A fully coupled land surface hydrologic model, Flux-PIHM, is developed by incorporating a land surface scheme into the Penn State Integrated Hydrologic Model (PIHM). The land surface scheme is adapted from the Noah land surface model. Because PIHM is capable of simulating lateral water flow and deep groundwater at spatial resolutions sufficient to resolve upland stream networks, Flux-PIHM is able to represent heterogeneities due to topography and soils at high resolution, including spatial structure in the link between groundwater and the surface energy balance (SEB). Flux-PIHM has been implemented at the Shale Hills watershed (0.08 km²) in central Pennsylvania. Multistate observations of discharge, water table depth, soil moisture, soil temperature, and sensible and latent heat fluxes in June and July 2009 are used to manually calibrate Flux-PIHM at hourly temporal resolution. Model predictions from 1 March to 1 December 2009 are evaluated. Both hydrologic predictions and SEB predictions show good agreement with observations. Comparisons of model predictions between Flux-PIHM and the original PIHM show that the inclusion of the complex SEB simulation only brings slight improvement in hourly model discharge predictions. Flux-PIHM adds the ability of simulating SEB to PIHM and does improve the prediction of hourly evapotranspiration, the prediction of total runoff (discharge), and the predictions of some peak discharge events, especially after extended dry periods. Model results reveal that annual average sensible and latent heat fluxes are strongly correlated with water table depth, and the correlation is especially strong for the model grids near the stream.

1. Introduction

The predictability of the atmosphere is limited by the chaotic nature of atmospheric turbulence to a time span of the order of one week or less (Lorenz 1969; Smagorinsky 1969; Lorenz 1982). Earth’s surface, however, has “memories” much longer than those of the atmosphere. Significant improvements in short-term climate forecasts as well as weather forecasts can be found by including the modeling of Earth’s surface, for example, land surface processes and sea surface temperature (SST), in predictive models (Palmer and Anderson 1994; Beljaars et al. 1996; Koster et al. 2000; Goddard et al. 2001; Ek et al. 2003; Mitchell et al. 2004; Kumar et al. 2008). While covering only 30% of Earth’s surface, the land surface plays a distinctive role in weather and climate because of its considerable heterogeneity; its dynamic hydrologic cycle and strong variations of temperature; and its highly variable land use, land cover, and topography (Yang 2004). Land surface processes are critical in the growth of the atmospheric boundary layer (ABL), the formation of clouds and precipitation, and the budgets of heat, momentum, and moisture within the atmosphere. Weather and climate models currently rely on land surface models (LSMs) to represent land surface processes. LSMs provide lower boundary conditions and initialize ground states for numerical weather prediction (Ookouchi et al. 1984; Liang et al. 1994; Betts et al. 1997; F. Chen et al. 1997; Xi and Pleim 2001; Ek et al. 2003; Kumar et al. 2006; Chen et al. 2007; Niu et al. 2011).

In the past few decades, LSMs have undergone significant development from the “bucket model” (Manabe 1969) to more sophisticated and more physical
parameterizations. The evolution of LSMs reflects the community’s effort to improve the representation of land surface fluxes, including the benefit of capturing the memory embedded in soil moisture (Ek et al. 2003; Mitchell et al. 2004). Groundwater has a longer memory than soil moisture and has been shown to influence the land surface and the atmosphere (Changnon 1987; Dooge 1992; Skoëien et al. 2003; Liang et al. 2003; Maxwell and Miller 2005; Yeh and Eltahir 2005; Kollet and Maxwell 2008). Subsurface waters, however, are not well described in most LSMs. Traditional LSMs are limited to vertical moisture transport in the soil column, and most of them ignore deeper soil moisture processes and lack physical representations of water table or the explicit lateral flow to streams. Therefore, those models have limited ability in representing the contribution of groundwater to the memory of land surface. This groundwater memory effect is particularly challenging to simulate in lower-order watersheds (e.g., first- and second-order streams and upland watersheds), which encompass a large fraction of the land surface (Shreve 1969; Benda et al. 2005), are highly heterogeneous, and thus require hydrologic modeling at spatial resolutions that resolve the complexity of these small-scale systems.

The water cycle, or hydrologic cycle, is central to the Earth system, and water resources are one of the critical environmental and political issues of the twenty-first century (National Research Council 2004). Hydrologic models are important tools to enhance the understanding of hydrological processes and to simulate and predict hydrological events for better decision making. They have evolved from early empirical rainfall–runoff models (e.g., Mulvany 1851; Sherman 1932; Horton 1935) to conceptual models (e.g., Linsley and Crawford 1960; Crawford and Linsley 1966; Dawdy and O’Donnell 1965; Sugawara 1969; Burnash et al. 1973; Burnash 1995; Arnold et al. 1998; Ma and Cheng 1998), to physically based models (e.g., Abbott et al. 1986a, b; Beven and Kirkby 1976, 1979; Beven et al. 1987; Grayson et al. 1992; Fortin et al. 2001a, b; Qu 2004; Qu and Duffy 2007; Kumar 2009). Physically based, spatially distributed hydrologic models have the advantages of taking into account spatial heterogeneity of inputs and characterizing hydrologic variables in space (Beven 1985; Smith et al. 2004; Wood and Lettenmaier 2006; Li et al. 2009). They can also be applied to the watersheds where no adequate data are available for calibrating empirical and conceptual models. Physically based hydrologic models are therefore extremely important for flood/drought forecasting at small-scale, low-order watersheds and for the study and understanding of hydrologic processes.

Evapotranspiration is an important component of the water cycle and an important process in hydrologic models. The accurate forecasting of the timing and magnitude of peak discharge events depends on the forecasting of evapotranspiration, especially after extended dry periods (Kampf 2006). Hydrologic models, however, usually have simplified representations of evapotranspiration. They either take potential evapotranspiration as an external atmospheric forcing (e.g., Beven et al. 1987; Grayson et al. 1992; Burnash 1995; Yeh et al. 2006) or use simplified formulations and parameterizations for calculating evapotranspiration (e.g., Abbott et al. 1986b; Kumar 2009). Studies show that applying different methods for calculating potential evapotranspiration in hydrologic models produces significant differences in evapotranspiration and, consequently, in runoff simulations (Vörösmarty et al. 1998; Kampf 2006; Weiß and Menzel 2008).

Coupled models of land surface and subsurface, which incorporate hydrologic components into LSMs or couple deeper subsurface with the atmosphere, may yield improvements in weather and short-term climate forecasting and flood/drought forecasting. There has been recent interest in incorporating a groundwater component into LSMs or coupling deeper subsurface with the atmosphere to improve the representation of soil moisture at the land surface (e.g., York et al. 2002; Seuffert et al. 2002; Mölders and Rühaak 2002; Liang et al. 2003; Yeh and Eltahir 2005; Maxwell and Miller 2005; Gulden et al. 2007; Maxwell et al. 2007; Kollet and Maxwell 2008; Rosero et al. 2011; Niu et al. 2011). Effects of groundwater and soil moisture on the ABL (Maxwell et al. 2007), as well as improvement in energy fluxes and rainfall predictions (Seuffert et al. 2002; Mölders and Rühaak 2002) were found. Rigon et al. (2006) developed a distributed hydrological model with coupled water and energy budgets (GEOTOP) and tested the model performance at a midsized watershed. Bouillon et al. (2010) coupled a soil–vegetation–atmosphere transfer model [Soil, Biosphere, and Atmosphere (ISBA)] with a conceptual hydrologic model (TOPMODEL) and evaluated the model predictions of discharge at three midsized watersheds. Livneh et al. (2011) developed a unified land model by combining a 1D LSM (Noah LSM) with a spatially lumped conceptual hydrologic model [Sacramento Soil Moisture Accounting model (SAC-SMA)] and tested the coupled model at six midsized watersheds. In the Noah LSM with multiparameterization option (Noah-MP), Niu et al. (2011) implemented a simple groundwater model and a TOPMODEL-based runoff model into the original Noah LSM and evaluated Noah-MP performance at various small watersheds. These coupled modeling studies did not evaluate the subsurface hydrology and land surface fluxes simultaneously. Livneh et al. (2011) and
Niu et al. (2011) evaluated subsurface hydrology in some watersheds and land surface fluxes in different watersheds. Rigon et al. (2006) used different simulations to evaluate each model component. A comprehensive evaluation of coupled model performance is needed.

