



# Inter-comparison of three distributed hydrological models with respect to seasonal variability of soil moisture patterns at a small forested catchment

Julian Koch <sup>a,b,\*</sup>, Thomas Cornelissen <sup>c</sup>, Zhufeng Fang <sup>d</sup>, Heye Bogena <sup>d</sup>, Bernd Diekkrüger <sup>c</sup>, Stefan Kollet <sup>d,e</sup>, Simon Stisen <sup>b</sup>

<sup>a</sup> Department of Geosciences and Natural Resources Management, University of Copenhagen, Denmark

<sup>b</sup> Department of Hydrology, Geological Survey of Denmark and Greenland, Denmark

<sup>c</sup> Department of Geography, University of Bonn, Germany

<sup>d</sup> Agrosphere Institute, IBG-3, Forschungszentrum Jülich, Germany

<sup>e</sup> Centre for High-Performance Scientific Computing in Terrestrial Systems, HPSC TerrSys, Geoverbund ABC/J, Germany

## ARTICLE INFO

### Article history:

Received 6 May 2015

Received in revised form 19 August 2015

Accepted 2 December 2015

Available online 15 December 2015

This manuscript was handled by Peter K. Kitaniidis, Editor-in-Chief, with the assistance of Ty Ferre, Associate Editor

### Keywords:

Distributed hydrological modeling

Model inter-comparison

Soil moisture

Spatial patterns

Soil parametrization

Spatial model evaluation

## SUMMARY

The objective of this study is to inter-compare three spatially distributed hydrological models (HydroGeoSphere, MIKE SHE and ParFlow-CLM) by means of their ability to simulate soil moisture patterns. This study pools the catchment modeling efforts which have been undertaken at the Wüstebach catchment; one of TERENO's hydrological observatories. The catchment is densely instrumented with a wireless sensor network (SoiNET) which allows continuous measurements of the spatio-temporal soil moisture dynamics. This unique dataset is ideal to benchmark hydrological models as it poses distinct challenges like seasonality and spatial heterogeneity. Two scenarios of soil parametrization assess the modeling implications of moving from homogeneous to heterogeneous porosity. The three given models perform well in terms of discharge and accumulated water balance components. However, their ability to predict soil moisture is found to be more diverging. Interpretations are ambiguous and depend on what performance metric and what level of spatial aggregation is chosen. In comparison to the other models, ParFlow-CLM performs more accurate at predicting the temporal dynamics and the heterogeneity aggregated to catchment scale. Nevertheless, at local scale HydroGeoSphere and MIKE SHE provide more detailed soil moisture predictions. Overall, a clear increase in performance can be attested to the scenario that includes heterogeneous porosity. Next to soil parametrization, topography is among the main drivers of soil moisture variability which was found to have an overemphasized feedback in ParFlow-CLM compared to the other models. This study stresses that further efforts toward spatially distributed input data need to emerge alongside a more suitable soil parametrization that can account for the observed heterogeneity and seasonality of soil moisture.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

There is a growing number of distributed hydrological models with varying degrees of complexity that are used to address a wide range of scientific questions (Kampf and Burges, 2007; Smith and Gupta, 2012). These models are utilized to predict not only discharge at the outlet and along the stream network but also to make spatially explicit predictions like hydrologic responses to land use

or land cover changes, land atmosphere interactions, water contamination and many others. The modeling community is faced with an overabundance of models (Clark et al., 2011) thus inter-comparison and benchmark studies gain an increasing interest. Moreover, there is a prevailing demand for guidance to which model with what degree of complexity to choose for a certain application. A recent effort toward this direction is the model inter-comparison with respect to future discharge predictions in a West African catchment undertaken by Cornelissen et al. (2013).

From the most abstract perspective, hydrological models can be compared on the basis of a synthetic catchment which allows reliable and transparent interpretations due to simplified conditions as presented by Maxwell et al. (2014). It was found that the

\* Corresponding author at: Department of Hydrology, Geological Survey of Denmark and Greenland (GEUS), Øster Voldgade 10, Copenhagen, Denmark. Tel.: +45 38142768.

E-mail address: [juko@geus.dk](mailto:juko@geus.dk) (J. Koch).

agreement between seven distributed hydrological models clearly diverges with an increasing conceptual complexity of the catchment. Closer to reality but still simplified are inter-comparison studies on man-made artificial catchments like the Chicken Creek (Hölzel et al., 2011). Holländer et al. (2009, 2014) compared ten conceptually different hydrological models with respect to simulated water balance components and discharge predictions at the outlet of the artificial Chicken Creek. They concluded that besides the choice of model, the modeler itself is an intrinsic part of any modeling study which is often not taken into consideration when comparing models. The Distributed Model Inter-comparison Project (DMIP) is an outstanding example of inter-comparison of hydrological models by means of real world catchments (Reed et al., 2004). The second phase of the DMIP incorporates discharge at both, the outlet and interior stream gauges as well as soil moisture to benchmark 14 different distributed hydrological models at three catchments in the Oklahoma region (Smith et al., 2012a). The participating models all performed well in generating streamflow at interior gauges and maintain the overall correct water balance while only two of the models provided reasonably good estimates of soil moisture (Smith et al., 2012b).

In general, most inter-comparison studies of hydrological models focus on the comparison of simulated discharge. Discharge is an important hydrological variable but nevertheless it is an aggregated measure that is found to be insensitive to spatially explicit hydrological states and fluxes as presented by Stisen et al. (2011). Furthermore, Clark et al. (2011) identify aggregated measures in inter-comparison studies as a major drawback to successfully evaluate distributed model alternatives. Along these lines, Pokhrel and Gupta (2011) underline the importance of adding spatially distributed data to the validation of distributed hydrological models. In order to undertake a true and reliable spatial validation of a distributed hydrological model it is important to have distributed observations at a similar scale as the distributed predictions of the model. Previous approaches have incorporated satellite based or densely sampled soil moisture observations (Wealands et al., 2005) or remotely sensed land surface temperature data (Koch et al., 2015). A recent example of a real world catchment that is subject to a model inter-comparison study that goes beyond the assessment of discharge and includes distributed soil moisture data is presented by Noh et al. (2015). Models of varying complexity are assessed against sparse soil moisture data and it is found that the models generally capture the temporal dynamics but miss the spatial variability of the soil moisture patterns.

This study focuses on the inter-comparison of spatio-temporal soil moisture variability of three state of the art distributed hydrological models (HydroGeoSphere, MIKE SHE and ParFlow-CLM). The most prominent model distinction is that HydroGeoSphere and ParFlow-CLM are fully integrated with 3D flow over the entire domain opposed to the coupling between 1D unsaturated flow and 3D saturated flow in MIKE SHE. Additionally, ParFlow-CLM incorporates the common land model (CLM) which represents a more detailed physical description of the land surface energy fluxes compared to the common practice of applying potential evapotranspiration as external forcing in the other two models. All of the above mentioned models fall under the category of physically based models that upscale well understood physical governing equations of small scale water movement to the field scale. This premise has been subject to a long and still ongoing discussion among the community (Grayson et al., 1992; Gupta et al., 2012; Kirchner, 2006; Semenova and Beven, 2015).

The catchment that is subject to the inter-comparison study is the densely instrumented Wüstebach catchment (Bogena et al., 2015) which is part of the network of hydrological observatories of TERENO (Zacharias et al., 2011). At the 38.52 ha large Wüstebach site, soil moisture is monitored at 150 locations (three sensor

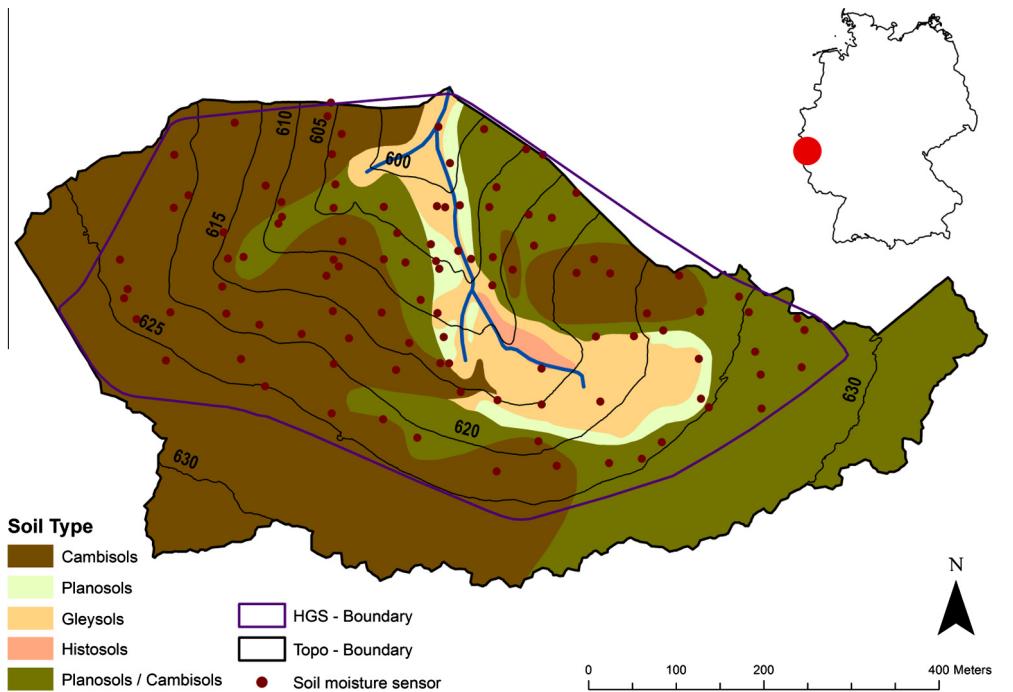
depths per location) which allows for a detailed analysis of the spatio-temporal variability of soil moisture and provides a unique validation dataset for distributed hydrological models. Previous modeling studies of the Wüstebach site that exploit this comprehensive soil moisture dataset for the spatial model validation were conducted by Cornelissen et al. (2014) and Fang et al. (2015).

Soil moisture is regarded the key hydrological state variable in the terrestrial system because it controls the fluxes of water and energy between the land surface and atmosphere with complex and nonlinear mechanisms (Sheffield and Wood, 2008; Vereecken et al., 2014). At smaller scales, soil moisture governs soil water movement and thus it is the main driver for runoff generation processes at hillslope catchments (Blume et al., 2009; Brocca et al., 2007) which was recently studied by Stockinger et al. (2014) at the Wüstebach site. The spatial-temporal variability of soil moisture is regarded complex which hinders straightforward monitoring and analysis of observed soil moisture data (Brocca et al., 2010; Graf et al., 2014; Wang et al., 2015). Typically, the Richards' equation represents the physical basis of modeling temporal changes in soil moisture at local scale in single- or multi-dimensional space (Martinez et al., 2013; Romano, 2014). Although the governing equations are well understood integrated multi-dimensional modeling of distributed variables such as soil moisture at catchment scale still remains challenging (Wood et al., 2011). Chaney et al. (2014) and Korres et al. (2010) identify topography, soil properties, vegetation and precipitation as the main drivers of spatial soil moisture variability and their correct acquisition poses a major challenge to distributed modeling. As topography, vegetation and precipitation are well described at the small-scale Wüstebach catchment this study will focus on soil properties by means of two scenarios with different degrees of heterogeneity. In order not to bias the inter-comparison by using model specific parameters the soil properties are estimated in a surrogate model, which is not part of the inter-comparison. HYDRUS 1D solves the Richards' equation in the vertical dimension and is utilized in this context to estimate the soil hydraulic parameters that are passed onto the three hydrological models that are subject to the inter-comparison. Recently, Castillo et al. (2015) showed that a 1D soil column model and a distributed 3D hydrological model simulate very comparable soil moisture predictions.

This study aims at a thorough inter-comparison of three state of the art distributed hydrological models, namely HydroGeoSphere, MIKE SHE and ParFlow-CLM. The focus lies on (1) annual water balance components and daily discharge dynamics and (2) spatio-temporal soil moisture dynamics where two scenarios investigate the effect of moving from homogeneous to heterogeneous soil properties. Lastly, (3) the correlation between the errors of the individual models is assessed to gain insight into how independent the models are. Hence, the implications of the differences in model structure between the three models are investigated with respect to simulated discharge and soil moisture.

