ streamflow at a time step t, respectively. $\overline{Q}$ is the mean of the observed ~~discharge~~ streamflow over all time steps T.

The period of 5 years from 2000 to 2004 is chosen for model calibration. This time period reflects various hydrologic conditions ranging from a high-impact flood event in Central Europe in August 2002 to a significant drought event in 2003. The remaining 35 years of available data (1965-1999) are used for model evaluation. All simulations are conducted with a 5-year spin-up period to abrogate the influence of initial conditions.

One hundred independent calibration runs are performed for each of the seven ~~catchments~~ basins (Figure 1). Using 2,000 model iterations per calibration run led to a large number of model evaluations per ~~catchment~~ basin (200,000). Finally, 100 of the 700 parameters sets are retained to derive nationwide ensemble simulations of water fluxes and states at a daily resolution.

7

3.3 Validation Data

In addition to ~~discharge~~ streamflow in the seven major German river basins, the model performance is evaluated against ~~discharge~~ streamflow in 222 additional ~~catchments~~ basins and complementary data sets including evapotranspiration, soil moisture and groundwater recharge. The cross-validation of ensemble parameter sets in ~~catchments~~ basins that have not been
5   used for parameter inference should prove the ability of the model to satisfactorily estimate ~~discharge~~ streamflow in various regions of Germany with differing hydrologic characteristics.

The ~~catchments~~ basins for cross-validation are distributed all over Germany and range in size from 100 km^2 to 8,500 km^2. A detailed characterization of these basins is given in Table S3 within the supplemental material. A subset of these ~~catchments contains sub-catchments~~ basins contains sub-basins of seven major basins. The simulation time period is adopted for the
10   available ~~discharge~~ streamflow observations but is at least 10 years. The mean simulation time period of all 222 ~~catchments~~ basins is 42 years. The ~~discharge~~ streamflow estimation in these ~~catchments~~ basins is evaluated using the ensemble median NSE, and its uncertainty is characterized by the range between the 5th and 95th percentiles of NSEs of the ensemble simulation.

~~Evapotranspiration~~ Local evapotranspiration observations are available at seven eddy covariance towers located in Germany (Figure 1, www.europe-fluxdata.eu). ~~These towers are designed to observe carbon~~ Carbon and water fluxes as well as all ~~fluxes~~
15   components of the energy balance, ~~i.e.,~~ latent heat (or evapotranspiration E_a), sensible heat H, ground heat flux G and net radiation R_n. ~~However, the observed fluxes have discrepancies in fulfillment of the energy balance ($R_n = E_a + H + G$), called the energy balance closure gap (Foken, 2008) . The source of the energy balance closure gap is still a subject of research. It is closed by applying mathematical corrections to the latent heat and sensible heat flux to satisfy the energy balance equation. Here, we~~ , are measured at the towers. The energy balance is, however, often not closed at the towers (Foken, 2008; Leuning et al., 2012) so
20   that the observed fluxes usually underestimate the real values, which needs to be corrected before comparison with a model conserving the water balance. We apply a correction ~~method to preserve the fractions of latent and sensible heat~~ to the observed fluxes similar to Kessomkiat et al. (2013) . The corrected evapotranspiration values at the eddy sites are compared with the corresponding model estimates based on the root mean squared error (RMSE), the Pearson correlation coefficient (ρ) and the bias.

25   Additionally, soil moisture observations, undertaken at eddy covariance stations, are used to evaluate modeled soil moisture. Soil moisture is measured using TDR or FDR sensors, which have a control volume of a few cm^3. This is much smaller than the model resolution of 100×100 m^2. A direct comparison between observed and simulated soil moisture may therefore be misleading due to differences in spatial representativeness and sampling depth. Here we aim to analyze the temporal dynamics of soil moisture by normalizing the respective soil moisture time series (Koster et al., 2009). The anomalies are calculated as

30
$$z(t) = \frac{SM(t) - \mu}{\sigma} \tag{4}$$

where μ is the mean and σ is the standard deviation of the entire soil moisture time series SM at a daily resolution. It is not possible to use deseasonalized values (normalization with monthly values) because the time periods of the available

observations are too short ($\approx$ 6 years). The modeled soil moisture is defined herein as the fraction of porosity, i.e., the soil water content divided by porosity.

The mHM simulation for comparing the observations at the location of the eddy covariance stations is conducted with deactivated lateral processes on a single grid cell. The model resolution (100×100 m^2) is adapted to the size of the footprint of the energy flux measurements, which is typically several tens to hundreds of meters. Rather than downscaling the model results, the hydrologic processes are modeled at the resolution of the observations. The transferability of mHM across scales is presented in Samaniego et al. (2010) and Kumar et al. (2013b).

~~We evaluated the model performance against long-term estimates of~~ The model is evaluated with spatially distributed data, i.e., evapotranspiration and groundwater recharge, besides the evaluation of the model at the point or local scale. A remote sensing based dataset is used for evaluating the monthly modeled evapotranspiration between 2001-2010. For this purpose we used the gridded ET dataset based on the Moderate Resolution Imaging Spectroradiometer (MODIS) , which was acquired from the Numerical Terradynamic Simulation Group at the University of Montana (Mu et al., 2007, 2011) . The spatial resolution is approximately 5×5 km^2 ($0.05°$) which is close to the model resolution of 4×4 km^2. The evapotranspiration estimates are based on the Penman-Monteith energy balance equation using global daily temperature, actual vapor deficit, incoming solar radiation as well as remotely sensed leaf area index, fraction of photosynthetic active radiation, albedo, and land cover characteristics. The meteorological variables are based on the reanalysis product from the Global Modeling and Assimilation Office whereas vegetation products are derived from MODIS. Interested readers may refer to ~~Mu et al.~~ Mu et al. (2007, 2011) for detailed description of the MODIS ET product.

As a second spatial dataset we utilize a long-term estimate of annual recharge over Germany (1961-1990). Due to the lack of observations, the estimated recharge from the Hydrologic Atlas of Germany (Federal Ministry for the Environment Nature Conservation Building and Nuclear Safety, 2003) is taken here as a reference. This recharge estimate is obtained using a multiple regression model accounting for ~~terrain properties (e.g., land cover), locally observed baseflow indices and depths~~ long-term estimated generated runoff, depth of the groundwater table~~among other variables~~ , and regionalized baseflow indices (Neumann and Wycisk, 2003). The regionalized baseflow indices are estimated with a linar regression based on the ratio between direct runoff and total runoff as well as terrain properties, such as slope and land cover among others. Due to the various assumptions and mathematical fittings behind this recharge estimate, it is taken as an indication for model evaluation rather than an evidence. The gridded recharge estimate is available at a 1×1 km^2 spatial resolution, which is remapped to a 4×4 km^2 resolution using bilinear interpolation to be comparable to the model estimates.

3.4 Uncertainty of Ensemble Model Simulations

The uncertainty of the modeled evapotranspiration, groundwater recharge, grid-cell-generated runoff and soil moisture is assessed by two different criteria. First, the spatially distributed uncertainties are presented as maps showing the coefficient of variation c_v, which is defined as

$$c_v = \frac{\sigma}{\mu} \tag{5}$$

in which μ is the mean and σ the standard deviation of the ensemble simulations. A large c_v describes a large variation in the modeled flux or state normalized with ~~its mean. The mean~~ μ. μ and ~~standard deviation~~ σ are derived from the 100 ensemble realizations of the hydrologic model mHM on every grid cell. The variances within the ensemble simulation are caused by predictive uncertainties. These uncertainties stem from the parametric uncertainty itself and from the transfer of parameters to locations that have not been used for model calibration. In the following, the variations of the ensemble simulations are denoted as uncertainty.

Second, to assess the temporal variation of the uncertainty throughout a year, the range and normalized range of the respective flux or state are considered. The range is defined as the difference between the 5th (p_5) and 95th (p_{95}) percentiles of the ensemble simulation, whereas the normalized range is defined as

$$r = \frac{p_{95} - p_5}{p_{50}} \tag{6}$$

where $p_{50}(x)$ denotes the median value of the ensemble simulation (50th percentile). The 5th and 95th percentiles are chosen to exclude potential outliers from the analysis.

4 Results and Discussion

The model simulations are evaluated against multiple variables available at different spatial and temporal resolutions. These include daily and monthly time-series of streamflow measured at the ~~catchment~~ basin outlets, soil moisture and evapotranspiration at ~~several~~ seven eddy covariance sites, monthly fields of satellite retrieved evapotranspiration, and a long-term, annual recharge map. mHM simulations are carried out at an hourly time scale at two spatial resolutions, i.e., 100×100 m^2 at the eddy covariance stations and 4×4 km^2 at the ~~catchment~~ basin level and for the nationwide ensemble simulations. Finally, an analysis of the model runs for the nationwide water fluxes and states, including grid-cell-generated runoff (Q_G), evapotranspiration (E_a), groundwater recharge (R) and soil moisture (SM), is presented. The ~~soil moisture is defined herein as the fraction of porosity, i.e., the soil water content divided by porosity. The~~ focus here is to provide a comprehensive overview of regional-scale water fluxes and states over Germany and analyze the uncertainty in modeled variables due to an ensemble of model parameters. The uncertainties are investigated with respect to their temporal and spatial distributions and their triggering sources. Finally, the interaction of uncertainties through the different model states and fluxes is analyzed.

4.1 ~~Discharge~~ Streamflow Evaluation in Major German River Basins

~~The discharge simulations of the hydrological model mHM are evaluated based on the NSE of the~~ In this section we present the evaluation of mHM simulated streamflow with observations in terms of NSEs at daily and monthly ~~discharge values~~ timescale for a validation (1965-1999) and a calibration (2000-2004) period. ~~The daily discharge of~~ Additionally, we show the hydrographs resulting from the ensemble parameter sets in comparison with observed streamflow.

The daily streamflow dynamics in the major German ~~catchments is sufficiently estimated revealing mean NSEs~~ basins is satisfactorily captured by the model revealing a mean NSE of 0.89 and 0.84 using the on-site calibrated parameters in the calibration and validation ~~period~~periods, respectively (Figure 2 ~~). Note that the ensemble parameter sets are common to all basins (grey boxes~~in Figure2). ~~They are chosen as compromise parameter sets, which should perform well in all seven basins~~ ~~(see section3.2). The median model performance of the ensemble parameters drops by approximately 6% compared to on-site estimated parameters. This performance loss can be attributed to changes in the basin climatic and land-surface conditions including terrain, soil, and vegetation properties. The ranges of NSEs, which correspond to the 100 on-site and ensemble parameter sets, are comparable across the investigated basins, which indicates that the application of the ensemble parameter sets did not significantly increase the uncertainty of the estimated discharge.~~

~~The model performance~~ white boxes). The model performance is lower during the validation period in comparison to the calibration period~~(Figure2)~~. Such a deterioration of model performance, which is common to other hydrological model applications, is caused by differences in hydro-meteorological regimes between the calibration and validation periods (Merz and Blöschl, 2004; Me constraining (over-fitting) of the parameters to compensate for errors in the model structure ~~. The~~ (Clark and Vrugt, 2006) . Using the on-site calibrated parameter sets, the model exhibited improved performance for monthly streamflow simulations with an average median NSE of 0.97 and 0.92 ~~for on-site calibrated parameter sets~~ during the calibration and validation period, respectively.

The ensemble parameter sets, which are depicted as the grey boxes in Figure 2, also reveal appropriate model performance. The median NSE corresponding to the ensemble parameter sets is 0.80 for daily streamflow in the validation period averaged across the seven basins. The median NSE of the ensemble parameters drops by approximately 6% compared to that of the on-site estimated parameters. This loss is reasonable considering that the ensemble parameter sets are a compromise solution, which should perform well across all seven basins (see section 3.2). The performance loss can be attributed to changes in the specific basin climatic and land-surface conditions including terrain, soil, and vegetation properties.

Changes in the predictive uncertainty corresponding to on-site and ensemble parameter sets are assessed using the range of model performance. The ~~corresponding NSEs with the transferred parameter sets were 0.94 and 0.87, respectively. The~~ spread of NSEs for the monthly streamflow is considerably narrower compared to the daily flows (Figure 2). ~~Unsurprisingly, the high variabilities~~ The high temporal variability of the daily streamflow ~~are~~ is smoothed when averaged over a longer (monthly) time scale leading to an overall better correspondence between observed and simulated flows.

