Improved Characterisation of Fluxes Across Compartmental Interfaces of the Terrestrial System

B. Schalge, C. Simmer, H.-J. Hendricks Franssen, V. Haefliger, B. Haese, S. Kollet

published in

NIC Symposium 2018
K. Binder, M. Müller, A. Trautmann (Editors)

Forschungszentrum Jülich GmbH,
John von Neumann Institute for Computing (NIC),
Schriften des Forschungszentrums Jülich, NIC Series, Vol. 49,
ISBN 978-3-95806-285-6, pp. 349.
http://hdl.handle.net/2128/17544

Improved Characterisation of Fluxes Across Compartmental Interfaces of the Terrestrial System

Bernd Schalge[1,4], Clemens Simmer[1,4], Harrie-Jan Hendricks Franssen[2,4], Vincent Haefliger[1,4], Barbara Haese[3], and Stefan Kollet[2,4]

[1] Meteorological Institute, University of Bonn, 53121 Bonn, Germany
E-mail: {csimmer, bschalge, chaeflig}@uni-bonn.de

[2] Institute of Bio- and Geosciences: Agrosphere (IBG-3), Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: {h.hendricks-franssen,s.kollet}@fz-juelich.de

[3] Institute of Geography, University of Augsburg, 86135 Augsburg, Germany
E-mail: barbara.haese@geo.uni-augsburg.de

[4] Centre for High-Performance Scientific Computing in Terrestrial Systems,
`http://www.hpsc-terrsys.de`

Combining numerical models, which simulate water and energy fluxes in the subsurface-land surface-atmosphere system in a physically consistent way, becomes increasingly important to understand and study processes at compartmental boundaries and interdependencies of states across these boundaries. Complete state evolutions generated by such models, when run at highest computationally possible resolutions while incorporating as many processes as attainable, may be regarded as a proxy of the real world, which can be used to test hypotheses on functioning of the coupled terrestrial system and may serve as source for virtual measurements to develop data-assimilation methods. Such simulation systems, however, face severe problems caused by the vastly different scales of the processes acting in the compartments of the terrestrial system. This study is motivated by the development of cross-compartmental data-assimilation methods, which face the difficulty of data scarcity in the subsurface when applied to real data. With appropriate and realistic measurement operators, the virtual reality not only allows taking virtual observations in any part of the terrestrial system at any density, thus overcoming data-scarcity problems of real-world applications, but also provides full information about true states and parameters aimed to be reconstructed from the measurements by data assimilation. We have used the Terrestrial Systems Modelling Platform TerrSysMP to set up the virtual catchment for a regional terrestrial system oriented at the Neckar catchment in southwest Germany. We find that the simulated catchment is in many aspects quite close to reality concerning, e.g., atmospheric boundary-layer height, precipitation, and runoff. But also discrepancies become apparent both in the ability of such models to correctly simulate certain processes and the realism of the results of some observation operators like the SMOS and SMAP sensors. In a succeeding step, we will use this virtual reality to generate observations in all compartments of the system for coupled data assimilation. The data assimilation will have to rely on a coarsened and simplified version of the model system because ensembles with up to 100 members must be generated in parallel.

1 Introduction

Numerical models, which simulate water and energy fluxes in the subsurface-land surface-atmosphere system in a physically consistent way, are important ingredients to climate and weather prediction, flood forecasting, water resources management, agriculture, and water quality control (Shrestha *et al.* 2014; Larsen *et al.*, 2014; Simmer *et al.*, 2015)[1–3]. Recent developments embrace the combination of coupled land surface-subsurface models with data-assimilation (DA) techniques in order to obtain real-time forecasts of the development of the coupled system (Shi *et al.*, 2014, 2015; Ridler *et al.*, 2014; Kurtz *et al.*, 2016)[4–7]. Different data-assimilation philosophies and methods exist encompassing nudging, variational, and sequential ensemble-based methods (e.g., Kalnay 2003)[8]. Yet the development and choice of appropriate data-assimilation methods is limited by the availability of observational information across all compartments as well as by appropriately characterised real-world systems. The use of virtual reality systems that provide a virtual truth and deliver full information about the subsurface-land surface-atmosphere system can be key for the development of a consistent data-assimilation framework.

