Presentation and discussion of the high resolution atmosphere-land surface-subsurface simulation dataset of the simulated Neckar catchment for the period 2007-2015

Bernd Schalge[1], Gabriele Baroni[2], Barbara Haese[3], Daniel Erdal[4], Gernot Geppert[5], Pablo Saavedra[1], Vincent Haefliger[1], Harry Vereecken[6,7], Sabine Attinger[8,9], Harald Kunstmann[3,10], Olaf A. Cirpka[4], Felix Ament[11], Stefan Kollet[6,7], Insa Neuweiler[12], Harrie-Jan Hendricks Franssen[6,7], Clemens Simmer[1,7]

[1]Institute for Geosciences, University of Bonn, Bonn, Germany

[2] Department of Agricultural and Food Sciences, University of Bologna, Bologna, Italy

[3]Institute of Geography, University of Augsburg, Augsburg, Germany

[4]Center for Applied Geoscience, University of Tübingen, Tübingen, Germany

[5]Department of Meteorology, University of Reading, Reading, England

[6]Forschungszentrum Jülich GmbH, Agrosphere (IBG-3), Jülich, Germany

[7]Centre for High-Performance Scientific Computing (HPSC-TerrSys), Geoverbund ABC/J, Jülich, Germany

[8]Institute of Earth and Environmental Science, University of Potsdam, Potsdam, Germany

[9]Helmholtz-Center for Environmental Research, Leipzig, Germany

[10]Institute of Meteorology and Climate Research (IMK-IFU), Karlsruhe Institute of Technology (KIT), Garmish-Partenkirchen, Germany

[11]Meteorological Institute, University of Hamburg, Hamburg, Germany

[12]Hannover, Institut für Strömungsmechanik und Umweltphysik im Bauwesen, Leibniz Universität Hannover, Germany

Correspondence to: Bernd Schalge (bschalge@uni-bonn.de)

Abstract.

Coupled numerical models, which simulate water and energy fluxes in the subsurface-land surface-atmosphere system in a physically consistent way are a prerequisite for the analysis and a better understanding of heat and matter exchange fluxes at compartmental boundaries and interdependencies of states across these boundaries. Complete state evolutions generated by such models may be regarded as a proxy of the real world, provided they are run at sufficiently high resolution and incorporate the most important processes. Such a simulated reality can be used to test hypotheses on the functioning of the coupled terrestrial system. Coupled simulation systems, however, face severe problems caused by the vastly different scales of the processes acting in and between the compartments of the terrestrial system, which also hinders comprehensive tests of their realism. We used the Terrestrial Systems Modeling Platform TerrSysMP, which couples the meteorological model COSMO, the land-surface model CLM, and the subsurface model ParFlow, to generate a simulated catchment for a regional

terrestrial system mimicking the Neckar catchment in southwest Germany, the virtual Neckar catchment. Simulations for this catchment are made for the period 2007-2015, and at a spatial resolution of 400m for the land surface and subsurface and 1.1km for the atmosphere. Among a discussion of modelling challenges, the model performance is evaluated based on observations covering several variables of the water cycle. We find that the simulated catchment behaves in many aspects quite close to observations of the real Neckar catchment, e.g. concerning atmospheric boundary-layer height, precipitation, and runoff. But also discrepancies become apparent, both in the ability of the model to correctly simulate some processes which still need improvement such as overland flow, and in the realism of some observation operators like the satellite based soil moisture sensors. The whole raw dataset is available for interested users. The dataset described here is available via the CERA database (Schalge et al, 2020): https://doi.org/10.26050/WDCC/Neckar_VCS_v1

1 Introduction

Earth environmental models are becoming increasingly important for climate and weather prediction, flood forecasting, water resources management, agriculture, and water quality control (e.g. Shrestha et al. 2014; Larsen et al. 2014; Simmer et al., 2015). Assuming that that the models are able to resemble the real-world based on state-of-the-art understanding of the system processes, the models are also used as "virtual realities" for hypothesis testing and decision support systems in many scientific disciplines (Clark et al., 2015, Semenova & Beven, 2015).

Virtual or simulated realities have been used for specific compartments of the terrestrial system in many studies (see Fatichi et al., 2016, and reference herein) and several advantages have been recognized. Bashford et al. (2002) computed simulated remote-sensing observations with 1 km resolution to derive, among others, process parameterizations for evapotranspiration in a hydrological model operating on the same scale as the remote sensing data. Weiler and McDonnel (2004) used a simulated-reality approach on the hill-slope scale to detect and quantify the major controls on subsurface flow processes and derive tunable parameters for conceptual models. Similar experiments allowed Schlueter et al. (2012) to explore the relationship between soil architecture and hydraulic behavior and Chaney et al. (2015) to testing sampling designs. Hein et al. (2019) explores the relative importance of different factors in the hydrologic response of a catchment. Simulated realities are also often used to overcome limitations on the data-scarce observations. In this context, Ajami and Sharma (2018) used simulations results to test disaggregation method for soil moisture observations. In subsurface hydrology it is a standard procedure to test inverse modeling and data-assimilation approaches on simulated aquifers (e.g., Zimmermann et al., 1998; Hendricks Franssen et al., 2009), which are used to generate realistic aquifer data with exactly known hydraulic and geochemical properties at every point (e.g., Schaefer et al., 2002).

More recently, it has been highlighted that the terrestrial systems should be better exploited by the use of integrated models which are able to simulate water and energy fluxes in the subsurface-land surface-atmosphere system in a physically consistent way (Clark et al., 2015, Davison et al, 2018). For this reason, these integrated modeling approaches have also

been considered to generate simulated realities (Mackay et al., 2015). However, despite the increasing computational capability and availability of infrastructures, these modelling approaches are generally more technically demanding. In addition, the use of these types of integrated models requires different expertises that are not usually cover within one single scientific group but requires strong interdisciplinary collaborations among different partners. For these reasons, the use of these types of models is still not commonly foreseen.

To overcome this limitation, in this paper we present the development, the testing and the data of a simulated reality of a mesoscale catchment based on a fully integrated terrestrial model system. Our virtual Neckar Catchment encompasses the terrestrial system from the bedrock to the upper atmosphere covering the catchment of a higher-order river (length ≈ 380km, area ≈ 14000km^2) including a buffer zone surrounding it, in which we simulate - as realistically as currently possible - the multi-year evolution of states including the water and energy fluxes in and between all its compartments. We specifically venture to represent the strong spatial variability of the land components, which affects the overall system behavior due to nonlinear couplings and feedbacks. Since a simulated catchment with no resemblance to a real world catchment hardly allows for evaluating its realism, we base our simulation loosely on the Neckar catchment in southwest Germany that contains quite variable topography, different land cover, high and low precipitation regions, deep and shallow water tables and regions prone to flooding events. (see Figure 1). The model does not aim at exactly reproducing the catchment's response to hydro-climatic forcing, instead we only require that the simulated response is realistic with respect to typical spatial and temporal characteristics. For this reason, we discuss the model realism in comparison with observations of the real catchment, but also its limitations, particularly in relation to the chosen resolutions which balance the detail in process representation and computational feasibility.Despite these simplifications we believe this dataset will be useful in a variety of ways, such as data assimilation, model comparison studies and model development studies as well as focused impact studies. In the discussion section at the end we go more into detail how this dataset can potentially be used and what the limits of applicability are.

The remainder of the paper is structured as follows. In section 2, we introduce the simulation platform TerrSysMP, while Section 3 describes in detail the surface and subsurface parameters for topography, soils and aquifers, land use, vegetation, and the river network. In Section 4, we show snapshots and time series of state variables or system parameters extracted from the simulated catchment and compare them to observations in the real Neckar catchment to demonstrate how well the most important requirements are met. These results as well as possible ways to improve them are discussed in Section 5 together with several issues, which came up during the development phase. We provide conclusions and an outlook in Section 6.

2 The Terrestrial Systems Modeling Platform (TerrSysMP)

We used the Terrestrial System Modeling Platform (TerrSysMP, see Shrestha et al. 2014; Gasper et al. 2014; Sulis et al. 2015) developed within the Transregional Collaborative Research Centre TR32 (Simmer et al. 2015) for the generation of

the simulated catchment. TerrSysMP couples (Figure 1 in Gasper et al., 2014) the hydrologic flow model ParFlow v693 (Ashby and Falgout, 1996; Jones and Woodward, 2001; Kollet and Maxwell, 2006), the land-surface model Community Land Model, CLM v3.5 (Oleson et al., 2008), and the atmospheric model Consortium for Small Scale Modeling (COSMO v4.21, Baldauf et al., 2011) via the Ocean Atmosphere Sea Ice Coupling framework OASIS3 (e.g. Valcke et al., 2006), using a dynamical two-way approach including down- and upscaling algorithms for fluxes and state variables between computational grids of different resolution.

ParFlow is a variably saturated watershed flow model, which solves the three-dimensional Richards equation to model saturated and unsaturated flow in the subsurface, and the fully integrated kinematic wave equation to model two-dimensional overland flow. Also other global and regional hydrological models use the latter to route overland flow, e.g. MODCOU (Haefliger et al., 2015) and TRIP (Alkama et al., 2012). Advanced Newton-Krylov multigrid solvers are used that are especially suitable for massively parallel computer environments. Excellent model performance and parallel efficiency have been documented by Jones and Woodward (2001), Kollet and Maxwell (2006), and Kollet et al. (2010). A unique feature of ParFlow is the use of an advanced octree data structure for rendering overlapping objects in 3-D space, which facilitates modeling complex geology and heterogeneity as well as the representation of topography based on digital elevation models and watershed boundaries.

CLM is a single column biogeophysical land-surface model released by the National Center for Atmospheric Research (NCAR), which considers coupled snow, soil, and vegetation processes. Land surface heterogeneity is represented as a nested sub-grid hierarchy in which grid cells are composed of multiple land units (glacier, lake, wetland, urban, and vegetation), snow/soil columns (to capture variability in snow and soil states within each land unit), and Plant Functional Types (PFTs) to capture the biogeophysical and biogeochemical differences between broad categories of plants in terms of their functional characteristics. In TerrSysMP, the 1-D Richards-equation model included in CLM is replaced by ParFlow.

COSMO is a limited-area, non-hydrostatic numerical weather prediction model, which operationally runs at the German weather service DWD, among others, for Numerical Weather Prediction (NWP) and various scientific applications on the meso-β and meso-γ scale. COSMO is based on the primitive thermo-hydrodynamical equations describing compressible flow in a moist atmosphere. As a limited-area model, COSMO needs lateral boundary conditions from a driving larger-scale model. We impose the lateral conditions by nesting COSMO in COSMO-DE which spans Germany. At the lateral boundaries a relaxation technique is used in which the internal model solution is nudged against an externally specified solution over a narrow transition zone between the two domains. Version 3.5 of CLM that is used here is already relatively old. Even though version 5 was not yet available when we started our work, it is now and comparison is warranted. Newer versions of CLM have several major improvements over 3.5. The first one is a more sophisticated routing scheme leading to much improved soil moisture profiles. In our case we replace this part with ParFlow anyway so our older version is not a disadvantage in that regard. Other improvements are the inclusion of carbon and nitrogen cycles as well as more options for crop type vegetation. Here we purposely simplify our setup as we not only have and want static land use but also use a blend type of crop with no sharp changes in LAI due to harvests. Instead, we assume harvest to be an ongoing process all

throughout autumn. Thus, all these improvements do not downgrade the simulation results presented and discussed in this study.

Within OASIS3, the upscaling algorithm uses the mosaic or explicit sub-grid approach (Avissar and Pielke, 1989) in which high-resolution land surface fluxes are averaged and transferred to the coarser resolution of the atmospheric model component. The implemented Schomburg scheme (Schomburg et al., 2010, 2012) downscales atmospheric variables of the lowest atmospheric model layer to the higher-resolved land surface model. The scheme involves (i) spline interpolation while conserving mean and lateral gradients of the coarse field, (ii) deterministic downscaling rules to exploit empirical relationships between atmospheric variables and surface variables, and (iii) the addition of high-resolution variability (i.e. noise) in order to honor the non-deterministic part and to restore spatial variability.

TerrSysMP allows simulating the terrestrial water, energy, and biogeochemical cycles from the deeper subsurface including groundwater (ParFlow) across the land-surface (CLM) into the atmosphere (COSMO). Water and energy cycles are coupled via evaporation and plant transpiration; these processes are modeled by CLM with a non-linear coupling to ParFlow through soil-water availability and root-water uptake (Figure 2). The two-way coupling between CLM and COSMO encompasses radiation exchange and turbulent exchanges of moisture, energy, and momentum. OASIS3 allows for different temporal and spatial resolutions of the coupled model components. For example, a temporal resolution of 15 minutes is sufficient for the subsurface and land-surface components, whereas time steps as small as 5 seconds are needed for the atmosphere. A higher spatial resolution can be assigned for the surface and subsurface parts to allow for a better representation of soil and land-use heterogeneity.

Since high-resolution and long time-series of the fully coupled system are needed to satisfy our need to check the statistical behavior of the system, the models were run on the IBM/BlueGeneQ System JUQUEEN at the Jülich Supercomputing Centre (Jülich Supercomputing Centre, 2015). JUQUEEN has a total of 28672 nodes with 16 cores each. Our configuration involved using 256 nodes for 12hours, restarting the simulation every 7 simulation days. This is necessary as the runtime for Parflow can vary greatly depending on the conditions in the catchment. The total number of grid cells for the domain is 323,675 per model layer with 10 layers for CLM and 50 layers for ParFlow, and 58,420 grid-points for the 50 COSMO layers resulting in 22.3 million grid cells. We ran the fully-coupled model for a period of nine years (2007-2015) as 2007 was the first full year where high resolution atmospheric forcings were available and nine years was the maximum possible simulation length given constraints on compute resources. On average the actual runtime was approximately eight hours. This means that for one year of simulation roughly 1.7 million core-hours are needed. For the full nine-year time-series that is about 12 million core-hours; another ~8 million hours were needed for the spin-up. We used an output interval of 15 minutes, which results in a total output of 38.5TB of data for the full time-series, where about half was produced by COSMO and a quarter each by CLM and ParFlow.

3 Description of the virtual Neckar Catchment

Our simulated catchment is based on the Neckar catchment in southwestern Germany (see Figure 1), east of the Black Forest mountain range and north of the Jurassic ridge of the Swabian Alb. The catchment has a varying topography including mountains up to 1050 m a.s.l., river valleys, different land use types, i.e. grassland, cropland (majority of the area), broadleaf and needle leaf forest (see Figure 3), and relatively large soil spatial variability. Annual mean precipitation over the real catchment ranges between 500 and 2000mm (see Section 5.1) with highest values over the Black Forest. Inter-annual variability of precipitation can reach up to one third of the mean value. Monthly precipitation can vary largely and its mean annual cycle is weak with slightly lower values in spring and autumn. While summer precipitation is dominated by convection, winter precipitation is predominantly related to fronts of extra-tropical cyclones with enhanced precipitation over the mountains due to orographic lift. Daily average temperatures vary with altitude between -5°C and 0°C in January and between 13 and 18°C in July. Land use and cover in the lower elevations are dominated by agriculture while the Black Forest features mainly needle-leaf trees. Broad-leaf trees can be found over smaller areas throughout the catchment. The distance to groundwater is in large parts of the area restricted to a few meters, in particular in lowland areas, which assures strong coupling between groundwater table and evapotranspiration (Maxwell et al., 2007). These typical central European catchment features in addition to the relatively shallow groundwater tables (implying a stronger possible feedback of groundwater on atmospheric conditions) were the basis to select the Neckar catchment for our simulation.