~~Heavy human interactions lead to lower model performances for~~ The ranges of NSEs corresponding to the 100 on-site and ensemble parameter sets are comparable across the investigated basins with exception of Main and Danube. In these two basins

the ensemble parameter sets provided a relatively larger range of NSEs. The relatively higher spread in NSE in those basins is likely to stem from the fact that different basins are sensitive to different parameters. For example, the Ems basin, located in the maritime-influenced north, is not as sensitive to snow parameters as the alpine-influenced Danube basin. Consequently, parameters that originate from the Ems basin potentially deteriorate ensemble predictions in the Danube basins. A simultaneous calibration of multiple, distinct basins would be beneficial for deriving hydrological fluxes and states at national or continental scales.

Examples of the modeled streamflow time series are given in Figure 3. In general, the model is able to adequately capture the discharge dynamics across the investigated basins. A relatively lower mode skill in capturing the discharge dynamics in the Saale river basin ~~, especially on the daily timescale~~ can be attributed to heavy human interactions. The highly regulated ~~discharge~~ streamflow in the headwaters of the Saale river (see section 2) is difficult to capture and thus leads to lower ~~performances~~ performance because mHM includes no reservoir operation. The main discharge mechanisms of Saale are considered to be adequately captured because the median NSEs are exceeding 0.85 and 0.7 at the monthly and daily resolutions for the ensemble parameter sets, respectively (Figure 2).

Interestingly, this ~~catchment~~ basin shows equal or higher performance for the ensemble parameter sets compared to the on-site parameter sets in the evaluation period. A similar behavior can be observed for the Weser ~~catchment~~basin. We conclude that ~~discharge~~ streamflow simulations in some ~~catchments~~ basins improve by gaining knowledge from remote locations.

The ~~filtering of transferred parameters to determine the ensemble parameters introduces a notable degree of uncertainty in some of the catchments, e.g., the Danube. This stems from the fact that different catchments are sensitive to different parameters. For example, the Ems, located in the maritime-influenced north, is not as sensitive to snow parameters as the alpine-influenced Danube is. Consequently, parameters that originate from the Ems deteriorate ensemble predictions in the Danube. A simultaneous calibration of multiple, distinct catchments would be beneficial for deriving hydrological fluxes and states at national or continental scales.~~

~~The~~ Mulde basin has a tendency to underestimate peak flows (Figure 3). This could be attributed to ~~our~~ the precipitation product. The headwaters of the Mulde basin are located in the Ore mountains at the border between Germany and the Czech Republic (Figure 1). In addition to a sparse network of rain gauges in these mountainous area, a lack of information on meteorological variables from the neighboring country (i.e., the Czech Republic) leads to an underestimation of precipitation ~~by the interpolation~~ in the interpolation process, especially for orographic-driven events. ~~In other basins, the model is able to adequately capture both high and low flows (Figure3)~~The model performance for the Mulde is comparably superior to those found by other studies, like Fleischbein et al. (2006) or Huang et al. (2010) .

The results presented in this section show that the method for determining ensemble parameter sets (section 3.2) leads to satisfactory estimations of ~~discharge in the catchments~~ streamflow in the basins used for parameter inference. ~~However~~Overall, the model ~~performances shown within this section compare~~ performance shown herein compares well to those of other studies, such as ~~that by Huang et al. (2010)~~ Lohmann et al. (1998); Strasser and Mauser (2001); Menzel et al. (2006); Fleischbein et al. (2006) ; and Huang et al. (2010) . A further investigation of the applicability of the ensemble parameter sets on additional, smaller ~~catchments~~ basins is shown in the ~~next~~ following section.

4.2 ~~Discharge~~ Streamflow Evaluation at Non-calibrated Basins

Following Klemeš (1986), the model performance is evaluated across 222 ~~catchments~~ basins diverging in size and geographical location. The streamflow data of these proxy locations have not been used during the model calibration. This cross-validation test focuses on evaluating the model performance against ~~discharge~~ streamflow simulations along a diverse range of climatic and land-surface conditions. The evaluations shown in Figure 4 indicate a satisfactory agreement between simulations and observations. The daily ~~discharge~~ streamflow simulations (Figure 4, panels A, B) reveal a median NSE value of at least 0.5 across the investigated basins based on the ensemble parameter sets. The overall average NSE value is 0.68. Expectedly, the model exhibits better skill in capturing monthly ~~discharge~~ streamflow dynamics, with an ensemble median NSE averaged across all basins of approximately 0.81 (Figure 4, panels D, E). Furthermore, the ensemble median NSE exceeded a value of 0.75 in more than 20% of the basins for the daily flows and 80% for the monthly flows. The spatial variability of the median NSE across the investigated basins is low with a standard deviation of approximately 0.09 for both daily and monthly flows.

To illustrate different climatic regimes of the 222 ~~catchments~~ basins, we make use of the dryness index E_p/P (Budyko, 1974). Various studies describe the relationship between the dryness and evaporative index E_a/P (Schreiber, 1904; Ol'dekop, 1911; Budyko, 1974; Gerrits et al., 2009) and span an uncertainty band around Budyko's curve. The model ~~performances~~ performance of the 222 ~~catchments are~~ basins is plotted in panels A and D of Figure 4 using these indexes. It separates the ~~catchments~~ basins into energy- ($E_p/P < 1$) and water-limited conditions ($E_p/P > 1$). The simulated evapotranspiration E_a is used to derive the Budyko plot to identify potential errors in the water balance closure (Figure 4 panels A, D). All ~~catchments~~ basins under investigation lie perfectly within the uncertainty ranges of the reported theoretical curves. ~~In~~ Please note that energy limited basins are closer to the lower uncertainty line of the reported curves, whereas water limited basins tend to the upper curve. In consequence basins with energy limitation tend to underrepresent the original Budyko curve and develop to overrepresentation for water limited locations. In conclusion, the water balances of those basins are well closed, with a mean closure error of 1% for the median simulation. The ~~performances are comparable for catchments~~ performance is comparable for basins in different climatic regimes. Such behavior is not obvious as studies like Newman et al. (2015b) and Xia et al. (2012a) found a significant dependency on the climatic regime. However, a tendency to perform better in large ~~catchments~~ basins is observed. A similar conclusion was drawn by McMillan et al. (2016) .

We further analyzed the relationship between model performance and physiographic attributes (e.g, terrain or land cover characteristics). These analyses did not show any significant relationship (see Figure B1). The absence of pairwise relationships between model performance and climatic or land surface characteristics confirms the validity of the derived ensemble parameters for the national scale. In contrast, Newman et al. (2015b) and McMillan et al. (2016) observed significant dependencies between model performance and basin characteristics, such as aridity or basin area.

The uncertainty for the individual basins caused by the ensemble parameter sets is expressed as the range between the 5th and 95th percentiles of NSEs (Figure 4, panels C, F). Substantial performance differences occur in 70% (45%) of the basins exceeding a range of 0.1 NSE for the daily (monthly) flow simulations. A geographical dependency of the uncertainty cannot be found as no spatial clustering is observed~~, whereas~~. Whereas daily flows show almost no relation between median NSEs and

13

the uncertainty range, i.e., worse performing ~~catchments~~ basins reveal high uncertainties, ~~and monthly NSEs are less uncertain if their NSE~~ the monthly NSEs show less uncertainty if the corresponding model performance is high.

The evaluation of the ensemble parameter sets presented in this section supports the hypothesis that the ensemble parameter sets are valid on the national scale. Studies like Perrin et al. (2008); Xia et al. (2012a); Cai et al. (2014); McMillan et al. (2016) and Hostetler and Alder (2016) validate their models based on streamflow over a large sample of basins and observed similar or lower NSEs. In the following section, evapotranspiration, soil moisture, and groundwater recharge estimates are evaluated.

4.3 Evapotranspiration and Soil Moisture Evaluation at Eddy Covariance Stations

The ensemble model simulations are further evaluated with the evapotranspiration (E_a) and soil moisture (SM) observed at seven eddy covariance stations (Figure 1) to assess the model's ability to represent other fluxes and states next to streamflow. The ensemble median of the daily sum of evapotranspiration is plotted against the corresponding observations in Figure 5, and the resulting error ~~statistics~~ metrics are summarized in Table 2.

The ~~results of the scatter plot,~~ scatter plots shown in Figure 5 ~~,~~ indicate no systematic over- or underestimation of the observed evapotranspiration. The highest deviation in terms of RMSE is observed during summer, when the highest fluxes occur, and the lowest during winter, in which the contribution of E_a is lowest among all seasons. The average bias estimated across all stations during spring is 0.34 mm d^{-1}, whereas it is 0.08 mm d^{-1}, 0.04 mm d^{-1} and 0.04 mm d^{-1} for winter, summer and autumn, respectively. The slight overestimation of the modeled E_a during spring is likely caused by the lack of a dynamic vegetation growth module in mHM. Thus, the onset of the vegetation period may not be captured adequately by the model. With respect to the vegetation class, the stations E1 and E6 covered by crops have the largest errors, with E_a RMSEs of ~~E_a of~~ 19.4 mm mon^{-1} and 15.4 mm mon^{-1} for monthly evapotranspiration, respectively (Table 2). These errors arise because of the high impact of human interactions on croplands, e.g., due to seeding, harvesting or irrigation, compared to other vegetation classes. Additionally, the land cover class cropland is not explicitly represented within the model; rather, it is generalized within a mixed land cover class, representing all land cover types different from sealed and forest. Varying goodness of fit for different land covers and seasons for evapotranspiration at eddy flux towers were found for the four land surface models used in NLDAS (Xia et al., 2015) and thus are not uniquely observed for mHM.

In general, errors of local evapotranspiration estimates can be attributed to limitations of the Hargreaves-Samani approach for estimating the potential evapotranspiration. This approach may be inappropriate for local weather conditions. Because this method approximates the net radiation based on the minimum and maximum daily temperatures, local phenomena such as short term cloudiness, e.g., due to convective precipitation cells, are not accounted for. This effect is especially high in summer, which causes the lowest correlations between observations and simulations during this period. Unfortunately, only temperature based methods are supported by the available input data. Please notice that the observational error caused by the energy balance closure gap ~~was~~ is, on average, 33% for the herein considered stations before applying the above-mentioned mathematical corrections.

In terms of temporal dynamics, the model is able to capture the observed evapotranspiration quite well across the different eddy covariance sites, as exemplarily shown in the upper panel of Figure 6. The model is able to adequately represent the

14

observed monthly dynamics with an average correlation of approximately 0.93 (Table 2). The correlation between the observed and the simulated daily evapotranspirations is at least 0.77, with the exception of the cropland site E1.

The lower panel of Figure 6 shows the performance of mHM in representing the daily soil moisture anomalies, which are generally in good correspondence with observations. The temporal dynamics of observed soil moisture anomalies during the wetting and drying phases are well captured by the model. The resulting correlation shown in Table 2 at different eddy stations ranges between 0.53 and 0.93. These correlations were similar to those of other studies, such as Cai et al. (2014) . The lowest values are observed at cropland sites, which is due to the above-mentioned human interaction and land cover class representativeness. The amplitude of the observed soil moisture anomalies is adequately captured by the model. Still, some peaks are not reproduced satisfactorily, which could be due to the non-representativeness of the 100×100 m^2 model grid cell for TDR/FDR soil moisture measurements. Thus, the simulated soil moisture is smoother compared to the observation because it represents the effective soil moisture of the entire grid cell.

4.4 Evaluation ~~of Groundwater Recharge~~ with Spatially Distributed Data

~~Finally, the ensemble simulations are evaluated with the~~

In this section, we present results of the model skill in representing gridded fluxes over the entire German domain. The first comparison is conducted for the assessment of reproducing the monthly fields of modeled ET against the remotely sensed MODIS ET product. The results are summarized in Figure 7 in terms of three key metrics: relative bias, correlation and root mean square error (RMSE). The analysis is conducted using the ensemble mean of ET from the 100 model simulations. The modelled ET is able to adequately capture the spatio-temporal features of the MODIS derived ET product with the majority of grid cells (74%) having a relative absolute bias of less than 10%. Notable differences among these two ET datasets are appearing in lowland areas along the Danube river basin in South Germany, where the modeled ET exhibited a dry bias compared to the MODIS ET. An opposite trend of postive bias in modeled ET is observed for grid cells lying along the coastal region in north Germany. The temporal correspondence between both evapotranspiration datasets is also remarkably high with an average Pearson correlation coefficient of 0.96 (standard deviation 0.02). Notably both ET datasets exhibit pronounced seasonal variability leading to a high temporal correspondence between them.