## 2. Study site

The study site for the inter-comparison study is the Wüstebach catchment which is part of the larger Rur catchment and it is located in the Eifel National Park in Germany near the Belgian border (Fig. 1). The Wüstebach catchment is one of the intensive hydrological observatories of TERENO (Zacharias et al., 2011). Its size amounts to 38.52 ha with altitude ranging from 595 to 629 m a.s.l. and average slopes of 3.6%. The catchment is homogeneously forested with more than 90% Norwegian spruce trees with an average height of 25 m. The geology consists of shallow soils with an average depth of 1.6 m that overlay a low permeable bedrock formation of Devonian shales. The predominant soil



**Fig. 1.** Soil map of the Wüstebach catchment and the wireless sensor network (SoilNET: 104 locations) used for this study.

texture is silty clay loam with cambisols and planosols on the hill-slopes and groundwater influenced gleysols and half-bogs in the riparian area in vicinity of the stream. A more detailed analysis of the spatial variability of soil texture is presented by [Sciuto and Diekkruger \(2010\)](#).

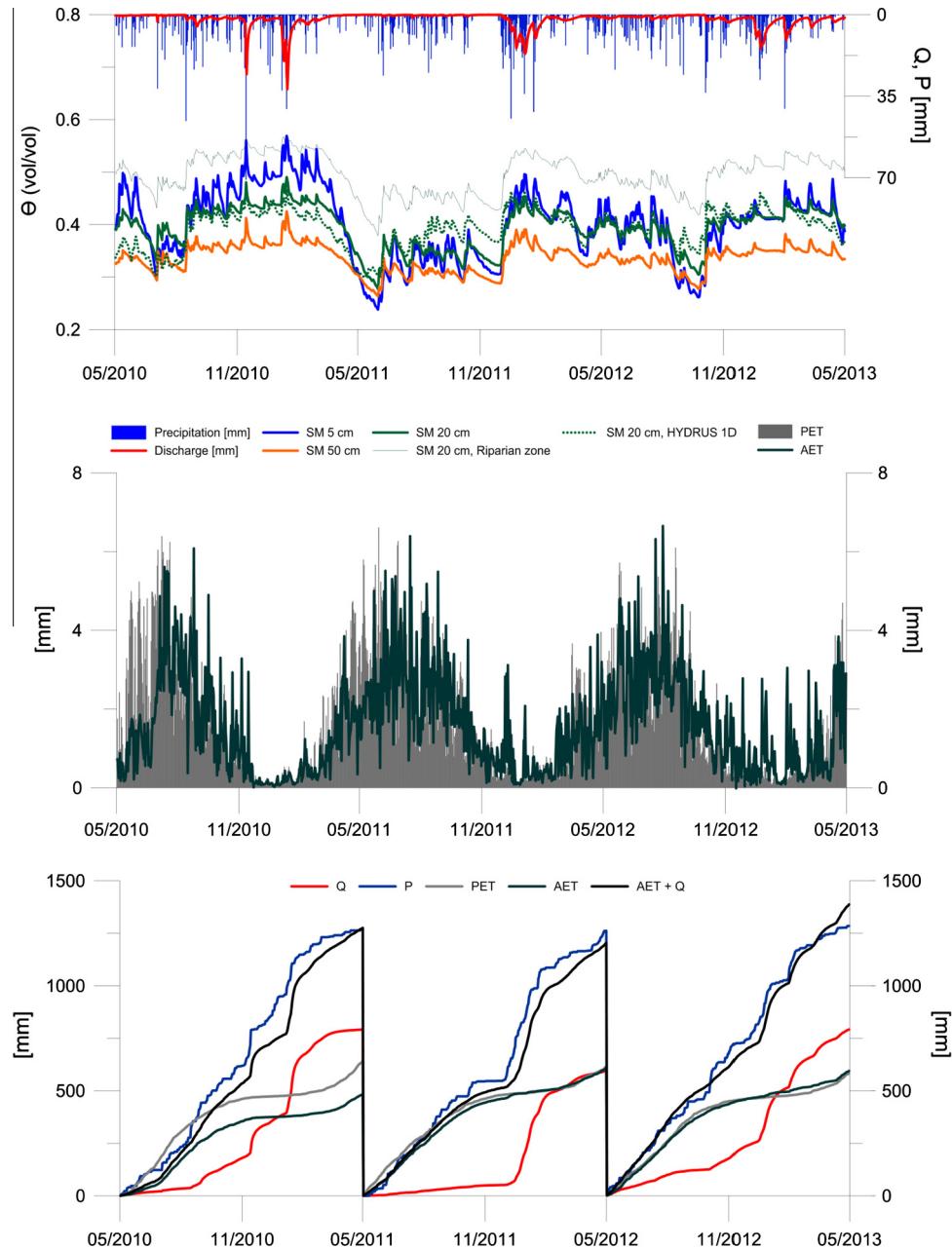
### 3. Hydrological data

High quality hourly data on all water balance components are available for the Wüstebach catchment ([Fig. 2](#)). Precipitation measurements are obtained from the nearest official meteorological station of the German weather service (Kalterherberg) which is located approximately 9 km to the west of the study site. Actual evapotranspiration (AET) data is measured at the top of a 38 m high tower directly in the Wüstebach catchment by the eddy-covariance technique. Potential evapotranspiration (PET) is calculated based on the FAO Penman–Monteith equation with input data measured at the nearby TERENO weather station Schönesieffen which is 3.5 km east of the Wüstebach catchment. Due to the small size of the catchment and the homogeneous land cover it is assumed that precipitation, actual- and reference-evapotranspiration are invariant in space over the entire catchment domain. Discharge is monitored at the outlet of the catchment by a combination of Parshall flumes and V-notch weirs. A more detailed description of measurements, corrections and equations of the four hydrological variables presented above is given by [Graf et al. \(2014\)](#) and [Bogena et al. \(2015\)](#).

[Fig. 2](#) presents observed data for three hydrological years in the period from 01-May-2010 to 30-April-2013. Discharge commonly peaks during winter and spring whereas summer and autumn are characterized by continuous low flows that originate from baseflow and quick preferential flow which emerges after strong precipitation events. On average 56% of the precipitation is turned into discharge and the remaining 44% account for AET. Overall, the Wüstebach catchment can clearly be characterized as an energy limited system where AET is

dominantly limited by the available energy input rather than by water availability. For the total three year period the residual between input (precipitation) and output (discharge plus AET) is 2% of the precipitation which emphasizes that changes in storage are negligible on an interannual scale and that groundwater percolation is insignificant. For the three individual hydrological years the residuals are –1.3%, 4.4% and –8.9% of the precipitation. Large water balance residuals can be explained by gap filling, spatial representation and uncertainties in precipitation- and AET-measurements ([Graf et al., 2014](#)).

Soil moisture (SM) within the Wüstebach catchment is measured at 150 locations using the wireless distributed sensor network SoilNET ([Bogena et al., 2010](#)). For this study, 104 of them are selected due to their quality and continuity of the data ([Fig. 1](#)). Each sensor measures soil temperature and SM at three depths (5, 20 and 50 cm). Further details on the network design, sensor calibration and signal processing are presented by [Bogena et al. \(2007, 2010\)](#). [Fig. 2](#) depicts the temporal dynamics of the daily averaged SM at all 104 stations at the three given depths. The general level of SM decreases with depth which reflects the increasing soil compaction with depth. Seasonality can be identified at all depths with relatively wetter periods during winter and spring. The SM is clearly in phase with large precipitation events with highest amplitudes at the 5 cm measurement sensor in the topsoil. The temporal changes are more alleviated at 20 cm and 50 cm depth because the effect of climate forcing is damped with depth. Additionally, a lower porosity, soil water redistribution and root water uptake reduce the temporal dynamics of the SM in the subsoil ([Rosenbaum et al., 2012](#)). Besides the temporal variability, there is a distinct spatial variability in the Wüstebach catchment which is caused by the spatial heterogeneity of the soil properties. The SM dynamics at 20 stations that are located in the riparian zone are wetter and smoother compared to the remaining 84 stations in the hillslope zone which is due to altered hydraulic soil properties alongside groundwater influence in the riparian zone. As indicated by the SM data, the spring and summer of 2011 is an exceptional dry period (European drought 2011).



**Fig. 2.** The top panel shows daily averaged soil moisture (volumetric water content) at the three sensor depths (5, 20 and 50 cm) along with daily precipitation and discharge. Additionally, the results of the inverse modeling in HYDRUS 1D are given for 20 cm depth. The middle panel depicts daily actual evapotranspiration and reference/potential evapotranspiration. The bottom panel presents the accumulated water budget components for three separate hydrological years.

Therefore, this dataset which contains such an extreme period is very suitable for an inter-comparison study because it poses distinct challenges to the hydrological models.

Following the findings by Qu et al. (2015), SM variability in the Wüstebach catchment can be well represented by a 1D unsaturated flow model based on the Van Genuchten Mualem (VGM) model. Therefore a numerical solution of the Richard's equation as implemented in HYDRUS 1D can be utilized to calibrate the VGM parameters (Qu et al., 2014). This approach was first undertaken by Bogena et al. (2013) and later applied by Fang et al. (2015) to estimate VGM parameters for an integrated catchment model. Fig. 2 shows that the inverse HYDRUS 1D model produces plausible results.

#### 4. Methods

##### 4.1. Model description

Many attempts have been made that bring forward frameworks to classify distributed hydrological models (Kampf and Burges, 2007). Most recently and most relevant for the models used in this study, Maxwell et al. (2014) suggest to classify models based on their solution scheme, coupling and grid orientation. This classification can offer a broader perspective to the three distributed hydrological models assessed in this study: HydroGeoSphere (HGS; Panday and Huyakorn (2004) and Therrien et al. (2010)), MIKE SHE (MSHE; Abbott et al. (1986) and Graham and Butts

**Table 1**

Hydraulic soil parameters for the Richard's equation with the Van Genuchten Mualem model. Parameter values are obtained from inverse modeling in HYDRUS 1D.

	Alpha [1/cm]	n	Ks [m/s]	Porosity	Residual saturation	Anisotropy factor
Litter layer	0.0264	1.29	2.31E–05	0.87	0	20
Layer 1	0.01	1.26	9.30E–05	0.57	0.12	
Layer 2		1.19	1.73E–04	0.49	0.15	
Layer 3		1.21	1.14E–05	0.43	0.12	

**Table 2**

Description of the homogeneous scenario (Sc#1) and heterogeneous scenario (Sc#2)

Scenario	Vertical	Horizontal – porosity	Horizontal – other soil properties	Vegetation
#1	Litter layer	Homogeneous	Homogeneous	Homogeneous
#2	+ 3 soil layers	Distributed soil types (31)		LAI = 4 RD = 0.5

(2005)) and ParFlow-CLM (PCLM; Ashby and Falgout (1996) and Kollet and Maxwell (2008)).

#### 4.1.1. HydroGeoSphere

HydroGeoSphere (HGS) is a fully integrated hydrological model that solves the 3D Richards' equation for subsurface flow with the numerical finite difference method. Surface and channel flow are expressed by 2D and 1D diffusion wave approximations of the Saint Venant equation, respectively (Table 3). The model domain of the Wüstebach is discretized with a triangulated network containing 805 nodes which represents an average resolution of 25 m. Additional 164 nodes are added to the riparian zone for a refined description of the channel topography. The combination of coarser elements at the hillslopes and a finer discretization along the river is found suitable to best represent the topography. The vertical discretization consists of 24 numerical layers with increasing thickness with depth (2.5–10 cm). A more detailed description of the HGS model setup is presented by Cornelissen et al. (2014). Another recent application of the HGS model is given by De Schepper et al. (2015).