The computational domain is a rectangular area of ~57,850km² encompassing the Neckar catchment of ~14,000km². The domain is larger than the Neckar catchment in order to allow the atmospheric model to develop its own internal dynamics. COSMO is run on a 1.1km horizonal grid with 230x254 grid points, which includes a 4 grid point-wide outer frame zone where only the lateral boundary forcing is used without coupling to the CLM, as well as 50 vertical layers in hybrid coordinates (terrain following at the surface, flat in the stratosphere). COSMO is set up identical to the operational COSMO-DE setup of the German national weather service (Deutscher Wetterdienst, DWD), e.g., the deep convection parameterization is switched off because at the chosen grid resolution convection is enabled by the dynamical core (see Section 2.1). In COSMO-DE, the operational resolution is 2.8km, so that the approximation regarding deep convection is even more appropriate in our simulations. Similar choices were taken by Smith et al. (2015), who simulated precipitation events of roughly the same domain using nested WRF models, where the cumulus parameterization was switched off at horizontal resolutions of 900m and 300m. Lateral boundary forcing and constant fields (topography, land-mask etc.) are provided by the COSMO-DE analysis fields, which are downscaled to the 1.1km grid by linear interpolation. The lateral relaxation zone, which moderates the jump from the lateral driving fields to the inner model area, is set to 12km.

A software restriction (unfixable bug specific to the supercomputing system we were using for our simulation runs r as described in the previous section) did not allow for cases with more than 4.2 million CLM columns as the model d id not initialize properly and crashed implying that a higher spatial resolution for CLM and ParFlow than 400 m could not be achieved for the Neckar catchment on the used system. So, ParFlow and CLM use the same horizontal grid with a resolution

of 400 m and 535x605 grid points. The vertical grid for both component models is partially the same, with CLM limited to 10 vertical layers up to a total depth of 3 m shared with ParFlow, which has in total 50 vertical layers reaching down to 100m. COSMO runs with a 5sec timestep while CLM and Parflow run at 15min timesteps, which is also the coupling frequency.

For setting up CLM, the European digital elevation model (DEM) by the European Environment Agency EEA (http://www.eea.europa.eu/data-and-maps/data/eu-dem) was projected to the latitude/longitude grid and bi-linearly interpolated to 400m from the original 30m spatial resolution. The same DEM is used to create the slope input files for ParFlow. A slight modification to the original DEM was made in order to ensure that the simulated Neckar River would flow in the correct valley, especially in the upper half of the catchment where the valley is not always properly resolved by the 400m resolution. In total, the elevation of 8 grid points was reduced to achieve proper routing for the Neckar River. The resulting elevation map is part of the CLM input data and available with the dataset as supplementary material. We have not considered rivers outside the Neckar catchment in these corrections; thus, there are cases where their routing is not identical to the real rivers.

Land use is taken from the 2006 Corine Land Cover Data Set (http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster-3) also provided by EEA. Since the latter dataset features many more land use types (at a resolution of 100m) than required by CLM, they were grouped according to the CLM (IGBP) Plant Functional Type classes (1) broad-leaf forests, (2) needle-leaf forests, (3) grassland, (4) cropland, and (5) bare soil. Urban areas are not considered in this setup and replaced by bare soil. Water surfaces (e.g., larger lakes like Lake Constance in the South of the domain) are also treated as bare soil in CLM while COSMO uses its own land-mask and specific calculations for water surfaces. Therefore, no values from CLM are used for water surfaces in COSMO. A few hundred grid cells feature shrubs (mostly areas that are re- or de-forested or areas at higher altitudes) which are treated as forests, and each grid cell features only one – the most dominant – plant functional type. The plant Leaf Area Index (LAI) is computed from MODIS (Myneni et al. 2002) as monthly averages for the year 2008 for each of the four vegetated land use classes. As a result, interannual variability is not considered in this simulation as we have the same LAI curve for each PFT each year. This somewhat limits the comparability to ET observations especially in spring. This LAI is increased for all plant functional types by 20 percent on average (more for forests and less for grassland and crops) in the summer months and significantly changed from factors less than 1 to 3.3 in winter-time (DJF average) for needle-leaf forests in order to account for known biases in the MODIS data (Tian et al. 2004) mostly related to snow cover and fractional land cover due to the satellite footprint which often includes other vegetation types or roads and other buildings, leading to an underestimation for a gridcell that is fully covered by just one type as we use them. The stem area index (SAI) is estimated from the LAI by a slightly modified (no dead leaves for crops, constant base SAI of 10 percent of the maximum LAI) formulation of Lawrence and Chase (2007) and Zeng et al. (2002) to better represent European tree types. Vegetation height was set to 7m for needle-leaf trees and 10m for broad-leaf trees to account for partial coverage by shrubs, to 20 - 120cm for crops, and to 10 - 60 cm for grass depending on the time of the year with low values in the winter months and largest values in July and August. Since we consider only one crop type, we do not

specify a harvest date when the plant height drops to its minimum, but assume a smooth decline between August and October.

For the representation of soils in CLM we use the 1:1,000,000 soil map (BUEK1000, roughly 1km resolution) provided by the Federal Institute for Geosciences and Natural Resources - BGR (http://www.bgr.bund.de/DE/Themen/Boden/Informationsgrundlagen/Bodenkundliche_Karten_Datenbanken/BUEK1000/buek1000_node.html). This soil map is available for entire Germany; thus only small areas in Switzerland and France are missing outside the Neckar catchment for which we assume a nearby soil class. BUEK1000 offers sand and clay percentages as well as carbon content for two to seven soil horizons down to a maximum depth of 3m for each soil type. The carbon content is used to infer soil color. For urban areas (modeled as bare soil, as mentioned above) a fixed soil color (class 8 in CLM) was used.

Since soil properties may vary substantially at scales smaller than the 1km for which BUEK1000 is appropriate, which might impact system dynamics (Binley et al. 1989, Herbst et al. 2006, Rawls 1983), the soil map is downscaled by artificially adding variability using the conditional points method recently presented in Baroni et al. (2017) as follows:

(1) The BUEK1000 soil map is randomly sampled at 1995 point locations with one sample every 5 km^2 on average, a minimum sample distance of 250 m, and at least one sample for each soil type of the original soil map which is realistic in the context of how soil maps are usually created. This strategy resulted from extensive testing by minimizing the tradeoffs between reproducing the main features of the original soil map and creating variability at finer resolution.

(2) The sample locations are used as conditional points for further interpolation. Here, texture, carbon content, and depth of the first three soil horizons are extracted from the BUEK1000 resulting in variable soil depth rather than the assumed unrealistic uniform soil depth. In addition, the sand content of the original map was increased by 20% (except for areas with very high sand content to avoid grid cells with >90% sand) resulting in a slightly higher hydraulic conductivity because previous simulations yielded too shallow unsaturated zones related to the spatial resolution of the simulation. Changing sand content increased the thickness of unsaturated zones and lowered groundwater tables, fixing most of the emerging biases

(3) Experimental variograms and cross-variograms are calculated for all variables and exponential models were fitted to all spatial structures.

(4) A texture map (sand and clay percentage) is generated using a single realization based on conditional co-simulation (Gomez-Hernandez and Journal, 1993) to provide the sub-scale variability (<1 km^2). Soil horizon depths and carbon content are, however, assumed to have a smoothed spatial variability; therefore, they are interpolated based on ordinary kriging as the removal of small-scale variability is not important for the depth and carbon content.

(5) Since ParFlow describes retention and hydraulic conductivity curves based on Mualem-van-Genuchten parameters, pedotransfer functions are applied to estimate these parameters. The pedotransfer function of Cosby et al. (1984) is used to estimate saturated hydraulic conductivity based on soil texture, the one from Rawls (1983) is used to estimate soil bulk density based on soil texture and organic matter and the one from Tóth et al. (2015) is used to estimate van Genuchten

parameters based on soil texture and bulk density. These have been selected based on data availability, applicability of the particular approaches, and previous evaluations conducted in the area (Tietje and Hennings, 1996).

In order to keep soil porosity identical between CLM and ParFlow, we replaced the porosity calculation within CLM (which uses a different pedotransfer function). The Manning's surface roughness was set to a constant value of 5.52×10^{-4} h/m$^{1/3}$ and the specific storage to 1×10^{-3}. The chosen surface roughness value results in a realistic base flow for the local rivers without calibration. Repercussions of this choice are discussed in Chapter 6. Slopes of the main rivers are additionally smoothed to avoid artificial ponded areas.

All these changes are part of the forcing files that are provided with the full dataset making it easy to reproduce our simulations (https://cera-www.dkrz.de/WDCC/ui/cerasearch/entry?acronym=Neckar_VCS_v1_FORCING)

In order to allow for realistic flow in the saturated zone, the 3-D geologic model of the geological survey of the state of Baden-Württemberg was used from which eleven rock types were defined for Baden-Württemberg (see Figure A1). Some characteristic features of the domain, such as middle Triassic and Jurassic karst aquifers, are not included to avoid the manifold hydrological challenges related to its modeling. While this can have significant impact on groundwater representation in the karst areas, for the rather short time period considered here we expect a limited impact on near-surface soil moisture content as the affected areas have in general deeper groundwater levels. For areas outside of Baden-Württemberg we extended the rock types at the boundary outwards to cover the full computational domain. Tab. 1 summarizes porosity and hydraulic conductivity used in the domain for the different stratigraphic units. Since karst features of limestones are not considered, porosities in stratigraphic units containing limestones and crystalline rocks are set considerably higher than in nature to somewhat counter this.

Not covered by the discussed data sets (not part of the soil and not large enough to be resolved in the geological map) are the large alluvial bodies filling large part of the Neckar valley throughout the domain (Riva et al., 2006). Up to 30% of the runoff takes place in the subsurface especially during periods of base flow according to a sub-catchment simulation performed for the year 2007. In that simulation we used measured precipitation and river discharge data together with the simulated evapotranspiration to calculate the water balance over a whole year. While our simulated evapotranspiration rates may be inaccurate, it is implausible that this can account for 30% of the precipitation as in this climate we are almost always energy limited and therefore ET errors will be smaller and mostly related to errors in atmospheric forcings and LAI. This implies that the water could only have left the domain through the subsurface. Thus, gravel channels are needed to account for this lateral flow. Since the valleys in the catchment are often small compared to the limited horizontal resolution of the model, we conceptualize the alluvial bodies as gravel layers underneath all river cells (cells with a mean pressure head >0.1m) and directly next to rivers (riverbanks, i.e., one grid point besides each river cell). The assumed gravel layers reach from beneath the soil down to a depth of 8m. The gravel cells are parameterized with a high hydraulic conductivity of 1 m/h, a porosity of 0.6 and van-Genuchten parameters of 2 for n and 4m^{-1} for α (residual saturation is 0.06 cm^3/cm^3). Our setup results in a reasonable distribution of surface and subsurface discharge at the outlet of the catchment and reasonable river – aquifer exchange fluxes. In addition to the gravel channels, we included a layer of weathered bedrock, which starts below the

soil and extents down to a depth of 6m. This layer is characterized by substantially larger porosity (0.4) and hydraulic conductivity (0.1 m/h) than the rock below. This layer was added to enhance subsurface flow and counter the common occurrence of too shallow water levels if these features are not included. Both these changes are realistic when compared to the actual morphology of the Neckar river valley. While it is quite narrow in many places, there are still significant alluvial deposits everywhere except the furthest upstream region (which are not considered for this anyway due to the pressure cutoff). The choice for the weathered bedrock layer is also reasonable given the temperature and moisture ranges leading to imperfections in the rock layers near the surface.

Since we enforce no-flow boundary conditions at the subsurface domain boundaries, all water has to eventually reach the surface in order to leave the domain. This happens predominantly in areas outside of the Neckar catchment, e.g. in the upper Rhine valley, thus soil-moisture values in this region may be too high.

4 Results

In the following, we present example results of the simulated-reality simulations in order to demonstrate its potential for a better understanding of the dynamics in coupled terrestrial systems. We will also show that the simulations quite well resemble observations in the real Neckar catchment, and thus can be used to develop and evaluate modelling and prediction strategies. Precipitation is the strongest hydrological driver in this region; thus its realistic spatial and temporal variability in the domain including its statistical relations with topography is important. Also, the state of the atmospheric boundary layer, which reflects the interaction of the land surface with the atmosphere is a critical component of the terrestrial system, which should be represented by the simulation with some confidence. Along with the comparisons we will also discuss the challenges experienced with such a modeling setup.

Even though we do not aim to be as close to reality as possible, we feel it important to show that the model system is behaving as expected and is thus suitable for the various use cases we discussed. Figure A2 shows as an example result a snapshot of the simulated three-dimensional distribution of cloud water/ice, precipitation density, and volumetric soil moisture. The soil exhibits different soil moisture layers, the variability of which is mainly connected to different soil hydraulic properties. Only clouds reaching high enough to have sufficient cloud ice produce precipitation, and some precipitation evaporates before it reaches the ground. Extended weather fronts moving through the domain (not shown), which are imposed by the boundary conditions, are also simulated realistically (timing, strength of wind gusts, change of wind direction, change in temperature and pressure) given the resolution of the atmospheric model.

4.1 Relation between water table depth and evapotranspiration

An important measure for hydro-meteorological interactions within a catchment is the relation between water availability and surface energy flux partitioning. Thus, the simulated catchment should capture the expected reduced evapotranspiration (ET) with increasing distance to groundwater (e.g., Maxwell et al., 2007; Shrestha et al., 2014). In Figure 3, we show daily

averaged evaporation (which here is equal to ET as all other contributors are zero) values over bare soil against distance to groundwater for 30th April and 31st July for the year 2007. These days were chosen as they were preceded with several dry days in almost the whole catchment leading to comparable states for the upper soil layers. April was almost completely dry (on average less than 3 mm precipitation over the domain), while July was much wetter, but the increased solar radiation and thus temperatures compared to April result in higher evaporation rates and thus a quicker drying of the top layer of the soil. Figure 3 indicates a reduction in evaporation when the distance to groundwater falls below 15 – 100cm, depending on soil properties with faster evaporation reduction for increasing soil sand contents. Such relations are less obvious for cells with significant plant cover: while trees show overall higher evaporation and almost no change with distance to groundwater due to their deep root zones, variability increases with larger distances to groundwater (not shown). Also crops and grassland show limited evaporation changes as a function of distance to groundwater, which can, however, be explained by the high water availability (no water stress) in the time period considered. Figure 3 also contains a small number of grid-points at water table depth of 7m or deeper with evaporation rates only slightly lower than in the shallow water table regions. These relate most likely to cells that retain high levels of upper-level soil moisture even during dry periods to support higher evaporation. This could be due to the way the water-table is calculated. We define the water-table as the deepest threshold between positive and negative pressure. Since there are some places where there is another saturated region closer to the surface, leading to higher water availability near the surface, the high value for water-table can be misleading with respect to near-surface soil moisture. Such a feature will only occur if the water-table is deep enough to begin with which is why we do not see this for water-tables of less than 10m.As a results volumetric soil moisture for these cells with deep water table but high evaporation is much more similar to cells with shallow water table than to cells with deep water table but low evaporation

We want to point out that in this region ET is almost always limited by atmospheric demand which is why we limit the analysis to bare-soil evaporation only. Since the upper-most layers can dry quickly the resulting drop in evaporation can be seen which is not the case for ET if there is an extended root zone as we have for crop,grassland and forests. These bare-soil areas are not a feature of the real catchment and as such can not be compared to real measurements.