The second assessment evaluates the modeled groundwater recharge with long-term ~~annual groundwater recharge~~ annual values from the Hydrologic Atlas of Germany (HAD) (Federal Ministry for the Environment Nature Conservation Building and Nuclear Safety, 2003). mHM's long-term recharge estimate implicitly represents the baseflow component of the total runoff based on the assumption that the underground ~~catchment~~ basin is closed and that there are no external losses (e.g., irrigation or pumping). Consequently, this analysis serves as a proxy for assessing the model skill for partitioning the total runoff into interflow and baseflow. The comparison of the spatial pattern of the recharge shows good accordance between the two maps with a correlation coefficient of approximately 0.8 (Figure 8). The spatial pattern of the recharge follows the known climatology of Germany with high recharge rates being observed in areas with high precipitation amounts (e.g., ~~Central Uplands or Alps~~ Alps - region 11 in Figure 10).

There are some significant differences between the modeled and HAD groundwater recharge, particularly at cells characterized by urbanization (i.e., Munich, Hamburg, Berlin, and the metropolises of Ruhrgebiet in the northwest). The model tends to underestimate the HAD recharges, with differences as high as approximately 200 mm a^{-1}. Notably, the herein used version of mHM treats sealed areas as almost impermeable, which is unrealistic. This issue has been resolved in recent mHM versions

5   (5.0 and higher). In general, the HAD estimate of recharge is, on average, 31 mm a^{-1} higher compared to the ensemble mean simulation. This mismatch arises from the differences in potential evapotranspiration (E_p), which were used for both estimates. The E_p estimates used for the HAD (Federal Ministry for the Environment Nature Conservation Building and Nuclear Safety, 2003) are lower than those used for mHM simulations and result in higher water amounts remaining in the underground. ~~In addition to~~ Besides these mismatches, the spatial pattern of the modeled groundwater recharge compares well with the HAD

10 estimates (Figure 8).

4.5   Spatial Patterns of Ensemble Means and Uncertainties

The estimated evapotranspiration (E_a) and grid-cell-generated runoff (Q_G), as well as their uncertainty, which is expressed as the coefficient of variation of the ensemble simulations, are presented in Figure 9. In addition to these simulation results, Figure 9 shows the mean annual precipitation, dryness index and land surface properties, i.e., porosity and dominating land

15 cover type. Thus, Figure 9 is used to analyze the spatial patterns of uncertainty and their main causes.

The high precipitation amounts above 1000 mm a^{-1} in panel A correspond to mountainous areas in Germany. The driest region is located in the northeastern part of Germany. This is, on the one hand, due to its distance to the sea (continental climate) and, on the other hand, due to the Central Uplands in the western and central part of Germany. These mountains, especially the Harz mountains (center of Germany), capture most of the precipitation events brought from the west. The low amounts

20 of precipitation in the east lead to lower amounts of evapotranspiration (Figure 9, panel B) and grid-cell-generated runoff (Figure 9, panel C) in this region compared to the rest of Germany. Thus, the northeastern part of Germany is characterized by high dryness indexes of 1.2 and above. The uppermost dryness indexes up to 1.4 are located in the lee of the Harz mountains. The average dryness index in Germany is 0.98. Another region characterized by high dryness indexes is the Upper Rhine Valley, which is known to have a locally warmer climate compared to its neighboring regions. Mountainous regions are characterized

25 by stronger energy limitation due to high precipitation amounts, which results in dryness indexes lower than 0.65.

The spatial ~~distributions of~~ distribution of the uncertainty, i.e., the coefficients of variation (see section 3.4), of the ~~simulated evapotranspiration (Figure9, panelF) and~~ grid-cell-generated runoff (Figure 9, panel G) ~~are~~ is mainly governed by the dryness index (Figure 9, panel D). ~~Both variables are prone to high uncertainties in regions of high dryness indexes, e. g., northeastern Germany. Additionally,~~ The Spearman rank correlation between both variables is 0.92. The uncertainty patterns of evapotran-

30 spiration ~~are connected~~ (Figure 9, panel F) have a closer relation to soil textural properties~~.~~, i.e, porosity (Figure 9, panel E), with a Spearman rank coefficient of 0.58 as compared to the dryness index (rank correlation=0.28). Locations of high uncertainty in E_a ~~correspond to regions of high porosities~~, e.g., northern Germany ~~(Figure9, panelsE and F)~~, correspond to regions of high porosities. Within this region, soils are dominated by sand and are highly conductive, which results in low water holding capacities. The modeled evapotranspiration is ~~very sensitive to~~ highly dependent on the soil parameterization because soil

16

water is the main source of evaporative water. In contrast, the ~~spatial structures of uncertainties of~~ uncertainty patterns of grid-cell-generated runoff, e.g., (Q_G) in the northeastern part of Germany and the Upper Rhine Valley, correspond to high values in the dryness index in those regions.

In conclusion, the spatial distribution of the uncertainty in evapotranspiration ~~in northern Germany is partially caused~~ is influenced by the parameterization of the soil ~~, whereas the main pattern is governed~~ whereas the runoff uncertainty pattern is dominated by the dryness index. The ~~amount of water that enters the soil is comparatively low in regions of high dryness. Thus, the model is highly sensitive to the partitioning of the water between the model internal reservoirs, i.e., the surface, soil water and ground water reservoir. In consequence, slight changes in the parameters affect the partitioning of water in the subsurface and lead to changes in the modeled fluxes and states.~~

~~The~~ patterns appearing in the evapotranspiration and grid-cell-generated runoff at the location of big cities (orange areas in panel H of Figure 9) are caused by the above-mentioned old representation of sealed areas for mHM versions prior to 5.0.

4.6 Spatio-temporal Distribution of Uncertainties

This section focuses on the spatio-temporal differences of uncertainties caused by the 100 ensemble parameter sets. Figure 10 shows the climatological dynamics and the normalized ranges (see section 3.4) of the respective variables, i.e., evapotranspiration (E_a), soil moisture (SM), groundwater recharge (R) and grid-cell-generated runoff (Q_G). The rows refer to different environmental zones in Germany (Federal Environmental Agency, 2005), which are depicted in the upper right corner of Figure 10. For comprehensibility only a selection of 5 environmental zones is depicted therein, representing the region of high dryness indexes in the north (zone 2), Central Germany including Central Uplands (zones 4 and 9), the foothills of the Alps (zone 10) and the Alps (zones 11).

The magnitude of the evapotranspiration uncertainty, i.e., the uncertainty range, is lowest among the four variables. Evapotranspiration is estimated by scaling the potential evapotranspiration with the water availability in several reservoirs, i.e., the interception storage, the surface ponds in sealed areas and the soil moisture. ~~It is highly dependent on the model input variable potential evapotranspiration, so its uncertainty magnitude is comparably low~~Notably, most of the area in Germany are characterized by humid and continental climate where the ET is constrained by available energy. The evapotranspiration is thus mainly driven by the potential evapotranspiration. A relatively large uncertainty in soil moisture does not directly propagate to evapotranspiration uncertainty. The highest uncertainties are observed for the groundwater recharge. This model's internal variable is neither closely related to the model input as E_a nor indirectly constrained by calibration as the generated ~~discharge~~streamflow. In consequence, its uncertainty is highest among the four variables.

The evapotranspiration uncertainty shows almost no dynamics during the course of the year. In contrast, the uncertainty of recharge and ~~generated discharge~~ ~~runoff~~ streamflow change significantly during the course of the year. Whereas the dynamics of the groundwater recharge and its uncertainty are positively correlated, the correlation for soil moisture and its uncertainty is negative. Thus, the recharge uncertainty is lowest for low recharge values, which occur in summer when the subsurface reservoirs are comparably dry. The low amplitude of the soil moisture uncertainty is reasoned in the high persistence of soil moisture. Regions of high porosity and low dryness indexes in northern Germany have more distinct dynamics compared to

southern locations. The uncertainty of the generated ~~discharge~~ runoff is a composite of the dynamics of soil moisture and recharge and thus shows the distribution of water among the model's internal reservoirs.

~~It is noticed that the highest uncertainty in recharge corresponds to the lowest uncertainty in soil moisture (zone11). The cause of this behavior is the parameters that control the snow accumulation and melting within mHM. Because the soils are almost close to saturation over the course of a year in this zone, water drains quickly to layers underneath the root zone. In this layer, the interflow and groundwater recharge is determined within mHM. Because the ensemble parameters have been derived in different regions of Germany, snow parameters from regions where snow does not have a large impact on the water balance are involved. Therefore, the uncertainty of groundwater recharge is highest during snow melting in spring for zone11 (Figure10).~~

5   Summary and Conclusion

In this study, we present the derivation and evaluation of a high-resolution ($4{\times}4$ km^2) dataset of hydrologic and meteorological fluxes and states for Germany covering the period 1951-2010, which is freely available. The dataset incorporates 100 spatially consistent ensemble simulations, which are analyzed regarding their uncertainty caused by the parameter estimation. The parameter sets of the ensemble simulations are determined by a two-step parameter selection method. The model is calibrated in seven ~~catchments~~basins, and the parameter sets are filtered based on the cross-validation results in all of the basins. Thus, the uncertainty is composed of the uncertainty in parameter estimation and the uncertainty stemming from transferring these parameters to remote locations. The ensemble simulations are evaluated with streamflow, evapotranspiration and soil moisture observations and recharge data.

A comparable study by Newman et al. (2015a) focuses on the provision of a 100 member ensemble dataset which is focusing on meteorological variables for major parts of North America. Similar to the study presented herein they evaluate the data in a large sample of basins, i.e., 671. We, however, conclude that 100 realizations is an appropriate sample size for an uncertainty assessment study.

The evaluation regarding ~~discharge~~ streamflow at 222 additional ~~catchments~~ basins revealed a median NSE of 0.68. Thus, the 100 ensemble parameter sets are considered to be representative for Germany. The evaluation with evapotranspiration from eddy covariance stations showed deficiencies in mHM. Especially in spring, deviations of the modeled and observed E_a indicate room for improving the representation of vegetation dynamics within mHM. The sites covered by cropland showed the largest deviations from evapotranspiration observations because croplands are highly human-influenced (seeding, harvest, or eventually irrigation), which makes it difficult to model their dynamics at the local scale. Additionally, cropland is generalized in a mixed land cover class in mHM. Soil moisture estimations at the same locations have been in good agreement with the observed dynamics.

The second part of the study focuses on the uncertainty of the simulated hydrological fluxes and states due to uncertainties in parameter estimation. It is shown that uncertainty varies in time, location and magnitude between hydrological variables. Among all of the variables, the uncertainty was lowest for evapotranspiration and highest for recharge. ~~Its spatial distribution~~

was similar The spatial distribution of runoff uncertainty is closely related to the spatial distribution of the dryness index. Only In contrast, the uncertainty patterns of evapotranspiration estimates are mostly connected to soil properties in some regions within Germany. In general, the highest uncertainties occur in the northeastern part of Germany, which is characterized by low precipitation amounts and high soil porosities. The temporal variation of uncertainties is almost constant for evapotranspiration, medium for grid-cell-generated runoff and soil moisture and high for groundwater recharge and depends on geographical location.

Based on these results we suggest incorporating additional data, e.g., in-situ soil moisture or satellite observations, into the calibration procedure to better constrain the model's internal states. The results of this study emphasize the importance of the considering parametric uncertainty for historical analysis, now- and forecasting in hydrology.

6 Data Availability and Data Format

The dataset consists of daily values of precipitation and minimum, maximum and average temperature, potential evapotranspiration, evapotranspiration, soil moisture, groundwater recharge and discharge generated runoff. Whereas the latter four are provided as ensemble of 100 simulations. The data format is the Net Common Data Format (NetCDF version 3) . and is based on the CF conventions (www.cfconventions.org). Additionally, the ensemble means and standard deviations are provided for download. The dataset is freely accessible under Creative Commons license at http://www.ufz.de/index.php?en=41160.