#### 4.1.2. MIKE SHE

MIKE SHE (MSHE) is a grid based hydrological model which comprises coupled modules that describe 3D groundwater flow (finite difference), 1D unsaturated flow (Richards' equation), 2D overland flow and river routing (Table 3). The model is not considered fully integrated; however full coupling between the modules is performed at each time step. For the Wüstebach model setup the unsaturated zone is discretized with an increasing thickness of the numerical layers with depth (5–20 cm). The lateral discretization

of the structured grid is 10 m. One major limitation of the MSHE model is the coupling between the saturated- and unsaturated-zone because numerical layers are defined independently for each zone. Recent applications of the MIKE SHE model are given by Ridler et al. (2014) or He et al. (2015)

#### 4.1.3. ParFlow-CLM

ParFlow-CLM (PCLM) is a grid based, fully integrated groundwater flow model that solves the Richards' equation in 3D. The coupled land surface model CLM simulates the land surface energy mass balance components and thus PCLM has a more detailed physical description than HGS and MSHE that use potential evapotranspiration as model input. PCLM simulates 2D surface flow by solving the kinematic wave equation (Table 3). The vertical discretization of the Wüstebach model in PCLM consists of 2.5 cm thick layers from terrain to bedrock. The lateral discretization of the structured grid is 10 m. In order to reduce the total number of computational cells the terrain following grid (Maxwell, 2013) was chosen. Fang et al. (2015) present a more elaborate description of the Wüstebach model setup in PCLM. Further, a very recent application of PCLM on the Rur catchment, of which the Wüstebach is a headwater catchment, can be found in Rahman et al. (2014).

#### 4.2. Model comparison

This section assesses similarities and main differences between the three hydrological models used in this study (Table 3). All models integrate subsurface, surface and land-surface processes. A key similarity of the models is their coupled solution of the Richards' equation describing flow in the subsurface and the Saint Venant equation describing surface flow. Maxwell et al. (2014) present the incorporated equations and coupling strategies in more detail. With respect to subsurface flow, HGS and PCLM are most advanced, as flow is fully integrated using the 3D Richards equation. Contrary, MSHE couples unsaturated 1D flow with saturated 3D flow in the subsurface. HGS and MSHE operate with potential evapotranspiration as forcing data and actual evapotranspiration and interception are estimated following the approach by Kristensen and Jensen (1975). On the other hand, ParFlow incorpo-

**Table 3**

Description of the modules incorporated in the three models assessed in the inter-comparison study.

Module	HGS	MSHE	PCLM
Interception	Storage approach – Function of LAI	Storage approach – Function of LAI	Storage approach – Function of LAI
Evapotranspiration	Potential ET: FAO Penman-Monteith equation	Potential ET: FAO Penman-Monteith equation	Common Land Model (CLM): Monin-Obukhov similarity theory
Unsaturated zone	Actual ET: Kristensen & Jensen	Actual ET: Kristensen & Jensen	3D Richards flow
Saturated zone	3D Richards flow	1D Richards flow	3D Richards flow
Overland flow	3D Richards flow	3D Darcy flow	Saint Venant-2D Diffusive Wave
Routing	Saint Venant-2D Diffusive Wave	Saint Venant-2D Diffusive Wave	1D Diffusive Wave
Numerical technique	1D Diffusive Wave	1D Diffusive Wave	Finite Difference – Terrain following
Grid	Finite Difference	Finite Difference	
Model structure	Unstructured	Structured	Structured
	Fully integrated	Fully coupled	Fully integrated

rates the Common Land Model (CLM; [Dai et al. \(2003\)](#)) which is driven by hourly climate data and couples land surface process with the subsurface to estimate actual evapotranspiration and interception. MSHE and PCLM use structured meshes for the lateral discretization whereas HGS utilizes an unstructured mesh. Unstructured grids provide more flexibility in representing complex features but are computationally more demanding than structured grids.

#### 4.3. Model setup

The three models have the same initial conditions, are set to a 3.5 year simulation period from 01-January-2010 to 30-April-2013 and are executed on hourly time steps. The quantitative [inter-comparison of the three models focuses only on the last two hydrological years \(01-May-2011 to 30-April-2013\)](#) which allows a [spin up period of 1.5 years](#). The models are initiated with a groundwater table map obtained from a MSHE simulation.

The topographical watershed boundary defines the model boundary in the MSHE and PCLM model. In contrast the HGS model domain is determined by the smaller test site boundary because the model was originally set up to focus on predicting SM dynamics ([Fig. 1](#)). The simulated discharge in the HGS model has thus to be scaled to the size of the topographic model domain ([Cornelissen et al., 2014](#)). This is unproblematic for the inter-comparison because climate and vegetation is regarded invariant in space and SM is not affected by the scaling.

#### 4.4. Scenario description

The distinct complexity of the spatio-temporal variability of the observed SM patterns at the Wüstebach catchment is investigated by [Korres et al. \(2015\)](#). This underlines that a thorough acquisition of spatial input data and parameters are required to allow an adequate simulation of the observed SM patterns by a hydrological model. [Stisen et al. \(2011\)](#) illustrate the importance of distributed model parameters by illustrating that homogeneous parameters perform well in predicting an aggregated hydrological variable such as discharge but perform poorly in predicting the patterns of spatial explicit variables.

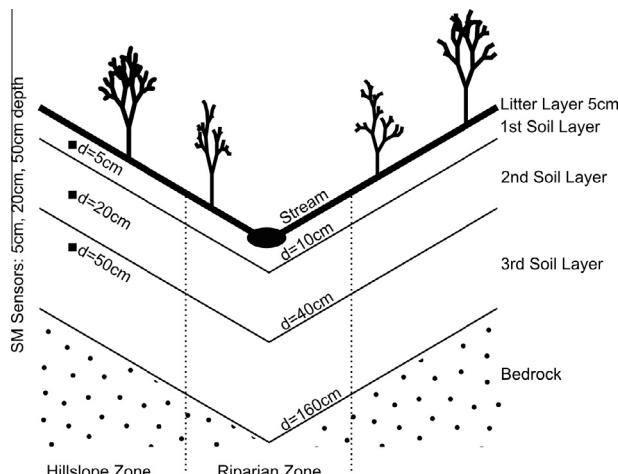
It is expected that the digital elevation model of the Wüstebach is well described by the 10 m resolution used in this study. Vegetation and climate forcing is anticipated to be homogeneous at

such a small scale as the Wüstebach catchment. This leaves the spatial distribution of soil parameters as the major unknown driver of spatially explicit model output. Therefore, this study investigates the effect of moving from homogeneous to heterogeneous soil parametrization in the three given models with respect to observed SM patterns. In this context, two scenarios are developed that focus on the lateral distribution of soil porosity. The sequence of soil horizons in the vertical dimension and the overlaying homogeneous litter layer are kept untouched in both scenarios. Following [Bogena et al. \(2013\)](#), the soil profile is differentiated into four soil types with specific hydraulic properties: A soil covering litter layer (+5 to 0 cm), a topsoil A horizon (0–10 cm), an intermediate B horizon (10–40 cm) and a subsoil C horizon (40–160 cm) which overlays the bedrock ([Fig. 3](#)). [Fang et al. \(2015\)](#) stress the importance of the litter layer with respect to both predicted discharge and SM dynamics and further conduct a sensitivity analysis of the anisotropy of the saturated hydraulic conductivity. An anisotropy factor of 20 was found best suitable to effectively represent the fast lateral interflows at the Wüstebach site.

The three given hydrological models utilize VGM parameters to solve the Richards' equation. These soil physical parameters for each soil layer are estimated by means of a HYDRUS 1D calibration as described by [Bogena et al. \(2013\)](#) and [Fang et al. \(2015\)](#). The obtained parameter values are given in [Table 1](#). This approach is favorable because the parameter calibration in a surrogate model will provide an unbiased parameter set to the inter-comparison. Additionally, the inverse estimation of the soil properties is computationally much more feasible in a 1D model than in a 3D model. The VGM parameters in the HYDRUS 1D model are inversely estimated against the SM dynamics aggregated to the entire catchment at three depths: 5 cm, 20 cm and 50 cm ([Fig. 2](#)). The same time period and the same meteorological forcing, as used in the inter-comparison, are applied to the calibration. The single VGM parameter not used in the calibration is porosity which is fixed to the maximum value of observed daily averaged SM during the study period for each soil layer. It is assumed that this approach is valid because saturation was reached at least once during the study period after strong rainfall events.

The first scenario (Sc#1) represents the homogeneous case as shown by [Table 2](#). The second scenario (Sc#2) is the same as Sc#1 besides additional porosity heterogeneity in all three soil layers. [Fig. 4](#) exemplifies how the porosity heterogeneity is derived from the SM data. Inverse distance weighted interpolation is conducted at all three depths based on the maximum observed SM. The interpolated map is then classified, individually for the riparian- and hillslope-zone, into three equally sized classes of porosity (low, medium and high) based on percentiles of the maximum SM values at the sensors. In the following, the classified porosity maps at all three depths are stacked on top of each other and same sequences of low, medium and high porosity are grouped into 31 soil types ([Fig. 4](#)).

It can be anticipated that changes in porosity will not affect the SM dynamics as such, but rather it functions as a scaling factor. In contrast, the remaining VGM parameters certainly govern the temporal SM dynamics, as they alter the shape of the water retention curve. Nevertheless, porosity is selected for the heterogeneous scenario (Sc#2) because trustworthy data exist on its spatial distribution.

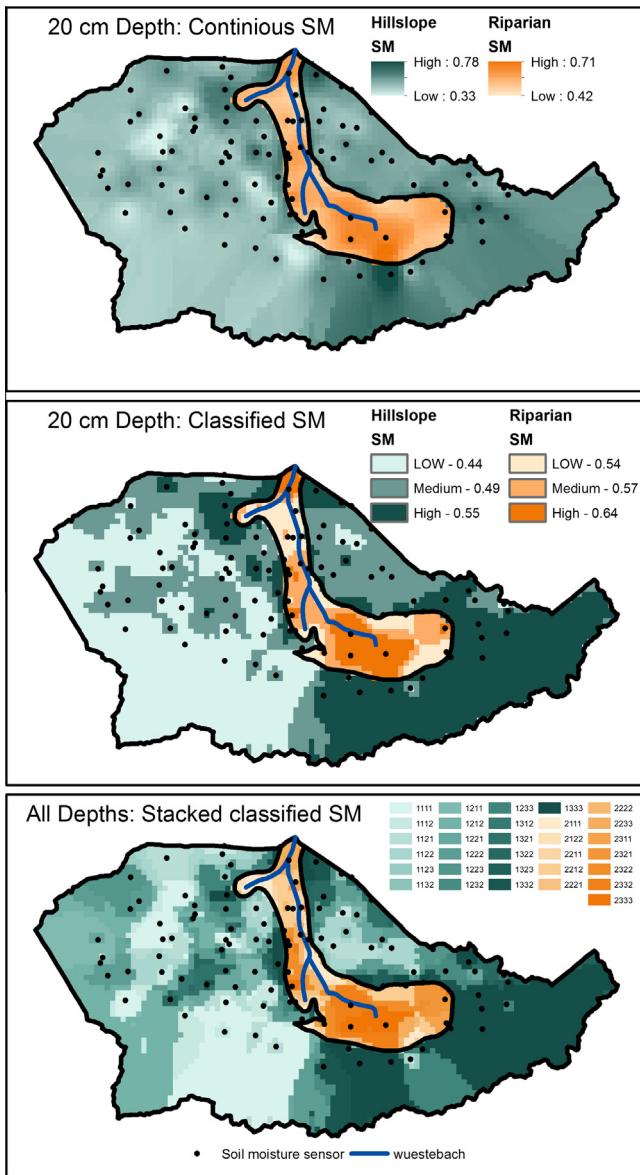


**Fig. 3.** The conceptual geological model for the forested Wüstebach catchment: Litter layer and three soil layers overlaying the bedrock formation. Soil moisture is measured at three depths; one sensor in each soil layer and the soils can generally be classified into a hillslope- and a riparian zone.