Maxwell and Miller (2005), Maxwell et al. (2007), Kollet and Maxwell (2008), and Rihani et al. (2010) have incorporated a fully coupled, three-dimensional (subsurface and overland) groundwater flow model (ParFlow), with a land surface model [Community Land Model (CLM)] on one hand and a mesoscale atmospheric model [Advanced Regional Prediction System (ARPS)] on the other hand, to study the influences of lateral subsurface water flow and shallow water table depth on the surface energy balance. Results indicate that the coupled land surface groundwater model behaves differently from uncoupled land surface model in synthetic data experiments and reveals strong correlation between land surface and subsurface. Because of the complexity of optimization, the models in those studies were not well calibrated. The improvement in forecasts over uncoupled models, if any, and the interaction between land surface and subsurface, however, still needs further exploration.

Hydrologic model parameters generally need to be calibrated for the model system to reproduce observed hydrologic responses even when a priori parameters can be estimated from existing data (e.g., national soil survey and hydrogeologic databases). Calibration of hydrologic models has been the focus of many studies (e.g., Ibbitt 1970; Johnston and Pilgrim 1976; Pickup 1977; Gupta and Sorooshian 1985; Duan et al. 1992; Sorooshian et al. 1993; Franchini 1996; Vrugt et al. 2003; Wagener et al. 2003; Kollat and Reed 2006; Xie and Zhang 2010).

Most of the previous studies focus on the calibration of model discharge, and sometimes water table depth, but tend to neglect other observations. However, “an acceptable model prediction might be achieved in many different ways, i.e., different model structures or parameter sets” (Beven 1993). This nonuniqueness of numerical model parameters and structures is called “equifinality.” Equifinality makes model calibration difficult using only one type of observation. If multiple parameter sets produce equally good discharge predictions, it is very difficult to find the optimal parameter set. One possible solution to equifinality is to use multistate observations to constrain the parameters. Some studies also found that using observations of subsurface conditions (e.g., soil moisture) in addition to discharge observations improves the forecast of streamflow (e.g., Oudin et al. 2003; Aubert et al. 2003; Francois et al. 2003; Camy-Peyret et al. 2009; Lee et al. 2011).

Calibrating model parameters using high temporal resolution observations may improve model representation of important hydrologic processes in low-order watersheds. Most previous studies calibrated and evaluated hydrologic models at a daily time scale, ignoring the subdaily changes of hydrologic variables. Because of the impacts of losing streams, infiltration, precipitation, evapotranspiration, and freeze–thaw processes, discharge and groundwater level have diurnal cycles (Lundquist and Cayan 2002; Griebovszki et al. 2010). The diurnal fluctuation of groundwater table is most significant in summer and in highly forested areas and reaches up to 11 cm (Thal-Larsen 1934) in some areas. In low-order watersheds, streamflow and groundwater level exhibit even more temporal variability than in larger basins (Reed et al. 2004) and are influenced by the complication of rapidly varying climatic gradients and topographic effects, as well as complex hydrologic and geomorphic conditions that control basin storage and runoff. Calibration using high temporal resolution observations could help capture the rapidly changing processes and improve the forecasting in low-order watersheds.

In this study, a coupled land surface hydrologic modeling system is developed and evaluated at a small-scale forested watershed. In particular, the Penn State Integrated Hydrologic Model (PIHM; Qu 2004; Qu and Duffy 2007; Kumar 2009) is coupled with the land surface schemes in the Noah LSM (Chen and Dudhia 2001; Ek et al. 2003). PIHM is a fully coupled, spatially distributed, and physically based hydrologic model. This model has advanced model physics, for example, fully coupled surface and subsurface flow, lateral surface and subsurface water flow, and macropore flow, and is capable of resolving small-scale hydrologic modeling at low-order watersheds. The model is implemented at a first-order watershed, the Shale Hills watershed (0.08 km²) in central Pennsylvania. A National Science Foundation (NSF)-sponsored Critical Zone Observatory (CZO), the Susquehanna Shale Hills Critical Zone Observatory (SSHCZO), now exists in this watershed. CZOs are operated at watershed scale to advance the understanding of Earth’s surface processes. SSHCZO brings together multiple research disciplines to observe and quantify Earth’s surface processes at hillslope to small-watershed scales. The broad array of observations at SSHCZO enables an unprecedented investigation of subsurface–land surface–atmosphere interactions and makes SSHCZO an ideal site for the coupled model test. Multistate observations, including discharge, water table depth, soil moisture, soil temperature, and surface heat fluxes, are used for the comprehensive model calibration and evaluation. The performance of the original PIHM version 2.0 (PIHM V2) is compared with the coupled land surface hydrologic model. Model predictions of hydrologic variables and land surface variables are evaluated at hourly resolution.
2. Description of the coupled land surface hydrologic model

a. The Penn State Integrated Hydrologic Model

PIHM is a multiprocess and multiscale hydrologic model. It simulates evapotranspiration, infiltration, recharge, overland flow, groundwater flow, and channel routing in a fully coupled scheme. The model has been tested at both small-sized (e.g., Qu 2004; Li 2010) and midsized watersheds (e.g., Kumar 2009). Here we summarize the major methodological elements of PIHM. Triangular and rectangular elements are used in PIHM. The land surface is decomposed into unstructured triangular elements while rivers are represented by rectangular elements. The use of an unstructured mesh provides an optimal representation of local heterogeneities in parameters and process dynamics with the least number of elements. It also provides better representation of linear features within the model domain such as river channels and watershed boundaries. Triangular and rectangular elements at the land surface are projected vertically down to bedrock to form prismatic volumes. Surface water is able to flow along topography, infiltrate into soil, or evaporate. Channel flow and overland flow are governed by the 1D (channel flow) and 2D (overland flow) St. Venant equations (Saint-Venant 1871). Infiltration is calculated through the top 10-cm-layer of soil. The subsurface prismatic volume is further subdivided into unsaturated and saturated zones. Soil water is restricted to vertical transport in the unsaturated zone. In the saturated zone, groundwater flow is horizontal with dynamic coupling to the unsaturated zone across the free surface (water table). Unsaturated flow and groundwater flow are described using the Richards equation, and the unsaturated hydraulic conductivities are calculated using the van Genuchten (1980) equations. Macropore flows are simulated in PIHM to account for the rapid water flows through root holes and soil cracks, that is, macropores. Macropores penetrate the soil from land surface to a depth into the soil, defined as the macropore depth. The effective hydraulic conductivity of the subsurface is considered as a weighted average conductivity of the macropores and the soil matrix. The semidiscrete finite volume method (Qu and Duffy 2007) is applied, and the system of governing differential equations is solved simultaneously. PIHM also includes a simple temperature index representation of snowmelt (Kumar 2009). Detailed descriptions and formulations of PIHM are provided by Qu (2004), Qu and Duffy (2007), Kumar (2009), and Li (2010).