The application of the virtual-reality concept for developing data-assimilation systems has the following main advantages:

(I) For the prediction of fluxes and states in a real-world subsurface-land surface-atmosphere system, we do not know whether the errors in our predictions are due to incomplete model specification (model structural errors) like wrong parameterisations, erroneous model parameters or measurement errors, or defects in data assimilation. Working with a virtual truth allows separating model structural error, both in the coupled terrestrial model and in the measurement operators from other error sources, like the generation of the ensemble or the choice of the data assimilation filters. Working only with real-world data would hardly allow the attribution of prediction errors to their causative sources.

(II) The immense scale complexity of terrestrial systems renders already state estimation very challenging despite the wealth of observations available nowadays from *in situ* and remote sensing. Available measurements never cover all locations and times, thus only parts of the heterogeneities can be probed. It is not always clear whether observations contain enough information with the appropriate spatial and temporal resolution to substantially improve predictions. With the virtual catchment we circumvent data scarcity in data-assimilation problems. The virtual-reality approach allows generating arbitrarily long time series and high-density observations to work with, and it facilitates comparing estimated states and parameters to the truth everywhere at all times. This is especially true for coupled subsurface-land surface-atmosphere systems, for which data assimilation has not been evaluated and tested until now.

(III) Studies on data assimilation using different observation densities, observation errors, and data types will provide insight into data value and thus contribute to the development of better monitoring designs for improved predictions of states and fluxes.

(IV) Virtual realities are unique testbeds to develop and validate data-assimilation methods as well as the impact of the underlying assumptions (e.g., Gaussianity). In particular, weakly and fully coupled data assimilation can be compared. In weakly coupled data assimilation, sequential data assimilation is done for multiple compartments of the terrestrial system (e.g., groundwater, land surface, and atmosphere), but observations made for a given compartment only update states (and possibly parameters) within that compartment and not in other compartments. In fully coupled data assimilation, by contrast, observations are available for multiple compartments of the terrestrial system and are used to update states (and possibly parameters) across different compartments of the terrestrial system through cross-correlations, which may be inferred from ensemble statistics of the fully coupled system.

We present the development and testing of a catchment-based virtual-reality system as

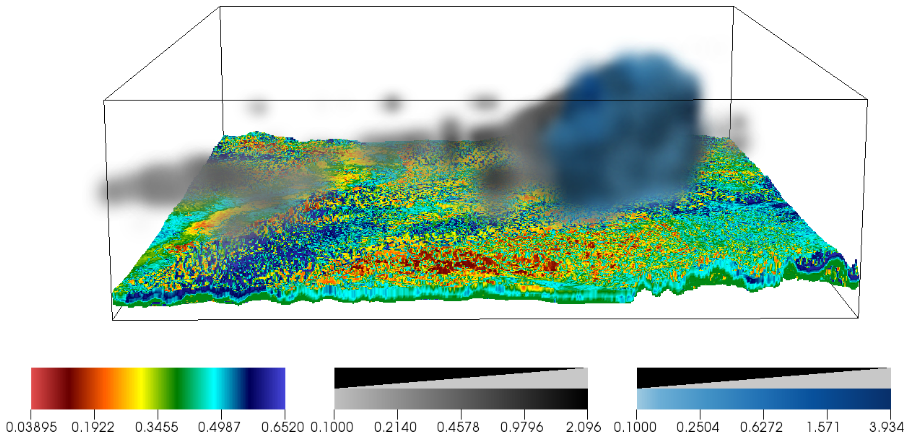

Figure 1. Snapshot of the modelling system showing clouds (grey), precipitation (blue) and volumetric soil moisture (coloured) for the entire domain. In the atmosphere the rain appears near the centre of the domain, while some smaller clouds are visible elsewhere. The larger differences in soil moisture are caused by differences in the physical properties of the soil; in addition regions that have not received any rain for a while are generally more dry near the surface.

a first step in the development of a (weakly or strongly) coupled data-assimilation system. Our virtual reality (VR) represents the fully coupled terrestrial system from the bedrock to the upper atmosphere covering the catchment of a higher-order river (length 380 km, area 14000 km^2) with a buffer zone surrounding it, in which we simulate - as realistically as possible - the evolution of states in all compartments and the water and energy fluxes throughout the system.

This virtual reality provides an artificial truth, from which arbitrary (virtual) observations can be made and which avoids the lack of information for a thorough evaluation of data-assimilation systems. Towards this ends, the model of the subsurface-land surface-atmosphere system needs to represent the strong spatial variability especially of the land component, which affects the overall system behaviour due to nonlinear couplings and feedbacks. Setting up and executing a fully coupled simulation comes with challenges. Without a sufficiently detailed representation of the relations and dependencies of fluxes, states, and parameters in different compartments, the applicability of any data-assimilation method derived from the simulation will be questionable. Since for a chosen size of the catchment, IT-requirements limit the spatial resolution of the fully coupled model, a balance needs to be found between the detail in process representation and computational feasibility. Accordingly, for some processes appropriate upscaling rules or apparent parameters need to be introduced and applied in order to guarantee that the interrelations between different observations and forcings are adequately represented.