#### 4.5. Spatio-temporal analysis

##### 4.5.1. EOF analysis

The EOF analysis is a widespread statistical method for the analysis of large spatio-temporal datasets and it is frequently applied



**Fig. 4.** The distributed porosity map is derived by (top) interpolating the maximum soil moisture values (volumetric water content) at the stations. Followed by, (middle) grouping the map into low, middle and high porosity and lastly, (bottom) stacking the grouped porosity maps of all three depths to generate one single map of 31 soil types. First digit 1 = hillslope or 2 = riparian. Remaining digits 1 = low, 2 = middle or 3 = high porosity for the 1st, 2nd and 3rd soil layer, respectively.

on soil moisture data (Korres et al., 2010; Perry and Niemann, 2007). It can decompose the variability of a dataset into a set of orthogonal spatial patterns (EOFs) which are invariant in time and a set of loadings which are invariant in space. Typically, the

EOF analysis is applied to decompose large observational dataset however the methodology receives more recent attention in spatial model validation of distributed hydrological models as well (Fang et al., 2015; Koch et al., 2015). Graf et al. (2014) conduct an EOF analysis on the observed soil moisture data at the Wüstebach site and utilized the results in a cluster analysis of the loadings to identify a “wet” and a “dry” cluster. This classification of “wet” and “dry” days is presented in the result section and is used to assess the seasonality of the soil moisture patterns through a separated validation.

#### 4.5.2. Taylor diagrams

Three different models with two scenarios each and data at 104 stations at three depths generate large amounts of data which poses a major challenge to the presentation of the results but also to which performance criteria to choose. Taylor diagrams (TDs) are employed in the results section of this study for a complete assessment of all data. TDs integrate several performance metrics of various models in a single plot and thus they are perfectly suitable for an inter-comparison study. A thorough description of TDs is given by Taylor (2001). Recently TDs receive increasing attention in inter-comparison studies of e.g. hydrological models (Orth et al., 2015) and of atmospheric climate models (Stevens et al., 2013; Volodko et al., 2013). TDs are computed and plotted using the R package “openair” (Carslaw, 2012). The performance metrics that are addressed in a TD are (1) the comparison of the standard deviation in the observed and the predicted dataset (2) the correlation between them and lastly (3) a bias corrected root mean squared error between both datasets.

#### 4.5.3. Variogram analysis

Vereecken et al. (2014) suggest to use geostatistics such as semivariograms for the analysis of spatio-temporal soil moisture patterns at field scale. Semivariograms are based on two point spatial statistics and describe the degree of spatial autocorrelation of a variable. The semivariances are calculated with the R package “gstat” (Pebesma, 2004) and exponential variogram models are found to give the best fit to the experimental semivariances. Rosenbaum et al. (2012) conducted an analysis of the temporal changes of the semivariances for selected events in the Wüstebach soil moisture data. A more complete temporal analysis of the dynamic changes in the spatial autocorrelation structure of the 2011 soil moisture data in the Wüstebach catchment is presented by Korres et al. (2015). Generally, it was found that spatial variability is elevated in drier periods and that it decreases with depth. Additionally, the semivariances of simulated soil moisture patterns using HGS (Cornelissen et al., 2014) were incorporated in the study and overall lower semivariances were detected in the predicted data because the model underestimated small scale heterogeneities in the soil moisture patterns.

**Table 4**

Accumulated water balance components for two hydrological years. The residuals (in bold) are given in percentage of the precipitation input.

		May 11–April 12				May 12–April 13			
		ET [mm]	Q [mm]	P [mm]	Res [%]	ET [mm]	Q [mm]	P [mm]	Res [%]
HGS	#1	604	562	1260	<b>7.5</b>	563	746	1274	<b>-2.7</b>
	#2	610	552		<b>7.8</b>	562	748		<b>-2.8</b>
MSHE	#1	581	567	1260	<b>8.9</b>	588	737	1274	<b>-4.0</b>
	#2	582	562		<b>9.2</b>	588	723		<b>-2.9</b>
PCLM	#1	643	662	1260	<b>-3.6</b>	622	772	1274	<b>-9.4</b>
	#2	643	649		<b>-2.5</b>	622	763		<b>-8.7</b>
Observed		609	596	1260	<b>4.4</b>	595	793	1274	<b>-8.9</b>

## 5. Results and discussion

The following section contains the presentation and discussion of the main findings in the inter-comparison study. First, water balance components and discharge dynamics are presented briefly and afterward the spatio-temporal SM dynamics are studied in more detail. Lastly, the correlation between the model errors is investigated to assess the implications of the different model structures.

### 5.1. Water balance

**Table 4** shows the predicted water balance components actual evapotranspiration (aET) and runoff discharge ( $Q$ ) and compares accumulated values with observations and the respective precipitation ( $P$ ) in the forcing data. The residuals in the observed data are 4.4% and –8.9% of the precipitation for the first (May 2011 to April 2012) and the second (May 2012 to April 2013) hydrological year, respectively. [Graf et al. \(2014\)](#) expect either measurement error or storage term depletion as a possible explanation of the large residual of 114 mm in the second hydrological year. The models react analogously with positive and negative residuals in the first and second hydrological year, respectively. Only PCLM opposes the trend with negative residuals in both scenarios in the first hydrological year. The intra-model differences between both scenarios are marginal for all models. The simulated aET is largest for PCLM but shows almost identical dynamics with the aET predicted by the Kristensen & Jensen approach in MSHE with a correlation of 0.95.

### 5.2. Discharge

**Fig. 5** depicts the hydrographs originating from the three hydrological models with two scenarios each. All simulated hydrographs capture the overall dynamics correctly and the differences between the models dominate over the intra-model differences of the two scenarios. This indicates that discharge is an aggregated measure with limited sensitivity to spatially distributed input, as already identified by [Stisen et al. \(2011\)](#). As expected, observed peaks in the dry periods which are initiated by preferential flow are mostly missed by the models. PCLM shows quick responses to single rainfall events and an overall overestimated flow in dry periods.

**Table 5**

Nash Sutcliffe Efficiency based on untransformed and log transformed discharge values.

Scenario	$Q$			Log ( $Q$ )			
	HGS	MSHE	PCLM	HGS	MSHE	PCLM	
Nash Sutcliffe	#1	0.77	0.83	0.75	0.27	0.76	0.58
Efficiency	#2	0.78	0.86	0.79	0.58	0.84	0.63

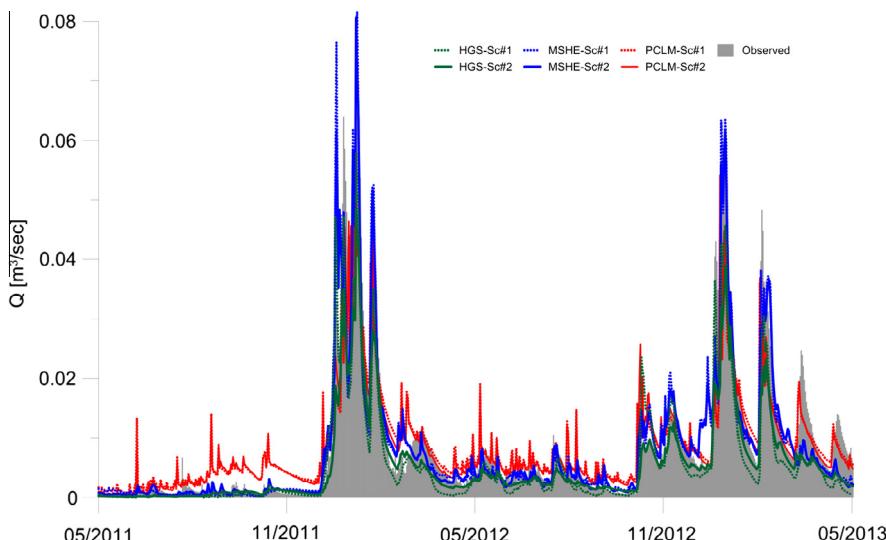
**Table 5** states the model performance in terms of Nash Sutcliffe Efficiency (NSE) which is computed based on untransformed and logarithmic discharge values. The evaluation of logarithmic discharge supplements the evaluation because it emphasizes the low flows. All models give satisfying results in terms of NSE, however the NSE on log data underlines the limitations of PCLM with respect to low flows. In comparison, the performance in discharge increases only to a minor degree between Sc#1 and Sc#2. This underlines that the inclusion of soil heterogeneity (Sc#2) does not significantly affect the runoff generation processes in any of the three models.

### 5.3. Soil moisture

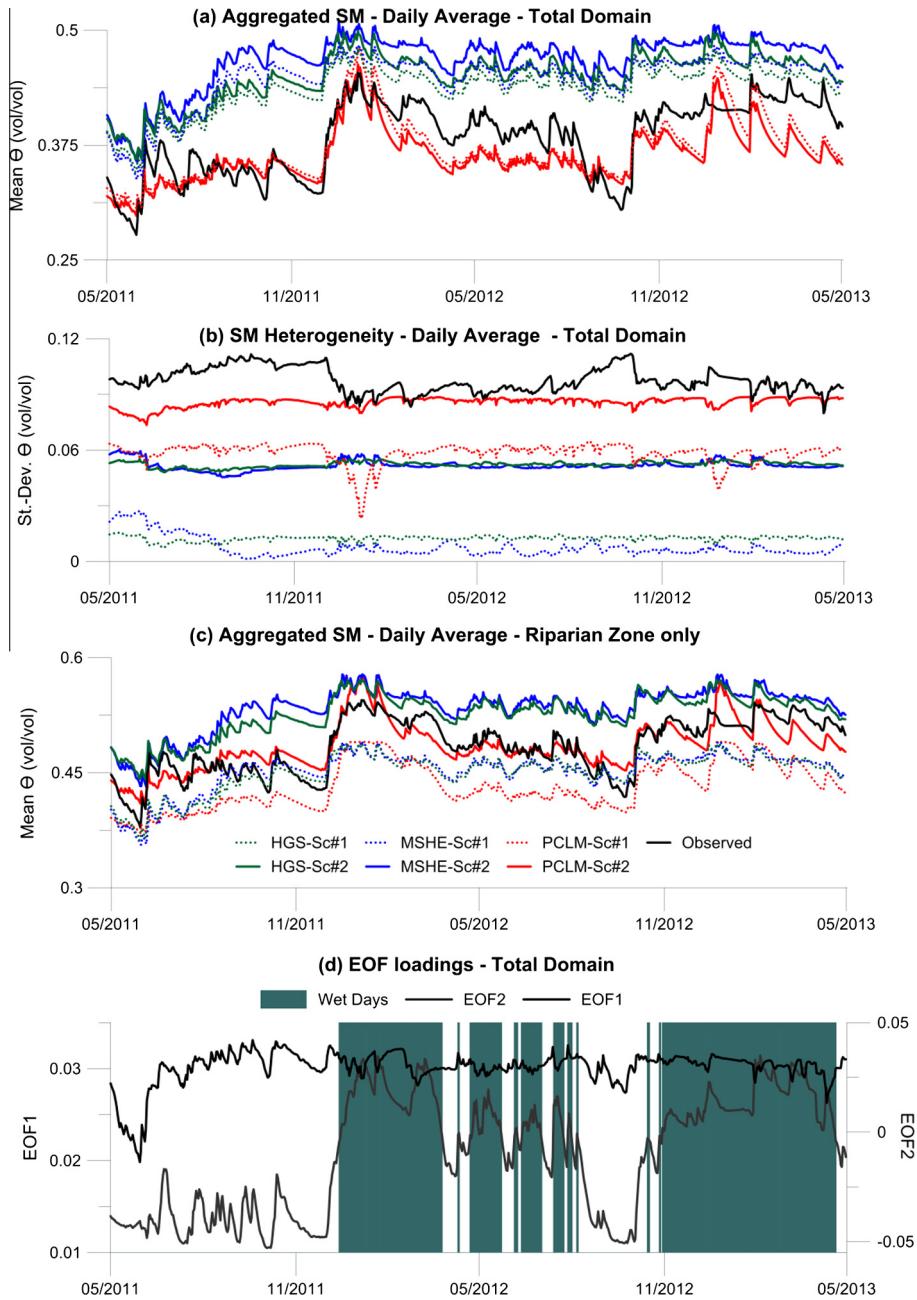
#### 5.3.1. Catchment scale dynamics

The spatio-temporal variability of soil moisture (SM) patterns is the main focus of this inter-comparison study. The temporal variability is assessed by comparing the SM dynamics aggregated to daily averages based on all 104 SM stations. Additionally, the heterogeneity is evaluated by means of the daily standard deviation of SM among all 104 stations. The aggregated SM over the entire catchment and the heterogeneity are presented for both scenarios at 20 cm depth in **Fig. 6(a)** and (b). The performance at all depths is given in **Table 6** by a set of statistical measures.