b. The land surface scheme

The new land surface hydrologic model Flux-PIHM incorporates a land surface scheme into PIHM. The land surface scheme is mainly adapted from the Noah LSM (Chen and Dudhia 2001; Ek et al. 2003), which has undergone extensive testing (e.g., Chen et al. 1996; T. H. Chen 1997; Liang et al. 1998; Chen and Mitchell 1999; Koren et al. 1999; Schlosser et al. 2000; Ek et al. 2003; Boone et al. 2004) and has been implemented into mesoscale atmospheric models, for example, the fifth-generation Pennsylvania State University–National Center for Atmospheric Research (NCAR) Mesoscale Model (MM5) and the Weather Research and Forecast Model (WRF). The Noah LSM integrates the Penman potential evaporation scheme of Mahrt and Ek (1984), the multiple-layer soil model of Mahrt and Pan (1984), the canopy model of Pan and Mahrt (1987), and the canopy resistance approach of Noilhan and Planton (1989) and Jacquemin and Noilhan (1990) (Chen et al. 2007). The Noah LSM has one canopy layer and four soil layers. It is able to capture diel and seasonal variations in surface heat fluxes and temperature with relatively few parameters. Its relatively simple formulation makes it easy to couple to PIHM.

The Noah LSM is adapted and coupled with PIHM. The soil water retention model, the canopy water capacity formulation, and the canopy drip rate formulation in the Noah LSM are replaced by the PIHM formulations in the coupled land surface scheme. To keep in accordance with PIHM, the Cosby et al. (1984) soil water retention model used in the Noah LSM is replaced by the van Genuchten (1980) soil model in PIHM; in the van Genuchten (1980) equation,

\[
\theta = \left( \frac{1}{1 + (\alpha h)^\beta} \right)^{1-1/\beta},
\]

where \(\theta\) is the saturation ratio of soil, \(\alpha\) and \(\beta\) are the van Genuchten soil parameters, and \(h\) is the hydraulic head. The maximum canopy water capacity \(W_{c,max}\) is a constant in the Noah LSM, is formulated as a linear function of the leaf area index (LAI):

\[
W_{c,max} = S LAI,
\]

where \(S\) is the reference canopy water capacity, and its default value is set to 0.2 mm. The canopy drip rate is formulated as

\[
Dr = \begin{cases} 
  k_D \exp\left(b \frac{W_c}{W_{c,max}}\right), & 0 < W_c \leq W_{c,max} \\
  k_D \exp(b) + \frac{W_c - W_{c,max}}{\Delta t}, & W_c > W_{c,max}
\end{cases}
\]
where the reference drip rate $k_D = 5.65 \times 10^{-2} \text{ m day}^{-1}$, $W_c$ is the water storage in the canopy, $\Delta t$ is the time step, and the parameter $b$ has a range from 3.0 to 4.6 as suggested by Rutter and Morton (1977) and is set to be 3.89 here.

The prismatic volume of each grid is divided into several soil layers depending on the depth of the prism. From the ground surface to the bottom, the standard thicknesses of the top four layers are 0.1, 0.3, 0.6, and 1.0 m, respectively, as in the Noah LSM. If the bedrock depth is less than 2 m, the number of soil layers and the thickness of the lowest layer are adapted to match the depth of bedrock. If the bedrock depth is larger than 2 m, additional soil layers are added as needed. Soil moisture contents and soil temperatures of those multiple layers are simulated. The volumetric soil moisture ($\Theta$) equations in the Noah LSM are adapted to allow coupling with PIHM. For the top soil layer,

$$d_z \frac{\partial \Theta}{\partial t} = -D_1 \left( \frac{\partial \Theta}{\partial z} \right)_{z_1} - K_{z_1} + I - E_{\text{soil}} - E_{11}; \quad (4a)$$

for the bottom soil layer $J$ above the water table,

$$d_z \frac{\partial \Theta}{\partial t} = D_J \left( \frac{\partial \Theta}{\partial z} \right)_{z_{j-1}} + K_{z_{j-1}} - R - E_{ij}; \quad (4b)$$

and for the layer $j$ in between,

$$d_z \frac{\partial \Theta}{\partial t} = D_{j-1} \left( \frac{\partial \Theta}{\partial z} \right)_{z_{j-1}} - D_j \left( \frac{\partial \Theta}{\partial z} \right)_{z_j} + K_{z_{j-1}} - K_{z_j} - E_{ij}, \quad (4c)$$

where $z$ represents the vertical direction, $d_z$ is the thickness of the $j$th soil layer, $K$ is the hydraulic conductivity, $D = K(\partial h/\partial \Theta)$ is the soil water diffusivity, $E_{\text{soil}}$ is the evaporation from soil, and $E_{ij}$ represents the canopy transpiration taken by root in the $j$th layer. The variable $I$ is the surface water infiltration, and $R$ is the groundwater recharge. When coupled with PIHM, both $I$ and $R$ are provided by the hydrologic model. The Noah LSM uses a parameterized infiltration scheme and a gravitational free-drainage subsurface runoff scheme (Schaake et al. 1996). In comparison, infiltration and groundwater recharge in this model [Eqs. (4a) and (4b)] are calculated using a physical hydrologic model, which should provide more realistic boundary conditions for the LSM. Further, the Noah LSM evapotranspiration formulation, including extraction of transpiration from the rooting zone ($E_r$) should provide more realistic upper-boundary conditions for groundwater hydrology. Equation (4) therefore provides the core innovation in this coupled model, Flux-PIHM.

**c. Fully coupled land surface hydrologic modeling system**

Flux-PIHM is thus a fully coupled land surface hydrologic modeling system. The surface energy balance scheme described in section 2b completely replaces the original evapotranspiration formulation in PIHM. The land surface and hydrologic components are coupled by exchanging water table depth, soil moisture, infiltration rate, recharge rate, net precipitation (i.e., precipitation that is not intercepted by canopy) rate, and evapotranspiration rate between each other, as described in Fig. 1. The 1D land surface scheme is integrated into every model grid, and the passing of information is performed at every time step. The hydrologic component and the land surface component share the same model parameters at subsurface (e.g., hydraulic conductivities). At each time step, the hydrologic component provides the land surface component with water table depth, infiltration rate, recharge rate, and integrated soil water storage over soil column. The soil layers below the water table are set to saturation. Soil moisture equations [Eq. (4)] are applied to the soil layers above the water table. Infiltration rate and recharge rate calculated by the hydrologic component are used as top and bottom boundary conditions for soil moisture transport. PIHM only predicts the total soil water storage within each soil column. The volumetric soil moisture contents simulated by the land surface component are rescaled using...
the integrated soil water storage provided by PIHM to guarantee mass conservation:

$$\theta_j' = \frac{SWC}{\sum_j \theta_j d_j} \sum_j \theta_j,$$

(5a)

and the average volumetric soil moisture content of the unsaturated zone SWC is calculated as

$$SWC = \frac{h_{\text{unsat}}}{BRD - h_{\text{sat}}} (\theta_{s} - \theta_{r}) + \theta_{r},$$

(5b)

where $\theta_j$ is the volumetric soil moisture at the $j$th soil layer predicted by the land surface module, $\theta_j'$ is the volumetric soil moisture after rescaling, $h_{\text{unsat}}$ is the integrated soil water storage predicted by PIHM, BRD is the bedrock depth, $h_{\text{sat}}$ is the groundwater level predicted by PIHM, $\theta_s$ is the soil porosity, and $\theta_r$ is the residual porosity. The land surface component then starts the surface energy balance simulation and provides the hydrologic component with the net precipitation rate and evapotranspiration rate.