2 Methods and Technical Aspects

The virtual catchment we chose for our studies is the Neckar catchment in southwestern Germany which was chosen as a typical central European river catchment of medium size. Other advantages include that the catchment has an average groundwater-table depth that is in the root-zone of most larger plants, but can be too deep for smaller grasses and crops to connect to it for water consumption. This leads to distinctive behaviours of the system, particularly under dry conditions. Furthermore, the Neckar flows through various different landscapes with its source located between the black forest and Swabian Alb mountains, then flowing mainly through hilly terrain to the flat upper Rhine valley where it reaches the Rhine near Mannheim. The land-use in the region is also typical and dominated by agriculture while forests can be found especially in the mountainous regions. Urban areas, which are specially difficult to model are replaced by simple bare-soil patches which may not be realistic, but for the purpose of our studies not critical. In order to use the VR as a surrogate reality we need a time-series of several years to cover all relevant different weather conditions, including floods and droughts in order to establish a system that is able to simulate all compartments and fluxes between them with reasonable accuracy.

To create the virtual catchment, we used the Terrestrial Modelling Platform (TerrSysMP) which couples the atmospheric model COSMO, which is used by the German Weather Service (DWD) for weather forecasts over central Europe, to the land-surface model CLM (Community Land Model, developed at NCAR), and the 3D-groundwater model ParFlow (developed at the Colorado School of Mines). The coupling is realised via the external coupler OASIS3-MCT. The whole system was run on JUQUEEN, which was chosen not only because the VR already needs extensive compute time, but also because the ensemble simulations needed for the actual DA experiments require a massively parallel infrastructure. Since a long time-series was needed for the VR, we continuously had jobs running on JUQUEEN, automatically restarting after seven days of simulation were completed which took about 6-10 hours. This range in simulation time is caused mainly by the hydrological model ParFlow which uses an iterative approach to solve the Richards equations, which requires more compute time during periods with strong rainfall events in order to handle the evolving strong gradients in soil moisture. Of all the model components CLM is the fastest component, needing only few cores to run on while outpacing COSMO and ParFlow. COSMO scales very well and was run using the same number of cores as ParFlow. Except for very dry conditions, COSMO would still slightly outpace ParFlow. In total we used 256 nodes, which is still rather small for JUQUEEN projects, but it was run almost continuously (not considering waiting times) for about 150 days to create a seven-year timeseries of the virtual reality. A 5 year spin-up (the bare minimum necessary) had to be performed before the system is in approximate equilibrium and the production run could be started.

The actual setup was realised with a rectangular COSMO domain covering an area of 57,850 km^2, which encompasses the whole Neckar catchment of 14,000 km^2. This results in 256×230 grid-points with a spacing of 0.01 deg (about 1.1 km) with 50 vertical layers which allows for directly simulating deep convection. CLM and ParFlow domains have to match exactly because both share the same grid for the upper 10 soil layers and were chosen slightly smaller to exclude the boundary zone of COSMO where the lateral forcing of the atmosphere by another, larger-scale atmospheric model dominates. CLM

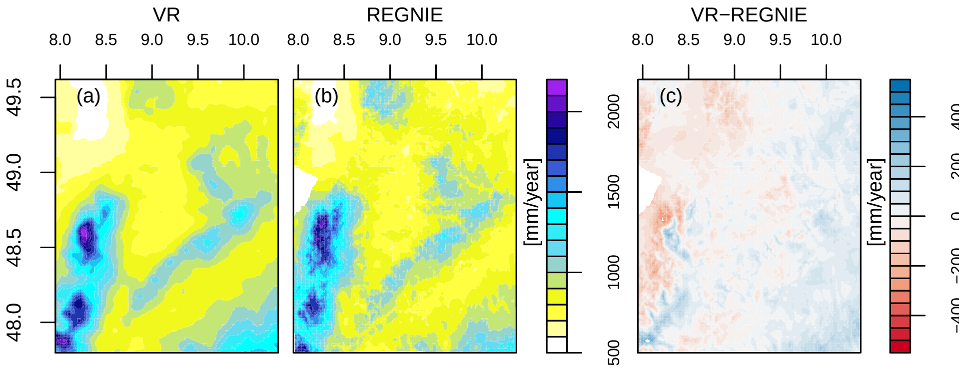

Figure 2. Results from the seven year timeseries of hourly precipitation. The left image shows average yearly precipitation for the virtual catchment. The middle image shows the same for the raster dataset REGNIE[12] which is based on observations while the right image shows the differences between the VR and REGNIE.

and ParFlow were run at 400 m grid spacing resulting in 605 $\times$ 535 grid-points and 50 vertical layers for ParFlow. In total the VR has 22.364.500 gridpoints.