HGS and MSHE have a distinct systematic positive bias with an overall good estimation of the SM dynamics at catchment scale (**Fig. 6(a)**), which gets supported by the correlation coefficient in **Table 6**. Differences might be caused by differences in aET or by the differences in vertical discretization. PCLM gives an unbiased prediction of the SM dynamics at field scale with clear differences in the temporal dynamics compared to the other two models. Besides the bias, all models appear to have distinct problems with the drying out of the soil. Analogue to the discharge results, the differences between the two scenarios are marginal at field scale



**Fig. 5.** Simulated discharge at the outlet of the Wüstebach catchment for both scenarios in HGS, MSHE and PCLM.



**Fig. 6.** (a) Daily aggregated soil moisture at 20 cm depth simulated by both scenarios in HGS, MSHE and PCLM. (b) Daily standard deviation as a measure of heterogeneity at 20 cm depth. (c) Daily aggregated soil moisture at 20 cm depth for a subset of 20 stations located in the riparian zone. (d) Time series of loadings originating from the EOF analysis and the resulting classification into wet and dry days derived from a cluster analysis.

which underlines that aggregated SM is not sensitive to distributed input either. Table 6 supports this by means of a set of statistical measures that detect only little change in performance between the scenarios at all depths. Overall, the performance does not change significantly with depth aside from PCLM where the mean error changes from positive to negative.

The standard deviation, as a measure of heterogeneity, clearly emphasizes that the variability increases between the scenarios (Fig. 6(b)). The heterogeneity in PCLM is predicted with more accuracy than in the two other models. The predicted SM patterns in PCLM Sc#1 are entirely topography driven. In contrast, topography generates a similar degree of spatial SM heterogeneity in PCLM than topography plus porosity heterogeneity (Sc#2) in HGS and MSHE. This indicates an enhanced topography feedback in PCLM

compared to the other models. The overall temporal dynamics in SM heterogeneity, expressed through the standard deviation, are completely missed by all models which is an indicator for a wrong representation of the seasonality. The seasonality of the observations is characterized by more homogeneous conditions during the wetter periods.

Large differences in hydraulic properties between the hillslope and the riparian zone are anticipated to cause significantly diverging SM patterns. Therefore, SM is assessed for a subset of 20 stations that are located in the riparian zone. Fig. 6(c) shows the daily averaged SM predicted by the three models for the subset of stations at 20 cm depth and compares it to the observed SM dynamic in the riparian zone. Sc#1 simulates dryer SM dynamics due to a lower porosity. On the other hand, the porosity

**Table 6**

Mean error (ME), root mean square error (RMSE) and correlation coefficient (*R*) for aggregated daily soil moisture dynamics.

Depth	Model	Aggregated SWC		
		ME	RMSE	<i>R</i>
5 cm	HGS-Sc#1	0.10	0.10	0.80
	HGS-Sc#2	0.11	0.11	0.81
	MSHE-Sc#1	0.11	0.12	0.75
	MSHE-Sc#2	0.14	0.14	0.76
	PCLM-Sc#1	0.08	0.08	0.82
	PCLM-Sc#2	0.08	0.09	0.81
20 cm	HGS-Sc#1	0.05	0.06	0.80
	HGS-Sc#2	0.07	0.07	0.82
	MSHE-Sc#1	0.06	0.07	0.72
	MSHE-Sc#2	0.08	0.09	0.72
	PCLM-Sc#1	-0.01	0.03	0.80
	PCLM-Sc#2	-0.02	0.03	0.81
50 cm	HGS-Sc#1	0.07	0.07	0.80
	HGS-Sc#2	0.07	0.08	0.82
	MSHE-Sc#1	0.07	0.07	0.70
	MSHE-Sc#2	0.08	0.09	0.71
	PCLM-Sc#1	-0.03	0.03	0.84
	PCLM-Sc#2	-0.04	0.04	0.85

heterogeneity in Sc#2 generates overall wetter conditions in the riparian zone which is in agreement with the observations. Comparing Fig. 6(a) and (c) clearly underlines that porosity is simply a scaling parameter that shifts the SM up or down and does not affect the dynamics as such.

Fig. 6(c) depicts the resulting time series of loadings from the EOF analysis. The first EOF explains 85.8% of the variance in the SM dataset while the second EOF accounts for additional 6.2%. The loadings for the first EOF are entirely positive and show similar dynamics as the mean soil moisture in Fig. 6(c). The associated loadings of the second EOF occur with both positive and negative sign. Following the cluster analysis conducted by Graf et al. (2014), the loadings can be grouped into two clusters that are interpreted as “wet” and “dry” days. This classification is applied in the following to emphasize on seasonality in the model assessment. Furthermore, Stockinger et al. (2014) present a study on transient time distribution using stable isotopes and found a critical SM value of 35 % at field scale for the Wüstebach catchment where the catchment switches between two hydrological responses. The threshold value coincides well with the wet and dry differentiation by means of the EOF analysis.

### 5.3.2. Taylor diagrams

As introduced before, Taylor diagrams (TDs) address multiple performance metrics in a single plot. The radial distance from the origin depicts the standard deviation of the modeled soil moisture and the observed standard deviation is represented by a black dashed circle. Thus, the relative position of the model to the dashed line gives information on the relative magnitude of variation between model and observation. The distance to the observation constitutes the centered (bias corrected) root mean squared error (RMS) and the golden concentric dashed lines that emanate from the observation show the RMS values and aid interpretation. Finally, the correlation coefficient (*R*) is represented by the arc and models that are closer to the *x*-axis have a higher correlation and gray lines help to underline specific *R* values. Opposed to Fig. 6 that shows results derived from daily averages at the sensor in 20 cm depth, the TDs in Fig. 7 are based on all available data (730 days \* 104 stations with 3 depths each). In order to focus on the seasonality, the data is separated into wet (358) and dry (372) days in the period of May 2011 to April 2013.

The previous results stress that aggregated SM data is a suitable measure for temporal dynamics but an ambiguous measure for spatial performance. Therefore, the TDs are based on all individual observations at the individual stations. Unlike the aggregated SM data in Fig. 6(a), the TDs underline a significant increase in performance between Sc#1 and Sc#2 for all models at all depths. The topsoil at 5 cm and 20 cm depth is affected by climate forcing and the models have difficulties to represent the observed seasonality accordingly which results in a more accurate performance for the wet days. On the other hand, seasonality in the subsoil (50 cm depth) is less distinct and hence the models show similar performance for the wet and the dry days. The TDs underline a distinct similarity between the SM predictions by HGS and MSHE. PCLM is associated with a higher standard deviation which is closer to the observation but at the same time RMS and *R* attest PCLM a poorer performance compared to the other two models. This emphasizes that PCLM simulates an appropriate range from wet to dry, yet the spatial allocation is imperfect. The effect of topography as the sole driver of spatial variability (Sc#1) is very evident in the PCLM predictions, whereas HGS and MSHE only show a topography feedback during the dry days.

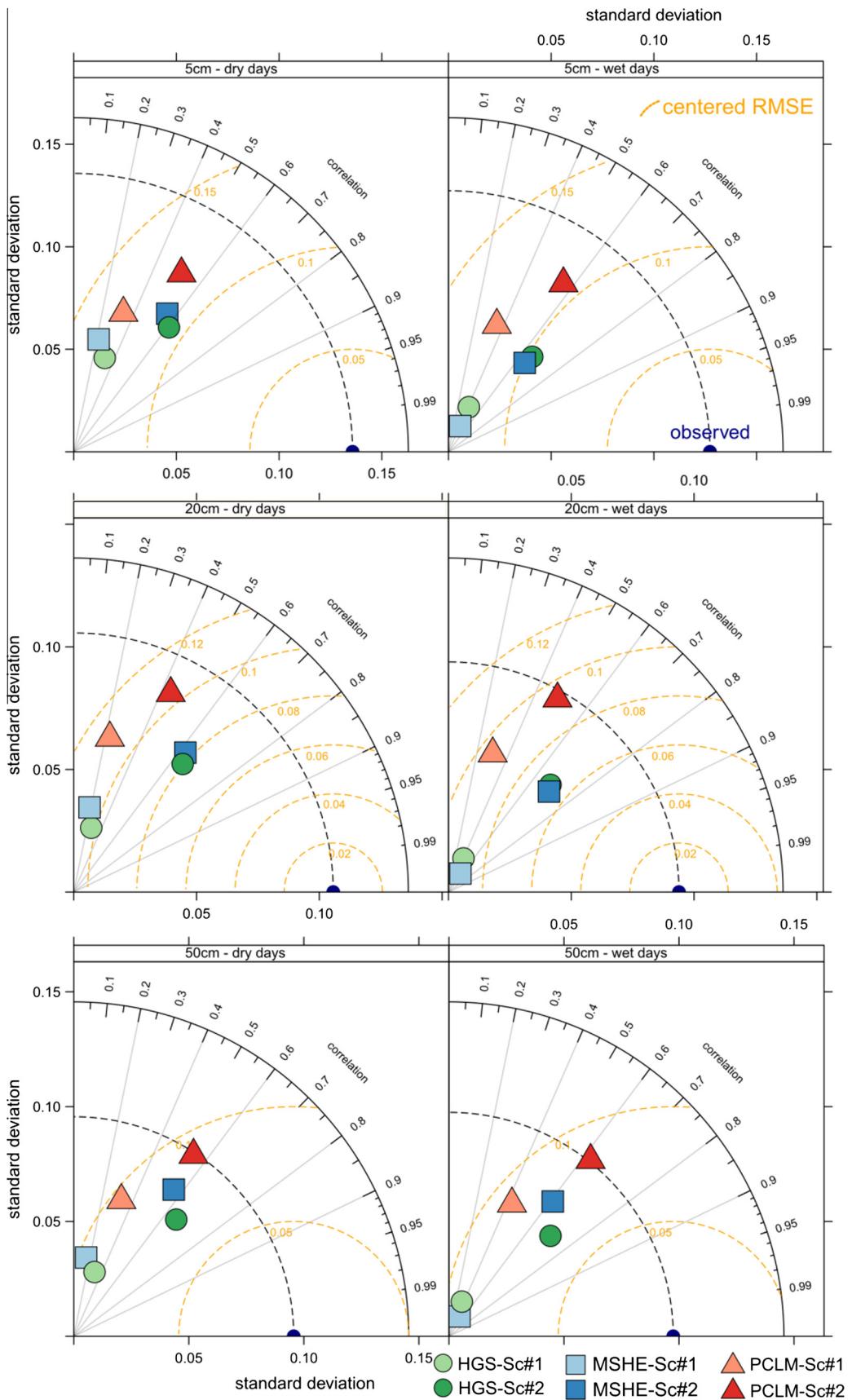
### 5.3.3. Variogram analysis

Fitted exponential variogram models are presented in Fig. 8 for the average soil moisture patterns in all wet and all dry days at 20 cm depth. The experimental semivariances are only given for the observed data. The seasonal distinction into a wet and a dry period results in differences between the fitted variogram models for the two periods. Analogous to the findings of Korres et al. (2015), a higher spatial variability is evident in the observed SM patterns at dry conditions. Overall, the models fail to reflect the degree of spatial correlation correctly; nevertheless a clear increase is apparent from Sc#1 to Sc#2. Only PCLM shows a measurable variability in Sc#1 which is caused by the increased topography feedback opposed to HGS and MSHE that show no spatial autocorrelation at all in Sc#1. All models are missing the enhanced spatial variability in the dry period and have more or less similar variogram models in both seasons. These findings are in line with Korres et al. (2015), where missing heterogeneity was identified to be accountable for the lack of spatial variability in the model output.

