Coupling Groundwater, Vegetation, and Atmospheric Processes: A Comparison of Two Integrated Models

MAURO SULIS, JOHN L. WILLIAMS, PRABHAKAR SHRESTHA, AND MALTE DIEDERICH
Meteorological Institute, University of Bonn, Bonn, Germany

CLEMENS SIMMER
Meteorological Institute, University of Bonn, Bonn, and Centre for High-Performance Scientific Computing in Terrestrial Systems, Geoverbund Aachen, Bonn, Cologne, and Jülich (ABC/J), Jülich, Germany

STEFAN J. KOLLET
Institute for Bio- and Geosciences: Agrosphere (IBG-3), Research Centre Jülich, and Centre for High-Performance Scientific Computing in Terrestrial Systems, Geoverbund Aachen, Bonn, Cologne, and Jülich (ABC/J), Jülich, Germany

REED M. MAXWELL
Integrated GroundWater Modeling Center, Department of Geology and Geological Engineering, Colorado School of Mines, Golden, Colorado

(Manuscript received 7 July 2016, in final form 14 February 2017)

ABSTRACT
This study compares two modeling platforms, ParFlow.WRF (PF.WRF) and the Terrestrial Systems Modeling Platform (TerrSysMP), with a common 3D integrated surface–groundwater model to examine the variability in simulated soil–vegetation–atmosphere interactions. Idealized and hindcast simulations over the North Rhine–Westphalia region in western Germany for clear-sky conditions and strong convective precipitation using both modeling platforms are presented. Idealized simulations highlight the strong variability introduced by the difference in land surface parameterizations (e.g., ground evaporation and canopy transpiration) and atmospheric boundary layer (ABL) schemes on the simulated land–atmosphere interactions. Results of the idealized simulations also suggest a different range of sensitivity in the two models of land surface and atmospheric parameterizations to water-table depth fluctuations. For hindcast simulations, both modeling platforms simulate net radiation and cumulative precipitation close to observed station data, while larger differences emerge between spatial patterns of soil moisture and convective rainfall due to the difference in the physical parameterization of the land surface and atmospheric component. This produces a different feedback by the hydrological model in the two platforms in terms of discharge over different catchments in the study area. Finally, an analysis of land surface and ABL heat and moisture budgets using the mixing diagram approach reveals different sensitivities of diurnal atmospheric processes to the groundwater parameterizations in both modeling platforms.

1. Introduction
The importance of soil moisture in regulating short-term weather patterns and in influencing long-term climate dynamics has been widely recognized in modeling studies and observational analyses in recent years (e.g., Delworth and Manabe 1988; Brubaker and Entekhabi 1996; Findell and Eltahir 1997; Eltahir 1998; Taylor and Lebel 1998; Koster et al. 2003, 2004; Ek and Holtslag 2004; Schär et al. 2004; Seneviratne et al. 2006; Santanello et al. 2009; Taylor et al. 2011; Orth and Seneviratne 2012). The interaction between soil moisture and the atmosphere evolves through exchanges of energy, mass, and momentum at the land surface, which are...
in turn influenced by topography, soil properties, and vegetation characteristics (Sellers et al. 1995; Pielke 2001; Teuling et al. 2006, 2010; Los et al. 2006). Recently, studies suggested that groundwater may function as a spatial organizor of soil moisture by maintaining wetter conditions in low-lying areas via lateral flow processes, for example, along river corridors (York et al. 2002; Bierkens and van den Hurk 2007; Anyah et al. 2008; Maxwell and Kollet 2008; Kollet and Maxwell 2008; Decker et al. 2013; Fan et al. 2013), and as a temporal buffer for soil moisture and river discharge by its delayed and small-amplitude response to weather and climate fluctuations (Lo and Famiglietti 2010; Miguez-Macho and Fan 2012; Sulis et al. 2012; Rahman et al. 2014).

Growing attention has been paid to the linkage between groundwater and root-zone soil moisture with an increased focus on the representation of subsurface hydrology in large-scale land surface models (LSMs) using more sophisticated parameterizations (Lee and Abriola 1999; Yeh and Eltahir 2005; Niu et al. 2007; Miguez-Macho et al. 2007). These LSMs typically use a prescribed groundwater reservoir that is either loosely or dynamically coupled to surface storage (e.g., rivers, wetlands, and soil moisture). An increasing number of subsurface models, which are based on three-dimensional or mixed equations (one-dimensional for the vadose zone and two-dimensional for groundwater) for variably saturated flow, have been coupled with detailed land surface parameterizations for surface heat, momentum, and vapor transfer (Maxwell and Miller 2005; Rigon et al. 2006; Kollet and Maxwell 2008; Shen et al. 2013; Niu et al. 2014). In these studies the atmosphere is represented only by offline forcings for the land and subsurface, and two-way interactions between surface and atmosphere are neglected.

Modeling efforts have recently led to a new simulation paradigm where the dynamic feedback between land and atmosphere is described by solving adapted forms of the three-dimensional equations governing both compartments (Maxwell et al. 2007, 2011; Göchis et al. 2013; Shrestha et al. 2014; Butts et al. 2014). Using different coupling technologies [e.g., Ocean Atmosphere Sea Ice Soil (OASIS), Model Coupling Toolkit (MCT), Open Modeling Interface (OpenMI), and WRF-Hydro] and strategies related to, for example, coupling frequency, data exchange, and spatial mapping, these integrated modeling systems dynamically simulate soil moisture variability produced by terrain slope and its effect on boundary layer evolution and mesoscale circulations (Shrestha et al. 2014), effects of subsurface processes on atmospheric boundary layer (ABL) development (Maxwell et al. 2007), and the influence of groundwater processes on the regional climate (Butts et al. 2014). A small but growing body of studies uses these systems to investigate aspects of coupled hydrological and meteorological processes. For instance, Williams and Maxwell (2011) showed how uncertainty in subsurface hydraulic conductivity propagates into atmospheric variables, specifically wind speed. Williams et al. (2013) presented a wind speed forecasting system using fully coupled surface–subsurface and atmospheric simulations together with an ensemble Kalman filter data assimilation technique. Larsen et al. (2014) assessed the influence of the data transfer interval between the atmospheric (HIRHAM regional climate model) and subsurface (MIKE-SHE hydrological model) components. Rahman et al. (2015) investigated the sensitivity of the atmosphere to subsurface hydrodynamics under convective conditions by performing fully coupled ensemble simulations with different model configurations for water-table dynamics.

Intercomparison studies in the atmospheric (Gates et al. 1999; Koster et al. 2006), land surface (Henderson-Sellers et al. 1995; Yang et al. 1995; Dirmeyer et al. 1999; Chen et al. 2014; Best et al. 2015), and subsurface hydrologic communities (Smith et al. 2004; Sulis et al. 2010; Maxwell et al. 2014; Kollet et al. 2017) have pointed out similarities and differences in formulations for and understanding of the complex processes within the different parts of the terrestrial system and documented some distinguishing features of the used modeling tools (e.g., land surface parameterizations, domain discretization, and coupling approaches). These studies informed users about key model features and led to a better understanding of their limitations when used, for example, as members of a multimodel ensemble (Knutti et al. 2010).

This study compares two fully integrated (from groundwater to atmosphere) models, ParFlow.WRF (PF.WRF) and the Terrestrial Systems Modeling Platform (TerrSysMP), that have been recently presented in the literature by Maxwell et al. (2011) and Shrestha et al. (2014), respectively. PF.WRF couples the atmospheric Weather Research and Forecasting (WRF) Model with the three-dimensional variably saturated ParFlow hydrologic model. TerrSysMP consists of the regional climate and weather forecasting Consortium for Small-Scale Modeling (COSMO) model sequentially coupled with ParFlow via the Community Land Model, version 3.5 (CLM3.5). Whereas differences in the land surface energy partitioning and atmospheric boundary layer parameterizations in the two integrated models have been documented in previous works using coupled (Jin et al. 2010; Lu and Kueppers 2012) and offline atmosphere (Schwitalla et al. 2011) simulations, this study attempts to provide such an assessment by focusing on their sensitivity to water-table depth and groundwater formulations (i.e., with/without lateral flow). Additional objectives are (i) to assess the degree of realism by
comparing the model responses against measurements and (ii) to evaluate their integrated rainfall–runoff response at the catchment scale. In this context, it is important to remark that the guiding principle of comparing the two models is to use differences and similarities in their simulated results to get an increased understanding of the complex coupled processes of the integrated energy and water cycles. Thus, no attempts are made to identify a superior model, as this would require systematic benchmarking across different weather situations, climate regimes, models configurations, and exhaustive observational dataset.

