

Using a spatially-distributed hydrologic biogeochemistry model with a nitrogen transport module to study the spatial variation of carbon processes in a Critical Zone Observatory

Yuning Shi^{a,*}, David M. Eissenstat^a, Yuting He^b, Kenneth J. Davis^b

^a Department of Ecosystem Science and Management, The Pennsylvania State University, University Park, PA, USA

^b Department of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, PA, USA

ARTICLE INFO

Keywords:

Ecohydrology
Spatial patterns
Critical Zone Observatory
Nitrogen transport

ABSTRACT

Terrestrial carbon processes are affected by soil moisture, soil temperature, nitrogen availability and solar radiation, among other factors. Most of the current ecosystem biogeochemistry models represent one point in space, and have limited characterization of hydrologic processes. Therefore, these models can neither resolve the topographically driven spatial variability of water, energy, and nutrient, nor their effects on carbon processes. A spatially-distributed land surface hydrologic biogeochemistry model, Flux-PIHM-BGC, is developed by coupling the Biome-BGC model with a physically-based land surface hydrologic model, Flux-PIHM. In the coupled system, each Flux-PIHM model grid couples a 1-D Biome-BGC model. In addition, a topographic solar radiation module and an advection-driven nitrogen transport module are added to represent the impact of topography on nutrient transport and solar energy distribution. Because Flux-PIHM is able to simulate lateral groundwater flow and represent the land surface heterogeneities caused by topography, Flux-PIHM-BGC is capable of simulating the complex interaction among water, energy, nutrient, and carbon in time and space. The Flux-PIHM-BGC model is tested at the Susquehanna/Shale Hills Critical Zone Observatory. Model results show that distributions of carbon and nitrogen stocks and fluxes are strongly affected by topography and landscape position, and tree growth is nitrogen limited. The predicted aboveground and soil carbon distributions generally agree with the macro patterns observed. Although the model underestimates the spatial variation, the predicted watershed average values are close to the observations. The coupled Flux-PIHM-BGC model provides an important tool to study spatial variations in terrestrial carbon and nitrogen processes and their interactions with environmental factors, and to predict the spatial structure of the responses of ecosystems to climate change.

1. Introduction

The future of the Earth's climate is extremely sensitive to the changes in land surface (Friedlingstein et al., 2014) because of its ability to take up or emit large amounts of carbon dioxide and its impact on water and energy cycling. The terrestrial carbon cycle is a major contributor to uncertainties in future climate projections (Bodman et al., 2013). Terrestrial biogeochemistry models, which simulate ecosystem biogeochemical cycling of water, carbon, and nutrient, are therefore important in predicting the future of the Earth's climate, and have been included in the new generation of land models (i.e., the land components of Earth system models, or ESMs) and land surface models (e.g., Oleson et al., 2008; Niu et al., 2011).

Hydrologic processes have strong impacts on the terrestrial carbon cycle through their controls on photosynthesis, organic matter decomposition and nutrient transport (Rodríguez-Iturbe et al., 2001; Ju et al.,

2006; Oleson et al., 2008; Lohse et al., 2009). Topographically driven lateral water flow and associated nutrient transport result in heterogeneously distributed soil water and nutrients availability, which lead to the spatial heterogeneity of land surface processes and biogeochemical processes. The description of hydrologic processes in terrestrial biogeochemistry models, however, is often highly simplified. Fisher et al. (2014) examined a total of 75 terrestrial biosphere models (TBMs), and indicated that most TBMs still use tipping or leaky bucket-based approaches for hydrology. Clark et al. (2015) reviewed hydrologic modeling advances in the land modeling community, and found that both upper (i.e., infiltration) and lower (i.e., recharge or subsurface runoff) boundary conditions for subsurface hydrology are highly simplified in those models, deep soil water or groundwater dynamics are usually neglected, and lateral flows are not explicitly accounted for. These models therefore have limited ability in representing the impact of hydrologic processes on biogeochemical processes. Clark et al.

* Corresponding author.

(2015) concluded that the development of land model can be improved by incorporating recent advances in hydrologic modeling. Improving the representation of hydrology will be one of the foci of TBM development over the next few years, as revealed by a survey to TBM developers (Fisher et al., 2014).

Terrestrial biogeochemistry models and land models usually use a one-dimensional model to represent the average carbon fluxes and stocks (i.e., quantities of carbon contained in carbon pools) over a large spatial area. Even if they can accurately estimate the average soil moisture and soil temperature over the large spatial area, estimating the spatial average of carbon fluxes and stocks is difficult since the interaction between hydrological processes and biogeochemical processes are nonlinear and each model grid is unique in sub-grid topography, soil texture, and land cover distributions. These models cannot represent the fine scale (e.g., 10^1 – 10^2 m) spatial variability in terrestrial carbon distribution, which can exceed the variability in magnitude of carbon stock and flux at larger scales (Houghton, 2005).

Coupling physically-based high-resolution spatially-distributed hydrologic models with terrestrial biogeochemistry models may yield improvements in terrestrial carbon cycle predictions (Yu et al., 2015). Recently there have been attempts to develop spatially-distributed ecohydrological models that contain physically based hydrologic components, or couple hydrologic models with land models or land surface models with biogeochemistry components, to improve the representation of hydrologic processes in biogeochemical modeling. Readers are referred to Faticchi et al. (2016) for a list of ecohydrological models.

Ivanov et al. (2008) coupled the Vegetation Generator for Interactive Evolution (VEGGIE) model to a spatially-distributed physically based hydrological model, the TIN-based Real-time Integrated Basin Simulator, tRIBS (Tucker et al., 2001). Faticchi et al. (2012) developed a spatially-distributed ecohydrological model, Tethys-Chloris. Nutrient dynamics (e.g., nitrogen dynamics) and soil thermodynamics, however, are neglected in both VEGGIE+tRIBS and Tethys-Chloris; thus, those models are likely to have difficulty simulating nutrient- or radiation-limited environments.

Niu et al. (2014) developed an integrated catchment-scale ecohydrological model by coupling a physically-based 3-D hydrological model, CATchment HYdrology (CATHY) (Camporese et al., 2010), to a land surface model with leaf dynamics, NoahMP (Niu et al., 2011). The coupled CATHY/NoahMP model has been calibrated and tested at two small first-order watersheds with high spatial resolution (10^0 – 10^1 m), and showed good ability in simulating surface energy and water fluxes. The simulated watershed average CO_2 fluxes in spring, however, did not compare well with the observations, due to the lack of soil carbon processes in the NoahMP model.