Because PIHM is capable of simulating lateral groundwater flow that explicitly resolves low-order channel networks, Flux-PIHM is able to represent the land surface heterogeneities caused by topography and capture the time scales that are generally missing in current land surface models. At the same time, the robust land surface scheme provides accurate sensible heat flux and latent heat flux (evapotranspiration) simulations at all scales in the watershed or river network. The coupled hydrologic and land surface schemes guarantee mass conservation at the subsurface, as well as at the land surface, and conserve energy balance at land surface, which provides physical constraints to surface heat fluxes and subsurface water movement.

3. Study site and data

a. The Shale Hills critical zone observatory

The Shale Hills watershed is a 0.08 km² watershed located in the valley and ridge physiographic province of central Pennsylvania, at 40°39′.87″N, 77°54′.40″W. The watershed is a small, forested, and temperate-climate catchment carved in shale bedrock. The catchment is V-shaped and is characterized by relatively steep slopes (25%–48%) and narrow ridges. Surface elevation varies from 256 m MSL at the outlet to 310 m MSL at ridge top. The first-order stream within the catchment is a tributary of Shaver’s Creek that eventually reaches the Juniata River. The sloping areas and ridges of the watershed are covered by several typical deciduous species.

The valley floor and north-facing ridge top are covered by some evergreen species (Wubbels 2010). Five soil series are identified within the watershed (Lin 2006; Lin and Zhou 2008). SSHCZO has been the focus of hydrologic research for several decades (e.g., Lynch 1976; Qu and Duffy 2007; Lin and Zhou 2008; Ma et al. 2010). Extensive multidisciplinary field surveys have been conducted at SSHCZO, including soil survey, tree survey, surface elevation measurements, and bedrock depth measurements. A real-time hydrologic monitoring network (RTHnet) is operating in SSHCZO. RTHnet provides real-time and high-frequency observations from bedrock to ABL. RTHnet instrument arrays include sensors for soil moisture, soil temperature, soil matric potential, groundwater level and temperature, and snow depth measurement. The arrays also include sap flux measurements, a single above-canopy eddy-covariance flux tower, and a weather station. All available data can be found at http://www.czo.psu.edu/.

b. Model setup and model parameters

The model simulation period is from 0000 UTC 1 January 2009 to 0000 UTC 1 January 2010 with a model time step of 1 min and an output interval of 1 h. The model domain is prepared using the PIHM Geographic Information System (PIHM GIS). The model domain is decomposed into a triangular irregular network of 535 grids with 299 nodes. The average grid size is 157 m². The stream channel is represented by 20 stream segments. The model grids are presented in Fig. 2. Depending on the bedrock depth, the model grids have either four (if the bedrock depth is equal to or smaller than 2.0 m) or five (if the bedrock depth is larger than 2.0 m) vertical soil layers. A no-flow boundary condition is applied to the watershed boundary (except for the stream outlet) and the aquifer lower boundary. A
zero-depth gradient boundary condition (Morris 1979) is applied at the last stream segment, which is the outlet of the watershed.

Figure 3 shows the vegetation type, soil type, surface elevation, and bedrock depth defined within the model domain. The surface terrain of the Shale Hills watershed is represented by a 1-m resolution digital elevation model (DEM) digitized from airborne light detection and ranging (lidar) data (http://pihm.ics.psu.edu/CZO_NOSL/). The high-resolution DEM is used to develop the numerical mesh necessary to resolve upland surface and subsurface processes.

A bedrock depth map was obtained from a field campaign conducted in 2003 (Lin et al. 2006). Bedrock depths range from less than 0.25 m on ridge tops to greater than 3 m along the channel. To account for flow through a deeper weathered shale, an extra 1.5 m is added to measured bedrock depths for all model mesh points. The effective bedrock depths are shown in Fig. 3. The soil map and some soil matrix properties, including vertical saturated hydraulic conductivity, horizontal saturated hydraulic conductivity, and porosity, are also obtained from the field campaign in 2003 (Lin et al. 2006; Lin 2006). Other soil matrix properties used in this study are acquired from the Soil Survey Geographic (SSURGO) national database (http://www.HydroTerre.psu.edu). For soil wilting point ($\Theta_w$) and field capacity ($\Theta_{ref}$), the methods in the Noah LSM are adapted. They are calculated as

$$\Theta_w = 0.25(\Theta_s - \Theta_r) \left( \frac{1}{1 + (200\alpha)^{1/\beta}} \right)^{1-1/\beta} + \Theta_r, \quad (6)$$

and

$$\Theta_{ref} = \left( \frac{1}{3} + \frac{2}{3} \theta_{cr} \right) \left( \Theta_s - \Theta_r \right) + \Theta_r, \quad (7)$$

where $\theta_{cr}$ is the saturation ratio at which soil matrix vertical hydraulic conductivity equals 0.5 mm day$^{-1}$.

The values of $\Theta_w$ and $\Theta_{ref}$ for different soil types at the Shale Hills watershed can be found in Table 1. Because of the lack of measurements, empirical values are used for soil macropore properties and streambed properties, and those values are subject to calibration. Soil horizontal macropore hydraulic conductivity and soil vertical macropore hydraulic conductivity are assumed to be 100 times their corresponding soil matrix conductivities (Li 2010) and are subject to calibration. The vegetation map is obtained from a tree survey conducted in 2008. Vegetation parameters for different land cover types are obtained from the modified International Geosphere–Biosphere Programme (IGBP) Moderate Resolution Imaging Spectroradiometer (MODIS) 20-category vegetation (land use) data (http://www.ral.ucar.edu/research/land/technology/lsm/parameters/), which are also used in the Noah LSM. Root zone depths are assumed to be 0.6 m for all vegetation types. The model parameters and their a priori values are listed in Tables 1, 2, and 3.

c. Forcing data and evaluation data

Meteorological data from in situ measurements and remote sensing datasets are used to drive and evaluate the model (Table 4). Given the small scale (0.08 km$^2$) of the watershed, spatially uniform forcing is used for this study. To enable the model to respond to realistic
temporal variations in meteorology, as many in situ measurements as possible are used to drive the model. Precipitation, air temperature, and relative humidity data are obtained from the RTHnet weather station (http://cataract.c ee.psu.edu/czo/rth/). Precipitation is measured with a Hach OTT precipitation gauge, while air temperature and relative humidity are measured with a Campbell Scientific HMP45C temperature and relative humidity probe. Those 10-min interval data are aggregated into hourly data. Downward longwave radiation, downward solar radiation, and surface pressure (prior to April 2009) are not available at SSHCZO and are obtained from the Surface Radiation Budget Network (SURFRAD) Penn State University station. The SURFRAD Penn State University station is located at 40.72°N, 77.93°W, which is 6.48 km away from the Shale Hills watershed. It is assumed that the radiation and surface pressure at the Penn State SURFRAD station represent the conditions at the Shale Hills watershed. Wind speed data from January 2009 to March 2009 are obtained from the Penn State SURFRAD station. An above-canopy eddy-covariance flux tower was installed in SSHCZO in April 2009. Wind speed and air pressure from the flux tower are used for the period from April to December 2009.