To address the aforementioned objectives, we compare both idealized diurnal simulations and hindcast runs over a domain located in western Germany. The idealized simulations with bare soil and crop land surface are used to study the propagation of heterogeneities between compartments and to investigate the sensitivity of land surface and atmospheric processes to fluctuations in the water-table depth. The hindcast simulations using a small ensemble of initial conditions are compared to observations from eddy covariance towers, soil moisture retrieved from cosmic-ray probes, and rainfall estimates from polarimetric radars. Finally, the influence of groundwater dynamics simulated with and without lateral flow on diurnal atmospheric processes is studied by evaluating the energy and moisture budgets of the atmospheric boundary layer using the vector representation of the mixing diagrams for different groundwater configurations of both integrated models.

2. Description of the integrated models

This section briefly presents the two integrated models, discusses their coupling features, and summarizes the main differences in the atmospheric and land surface compartments.

a. PF.WRF

PF.WRF (Maxwell et al. 2011) combines the ParFlow hydrologic model (Ashby and Falgout 1996; Jones and Woodward 2001; Kollet and Maxwell 2006) with the mesoscale atmospheric WRF Model (Skamarock and Klemp 2008) via the Noah land surface model (Chen and Dudhia 2001; Ek et al. 2003). WRF is a numerical weather prediction code that solves the fully compressible nonhydrostatic Navier–Stokes equations in three dimensions, vertically discretized in terms of pressure levels. PF.WRF uses the Advanced Research version of WRF (ARW). The Noah LSM, one of several options available in the WRF package, implements a combined surface layer of soil and vegetation with a stress function limited by potential evaporation calculated based on a Penman energy balance approach to simulate terrestrial moisture and energy fluxes. The coupled surface and subsurface hydrodynamics are described in ParFlow by the three-dimensional Richards equation coupled via an overland flow boundary condition with the kinematic wave approximation for surface water flow (Ashby and Falgout 1996; Kollet and Maxwell 2006; Maxwell 2013). The hydrologic model uses the same or a larger time step as the atmospheric model in order to keep computational efficiency while restricting the loss of accuracy, as shown by Maxwell et al. (2011).

b. TerrSysMP

TerrSysMP (Shrestha et al. 2014) is a modular modeling platform comprising the numerical weather prediction COSMO model in a convection-permitting configuration (Baldauf et al. 2011), CLM3.5 (Oleson et al. 2008), and ParFlow. The dynamical core of COSMO solves the compressible Euler equations in terrain-following coordinates with variable vertical discretization using an Arakawa C grid. In CLM3.5, land surface energy fluxes are described using separated vegetation canopies and surface layers discriminating between shaded and sunlit components. The external coupler OASIS, version 3 (OASIS3: Valcke 2013), is used to drive TerrSysMP and control the exchange of fluxes between each component model. OASIS3 [recently updated to OASIS3-MCT by Gasper et al. (2014)] handles the coupling sequence and frequency, exchange of data structures and spatial grids, and the type of spatial transformations (e.g., interpolation techniques) used in the two-dimensional coupled fields. Fluxes can be exchanged taking into account differing spatial and temporal resolutions of the component models. TerrSysMP can run with different combinations of the component models, which are compiled and executed as stand-alone models. Thus, COSMO can be coupled and compiled in a “plug and play” fashion with CLM, CLM with ParFlow (using offline atmospheric forcing), and COSMO coupled both with CLM and ParFlow.

c. Coupling of exchange fluxes

PF.WRF and TerrSysMP use an operator-splitting approach to replace the subsurface hydrology of the LSM with three-dimensional moisture fields calculated by ParFlow. This exchange involves the top four layers in Noah (PF.WRF) and the top 10 layers in CLM (TerrSysMP) with a total soil depth of 2.0 and 3.0 m, respectively. The vertical discretization employed by PF.WRF and TerrSysMP reflects the standard configuration of the respective vegetation schemes (i.e., Noah and CLM), which is also consistent with the way both LSMS parameterize the root distribution. The LSMS
calculate the source/sink term (net precipitation/evapotranspiration) for the Richards equation, which is passed to ParFlow at the coupling time steps.

The two systems differ in their parameterizations and coupling with the land surface. In this study, PF.WRF uses the Mellor–Yamada–Janjic (Mellor and Yamada 1982) ABL scheme, which employs a level-2.5 turbulent kinetic energy closure with local-K vertical mixing. The scheme is tied with the Eta surface-layer similarity scheme (Janjic 1990) to calculate the surface exchange coefficients and the sensible heat, latent heat, and momentum fluxes, which provide the lower boundary condition for the ABL scheme. The land surface scheme provides to the atmospheric model the reflected solar and emitted terrestrial radiation; the latter is calculated using the skin temperature estimated via a single linearized surface energy balance equation representing the combined surface layer of vegetation and soil surface. In TerrSysMP (COSMO), the turbulence scheme for vertical transport is based on a combination of the traditional K closure and a prognostic equation for turbulent kinetic energy (TKE), which is similar to the level-2.5 scheme of Mellor and Yamada (1982). The equations of all second-order moments are taken into account except the time derivatives and transport terms (Baldauf et al. 2011; Raschendorfer 2001). This TKE closure is complemented by the COSMO surface transfer scheme, which is formulated based on the TKE closure scheme itself. The transfer layer extends from the rigid surface at the ground to the first model level. This layer is divided into surface layer and roughness layer, which is further divided into a turbulent roughness layer and a laminar layer. The total resistance of the transfer layer is the sum of the resistances of its three sublayers, from which the transfer coefficients are computed. In TerrSysMP, these transfer coefficients are updated by CLM and computed based on the Monin–Obukhov similarity theory (Shrestha et al. 2014). The albedo (direct and diffuse) and the emitted terrestrial radiation estimated from the bare ground surface and vegetation temperature. The most important differences concerning the discretization, parameterizations, and formulations employed in the PF.WRF and TerrSysMP setups in this study are summarized in Table 1.

### 3. Test cases

#### a. Idealized test cases

Clear-sky simulations are performed for a full diurnal cycle for the surface land-cover types bare soil and crop. The simulation domain covers 64 km × 64 km in the horizontal with a grid spacing of 1 km for each component of PF.WRF and TerrSysMP. For both systems, the atmosphere is vertically discretized into 50 layers with identically stretched grids and a near-surface resolution of 20 m. The subsurface domain in PF.WRF is vertically discretized into 40 layers with 4 layers of increasing depth from 0.1 to 1.0 m near the surface followed by 6 (1.0 m) and 30 (1.35 m) layers of constant depth extending to 40 m below the surface. In TerrSysMP, the same total soil depth (40 m) is discretized using 10 near-surface layers with increasing thickness from 0.02 to 0.10 m followed by 20 constant-depth layers (1.35 m). As discussed in section 2c, the vertical discretization in the upper layers of the subsurface model domains replicates the standard configurations of CLM (TerrSysMP) and Noah (PF.WRF), respectively.