Tague and Band (2004) developed a semi-distributed ecohydrological model, the Regional Hydro-Ecological Simulation System (RHESSys), which has been used in a number of ecohydrological studies. Shen et al. (2013) coupled a process-based quasi-3-D hydrologic model, Process-based Adaptive Watershed Simulator (PAWS) (Shen and Phanikumar, 2010), to the Community Land Model (CLM) (Oleson et al., 2010), and used the coupled PAWS+CLM model to study the impact of hydrologic processes on land surface and carbon dynamics (Shen et al., 2016). Kollet and Maxwell (2008) and Shrestha et al. (2014) coupled a three-dimensional variably saturated groundwater flow model, Parallel Flow (ParFlow) (Kollet and Maxwell, 2006) to CLM, and the coupled model has recently been applied at continental scale (Maxwell et al., 2015). Although lateral water flow is simulated in RHESSys, PAWS+CLM and CLM-ParFlow, nutrient transport with lateral water flow is not accounted for; RHESSys does not simulate surface energy balance or soil thermodynamics. In addition, none of the above models, except for Tethys-Chloris, simulates topographic solar radiation, which is an important factor in determining the spatial patterns of forest carbon dynamics (Smith et al., 2017) and other critical zone processes (Pelletier et al., 2018). There are also reactive solute transport models, e.g., the Unsaturated-Zone Flow-Reactive Transport in 3

Dimensions (UZF-RT3D; Bailey et al., 2013, 2015) model, that include the simulation of carbon and nitrogen cycling in the soil-plant system. These models, however, usually ignore or have highly simplified descriptions of aboveground processes.

A high-resolution, spatially-distributed, coupled hydrologic-land surface-terrestrial biogeochemistry model with nutrient transport simulation, which can reflect the impact of soil water, soil temperature, topographic solar radiation, and nutrient availability on ecosystem carbon processes, remains elusive to date. In addition, those spatially-distributed models are usually evaluated using watershed average measurements. The spatial patterns of carbon stocks and fluxes are often excluded from evaluation.

In this study, a coupled land surface hydrologic terrestrial biogeochemistry modeling system is developed by coupling the Penn State Integrated Hydrologic Model with a surface heat flux module (Flux-PIHM) (Shi et al., 2013) to the Biome-BGC model (Thornton et al., 2002). Flux-PIHM is a coupled land surface hydrologic model, which incorporates the Noah land surface model (Chen and Dudhia, 2001; Ek et al., 2003) into the Penn State Integrated Hydrologic Model (PIHM) (Qu and Duffy, 2007), a physically based spatially-distributed hydrologic model. Flux-PIHM has fully coupled surface and subsurface flow, lateral surface and subsurface water flow, macropore flow, and fully explicit river flow. The simulation of those processes has been identified as the key candidate areas to improve the hydrologic processes in land models by Clark et al. (2015). Flux-PIHM is able to represent the link between groundwater and the surface energy balance, as well as the land surface heterogeneities caused by topography at high spatial resolution (Shi et al., 2015a). This model is therefore an ideal candidate to improve the representation of hydrologic processes in TBMs. Biome-BGC is a process-based mechanistic terrestrial biogeochemistry model, and is the prototype of the carbon-nitrogen (CN) model in the CLM.

Past research has focused largely on the spatial variations of carbon processes with respect to precipitation, air temperature, nitrogen deposition, land cover, and land use (Raich and Schlesinger, 1992; Jobbágy and Jackson, 2000; Houghton, 2005). The Shale Hills watershed (0.08 km^2) is relatively homogeneous with respect to these properties, and enables study of the impact of additional processes including topography, watershed hydrology, soil physical properties and shading on the carbon cycle of a first-order watershed. We also note that low-order watersheds cover a large fraction of the landscape (Shreve, 1969; Benda et al., 2005). Therefore, we test the coupled model at the Shale Hills watershed, a first-order, monolithologic, forested watershed. The watershed is located in central Pennsylvania, and is one of the experimental sites of the Susquehanna Shale Hills Critical Zone Observatory (SSHCZO). The broad array of hydrological, land surface, and biogeochemical observations, including discharge, soil moisture, soil temperature, meteorological conditions, above ground carbon stocks and productivity, soil carbon stocks, leaf area index (LAI) and catchment-scale net ecosystem-atmosphere carbon fluxes, makes the Shale Hills watershed an ideal site for the coupled model test. The predicted spatial patterns of carbon stocks are evaluated using field measurements. We also examine the spatial patterns of carbon fluxes, as well as investigate the interaction of abiotic factors with vegetation carbon.

2. Description of the coupled modeling system

2.1. The Penn State Integrated Hydrologic Model with a surface heat flux module

The Penn State Integrated Hydrologic model with a surface heat flux module (Flux-PIHM) (Shi et al., 2013) is a coupled land surface hydrologic model. In Flux-PIHM, the land surface is decomposed into unstructured triangular grids and river channels are represented by rectangular elements. Channel flow and surface water flow calculations are handled by PIHM, using the 1-D (channel flow) and 2-D (surface

water flow) Saint-Venant (1871) equations. Subsurface horizontal flow is restricted to the saturated zone, and is calculated by PIHM using Darcy's law. The vertical water flow in the subsurface prismatic volume is calculated by both the PIHM and the Noah LSM: at each time step, PIHM solves for the total storage in the unsaturated zone and the saturated zone, by calculating infiltration rate, subsurface horizontal flow rate, and recharge rate; then the land surface module (adapted from the Noah LSM) uses a series of Richards equations to calculate the soil moisture content in up to 11 soil layers in each model grid, using PIHM simulated infiltration rate, lateral subsurface flow rate, and position of water table as boundary conditions. Surface energy balance, soil temperature, snow melt, interception, and drip are simulated in the LSM. Detailed descriptions and formulations of PIHM and Flux-PIHM are provided by Qu and Duffy (2007), Shi et al. (2013), and Shi et al. (2014a).

One of the distinguishing features of Flux-PIHM is the simulation of lateral groundwater flow. The lateral groundwater flux from element i to its neighbor at edge j is calculated as

$$q_{\text{sub}ij} = A_{ij} K_{\text{eff}} \frac{(H_g)_i - (H_g)_j}{D_{ij}}, \quad (1)$$

where A_{ij} is the projection area of saturated zone between elements i and j , K_{eff} is the effective horizontal hydraulic conductivity, D_{ij} is distance between the circumcenters of elements i and j , and $(H_g = h_g + z)_i$ is hydraulic head where z is elevation of datum, and h_g is the groundwater level.