MODIS LAI data are introduced to represent vegetation phenology. MODIS product MOD15A2 provides 8-day composite LAI data at 1-km resolution (Knyazikhin et al. 1999; Myneni et al. 2002). LAI field measurements have also been collected regularly from 25 April to 31 October 2010 at SSHCZO, using a Decagon AccuPAR meter and a LI-COR 2200 leaf area. The MODIS LAI data in 2009 are rescaled based on the comparison between the MODIS product and the CZO field measurements to drive the model:

\[
LAI = \left( \frac{LAI_{\text{MODIS}} - LAI_0}{\max(LAI_{\text{CZO}}) - LAI_0} \right) \max(LAI_{\text{MODIS2010}}) - LAI_0 + LAI_0,
\]

where LAI is the rescaled MODIS LAI, LAI_{MODIS} is the MODIS LAI measurements, LAI_{CZO} is the watershed-averaged SSHCZO field-measured LAI in 2010, and LAI_0 is a reference value set to be 0.5 m^2 m^{-2}.

Evaluation data include outlet discharge, groundwater level, soil moisture, soil temperature, and eddy-covariance surface heat flux measurements (Table 4) of the year 2009. A V-notch weir measures water level at the outlet of catchment. The measured stream stage is converted to discharge rate using a rating curve developed by Nutter (1964) for the V-notch weir at the Shale Hills watershed. Groundwater level and soil moisture data are collected from three RTHnet wells drilled near

Table 1. Soil parameters used for the Shale Hills watershed domain. Listed values are the a priori (uncalibrated) parameter values. All parameters in this table are calibrated in the optimization process.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Source</th>
<th>Weikert</th>
<th>Berks</th>
<th>Rushtown</th>
<th>Blairton</th>
<th>Ernest</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{satV}$</td>
<td>Vertical saturated hydraulic conductivity of infiltration layer (m day^{-1})</td>
<td>Lin (2006)</td>
<td>18.11</td>
<td>9.09</td>
<td>5.20</td>
<td>0.87</td>
<td>3.93</td>
</tr>
<tr>
<td>$K_V$</td>
<td>Vertical saturated hydraulic conductivity (m day^{-1})</td>
<td>Lin (2006)</td>
<td>5.89</td>
<td>0.89</td>
<td>1.15</td>
<td>0.26</td>
<td>0.86</td>
</tr>
<tr>
<td>$K_H$</td>
<td>Horizontal saturated hydraulic conductivity (m day^{-1})</td>
<td>Lin (2006)</td>
<td>10.09</td>
<td>2.68</td>
<td>65.09</td>
<td>9.35</td>
<td>2.68</td>
</tr>
<tr>
<td>$\Theta_s$</td>
<td>Porosity (m^3 m^{-3})</td>
<td>Lin (2006)</td>
<td>0.53</td>
<td>0.37</td>
<td>0.38</td>
<td>0.34</td>
<td>0.40</td>
</tr>
<tr>
<td>$\Theta_r$</td>
<td>Residual porosity (m^3 m^{-3})</td>
<td>SSURGO</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>$\Theta_{ref}$</td>
<td>Field capacity (m^3 m^{-3})</td>
<td>Eq. (7)</td>
<td>0.41</td>
<td>0.32</td>
<td>0.29</td>
<td>0.28</td>
<td>0.31</td>
</tr>
<tr>
<td>$\Theta_w$</td>
<td>Wilting point (m^3 m^{-3})</td>
<td>Eq. (6)</td>
<td>0.09</td>
<td>0.07</td>
<td>0.06</td>
<td>0.06</td>
<td>0.07</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Van Genuchten soil parameter (m^{-1})</td>
<td>SSURGO</td>
<td>2.46</td>
<td>2.51</td>
<td>2.84</td>
<td>2.79</td>
<td>3.27</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Van Genuchten soil parameter (-)</td>
<td>SSURGO</td>
<td>1.20</td>
<td>1.21</td>
<td>1.33</td>
<td>1.33</td>
<td>1.32</td>
</tr>
<tr>
<td>$K_{macV}$</td>
<td>Vertical macropore hydraulic conductivity (m day^{-1})</td>
<td>Empirical</td>
<td>100 $K_{satV}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{macH}$</td>
<td>Horizontal macropore hydraulic conductivity (m day^{-1})</td>
<td>Empirical</td>
<td>100 $K_H$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f_V, f_H$</td>
<td>Vertical and horizontal area fraction of macropores (m^2 m^{-2})</td>
<td>Empirical</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D_{mac}$</td>
<td>Macropore depth (m)</td>
<td>Empirical</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Parameters used for the river segments. Listed values are the uncalibrated parameter values. All parameters in this table are calibrated in the optimization process.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Source</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_{riv}$</td>
<td>Manning’s roughness coefficient (s m^{-1/3})</td>
<td>Empirical</td>
<td>0.04</td>
</tr>
<tr>
<td>$K_{rivH}$</td>
<td>Riverside hydraulic conductivity (m day^{-1})</td>
<td>Empirical</td>
<td>0.1</td>
</tr>
<tr>
<td>$K_{rivV}$</td>
<td>Riverbed hydraulic conductivity (m day^{-1})</td>
<td>Empirical</td>
<td>1.0</td>
</tr>
</tbody>
</table>
the stream. Each well is equipped with one water level sensor (Druck pressure transducer CS420-L manufactured by Campbell Scientific) and three soil moisture sensors (Decagon Echo2 probes). The three soil moisture sensors at each well are located at three different levels below ground, with depths of 0.1, 0.3, and 0.5 m below the surface, respectively. Among the three RTHnet wells, Well 1 (Fig. 2) is not deep enough to capture the change of the water table in the summer months, when the water table drops below the screened well depth. For those time periods, only measurements from RTHnet Wells 2 and 3 are used. Two of the nine soil moisture sensors continuously show measurements larger than 1 m$^3$ m$^{-3}$. Therefore, data from those two sensors are not used in this study. The water table depth (WTD; distance from the land surface to the groundwater table) measurements collected in different wells are averaged to represent the observed WTD at RTHnet wells. The multiple volumetric soil moisture content observations from those three wells are averaged to represent the observed soil water content (SWC) over the soil column. The standard deviations from different water level and soil moisture sensor measurements are calculated as the uncertainty range for observed WTD and SWC. In Flux-PIHM, the model domain is discretized as such that the three RTHnet wells are located at three vertices of one model grid for the convenience of model–data comparison. This model grid surrounded by RTHnet wells is selected for the comparison of WTD and SWC. Soil temperature measurements are collected at different locations and different vertical levels in 2009. Soil temperatures are measured with 229 probes manufactured by Campbell Scientific and 5TE probes manufactured by Decagon. The locations of the five soil temperature sites are shown in Fig. 2. For those five sites, the soil temperature data measured at 5 cm below surface are averaged. Model predictions of soil temperature at the top soil layer (0–10 cm) from the five corresponding model grids are averaged to be compared with the observations.