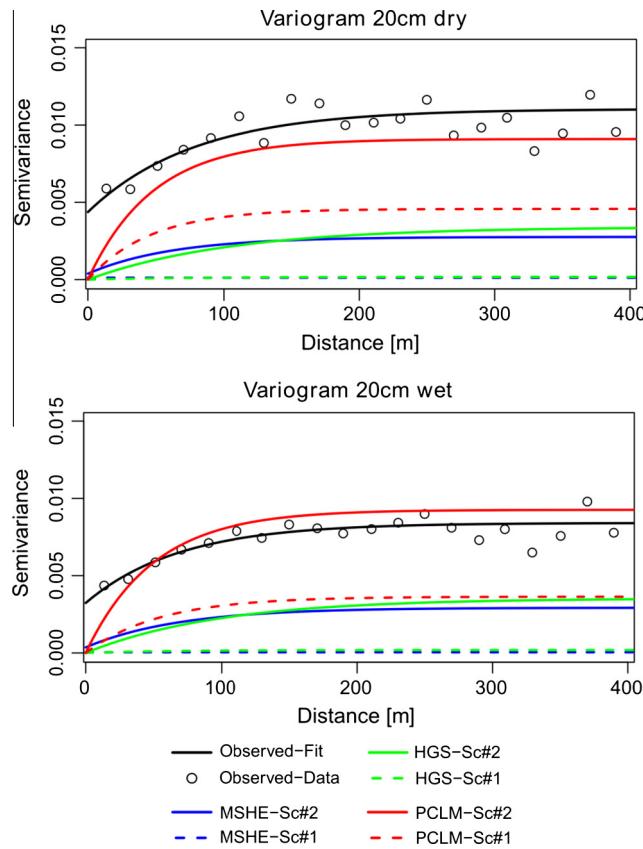
### 5.3.4. Soil moisture maps

Interpolated SM maps at 20 cm depth are shown in Fig. 9 which allows a better visual assessment of the differences in predicted SM patterns of the individual models and the scenarios. Although all models have a spatially distributed output it was chosen to interpolate the SM data at the stations in the same way the observed SM data is interpolated. This enables a fair comparison because all data are treated in the same manner for visualization. In order to underline the seasonality in the SM variability a distinct wet day (25-12-2011) and dry day (10-Sep-2012) are shown. The differences between the observed SM patterns are striking; the wet day shows a dry pattern west of the stream and a wet part east of the stream. Opposed, the entire hillslope is homogeneously dried out at the dry day. A distinct similarity between both maps is locations of high saturation at local source areas in the riparian zone and at the outlet of the stream.

A first visual assessment reveals the distinct positive bias in the predicted SM patterns by HGS and MSHE and that the range from dry to wet is most accurately represented by the SM fields predicted by PCLM. Sc#1 in HGS and MSHE which is solely driven by topography has a minimal spatial variability in the simulated SM patterns and thus they are in poor agreement with the observed SM maps. PCLM Sc#1 simulates a more adequate spatial variability but wet areas are wrongly allocated in topographical sinks. Sc#2 introduces



**Fig. 7.** Model evaluation by means of Taylor Diagrams on daily stations based soil moisture data for both scenarios in HGS, MSHE and PCLM. All three sensor depths are included and the assessment is separated into wet and dry days to focus on seasonality.



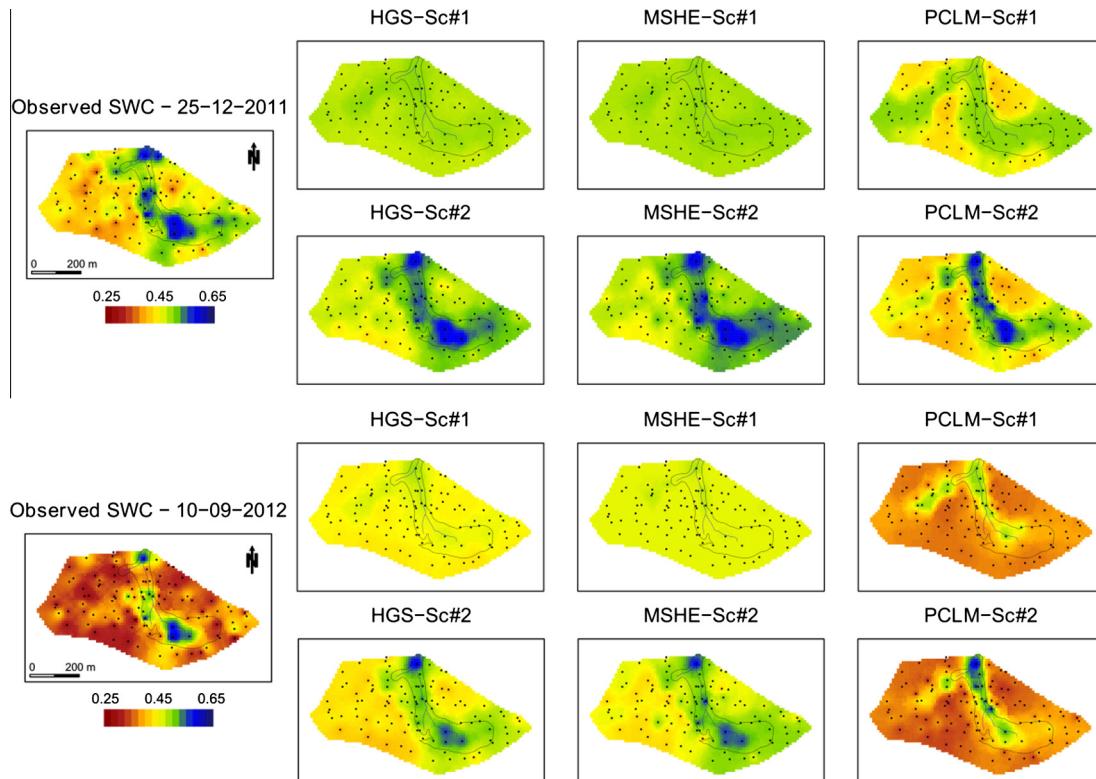
**Fig. 8.** Exponential variogram models for observed soil moisture data and both scenarios in HGS, MSHE and PCLM for the mean dry and wet soil moisture pattern. The experimental semivariances are only shown for the observed data.

additional heterogeneity in porosity and clearly increases the spatial performance at both days for all models. The pronounced wetness at the source areas in the riparian zone and the outlet are better represented. Also, the drier conditions west of the stream are adequately presented at the wet day. However, the dry day lacks the wet source areas in PCLM Sc#2.

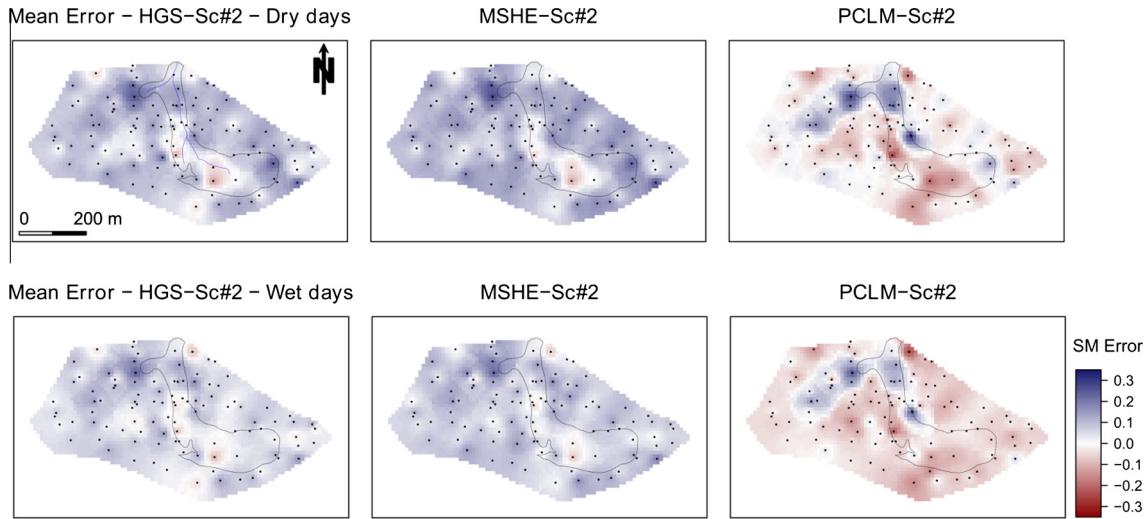
Interpolated maps of the mean error during the wet and dry days are given in Fig. 10 for Sc#2 at 20 cm depth. A noticeable similarity of all maps is that the source area is persistently simulated too dry by all models. The PCLM error maps are associated with a systematic wet bias at the topographical sink west of the stream at the hillslope. Opposed, other parts of the hillslope are systematically predicted too dry. HGS and MSHE have a distinct wet bias throughout both seasons. The error in the wet period is clearly lower than the error in the dry period but the spatial pattern of the error is consistent. In contrast, the pattern of the error map for PCLM changes between the dry and the wet season; some areas with a systematic dry error in the wet season change into a systematic wet error during the dry season.

#### 5.4. Error correlation

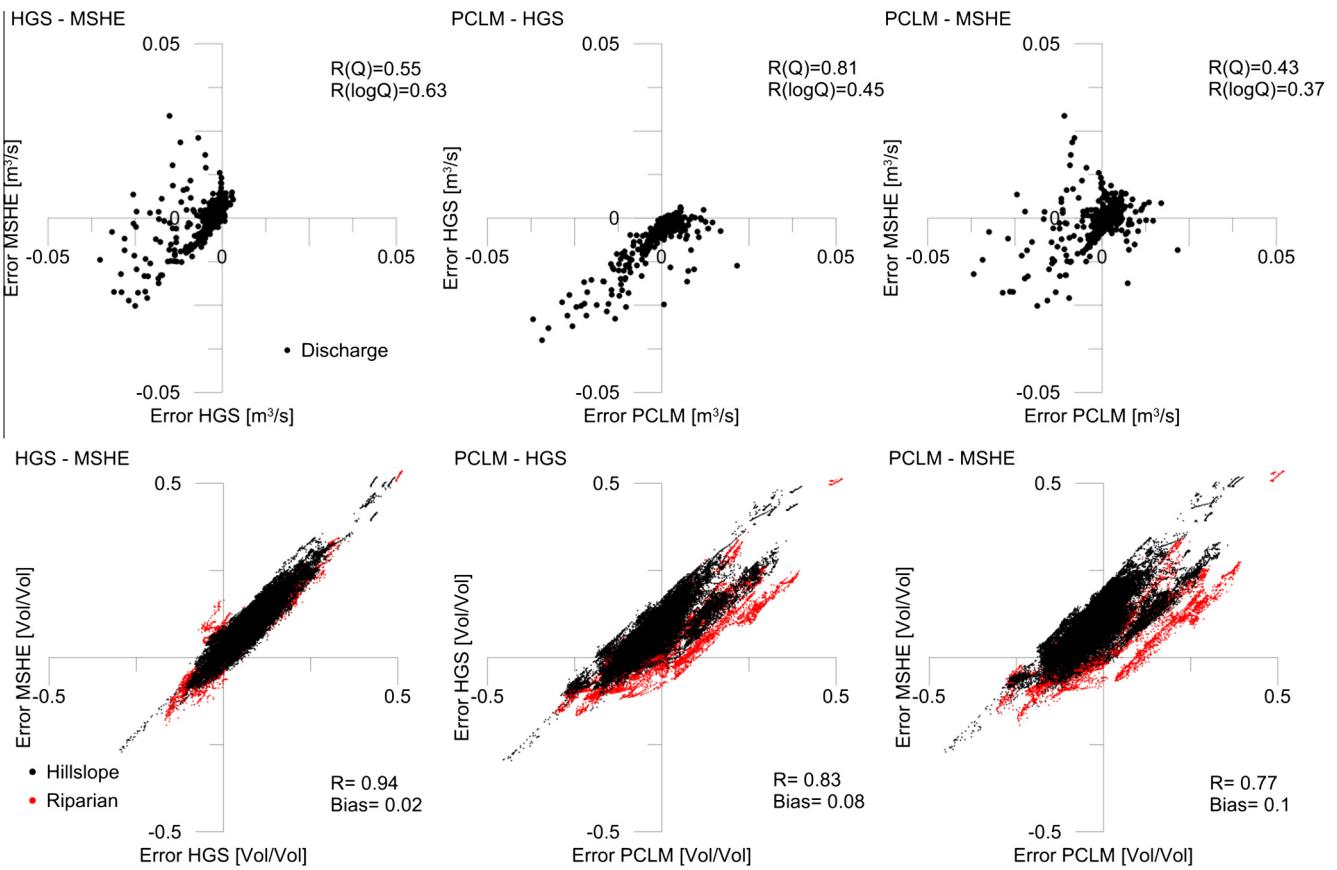
The three hydrological models that are assessed in this study are based on a comparable understanding of the physics but differ in their model structures (see Section 4.1). As models are always a simplification there will always be associated errors. In an extreme case where the errors of two models are independent, due to highly diverging model structures, the correlation between the errors will be very weak. Vice versa a strong correlation will be apparent for dependent errors caused by similar model structures. Fig. 11 depicts the error correlation for discharge and SM and quantifies the dependencies by means of the correlation coefficient.