In this study, a topographic solar radiation module is added to Flux-PIHM, taking into account the slopes (s) and aspects (a) of model grids. The solar position algorithm (SPA) developed by Reda and Andreas (2004) has been incorporated into Flux-PIHM. The SPA calculates the solar zenith and azimuth angles in the period from year 2000 BC to 6000 AD, with uncertainties of $\pm 0.0003^\circ$, based on date, time, and location on Earth (longitude, latitude, and elevation). The atmospheric refraction correction is taken into account, given the annual average local pressure and temperature. At each LSM time step, the SPA is used to estimate the solar zenith (φ) and azimuth angles (ψ). When the Sun is above the local horizon, the total solar radiation S received in each model grid is then calculated as the sum of direct radiation on the slope, diffuse radiation from the sky, and reflection from the surrounding terrain (Dozier and Frew, 1990; Tian et al., 2001):

$$S = S_{\text{dir}} \cos \theta + V_d S_{\text{dif}} + \bar{A} C_t (S_{\text{dir}} \cos \varphi + S_{\text{dif}}). \quad (2)$$

Here S_{dir} is the measured direct solar radiation, S_{dif} is the measured diffuse solar radiation, θ is the Sun incidence angle, i.e., the angle between the normal vector of the grid and the Sun vector, V_d is the sky view factor, which is the ratio of the sky visible at a point to that on an unobstructed horizontal surface, $\bar{A}$ is the average albedo of the adjacent grids, and C_t is the terrain configuration factor representing the contribution from the surroundings. When the Sun is below the local horizon, the first term on the right-hand-side of Eq. (2) becomes zero. The calculations of V_d and C_t used by Dozier and Frew (1990) are adopted:

$$V_d = \frac{1}{2\pi} \int_0^{2\pi} [\cos s \sin^2 H_\gamma + \sin s \cos(\gamma - a)(H_\gamma - \sin H_\gamma \cos H_\gamma)] d\gamma, \quad (3)$$

and

$$C_t = \frac{1 + \cos s}{2} - V_d, \quad (4)$$

where H_γ is the angle from the zenith downward to the local horizon at a discrete set of 36 directions γ_{1-36} around the circle. Note the topographic solar radiation module does not take into account the height of canopy, nor the effects of terrain outside the model domain. The topographic solar radiation module enables the quantitative investigation of the effects of aspect on terrestrial biogeochemical processes,

especially at small watersheds.

2.2. Biome-BGC

Biome-BGC (Thornton et al., 2002) is a 1-D mechanistic biogeochemistry model that simulates daily states and fluxes of carbon, nitrogen, and water using prescribed soil conditions and daily meteorology. It represents one point in space (ranging from 1 m^2 to the entire globe) with all carbon, nitrogen, and water fluxes and stocks scaled to a unit area basis. The water processes include canopy interception of precipitation, snow melt and sublimation, canopy evapotranspiration, soil evaporation, and water outflow. The carbon processes include growth respiration, maintenance respiration, photosynthesis (for both sun and shade leaves), decomposition, allocation, and mortality. Lastly, the nitrogen processes include deposition, fixation, leaching, nitrogen plant uptake, mineralization, immobilization, and volatilization. The gross primary production (GPP) is simulated with the Farquhar photosynthesis model (Farquhar et al., 1980), and the net primary productivity (NPP) is GPP minus maintenance respiration (a Q_{10} model) and growth respiration (a constant fraction of GPP). Detailed descriptions and formulations of Biome-BGC can be found in White et al. (2000) and Thornton et al. (2002).

2.3. Coupled land surface hydrologic biogeochemistry modeling system

Flux-PIHM-BGC is a coupled land surface hydrologic biogeochemistry modeling system that has a land surface hydrologic component (Flux-PIHM) and a biogeochemistry (BGC) module (adapted from Biome-BGC). Several changes are made to Biome-BGC in order to be coupled to Flux-PIHM.

The first change is Biome-BGC's phenology model. Biome-BGC uses a "look-ahead" approach to calculate the appropriate onset and senescence dates, by looking ahead at the soil temperature, which is calculated as the 11-day running weighted average of daily average temperature. In the coupled system, the BGC module can no longer look ahead because soil temperature is predicted by Flux-PIHM. Therefore, we replace the Biome-BGC phenology model using the CLM-CN prognostic phenology model. In this phenology model, the seasonal deciduous trees have a single annual growing season. Leaf onset is initiated when the degree-day summation of soil temperatures is above a critical value defined by White et al. (1997). Leaf senescence starts when day length is shorter than 10 h and 55 min. The period for carbon and nitrogen transfer from storage to new growth is set to 30 days and the period for complete leaf litter fall is set to 15 days as in CLM-CN.

The second change is Biome-BGC's stomatal resistance model. Both Flux-PIHM and Biome-BGC use Jarvis (1976) type models for stomatal resistance, which is controlled by photosynthetically active radiation, soil moisture, air temperature, and air humidity. The models, however, are slightly different. To make sure that transpiration (simulated by Flux-PIHM) and photosynthesis (simulated by the BGC module) use the same stomatal resistance, the Biome-BGC stomatal resistance is replaced by the Flux-PIHM (Noah LSM) stomatal resistance for the calculation of daily photosynthesis.

The third change is Biome-BGC's soil matric potential model. The Cosby et al. (1984) soil matric potential equation is replaced by the van Genuchten (1980) equation used in Flux-PIHM.

The fourth and the most important change is Biome-BGC's nitrogen deposition and leaching simulation. In Biome-BGC, atmospheric nitrogen deposition directly enters the soil mineral nitrogen pool, and 10% of the soil mineral nitrogen leaches with the occurrence of water outflow (Thornton and Rosenbloom, 2005). In the coupled model, a surface nitrogen pool is added. Atmospheric nitrogen deposition enters the surface nitrogen pool first, and then either enters the soil mineral nitrogen pool with water infiltration, or moves laterally when overland flow occurs. Soil mineral nitrogen in the subsurface pool also moves laterally with subsurface groundwater flow. The lateral transport of

nitrogen occurs entirely via advection in the aqueous phase. Applying the semi-discrete approach to the advection equation by integrating over a three dimensional control volume gives the change of nitrogen storage at both surface and subsurface (Bao et al., 2017):

$$V_{\text{surf}i} \frac{d}{dt} C_{\text{surf}i} = \sum_j (-mq_{\text{surf}ij} C_{\text{surf}ij}) - N_{\text{infil}} + N_{\text{dep}}, \quad (5a)$$

and

$$V_{\text{sub}i} \frac{d}{dt} (\theta_i C_{\text{sub}i}) = \sum_j (-mq_{\text{sub}ij} C_{\text{sub}ij}) + N_{\text{infil}} + F_{\text{sub}}. \quad (5b)$$