Appendix A: Interpolation of Meteorological Data

A1 Variogram Estimation

The variogram for the German domain is estimated based on two different approaches. In the first approach regionalized variograms for rectangular sub-domains (blocks) were estimated (Figure A1). The interpolation of meteorological variables based on these regionalized variograms, however, lead to discontinuous fields of these meteorological variables. This result contradicted the aim of deriving seamless fields of hydro-meteorological fluxes and states for entire Germany. In consequence, continuous meteorological interpolations have been the prerequisite for the next approach. In the second approach, a compromise variogram for entire Germany is estimated by considering all available data from all meteorological stations, e.g., approximately 5700 stations for precipitation, for the estimation of an empirical variogram. An exponential, theoretical variogram is fitted to this empirical variogram. The fitted variogram curves of both methods are presented exemplarily for precipitation in Figure A1. The empirical variogram is well represented by the theoretical variogram with a root mean squared error of 0.02. The consecutive estimation of meteorological fields is based on the second approach using a compromise variogram for Germany.

A2 Interpolation Error

The interpolation error was assessed by a leave-one out strategy, i.e., the Jacknife method. This cross-validation informs about the ability of the external drift kriging to estimate meteorological variables at locations where observations are available. The algorithm is as follows:

1. Exclude one station from the set of observations.

5
2. Estimate the meteorological time series at this location using external drift kriging.

3. Compare the interpolated time series with the observation and assess the interpolation error at each station.

4. Interpolate the Jacknife-error estimates over the Germany domain using ordinary kriging to obtain error maps for visualization purposes.

The error at each station is characterized by the bias, relative bias, root mean squared error, and Pearson correlation coefficient
10 (Figure A2). Exemplarily we present the errors of the precipitation interpolation because this variable has the highest spatial and temporal variability among the interpolated variables (precipitation; minimum, maximum, and average temperature). The average and the standard deviation for the different errors assessments over all stations are 0.01 and 0.15 mm d^{-1} for the bias, 0.64% and 5.60% for the relative bias, 0.93 and 0.03 for the Pearson correlation coefficient, and 1.75 and 0.48 mm d^{-1} for the root men squared error. Reviewing these values the chosen interpolation approach is seen appropriate.

15 **Appendix B: Relation of Model Performance and Land Surface and Hydro-climatic Characteristics**

The analysis for identifying relations between land surface and hydro-climatic characteristics and model performance is presented in Figure B1. This analysis does not reveal any hydro-meteorological or morphological conditions which explain different model performance in distinct basins. In conclusion, the retrieved parameter sets are representative for various climatic and physiographic conditions.

References

Berg, A. A., Famiglietti, J. S., Rodell, M., Reichle, R. H., Jambor, U., Holl, S. L., and Houser, P. R.: Development of a hydrometeorological forcing data set for global soil moisture estimation, International Journal of Climatology, 25, 1697–1714, doi:10.1002/joc.1203, 2005.

Bergrström, S.: Development and application of a conceptual runoff model for Scandinavian catchments, Tech. rep., University of Lund, Norrköping, Sweden, 1976.

Beven, K.: Prophecy, reality and uncertainty in distributed hydrological modelling, Advances in Water Resources, 16, 41–51, 1993.

Bierkens, M. F. P., Bell, V. a., Burek, P., Chaney, N., Condon, L. E., David, C. H., de Roo, A., Döll, P., Drost, N., Famiglietti, J. S., Flörke, M., Gochis, D. J., Houser, P., Hut, R., Keune, J., Kollet, S., Maxwell, R. M., Reager, J. T., Samaniego, L., Sudicky, E., Sutanudjaja, E. H., van de Giesen, N., Winsemius, H., and Wood, E. F.: Hyper-resolution global hydrological modelling: what is next?, Hydrological Processes, 29, 310–320, doi:10.1002/hyp.10391, 2015.

Budyko, M. I.: Climate and Life, Academic Press, New York, 1974.

Cai, X., Yang, Z.-l., Xia, Y., Huang, M., Wei, H., Leung, L. R., and Ek, M. B.: Assessment of simulated water balance from Noah, Noah-MP, CLM, and VIC over CONUS using the NLDAS test bed, Journal of Geophysical Research: Atmospheres, 119, 13,751–13,770, doi:10.1002/2014JD022113, 2014.

Clark, M. P. and Vrugt, J. a.: Unraveling uncertainties in hydrologic model calibration: Addressing the problem of compensatory parameters, Geophysical Research Letters, 33, L06 406, doi:10.1029/2005GL025604, 2006.

Day, G. N.: Extended Streamflow Forecasting Using NWSRFS, Journal of Water Resources Planning and Management, 111, 157–170, doi:10.1061/(ASCE)0733-9496(1985)111:2(157), 1985.

Dee, D., Fasullo, J., Shea, D., Walsh, J., and National Center for Atmospheric Research Staff: The Climate Data Guide: Atmospheric Reanalysis: Overview & Comparison Tables, 2016.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, doi:10.1002/qj.828, 2011.

Deutscher Wetterdienst (DWD): REGNIE: Verfahrensbeschreibung und Nutzeranleitung, 2013.

Deutscher Wetterdienst (DWD): Climate station data, 2015.

Duckstein, L.: Multiobjective Optimization in Structural Design: The Model Choice Problem, in: New directions in optimum structural design, edited by Atrek, E., Gallagher, R. H., Ragsdell, K. M., and Zienkiewicz, O. C., pp. 459–481, John Wiley, New York, doi:10.1002/oca.4660060212, 1984.

European Commission - Joint Research center (JRC): CCM River and Catchment Database, http://ccm.jrc.it/, 2007.

European Environmental Agency (EEA): CORINE Land Cover 1990, 2000 and 2006, http://www.eea.europa.eu, 2009.

European Water Archive (EWA): Runoff data, http://ne-friend.bafg.de, 2011.

Fan, Y. and van den Dool, H.: Climate Prediction Center global monthly soil moisture data set at 0.5° resolution for 1948 to present, Journal of Geophysical Research D: Atmospheres, 109, D10 102, doi:10.1029/2003JD004345, 2004.

Federal Agency for Cartography and Geodesy (BKG): Digital Elevation Model (DEM), 2010.

Federal Environmental Agency: Climate Change in Germany: Vulnerability and Adaption of Climate sensitive Sectors: Research Report 201 41 253, Tech. rep., Federal Environmental Agency (Umweltbundesamt), Dessau, Germany, 2005.

Federal Institute for Geosciences and Natural Resources (BGR): Digital soil map of Germany 1:1,000,000 (BUEK 1000), 1998.

Federal Institute for Geosciences and Natural Resources (BGR): Hydrogeological map of Germany 1:200,000 (HUEK 200), 2009.

5　Federal Ministry for the Environment Nature Conservation Building and Nuclear Safety: Hydrological Atlas of Germany (HAD), Bonn, 2003.

Fleischbein, K., Lindenschmidt, K.-E., and Merz, B.: Modelling the runoff response in the Mulde catchment (Germany), Advances in Geosciences, 9, 79–84, doi:10.5194/adgeo-9-79-2006, 2006.

FLUXNET: accessed at April 2016, https://fluxnet.ornl.gov, 2007.

10　Foken, T.: The energy balance closure problem: An overview, Ecological Applications, 18, 1351–1367, doi:10.1890/06-0922.1, 2008.

Gerrits, A. M. J., Savenije, H. H. G., Veling, E. J. M., and Pfister, L.: Analytical derivation of the Budyko curve based on rainfall characteristics and a simple evaporation model, Water Resources Research, 45, 1–15, doi:10.1029/2008WR007308, 2009.

Global Runoff Data Centre (GRDC): Runoff data, http://www.bafg.de/GRDC, 2011.

Hargreaves, G. and Samani, Z.: Reference crop evapotranspiration from ambient air temperature, American Society of Agricultural Engineers, 1, 96–99, 1985.

15

Huang, S., Krysanova, V., Österle, H., and Hattermann, F. F.: Simulation of spatiotemporal dynamics of water fluxes in Germany under climate change, Hydrological Processes, 24, 3289–3306, doi:10.1002/hyp.7753, 2010.

Hundecha, Y. and Bárdossy, A.: Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model, Journal of Hydrology, 292, 281–295, doi:10.1016/j.jhydrol.2004.01.002, 2004.

20　Kessomkiat, W., Franssen, H.-J. H., Graf, A., and Vereecken, H.: Estimating random errors of eddy covariance data: An extended two-tower approach, Agricultural and Forest Meteorology, 171-172, 203–219, doi:10.1016/j.agrformet.2012.11.019, 2013.

Klemeš, V.: Operational testing of hydrological simulation models, Hydrological Sciences Journal, 31, 13–24, doi:10.1080/02626668609491024, 1986.

Koster, R. D., Guo, Z., Yang, R., Dirmeyer, P. a., Mitchell, K., and Puma, M. J.: On the Nature of Soil Moisture in Land Surface Models, Journal of Climate, 22, 4322–4335, doi:10.1175/2009JCLI2832.1, 2009.

25

Kumar, R., Samaniego, L., and Attinger, S.: The effects of spatial discretization and model parameterization on the prediction of extreme runoff characteristics, Journal of Hydrology, 392, 54–69, doi:10.1016/j.jhydrol.2010.07.047, 2010.

Kumar, R., Livneh, B., and Samaniego, L.: Toward computationally efficient large-scale hydrologic predictions with a multiscale regionalization scheme, Water Resources Research, 49, 5700–5714, doi:10.1002/wrcr.20431, 2013a.

30　Kumar, R., Samaniego, L., and Attinger, S.: Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations, Water Resources Research, 49, 360–379, doi:10.1029/2012WR012195, 2013b.

Kumar, R., Musuuza, J. L., Van Loon, A. F., Teuling, A. J., Barthel, R., Ten Broek, J., Mai, J., Samaniego, L., and Attinger, S.: Multiscale evaluation of the Standardized Precipitation Index as a groundwater drought indicator, Hydrology and Earth System Sciences, 20, 1117–1131, doi:10.5194/hess-20-1117-2016, 2016.

35　Hartmann, A., Gleeson, T., Rosolem, R., Pianosi, F., Wada, Y., and Wagener, T.: A large-scale simulation model to assess karstic groundwater recharge over Europe and the Mediterranean, Geoscientific Model Development, 8, 1729–1746, doi:10.5194/gmd-8-1729-2015, 2015.

Henriksen, H. J., Troldborg, L., Nyegaard, P., Sonnenborg, T. O., Refsgaard, J. C., and Madsen, B.: Methodology for construction, calibration and validation of a national hydrological model for Denmark, Journal of Hydrology, 280, 52–71, doi:10.1016/S0022-1694(03)00186-0, 2003.

Hostetler, S. W. and Alder, J. R.: Implementation and evaluation of a monthly water balance model over the US on an 800 m grid, Water Resources Research, 52, 1–20, doi:10.1002/2016WR018665, 2016.

Leuning, R., van Gorsel, E., Massman, W. J., and Isaac, P. R.: Reflections on the surface energy imbalance problem, Agricultural and Forest Meteorology, 156, 65–74, doi:10.1016/j.agrformet.2011.12.002, 2012.

Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models, Journal of Geophysical Research, 99, 14 415–14 428, doi:10.1029/94JD00483, 1994.

Liu, Y., Dorigo, W., Parinussa, R., de Jeu, R., Wagner, W., McCabe, M., Evans, J., and van Dijk, A.: Trend-preserving blending of passive and active microwave soil moisture retrievals, Remote Sensing of Environment, 123, 280–297, doi:10.1016/j.rse.2012.03.014, 2012.

Livneh, B., Rosenberg, E. a., Lin, C., Nijssen, B., Mishra, V., Andreadis, K. M., Maurer, E. P., and Lettenmaier, D. P.: A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States: Update and Extensions, Journal of Climate, 26, 9384–9392, doi:10.1175/JCLI-D-12-00508.1, 2013.

Livneh, B., Bohn, T. J., Pierce, D. W., Munoz-Arriola, F., Nijssen, B., Vose, R., Cayan, D. R., and Brekke, L.: A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950–2013, Scientific Data, 2, 150 042, doi:10.1038/sdata.2015.42, 2015.

Lohmann, D., Raschke, E., Nijssen, B., and Lettenmaier, D. P.: Regional scale hydrology: II. Application of the VIC-2L model to the Weser River, Germany, Hydrological Sciences Journal, 43, 143–158, doi:10.1080/02626669809492108, 1998.

Maurer, E. P., Wood, A. W., Adam, J. C., Lettenmaier, D. P., and Nijssen, B.: A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States, Journal of Climate, 15, 3237–3251, doi:10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2, 2002.