The above-canopy eddy-covariance flux tower measures wind speed and air temperature with a Campbell Scientific CSAT3 three-dimensional sonic anemometer. Carbon dioxide and water vapor concentrations are measured with a LI-COR LI-7500 CO$_2$/H$_2$O analyzer. Eddy-covariance fluxes are calculated following the quality control methods documented by Vickers and

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{\text{cmin}}$</td>
<td>Minimum stomatal resistance (s m$^{-1}$)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$R_{\text{g}}$</td>
<td>Reference visible solar radiation (W m$^{-2}$)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$A_{\text{min}}$</td>
<td>Minimum albedo (–)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$A_{\text{max}}$</td>
<td>Maximum albedo (–)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$e_{\text{min}}$</td>
<td>Minimum emissivity (–)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$e_{\text{max}}$</td>
<td>Maximum emissivity (–)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$z_0$</td>
<td>Minimum roughness length (m)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$h_s$</td>
<td>Maximum roughness length (m)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$h_v$</td>
<td>Water vapor exchange coefficient (–)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$d_{\text{root}}$</td>
<td>Root zone depth (m)</td>
<td>Empirical</td>
</tr>
<tr>
<td>$C_{\text{zil}}$</td>
<td>Zilitinkevich parameter (–)</td>
<td>Empirical</td>
</tr>
</tbody>
</table>

*Zilitinkevich (1995)*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{\text{cmin}}$</td>
<td>Minimum stomatal resistance (s m$^{-1}$)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$R_{\text{g}}$</td>
<td>Reference visible solar radiation (W m$^{-2}$)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$A_{\text{min}}$</td>
<td>Minimum albedo (–)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$A_{\text{max}}$</td>
<td>Maximum albedo (–)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$e_{\text{min}}$</td>
<td>Minimum emissivity (–)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$e_{\text{max}}$</td>
<td>Maximum emissivity (–)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$z_0$</td>
<td>Minimum roughness length (m)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$h_s$</td>
<td>Maximum roughness length (m)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$h_v$</td>
<td>Water vapor exchange coefficient (–)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$d_{\text{root}}$</td>
<td>Root zone depth (m)</td>
<td>Empirical</td>
</tr>
<tr>
<td>$C_{\text{zil}}$</td>
<td>Zilitinkevich parameter (–)</td>
<td>Empirical</td>
</tr>
</tbody>
</table>

*Zilitinkevich (1995)*

Table 3. Vegetation parameters used for the Shale Hills watershed domain. Only the parameter $R_{\text{cmin}}$ is calibrated in the optimization process.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{\text{cmin}}$</td>
<td>Minimum stomatal resistance (s m$^{-1}$)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$R_{\text{g}}$</td>
<td>Reference visible solar radiation (W m$^{-2}$)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$A_{\text{min}}$</td>
<td>Minimum albedo (–)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$A_{\text{max}}$</td>
<td>Maximum albedo (–)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$e_{\text{min}}$</td>
<td>Minimum emissivity (–)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$e_{\text{max}}$</td>
<td>Maximum emissivity (–)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$z_0$</td>
<td>Minimum roughness length (m)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$h_s$</td>
<td>Maximum roughness length (m)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$h_v$</td>
<td>Water vapor exchange coefficient (–)</td>
<td>Modified IGBP MODIS</td>
</tr>
<tr>
<td>$d_{\text{root}}$</td>
<td>Root zone depth (m)</td>
<td>Empirical</td>
</tr>
<tr>
<td>$C_{\text{zil}}$</td>
<td>Zilitinkevich parameter (–)</td>
<td>Empirical</td>
</tr>
</tbody>
</table>

*Zilitinkevich (1995)*

Table 4. Forcing and evaluation data used for simulation.

<table>
<thead>
<tr>
<th>Data</th>
<th>Usage</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precipitation, 2-m air temperature, and relative humidity</td>
<td>Forcing</td>
<td>RTHnet weather station</td>
</tr>
<tr>
<td>Downward longwave radiation, downward solar radiation</td>
<td>Forcing</td>
<td>SURFRAD</td>
</tr>
<tr>
<td>Wind speed and surface pressure</td>
<td>Forcing</td>
<td>RTHnet flux tower and SURFRAD (prior to 1 April 2009)</td>
</tr>
<tr>
<td>LAI</td>
<td>Forcing</td>
<td>Modified MODIS data</td>
</tr>
<tr>
<td>River discharge</td>
<td>Evaluation</td>
<td>RTHnet outlet gauge</td>
</tr>
<tr>
<td>Water table depth and soil water content</td>
<td>Evaluation</td>
<td>RTHnet wells</td>
</tr>
<tr>
<td>Soil temperature</td>
<td>Evaluation</td>
<td>RTHnet soil temperature sites</td>
</tr>
<tr>
<td>Surface heat fluxes</td>
<td>Evaluation</td>
<td>RTHnet flux tower</td>
</tr>
</tbody>
</table>
Mahrt (1997). The gaps in processed 30-min sensible and latent heat fluxes are filled using the lookup table method suggested by Falge et al. (2001). The gap-filled sensible and latent heat fluxes are aggregated into hourly data to be compared with hourly predictions of sensible and latent heat fluxes averaged over the model domain.

4. Optimization and evaluation of Flux-PIHM

a. Model optimization and spinup

Flux-PIHM has a large number of tunable model parameters. Some of those parameters are dependent on soil type or land cover type. Thus, calibration could be very complicated if the model domain contains a large number of soil types or land cover types. To simplify the calibration problem, the single global calibration multiplier method (or one-factor method) (Pokhrel and Gupta 2010; Wallner et al. 2012) is adopted. One single global calibration coefficient is used for each model parameter regardless of soil type or land cover type. The global calibration coefficient is a multiplier acting on the corresponding soil- or vegetation-related parameter for all soil or vegetation types. By applying global multiplier calibration coefficients, the dimension of parameter space for calibration is reduced and the ratios between uncalibrated a priori parameters of different soil/vegetation types are preserved. The default value for calibration coefficient is 1.0.

A comprehensive manual calibration is performed using the “trial and error” strategy. In the comprehensive calibration process, outlet discharge, water table depth, soil water content, soil temperature at 5 cm below surface, and eddy-covariance surface heat flux data from June to July 2009 are used to adjust model parameters. This period is chosen because it covers both peak discharge events and a base flow period and exhibits a considerable rate of evapotranspiration. The model parameters that are calibrated and their corresponding calibration coefficients are listed Table 5.

After calibration, Flux-PIHM spinup uses atmospheric forcing from 0000 UTC 20 October 2008 to 0000 UTC 1 January 2009, initialized from the “relaxation mode,” in which the model relaxes from a saturation state. In wet periods such as the winter conditions used for this spinup, it takes approximately three weeks for the models to reach an equilibrium state at the Shale Hills watershed. This 2-month spinup period is used to ensure the model eliminating the effects of initial conditions.

Initial surface skin temperature is set to the air temperature and the soil temperatures are obtained from a linear interpolation between the surface skin temperature and the bottom boundary condition.

b. Water budget

Predicted water budgets from Flux-PIHM and PIHM V2 are compared for the period from 1 March to 1 December 2009 and are presented in Fig. 4. Snow-covered
conditions are excluded from the model evaluation because Flux-PIHM is using simple snow physics that are not sufficient for accurate hydrologic prediction in winter. Therefore, this study focuses on the warm seasons. Runoff in Fig. 4 is calculated as the total discharge divided by the total area of the watershed. Predicted total runoffs from both models are close to the observation, and Flux-PIHM provides a closer total runoff prediction compared with PIHM V2. Flux-PIHM overestimates the total runoff by only 0.4% (1 mm) while PIHM V2 overestimates the total runoff by 4.1% (11 mm).