4.2 Precipitation

We compare the simulated precipitation with the 1x1 km gridded REGNIE product of DWD, derived from in-situ precipitation observations (Rauthe et al., 2013). For the evaluation of seasonal daily precipitation cycles hourly observations of 71 DWD observational stations are used. The simulated seasonal mean precipitation (Figure 5) and the annual mean precipitation (not shown) are governed by the orographic structures of the Black Forest and Swabian Alb. Values range between approximately 520 mm/year around Mannheim and 2105 mm/year over the Black Forest in good accordance with REGNIE concerning the overall pattern and range (510 mm/year – 2130 mm/year). Overall the simulation shows about 10% higher annual precipitation in the east and south and about 25% lower in the north and west compared to REGNIE. During winter (December to February) precipitation is dominated by advection from the west, which result in maxima over the

upwind and peak zones of the mountains and leeward minima. The simulated winter pattern (j) compares well with REGNIE (k), but the model underestimates precipitation in the northwestern part of the catchment (l). Over the mountains a slight lateral shift of this kind of precipitation pattern results in neighboring areas with under- and overestimation also found for COSMO simulations coupled to its own TERRA land surface model (e.g., Dierer et al., 2009; Lindau and Simmer, 2013). In fall, the difference pattern between simulations and REGNIE (i) is similar to the winter pattern, but has smaller contrasts. In spring, the simulated precipitation is higher compared to REGNIE. In the summer (June to August), cloud bases are usually higher and reduce the patterns caused by the luff-lee effects. Moist air extends further to the east and south and gets staunched by the alpine upland leading to enhanced precipitation there. The simulated summer precipitation pattern, which is dominated by convective precipitation, resembles the REGNIE pattern but exceeds the latter by about 20% lower over large parts of the catchment (Figure 4).

The mean seasonal diurnal precipitation cycles (Figure 5) reflect the dominating precipitation types. While observed and simulated winter precipitation (Figure 5b) do not show a diurnal cycle, summer precipitation (Figure 5a) increases over the afternoon reaching a maximum at about 7pm in accordance with the maximum of convective precipitation. The simulations reproduce this pattern but exhibit a weak second peak between 6am and 12am while the afternoon/evening increase is delayed by about two hours. The simulated daily precipitation distribution fits the observations especially in late afternoon and night while it overestimates precipitation during the late morning and underestimates it in early afternoon in summer. In winter this effect is much less pronounced. This behavior is related to the representation of convective showers in the atmospheric model. The responsible parametrization was not designed for the km scale and application at this resolution results in a too early onset of convective precipitation. While the simulated catchment has somewhat less dry and low precipitation days than REGNIE, the number of days between 4 and 10 mm are higher than in REGNIE (not shown). The simulated and observed seasonal precipitation cycles (Figure 6) compare very well and mean precipitation is nearly identical between simulations and observations. The model reproduces the seasonal cycle of maximum daily precipitation well, however with larger differences in the summer (see also Dierer et al. 2009).

4.3 Atmospheric State Variables and Surface Radiation

We compare the atmospheric boundary layer (ABL) of the simulated catchment to observations from the meteorological tower at Karlsruhe Institute of Technology (KIT; Kalthoff and Vogel, 1992) and with DWD radiosonde observations in Stuttgart (STG) (see Figure 1 for locations and Table A1 for details of observed quantities). To avoid a biased comparison related to land-cover mismatches between the simulation and the actual land use at the observation sites, the simulation results are averaged over five-by-five atmospheric grid boxes centered around the observation sites thus giving approximately the same fractional land cover as is present at the observation location.

The 10m mean diurnal minimum temperatures in the catchment are between 0.5 K (January) and 2.5 K (August) higher than observed (Figure 7, top) and are reached approximately one hour later than observed with the subsequent morning temperature rise shifted accordingly. The simulated diurnal temperature maxima are on average 0.7 K lower than in the

observations and are reached 30 minutes later than measured. The morning temperature gradient in the simulation ranges from 0.10 K/h in December to 0.31 K/h in April, which compares reasonably well with the observations (0.13/0.52 K/h in January/April). The evening cooling, however, progresses too slowly and results in too high minimum temperatures. At 100 m above ground, diurnal maximum temperatures agree within 0.7 K while the warm bias of diurnal minimum temperatures (0.9 K) is smaller than at 10m height (Figure 7, bottom). Also at 100 m a 1h shift between the diurnal minimum temperatures and the morning temperature rise are found. In 200 m height, the simulated monthly mean diurnal cycles are practically identical to the KIT observations (not shown). The simulated temperature standard deviations (mean absolute difference for each time of day between the specific daily value and the corresponding monthly mean, see appendix Formula A1 for details) are somewhat smaller than observed, especially in afternoons in the summer half year with underestimations of the temperature standard deviation larger than 20%.

COSMO in TerrSysMP estimates ABL heights via the bulk Richardson number criterion with a threshold of 0.22 for unstable and 0.33 for stable conditions (Szintai and Kaufmann, 2008). Both seasonal and diurnal variations of the mean ABL height at 0 and 12h local time agree well with the observations using the same criterion (Figure 8), but the simulation tends to overestimate ABL heights at nighttime by up to 150 m and underestimate it at daytime by up to 200 m in March. Figure 9 compares simulated mean vertical profiles of temperature, virtual potential temperature, and specific humidity with radiosonde observations at 0 h and 12 h local time in Stuttgart (STG) including the mean differences (bias) and the standard deviation of the differences. Simulations are up to 0.9 K warmer close to the surface at 0 h and up to 0.5 K colder at 12 h. At larger heights, the simulations are up to 0.5 K warmer depending on land cover. Specific humidity profiles at 0 h are approximately 0.2 g/kg too dry close to the surface and 0.2 g/kg too wet above 1500 m. At 12 h profiles are up to 0.3 g/kg too wet throughout. The simulations have smaller virtual potential temperature gradients and are thus less stable close to the surface at 0 h. At 12 h, the decreasing virtual potential temperature close to the surface is not captured and tends towards a more neutral instead of unstable profile at low heights.

At KIT (STG) the land surface receives on average 20 W/m^2 (5.3 W/m^2) more incoming shortwave radiation and 18 W/m^2 (8 W/m^2) less incoming longwave radiation indicating a somewhat lower cloud cover (or lower cloud optical depth) as observed. At daytime (6 h – 22 h), the mean outgoing longwave radiation matches the KIT observations, while at nighttime (22 h – 6 h) values are 7.2 W/m^2 larger than observed, which corresponds to a higher surface temperature of approximately 1.4 K.

Overall, the atmospheric profiles, including the ABL heights, are very close to observations during the day and at heights above 10m. Noteworthy differences only occur close to the surface with too high nighttime temperatures (up to 2.5 K in summer) and subsequently too small morning temperature gradients. Somewhat higher incoming shortwave and lower incoming longwave radiation at the surface indicate less cloud cover (or lower cloud optical depths) compared to the observations. These results are in line with a previous evaluation of a 2.2 km COSMO simulation (Ban et al. 2014). In addition, we note somewhat reduced unstable conditions at daytime close to the surface in the simulations.

4.4 Passive Microwave Observations

The most direct area-covering observations of soil moisture are currently provided by L-Band (1.4 GHz) passive microwave observations from satellites. The Community Microwave Emission Model (CMEM) is used as a forward operator to simulate the brightness temperatures (TB) at this frequency in vertical and horizontal polarization (de Rosnay et al., 2009). CMEM simulates brightness temperatures at the top of the atmosphere resulting from microwave emission and interaction by soil, vegetation, and atmosphere based on the state variables of the simulated catchment. Input to CMEM are the percentages of clay and sand in the soil, the coverage with open water surfaces, the profiles of soil moisture and soil temperature, vegetation types, and leaf area index (LAI). Satellite orbit geometry, antenna pattern, foot-print and incidence angle are taken into account following the ESA SMOS (Soil Moisture Ocean Salinity) instrument specifications, i.e. a full-width-half-maximum field of view leading to a footprint of 40km across-orbit and 47km along-orbit at multiple incidence angles (Kerr 2001) is applied. This antenna pattern weighs the grid-cell simulated brightness temperatures (Figure 11, left) in order to obtain simulated SMOS observations. Finally, these synthetic observations are rendered according to pixels based on the Icosahedral Snyder Equal Area (ISEA) projection at a spatial separation of about 15 km similar to the SMOS L1C TB data product (Figure A4, right), which can then be compared with observations for an indirect evaluation of the simulation. Every pixel corresponds to a fixed geo-location of the real SMOS L1C data product over the modeled area. Optionally, the satellite observation operator in TerrSysMP is able to also replicate the NASA SMAP (Soil Moisture Active Passive) radiometer (Saavedra et al., 2016) for years beyond 2015 since when SMAP data is available.

We evaluate the simulated brightness temperature distribution over the domain with real SMOS observations between April 2011 and September 2011. The SMOS observations are corrected from radio-frequency interference (RFI) effects over the region following Saavedra et al. (2016). Initial results with CMEM adapted parameters for surface roughness and vegetation optical thickness (which needed to be increased from its standard values found in the literature), lead to a systematic underestimation of the brightness temperature of about -20K on average (see orange line in Figure A3, which compares real SMOS observations with the simulated brightness temperatures) and maximum and minimum differences of -33K and -6K, respectively, for an incidence angle of 30°. A similar underestimation of -14K resulted for the 40° incidence angle with maximum and minimum values of -34K and +15K (lower plot in Figure 10). Those differences are mainly caused by the too large near-surface soil moisture values in the simulated catchment. The cumulative distribution functions of the satellite-derived soil moisture products and the simulated soil moisture suggests an about 63% higher near-surface soil moisture compared to the satellite estimates (Saavedra et al. 2016, Figure 6) with extremes of 44% and 95%. With that, a daily matching of the cumulative distribution functions of the simulated catchment and satellite retrieved soil moisture is performed to find a factor which then is assumed to be the soil-moisture bias of the simulation and is applied as a correction factor. Figure 10 compares true SMOS observations with simulated brightness temperatures obtained without and with day-to-day correction for the assumed soil-moisture bias of the simulation. The correction decreases the average bias in brightness temperature from -20K(-14K) to about -3K (-2K) for the incidence angle of 30° (40°) at horizontal polarization.

Similar results are found when the simulations were statistically compared with observations of later years from the NASA SMAP (Fig. 3 in Saavedra et al. 2016). The remaining bias can probably be further reduced by fine tuning radiation interaction parameters in CMEM, and by including orographic effects on the effective incidence angle. These biases will be addressed by an improved exploitation of the uncertainty of the radiation interaction parameters and by including in CMEM a two-stream approximation to better simulate cases with dense vegetation in the future.

The microwave observations retrieved from the simulated catchment show a typical situation encountered in data assimilation; more often than not there are biases between simulated and remote sensing observations. This discrepancy usually has multiple causes, which can relate to the observations themselves, assumptions in the observation operator used to simulate the observations, and in the model used to generate the systems state variables entering the observation operator. Even if these differences cannot be removed, such observations can be highly valuable for data assimilation as long as temporal tendencies are meaningful information. Usually, the bias is statistically corrected and thus only the information in the temporal and (if meaningful) spatial variability of the observations is exploited for moving the model states towards the true states.

4.5 Evaluation of River Discharge

We compare river discharge in the simulated catchment with observations made in the Neckar catchment at the gaging stations Rockenau, Lauffen, and Plochingen for a three-year period from 2007 to 2009 (Figure 11). The range of the hydrological responses to precipitation in the simulated catchment is similar to the observations and also during dry periods the behavior is similar, which is noteworthy since no calibration to runoff data has been applied to the model. The simulated discharge peaks are, however, higher and delayed by one to three days compared to the observations. A reason could be a too large Manning´s coefficient and the model resolution. In the discussion we suggest a scaling of Mannings coefficient to account for the mismatch between true river width and the model resolution in order to better represent realistic flood dynamics. In spring and summer, the response to precipitation is significantly smoother than observed and peak amplitudes vary with respect to peak amplitudes of the observations. The differences between observed and simulated precipitation discussed above and the effects of the less predictable convective events during these seasons may also play a significant role. Convective events will often be displaced in space and time compared to the observations and may even show different individual life cycles including lifetime and amplitude. Finally, the base flow is much lower compared to the real catchment during dry periods, most likely because the grid resolution is considerably larger than the actual river width and the unresolved subsurface spatial heterogeneity. An increased hydraulic conductivity via an increased soil sand content may reduce the base flow further as infiltration increases.

The results are further evaluated comparing the flow duration curve and the monthly run off coefficient. The former represents the statistical probability to exceed a specific discharge value within a given time period while the latter is the ratio between runoff and precipitation over the catchment area. Figure A4 shows the lower exceedance probability compared

to the observations, in particular for low discharge rates, a behavior attributed to the lower base flow component and confirmed by the too low runoff coefficients in spring and summer but similar coefficients during the rest of the year (Figure 12). We hypothesize that in this period the simulation has a lower hydrological response also due to missing subsurface heterogeneity. As stated above, we have neglected karst features, which are known to produce fast lateral subsurface flows. Overall, the model captures the general statistical features of the catchment including the typical seasonal trends quite well, while differences are noted related to hydrological extremes and base flow. These differences could be reduced by model calibration from which we refrain because hydrological extremes are not primary the objective of this study. We discuss options to improve the representation of river discharge further below.

4.6 Groundwater

A plausibility check of the groundwater levels is performed in two steps. First, we visually inspect the groundwater depth map, shown in Figure 13a. Accordingly, the model shows a reasonable split between shallower and deeper (5 meter and below) groundwater tables compared to expected values from observations with shallower levels overall. Furthermore, the deeper sections are found in the mountainous areas of the model domain, which corresponds well with the real situation. It has to be noted though that regions with shallow groundwater levels often show very small values, likely not to be found in the real catchment where the unsaturated zone is usually thicker. In a second step, we compare simulated hydraulic heads with available data. The environmental protection agency of the state of Baden-Württemberg (Landesanstalt für Umwelt, Messungen und Naturschutz – LUBW) operates 33 continuous groundwater observation wells. Comparing those point measurements to simulation results of an uncalibrated model with 400m grid resolution makes little sense. Instead, we compare (1) the magnitude of the fluctuation in the groundwater table throughout the catchment during a year (calculated as the groundwater observation minus its yearly mean, shown in Figure 13b) and (2) the average trend of the groundwater level in the full model domain (calculated after subtracting the mean and scaling the fluctuations to have the same magnitude). This means we are comparing standardized anomalies for the observed and simulated groundwater levels. According to Figure 13b, the magnitude of the groundwater fluctuations is within similar ranges as the observations (Figure 13b), while a few observation wells show larger fluctuations. Also, the fluctuations overall follow similar long-term patterns over the year (Figure 13c). Hence, the groundwater, given the coarse resolution of the model in comparison with the compared point measurements, shows a reasonable behavior.

5 Discussion

The size of the catchment and resolution considered (400 m) pose an enormous challenge in terms of required CPU-time. Still, the applicability of Darcy's law with laboratory-based parameters can be debated as we have to resort to apparent model parameters to produce realistic mass fluxes in the compartments. By compromising these technical and physical aspects in the setup of the virtual Neckar catchment, we experienced several challenges; three of them will be discussed

which we believe to be inherent to simulating energy and mass fluxes across compartment boundaries with partial-differential-equation-based, high-resolution coupled models.