In both model systems, the atmospheres are initialized horizontally with a homogeneous mean wind of $U = 0.5 \text{ m s}^{-1}$ and a slightly stable temperature profile ($T = 300 - 0.005 z \Delta z$ K) for $z < 10 \text{ km}$. The relative humidity is set to 50% for $z < 1.5 \text{ km}$. In the vegetated case, a fixed leaf area index (LAI) value of 2.0 is used in both models. In PF.WRF, the vegetation fraction $\sigma_v$ is consistent with the partitioning of total evaporation into soil evaporation, canopy evaporation, and canopy transpiration, that is, internally calculated in TerrSysMP according to Lawrence et al. (2007). The soil hydraulic characteristics

#### Table 1. Summary of the main differences between PF.WRF and TerrSysMP.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PF.WRF</th>
<th>TerrSysMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiation scheme</td>
<td>CAM3</td>
<td>COSMO</td>
</tr>
<tr>
<td>Atmospheric boundary layer parameterizations</td>
<td>Mellor–Yamada–Janjic (level-2.5 turbulence parameterization)</td>
<td>Mellor–Yamada (level-2.5 turbulence parameterization)</td>
</tr>
<tr>
<td>Land surface parameterizations</td>
<td>Ground evaporation ($\beta$ factor), stomatal resistance (Jarvis type), and surface emission (combined ground and vegetation)</td>
<td>Ground evaporation (soil resistance), stomatal resistance (Ball–Berry type), and surface emission (separated ground and vegetation)</td>
</tr>
<tr>
<td>Top soil vertical discretization</td>
<td>4 layers (0–2.0 m; 0.1/0.3/0.6/1.0)</td>
<td>10 layers (0–3.0 m; 0.02/0.03/0.05/0.07/0.13/0.20/0.30/0.50/0.70/1.0)</td>
</tr>
</tbody>
</table>
are described using the van Genuchten relationships between soil moisture content and pressure head (van Genuchten and Nielsen 1985), with a saturated hydraulic conductivity $K_{sat} = \text{9.1} \times 10^{-6}$ m s$^{-1}$, porosity $\phi = 0.434$, residual moisture content $\phi_{res} = 0.04$, $a = 2.7$, and $n = 1.4$. Hydraulic pressure is initialized with a hydrostatic profile defined by a water table at 2.5 m depth below the land surface. Periodic lateral boundary conditions are used for the atmospheric models, and a no-flow lateral boundary condition is used for the subsurface. For both modeling systems, a time step of 9 s is used for the atmospheric, land surface, and subsurface component (i.e., no temporal subcycling).

b. Real test cases

1) EXPERIMENTAL DESIGN

Simulations are performed over a domain of 150 km $\times$ 150 km encompassing most of the North Rhine–Westphalia region in western Germany, including parts of Belgium, the Netherlands, and Luxembourg (Fig. 1), using the horizontal and vertical discretizations from the idealized case. Heterogeneous land-use information from the MODIS database (land-cover type 5 provided by the Oak Ridge National Laboratory Distributed Active Archive Center) and from a phenology study (Shrestha et al. 2014) is used to characterize surface land cover. Shallow subsurface soil types are taken from the digital soil map of the Food and Agriculture Organization of the United Nations. Soil characteristics are represented using van Genuchten functions for different soil types in the model (Table 2), with parameter values obtained from Schaap and Leij (1998). To alleviate a wet bias in soil moisture in ParFlow due to the coarse lateral discretization (Shrestha et al. 2015), the horizontal hydraulic conductivity is scaled (increased) by two orders of magnitude. The global database of Gleeson et al. (2011) provides for the deeper subsurface characterization. In TerrSysMP, a time step of 10 s is used for the atmospheric components and 900 s for the land surface and subsurface, which matches the update frequency of the radiation parameterization in the atmospheric component. PF.WRF uses a time step of 6 s in the atmosphere and 300 s in the

---

**TABLE 2. Soil hydraulic parameters for shallow subsurface of the North Rhine–Westphalia domain:** $K_{sat}$ is the saturated hydraulic conductivity, $\phi$ is the porosity, $\phi_{res}$ is the relative residual moisture content, and $a$ and $n$ are the van Genuchten parameters.

<table>
<thead>
<tr>
<th>Soil texture</th>
<th>$K_{sat}$ (m s$^{-1}$)</th>
<th>$\phi$</th>
<th>$a$</th>
<th>$n$</th>
<th>$\phi_{res}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clay</td>
<td>$1.72 \times 10^{-6}$</td>
<td>0.4701</td>
<td>2.1</td>
<td>2.0</td>
<td>0.21</td>
</tr>
<tr>
<td>Clay loam</td>
<td>$9.44 \times 10^{-7}$</td>
<td>0.4449</td>
<td>2.1</td>
<td>2.0</td>
<td>0.17</td>
</tr>
<tr>
<td>Loam</td>
<td>$1.38 \times 10^{-6}$</td>
<td>0.4386</td>
<td>2.0</td>
<td>2.0</td>
<td>0.15</td>
</tr>
<tr>
<td>Sandy loam</td>
<td>$4.38 \times 10^{-6}$</td>
<td>0.4071</td>
<td>2.7</td>
<td>2.0</td>
<td>0.10</td>
</tr>
</tbody>
</table>
subsurface, with radiation updates every 60 s. In each case, atmospheric variables are averaged over the time step of the hydrological component and passed at each subcycle. Initial and lateral boundary conditions for both atmospheric models are obtained from the operational weather forecast model COSMO-DE of the German Weather Service (DWD).

2) LAND SURFACE–SUBSURFACE INITIALIZATION

Initial soil moisture and soil temperature were obtained for both models from spinup runs with the stand-alone hydrological model ParFlow.CLM. ParFlow.CLM was run repeatedly using the COSMO-DE reanalysis atmospheric forcing data from 2011 to 2012 until a dynamic equilibrium was reached. The changes in soil water storage and subsurface temperature between two successive iterations were examined at two specific dates (21 May 2012 and 1 April 2011) 2 months prior to the short-term events of interest for this study (21–26 July 2012 and 1–5 June 2011, respectively). A spatially distributed analysis revealed that for both dates of examination, changes in total soil water storage and temperature were below 0.4% and 0.001% for approximately 90% and 98% of the model domain. The initial states obtained from the stand-alone ParFlow.CLM spinup runs were then used to initialize 2-month-long, fully coupled simulations with both model systems to obtain model-adjusted initial conditions. In the case of PF.WRF, the initial states obtained from the ParFlow.CLM spinup were interpolated onto the coarser vertical discretization used in PF.WRF. To account for the uncertainty in this initialization process, ensemble initial conditions for both short-term event runs were created by sampling backward in time (0, 24, 48, 72, and 96 h) from the fully coupled, 2-month runs.

3) OBSERVATIONS

Observations were collected from sites within the Rur catchment (Fig. 1), which constitutes the central research area for the Transregional Collaborative Research Centre 32 (TR32; Graf et al. 2013; Mauder et al. 2013; Simmer et al. 2015). Energy balance components (i.e., radiation and response fluxes) were monitored at Selhausen (crop at 50°52′N, 6°26′E), Merzenhausen (crop at 50°55′N, 6°17′E), and Rollesbroich (meadow at 50°55′N, 6°17′E). Radiation data were retrieved from four-component net radiometers observing downward and upward shortwave and longwave radiation (CNR04, Kipp and Zonen, Delft, the Netherlands) at 2.5 m above ground every second. Ten-minute averages derived from these measurements were used in this study. At the same stations, eddy covariance systems continuously monitored latent and sensible heat fluxes, which were analyzed for 30-min intervals. Wind speed and air temperature were measured with an ultrasonic anemometer (CSAT3, Campbell Scientific, Inc., Logan, Utah, United States). Water vapor was observed with an open-path infrared gas analyzer (LI7500, LI-COR Inc., Lincoln, Nebraska, United States) at a sampling rate of 20 Hz.

Neutron flux observations with cosmic-ray probes (type CRS1000; HydroInnova LLC 2009) were used to derive soil water content, calibrated by gravimetric measurements from field campaigns (Bogena et al. 2013; Baatz et al. 2014). The probes were installed at field sites representative for the main land-use types (coniferous and deciduous forest, grassland, and crops) within the Rur catchment. The cosmic-ray probe measurements have an effective horizontal footprint with a radius of 300 m and a variable vertical extent (between 10 and 15 cm for most of the stations) estimated with the formulation proposed by Franz et al. (2012). Hourly values of soil water content were averaged over the day for noise reduction.