The total evapotranspiration predicted by both models are close and only differ by 8 mm. The partitioning of total evapotranspiration, however, is very different in the two models. Flux-PIHM predicts higher canopy evaporation and transpiration compared with PIHM V2, while PIHM V2 predicts higher soil evaporation. Soil evaporation and transpiration extract water from different subsurface layers. Soil evaporation only has access to the top soil layer, while transpiration can transport deep soil water or even groundwater to the atmosphere, depending on the root zone depth and the water table depth. For the six model grids near the stream outlet, transpiration that is extracted directly from the groundwater is calculated. For those six model grids, in Flux-PIHM 34.8% of total evapotranspiration is extracted directly from groundwater, while in PIHM V2, this fraction is only 22.1% because of the lower transpiration prediction in PIHM V2. Such differences will produce different soil moisture profiles and thus affect evapotranspiration and flood/drought forecasting.

c. Evaluation of hydrologic predictions: Discharge, water table depth, and soil water content

To test the model’s ability to predict the hydrological states of the watershed, model discharge, WTD, and SWC predictions are compared with in situ measurements. Several criteria are used to quantify model predictions of discharge, including the total bias, the Nash–Sutcliffe model efficiency coefficient (NSE; Nash and Sutcliffe 1970), and the correlation coefficient with observation. The total bias is defined as

$$\text{Total bias} = \frac{\sum_{t=1}^{T} (Q'_m - Q'_o)}{\sum_{t=1}^{T} Q'_o} \times 100\% ,$$

where $Q_o$ is the observed discharge, $Q_m$ is the predicted discharge, $T$ is the length of observation, and a superscript $t$ indicates observation or modeled discharge at the time step $t$. NSE is the most commonly used coefficient to evaluate hydrologic models. It is defined as

$$\text{NSE} = 1 - \frac{\sum_{t=1}^{T} (Q'_m - Q'_o)^2}{\sum_{t=1}^{T} (Q'_o - \overline{Q'_o})^2}.$$  

(10)

The correlation coefficient is defined as

$$R = \frac{\sum_{t=1}^{T} (Q'_m - \overline{Q'_m})(Q'_o - \overline{Q'_o})}{\sqrt{\sum_{t=1}^{T} (Q'_m - \overline{Q'_m})^2 \sum_{t=1}^{T} (Q'_o - \overline{Q'_o})^2}} .$$

(11)

Figure 5 presents the comparison of hourly outlet discharge among RTHnet measurements and two models from March to December 2009. The statistics of discharge predictions are presented in Table 6.

Flux-PIHM predictions agree well with measurements for both dry period flows and peak flows as illustrated in Fig. 5. Though June and July data are used for calibration, Flux-PIHM still overestimates the peak event in June. In March, April, and May, Flux-PIHM underestimates base flows but overestimates some discharge peaks. The largest errors occur in October, when Flux-PIHM underestimates the highest peak as well as overestimates a series of peak events before that.

Flux-PIHM and PIHM V2 discharge simulations are almost indistinguishable for most of the period. The only notable differences are for the peak events in May, June, and October. The predictions of the two models differ substantially for the 15–19 October flood event, which is highlighted in Fig. 5. The observed peak discharge rate is 277 m$^3$ day$^{-1}$, the Flux-PIHM predicted peak discharge rate is 778 m$^3$ day$^{-1}$, and the PIHM V2 predicted peak discharge rate is 1516 m$^3$ day$^{-1}$. The timings of the predicted peak discharge also differ by 6 h in the two models. As for the average performance, Table 6 shows that Flux-PIHM has improved performance in discharge prediction over PIHM V2, especially at the daily time step.

The comparisons of WTD and SWC among RTHnet well measurements and Flux-PIHM and PIHM V2 predictions are presented in Fig. 6. The biases, correlation coefficients, and root-mean-square errors (RMSEs) of WTD and SWC predictions compared with RTHnet wells are presented in Table 7. Figure 6 shows that Flux-PIHM predictions of WTD and SWC agree well with RTHnet measurements. For most of the evaluation
period, the differences between Flux-PIHM predictions and RTHnet measurements are within the observation uncertainty ranges for both WTD and SWC. Like discharge predictions, the largest errors appear in October, when model errors in WTD and SWC fall out of the observation uncertainty ranges. In October, Flux-PIHM continuously underestimates WTD and overestimates SWC, when the error in WTD reaches about $-0.5$ m and the error in SWC reaches up to $0.1 \text{ m}^3 \text{ m}^{-3}$. In spring (March–May), although model errors of WTD and SWC are always within the observation uncertainty ranges, the model predictions have a consistent wet bias (low bias in WTD and high bias in SWC) compared with observed average WTD and SWC. Figure 6 also shows that the PIHM V2 predictions in WTD and SWC follow the observations better than Flux-PIHM in March, but PIHM V2 is outperformed by Flux-PIHM in almost all following months. The largest differences between Flux-PIHM and PIHM V2 appear in October, when the differences between the two models reach up to 0.25–0.30 m. As indicated in Table 7, Flux-PIHM provides better overall predictions of WTD and SWC than PIHM V2, although PIHM V2 WTD bias is slightly smaller.

d. Evaluation of surface energy balance predictions: Sensible and latent heat fluxes and soil temperature

The aggregated hourly sensible ($H$) and latent ($LE$) heat fluxes measured at SSHCZO are compared with model hourly predictions averaged over the model domain. Both $H$ and $LE$ measurements at SSHCZO are available from 1 April 2009. The period from 1 April 2009 to 1 January 2010 is divided into four subperiods to evaluate surface heat flux predictions for different seasons: 1 April to 1 June 2009 as spring, 1 June to 1 September 2009 as summer, 1 September to 1 December 2009 as fall, and 1 December 2009 to 1 January 2010 as winter. Plots of hourly $H$ and $LE$ are not presented to save space: instead, Table 8 lists the statistics for each season based on the comparison between hourly predictions and observations of $H$ and $LE$. Besides, Fig. 7 presents the comparisons of surface heat fluxes among Flux-PIHM, PIHM V2, and RTHnet eddy-covariance measurements as averaged daily cycles for each season. Flux-PIHM predicts both $H$ and $LE$, but PIHM V2 only predicts evapotranspiration rate, which is equivalent to $LE$.

Figure 7 shows that Flux-PIHM predictions of midday $H$ and $LE$ have consistent high biases during all seasons.

### Table 6. Bias, NSE, and correlation coefficient of Flux-PIHM and PIHM V2 hourly discharge predictions compared with RTHnet measurements from 1 March to 1 December 2009. NSE is defined as $\text{NSE} = 1 - \frac{\sum_{t=1}^{T}(Q_{m} - Q_{o})^2}{\sum_{t=1}^{T}(Q_{o} - \overline{Q}_{o})^2}$.

<table>
<thead>
<tr>
<th>Model</th>
<th>Calibration period</th>
<th>Total bias (%)</th>
<th>NSE (hourly)</th>
<th>$R$ (hourly)</th>
<th>NSE (daily)</th>
<th>$R$ (daily)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flux-PIHM</td>
<td>2 months</td>
<td>0.30</td>
<td>0.73</td>
<td>0.90</td>
<td>0.85</td>
<td>0.97</td>
</tr>
<tr>
<td>PIHM V2</td>
<td>2 months</td>
<td>4.13</td>
<td>0.69</td>
<td>0.89</td>
<td>0.75</td>
<td>0.95</td>
</tr>
</tbody>
</table>
It needs to be noted that Flux-PIHM closes the energy budget at the land surface, while eddy-covariance measurements typically fail to close the energy budget, with the measured $H + LE$ tending to be less than $R_n - G$ (McNeil and Shuttleworth 1975; Fritschen et al. 1992; Twine et al. 2000; Foken 2008). Therefore, in the calibration process, Flux-PIHM is subjectively calibrated to yield higher $H$ and $LE$, assuming the eddy-covariance method underestimates surface heat fluxes. Generally, Flux-PIHM prediction of $LE$ has a larger bias in the afternoon than in the morning, while $H$ has a larger bias in the morning than in the afternoon (Fig. 7). Except for spring and winter $LE$, Flux-PIHM hourly predictions of $H$ and $LE$ have high correlations with RTHnet measurements, with correlation coefficients between 0.91 and 0.94 (Table 8), which shows that Flux-PIHM is capable of capturing the variation of surface heat fluxes well at hourly resolution. Flux-PIHM bias of $LE$ in spring is much higher than in other seasons (Fig. 7). The RMSE for spring $LE$ is larger than the RMSEs in other seasons, and the correlation coefficient for $LE$ in spring is also relatively lower than in summer and fall (Table 8). This substantial high bias in spring agrees with the signs of biases in WTD and SWC predictions in the corresponding months (Fig. 6). Because the calibration of Flux-PIHM only focused on summer months, the seasonality of evapotranspiration is not taken into account in calibration. Using a longer calibration period or including more parameters for calibration would improve the model prediction of the seasonality of evapotranspiration. It is also possible that this high bias in spring is caused by model structural errors or forcing data errors. The correlation coefficient of winter $LE$ is as low as 0.55, but $LE$ measurements of an open-path gas analyzer in winter are less reliable than in other seasons (Burba et al. 2008). The insufficient winter physics in Flux-PIHM may also contribute to the low correlation of $LE$ in winter.