**Fig. 9.** Interpolated (inverse distance weighted) soil moisture maps (volumetric water content) for one typical wet day (25-12-2011) and one typical dry day (10-09-2012).



**Fig. 10.** Interpolated maps of the mean error (volumetric water content) of all dry (top panel) and wet (bottom panel) days at 20 cm depth for the second scenario. Blue colors indicate that the model predicts too wet conditions and vice versa represent red colors too dry conditions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

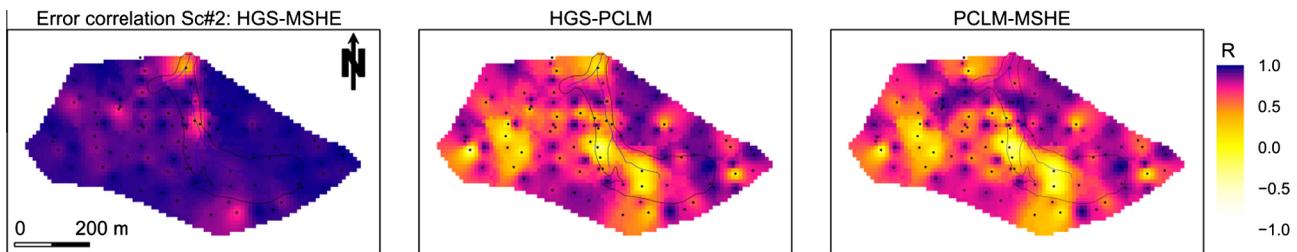


**Fig. 11.** Error correlation between the three models for daily discharge (top) and daily station based soil moisture (volumetric water content) at 20 cm depth (bottom). Additionally, the correlation coefficient ( $R$ ) and bias is given.

The error correlation of simulated discharge is highest between PCLM and HGS and the most distinct disagreement is evident between PCLM and MSHE. The error correlation of log transformed discharge allows emphasizing on low flows and it has its strongest correlation between HGS and MSHE because PCLM clearly over predicts low flows. The models largely agree well on negative errors which might indicate that some large discharge events in

the observed data do not coincide with large precipitation events in the forcing data.

Errors in the simulated SM at 20 cm depth for HGS and MSHE show an almost perfect correlation. Opposed, the errors of both models show a weaker correlation to the SM error in PCLM. Especially the error correlation between PCLM and MSHE shows most scatter. There is a systematical bias between the errors at the



**Fig. 12.** Interpolated map showing the error correlation per station over the 2-year evaluation period at 20 cm depth.

hillslope stations and the stations in the riparian zone. The previously found enhanced topography feedback in PCLM is more evident at the hillslopes because the hillslope zone has larger topographical gradients. All models are associated with extremely positive and negative errors but at the same time show a strong correlation among the large errors. This is caused by either inadequate spatial parametrization, small scale processes like preferential flow that the models do not incorporate or inconsistency between forcing- and observation-data (e.g. a large precipitation event is missed or SM measurement errors).

The error correlation at each station between the models can be interpolated to visualize the spatial significance of varying model structures (Fig. 12). HGS and MSHE show a strong error correlation at all stations besides the outlet which underlines that overall the differences in model structure between the models do not cause significant differences in the predictions. PCLM shows weak error correlations to HGS and MSHE at the source area and at the western part of the hillslopes, where high topographical gradients are present. The eastern part of the hillslope with lower gradients is less affected by the differences in model structure and thus it is associated with higher error correlations between PCLM and the other two models.

## 6. Conclusion

This study presents a thorough inter-comparison of three state of the art distributed hydrological models (HydroGeoSphere, MIKE SHE and ParFlow-CLM) at a small headwater catchment with focus on spatio-temporal soil moisture (SM) dynamics. The densely instrumented Wüstebach catchment poses major challenges to the models in terms of SM heterogeneity and seasonality.

Water balance components, discharge dynamics at the outlet and daily averaged SM at catchment scale are aggregated performance measures and thus they are found not to be sensitive to spatial input data. These measures are well suited for the evaluation of global temporal dynamics at the catchment scale but allow limited and ambiguous interpretations if spatial evaluation is of interest.

Including heterogeneous porosity clearly improves all models in terms of SM heterogeneity, temporal correlation as well as the spatial correlation structure of the predicted SM fields. Opposed, aggregated measures are only marginally improved by porosity heterogeneity. Nevertheless, porosity is simply a scaling factor of the SM dynamics. Future work must focus on other possibilities to further distribute the remaining VGM parameters.

The observed spatial patterns of SM and the hydrological behavior of the catchment are clearly affected by seasonality. Therefore, a separated evaluation of dry and wet days is advisable. The period is separated into dry and wet days following the results of an EOF analysis. All models show a poorer performance during the dry periods. This problem can be attested to the water retention curve in the Van Genuchten Mualem model which is essentially a wetting curve and does not consider soil hysteresis correctly.

The most prominent distinction of the simulated SM patterns is the enhanced topography feedback in PCLM. Topography is a well known driver of soil moisture patterns and surprisingly it does not produce a noticeable pattern in the homogeneous Sc#1 in HGS and MSHE. On the other hand the topography effect in PCLM Sc#1 is overemphasized. Opposed to topography, distributed porosity (Sc#2) clearly affects the predicted SM patterns in all three models.

The correlation of the errors in discharge and SM between the models allows further investigations of the implications of diverging model structures. The enhanced topography feedback of the predicted SM by PCLM is found to be more significant at the hillslopes where larger topographical gradients are found. In general, the SM errors between the models show a strong correlation which allows drawing the conclusions that the spatial soil parametrization is inadequate, small scale processes like preferential flow are missed by the model structures or there is inconsistency between forcing- and observation-data. The distinct similarity between HGS and MSHE promotes that a 3D representation of flow processes in the unsaturated zone does not necessarily improve spatio-temporal SM predictions at the Wüstebach catchment.

The parameter transfer from HYDRUS 1D to the more complex multi-dimensional models proves to be limited because the HYDRUS 1D results clearly diverge from the distributed hydrological models. Also the obtained parameters appear to ignore the drying out of the soil. Apparently, model specific effective parameters are needed to better predict the spatio-temporal SM dynamics at the Wüstebach site.

In order to further improve the Wüstebach models, new ways to parametrize the soil in a realistic spatially distributed manner have to emerge in a spatial calibration framework. Spatial regularization methods (Pokhrel and Gupta, 2010; Samaniego et al., 2010) can be employed to reduce the number of free parameters while providing enough freedom to the patterns to be adjusted during the calibration. To fully exploit the unique SM dataset more advanced spatial metrics (Koch et al., 2015; Wealands et al., 2005) have to be incorporated in a multi-objective calibration framework which utilizes all available data of the catchment.

## Acknowledgments

The work has been carried out under the HOBE (Center for Hydrology) and SPACE (SPATial Calibration and Evaluation in distributed hydrological modeling) project, both founded by the Villum Foundation. Further the authors would like to acknowledge the TERENO (Terrestrial Environmental Observatories) project, funded by the Helmholtz Association of German Research Centers. The authors gratefully acknowledge the support by the SFB-TR32 project funded by the Deutsche Forschungsgemeinschaft (DFG), the MIKLIP project funded by the Federal Ministry of Education and Research of Germany (BMBF) and the support of Centre for High-Performance Scientific Computing in Terrestrial Systems, HPSC TerrSys.