McMillan, H., Booker, D., and Cattoën, C.: Validation of a national hydrological model, Journal of Hydrology, 541, 800–815, doi:10.1016/j.jhydrol.2016.07.043, 2016.

Menzel, L., Thieken, A. H., Schwandt, D., and Bürger, G.: Impact of climate change on the regional hydrology - Scenario-based modelling studies in the German Rhine catchment, Natural Hazards, 38, 45–61, doi:10.1007/s11069-005-8599-z, 2006.

Merz, R. and Blöschl, G.: Regionalisation of catchment model parameters, Journal of Hydrology, 287, 95–123, doi:10.1016/j.jhydrol.2003.09.028, 2004.

Merz, R., Parajka, J., and Blöschl, G.: Time stability of catchment model parameters: Implications for climate impact analyses, Water Resources Research, 47, W02 531, doi:10.1029/2010WR009505, 2011.

Mitchell, K. E.: The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system, Journal of Geophysical Research, 109, D07S90, doi:10.1029/2003JD003823, 2004.

Mu, Q., Heinsch, F. A., Zhao, M., and Running, S. W.: Development of a global evapotranspiration algorithm based on MODIS and global meteorology data, Remote Sensing of Environment, 111, 519–536, doi:10.1016/j.rse.2007.04.015, 2007.

Mu, Q., Zhao, M., and Running, S. W.: Improvements to a MODIS global terrestrial evapotranspiration algorithm, Remote Sensing of Environment, 115, 1781–1800, doi:10.1016/j.rse.2011.02.019, 2011.

Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I — A discussion of principles, Journal of Hydrology, 10, 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.

Neumann, J. and Wycisk, P.: Mean annual ground water recharge, in: Hydrologischer Atlas von Deutschland (HAD), edited by Federal Institute for Geosciences and Natural Resources (BGR), p. 5.5, Federal Ministry for the Environment Nature Conservation Building and Nuclear Safety, Freiburg i. Br., 2003.

Newman, A. J., Clark, M. P., Craig, J., Nijssen, B., Wood, A., Gutmann, E., Mizukami, N., Brekke, L., and Arnold, J. R.: Gridded Ensemble Precipitation and Temperature Estimates for the Contiguous United States, Journal of Hydrometeorology, 16, 2481–2500, doi:10.1175/JHM-D-15-0026.1, 2015a.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T., and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance, Hydrology and Earth System Sciences, 19, 209–223, doi:10.5194/hess-19-209-2015, 2015b.

Nijssen, B., Schnur, R., and Lettenmaier, D. P.: Global Retrospective Estimation of Soil Moisture Using the Variable Infiltration Capacity Land Surface Model, 1980–93, Journal of Climate, 14, 1790–1808, doi:10.1175/1520-0442(2001)014<1790:GREOSM>2.0.CO;2, 2001.

Ol'dekop, E. M.: On evaporation from the surface of river basins, Transactions on Meteorological Observations, 4, 200, 1911.

Perrin, C., Andréassian, V., Rojas Serna, C., Mathevet, T., and Le Moine, N.: Discrete parameterization of hydrological models: Evaluating the use of parameter sets libraries over 900 catchments, Water Resources Research, 44, W08447, doi:10.1029/2007WR006579, 2008.

Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., Attinger, S., Schäfer, D., Schrön, M., and Samaniego, L.: Multiscale and Multivariate Evaluation of Water Fluxes and States over European River Basins, Journal of Hydrometeorology, 17, 287–307, doi:10.1175/JHM-D-15-0054.1, 2016.

Rauthe, M., Steiner, H., Riediger, U., Mazurkiewicz, A., and Gratzki, A.: A Central European precipitation climatology – Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS), Meteorologische Zeitschrift, 22, 235–256, doi:10.1127/0941-2948/2013/0436, 2013.

Saha, S., Moorthi, S., Pan, H. L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y. T., Chuang, H. Y., Juang, H. M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., Van Den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J. K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C. Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R. W., Rutledge, G., and Goldberg, M.: The NCEP climate forecast system reanalysis, Bulletin of the American Meteorological Society, 91, 1015–1057, doi:10.1175/2010BAMS3001.1, 2010.

Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resources Research, 46, W05523, doi:10.1029/2008WR007327, 2010.

Samaniego, L., Kumar, R., and Zink, M.: Implications of Parameter Uncertainty on Soil Moisture Drought Analysis in Germany, Journal of Hydrometeorology, 14, 47–68, doi:10.1175/JHM-D-12-075.1, 2013.

Samaniego, L., Kumar, R., Breuer, L., Chamorro, A., Flörke, M., Pechlivanidis, I. G., Schäfer, D., Shah, H., Vetter, T., Wortmann, M., and Zeng, X.: Propagation of forcing and model uncertainties on to hydrological drought characteristics in a multi-model century-long experiment in large river basins, Climatic Change, pp. 1–15, 2016.

Schreiber, P.: Über die Beziehungen zwischen dem Niederschlag und der Wasserführung der Flüsse in Mitteleuropa, Zeitschrift für Meteorologie, 21, 441–452, 1904.

Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling, Journal of Climate, 19, 3088–3111, doi:10.1175/JCLI3790.1, 2006.

Sheffield, J. and Wood, E. F.: Characteristics of global and regional drought, 1950–2000: Analysis of soil moisture data from offline simulation of the terrestrial hydrologic cycle, Journal of Geophysical Research, 112, D17 115, doi:10.1029/2006JD008288, 2007.

5 ~~Sheffield, J. , Goteti, G., and Wood, E. F.~~

Strasser, U. and Mauser, W.: ~~Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling~~Modelling the spatial and temporal variations of the water balance for the Weser catchment 1965–1994, Journal of ~~Climate, 19, 3088–3111, , 2006.~~Hydrology, 254, 199–214, doi:10.1016/S0022-1694(01)00492-9, 2001.

Thober, S., Mai, J., Zink, M., and Samaniego, L.: Stochastic temporal disaggregation of monthly precipitation for regional gridded data sets,
10 Water Resources Research, 50, 8714–8735, doi:10.1002/2014WR015930, 2014.

Thober, S., Kumar, R., Sheffield, J., Mai, J., Schäfer, D., and Samaniego, L.: Seasonal Soil Moisture Drought Prediction over Europe Using the North American Multi-Model Ensemble (NMME), Journal of Hydrometeorology, 16, 2329–2344, doi:10.1175/JHM-D-15-0053.1, 2015.

Tolson, B. A. and Shoemaker, C. A.: Dynamically dimensioned search algorithm for computationally efficient watershed model calibration,
15 Water Resources Research, 43, 1–16, doi:10.1029/2005WR004723, 2007.

Troy, T. J., Wood, E. F., and Sheffield, J.: An efficient calibration method for continental-scale land surface modeling, Water Resources Research, 44, W09 411, doi:10.1029/2007WR006513, 2008.

Vereecken, H., Huisman, J. A., Bogena, H., Vanderborght, J., Vrugt, J. A., and Hopmans, J. W.: On the value of soil moisture measurements in vadose zone hydrology: A review, Water Resources Research, 44, 1–21, doi:10.1029/2008WR006829, 2008.

20 Vogt, J., Soille, P., de Jager, A., Rimaviciute, E., Mehl, W., Foisneau, S., Bodis, K., Dusart, J., Paracchini, M.-L., Haastrup, P., and Bamps, C.: A pan-European river and catchment database, European Commission - JRC, Luxembourg, 2007.

Wood, A. W. and Lettenmaier, D. P.: An ensemble approach for attribution of hydrologic prediction uncertainty, Geophysical Research Letters, 35, L14 401, doi:10.1029/2008GL034648, , 2008.

Wood, A. W., Leung, L. R., Sridhar, V., and Lettenmaier, D. P.: Hydrologic Implications of Dynamical and Statistical Approaches to Down-
25 scaling Climate Model Outputs, Climatic Change, 62, 189–216, doi:10.1023/B:CLIM.0000013685.99609.9e, 2004.

Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., de Roo, A., Döll, P., Ek, M., Famiglietti, J., Gochis, D., van de Giesen, N., Houser, P., Jaffé, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., Sheffield, J., Wade, A., and Whitehead, P.: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, Water Resources Research, 47, 1–10, doi:10.1029/2010WR010090, 2011.

30 Xia, Y., Mitchell, K., Ek, M., Cosgrove, B., Sheffield, J., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Duan, Q., and Lohmann, D.: Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow, Journal of Geophysical Research: Atmospheres, 117, n/a–n/a, doi:10.1029/2011JD016051, 2012a.

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Lettenmaier, D.,
35 Koren, V., Duan, Q., Mo, K., Fan, Y., and Mocko, D.: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products, Journal of Geophysical Research: Atmospheres, 117, D03 109, doi:10.1029/2011JD016048, 2012b.

Xia, Y., Hobbins, M. T., Mu, Q., and Ek, M. B.: Evaluation of NLDAS-2 evapotranspiration against tower flux site observations, Hydrological Processes, 29, 1757–1771, doi:10.1002/hyp.10299, 2015.

Zappa, M., Jaun, S., Germann, U., Walser, A., and Fundel, F.: Superposition of three sources of uncertainties in operational flood forecasting chains, Atmospheric Research, 100, 246–262, doi:10.1016/j.atmosres.2010.12.005, 2011.

Zhang, X.-J., Tang, Q., Pan, M., and Tang, Y.: A Long-Term Land Surface Hydrologic Fluxes and States Dataset for China, Journal of Hydrometeorology, 15, 2067–2084, doi:10.1175/JHM-D-13-0170.1, 2014.

Zhu, C. and Lettenmaier, D. P.: Long-Term Climate and Derived Surface Hydrology and Energy Flux Data for Mexico: 1925–2004, Journal of Climate, 20, 1936–1946, doi:10.1175/JCLI4086.1, 2007.

Zink, M., Samaniego, L., Kumar, R., Thober, S., Mai, J., Schäfer, D., and Marx, A.: The German drought monitor, Environmental Research Letters, 11, 074 002, doi:10.1088/1748-9326/11/7/074002, 2016.

Figure 1. Study area showing the seven ~~catchments~~ basins used for estimation of ~~common~~ the ensemble parameter sets for Germany. The different colors are making the basins better distinguishable. The points E1-E7 denote eddy covariance stations which are used for the evaluation of evapotranspiration and soil moisture.
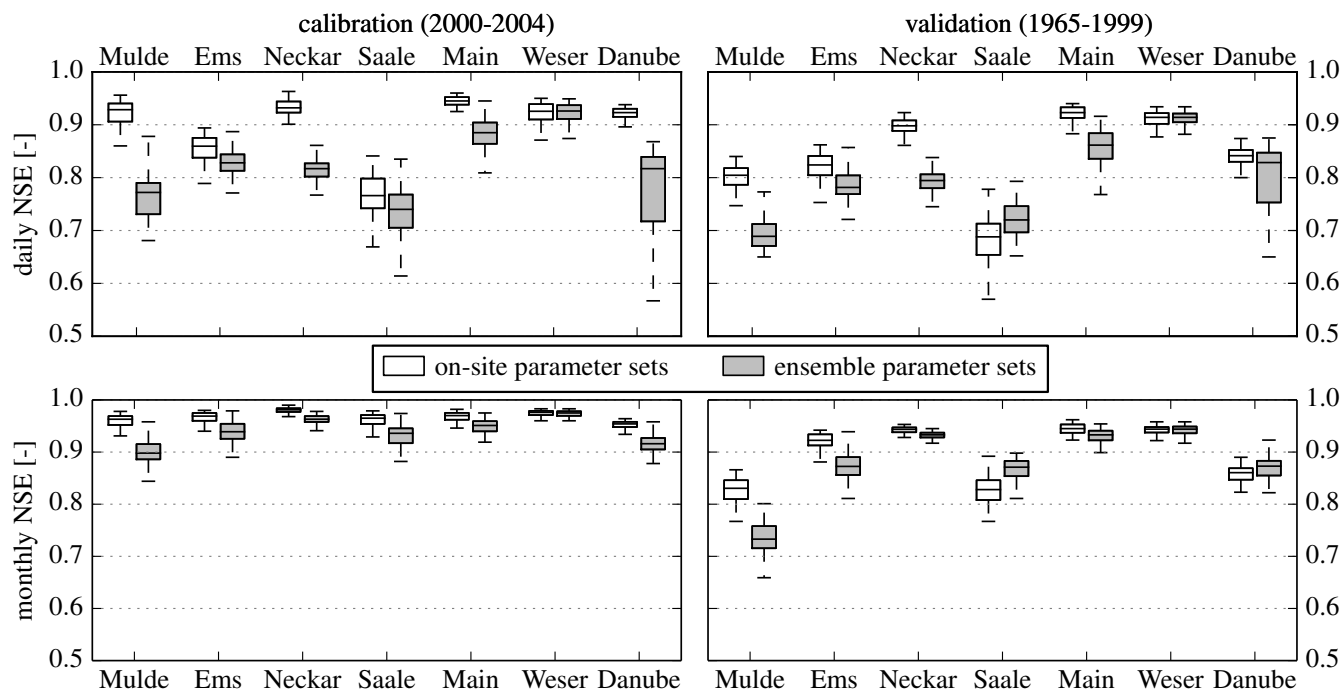
Figure 2. Model performance expressed as Nash Sutcliffe Efficiency (NSE) at daily (upper row) and monthly (lower row) resolutions for the calibration period 2000-2004 (left-hand side) and validation period 1965-1999 (right-hand side). The white box plots show the results of the on-site calibration, whereas the gray box plots are simulations using the 100 ensemble parameter sets for Germany. Please note that the y-axis starts at NSE=0.5

Figure 3. Observed and modeled monthly ~~discharge~~ streamflow for the seven ~~catchments~~basins, which were used for parameter inference. The figure shows one decade (1990-1999) of the evaluation period. The solid dark gray line depicts the median model results and the light gray band depicts the range between the 5th and 95th percentile of the 100 ensemble simulations.