where $V_{\text{surf}i}$ is the total volume of surface water storage of element i , $V_{\text{sub}i}$ is the total volume of subsurface of element i , $C_{\text{surf}i}$ and $C_{\text{sub}i}$ are the mineral nitrogen concentration at the surface and subsurface of element i , respectively, θ_i is the soil water content of element i , $m = 10\%$ is the fraction of mineral nitrogen that is in soluble form and mobile, $q_{\text{surf}ij}$ and $q_{\text{sub}ij}$ are the lateral water flow at the surface and subsurface between element i and its neighbor at edge j , N_{dep} is the rate of atmospheric nitrogen deposition, N_{infil} is the rate of nitrogen entering the subsurface pool via infiltration, and F_{sub} is a source/sink term of soil mineral nitrogen, including mineralization, denitrification (volatilization), immobilization, and plant uptake. When $q_{\text{surf}ij} > 0$, i.e., element i is losing water from edge j , $C_{\text{surf}ij} = C_{\text{surf}i}$; when $q_{\text{surf}ij} < 0$, i.e., element i is gaining water from edge j , $C_{\text{surf}ij}$ is the concentration of surface mineral nitrogen concentration of the neighbor element at edge j . The same applies to subsurface calculation.

The BGC module is integrated into every Flux-PIHM model grid, and all grids are connected by lateral surface and subsurface flow and the nitrogen transport module. Nitrogen transport is also simulated in stream flow, in which case the fraction of soluble and mobile nitrogen $m = 1$. Note that although Flux-PIHM has multiple soil layers, the BGC module has only one bulk soil layer for different soil carbon and nitrogen pools as in Biome-BGC. Therefore, the average subsurface soil mineral nitrogen concentration is used as C_{sub} for subsurface nitrogen transport (Eq. (5)).

Flux-PIHM and the BGC component are coupled by exchanging root zone soil moisture content, total soil moisture storage, top layer soil temperature, surface albedo, aerodynamic resistance, incoming solar radiation, vapor pressure deficit, surface pressure, daily maximum and minimum air temperature, daily average air temperature, daytime and nighttime air temperature, lateral water flow among grids, and LAI between each other (Fig. 1). The nitrogen transport is simulated at every hydrologic time step, while biogeochemical processes are simulated every day.

Flux-PIHM completely replaces the hydrologic and land surface processes in Biome-BGC, including soil moisture, soil temperature, and water outflow calculations. Biome-BGC calculates the soil moisture of one bulk soil layer using a bucket model and uses the 11-day running weighted average of daily average temperature as the soil temperature. When coupled to Flux-PIHM, Flux-PIHM simulated daily root zone soil moisture is used to calculate a daily stomatal conductance for photosynthesis in the BGC module, the total soil moisture is used to calculate the soil nitrogen concentration and constrain soil decomposition, and the top layer soil temperature replaces the empirical soil temperature calculation in the original Biome-BGC.

In Biome-BGC, surface albedo and leaf boundary conductance are set as constants. In the coupled system, Flux-PIHM simulated surface albedo, which reflects the impact of canopy growth and snow accumulation, is transferred to the BGC module to calculate available solar energy for photosynthesis. Flux-PIHM simulated aerodynamic conductance replaces the constant leaf boundary conductance. In addition, Flux-PIHM calculates daily meteorological conditions (i.e., daytime average incoming solar radiation, daily average vapor pressure deficit, daily average surface pressure, daily average, maximum and minimum air temperatures, daytime and nighttime average air temperatures)

from the imposed hourly forcing to be used in the BGC module.

The LAI simulated by the BGC module replaces the prescribed LAI forcing in Flux-PIHM.

The capability of Flux-PIHM to represent the land surface heterogeneities caused by topography at high spatial and temporal resolutions enables Flux-PIHM-BGC to represent the impact of abiotic environmental factors on terrestrial biogeochemical processes. At the same time, the BGC module enables prognostic prediction of vegetation dynamics.

It is computationally expensive for the coupled modeling system to reach equilibrium using the model's native dynamics (Thornton and Rosenbloom, 2005). An accelerated spin-up mode is implemented using the “modified accelerated decomposition” method (Koven et al., 2013) based on Thornton and Rosenbloom (2005), which is used in the CLM to equilibrate the CN model. In the modified accelerated decomposition method, the base decomposition rates for the slow microbial recycling pool and the recalcitrant soil organic matter pool are accelerated by 5 and 70 times, respectively. When exiting the accelerated spin-up mode, the carbon and nitrogen stocks in the slow microbial recycling pool and the recalcitrant soil organic matter pool are multiplied by 5 and 70, respectively, to compensate for the accelerated decomposition rates. The accelerated spin-up can be followed by a normal spin-up, i.e., the native dynamics method (Thornton and Rosenbloom, 2005), for the model to reach equilibrium under the normal decomposition rates.

3. Site and data

3.1. The Shale Hills watershed

The Shale Hills watershed is a first-order forested watershed in the valley and ridge physiographic province of central Pennsylvania (Fig. 2). This watershed is part of the Susquehanna Shale Hills Critical Zone Observatory (SSHCZO), and has been a test site for a suite of PIHM models (e.g., Shi et al., 2013; Yu et al., 2014; Zhang et al., 2016). Shale Hills is a temperate-climate watershed with a mean annual temperature around 10°C and a mean annual precipitation around 107 cm. The watershed is V-shaped with almost true north-facing and south-facing slopes, which are interrupted by seven swales (depressional areas). Five soil series are identified within the watershed (Lin, 2006; Lin and Zhou, 2007). The hill slopes and south-facing ridge top of the watershed are covered by typical deciduous species. The valley floor and north-facing ridge top are covered by some evergreen species (Wubbels, 2010). The broad array of meteorological, hydrological, and ecological observations (e.g., discharge, soil moisture, soil temperature, groundwater level, snow depth, leaf area index, eddy covariance measurements) and extensive multidisciplinary field surveys and campaigns (e.g., soil survey, tree survey, lidar surface elevation measurements (Guo, 2010), bedrock depth measurements) provide important input and evaluation data for Flux-PIHM-BGC.