Rainfall estimates were obtained using observations from two polarimetric weather radars operating at X-band wavelength (BoXPol and JuXPol). The radars were calibrated and corrected for partial beam blockage using consistent relations between observed specific differential phase shift $K_{DP}$ and radar reflectivity at horizontal polarization $Z_H$ in rainfall over several months (Diederich et al. 2015a). Differential phase shift was also used to correct for path-integrated attenuation in the $Z_H$ measurement. Rain rates were calculated using power-law relations between rain rate and $K_{DP}$ (for rain rates above 8 mm h$^{-1}$) and $Z_H$ (for rain rates below 8 mm h$^{-1}$), following Diederich et al. (2015b). Comparisons with local rain gauge measurements showed a slightly larger rain-rate estimate from the radar data by approximately 20% on 5 and 6 June 2011, which can be explained by melting layer contamination at radar beam height and potential rain gauge undercatch due to wind. The estimates from both radar systems were merged onto the COSMO-DE grid taking into account for each radar the signal sensitivity above background noise, attenuation, and the distance of the observation from the ground. Rainfall accumulations were generated by temporally interpolating the 5-min measurements via advection estimated by cross-correlating subsequent scans.

4) Results

a) Idealized test cases

The simulations of land surface energy fluxes and ABL thermodynamics by PF.WRF and TerrSysMP are compared separately for bare and vegetated soil in order to capture potential systematic differences between
both model systems in representing land surface energy partitioning. It is important to note that the LSM (i.e., Noah and CLM3.5) is one of the key differences between the two integrated models and that their relative response has already been compared (e.g., Lu and Kueppers 2012) for multiyear simulations. The idealized test cases presented in this work are designed to quantify such differences in a fully integrated manner by tracking their propagation through all components of the soil–vegetation–atmosphere system and allowing the evaluation of their sensitivity to water-table variations.

1) BARE SOIL LAND COVER

Figure 2 compares the simulated diurnal flux variations for net radiation (shortwave and longwave components) and latent heat (LE) and sensible heat $H$ for bare soil. Net shortwave radiation, which depends on the soil via its albedo and the state of the atmosphere via cloud cover in particular, matches well between PF.WRF and TerrSysMP. Daytime net longwave radiation differs slightly because the soil temperature on which the upward longwave component depends differs between both modeling systems; that is, the skin temperature in PF.WRF is higher than the 2-cm topsoil layer temperature in TerrSysMP.

Latent and sensible heat differ considerably between PF.WRF and TerrSysMP, with the former simulating a much lower LE and the latter a much lower $H$. This discrepancy results from different formulations for soil evaporation. PF.WRF (via Noah) implements a soil-moisture-dependent $\beta$-factor approach to compute actual evaporation from potential evaporation (Chen and Dudhia 2001). TerrSysMP (via CLM3.5) uses an empirical soil resistance term as an exponential function of soil wetness in the topsoil layer (Sellers et al. 1992b). Bare-soil evaporation formulations are continuously revised in LSMs (Sakaguchi and Zeng 2009; Tang and Riley 2013b), and their strong impact on the regional and global surface energy and water budgets were demonstrated recently by Tang and Riley (2013a). Our results (Fig. 2) also reflect such an impact in significantly different Bowen ratios: $b = 3.2$ for PF.WRF and $b = 1.2$ for TerrSysMP at 1200 UTC. In TerrSysMP, soil moisture at the land surface exhibits a stronger diurnal variation (results not shown for the sake of brevity) and a more pronounced capillary rise effect at the end of the day. This response, which is triggered by the strong soil pressure head gradient induced by the higher diurnal variation of LE, is amplified by the finer vertical discretization near the surface used in TerrSysMP compared to PF.WRF (see section 2c).

There is also a temporal lag in $H$ between PF.WRF and TerrSysMP during the rising part of the diurnal cycle caused by different soil–atmosphere temperature gradients in the models at 0600 UTC as a consequence of a more pronounced nighttime atmospheric cooling (with respect to the soil) in PF.WRF. The different cooling tendencies result from different vertical mixing lengths (i.e., the scaling of turbulent diffusion coefficient for momentum and scalar fluxes) under the usually stable nighttime conditions (PF.WRF imposes a lower limit to the turbulent length scale) that distribute energy differently within the ABL. Accordingly, a weaker but wider inversion layer is simulated by COSMO compared to WRF.
The diurnal evolution of potential temperature interpolated at fixed pressure levels indicates that PF.WRF simulates warmer atmospheric conditions compared to TerrSysMP (Fig. 3a), with a stronger nighttime temperature inversion. The warmer vertical profiles simulated by PF.WRF during daytime (Fig. 3b) result from the higher $H$ flux, which also produces a deeper ABL of 1500 m compared to 1150 m for TerrSysMP at 1500 UTC. PF.WRF simulates a drier and higher ABL throughout the day with an inversion between 810 and 860 hPa (840 and 870 hPa in TerrSysMP) at 1500 UTC (Fig. 3b). The higher ABL is also reflected by the higher TKE simulated by PF.WRF (volume averaged TKE at 1500 UTC $\langle \text{TKE} \rangle_V = 0.30 \text{ m}^2 \text{s}^{-2}$) compared to TerrSysMP ($\langle \text{TKE} \rangle_V = 0.20 \text{ m}^2 \text{s}^{-2}$). This higher TKE is generated by more buoyant thermals triggered by higher surface $H$ fluxes. Note that the stronger LE feedback in TerrSysMP propagates upward into the atmosphere, generating a small discontinuity at 1800 UTC in the humidity profile at the lowest model level (Fig. 3b). This effect is clearly visible at this time of the day as the turbulence decreases and the land surface processes start to decouple from the upper-atmosphere dynamics.

2) VEGETATED LAND COVER

For the vegetated surface, PF.WRF and TerrSysMP also agree well in the simulated net solar radiation (Fig. 4). Thus, both models are again constrained by a similar solar radiation energy budget at the land surface. The net longwave radiation during the daytime compares better between the models than in the bare-soil case, with the skin temperature simulated by PF.WRF
now better matching the vegetation/ground temperature used in TerrSysMP to calculate outgoing longwave radiation. The simulated LE and $H$ fluxes, however, differ again in magnitude and partitioning. The diurnal cycle of LE simulated by TerrSysMP indicates stomatal closure in the late morning (~1100 UTC) and afternoon (~1500 UTC), which is not modeled by PF.WRF. PF.WRF uses a meteorological (or Jarvis type) approach extended by Jacquemin and Noilhan (1990) that attempts to modify a minimum stomatal resistance, defined a priori, through external factors such as moisture and radiation availability. TerrSysMP instead uses the physiological approach included in CLM based on the coupling of leaf photosynthesis and stomatal resistance (Sellers et al. 1992a). Its two-peak diurnal cycle (Fig. 4) is a result of a faster response to atmospheric conditions (e.g., low wind) and high vegetation temperature that trigger a thermal breakdown of metabolic processes (Niyogi and Raman 1997; Dai et al. 2004) with a subsequent stomatal closure. Note that despite being also accounted for in the Jarvis-type approach, such temperature control on stomatal resistance is less pronounced for the setup used and conditions experienced in this work. Moreover, the different plant dynamics should be interpreted in light of a combined effect of using a finer vertical discretization in TerrSysMP (i.e., drier soil matrix potentials in the root zone), which in turn influence the photosynthesis process via the wilting factor.

Figure 5 shows the diurnal evolution of the vertical profiles of potential temperature and vapor mixing ratio at two times of the day. For vegetated land, differences between both models are less pronounced, as also suggested by the Bowen ratios ($b = 0.6$ for PF.WRF and $b = 0.9$ for TerrSysMP at 1200 UTC). The propagation of the land surface signature—as simulated in TerrSysMP (Fig. 4)—into the ABL is clearly visible in Fig. 5a (right). Specifically, the slower heating in the rising part of the diurnal cycle followed by an earlier decrease in the afternoon limits the ABL development in TerrSysMP, which remains shallower and less mixed compared to PF.WRF. A stable boundary layer at the end of the day is evident in both PF.WRF and TerrSysMP simulations.