Figure 4. Budyko plot and performance maps for 100 ensemble parameter sets at 222 ~~catchments~~ basins spread over Germany. The upper row depicts evaluations based on daily values (panels A, B, and C), whereas the lower row depicts monthly ~~discharge~~ streamflow evaluations (panels D, E, and F). In the first column the ~~catchments~~ basins are presented as Budyko plots (panels A and D), which are color-coded based on the ensemble median NSE for daily (panel A) and monthly (panel ~~B~~D) ~~discharge~~ streamflow values. The gray band envelops different estimations of the Budyko curve (Schreiber, 1904; Ol'dekop, 1911; Budyko, 1974). A separation to energy- ($E_p/P < 1$) and water-limited basins ($E_p/P > 1$) can be made based on the x-axis. The center column depicts the location of the 222 ~~catchments~~ basins shown in the Bydyko plots using the same color code (panels B and E). The right column shows the range of the 5th and 95th ensemble percentiles for the NSE on daily (panel C) and monthly (panel F) basis. Panels A, B, D, and E share the left color bar, and panels C and F share the right color bar. The simulation period is adopted according to the available ~~discharge~~ streamflow observations but is at least 10 years (average=42 years).
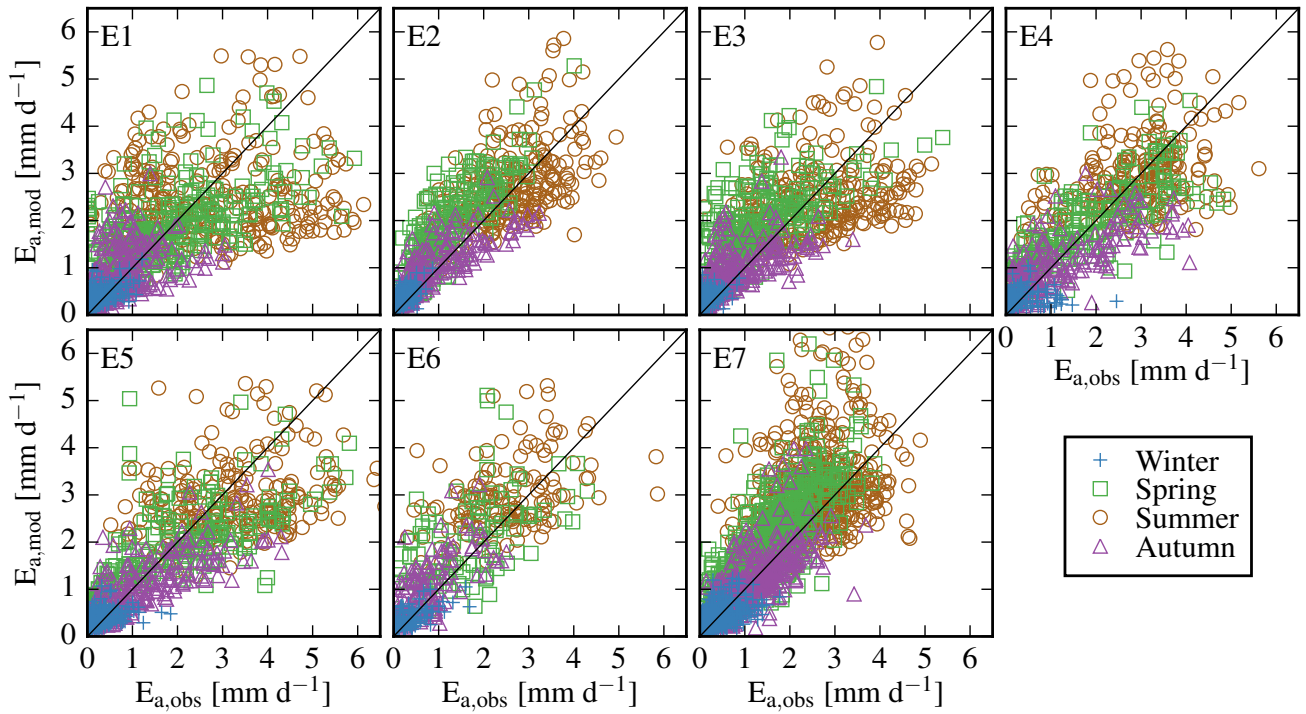
Figure 5. Observed ($E_{a,obs}$) versus ensemble median modeled evapotranspiration ($E_{a,mod}$) on daily basis at the seven eddy covariance stations (Figure 1, Table 2).
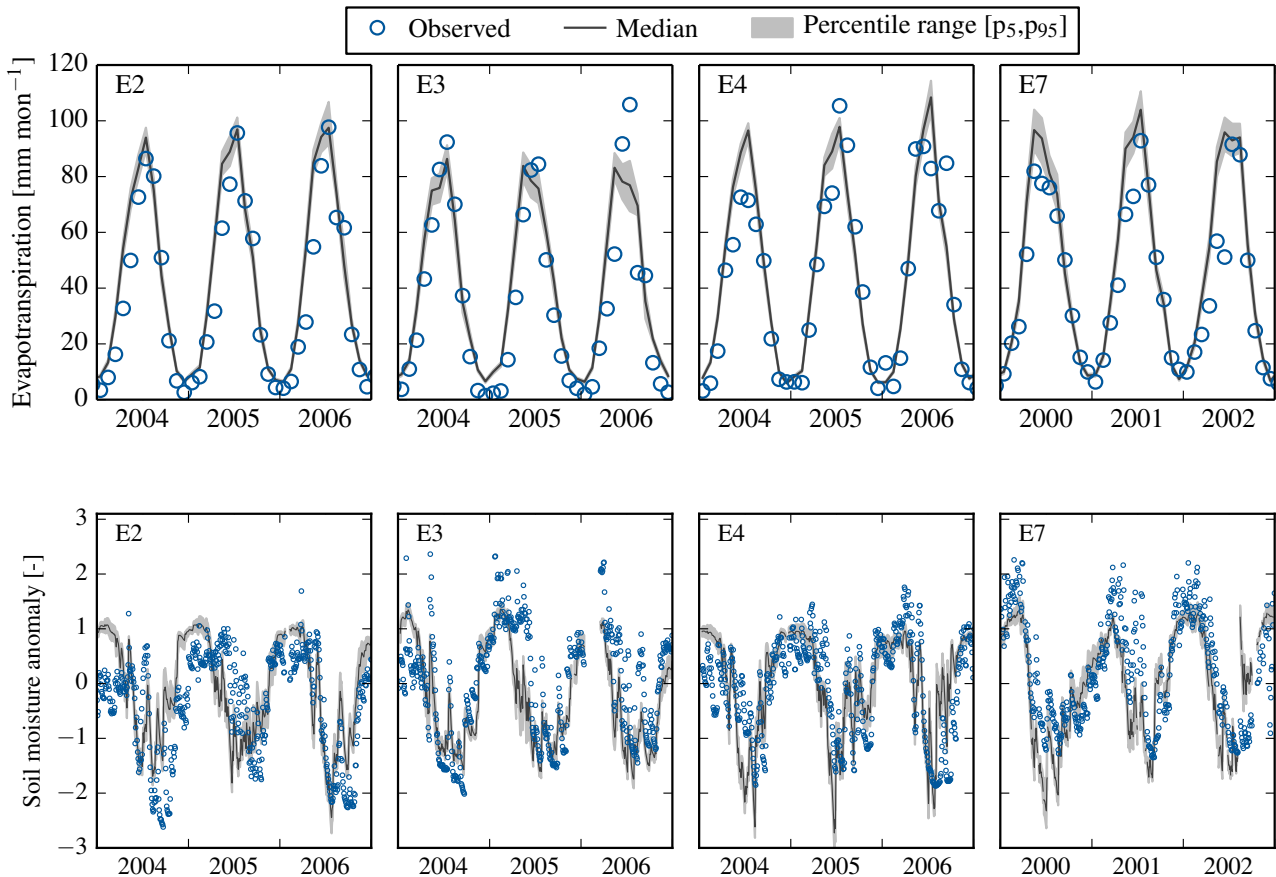
31