Representation of rivers and surface roughness: River flow in the ParFlow module of TerrSysMP is simulated by an overland flow module. Overland flow appears when hydraulic heads in the top cells are above the land surface. As there is no discrimination between overland flow and river flow, rivers in the simulation have the width of the grid resolution whereas the real rivers may be significantly narrower. Overland flow is represented in ParFlow with the kinematic wave approximation of the St. Venant equations with the surface friction parameterized by Manning's coefficient. Typical Manning's coefficients when assigned to e.g. to a 400 m grid cell while in fact the river is much narrower, would result in too high discharge values during rain events and far too low ones during dry periods. In both cases the always too low water levels caused by the too wide rivers result in a poor representation of river-subsurface exchange. Our current choice of Manning's coefficient in ParFlow (5.52×10^{-4} h/m$^{1/3}$) results in realistic average discharge throughout the year, albeit at too low flow velocities. In order to compensate for this inconsistency, the Manning's coefficient could be scaled such that the overland flow velocity in river cells equals the river flow velocity as proposed by Schalge et al. (2019), which improves the phasing between simulated and observed discharge and the discharge peak. Similarly, the hydraulic conductivity of the model top layer for river cells could be scaled in order to reduce the loss of too much surface water to the subsurface caused by the too wide river cells. These issues will become even more severe when model resolutions are reduced, e.g., for ensemble-based data assimilation because of the even higher demands for computing efficiency.

Coarsening of topography: The still coarse topography of the simulation reduces the true hill slopes where lateral flow on the surface and in the shallow subsurface takes place. This affects quick-flow components towards rivers. As shown by Shrestha et al. (2015), coarse topography directly impacts the storage of water in the unsaturated zone because drainage becomes less effective. This in turn can lead to an overestimation of latent and underestimation of sensible heat flux. Additionally, coarse-resolution model runs result in delayed and stretched discharge peaks in the rivers. The severity of this effect is proportional to the degree of topography smoothing, that is introduced by the coarser resolution; therefore, any change in subsurface parameters such as hydraulic conductivity will depend on the degree of coarsening and the location within a catchment. Especially in narrow valleys and in mountainous areas this will lead to an overestimation of soil moisture, which we have not yet compensated by changing other parameters. Recently a method has been proposed to improve these issues by scaling horizontal hydraulic conductivity (Foster and Maxwell, 2019).

Soil parameters: As outlined in section 2, the soil hydraulic parameters were generated based on soil maps of the real Neckar catchment. According to the maps, the soils in the catchment consist mainly of clay and silt, which have rather low saturated hydraulic conductivities and small air entry pressure values. In large areas of the domain, the water content in our first simulations was close to saturation, even for upper soil layers, and the infiltration velocities were unrealistically low. Reasons are the soil parameters, which do not capture the true soil heterogeneity; moreover, real infiltration often takes place in root channels, small fractures, and other small structures. Thus, infiltration is always underpredicted by models using observed soil parameters assuming homogeneity. Infiltration processes may be better captured with dual domain approaches,

556 which are, however, computationally demanding. A workaround would be to change the soil hydraulic parameters in order to
557 obtain stronger infiltration. Currently, we use an artificially increased sand percentage of the soils in order to stay consistent
558 with the concept of the pedotransfer functions used in CLM. We will also test known scaling rules (e.g., Ghanbarian et al.,
559 2015) to increase for example the saturated hydraulic conductivity for larger soil units. These rules should be applied on the
560 soil hydraulic parameters, estimated by the pedotransfer functions.

561 **6 Conclusions and Outlook**

562 In the present study we show the development and the data generated based on an integrated subsurface-land surface-
563 atmosphere system TSMP. Plausibility tests for the derived simulated reality which tries to mimic the Neckar catchment in
564 southwestern Germany, show that the virtual Neckar catchment is able to reproduce realistic behavior when compared to
565 measurements. Comparisons of simulated precipitation and ABL statistics show a very reasonable agreement with
566 observations. However, comparisons with observed passive microwave measurements by satellites shows clearly a
567 systematic bias which is probably related to a mixture of systematic errors in the latter, assumptions in the used forward
568 operator, parameterizations of land surface properties (soil parameters) in the simulation, and missing processes therein (e.g.,
569 preferential flow, hill-slope processes). The analysis also shows a realistic connection between evapotranspiration and
570 distance to groundwater in the simulation, while larger deviations from reality are found for river discharge dynamics. The
571 deficiencies could be traced to the model resolution, which limits the often much smaller river widths to multiples of the
572 model resolution, and to the way river discharge is handled in the ParFlow component of TerrSysMP. A new
573 parameterization scheme proposed by Schalge et al. (2019) will avoid such problems in future model simulations. The main
574 issues we face for the upper Neckar are too high soil moisture and shallow groundwater levels. Several ideas have been
575 proposed to improve the setup including scaling of the surface roughness and soil parameters in response to the results we
576 obtained here. While these changes would show improvements, they are likely marginal or very specific (river discharge
577 characteristics) and would therefore not warrant the great computational cost to re-run for such a long time. Future
578 developments of TerrSysMP may enable this option and it would be interesting to compare resulting datasets and quantify
579 the increase of simulation speed by using GPU compute technologies.
580 Overall the results are encouraging regarding the viability of the simulated reality as key input parameters to the land surface
581 and subsurface show very good agreement with observations. For these reasons, the analysis show that the results can be
582 used as a basis for the community for, among others, exploring feedbacks between compartments, identify in which
583 conditions simplification of the models could be done (Baroni et al., 2019) or develop and test methods for assimilating
584 observations across compartments . We encourage the scientific community to explore this data for the different
585 applications. Within the study we also highlighted some limitations mainly due to the still sever technical limitation and the
586 IT-requirements. We anticipate however that more sophisticated versions of simulated catchments (higher resolution,

improved parameterization of sub-scale processes as discussed above) are already in progress that could be also compared to this dataset in further study.

Finally, we want to address the applicability and usefulness of this dataset for various studies. As indicated, this dataset can be valuable for data assimilation both for testing new methods or algorithms and as a standard set for synthetic observations to pull from. It is thus possible to carry out data assimilation experiments with different conditioning datasets. Due to the long timeseries we have covered almost any possible weather regime (with the exception of truly extreme events) which can be a great advantage as some algorithms may work well for most conditions but may show weaknesses for other specific conditions (for instance the CMEM operator in combination with frozen soils). It also allows to investigate the impact of simplifications such as using a fixed atmospheric forcing instead of a model and thus disregarding feedback mechanisms. Next to data assimilation there are also model development and model analysis and comparison studies that can benefit from this dataset. If specific changes to the model system are made, for example testing a new cloud parametrization, all of the input files that are provided with this dataset can be used to quickly set up a working environment with known results to compare to. Here the length of the simulation is again an advantage since any development can be tested for relevant time slices. A detailed analysis of the dataset regarding compartment interactions is also of interest. We have shown the overall behavior of the system but we have not studied specific interesting events such as heatwaves, dry periods or floods in detail. It would also be of interest to perform longer term simulations to analyze climate change and analyze better inter-annual variability by considering yearly changes in the LAI cycle. Lastly, this setup can also be considered as a template for ensemble-based setups in the future. Right now, reduced resolutions are needed in order to run many members of such a coupled model system. As we have shown, even this higher resolved simulation still shows some biases that are directly related to resolution so increasing resolution also in ensembles will be logical step in the future to obtain better results. When this happens, the methods we used here to generate this simulation will be very useful as well as the analysis presented here to decide how an ensemble should be set up based on the goal (an ensemble for flood forecasts would benefit from a different strategy than an ensemble for drought monitoring).

7 Data Availability

The presented dataset is available in the CERA database of the German Climate Computing Center (DKRZ: Deutsches Klimarechenzentrum GmbH) (Schalge et al., 2020) at https://doi.org/10.26050/WDCC/Neckar_VCS_v1. The full nine-year time series (2007-2015) for all three compartments has a size of roughly 40TB in compressed netCDF4 format. Nevertheless, we encourage the use of this data set for investigations on data assimilation, but also the general functioning of catchments including cross-environmental interactions and predictability studies can profit from such complete state evolutions of the regional Earth system.

The TerrSysMP model is built in a modular way and users are supposed to get the component models by themselves while the coupling interface is provided through a git repository (https://git.meteo.uni-bonn.de). As of now, registration is required to access the TerrSysMP git and wiki page.

Both ParFlow (https://parflow.org/) and CLM (http://www.cgd.ucar.edu/tss/clm/distribution/clm3.5/) are freely available for download from their respective websites or repositories. COSMO is not available, but the DWD supplies it free of charge for research purposes upon request. More information on this process can be found in the TerrSysMP wiki.

Acknowledgements

This research is funded by the Deutsche Forschungsgemeinschaft (DFG, FOR2131: "Data Assimilation for Improved Characterization of Fluxes across Compartmental Interfaces"). The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUQUEEN at Jülich Supercomputing Centre (JSC). We thank the members of HPSC-TerrSys (http://www.hpsc-terrsys.de/hpsc-terrsys/EN/Home/home_node.html) and Klaus Goergen in particular for invaluable technical support with the JUQUEEN supercomputer. Furthermore, we thank Prabhakar Shresta and Mauro Sulis for their preliminary work and introduction to the TerrSysMP modeling platform. We also acknowledge work done on an earlier version of this script by Jehan Rihani.

References

Alkama, R., Papa, F., Faroux, S., Douville, H., and Prigent, C.: Global off-line evaluation of the ISBA–TRIP flood model, Climate Dyn., 38, 1389–1412, doi:10.1007/s00382-011-1054-9, 2012.

Ashby, S.F., and Falgout R. D.: A parallel multigrid preconditioned conjugate gradient algorithm for groundwater flow simulations, Nuclear Science and Engineering, 124(1), 145-159,1996.

Avissar, R., and Pielke, R. A.: A parameterization of heterogeneous land surfaces for atmospheric numerical models and its impact on regional meteorology. Monthly Weather Review, 117(10), 2113-2136, doi:10.1175/1520-0493(1989)117<2113:APOHLS>2.0.CO;2,1989.

Baldauf, M., Seifert, A., Foerstner, J., Majewski, D., Raschendorfer, M., and Reinhardt, T.: Operational convective-scale numerical weather prediction with the COSMO model: description and sensitivities. Monthly Weather Review, 139(12), 3887-3905, doi:10.1175/MWR-D-10-05013.1,2011.

Ban, N., Schmidli, J., and Schaer, C.: Evaluation of the convection-resolving regional climate modelling approach in decade-long simulations. J. Geophys. Res. Atmos., 119, 7889–7907, doi:10.1002/2014JD021478,2014.

Baroni, G., Schalge, B., Rakovec, O., Kumar, R., Schüler, L., Samaniego, L., Simmer, C., Attinger, S.: A comprehensive distributed hydrological modelling inter-comparison to support processes representation and data collection strategies. Water Resources Research. https://doi.org/10.1029/2018WR023941, 2019

Baroni, G., Zink, M., Kumar, R., Samaniego, L., Attinger, S.: Effects of uncertainty in soil properties on simulated hydrological states and fluxes at different spatio-temporal scales. Hydrol. Earth Syst. Sci. 21, 2301–2320. https://doi.org/10.5194/hess-21-2301-2017, 2017.

Bashford, K. E., Beven, K. J., and Young, P. C.: Observational data and scale-dependent parameterizations: explorations using a virtual hydrological reality. Hydrological Processes 16, 293–312. doi:10.1002/hyp.33,2002.

Binley, A., Elgy, J., and Beven, K.: A physically based model of heterogeneous hillslopes: 1. Runoff production. Water Resources Research, 25, 1219–1226. doi:10.1029/WR025i006p01219,1989.

Chan, S., Njoku, E., and Colliander, A.: SMAP L1C Radiometer Half-Orbit 36 km EASE-Grid Brightness Temperatures. Version 2. Boulder, Colorado USA: NASA National Snow and Ice Data Center Distributed Active Archive Center. doi:10.5067/Y9C3Q3060AZ5,2015.

Chaney, N. W., Roundy, J. K., Herrera-Estrada, J. E., and Wood, E. F. (2015): High-resolution modeling of the spatial heterogeneity of soil moisture: Applications in network design. Water Resources Research 51, 619–638. doi:10.1002/2013WR014964,2015.

Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., Hooper, R. P., Kumar, M., Leung, L. R., Mackay, D. S., Maxwell, R. M., Shen, C., Swenson, S. C. and Zeng, X.: Improving the representation of hydrologic processes in Earth system models. Water Resources Research, 51, 5929–5956. https://doi.org/10.1002/2015WR017096, 2015

Cosby, B. J., Hornberger, G. M., Clapp, R. B., and Ginn, T. R.: A Statistical Exploration of the Relationships of Soil Moisture Characteristics to the Physical Properties of Soils. Water Resour. Res., 0, 682–690. doi:10.1029/WR020i006p00682,1984.

de Rosnay, P., Drusch, M., Boone, A., Balsamo, G., Decharme, B., Harris, P., Kerr, Y., Pellarin, T., Polcher, J., and. Wigneron, J. P.: AMMA Land Surface Model Intercomparison Experiment coupled to the Community Microwave Emission Model: ALMIP-MEMP. Journal Geophysical Research, 114, D05108, doi:10.1029/2008JD010724,2009.

Davison, J.H., Hwang, H.-T., Sudicky, E.A., Mallia, D.V., Lin, J.C.: Full Coupling Between the Atmosphere, Surface, and Subsurface for Integrated Hydrologic Simulation. Journal of Advances in Modeling Earth Systems. https://doi.org/10.1002/2017MS001052, 2018

Dierer, S., Arpagaus, M., Seifert, A., Avgoustoglou, E., Dumitrache, R., Grazzini, F., Mercogliano, P., Milelli, M., and Starosta, K.: Deficiencies in quantitative precipitation forecasts: sensitivity studies using the COSMO model. Meteorologische Zeitschrift, 18 (6), 631-645, doi:10.1127/0941-2948/2009/0420,2009.

Foster, L.M., Maxwell, R.M., 2019. Sensitivity analysis of hydraulic conductivity and Manning's n parameters lead to new method to scale effective hydraulic conductivity across model resolutions. Hydrological Processes 33, 332–349. https://doi.org/10.1002/hyp.13327

Gasper, F., Goergen, K., Shrestha, P., Sulis, M., Rihani, J., Geimer, M., and Kollet, S.: Implementation and scaling of the fully coupled Terrestrial Systems Modeling Platform (TerrSysMP) in a massively parallel supercomputing environment-a case study on JUQUEEN (IBM Blue Gene/Q). Geoscientific model development discussions, 7, 3545-3573, doi:10.5194/gmdd-7-3545-2014,2014.