3) INFLUENCE OF WATER-TABLE DEPTH ON LAND SURFACE AND ATMOSPHERIC PROCESSES

To illustrate the connection between groundwater, land surface, and atmospheric processes simulated by both models, the idealized simulations are repeated while changing the initial water table between 0.5, 1.0, and 5.0 m from its default at 2.5 m. Figure 6 shows the sensitivity of land surface energy partitioning (i.e., Bowen ratio) and ABL height to the initial water-table depth for both land covers. In both PF.WRF and TerrSysMP, a deeper groundwater level consistently increases the Bowen ratio (i.e., larger contribution of $H$ with respect to LE) and the ABL height (i.e., larger turbulence). The initial water-table depth plays a larger role (via ground evaporation) for bare soil, whereas vegetated land mediates (via the root distribution) the soil moisture changes in the top soil layers. For both models, water-table-induced perturbations are dampened by atmospheric turbulence, which reduces changes of the ABL height with respect to variations in land surface energy partitioning.
Despite similar trends, both models exhibit different ranges of sensitivity to water-table variations. For the bare soil, energy partitioning in PF.WRF is much more sensitive to variations of deeper (between 2.5 and 5 m) water-table depth; this holds despite a lower topsoil moisture sensitivity to such variations due to the coarser vertical discretization in PF.WRF. This suggests that the different responses of energy flux partitioning are caused by the way soil moisture enters the soil resistance formulations for ground evaporation in both models. While ABL height increases with lowering the water-table depth in both models, these changes do not mirror the large differences in the sensitivity of the land surface flux partitioning. For vegetated land cover, PF.WRF and TerrSysMP show much the same connection between water-table depth and changes in the energy fluxes because the root distributions ameliorate effects of different near-surface soil moisture profiles. In TerrSysMP, variations in the energy fluxes by changing the water table between 0.5 and 2.5 m only marginally translates into the ABL height change. Thus, the two integrated models may result in different zones where variations in the land surface and atmospheric states are tightly coupled to the fluctuations of the water-table depth.

b. Real test cases

The first hindcast period (21–26 July 2012) covers six consecutive days with almost completely clear-sky conditions over the whole model domain and is thus optimally suited for comparing fluxes and states on a diurnal time scale. The second hindcast period (1–5 June 2011) starts with three clear-sky days followed by two rainy days, including strong convective precipitation (~20 mm h⁻¹) on the fifth day, and is thus well suited for comparing convection initiation and development. The response of the two models is evaluated by comparing for each variable of interest the mean and the spread of the ensemble simulations.
Figure 7 compares the simulated average daily cycle of the radiation balance components (net shortwave and net longwave) with observations from stations at Selhausen (cropland) and Rollesbroich (grassland, locations indicated in Fig. 1). Both models overestimate absorbed shortwave radiation at Selhausen, because the locally observed cloudiness was not reproduced. While TerrSysMP almost perfectly reproduces the observations at Rollesbroich (except for minor effects of observed but not simulated cloudiness during noon), PF.WRF significantly overestimates absorbed solar radiation. This discrepancy may be caused by the lower albedo for grassland (meadow) assumed by PF.WRF. This land-use type, which occupies a small portion of the whole domain (~5%), has an albedo of 0.25 in TerrSysMP and 0.15 in PF.WRF, which relates to the different ways the models calculate the albedo. PF.WRF (via Noah) interpolates between a minimum and maximum value in a lookup table based on the prevalent vegetation classes in the grid cell according to the areal fractional coverage of vegetation. TerrSysMP (via CLM) computes the ground and vegetation albedo for direct and diffuse visible and near-infrared radiation based on lookup tables, which differentiate between soil
color classes under dry and saturated soil conditions and depend additionally on plant physiological properties (i.e., canopy reflectance). Because of this albedo parameterization, TerrSysMP shows a larger influence—albeit rather small in terms of absolute values—to the perturbed initial soil moisture and temperature conditions compared to PF.WRF.

The amplitude of the simulated daytime net longwave radiation significantly exceeds the observations at Selhausen and Merzenhausen for both models because of the missed cloudiness, which leads to higher simulated soil (and vegetation) temperatures. Differences between the simulated and observed radiation components, however, partially cancel each other, resulting in quite well reproduced average diurnal variation in net radiation; daily averaged difference between observed and simulated net radiation ranges from 11 W m⁻² (Selhausen) to 40 W m⁻² (Rollesbroich) for PF.WRF and from 6 W m⁻² (Selhausen) to 15 W m⁻² (Rollesbroich) for TerrSysMP. Also, when averaged individually over the three dominant land-use classes (evergreen forest 12%, deciduous forest 31%, and cropland 35%) for the whole domain (excluding the buffer zone depicted in Fig. 1), PF.WRF and TerrSysMP agree well in daily net radiation within 5 and 44 W m⁻² for cropland and deciduous forest, respectively.

Models and observations disagree, however, in some characteristics of the daily averaged diurnal cycle of LE and H at the Selhausen and Merzenhausen crop sites (Fig. 8). At both stations, PF.WRF simulates higher LE fluxes compared to TerrSysMP and the measurements. The better performance of TerrSysMP can most probably be attributed to the recent implementation of crop-specific physiological properties for winter wheat, which outperforms the generic CLM crop parameterizations (Sulis et al. 2015). The larger LE values simulated by both models at Merzenhausen are likely a consequence of too high LAI values, which ignore the senescence period of winter wheat dominating the footprint of the eddy covariance station. Accordingly, both models compare well with these observations over an area a few grid points apart dominated by sparsely vegetated land cover; here both reproduce the large observed drop in LE (Fig. 8).

Both models overestimate H in daytime over Selhausen and underestimate it at Merzenhausen, while nighttime downward H is always overestimated. The daytime overestimation of H at Selhausen is probably again caused by an overestimation of soil temperature in both models due to missed cloudiness. There is also a temporal lag in the diurnal cycles between the models and the observations, which seems to be somewhat better captured by TerrSysMP during daytime, while PF.WRF better replicates the low observed nighttime LE values. As discussed already for the idealized test cases, the temporal lag might be related to the differences in the assumed turbulent length scale under nighttime stable conditions in both model systems. The differences between PF.WRF and TerrSysMP, and also between the models and the observations, are reflected in the Bowen ratios ($b_{obs} = 0.10$, $b_{PF.WRF} = 0.089$, and $b_{TerrSysMP} = 0.098$ in Selhausen and $b_{obs} = 1.0$, $b_{PF.WRF} = 0.10$, and $b_{TerrSysMP} = 0.22$ in Merzenhausen). When inspecting the mismatch between models and observations, one should also take into account the systematic underestimation of turbulent fluxes by the
Eddy covariance technique in general. A previous study by Eder et al. (2015) showed that the energy balance closure at these stations range from 0.76 ± 0.11 (Merzenhausen) to 0.85 ± 0.13 (Selhausen).

The daily averaged volumetric soil moisture content simulated by PF.WRF and TerrSysMP are compared to values extracted from cosmic-ray probes (Table 3). The model values are derived by averaging over the top 10–15-cm layers to which neutron counts are sensitive. Both models consistently simulate the drying process within the variability induced by the different initial conditions. They also compare reasonably well with the observations given the different scales covered by the models and the cosmic-ray probes. PF.WRF, however, simulates wetter soil conditions at some locations (Heisenberg, RurAue, and Wildenrath) compared to TerrSysMP, which better agrees with the observations at these locations. The lower performance of PF.WRF could be partially explained by its coarser vertical discretization of the subsurface, which requires higher LE to remove a larger amount of water stored in the topsoil layer. The cosmic-ray signal originates mostly from shallower soil layers, with contributions decaying exponentially with soil depth (Bogena et al. 2013; Baatz et al. 2014). Thus, soil moisture variations should be better captured with a higher vertical grid resolution. The evaluation of models against local measurements of soil moisture should, however, be interpreted with care. Both PF.WRF and TerrSysMP use a 1-km grid resolution, which reduces subsurface drainage efficiency and thus could explain the wet bias of the models for locations in the northern lowland region (Merzenhausen, Gevenich, Heisenberg, and Wildenrath). Some stations (e.g., Wuestebach) are located over soils with high organic

![Fig. 8. Observed and ensemble mean simulated average diurnal cycle of the latent and sensible heat fluxes at the](http://journals.ametsoc.org/doi/pdf/10.1175/JHM-D-16-0159.1)
content, which was not included in the parameterizations of both models. Accordingly, the porosity measured in the field differs from the static model fields. Finally, both models in our setup use one dominant land-cover type (a single plant functional type in CLM) for each grid cell, which might disagree with the true land use within the cosmic-ray footprints.