The different models need several inputs for their setup. COSMO needs lateral forcings to drive the model, which we got directly from DWD from their 2.8 km forecast analysis archive which also includes all static fields such as topography. For CLM we used the topography provided by the European Environment Agency EEA[9] and the 2006 Corine Land Cover Data Set[11]. From the MODIS database we derived leaf area index values for four of the land-use classes that CLM uses: broad-leaf and needle-leaf trees, crops and grasslands. For the soil we used the BÜK soil map from the Federal Institute for Geosciences and Natural Resources - BGR[10] from which the soil parameters needed for ParFlow and CLM were derived.

3 Results

In order to showcase the VR we present a snapshot in Fig. 1 which shows parts of all compartments with clouds, rain and soil moisture showcased. One focus of the analysis of the extensive VR ($\sim$35 TB are required for hourly system states) due to the hourly output, was to make sure, that this dataset is suitable to be used as an artificial reality. In some areas we expected rather poor results such as river discharge, because the Neckar river is heavily managed, which is not considered in our setup and also because in ParFlow rivers develop on the gridsize which at 400 m is much larger than most rivers in the domain. However, these problems can be mostly solved by going to a smaller gridsize, which is planned in the future. Differences in discharge will, however, have only very limited impact on the behaviour of the system, while other variables, especially related to the atmosphere, much more impact the system state as they act as the primary forcing for both CLM and ParFlow. Precipitation in particular is very important as it impacts many different processes and terrestrial system states, from evaporation to soil moisture and groundwater recharge.

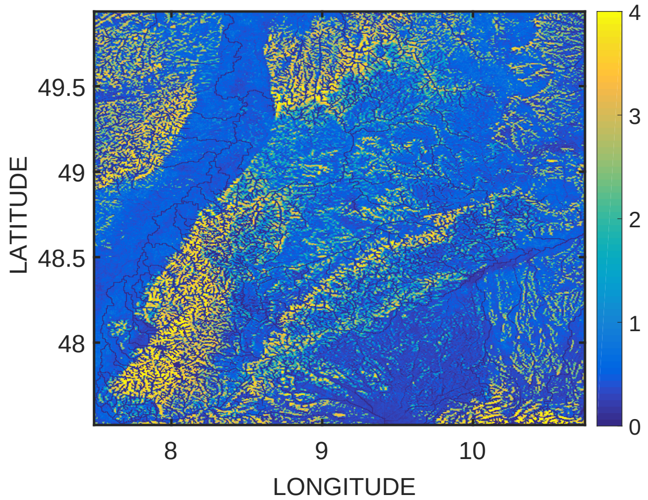

Figure 3. Map of the distance to groundwater for the entire domain for May 2007. The colour scale is modified and shows log(distance to groundwater+1) in order to give more detail at the shallow end of the spectrum.

In Fig. 2 the results from our VR are compared to the observation-based gridded dataset REGNIE[12]. While there are in general areas, especially near the boundary, with larger differences, they are rather small in most of the Neckar catchment. There some smaller discrepancies common to most atmospheric models, such as a slight underestimation of convective rainfall in the summer.

However, even the best precipitation input is useless if the setup of the subsurface is poor and does not produce realistic groundwater levels or soil moisture. In Fig. 3 the distance to groundwater is shown (with modified colour scale to better show differences at small values). Very shallow groundwater table depths are found in the river valleys and very deep groundwater in the mountain regions including steep slopes. There are also discrepancies between simulated soil moisture and soil moisture derived from satellites, which currently cannot yet be attributed to the model or the satellite retrieval algorithm.

In summary, the VR is able to fulfil its role and observations extracted from it would have a good enough quality to be used in the planned DA experiments.

4 Outlook

There is a follow-up version planned for the virtual catchment with the resolution of the surface and sub-surface components increased to 100 m. This will improve the modelling of discharge for individual hill-slopes and for narrow valleys. In order to realise this, we have to upgrade the CLM component to a newer version, as CLM3.5 is not able to run at the required resolution while also performing I/O. Production runs are planned in January 2018.