686 Gomez-Hernandez, J., Journel J., and Journel, A. G.: Joint Sequential Simulationof MultiGaussian Fields, in Troia '92,
687 Amilcar Soares (Ed.), volume 1, pages 85-94, Kluwer Academic Publishers,1993.
688 Fatichi, S., Vivoni, E. R., Ogden, F. L., Ivanov, V. Y., Mirus, B., Gochis, D., Downer, C. W., Camporese, M., Davison, J.
689 H., Ebel, B., Jones, N., Kim, J., Mascaro, G., Niswonger, R., Restrepo, P., Rigon, R., Shen, C., Sulis, M., and Tarboton, D.:
690 An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. J.
691 Hydrol. 537, 45–60. doi:10.1016/j.jhydrol.2016.03.026,2016.
692 Ghanbarian, B., Taslimitehrani, V., Dong,, G., and Pachepsky, Y. A.: Sample dimensions effect on prediction of soil water
693 retention curve and saturated hydraulic conductivity. Journal of Hydrology. 528,127-137,
694 doi:10.1016/j.jhydrol.2015.06.024,2015.
695 Haefliger, V., Martin, E., Boone, A., Habets, F., David, C. H., Garambois, P.-A., Roux, H., Ricci, S., Berthon, L., Thévenin,
696 A., and Biancamaria, S.: Evaluation of Regional-Scale River Depth Simulations Using Various Routing Schemes within a
697 Hydrometeorological Modeling Framework for the Preparation of the SWOT Mission, Journal of Hydrometeorology,
698 16, 1821-1842, doi:10.1175/JHM-D-14-0107.1,2015.
699 Hein, A., Condon, L., Maxwell, R M.: Evaluating the relative importance of precipitation, temperature and land-cover
700 change in the hydrologic response to extreme meteorological drought conditions over the North American High Plains.
701 Hydrology and Earth System Sciences 23, 1931–1950. https://doi.org/10.5194/hess-23-1931-2019, 2019
702 Hendricks Franssen, H. J., Alcolea, A., Riva, M., Bakr, M., Van der Wiel, N., Stauffer, F., and Guadagnini, A.: A
703 comparison of seven methods for the inverse modelling of groundwater flow. Application to the characterisation of well
704 catchments, Adv. Water. Resour., 32 (6), 851-872, doi:10.1016/j.advwatres.2009.02.011,2009.
705 Herbst, M., Diekkrueger, B., and Vanderborght J.: Numerical experiments on the sensitivity of runoff generation to the
706 spatial variation of soil hydraulic properties. Journal of Hydrology, 326, 43–58. doi:10.1016/j.jhydrol.2005.10.036,2006.
707 Jones, J. E., and Woodward C. S.: Newton–Krylov-multigrid solvers for large-scale, highly heterogeneous, variably
708 saturated flow problems. Advances in Water Resources, 24(7), 763-774, 2001.
709 Jülich Supercomputing Centre. JUQUEEN: IBM Blue Gene/Q Supercomputer System at the Jülich Supercomputing Centre.
710 Journal of large-scale research facilities, 1, A1. doi:10.17815/jlsrf-1-18, 2015.
711 Kalthoff, N., and Vogel, B.: Counter-current and channelling effect under stable stratification in the area of Karlsruhe.
712 Theor. Appl. Climatol., 45(2), 113-126, doi:10.1007/BF00866400,1992.
713 Kerr, Y H, Waldteufel, P, Wigneron, J P, Front, J and Berger, M: Soil moisture retrieval from space: The soil moisture and
714 ocean salinity (SMOS) mission. IEEE Trans. Geosci. Re. Sens., 39(8): 1729–1735. DOI:https://doi.org/10.1109/36.942551,
715 2001.
716 Kollet, S. J., and Maxwell, R. M.: Integrated surface–groundwater flow modeling: A free-surface overland flow boundary
717 condition in a parallel groundwater flow model. Advances in Water Resources, 29(7), 945-958,
718 doi:10.1016/j.advwatres.2005.08.006,2006.

719 Kollet, S. J., Maxwell, R. M., Woodward, C. S., Smith, S., Vanderborght, J., Vereecken, H., and Simmer, C.: Proof of
720 concept of regional scale hydrologic simulations at hydrologic resolution utilizing massively parallel computer resources.
721 Water Resources Research, 46(4), doi:10.1029/2009WR008730,2010.
722 Kurtz, W., He, G., Kollet, S. J., Maxwell, R. M., Vereecken, H., and H. Franssen, H.-J.: TerrSysMP-PDAF version 1.0): a
723 modular high-performance data assimilation framework for an integrated land surface-subsurface model. Geoscientific
724 Model Development, 9(4), 1341-1360, doi:10.5194/gmd-9-1341-2016,2016.
725 Larsen, M. A. D., Refsgaard, J. C., Drews, M., Butts, M. B., Jensen, K. H., Christensen, J. H., and Christensen, O. B.:
726 Results from a full coupling of the HIRHAM regional climate model and the MIKE SHE hydrological model for a Danish
727 catchment, Hydrol. Earth Syst. Sci., 18, 4733-4749, doi:10.5194/hess-18-4733-2014, 2014.
728 Lawrence, P. J., and Chase, T. N.: Representing a new MODIS consistent land surface in the Community Land Model (CLM
729 3.0). J. Geophys. Res., 112, G01023, doi:10.1029/2006JG000168,2007.
730 Lindau, R., and Simmer C.: On correcting precipitation as simulated by the regional climate model COSMO-CLM with daily
731 rain gauge observations. Meteorology and Atmospheric Physics, 119, 1-2, p.31-42, doi:10.1007/s00703-012-0215-7,2013.
732 Mackay, E. B., Wilkinson, M. E., Macleod, C. J. A., Beven, K., Percy, B. J., Macklin, M. G., Quinn, P. F., Stutter, M., and
733 Haygarth, P. M.: Digital catchment observatories: A platform for engagement and knowledge exchange between catchment
734 scientists, policy makers, and local communities. Water Resour. Res., 51(6), 4815-4822. doi:10.1002/2014WR016824,2015.
735 Maxwell R. M., Chow F. K., Kollet S. J.: The groundwater-land-surface-atmosphere connection: Soil moisture effects on the
736 atmospheric boundary layer in fully-coupled simulations, Advances in Water Resour., 30(12), 2447-2466, 2007
737 Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G.
738 R., Lotsch, A., Friedl, M., Morisette, J. T., Votava, P., Nemani, R. R., and Running, S. W.: Global products of vegetation
739 leaf area and fraction absorbed PAR from year one of MODIS data. Remote Sensing of Environment, 83(1), 214-231,
740 doi:10.1016/S0034-4257(02)00074-3,2002.
741 Oleson, K. W., Niu, G. Y., Yang, Z. L., Lawrence, D. M., Thornton, P. E., Lawrence, P. J., and Qian, T.: Improvements to
742 the community land model and their impact on the hydrological cycle. Journal of Geophysical Research: Biogeosciences
743 (2005–2012), 113(G1), doi:10.1029/2007JG000563,2008.
744 Poltoradnev, M., Ingwersen, J., and Streck, T.: Spatial and temporal variability of soil water content in two regions of
745 Southwest Germany during a three-year observation period. Vadose Zone Journal, 15(6),
746 doi:10.2136/vzj2015.11.0143,2016.
747 Rauthe, M., Steiner, H., Riediger, U., Mazurkiewicz, A., and Gratzki, A. (2013): A Central European precipitation
748 climatology – Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS), Vol. 22(3), p 235–256,
749 doi:10.1127/0941- 2948/2013/0436,2013.
750 Rawls, W. J.: Estimating soil bulk density from particle size analysis and organic matter content. Soil Sci., 135 (2), 123-
751 125,1983.

Ridler, M. E., van Velzen, N., Hummel, S., Sandholt, I., Falk, A. K., Heemink, A., and Madsen, H.: Data assimilation framework: linking an open data assimilation library (OpenDA) to a widely adopted model interface (OpenMI), Environ. Modell. Softw., 57, 76–89, doi:10.1016/j.envsoft.2014.02.008, 2014.

Riva, M., Guadagnini, L., Guadagnini, A., Ptak, T., and Martac, E. (2006): Probabilistic study of well capture zones distribution at the Lauswiesen field site. J. Contam. Hydrol. 88, 92–118, doi:10.1016/j.jconhyd.2006.06.005,2006.

Saavedra P., Simmer, C., and Schalge, B.: Evaluation of modeled high resolution soil moisture virtual brightness temperature compared to space-borne observations for the Neckar catchment. IEEE Xplore 14th Specialist Meeting on Microwave Radiometry and Remote Sensing of the Environment (MicroRad), pp. 85-90, doi:10.1109/MICRORAD.2016.7530510,2016.

Saavedra, P. and Simmer, C.: An Octave/MATLAB® Interface for Rapid Processing of SMOS L1C Full Polarization Brightness Temperature. Journal of Open Research Software, 6, 1, p.2, DOI: http://doi.org/10.5334/jors.165, 2018

Schaefer, D., Dahmke, A., Kolditz, O., and Teutsch, G.: Virtual Aquifers: A Concept for Evaluation of Exploration, Remediation and Monitoring Strategies. In: Calibration and Reliability in Groundwater Modelling: A Few Steps Closer to Reality (Proceedings of the ModelCARE 2002 Conference held in Prague, Czech Republic, June 2002), edited by K. Kovar, Z. Hrkal, IAHS Publication 277, IAHS Press Oxfordshire, pp. 52-59,2002.

Schalge, B., Haeflinger, V, Kollet, S, and Simmer, C: Improvement of surface runoff in the hydrological model ParFlow by a scale-consistent river parameterization. Accepted for publication in Hydrological Processes, 2019.

Schalge, Bernd; Baroni, Gabriele; Haese, Barbara; Erdal, Daniel; Geppert, Gernot; Saavedra, Pablo; Haefliger, Vincent; Vereecken, Harry; Attinger, Sabine; Kunstmann, Harald; Cirpka, Olaf A.; Ament, Felix; Kollet, Stefan; Neuweiler, Insa; Hendricks Franssen, Harrie-Jan; Simmer, Clemens (2020). Virtual catchment simulation based on the Neckar region. World Data Center for Climate (WDCC) at DKRZ. https://doi.org/10.26050/WDCC/Neckar_VCS_v1

Schlueter, S., Vogel, H.-J., Ippisch, O., Bastian, P., Roth, K., Schelle, H., Durner, W., Kasteel, R., and Vanderborght, J.: Virtual soils: assessment of the effects of soil structure on the hydraulic behavior of cultivated soils. Vadose Zone Journal 11, doi:10.2136/vzj2011.0174,2012.

Schomburg, A., Venema, V., Lindau, R., Ament, F., and Simmer, C.: A downscaling scheme for atmospheric variables to drive soil–vegetation–atmosphere transfer models. Tellus B, 62(4), 242-258, doi:10.1111/j.1600-0889.2010.00466.x,2010.

Schomburg, A., Venema, V., Ament, F., and Simmer, C.: Disaggregation of screen-level variables in a numerical weather prediction model with an explicit simulation of subgrid-scale land-surface heterogeneity. Meteorology and Atmospheric Physics, 116(3-4), 81-94, doi:10.1007/s00703-012-0183-y,2012.

Semenova, O., and Beven, K.: Barriers to progress in distributed hydrological modelling: Invited commentary. Hydrological Processes, 29(8), 2074–2078. https://doi.org/10.1002/hyp.10434

Shi, Y., Davis, K. J., Zhang, F., Duffy, C. J., and Yu, X.: Parameter estimation of a physically based land surface hydrologic model using the ensemble Kalman filter : A synthetic experiment, Water Resources Research, 50(1), 706-724, doi:10.1002/2013wr014070,2014.

Shi, Y., Davis, K. J., Zhang, F., Duffy, C. J., and Yu, X.: Parameter estimation of a physically-based land surface hydrologic model using an ensemble Kalman filter: A multivariate real-data experiment, Advances in Water Resources, 83, 421-427, doi:10.1016/j.advwatres.2015.06.009,2015.

Shrestha, P., Sulis,, M., Masbou, M., Kollet, S., and Simmer, C.: A scale-consistent terrestrial systems modeling platform based on COSMO, CLM, and ParFlow. Monthly Weather Review, 142(9), 3466-3483, doi:10.1175/MWR-D-14-00029.1,2014.

Shrestha, P., Sulis, M., Simmer, C., Kollet, S.: Impacts of grid resolution on surface energy fluxes simulated with an integrated surface-groundwater flow model, Hydrol. Earth Syst. Sci., 19, 4317-4326, doi:10.5194/hess-19-4317-2015,2015.

Simmer, C., Thiele-Eich, I., Masbou, M., Amelung, W., Crewell, S., Diekkrueger, B., Ewert, F., Hendricks Franssen, H.-J., Huisman, A. J., Kemna, A., Klitzsch; N., Kollet, S., Langensiepen, M., Loehnert, U., Rahman, M., Rascher, U., Schneider, K., Schween, J., Shao, Y., Shrestha, P., Stiebler, M., Sulis, M., Vanderborght, J., Vereecken, H., van der Kruk, J., Zerenner, T., and Waldhoff, G.: Monitoring and Modeling the Terrestrial System from Pores to Catchments - the Transregional Collaborative Research Center on Patterns in the Soil-Vegetation-Atmosphere System. Bulletin of the American Meteorological Society, 96, 1765–1787. doi:10.1175/BAMS-D-13-00134.1,2015.

Smith, V. H., Mobbs, S. D., Burton, R. R., Hobby, M., Aoshima, F., Wulfmeyer, V., and Di Girolamo, P.: The role of orography in the regeneration of convection: A case study from the convective and orographically-induced precipitation study. Meteorologische Zeitschrift, Vol. 24, No. 1, 83–97, doi:10.1127/metz/2014/0418,2015.

Sulis, M., Langensiepen, M., Shrestha, P., Schickling, A., Simmer, C., and Kollet, S. J.: Evaluating the influence of plant-specific physiological parameterizations on the partitioning of land surface energy fluxes. Journal of Hydrometeorology, 16(2), 517-533, doi:10.1175/JHM-D-14-0153.1,2015.

Szintai, B., and Kaufmann , P.: TKE as a measure of turbulence. COSMO Newsletter 8, 2-9,2008.

Tian, Y., Dickinson, R. E., Zhou, L., Zeng, X., Dai, Y., Myneni, R. B., Knyazikhin, Y., Zhang, X., Friedl, M., Yu, I., Wu, W., and Shaikh, M.: Comparison of seasonal and spatial variations of leaf area index and fraction of absorbed photosynthetically active radiation from Moderate Resolution Imaging Spectroradiometer (MODIS) and Common Land Model. J. Geophys. Res., 109, D01103, doi:10.1029/2003JD003777,2004.

Tietje, O., and Hennings, V.: Accuracy of the saturated hydraulic conductivity prediction by pedo-transfer functions compared to the variability within FAO textural classes. Geoderma, 69, 71–84. doi:10.1016/0016-7061(95)00050-X,1996.