2) Simulations of 1–5 June 2011

The second hindcast period allows us to compare the ability of the models to reproduce a series of rainfall events following 3 days of clear-sky conditions. Figure 9a shows the spatially averaged hourly values of precipitation during the last 2 days of the period between 80 and 120 h (the shaded area depicted in Fig. 1 excluded). The first rainfall event, which occurs between 85 and 103 h in the radar observations, is mostly related to a frontal passage, while the second event between 107 and 119 h is caused by local convective cells initiated during the afternoon hours in the middle of the domain and exhibiting a larger variability in the amount of precipitation due to the different initial soil moisture conditions. Both models underestimate the total amount of precipitation, with a cumulative value of 22 mm observed from the radar composites and 12 and 16 mm simulated by PF.WRF and TerrSysMP, respectively. This underestimation relates, however, mostly to the first (frontal) rainfall event while the second (convective) event is reproduced more consistently both in timing and magnitude by the models. The sensitivity of the simulated local convective events to the perturbed soil moisture initial conditions in the ensemble is similar for both models (see location of spread in Fig. 9a), with PF.WRF exhibiting a stronger development in the

<table>
<thead>
<tr>
<th>Location</th>
<th>21 Jul</th>
<th>22 Jul</th>
<th>23 Jul</th>
<th>24 Jul</th>
<th>25 Jul</th>
<th>26 Jul</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wuestebach</td>
<td>0.453</td>
<td>0.391</td>
<td>0.379</td>
<td>0.417</td>
<td>0.423</td>
<td>0.385</td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF.WRF</td>
<td>0.238 (0.009)</td>
<td>0.221 (0.005)</td>
<td>0.208 (0.003)</td>
<td>0.196 (0.002)</td>
<td>0.187 (0.002)</td>
<td>0.179 (0.001)</td>
</tr>
<tr>
<td>TerrSysMP</td>
<td>0.290 (0.017)</td>
<td>0.278 (0.013)</td>
<td>0.266 (0.011)</td>
<td>0.255 (0.010)</td>
<td>0.246 (0.009)</td>
<td>0.237 (0.008)</td>
</tr>
<tr>
<td>Rollesbroich</td>
<td>0.402</td>
<td>0.380</td>
<td>0.380</td>
<td>0.377</td>
<td>0.352</td>
<td>0.368</td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF.WRF</td>
<td>0.295 (0.005)</td>
<td>0.273 (0.002)</td>
<td>0.258 (0.001)</td>
<td>0.248 (0.001)</td>
<td>0.239 (0.001)</td>
<td>0.232 (0.001)</td>
</tr>
<tr>
<td>TerrSysMP</td>
<td>0.236 (0.014)</td>
<td>0.223 (0.012)</td>
<td>0.208 (0.011)</td>
<td>0.194 (0.010)</td>
<td>0.181 (0.010)</td>
<td>0.169 (0.010)</td>
</tr>
<tr>
<td>Merzenhausen</td>
<td>0.270</td>
<td>0.255</td>
<td>0.232</td>
<td>0.226</td>
<td>0.212</td>
<td>0.209</td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF.WRF</td>
<td>0.317 (0.011)</td>
<td>0.305 (0.008)</td>
<td>0.295 (0.007)</td>
<td>0.285 (0.006)</td>
<td>0.276 (0.006)</td>
<td>0.268 (0.006)</td>
</tr>
<tr>
<td>TerrSysMP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gevenich</td>
<td>0.250</td>
<td>0.237</td>
<td>0.211</td>
<td>0.213</td>
<td>0.200</td>
<td>0.190</td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF.WRF</td>
<td>0.335 (0.012)</td>
<td>0.301 (0.006)</td>
<td>0.284 (0.005)</td>
<td>0.271 (0.004)</td>
<td>0.261 (0.004)</td>
<td>0.253 (0.003)</td>
</tr>
<tr>
<td>TerrSysMP</td>
<td>0.270 (0.013)</td>
<td>0.250 (0.009)</td>
<td>0.232 (0.008)</td>
<td>0.216 (0.007)</td>
<td>0.202 (0.006)</td>
<td>0.188 (0.006)</td>
</tr>
<tr>
<td>Heinsberg</td>
<td>0.280</td>
<td>0.270</td>
<td>0.248</td>
<td>0.240</td>
<td>0.240</td>
<td>0.237</td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF.WRF</td>
<td>0.438 (0.00)</td>
<td>0.435 (0.002)</td>
<td>0.416 (0.010)</td>
<td>0.390 (0.009)</td>
<td>0.373 (0.006)</td>
<td>0.359 (0.005)</td>
</tr>
<tr>
<td>TerrSysMP</td>
<td>0.248 (0.016)</td>
<td>0.241 (0.012)</td>
<td>0.216 (0.010)</td>
<td>0.202 (0.009)</td>
<td>0.189 (0.009)</td>
<td>0.177 (0.008)</td>
</tr>
<tr>
<td>Kall</td>
<td>0.252</td>
<td>0.351</td>
<td>0.327</td>
<td>0.323</td>
<td>0.314</td>
<td>0.296</td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF.WRF</td>
<td>0.273 (0.010)</td>
<td>0.255 (0.006)</td>
<td>0.243 (0.004)</td>
<td>0.234 (0.003)</td>
<td>0.227 (0.002)</td>
<td>0.220 (0.002)</td>
</tr>
<tr>
<td>TerrSysMP</td>
<td>0.259 (0.017)</td>
<td>0.247 (0.014)</td>
<td>0.235 (0.012)</td>
<td>0.224 (0.011)</td>
<td>0.213 (0.001)</td>
<td>0.203 (0.009)</td>
</tr>
<tr>
<td>RurAue</td>
<td>0.324</td>
<td>0.302</td>
<td>0.300</td>
<td>0.300</td>
<td>0.290</td>
<td>0.283</td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF.WRF</td>
<td>0.444 (0.00)</td>
<td>0.438 (0.020)</td>
<td>0.424 (0.007)</td>
<td>0.394 (0.005)</td>
<td>0.375 (0.004)</td>
<td>0.359 (0.003)</td>
</tr>
<tr>
<td>TerrSysMP</td>
<td>0.30 (0.013)</td>
<td>0.291 (0.009)</td>
<td>0.281 (0.007)</td>
<td>0.272 (0.006)</td>
<td>0.264 (0.006)</td>
<td>0.256 (0.005)</td>
</tr>
<tr>
<td>Wildenrath</td>
<td>0.243</td>
<td>0.242</td>
<td>0.215</td>
<td>0.198</td>
<td>0.186</td>
<td>0.180</td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF.WRF</td>
<td>0.430 (0.005)</td>
<td>0.402 (0.008)</td>
<td>0.377 (0.004)</td>
<td>0.359 (0.003)</td>
<td>0.344 (0.003)</td>
<td>0.331 (0.003)</td>
</tr>
<tr>
<td>TerrSysMP</td>
<td>0.321 (0.006)</td>
<td>0.313 (0.004)</td>
<td>0.306 (0.004)</td>
<td>0.299 (0.003)</td>
<td>0.293 (0.003)</td>
<td>0.287 (0.003)</td>
</tr>
<tr>
<td>Aachen</td>
<td>0.397</td>
<td>0.380</td>
<td>0.372</td>
<td>0.357</td>
<td>0.340</td>
<td>0.319</td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF.WRF</td>
<td>0.300 (0.023)</td>
<td>0.272 (0.013)</td>
<td>0.256 (0.010)</td>
<td>0.244 (0.009)</td>
<td>0.234 (0.009)</td>
<td>0.227 (0.008)</td>
</tr>
<tr>
<td>TerrSysMP</td>
<td>0.290 (0.035)</td>
<td>0.266 (0.025)</td>
<td>0.247 (0.021)</td>
<td>0.230 (0.019)</td>
<td>0.216 (0.018)</td>
<td>0.202 (0.017)</td>
</tr>
</tbody>
</table>
earlier hours of the second event. Both models simulate higher percentages for low rainfall rates compared to the observations for the frontal passage but adequately reproduce the intensity frequencies for the convective event (Figs. 9b,c). In general, the higher similarity between the models compared to the observations suggests that the observation model difference is mainly caused by uncertainties of the lateral atmospheric boundary.