Observations used to evaluate the model predictions include aboveground carbon and soil carbon pool measurements. The diameters-at-breast-height (DBH) of all trees with DBH over 20 cm at the Shale Hills watershed were measured in 2008 and 2012. Smith et al. (2017) calculated the aboveground vegetation carbon for each tree based on empirical allometric equations (Jenkins et al., 2003), which convert tree DBH to biomass. Aboveground carbon was assumed to be 45% of tree biomass. Andrews et al. (2011) measured soil organic carbon (SOC) concentration and soil bulk density at different soil horizons in the top 1.1 m soil column at 56 locations within the Shale Hills watershed. Soil carbon storage is calculated using (Andrews et al., 2011)

$$\text{SOC storage} = \text{SOC concentration} \cdot \rho_b \cdot z, \quad (6)$$

where ρ_b is the soil bulk density and z is the thickness of each soil horizon. The SOC storage values at different soil horizons are summed up to get the total soil carbon storage within the soil column.

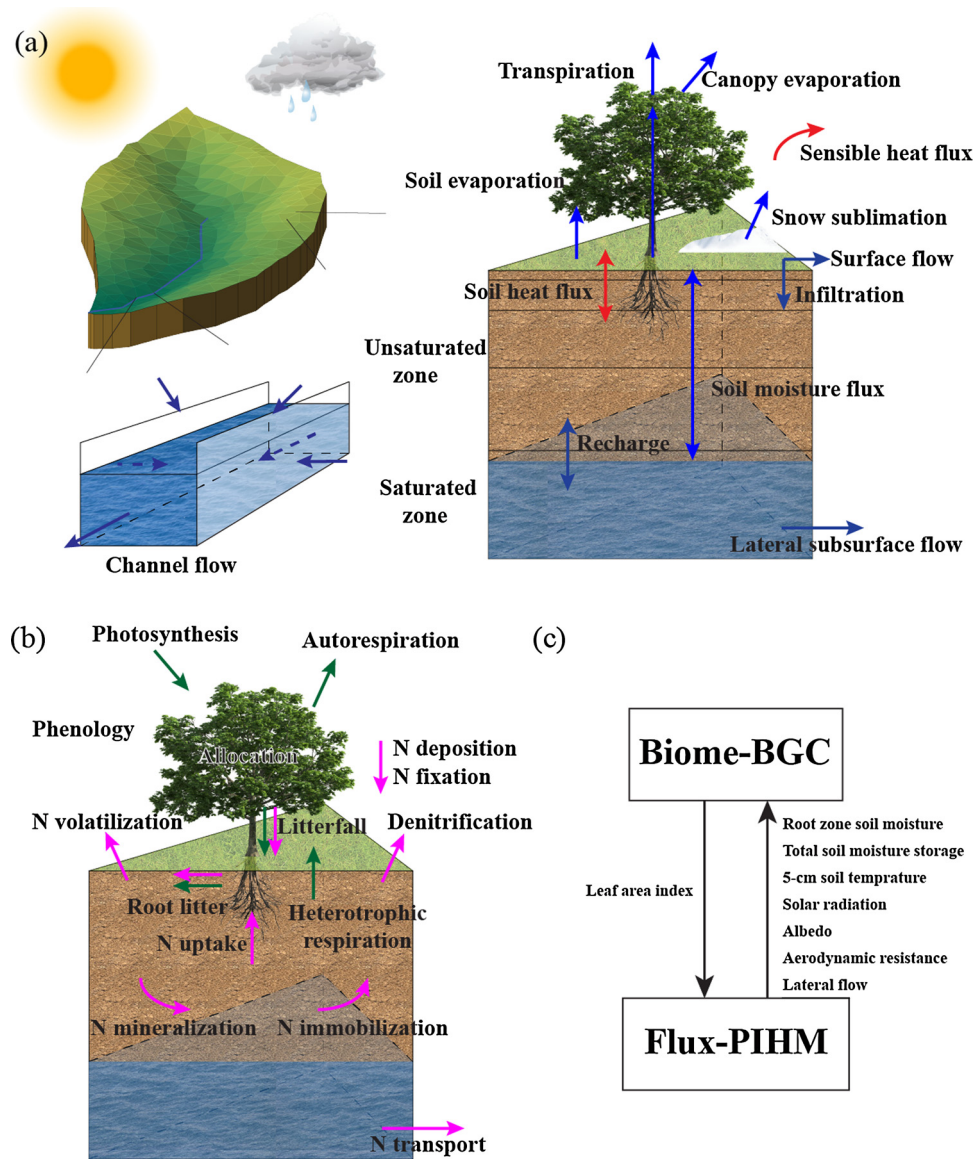


Fig. 1. Schematic illustration of (a) land surface and hydrologic processes simulated by Flux-PIHM, (b) carbon and nitrogen processes simulated by the BGC module (adapted from Biome-BGC), and (c) the coupling between Flux-PIHM and the BGC module.

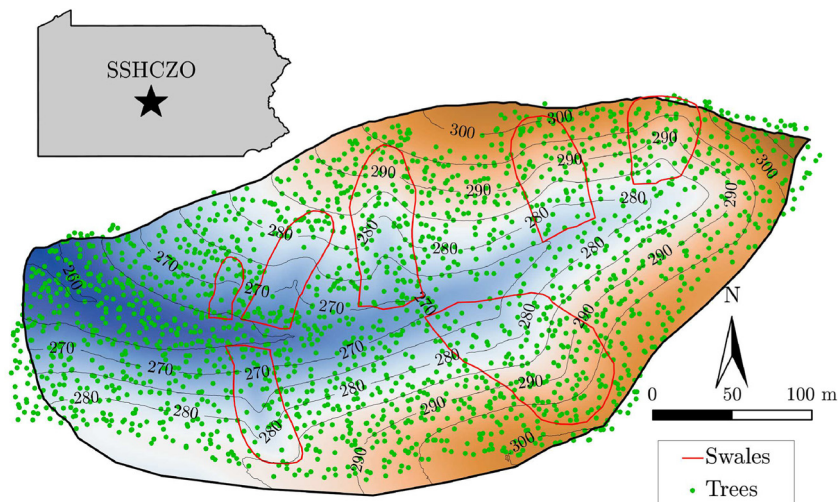


Fig. 2. The Shale Hills watershed with locations of the seven swales and all trees sampled. The inset shows the location of SSHCZO within Pennsylvania.