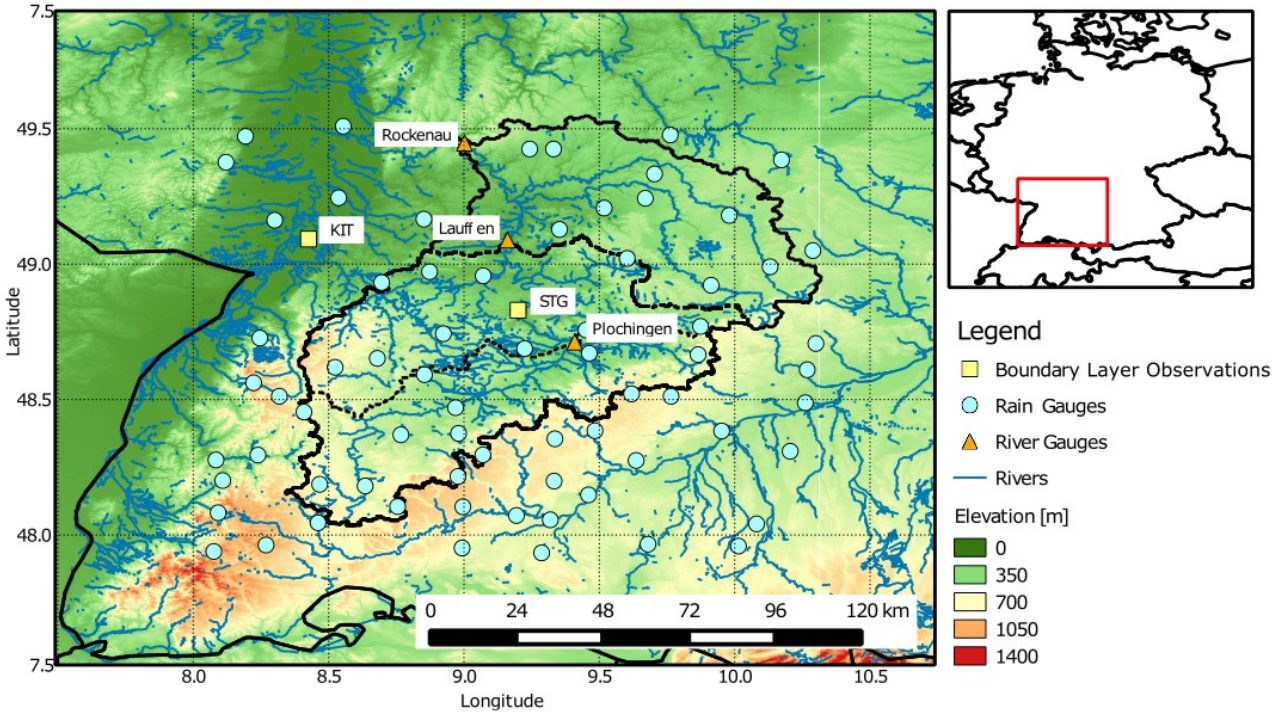
Tóth, B., Weynants, M., Nemes, A., Makó, A., Bilas, G., and Tóth, G.: New generation of hydraulic pedotransfer functions for Europe: New hydraulic pedotransfer functions for Europe. Eur. J. Soil Sci., 66, 226–238. doi:10.1111/ejss.12192,2015.

Valcke, S.: OASIS3 PRISM Model Interface Library: OASIS3 User Guide. PRIMS Support Initiative, Report No. 3,2006.

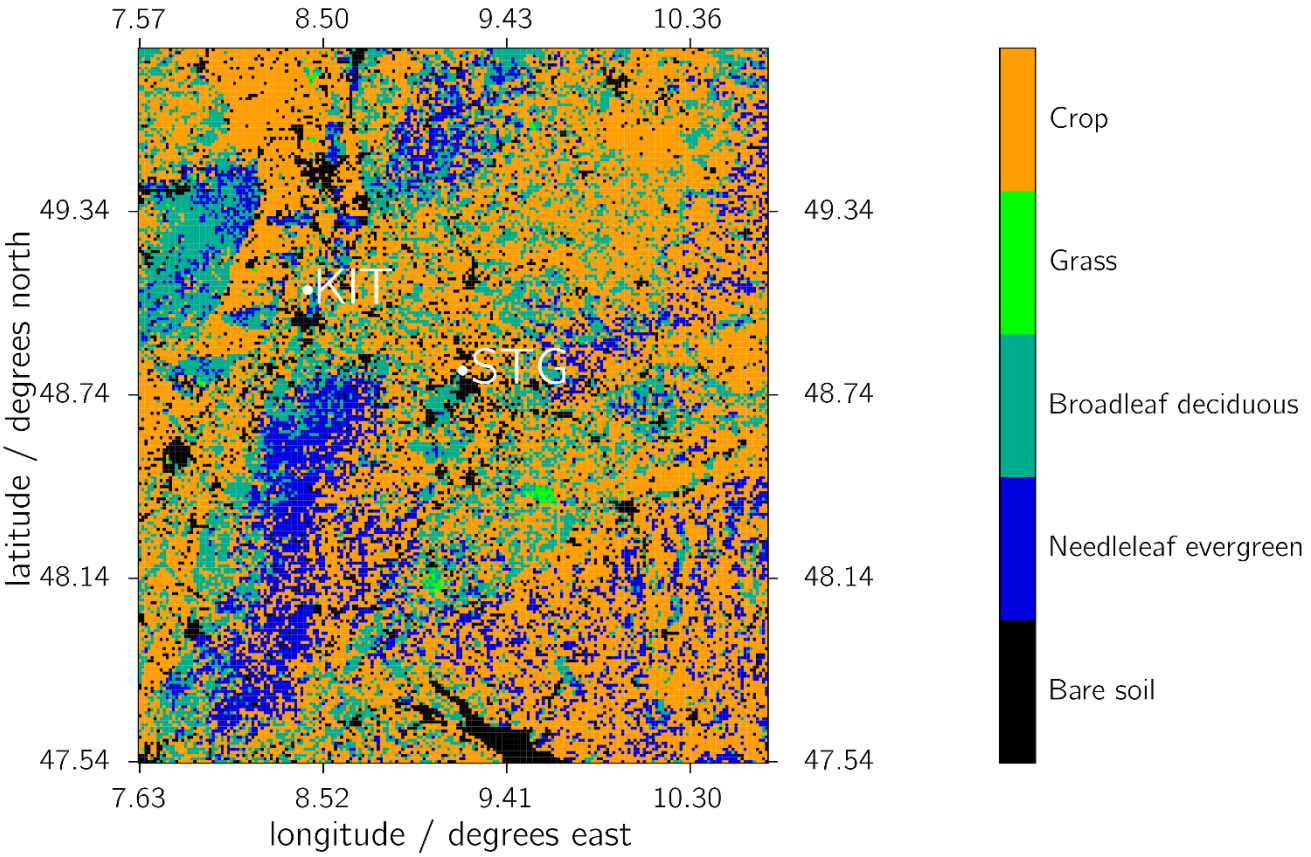
Weiler, M., and McDonnell, J.: Virtual experiments: a new approach for improving process conceptualization in hillslope hydrology. Journal of Hydrology 285, 3–18. doi:10.1016/S0022-1694(03)00271-3,2004.

Zeng, X., Shaikh, M., Dai, Y., Dickinson, R. E., Myneni, R.: Coupling of the common land model to the NCAR community climate model. J. Climate 15, 1832–1854, doi:10.1175/1520-0442(2002)015<1832:COTCLM>2.0.CO;2,2002.

820 Zimmerman, D. A., G. de Marsily, G., C.A. Gotway, C. A., M.G. Marietta, M. G., C.L. Axness, C. L., R.L. Beauheim, R. L.,
821 Bras, R. L., Carrera, J., Dagan, G., Davies, P. B., Gallegos, D. P., Galli, A., Gomez-Hernandez, J., Grindrod, P., Gutjahr, A.
822 L., Kitanidis, P. K., Lavenue, A. M., McLaughlin, D., Neuman, S. P., RamaRao, B. S., Ravenne, C., and Rubin, Y.: A
823 comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport
824 by groundwater flow, Water Resour Res 34(6), 1373-1413, doi:10.1029/98WR00003,1998.
825

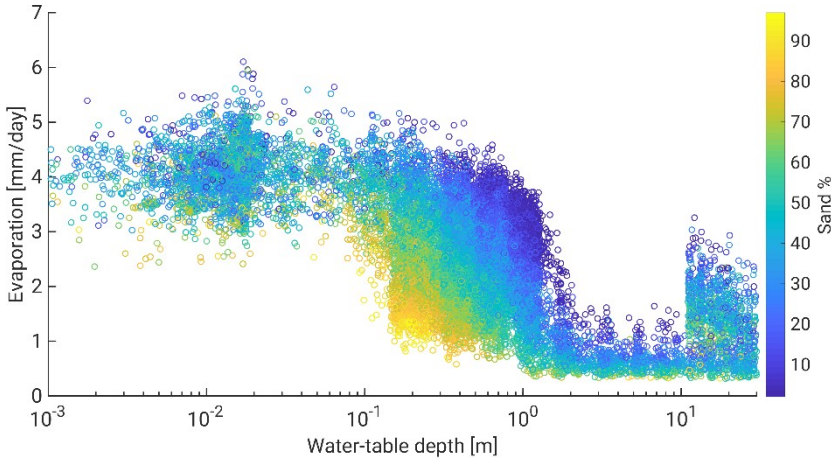


826
827 **Figure 1: Location of the Neckar catchment within SW Germany.**
828



829

830 **Figure 2: Land cover in the simulated domain covering the entire Neckar catchment and bounding areas. KIT: Karlsruhe**
831 **Institute of Technology (location of meteorological tower observations), STG: Stuttgart (location of radiosonde observations).**



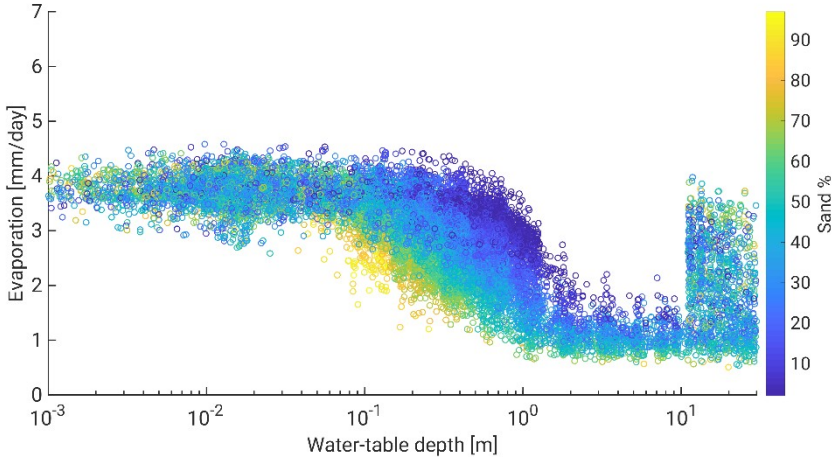
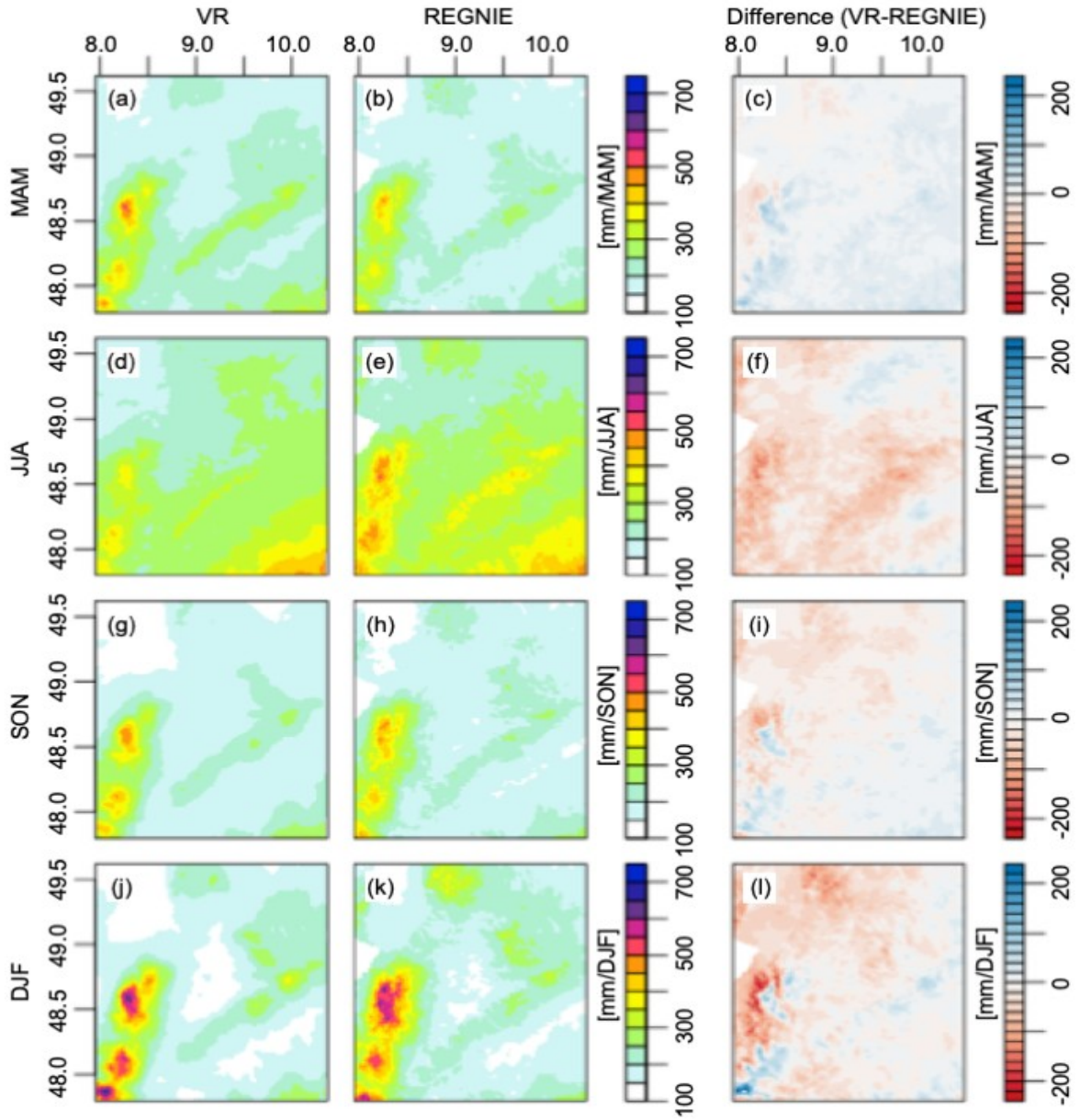


Figure 3: Daily average evaporation simulated for 30th April (left) and 31st July 2007 in [mm/day]. The color indicates soil sand percentage.



837

838 **Figure 4: Mean seasonal precipitation over the Neckar catchment between 2007-2013 in the simulated reality (VR, left column)**
839 **compared to the REGNIE data set (middle column). The difference between VR and REGNIE is shown in the right column.**
840 **Figure (a), (b), and (c) show the comparison for spring (March – May); (d), (e), and (f) for summer (June – August); (g), (h), and**
841 **(i) for fall (September – November); and (j), (k), and (l) for winter (December-February).**

842

843

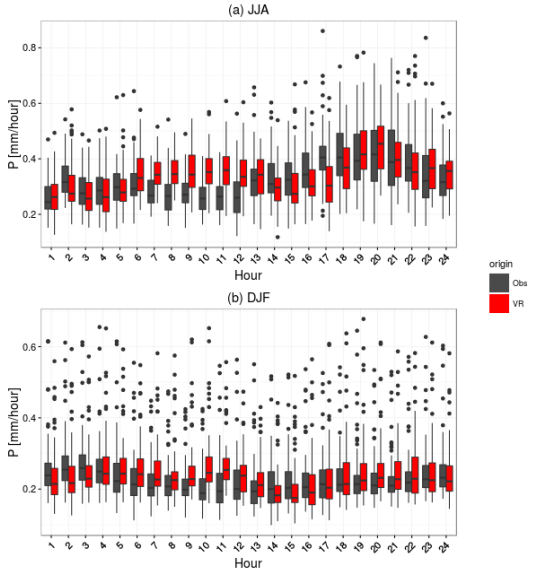


Figure 5: Mean diurnal precipitation cycle for the 71 DWD stations and the corresponding simulations for wet days (more than 1 [mm/day]) for June-August (a) and December-February (b) season. The upper and lower hinges correspond to the first and third quartile, the center black line the median, the upper whisker (analog for lower whisker) extends from the hinge to the highest value within 1.5*(interquartile range), and the black dots mark the outliers.