Even before this new VR is generated we will implement the first version of our data assimilation system, which will focus on experiments related to weakly coupled data assimilation, where measurements from one compartment only update states in the same compartment. This DA system will be based on an ensemble that we are currently testing to make sure that the ensemble spread (the states covered by the different members) is sufficient for all variables and compartments. The second version of the DA system is planned to start in early 2019 and will use virtual observations from the latest virtual catchment. This setup will include strongly coupled DA as well as parameter estimation experiments. Towards the end of 2019 apply all newly developed DA methods to a real word catchment ,the Rur catchment, in order to test our developments under real world conditions.

The ensemble runs and the data assimilation runs are even more demanding than the VR run, however, time periods will be shorter and parallel scaling will be even better.

Finally, we strive to make all datasets we produce available to everyone (with special focus on the VR and the possibility to create virtual observations).

Acknowledgements

Foremost, we want to acknowledge all other members of the FOR2131 research unit not contained in the author list (`http://for2131.de/`). Furthermore we like to thank Fabian Gasper, Ketan Kulkarni, Wendy Sharples, and Klaus Görgen from the SimLab TerrSys for their great technical support related to the use of the JSC supercomputers. The computations (of which we have shown only one sample) were performed with a grant of computer time provided by the Gauss Centre for Supercomputing; since 2013 we were awarded over 140 million core-h on JUQUEEN and 10 million core-h on JURECA for which we are very grateful.

References

1. P. Shrestha, M. Sulis, M. Masbou, S. Kollet, and C. Simmer, *A scale-consistent terrestrial systems modeling platform based on COSMO, CLM, and ParFlow*, Monthly Weather Review **142(9)**, 3466-3483, 2014, doi:10.1175/MWR-D-14-00029.1.
2. M. A. D. Larsen, J. C. Refsgaard, M. Drews, M. B. Butts, K. H. Jensen, J. H. Christensen, and O. B. Christensen, *Results from a full coupling of the HIRHAM regional climate model and the MIKE SHE hydrological model for a Danish catchment*, Hydrol. Earth Syst. Sci. **18**, 4733-4749, 2014, doi:10.5194/hess-18-4733-2014.
3. C. Simmer, I. Thiele-Eich, M. Masbou, W. Amelung, S. Crewell, B. Diekkrueger, F. Ewert, H. J. Hendricks Franssen, A. J. Huisman, A. Kemna, N. Klitzsch, S. Kollet, M. Langensiepen, U. Loehnert, M. Rahman, U. Rascher, K. Schneider, J. Schween, Y. Shao, P. Shrestha, M. Stiebler, M. Sulis, J. Vanderborght, H. Vereecken, J. van der Kruk, T. Zerenner, and G. Waldhoff, *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**, 17651787, 2015, doi:10.1175/BAMS-D-13-00134.1.
4. Y. Shi, K. J. Davis, F. Zhang, C. J. Duffy, and X. Yu, *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, 2014, doi:10.1002/2013wr014070.

5. Y. Shi, K. J. Davis, F. Zhang, C. J. Duffy, and X. Yu, *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, 2015, doi:10.1016/j.advwatres.2015.06.009.
6. M. E. Ridler, N. van Velzen, S. Hummel, I. Sandholt, A. K. Falk, A. Heemink, and H. Madsen. *Data assimilation framework: linking an open data assimilation library (OpenDA) to a widely adopted model interface (OpenMI)*, Environ. Modell. Softw. **57**, 7689, 2014, doi:10.1016/j.envsoft.2014.02.008.
7. W. Kurtz, G. He, S. J. Kollet, R. M. Maxwell, H. Vereecken, and H. J. Hendricks-Franssen, *TerrSysMP-PDAF version 1.0): a modular high-performance data assimilation framework for an integrated land surface-subsurface model*, Geoscientific Model Development **9(4)**, 1341-1360, 2016, doi:10.5194/gmd-9-1341-2016.
8. E. Kalnay, *Atmospheric modeling, data assimilation and predictability*, Cambridge University Press, 341, 2003.
9. `http://www.eea.europa.eu/data-and-maps/data/eu-dem`
10. `http://www.bgr.bund.de/DE/Themen/Boden/Informationsgrundlagen/Bodenkundliche_Karten_Datenbanken/BUEK1000/buek1000_node.html`
11. `http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster-3`
12. `https://www.dwd.de/DE/leistungen/regnie/regnie.html`