When averaged into daily cycles, Flux-PIHM and PIHM V2 $LE$ are almost indistinguishable in summer and winter (Fig. 7), but predictions of $LE$ by Flux-PIHM are higher in spring and fall. In spring, summer, and fall, Flux-PIHM $LE$ predictions have higher correlation coefficients and lower RMSEs compared to PIHM V2, which indicates that Flux-PIHM is more capable of reproducing the hourly variation in evapotranspiration.

Figure 8 presents the comparison of hourly soil temperature at 5 cm below the surface between Flux-PIHM predictions and RTHnet observations. From March to November, the average bias for soil temperature prediction at 5 cm below the surface is 0.81°C and the RMSE is 1.6°C. The correlation coefficient between model and measurement is 0.97. Figure 8 suggests that the soil temperature predicted by Flux-PIHM has a consistent high bias in summer, but the errors are random in other seasons. At the beginning of March, the

| Table 7. Bias, RMSE, and correlation coefficient of Flux-PIHM and PIHM V2 hourly WTD and SWC predictions compared with RTHnet well measurements from 1 March to 1 December 2009. |
|-----------------|-------|------|-----------------|-------|------|
|                | WTD   |      | SWC             |      |      |
|                | Bias  | $R$  | RMSE (m)        | Bias  | $R$  | RMSE (m$^3$ m$^{-2}$) |
| Flux-PIHM      | $-0.12$ | 0.84 | 0.21            | 0.01  | 0.76 | 0.02 |
| PIHM V2        | $-0.11$ | 0.80 | 0.23            | 0.01  | 0.73 | 0.03 |
error could be as large as ±5°C. This is probably because of the simple snow physics in Flux-PIHM, which may not simulate thaw well.

e. Correlation between surface heat fluxes and water table depth

The coupled land surface hydrologic model is a good tool for the study of the interaction between the subsurface and the land surface. Annual average sensible and latent heat fluxes as functions of WTD in each model grid from Flux-PIHM simulation for the year 2009 are presented in Fig. 9. Sensible heat fluxes are positively correlated with water table depth and latent heat fluxes are negatively correlated with water table depth. It is also clear that the change of surface heat fluxes with respect to WTD for the Ernest soil type (the diamond-shaped symbols) is much more pronounced than for the other soil types. Because the Ernest soil type is found primarily in the lowest portions of the valley (Fig. 3b), it indicates that impacts of groundwater are stronger on those grids near the stream than the other model grids. For the grids near the stream, the groundwater table is relatively shallow and roots have access to groundwater; thus, groundwater has direct impacts on the surface energy balance. For the grids far from the stream, the groundwater table is deep and roots have no access to groundwater; thus, the impacts of groundwater are weak. For those grids, the effect of groundwater on surface heat fluxes is indirect and is accomplished by affecting the soil moisture profile.

This correlation between surface heat fluxes and groundwater table has been found in other studies (e.g., Kollet and Maxwell 2008). Kollet and Maxwell (2008) also discovered that when the water table depth is smaller than 1 m or greater than 5 m in the Little Washita watershed, Oklahoma, the land surface and groundwater are decoupled. This decoupling of land surface and groundwater is not found in this study at the Shale Hills watershed.

5. Discussion and conclusions

This paper presents a physically based, land surface hydrologic model, Flux-PIHM, that performs well in a
Flux-PIHM simulates interactions between the surface energy balance and hydrologic processes that cannot be achieved without the coupling of these model components. The model results indicate that mean annual surface heat fluxes are highly correlated with water table depth at a significant subset of the grid points in the watershed. The coupling between groundwater and land surface is affected by soil type, land cover type, and topography. The strong coupling between SEB and water table depth implies that LSMs that do not simulate water table are prone (at least at some grid points) to long-term biases as the water table fluctuates.
Flux-PIHM is compared with PIHM V2, which does not have a surface energy balance simulation. Flux-PIHM is shown to have a significantly different water budget for the test watershed, while maintaining similar or slightly improved performance in the components simulated by both models. Flux-PIHM predicts 44% (372 mm) of the total water export from the watershed in March–December 2009 in the form of transpiration, compared to 37% (311 mm) for PIHM V2, and only 15% (127 mm) in the form of soil evaporation, compared to 23% (192 mm) for PIHM V2. These differences, though not dramatic, have the potential to result in substantially different soil water profiles and responses of the watershed to flood or drought events. Comparison to direct measurements upscaled sap flux and spatially resolved soil moisture measurements would enable future studies to determine which system yields a more accurate simulation of watershed function.

Results indicate that Flux-PIHM does slightly improve the overall hourly predictions of outlet discharge, WTD, and SWC compared with PIHM V2. Although the improvement of model discharge predictions are not clear in Fig. 5, the Flux-PIHM prediction has a smaller total bias, a higher correlation coefficient with observation, and a higher NSE compared with PIHM V2. Differences in discharge predictions are noticeable for some high discharge peaks, especially after the extended dry, low-flow summer period. For a flood event from 15 to 19 October, Flux-PIHM and PIHM V2 provide substantially different predictions in peak discharge rate and timing: the Flux-PIHM simulation is noticeably closer to the observed discharge. Because the hydrologic schemes in those two models are identical, the differences in the predictions of discharge rate and timing are likely caused by the different evapotranspiration calculations. These differences demonstrate the impact of evapotranspiration simulation on flood prediction. We need to point out that the comparisons between Flux-PIHM and PIHM V2 are made by using the same manually calibrated parameter values, and those parameter values may not represent the optimal parameters for both models, although good agreements between model predictions and observations have been shown.

The errors that remain in Flux-PIHM postcalibration illustrate areas for potential future improvements in the modeling system. Notable errors in Flux-PIHM predictions occur in March–May and October. From March to May, the model systematically overestimates SWC and underestimates WTD. Perhaps because of the high bias in soil moisture, Flux-PIHM produces a high bias in LE during this period. Spring simulations are challenging and could be improved by better winter period/snow physics. In October, the model underestimates the highest peak discharge event but overestimates some prior peak discharge events. The largest errors in WTD and SWC during the evaluation period occur in October as well. Improved simulation of the recovery of the watershed from dry, low-flow conditions is another topic of future research.

Acknowledgments. This research was supported by the National Oceanic and Atmospheric Administration through Grant NA10OAR4310166, and the National Science Foundation Susquehanna Shale Hills Critical Zone Observatory project through Grant EAR 0725019. We thank Dr. Henry Lin, Dr. David Eissenstat, and Dr. Kusum Naithani for providing the data and Kelly Cherrey and Colin Duffy for their support of the instrumentation and data collection.

REFERENCES