Figure 6. Exemplary time series of observed and modeled monthly evapotranspiration and daily soil moisture anomalies at four eddy covariance stations (Figure 1, Table 2). The four stations are chosen because they represent the two major mHM land cover classes (forest and mixed) and have three consecutive years of data without significant data gaps. Further the four station are spread over the three regions where eddy covariance observations are available. The solid dark gray line depicts the median model results and the light gray band depicts the range between the 5th and 95th percentile of the 100 ensemble simulations.
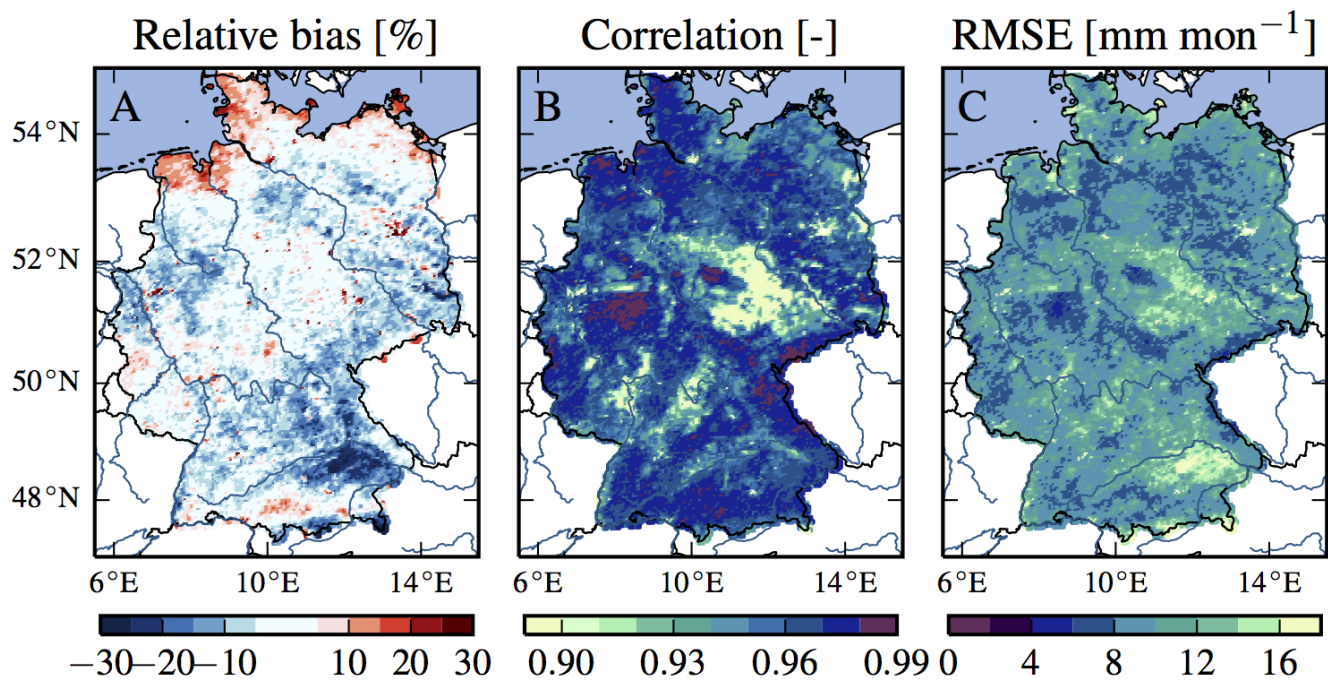
Figure 7. Comparison of monthly estimates of evapotranspiration from mHM and MODIS in the period 2001-2010. The ensemble is represented by the ensemble mean of 100 evapotranspiration estimates. The comparison is based on three metrics: A) relative bias, B) Pearson correlation coefficient, and C) root mean squared error (RMSE). The respective units are given in brackets.
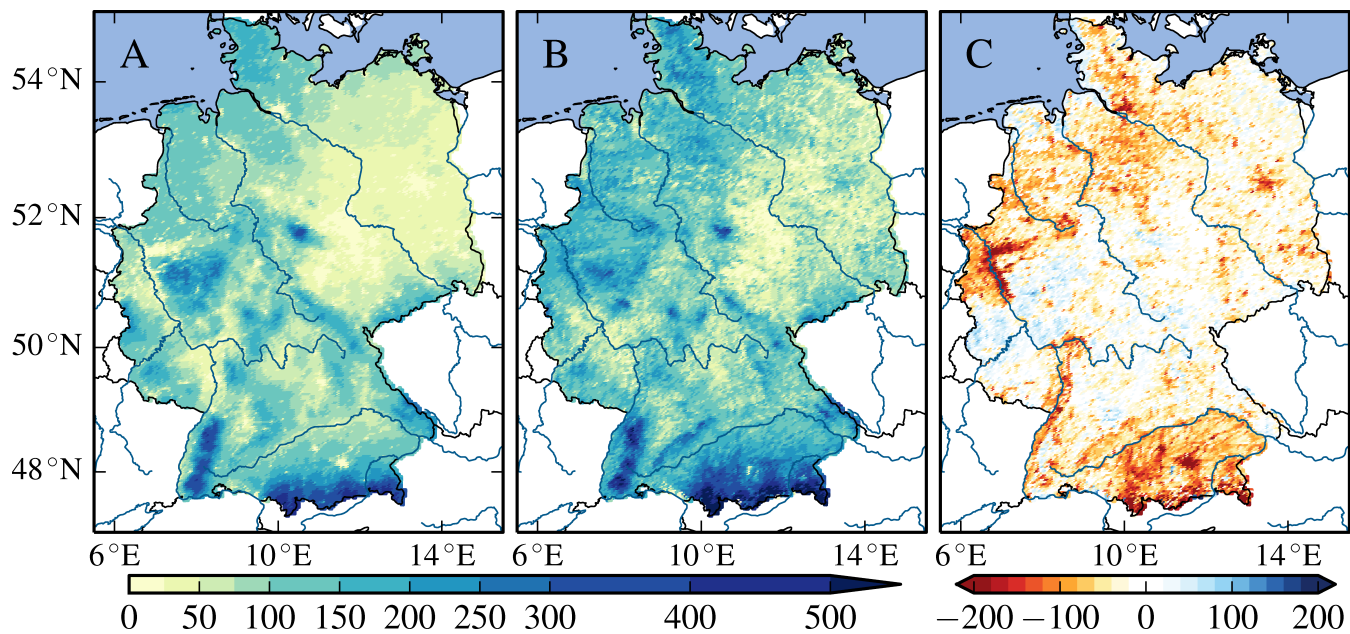
Figure 8. Comparison of mean annual groundwater recharge (R) modeled with A) mHM and from B) the Hydrologic Atlas of Germany (Federal Ministry for the Environment Nature Conservation Building and Nuclear Safety, 2003; Neumann and Wycisk, 2003). Panel C shows the difference (A-B) between the two data sets. The units are [mm a^{-1}] for all panels.
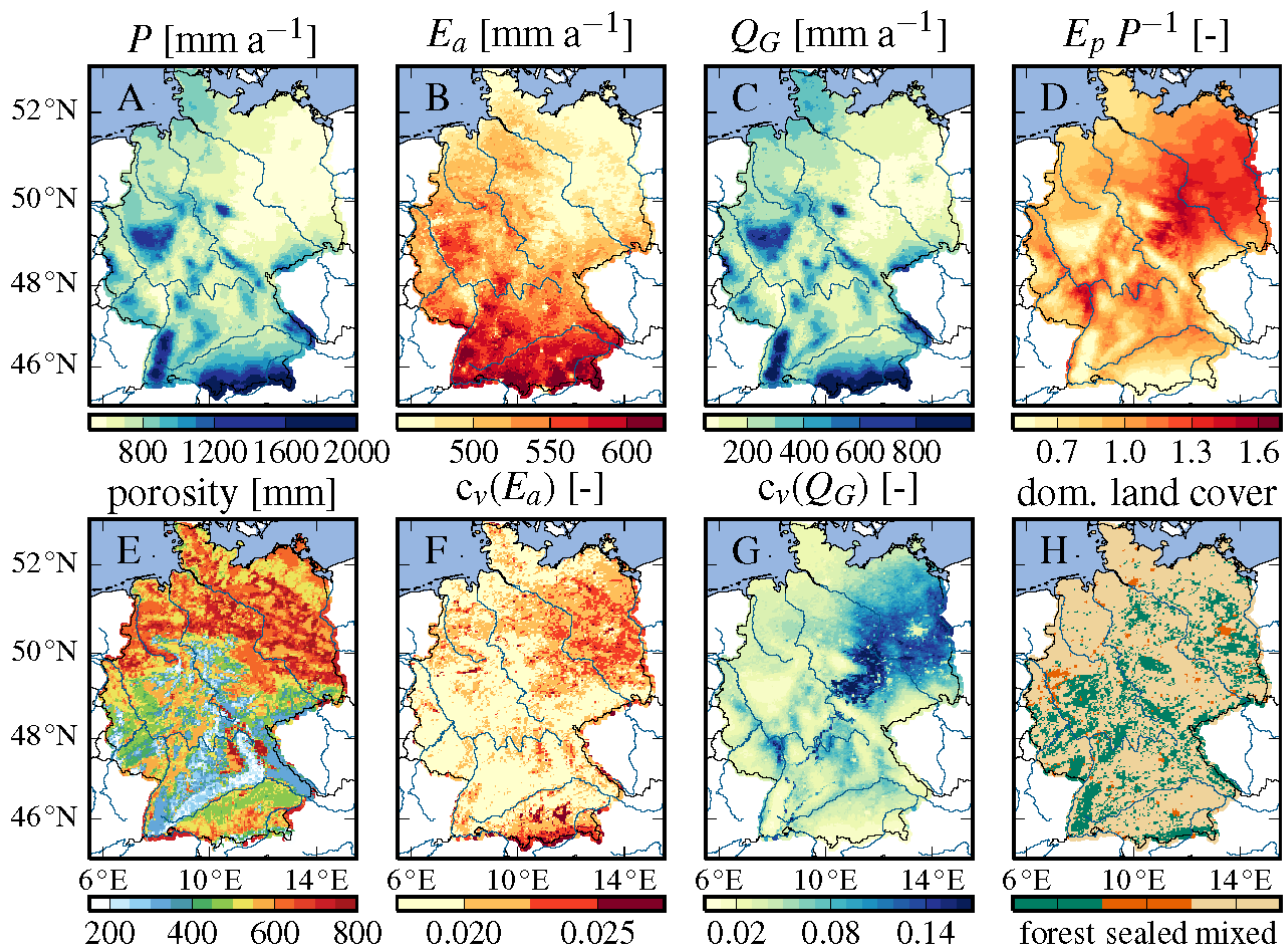
Figure 9. Water balance variables, their coefficients of variation and land surface characteristics for Germany. A) Mean annual precipitation P, B) ensemble mean annual evapotranspiration E_a, C) and grid-cell-generated runoff Q_G, D) dryness index E_p/P, E) sum of porosities (saturated soil water content) of all model layers, F) coefficient of variation ~~from~~ of the ensemble of annual evapotranspiration and G) ~~discharge~~ generated runoff, H) dominating land cover class on a $4{\times}4$ km^2 grid. The mean values and coefficients of variation are based on the period 1951-2010.
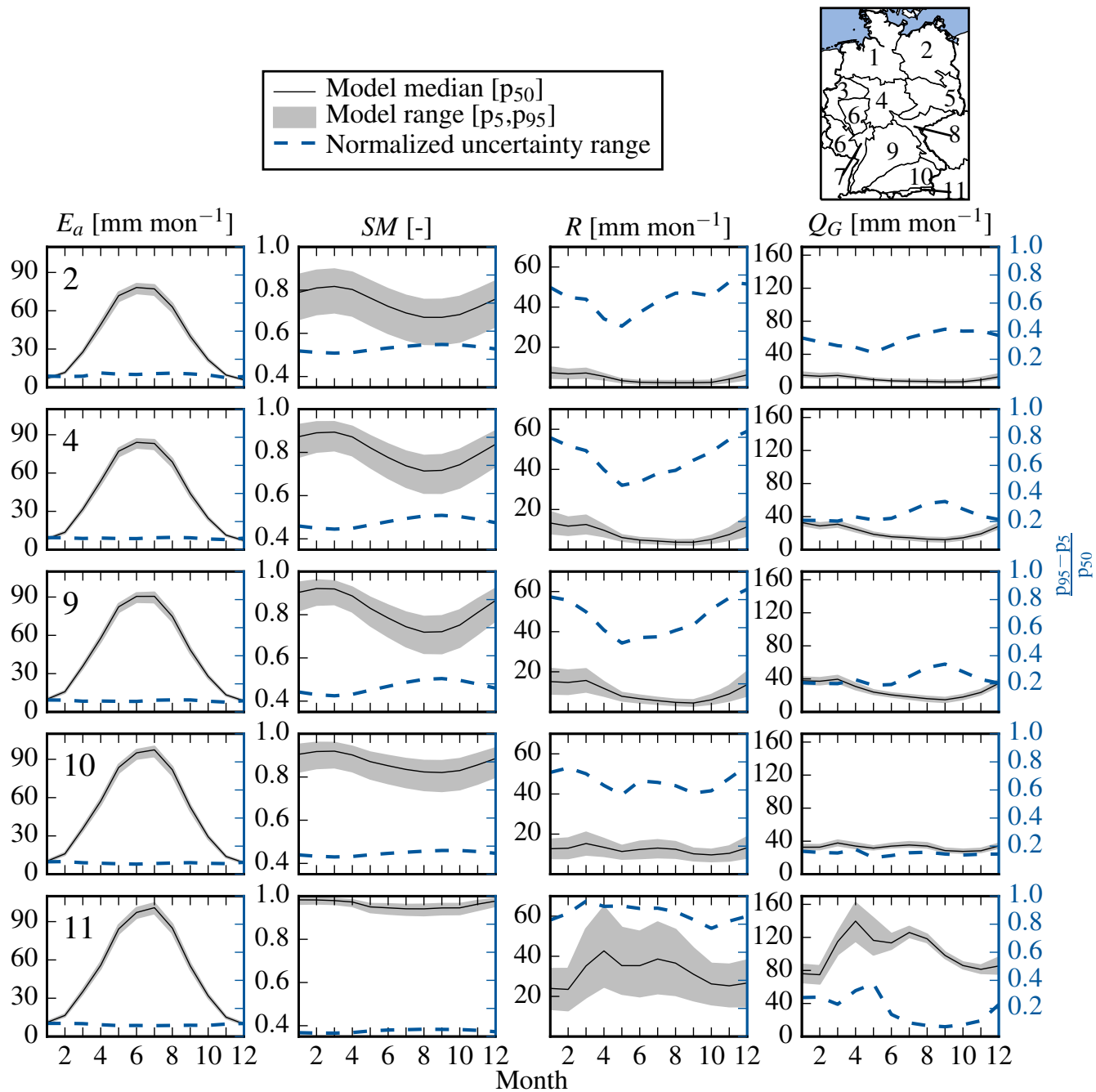
Figure 10. Spatio-temporal patterns of uncertainty for five different environmental zones in Germany. The locations of the different zones are depicted on the map on the upper right. The presented hydrologic variables are evapotranspiration (E_a), soil moisture (SM), recharge (R), and grid-cell-generated runoff (Q_G). The uncertainty ranges and the ensemble median refer to the left ordinate (black and grey), whereas the normalized uncertainty range refers to the right ordinate (blue). The reference period for the climatological values is 1951-2010.
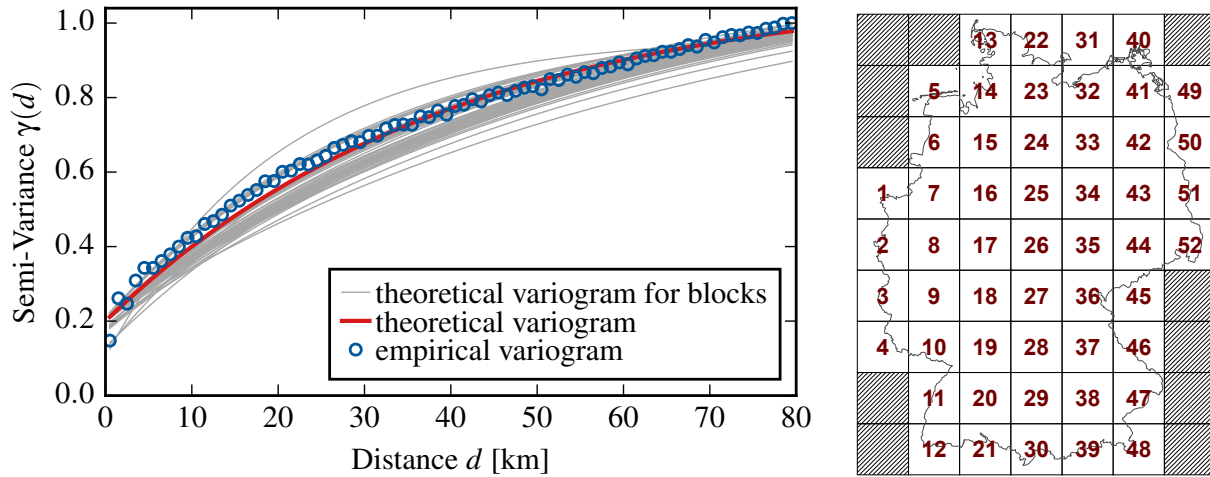
Figure A1. The panel on the left shows the empirical variogram (blue circles) and a fitted exponential variogram (red curve) for the entire domain of Germany as well as fittings for sub-domain (block) variaograms (grey lines). The 52 sub-domains (blocks) are depicted in the panel on the right.
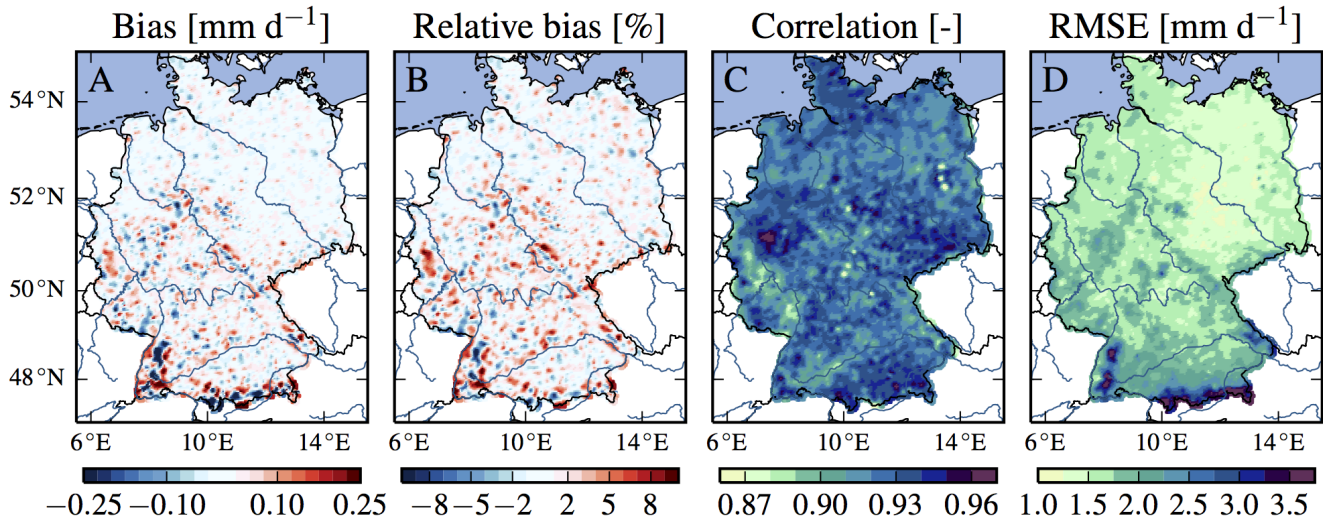