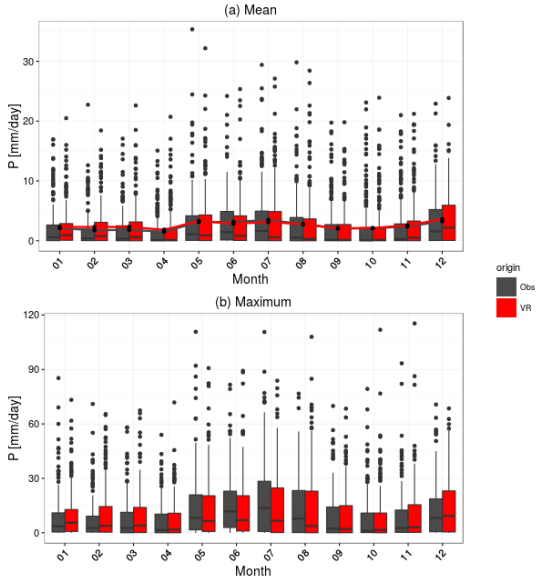


Figure 6: (a) Daily precipitation distribution on a monthly basis as observed (black) and simulated (red). The gray and red lines indicate the monthly mean precipitation. (b) Maximum daily precipitation for the given months for the 71 DWD stations and the corresponding simulation. Box sizes as explained in the caption of Figure 10.

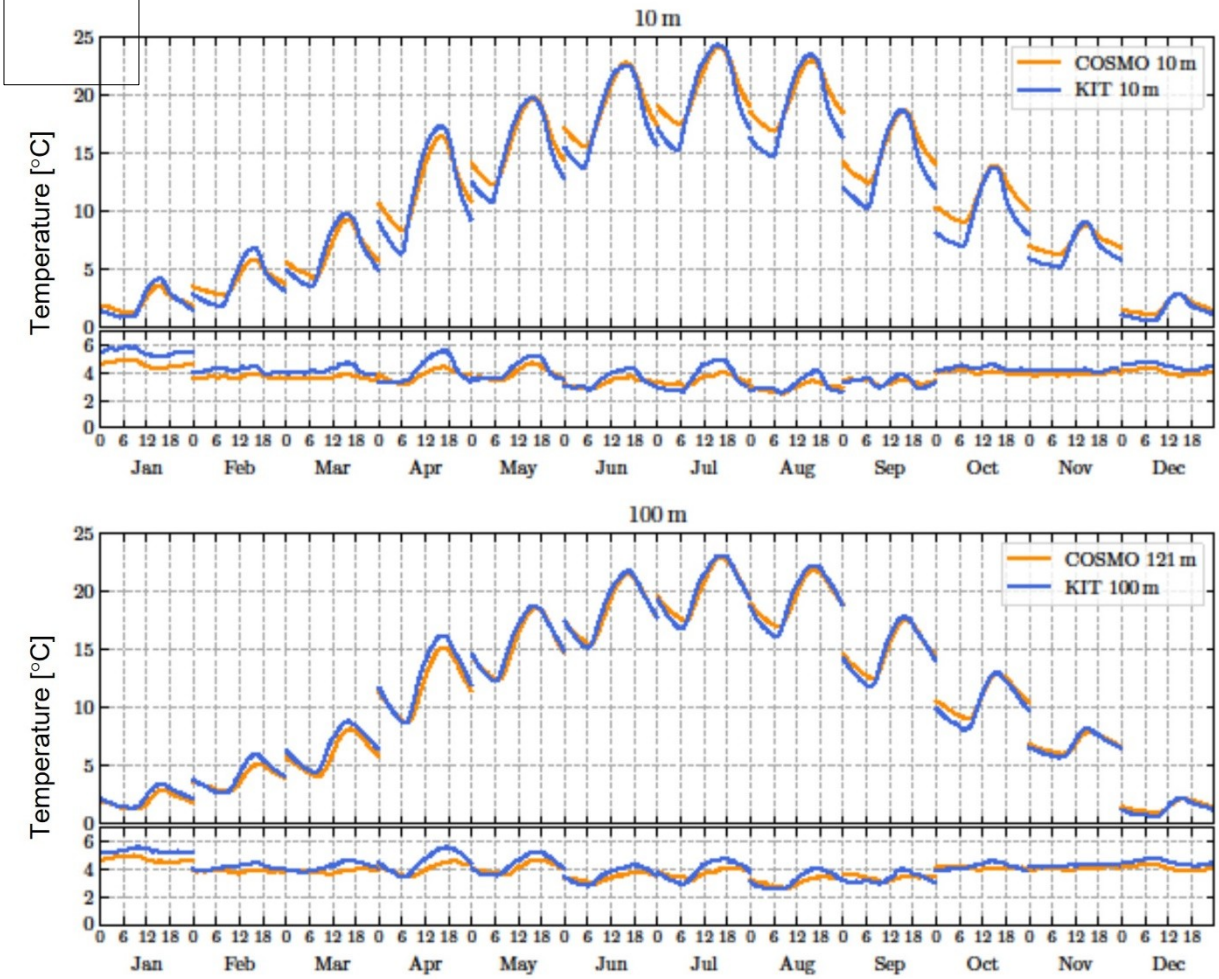
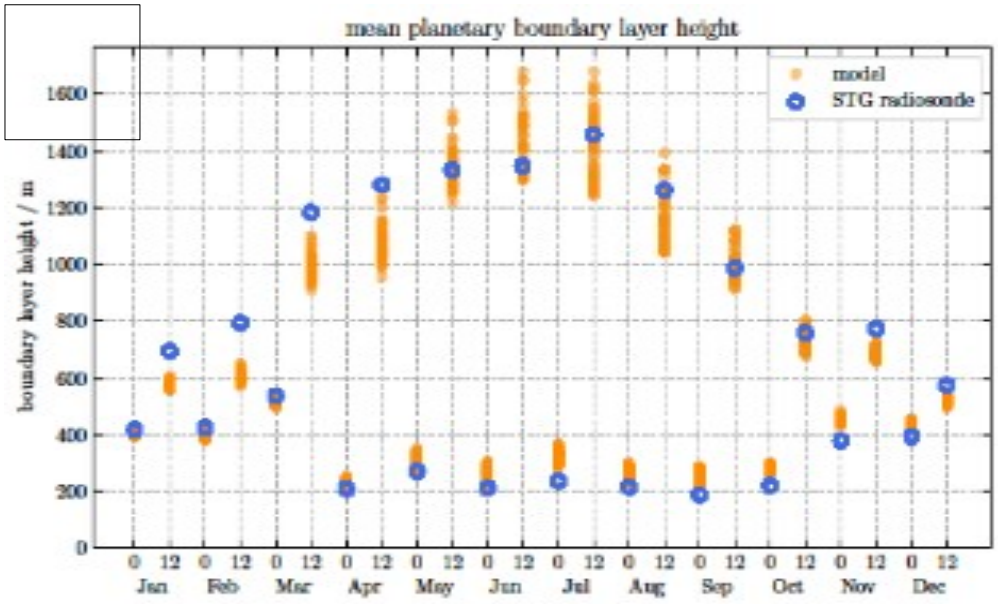


Figure 7: Monthly mean diurnal cycles (local time) and respective standard deviation (see text) for air temperature [°C] in 10 m (top) and 100 m (bottom) height at the KIT tower and for the COSMO grid boxes around the KIT location.



860

861 **Figure 8: Monthly mean boundary layer height at 0 h and 12 h local time for different land covers diagnosed from radiosonde**
862 **observations at Stuttgart STG and from atmospheric profiles above grid boxes of CLM.**

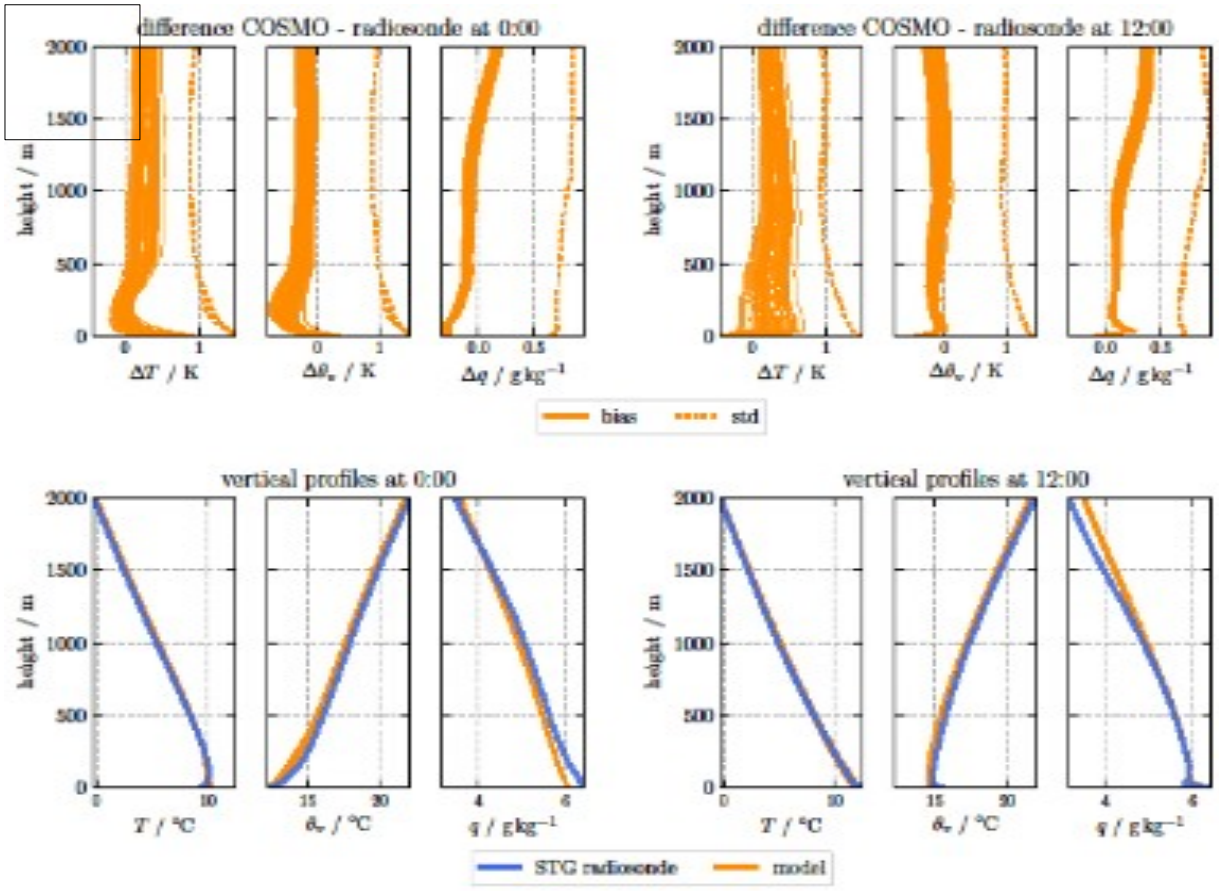
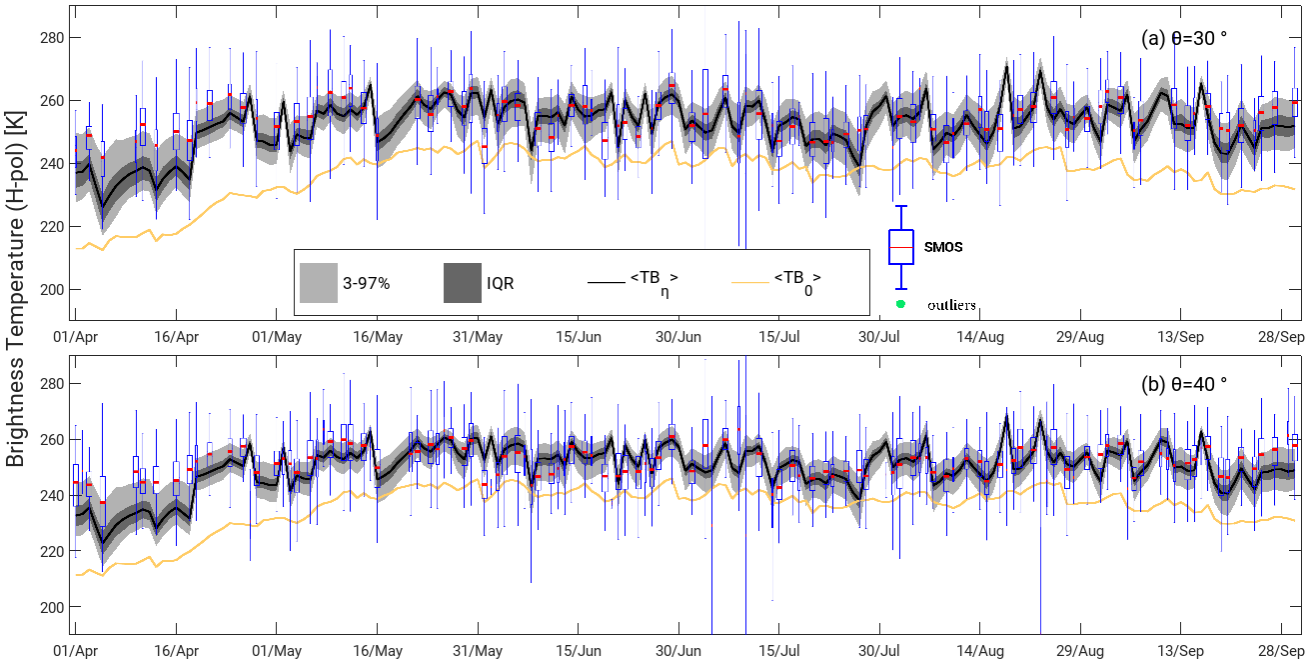
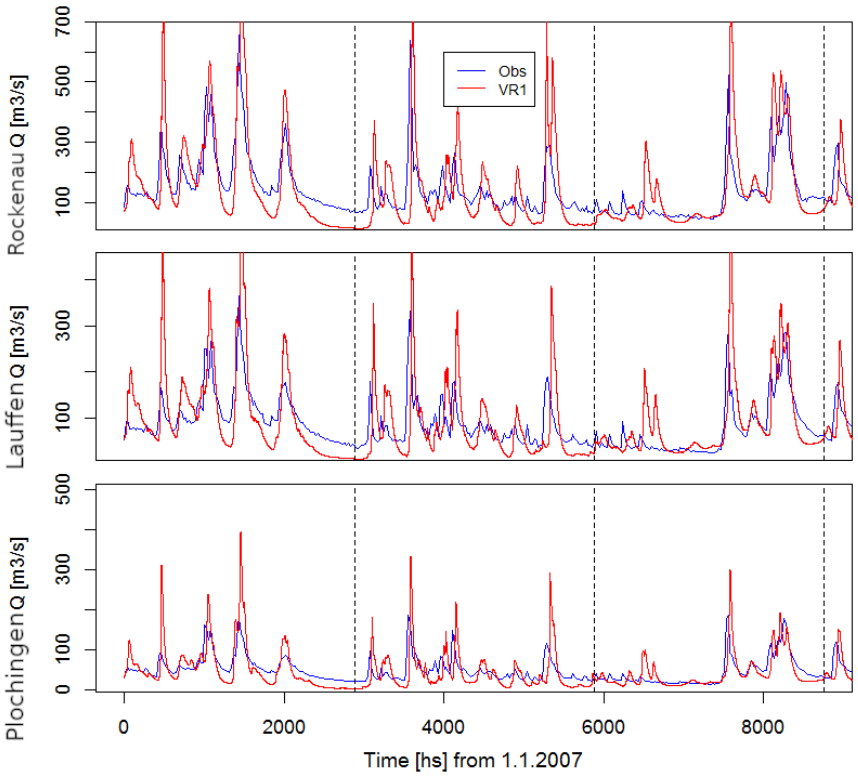


Figure 9: Mean vertical profiles of temperature, virtual potential temperature, and specific humidity (top), and mean differences between modelled and observed data including the standard deviation of the differences (bottom). The experimental data are from the radiosonde data at STG and the simulated data from the grid boxes of the simulated catchment with different land cover (left: 0 h local time, right: 12 h local time).