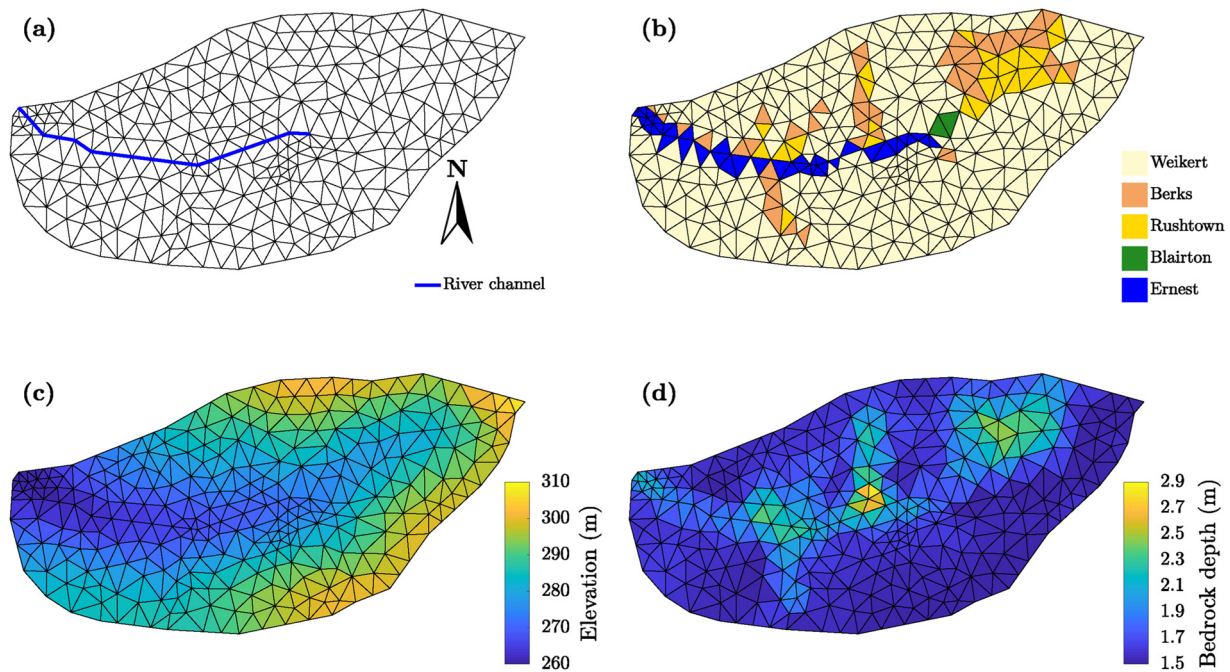


Fig. 3. The (a) grid and stream channel setting, (b) soil type, (c) surface elevation, and (d) bedrock depth defined in Flux-PIHM-BGC for the Shale Hills simulation. The bedrock depth shown in (d) is the observed bedrock depth plus an additional 1.25-m weathered rock layer.

3.2. Model setup and parameters

The Shale Hills watershed model domain is decomposed into a triangular network of 532 grids, and the river channel is represented by 19 rectangular segments (Fig. 3a). An airborne light detection and ranging (lidar) measured surface terrain map with 1-m horizontal resolution (Guo, 2010) is used to describe surface topography of the domain and determine the watershed boundary. A field campaign conducted in 2003 provides soil map and bedrock depth map for the watershed. To account for water flow through a deeper weathered shale layer below the measured bedrock depth, an extra 1.25 m is added to the measured bedrock depths for the whole watershed (Shi et al., 2015a). The surface terrain map, soil map, and bedrock depth map are projected to the model domain to describe the characteristics of each model grid, and are presented in Fig. 3b–d. Detailed descriptions of the geospatial data can be found in Shi et al. (2013, 2015a). For the sake of simplification, all model grids are assumed to be covered by deciduous trees. A small area of coniferous trees is located at the outlet of the watershed, but they have been shown to have relatively little impact on the soil hydrology (Shi et al., 2015a) and their biogeochemistry is very similar to deciduous trees. Since our focus is to examine interactions between the hydrology and vegetation, we have chosen to simplify the landscape with uniform forest type.

Flux-PIHM-BGC simulations require forcing time series of precipitation, air temperature, relative humidity, wind speed, surface pressure, downward longwave radiation, and direct and diffuse solar radiation (for topographic solar radiation). Precipitation, air temperature, and relative humidity data are obtained from the weather station located on the south-facing ridge of Shale Hills. Wind speed and surface pressure are measured at the flux tower near the weather station. Downward longwave radiation, and direct and diffuse solar radiation data are not available at SSHCZO for the study period and are obtained from the Surface Radiation Budget Network (SURFRAD) Penn State University station, which is 6.48 km away from the Shale Hills watershed. Gaps in the *in situ* precipitation, air temperature, relative humidity, wind speed, surface pressure and longwave radiation data are filled using the forcing data for Phase 2 of the North American Land Data Assimilation system (NLDAS-2) (Xia et al., 2012). To fill the gaps

in the direct and diffuse solar radiation data, we first derive a linear relationship (Tian et al., 2001) using available data:

$$\frac{S_{\text{dif}}}{S_{\downarrow}} = k \frac{S_{\downarrow}}{S_0} + b, \quad (7)$$

where $S_{\downarrow}$ is the downward solar radiation received (sum of direct and diffuse), S_0 is the extraterrestrial solar radiation, and k and b are fitting parameters. The extraterrestrial solar radiation S_0 is calculated using the National Renewable Energy Laboratory Renewable Resource Data Center's solar intensity calculator (<https://www.nrel.gov/midc/solpos/solpos.html>). Gaps in diffuse solar radiation are then filled by applying Eq. (7) to NLDAS-2 downward solar radiation forcing. Gaps in direct solar radiation are filled using:

$$S_{\text{dir}} = \frac{S_{\downarrow} - S_{\text{dif}}}{\cos \varphi}. \quad (8)$$

The soil hydraulic properties measured by Lin (2006) and Baldwin (2011) are used for the study. The Noah LSM vegetation parameters (Chen and Dudhia, 2001) are used to describe the physical properties (e.g., rooting depth, emissivity, and albedo) of the deciduous broadleaf forest. The ecophysiological parameters for deciduous trees are provided by the default Biome-BGC input file. The default annual whole-plant mortality fraction (0.005 year^{-1}) is replaced by the measured whole-plant mortality fraction, which is calculated as the ratio of annual aboveground productivity (based on the difference in aboveground carbon between 2008 and 2012) over aboveground biomass. The annual nitrogen deposition rate is obtained from the National Atmospheric Deposition Program (NADP) Leading Ridge site, which is about 3 km west to Shale Hills. In this study we use the nitrogen deposition rate from 2012. All ecophysiological parameters used in the simulations are listed in Table 1.

The experimental period is from 0000 UTC 1 January 2009 to 0000 UTC 1 January 2012. A 600-year accelerated spin-up using the modified accelerated decomposition method is performed for the coupled system to reach approximate steady state, and is accomplished by recycling the 3-year meteorological forcing. The accelerated spin-up is followed by a 600-year normal spin-up using native dynamics to reach equilibrium. To further accelerate the spin-up process, coarser time

Table 1
Ecophysiological parameters for deciduous trees.