Figure A2. Evaluation of the interpolation at precipitation stations based on a leave one out cross-validation strategy, i.e., the Jacknife method. The performance criteria from the individual stations are interpolated to a 4×4 km^2 grid using ordinary kriging. The panels denote different performance metrics: A) bias, B) relative bias, C) Pearson correlation coefficient, and D) root mean squared error (RMSE).
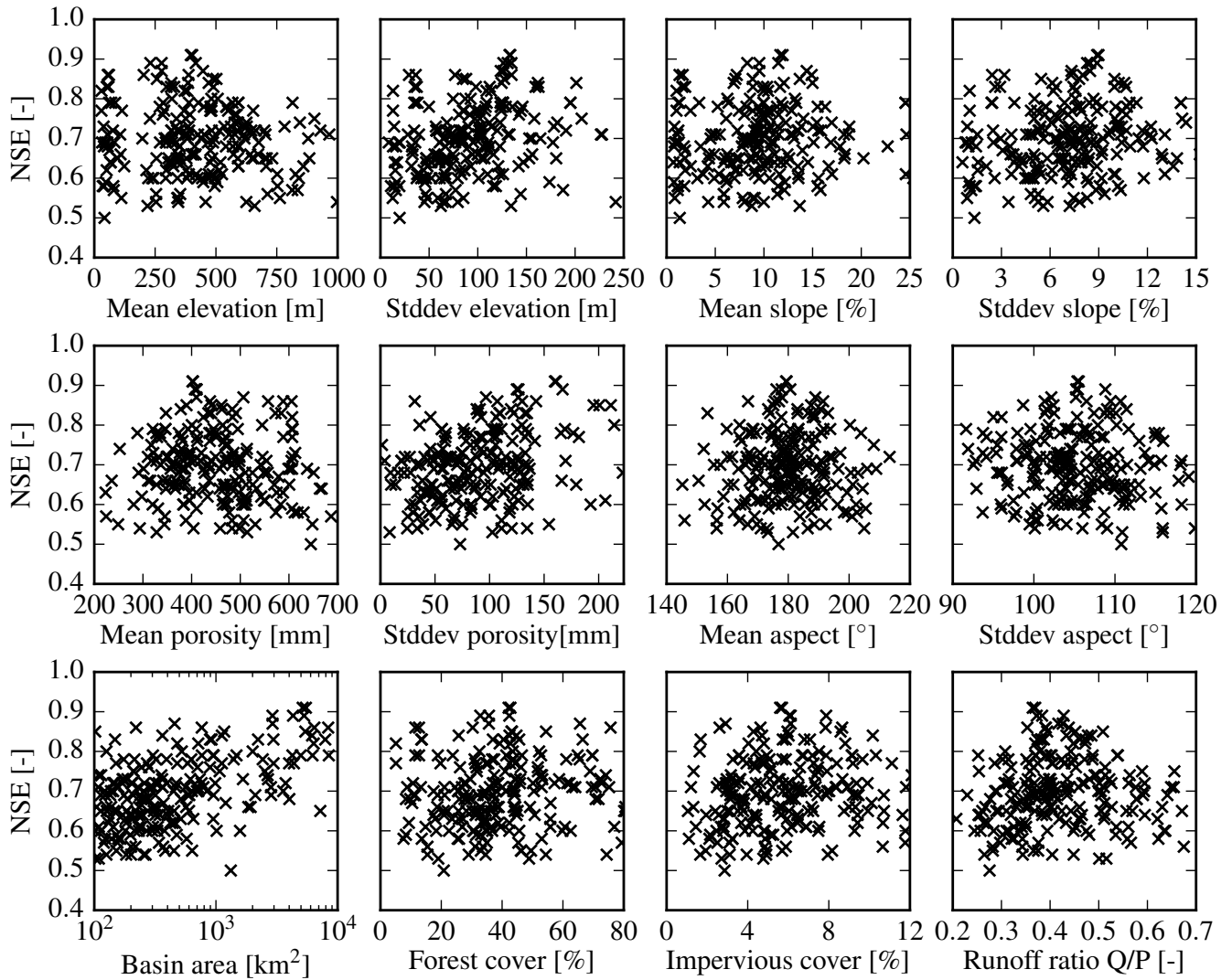
Figure B1. Relation between land surface and hydro-climatic conditions and model performance for the 222 river basins. The location of the basins is depicted in Figure 4. The mean and standard deviation (stddev) of a characteristic for the single basins are based on the morphological input data at the 100×100 m^2 resolution.

Table 1. ~~Catchment~~ Basin properties and water balance characteristics of the seven major German river basins. The geographical location of the ~~catchments~~ basins is depicted in Figure 1. Abbreviations: avg - average, std - standard deviation, min - minimum, max - maximum, P - precipitation, Q - streamflow, E_a - evapotranspiration (P-Q), E_p - potential evapotranspiration

Major Basins	~~catchment~~ Basin ~~area~~ Area [km^2]	Elevation [m]				Land Cover [%]			Water Balance [mm a^{-1}]			~~dryness~~ Dryness ~~index~~ Index [-]	~~runoff~~ Runoff ~~coeff~~ Coeff. [-]
		avg	std	min	max	forest	sealed	mixed	P	Q	E_a	E_p/P	Q/P
Mulde	6 200	386	201	75	1 212	26	10	64	798	344	454	0.88	0.43
Ems	8 400	60	36	10	383	13	8	79	802	312	490	0.89	0.39
Neckar	12 700	445	153	124	1 015	35	10	55	914	356	558	0.85	0.39
Main	23 700	356	113	93	1 044	39	6	55	793	247	546	0.97	0.31
Saale	24 800	287	162	56	1 139	23	8	69	645	161	484	1.13	0.25
Weser	37 700	223	165	8	1 116	34	7	59	781	276	505	0.91	0.35
Danube	47 500	558	170	302	2 329	32	6	62	948	469	479	0.80	0.49

Table 2. Evaluation of evapotranspiration E_a and soil moisture SM at seven eddy covariance stations. The evaluation is based on daily and monthly values for the available observation period. The location of the eddy stations is depicted in Figure 1. Abbreviations: RMSE - root mean squared error, ρ - Pearson correlation coefficient, E_a - evapotranspiration, SM - soil moisture

ID	Station Name	Period	Land Cover	Monthly E_a [mm mon^{-1}]		[-]	Daily E_a [mm d^{-1}]		[-]	~~daily~~ Daily SM [-]
				RMSE	~~BIAS~~ Bias	ρ	RMSE	~~BIAS~~ Bias	ρ	ρ
E1	Gebesee	2003-2008	cropland	19.14	0.61	0.85	1.01	0.02	0.67	0.62
E2	Hainich	2000-2007	DBF*	11.72	6.99	0.95	0.62	0.23	0.87	0.68
E3	Mehrstedt	2003-2006	grasland	12.44	5.78	0.94	0.74	0.18	0.79	0.80
E4	Wetzstein	2004-2008	ENF**	9.86	1.58	0.96	0.73	0.05	0.84	0.80
E5	Grillenburg	2004-2008	grasland	13.93	-4.19	0.94	0.89	-0.14	0.8	0.93
E6	Klingenberg	2004-2008	cropland	15.39	9.38	0.93	0.86	0.31	0.77	0.53
E7	Tharandt	1997-2008	ENF**	13.39	7.71	0.96	0.72	0.26	0.83	0.82

* deciduous broadleaf forest, ** evergreen needleleaf forest