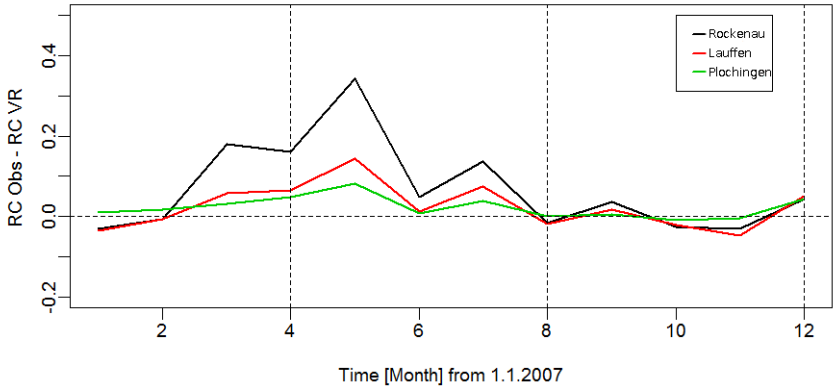
868

869 **Figure 10: Area-averaged L-band brightness temperature the period from April to September 2011 for an incidence angle of 30°**
870 **(top) and 40° (bottom). The boxplots indicate the real SMOS observations averaged over the same domain. The black line is the**
871 **median of the observations simulated with CMEM. The dark-gray area corresponds to the inter-quartile range (IQR) while the**
872 **light-gray area encompasses the 3 to 97% range. The orange continuous line indicates the brightness temperature without taking**
873 **into account an assumed bias in surface soil moisture content (see text).**



874

875 **Figure 11: Hourly values river discharge at the gauging stations Rockenau (P1), Lauffen (P2) and Plochingen (P3) for the year**
876 **2007. Blue: observed; red: simulated catchment.**



877

878 **Figure 12: Differences between the run off coefficient calculated for the three stations for the year 2007 based on observations and**
879 **simulation.**

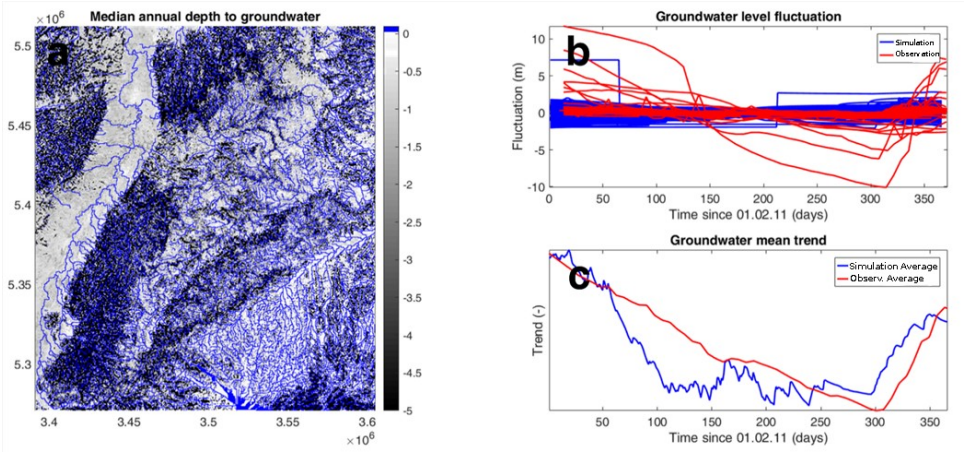


Figure 13:(a) Mean groundwater table depth of the entire domain for the year ranging from 01.02.11 to 01.02.12, (b) groundwater fluctuations around a zero mean and (c) the total mean of all model cells and all real data points superimposed on top of each other to show the annual average trend. Please note that for readability of the figure, subfigure (a) is limited to a maximum depth of -5 m, while the underlying data ranged down to -88 m

Appendix

7.1 Appendix Tables

Table A1: Values of porosity and hydraulic conductivity of rocks found in Baden-Wuerttemberg

Nr.	rock type	Ksat [m/h]	porosity fraction
1	Quarternary	0.00100	0.3
2	Tertiary	0.00100	0.3
3	Upper Jura	0.00720	0.3
4	Middle Jura	10^{-7}	0.3
5	Lower Jura	10^{-7}	0.3
6	Upper Triassic (Keuper)	0.00036	0.3
7	Middle Triassic (Muschelkalk)	0.00180	0.3
8	Lower Triassic (Buntsandstein)	0.02160	0.4
9	Upper Permian (Rotliegendes)	0.00360	0.3
10	New Red Conglomerate	0.00100	0.3
11	Bedrock/Granite	10^{-7}	0.3

Table A2: Observed atmospheric variables at KIT and STG. Local time at STG is UTC+01.

dataset	quantity	temporal resolution	height above ground	data coverage
KIT	temperature	10 min averages (resampled to 15 min)	10 m, 100 m	01/2007 – 12/2013
	Incoming and outgoing shortwave radiation		-	
	Incoming and outgoing longwave radiation		-	06/2011 – 12/2013
STG	temperature	12 h (11:45 h and 23:45 h local time)	vertical profiles (interpolated to model levels)	01/2007 – 12/2013
	dew point temperature			
	pressure			
	incoming shortwave radiation	1 h averages	-	
	incoming longwave radiation		-	

902 **Table A3: Strengths and Weaknesses of our simulation regarding several variables**

Variable	Strength	Weakness	Other comment
River discharge	Adequadely captures flood events	Underestimated base flow	Could be improved by higher resolution of the river cells (ParFlow in general)
ET	Shows expected behavior	Little variability in forests as they are never water limited	-
Precipitation	Good agreement with observations in general	Some specific areas show larger differences	Could be fixed by larger simulation area and new parametrizations for the 1km scale
ABL representation	Good fit to tower and radiosonde observations, especially during the daytime in summer	Overestimated nighttime temperature, especially in the summer	Related to COSMO parametrization of near surface inversion at night
Groundwater dynamics	Dynamics closely resembles measured dynamics	Absolute values are biased	Could be related to our simplification of not considering karst
Large-scale soil moisture (satellite)	Dynamics throughout the year are captured well	Large bias in absolute values	

903

904

7.2 Appendix Figures

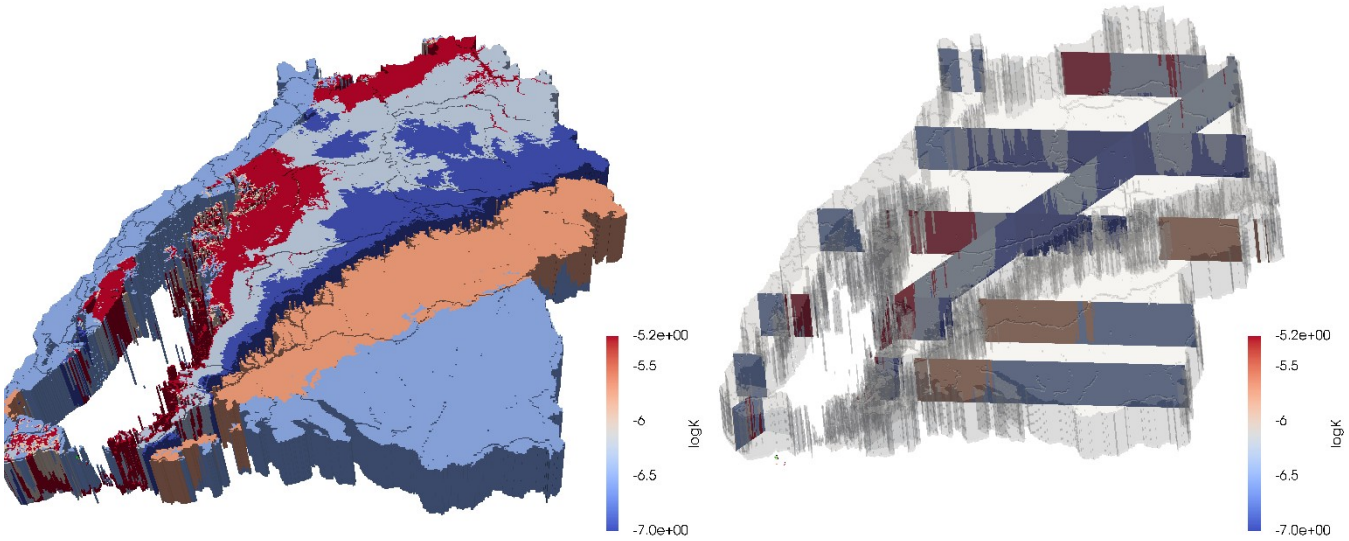
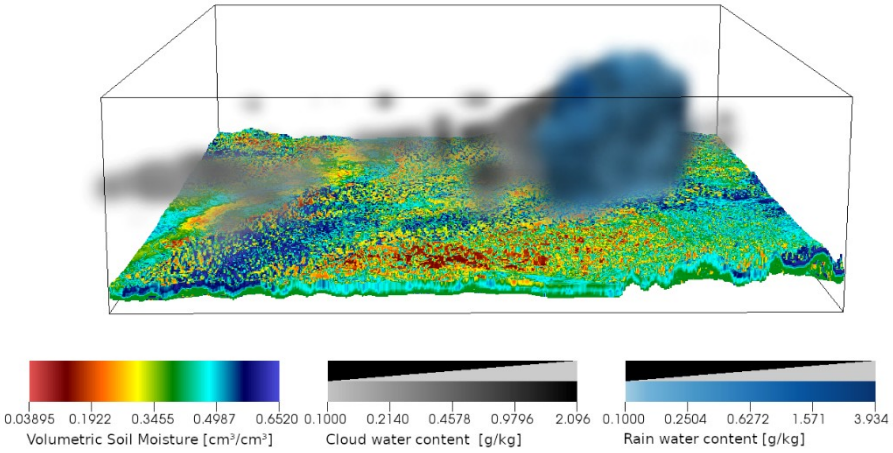


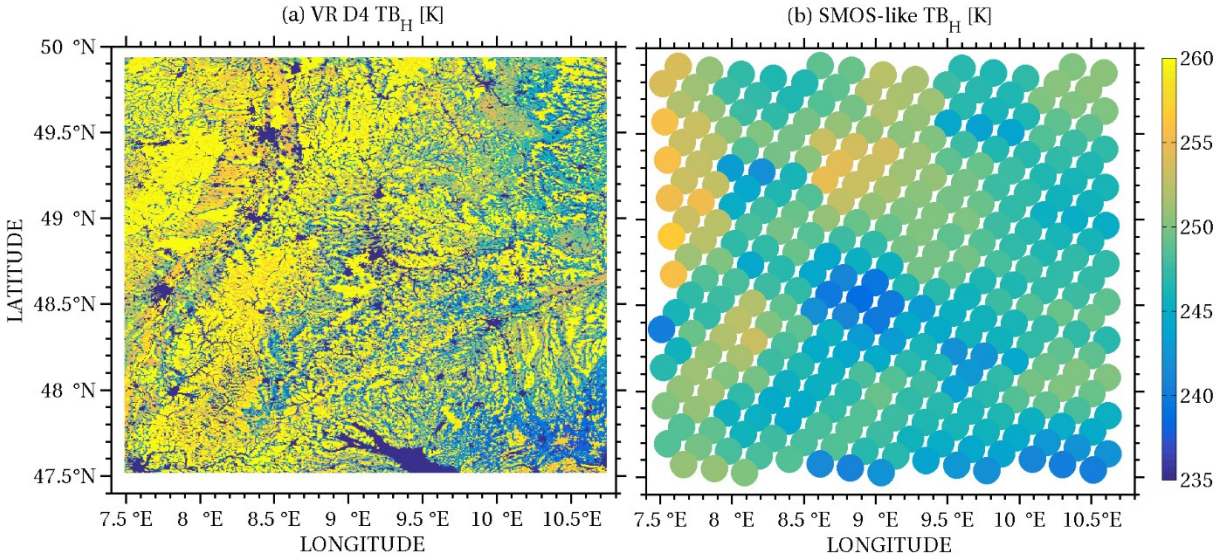
Figure A1: Stratigraphy in the state of Baden-Württemberg represented by its logarithmic conductivity. The left figure shows a 3-D view of the 100 m deep geological model used in this work, where the elevation has been neglected for readability and the transparent regions corresponds to low-permeable material. The right figure shows the same using cross-sections to better visualize the vertical heterogeneity.



911

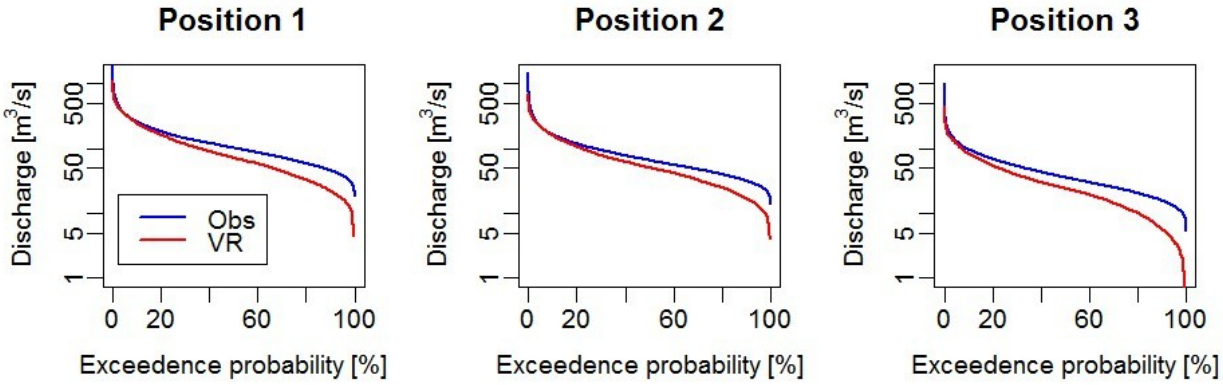
912 **Figure A2: Snapshot of the three dimensional distribution of cloud water/ice [g/kg] (greyscale), precipitation/rain water [g/kg]**
913 **(blue in foreground over cloud) and soil moisture [cm3/cm3] (colored) at a time point with a single rain cloud with light rain.**

914



915

916 **Figure A3: Brightness temperature calculated by the application of CMEM (H-polarization) on the simulated-reality output on**
917 **July 2nd 2011 (left) and its aggregation on the spatial resolution of the L1C data-product SMOS passive microwave radiometer**
918 **(right).**



919

920 **Figure A4: Flow duration curve for the three stations for the three-year time period based on. Blue: observations; red: simulated**
921 **catchment.**

922 7.3 Appendix Equations

923 $$\sigma_{T,t} = \frac{1}{days} \sum \left| T_{days,t} - \overline{T}_t \right|$$

924 Equation A1: σ is the temperature standard deviation and the subscript t denotes the time of day. This is calculated
925 separately for each month of the year to create the 12 profiles. The overbar for the temperature T denotes the monthly mean

926 temperature value for each time of the day, while the subscript days,t indicates that this is the daily value for the respective
927 time of day.