Parameter	Unit	Value
Annual leaf and fine root turnover fraction	year ⁻¹	1.0
Annual live wood turnover fraction	year ⁻¹	0.7
Annual whole-plant mortality fraction	year ⁻¹	0.045
Annual fire mortality fraction	year ⁻¹	0.0025
Allocation new fine root C: new leaf C	Unitless	1.0
Allocation new stem C: new leaf C	Unitless	2.2
Allocation new live wood C: new total wood C	Unitless	0.1
Allocation new coarse root C: new stem C	Unitless	0.23
Allocation current growth proportion	Unitless	0.5
C:N of leaves	kgC kgN ⁻¹	24.0
C:N of leaf litter, after retranslocation	kgC kgN ⁻¹	49.0
C:N of fine roots	kgC kgN ⁻¹	42.0
C:N of live wood	kgC kgN ⁻¹	50.0
C:N of dead wood	kgC kgN ⁻¹	442.0
Leaf litter labile proportion	Unitless	0.39
Leaf litter cellulose proportion	Unitless	0.44
Leaf litter lignin proportion	Unitless	0.17
Fine root labile proportion	Unitless	0.30
Fine root cellulose proportion	Unitless	0.45
Fine root lignin proportion	Unitless	0.25
Dead wood cellulose proportion	Unitless	0.76
Dead wood lignin proportion	Unitless	0.24
Canopy light extinction coefficient	Unitless	0.70
All-sided to projected leaf area ratio	Unitless	2.0
Canopy average specific leaf area (SLA)	m ² kgC ⁻¹	30.0
Ratio of shaded SLA: sunlit SLA	Unitless	2.0
Fraction of leaf N in Rubisco	Unitless	0.08
Maximum stomatal conductance	m s ⁻¹	0.01
Cuticular conductance	m s ⁻¹	0.0002
Leaf water potential at start of conductance reduction	MPa	-0.6
Leaf water potential at complete conductance reduction	MPa	-2.3
Vapor pressure deficit at start of conductance	Pa	930.0
Vapor pressure deficit at complete conductance	Pa	4100.0
Wet and dry atmospheric deposition of nitrogen	g N m ⁻² year ⁻¹	0.75

steps for hydrologic, nitrogen transport, and land surface processes are used. During both the accelerated and normal spin-up processes, hydrologic and nitrogen transport simulation time step is set to 10 min, and the land surface simulation time step is set to 1 h. The biogeochemical process time step is one day (24 h). In the accelerated spin-up, the total watershed soil carbon pool reaches steady states, i.e., the change of watershed average soil carbon over time is lower than 0.5 g m⁻² year⁻¹ (Thornton and Rosenbloom, 2005), after 369 years for Flux-PIHM-BGC. In the normal spin-up, it takes nine years for the coupled system to reach a new steady state under the normal decomposition rates. A three-year, coupled Flux-PIHM-BGC simulation is then performed, with a hydrologic and transport simulation time step of 1 min and a land surface simulation time step of 15 min. The biogeochemical processes are still simulated on a daily time step.

To demonstrate the effect of the nitrogen transport module, a Flux-PIHM-BGC simulation without the nitrogen transport module is executed. In this simulation, the nitrogen transport module is replaced by a simple nitrogen leaching module: nitrogen deposition directly enters the subsurface (soil mineral nitrogen pool) and nitrogen leaching occurs only if there is subsurface outflow. In each model grid, when there is subsurface outflow ($\sum_j q_{subij} > 0$), the leaching loss is a function of excess soil mineral nitrogen concentration (with 10% of the soil mineral nitrogen is assumed to be soluble) and subsurface outflow rate. The leaching loss is removed from the soil mineral nitrogen pool, but does not enter the downstream grids, i.e.,

$$V_{subi} \frac{d}{dt} (\theta_i C_{subi}) = \begin{cases} N_{dep} + F_{sub}, & \sum_j q_{subij} \leq 0, \\ -mC_{subi} \sum_j q_{subij} + N_{dep} + F_{sub}, & \sum_j q_{subij} > 0. \end{cases} \quad (9)$$

The Flux-PIHM-BGC without nitrogen transport model is also equilibrated using a 600-year accelerated spin-up, followed by a 600-

year normal spin-up.

4. Results

Model results from the three-year coupled Flux-PIHM-BGC simulation, after the 1200-year spinup, is presented below, to study the spatial variation of carbon stocks and fluxes at the Shale Hills watershed.

The spatial patterns of Flux-PIHM-BGC simulated available root zone soil moisture fraction (fraction between soil wilting point and soil porosity), soil moisture storage, incoming solar radiation, and 5 cm soil temperature, averaged over the simulation period are presented in Fig. 4. Note that the figures only show predictions from the triangular grids, but not the river channels. The valley floor and swales are wetter than the planar slopes, when measured using both available root zone soil moisture fraction and soil moisture storage (Fig. 4a and b). This is mainly caused by the differences in soil hydraulic properties and topography driven groundwater flow (Shi et al., 2015a). The north-facing planar slope is slightly wetter than the south-facing planar slope, especially in the root zone (Fig. 4a and b), because of the lower evapotranspiration rate on this slope. As for incoming solar radiation, the south-facing slope and the north-facing ridge top locations receive more solar radiation than the north-facing slope (Fig. 4c). On average, the south-facing slope receives almost 1 kWh m⁻² day⁻¹ more solar radiation than the north-facing slope. The swales also see less incoming solar radiation compared with the planar slopes. The simulated pattern of 5 cm soil temperature (Fig. 4d) is impacted by both soil moisture and incoming solar radiation. Soil temperature is lower near the stream and in the swales (wetter) than on the planar slopes (drier) due to the influence of soil moisture. Average soil temperature is about 1 °C lower on the north-facing slope than the south-facing slope due to the differences in incoming solar radiation.

Fig. 5 illustrates the average patterns of gross primary production (GPP), net primary production (NPP), net ecosystem production (NEP), and heterotrophic respiration (Rh) simulated by Flux-PIHM-BGC. Simulated watershed average GPP, NPP, NEP, and Rh values for each year are presented in Table 2. Total precipitation, average air temperature, and simulated average solar radiation are also shown in Table 2.