## References

- Abbott, M.B., Bathurst, J.C., Cunge, J.A., Oconnell, P.E., Rasmussen, J., 1986. An introduction to the European hydrological system – système hydrologique Européen, SHE. 2. Structure of a physically-based, distributed modeling system. *J. Hydrol.* 87 (1–2), 61–77.
- Ashby, S.F., Falgout, R.D., 1996. A parallel multigrid preconditioned conjugate gradient algorithm for groundwater flow simulations. *Nucl. Sci. Eng.* 124 (1), 145–159.
- Blume, T., Zehe, E., Bronstert, A., 2009. Use of soil moisture dynamics and patterns at different spatio-temporal scales for the investigation of subsurface flow processes. *Hydrol. Earth Syst. Sci.* 13 (7), 1215–1233.
- Bogena, H., Huisman, J., Oberdörster, C., Vereecken, H., 2007. Evaluation of a low-cost soil water content sensor for wireless network applications. *J. Hydrol.* 344 (1), 32–42.
- Bogena, H., Herbst, M., Huisman, J., Rosenbaum, U., Weuthen, A., Vereecken, H., 2010. Potential of wireless sensor networks for measuring soil water content variability. *Vadose Zone J.* 9 (4), 1002–1013.
- Bogena, H., Huisman, J., Baatz, R., Hendricks Franssen, H.J., Vereecken, H., 2013. Accuracy of the cosmic-ray soil water content probe in humid forest ecosystems: the worst case scenario. *Water Resour. Res.* 49 (9), 5778–5791.
- Bogena, H., Bol, R., Borchard, N., Brüggemann, N., Diekkrüger, B., Drüe, C., Groh, J., Gottselig, N., Huisman, J., Lücke, A., 2015. A terrestrial observatory approach to the integrated investigation of the effects of deforestation on water, energy, and matter fluxes. *Sci. China Earth Sci.* 58 (1), 61–75.
- Brocca, L., Morbidelli, R., Melone, F., Moramarco, T., 2007. Soil moisture spatial variability in experimental areas of central Italy. *J. Hydrol.* 333 (2–4), 356–373.
- Brocca, L., Melone, F., Moramarco, T., Morbidelli, R., 2010. Spatial-temporal variability of soil moisture and its estimation across scales. *Water Resour. Res.* 46 (2).
- Carslaw, D., 2012. The Openair Manual—Open-Source Tools for Analysing Air Pollution Data. Manual for Version 0.7–0. King's College London.
- Castillo, A., Castelli, F., Entekhabi, D., 2015. Gravitational and capillary soil moisture dynamics for distributed hydrologic models. *Hydrol. Earth Syst. Sci.* 19 (4), 1857–1869.
- Chaney, N.W., Roundy, J.K., Herrera-Estrada, J.E., Wood, E.F., 2014. High-resolution modeling of the spatial heterogeneity of soil moisture: applications in network design. *Water Resour. Res.*
- Clark, M.P., Kavetski, D., Fenicia, F., 2011. Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resour. Res.* 47 (9), W09301.
- Cornelissen, T., Diekkrüger, B., Giertz, S., 2013. A comparison of hydrological models for assessing the impact of land use and climate change on discharge in a tropical catchment. *J. Hydrol.* 498, 221–236.
- Cornelissen, T., Diekkrüger, B., Bogena, H.R., 2014. Significance of scale and lower boundary condition in the 3D simulation of hydrological processes and soil moisture variability in a forested headwater catchment. *J. Hydrol.* 516, 140–153.
- Dai, Y.J. et al., 2003. The common land model. *Bull. Am. Meteorol. Soc.* 84 (8), 1013–1023.
- De Schepper, G., Therrien, R., Refsgaard, J.C., Hansen, A.L., 2015. Simulating coupled surface and subsurface water flow in a tile-drained agricultural catchment. *J. Hydrol.* 521, 374–388.
- Fang, Z., Bogena, H., Kollet, S., Koch, J., Vereecken, H., 2015. Spatio-temporal validation of long-term 3D hydrological simulations of a forested catchment using empirical orthogonal functions and wavelet coherence analysis. *J. Hydrol.* 529, 1754–1767.
- Graf, A., Bogena, H.R., Drüe, C., Hardelauf, H., Pütz, T., Heinemann, G., Vereecken, H., 2014. Spatiotemporal relations between water budget components and soil water content in a forested tributary catchment. *Water Resour. Res.* 50 (6), 4837–4857.
- Graham, D.N., Butts, M.B., 2005. Flexible, integrated watershed modelling with MIKE SHE. *Watershed Models* 849336090, 245–272.
- Grayson, R.B., Moore, I.D., McMahon, T.A., 1992. Physically based hydrologic modeling: 1. A terrain-based model for investigative purposes. *Water Resour. Res.* 28 (10), 2639–2658.
- Gupta, H.V., Clark, M.P., Vrugt, J.A., Abramowitz, G., Ye, M., 2012. Towards a comprehensive assessment of model structural adequacy. *Water Resour. Res.* 48 (8).
- He, X., Höjberg, A.L., Jørgensen, F., Refsgaard, J.C., 2015. Assessing hydrological model predictive uncertainty using stochastically generated geological models. *Hydrol. Process.*
- Holländer, H., Blume, T., Bormann, H., Buytaert, W., Chirico, G., Extrayat, J.-F., Gustafsson, D., Hözel, H., Kraft, P., Stamm, C., 2009. Comparative predictions of discharge from an artificial catchment (Chicken Creek) using sparse data. *Hydrol. Earth Syst. Sci.* 13 (11), 2069–2094.
- Holländer, H., Bormann, H., Blume, T., Buytaert, W., Chirico, G., Extrayat, J., Gustafsson, D., Hözel, H., Krause, T., Kraft, P., 2014. Impact of modellers' decisions on hydrological a priori predictions. *Hydrol. Earth Syst. Sci.* 18 (6), 2065–2085.
- Hözel, H., Rössler, O., Diekkrüger, B., 2011. Grope in the Dark-Hydrological modelling of the artificial Chicken Creek catchment without validation possibilities. *Phys. Chem. Earth, Parts A/B/C* 36 (1), 113–122.
- Kampf, S.K., Burges, S.J., 2007. A framework for classifying and comparing distributed hillslope and catchment hydrologic models. *Water Resour. Res.* 43 (5).
- Kirchner, J.W., 2006. Getting the right answers for the right reasons: linking measurements, analyses, and models to advance the science of hydrology. *Water Resour. Res.* 42 (3), W03S04.
- Koch, J., Jensen, K.H., Stisen, S., 2015. Toward a true spatial model evaluation in distributed hydrological modeling: Kappa statistics, fuzzy theory, and EOF-analysis benchmarked by the human perception and evaluated against a modeling case study. *Water Resour. Res.* 51 (2), 1225–1246.
- Kollet, S.J., Maxwell, R.M., 2008. Capturing the influence of groundwater dynamics on land surface processes using an integrated, distributed watershed model. *Water Resour. Res.* 44 (2).
- Korres, W., Koyama, C.N., Fiener, P., Schneider, K., 2010. Analysis of surface soil moisture patterns in agricultural landscapes using Empirical Orthogonal Functions. *Hydrol. Earth Syst. Sci.* 14 (5), 751–764.
- Korres, W., Reichenau, T., Fiener, P., Koyama, C., Bogena, H., Cornelissen, T., Baatz, R., Herbst, M., Diekkrüger, B., Vereecken, H., 2015. Spatio-temporal soil moisture patterns – a meta-analysis using plot to catchment scale data. *J. Hydrol.* 520, 326–341.
- Kristensen, K.J., Jensen, S.E., 1975. A model for estimating actual evapotranspiration from potential evapotranspiration. *Nord. Hydrol.* 6 (3), 170–188.
- Martinez, G., Pachepsky, Y.A., Vereecken, H., Hardelauf, H., Herbst, M., Vanderlinden, K., 2013. Modeling local control effects on the temporal stability of soil water content. *J. Hydrol.* 481, 106–118.
- Maxwell, R.M., 2013. A terrain-following grid transform and preconditioner for parallel, large-scale, integrated hydrologic modeling. *Adv. Water Resour.* 53, 109–117.
- Maxwell, R.M., Putti, M., Meyerhoff, S., Delfs, J.O., Ferguson, I.M., Ivanov, V., Kim, J., Kolditz, O., Kollet, S.J., Kumar, M., 2014. Surface-subsurface model intercomparison: a first set of benchmark results to diagnose integrated hydrology and feedbacks. *Water Resour. Res.* 50 (2), 1531–1549.
- Noh, S.J., An, H., Kim, S., Kim, H., 2015. Simulation of soil moisture on a hillslope using multiple hydrologic models in comparison to field measurements. *J. Hydrol.* 523, 342–355.
- Orth, R., Staudinger, M., Seneviratne, S.I., Seibert, J., Zappa, M., 2015. Does model performance improve with complexity? A case study with three hydrological models. *J. Hydrol.* 523, 147–159.
- Panday, S., Huyakorn, P.S., 2004. A fully coupled physically-based spatially-distributed model for evaluating surface/subsurface flow. *Adv. Water Resour.* 27 (4), 361–382.
- Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. *Comput. Geosci.* 30 (7), 683–691.
- Perry, M.A., Niemann, J.D., 2007. Analysis and estimation of soil moisture at the catchment scale using EOFs. *J. Hydrol.* 334 (3–4), 388–404.
- Pokhrel, P., Gupta, H.V., 2010. On the use of spatial regularization strategies to improve calibration of distributed watershed models. *Water Resour. Res.* 46.
- Pokhrel, P., Gupta, H.V., 2011. On the ability to infer spatial catchment variability using streamflow hydrographs. *Water Resour. Res.* 47.
- Qu, W., Bogena, H., Huisman, J., Martinez, G., Pachepsky, Y., Vereecken, H., 2014. Effects of soil hydraulic properties on the spatial variability of soil water content: evidence from sensor network data and inverse modeling. *Vadose Zone J.* 13 (12).
- Qu, W., Bogena, H., Huisman, J., Vanderborght, J., Schuh, M., Priesack, E., Vereecken, H., 2015. Predicting sub-grid variability of soil water content from basic soil information. *Geophys. Res. Lett.*
- Rahman, M., Sulis, M., Kollet, S., 2014. The concept of dual-boundary forcing in land surface–subsurface interactions of the terrestrial hydrologic and energy cycles. *Water Resour. Res.* 50 (11), 8531–8548.
- Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J., Participants, D., 2004. Overall distributed model intercomparison project results. *J. Hydrol.* 298 (1), 27–60.
- Ridder, M.E., Madsen, H., Stisen, S., Bircher, S., Fensholt, R., 2014. Assimilation of SMOS-derived soil moisture in a fully integrated hydrological and soil-vegetation-atmosphere transfer model in Western Denmark. *Water Resour. Res.* 50 (11), 8962–8981.
- Romanow, N., 2014. Soil moisture at local scale: measurements and simulations. *J. Hydrol.* 516, 6–20.
- Rosenbaum, U., Bogena, H.R., Herbst, M., Huisman, J.A., Peterson, T.J., Weuthen, A., Western, A.W., Vereecken, H., 2012. Seasonal and event dynamics of spatial soil moisture patterns at the small catchment scale. *Water Resour. Res.* 48.
- Samaniego, L., Kumar, R., Attinger, S., 2010. Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resour. Res.* 46.
- Sciuto, G., Diekkrüger, B., 2010. Influence of soil heterogeneity and spatial discretization on catchment water balance modeling. *Vadose Zone J.* 9 (4), 955–969.
- Semenova, O., Beven, K., 2015. Barriers to progress in distributed hydrological modelling. *Hydrol. Process.*
- Sheffield, J., Wood, E.F., 2008. Global trends and variability in soil moisture and drought characteristics, 1950–2000, from observation-driven simulations of the terrestrial hydrologic cycle. *J. Clim.* 21 (3), 432–458.
- Smith, M., Gupta, H., 2012. The distributed model intercomparison project (DMIP) – phase 2 experiments in the Oklahoma Region, USA. *J. Hydrol.* 418–419, 1–2.
- Smith, M., Koren, V., Reed, S., Zhang, Z., Zhang, Y., Moreda, F., Cui, Z., Mizukami, N., Anderson, E.A., Cosgrove, B.A., 2012a. The distributed model intercomparison project-phase 2: motivation and design of the Oklahoma experiments. *J. Hydrol.* 418, 3–16.

- Smith, M., Koren, V., Zhang, Z., Zhang, Y., Reed, S.M., Cui, Z., Moreda, F., Cosgrove, B.A., Mizukami, N., Anderson, E.A., 2012b. Results of the DMIP 2 Oklahoma experiments. *J. Hydrol.* 418–419, 17–48.
- Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crucifix, T., Rast, S., Salzmann, M., Schmidt, H., Bader, J., Block, K., 2013. Atmospheric component of the MPI-M earth system model: ECHAM6. *J. Adv. Model. Earth Syst.* 5 (2), 146–172.
- Stisen, S., McCabe, M.F., Refsgaard, J.C., Lerer, S., Butts, M.B., 2011. Model parameter analysis using remotely sensed pattern information in a multi-constraint framework. *J. Hydrol.* 409 (1–2), 337–349.
- Stockinger, M.P., Bogen, H.R., Lücke, A., Diekkrüger, B., Weiler, M., Vereecken, H., 2014. Seasonal soil moisture patterns: controlling transit time distributions in a forested headwater catchment. *Water Resour. Res.* 50 (6), 5270–5289.
- Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.: Atmos.* (1984–2012) 106 (D7), 7183–7192.
- Therrien, R., McLaren, R., Sudicky, E., Panday, S., 2010. HydroGeoSphere: A Three-Dimensional Numerical Model Describing Fully-Integrated Subsurface and Surface Flow and Solute Transport. Groundwater Simulations Group, University of Waterloo, Waterloo, ON.
- Vereecken, H., Huisman, J., Pachepsky, Y., Montzka, C., Van Der Kruk, J., Bogen, H., Weihermüller, L., Herbst, M., Martinez, G., Vanderborght, J., 2014. On the spatio-temporal dynamics of soil moisture at the field scale. *J. Hydrol.* 516, 76–96.
- Voldoire, A., Sanchez-Gomez, E., y Méliá, D.S., Decharme, B., Cassou, C., Sénesi, S., Valcke, S., Beau, I., Alias, A., Chevallier, M., 2013. The CNRM-CM5.1 global climate model: description and basic evaluation. *Clim. Dyn.* 40 (9–10), 2091–2121.
- Wang, T., Franz, T.E., Zlotnik, V.A., You, J., Shulski, M.D., 2015. Investigating soil controls on soil moisture spatial variability: numerical simulations and field observations. *J. Hydrol.* 524, 576–586.
- Wealands, S.R., Grayson, R.B., Walker, J.P., 2005. Quantitative comparison of spatial fields for hydrological model assessment – some promising approaches. *Adv. Water Resour.* 28 (1), 15–32.
- Wood, E.F., Roundy, J.K., Troy, T.J., Van Beek, L., Bierkens, M.F., Blyth, E., de Roo, A., Döll, P., Ek, M., Famiglietti, J., 2011. Hyperresolution global land surface modeling: meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resour. Res.* 47 (5).
- Zacharias, S., Bogen, H., Samaniego, L., Mauder, M., Fuß, R., Pütz, T., Frenzel, M., Schwank, M., Baessler, C., Butterbach-Bahl, K., 2011. A network of terrestrial environmental observatories in Germany. *Vadose Zone J.* 10 (3), 955–973.