The spatial pattern of the cumulative precipitation is evaluated between 85 and 94 h (Fig. 10a) and between 107 and 119 h (Fig. 10b), which correspond to 1300–2200 UTC of the fourth day and to 1100–2300 UTC of the fifth day, respectively. Both models reproduce the passage of a southeastern front on day 4, consistent with the radars. On day 5, the radars detect four main convective cells located between 50°20′ and 50°45′N that produce a large amount of rainfall in a short time period. Both PF.WRF and TerrSysMP also produce these strong convective storms between 50°15′ and 50°45′N. For the convective event, the spatial pattern simulated by the two models differs most notably over the northern lowland region. We hypothesize that this difference is caused by the different forcings generated by the two land surface schemes. Figure 11 shows the frequency distribution of the turbulent energy fluxes and the convective available potential energy (CAPE) prior to the convective event in PF.WRF over an area around 51°N, 6°E (see shaded area in Fig. 10b), which does not develop in TerrSysMP. The higher soil moisture saturation in PF.WRF in the top 40 cm (0.84 compared to 0.67 in TerrSysMP) leads to lower sensible heat fluxes and higher latent heat fluxes, which might result in higher CAPE in PF.WRF compared to TerrSysMP in this area. CAPE is a measure of the potential convection intensity estimated from the parcel buoyancy between the level of free convection and the neutral level. In TerrSysMP, CAPE strongly peaks at low values around ~1900 J kg⁻¹ while PF.WRF exhibits a much smoother distribution with a rather flat peak value around 2300 J kg⁻¹ and considerable fractions even up to ~3000 J kg⁻¹, which might be responsible for the development of this event.

While land surface energy partitioning seems to influence the development of convective cells in both models, convective rainfall involves many processes between the surface, the ABL, and the free troposphere, including surface exchange, turbulence, convection, and cloud–radiation interactions. Thus, it is difficult to draw firm conclusions in the presented numerical experiments. This causality could be more adequately investigated by employing numerical experiments in a multiphysics/multimodel ensemble framework that allows us to systematically disentangle all the intricacies of coupled processes in the terrestrial system.

The propagation of spatial differences in the precipitation is analyzed by comparing the integrated discharge response at two catchments. One catchment (point A in Fig. 12) includes a mountainous region while the second (point B in Fig. 12) drains mostly the low-lying area of the model domain. The discharge response at point A consistently reproduces the frontal passage for both models, with TerrSysMP (in agreement with Fig. 9a) simulating an earlier and steeper-rising hydrograph. The hydrograph also signals a very localized convective event in TerrSysMP, which is not simulated by PF.WRF. The envelope of the hydrographs produced by the ensemble gains in spread for this event, demonstrating how the variability (induced by the perturbed moisture conditions) in convective rainfall propagates into the integrated model response. The discharge simulated by both model systems at the downstream point B integrates once again the frontal passage with the flow.

FIG. 9. (a) Spatially averaged time series and cumulative frequency distribution (b) between 85 and 100 h and (c) between 107 and 119 h of the observed and ensemble mean simulated precipitation for the simulations of 1–5 Jun 2011. Note the semilog scale for (b) and (c).
response delayed (~98 h) and dampened compared to the one simulated at point A. The much steeper discharge rise simulated by PF,WRF reflects the convective cell detected around 51°N, 6°80'E. In TerrSysMP, the rainfall signal associated with the convective event simulated upstream in the catchment around 50°45’N, 6°80’E is largely attenuated through the stream network. Overall, these results highlight how different and uncertain spatial precipitation patterns may lead to quite diverse (and thus unreliable) discharge predictions, which can only be ameliorated for short-term forecasts that assimilate observed radar observations (e.g., Milan et al. 2014; Bick et al. 2016).

3) INFLUENCE OF GROUNDWATER DYNAMICS ON DIURNAL ATMOSPHERIC BOUNDARY LAYER EVOLUTION

The relative differences in land–atmosphere coupling between the two systems were explored in terms of ABL heat and moisture budgets calculating the mixing diagrams for two groundwater configurations. The first configuration (3D) corresponds to the standard implementation of both models, with a dynamic calculation of water-table depth from the three-dimensional Richards equation. The second configuration (FD) mimics the description of soil water flow in classical LSMs by imposing free gravity drainage at the soil bottom (Campoy et al. 2013; Rahman et al. 2016). In the vector representation of the mixing diagrams (Santanello et al. 2009, 2011), the diurnal coevolution of 2-m humidity and potential temperature in the $\text{L}_q - \text{C}_p\theta$ plane is used to evaluate conditions and processes at the land surface and the top of the ABL, where $L$ is the latent heat of vaporization and $C_p$ is the specific heat at constant pressure. The relative contribution of the bottom and top boundaries to the ABL heat and moisture budgets can be assessed with the indices $b_{sfc} = H_{sfc}/LE_{sfc}$ and $b_{ent} = H_{ent}/LE_{ent}$, which quantify the partitioning between latent and sensible heat at the surface (sfc) and at the entrainment (ent) interface and the proportion of sensible and latent heat input to the ABL ($A_H = H_{ent}/H_{sfc}$; $A_{LE} = LE_{ent}/LE_{sfc}$) from both interfaces.

Figure 13 shows the mixing diagrams computed from the state variables over crop land use (see Fig. 1) and
with water-table depths below 1 m. The 3D and FD simulate the diurnal coevolution of $q$ and $\theta$ for both models in a qualitatively similar manner. The $q$ rises in the morning when the ABL is shallow and capped by relatively moister air from the previous day and decreases in the afternoon when the ABL entrains drier air from the free atmosphere. The larger rise of $\theta$ in PF.WRF reflects the faster increase of $H$ compared to TerrSysMP (Fig. 8). The shift toward moister atmospheric conditions from the second to the last day suggests the advection of moister air during the simulation period.

Table 4 summarizes the metrics extracted from the mixing diagrams following the methodology described in Santanello et al. (2009, 2011). Both models largely agree on energy partitioning at the land surface. The simulated Bowen ratio indicates a larger contribution of LE (low $b_{sc}$) over the surface at the beginning and the end of the simulation and a slightly larger contribution of LE in PF.WRF. The two models also show a similar evolution at the entrainment interface ($b_{ent}$), with PF.WRF simulating a higher $H$ input at the top of the ABL during the last simulation day (26 July). Both models agree on the energy partitioning at the land surface for the FD groundwater configuration. In this case, however, the larger increase of the Bowen ratio suggests a stronger sensitivity of PF.WRF to the groundwater configuration. This is also indicated by the changes in the evaporative fraction (EF; ratio of LE to total surface energy flux) simulated by PF.WRF for the FD configuration on the second (EF$_{PF.WRFFD}$ = 0.77, EF$_{PF.WRF3D}$ = 0.62, EF$_{TerrSysMPFD}$ = 0.71, and EF$_{TerrSysMP3D}$ = 0.67) and last day (EF$_{PF.WRF3D}$ = 0.81, EF$_{PF.WRFFD}$ = 0.70, EF$_{TerrSysMP3D}$ = 0.78, and EF$_{TerrSysMPFD}$ = 0.74), which leads to a different ABL growth (for PF.WRF 1078 and 850 m and for TerrSysMP 1203 and 826 m, respectively). It is interesting to observe that variations in the soil moisture over the top 10 cm induced by the two groundwater configurations for the point used in this analysis (cropland land use with a water table below 1 m) differ between the two models. PF.WRF shows variations between 40% and 38% from the second (22 July) to the last day (26 July) of the simulation, while TerrSysMP exhibits changes between 44% and 63% for the same time period. These differences can be explained by the coarser vertical discretization used in the subsurface part of PF.WRF, which tends to dampen soil moisture variability. Moreover, from the energy partitioning metrics previously discussed (e.g., EF), it also appears that such soil moisture variations evolve differently in PF.WRF and TerrSysMP because of their different land surface–atmosphere parameterizations.

5. Discussion

This work presents a first comparison of two fully integrated models that account for a 3D resolution of both groundwater and atmospheric dynamics. Special emphasis is placed on the identification of key distinguishing features in the land surface and ABL parameterizations, the sensitivity of integrated model responses to groundwater initialization and formulation, and the degree of realism in the comparison with
measurements. Model differences (and similarities) are explained without suggestions on the better model, as the scope of such intercomparison is to get a better understanding of the complex dynamics in the terrestrial systems. Nevertheless, comparing and interrogating different models also provides the opportunity for identifying venues of future developments in models physics, parameterizations, and coupling strategies. In light of the findings and caveats associated with this work, realizing this potential requires addressing important additional technical tasks and methodological improvements. In fact, there is the need of approaches, especially for real-case applications, for disentangling the interplay between groundwater, land surface, and atmospheric physics formulations. These should be explored using ensembles with variable land characterizations (as done here) and different ABL parameterizations and by simulating subsurface hydrological processes with different levels of complexity (e.g., 3D, hybrid-3D, and 1D). This could be achieved by designing generic multimodel coupling infrastructure with embedded multiphysics options, which would facilitate in demonstrating causality between models’ differences.

Processes in the terrestrial system interact in a highly nonlinear way, and we expect that the differences in model discretization, formulation, and parameterization amplify or attenuate state evolutions. To arrive at more general conclusions, this type of analysis should therefore be extended to other environmental conditions encompassing, for instance, persistent and prolonged dry or moist periods. Longer time series will allow for the use of metrics based on signals in the frequency domain. Such analyses could better reveal which component of the terrestrial system dominates the variability of model response, which would help in more systematically identifying the circumstances under

FIG. 12. Simulated discharge at upstream (point A) and downstream (point B) outlets for the simulations of 1–5 Jun 2011. Note the different scale used for the two hydrographs.
which different modeling approaches may provide contrasting or similar results.

6. Conclusions

This study compares two fully coupled hydrologic–atmospheric models, PF.WRF and TerrSysMP, which dynamically simulate interactions and feedback mechanisms in the aquifer–soil–vegetation–atmosphere continuum. This new class of models provides the opportunity to explore the influence of spatiotemporal subsurface hydrodynamics on atmospheric processes using a holistic and physically based approach. The knowledge of potential differences between such models, including the mechanisms and their evaluation with observations, is important for the hydrological and meteorological modeling community.

The first part of this intercomparison comprises an analysis of model results obtained using idealized configurations with uniform land-cover conditions (bare soil and vegetated) through one diurnal cycle. The results show the key role of land surface parameterizations (i.e., soil and stomatal resistances) in controlling the exchange fluxes and subsequent atmospheric thermodynamics in the models. PF.WRF generally simulates a deeper and more buoyant ABL because of a stronger (bare soil) and faster (vegetated) heating during the rising part of the diurnal cycle. The results indicate that the different treatment of the vertical mixing length for heat during nighttime, as well as different subsurface vertical discretizations, may considerably impact the diurnal evolution of land surface energy fluxes. A sensitivity analysis to water-table depth also shows the different connection between groundwater, land surface, and atmospheric processes simulated by both models. Overall, our results corroborate and complement some of the findings of earlier studies by providing an integrated assessment of formulations based on different assumptions and complexity.

Two hindcast simulations with clear sky only and mixed with rainy conditions extended the model comparison over an area in western Germany. These real-case simulations enabled a comparison with observations of soil moisture, radiation fluxes, latent and sensible heat fluxes, and precipitation. The comparisons revealed a generally good agreement between PF.WRF and TerrSysMP and with net radiation and total precipitation observations. Both models, however, slightly

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( b_{\text{dsc}} )</td>
<td>( b_{\text{ext}} )</td>
<td>( A_H )</td>
<td>( A_{\text{LE}} )</td>
<td>( \text{ABL}_{\text{height}} ) (m)</td>
</tr>
<tr>
<td>22 Jul</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3D PF.WRF</td>
<td>0.29</td>
<td>-0.51</td>
<td>1.33</td>
<td>-0.77</td>
<td>940</td>
</tr>
<tr>
<td>TerrSysMP</td>
<td>0.40</td>
<td>-0.48</td>
<td>1.24</td>
<td>-1.04</td>
<td>1070</td>
</tr>
<tr>
<td>FD PF.WRF</td>
<td>0.59</td>
<td>-0.46</td>
<td>0.86</td>
<td>-1.10</td>
<td>1078</td>
</tr>
<tr>
<td>TerrSysMP</td>
<td>0.49</td>
<td>-0.71</td>
<td>1.27</td>
<td>-0.88</td>
<td>1203</td>
</tr>
<tr>
<td>26 Jul</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3D PF.WRF</td>
<td>0.22</td>
<td>-0.42</td>
<td>0.95</td>
<td>-0.50</td>
<td>630</td>
</tr>
<tr>
<td>TerrSysMP</td>
<td>0.28</td>
<td>-0.33</td>
<td>0.93</td>
<td>-0.78</td>
<td>690</td>
</tr>
<tr>
<td>FD PF.WRF</td>
<td>0.43</td>
<td>-0.52</td>
<td>0.80</td>
<td>-0.65</td>
<td>850</td>
</tr>
<tr>
<td>TerrSysMP</td>
<td>0.35</td>
<td>-0.39</td>
<td>1.02</td>
<td>-0.92</td>
<td>826</td>
</tr>
</tbody>
</table>
overestimate net shortwave radiation and underestimate total precipitation for these periods. The models differ more for land surface energy fluxes and soil moisture. TerrSysMP slightly outperforms PF.WRF in the simulation of LE at the local scale because of crop-specific physiological vegetation parameterizations and better captures the soil moisture content with a finer vertical discretization in the subsurface layers.

Despite similar amounts and intensities of simulated rainfall, the models produce quite different spatial precipitation patterns. PF.WRF simulates the formation of more scattered convective precipitation cells, which probably reflect the differing land surface parameterizations. Such differences in atmospheric processes propagate into the stream discharge simulated at two catchments. The results also show that uncertainty in the initial soil moisture conditions, which is due to some limitations in implementing a consistent spinup approach for PF.WRF and TerrSysMP, propagates through the integrated rainfall (via convection initiation)–runoff response.

Finally, the evolution of the ABL energy and moisture balance simulated by both systems was analyzed for two groundwater configurations: the 3D that dynamically calculates water-table depth from the three-dimensional Richards equation and the FD that reproduces the classical one-dimensional description of soil water flow in LSMS. These numerical experiments reveal consistently different heat and moisture budgets of the ABL with a much higher H simulated for the FD configuration by both PF.WRF and TerrSysMP. The results also suggest a greater sensitivity of the energy partitioning to water-table-induced near-surface soil moisture variations in PF.WRF.

Acknowledgments. This research work was supported by the Sonderforschungsbereich/Transregional Collaborative Research Centre 32 (SFB/TR32) Patterns in Soil–Vegetation–Atmosphere Systems: Monitoring, Modeling and Data Assimilation project funded by the Deutsche Forschungsgemeinschaft (DFG). We thank Marius Schmidt, Alexander Graf, Heye Bogena, and Roland Baatz for providing meteorological, flux, and soil moisture data from the Eifel/Lower Rhine valley observatory of the Terrestrial Environmental Observations (TERENO) project, funded by the Helmholtz Association. We are grateful to Sebastian Knist and Mostaquimur Rahman for their useful comments on an early version of the manuscript. We also gratefully acknowledge the computing time (project HBN33) granted by the John von Neumann Institute for Computing (NIC) and provided on the supercomputer Jülich Research on Exascale Cluster Architectures (JURECA) at the Jülich Supercomputing Centre (JSC). The constructive criticism of Editor-in-Chief Christa Peters-Lidard and of two anonymous reviewers has been greatly beneficial for the quality of the manuscript.

REFERENCES


Larsen, M. A. D., J. C. Refsgaard, M. Drews, M. B. Butts, K. H. Jensen, J. H. Christensen, and O. B. Christensen, 2014: Results...


——, M. D. Heiser, and F. G. Hall, 1992b: Relations between surface conductance and spectral vegetation indices at intermediate (100 m2 to 15 km2) length scales. J. Geophys. Res., 97, 19 033–19 059, doi:10.1029/92JD01096.