The spatial patterns of GPP and NPP are similar (Fig. 5a and b), showing higher productivity on the south-facing slope, and higher productivity in the swales and on the valley floor. GPP and NPP are the highest in 2010, which has the highest temperature and solar radiation among the three years. The whole watershed is a carbon sink in 2009 and 2010, but becomes a carbon source in 2011 (Table 2). The year 2011 is an extremely wet year (Table 2), resulting in high decomposition rate which outweighs NPP and turns the forest into a carbon source. When averaged over the whole simulation period, the watershed is almost at equilibrium due to the spin-up process (Fig. 5c). Swales and the valley floor show higher NEP compared to the planar slopes. Heterotrophic respiration is stronger on the warmer south-facing slope, and in the wetter areas, i.e., on the valley floor and in the swales (Fig. 5d).

Fig. 6 presents the observed spatial patterns of aboveground carbon and soil organic carbon at the Shale Hills watershed. Aboveground carbon density is calculated for each Flux-PIHM-BGC grid, by summing the observed aboveground carbon storage over all trees within that model grid and dividing by the grid area. Fig. 6 shows that the south-facing slope and swales have higher aboveground carbon density than the north-facing slope and planar sites. The watershed average aboveground carbon at Shale Hills estimated by Smith et al. (2017) is 9.5 kg C m⁻². A similar spatial pattern is found for the soil carbon distribution. South-facing slopes and swales have higher soil carbon storage than north-facing slopes and planar sites. The observed average soil carbon storage is about 13 kg C m⁻².

Fig. 7a–d illustrates the spatial patterns of aboveground vegetation carbon, soil carbon, litter carbon, and soil mineral nitrogen predicted by Flux-PIHM-BGC averaged over the three-year simulation period.

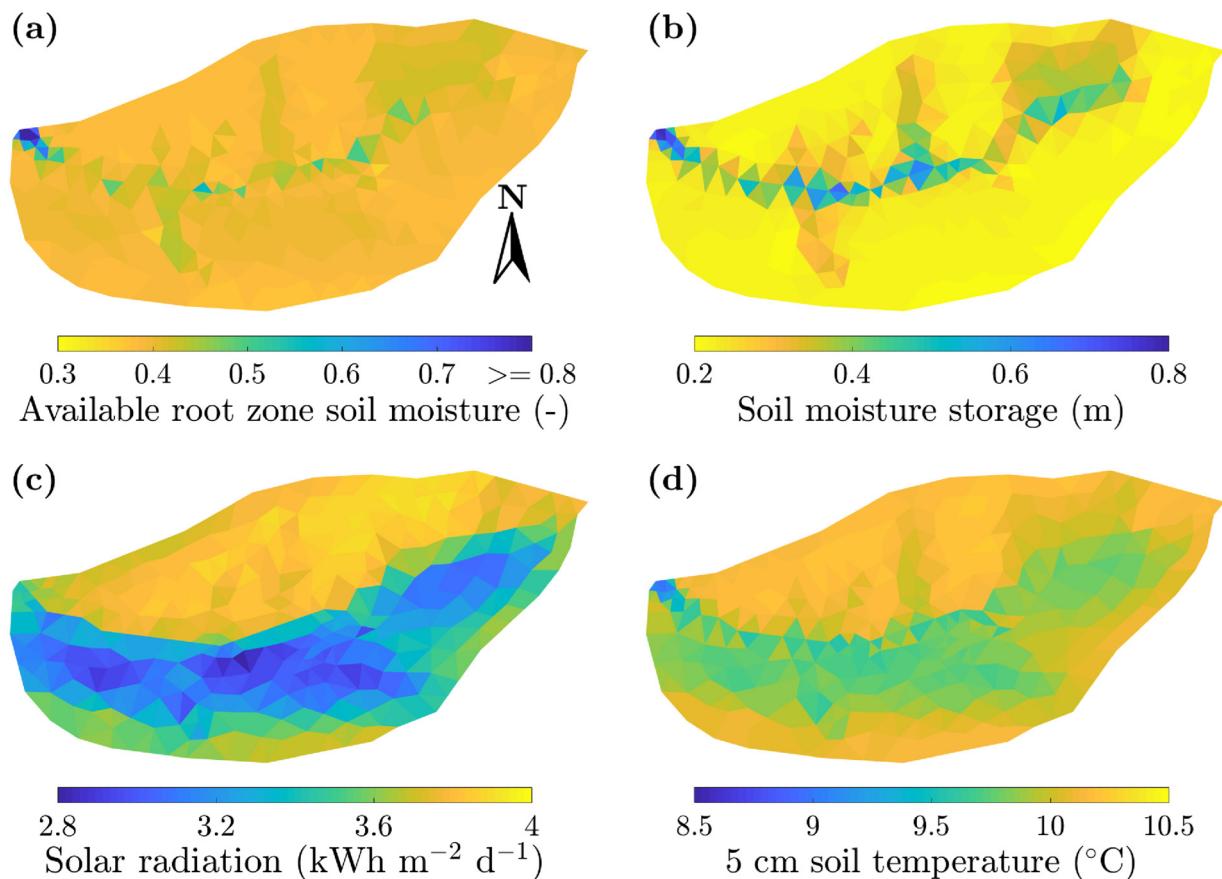


Fig. 4. (a) Average available root zone soil moisture (i.e., fraction between soil wilting point and soil porosity), (b) total soil moisture storage, (c) incoming solar radiation, and (d) 5 cm soil temperature patterns at Shale Hills simulated by Flux-PIHM-BGC. The spatial patterns are averaged over the simulation period (three years).

Flux-PIHM-BGC simulated distributions of aboveground carbon, soil carbon, litter carbon and soil mineral nitrogen are similar. Carbon and nitrogen storage is generally higher on the south-facing slope than the north-facing slope, because the south-facing slope is more favorable for plant growth due to its higher incoming solar radiation and higher soil temperature. On each slope, carbon and nitrogen storage is higher near the stream and in the swales, which are generally wetter than the planar slopes. The predicted watershed average aboveground vegetation carbon is 8.82 kg C m^{-2} . The predicted watershed average soil carbon is $13.38 \text{ kg C m}^{-2}$.

The comparison between Figs. 6 and 7 shows that Flux-PIHM-BGC is capable of reproducing the observed macro patterns of aboveground vegetation carbon and soil carbon at the Shale Hills watershed. The predicted watershed average aboveground vegetation carbon (8.82 kg C m^{-2}) and average soil carbon ($13.38 \text{ kg C m}^{-2}$) are close to the observed values (9.5 kg C m^{-2} and 13 kg C m^{-2} , respectively). But it is clear that the observed carbon stocks exhibit greater spatial variability than the simulation. It should be noted that the ecophysiological parameters are not calibrated.

Fig. 7e and f presents the average aboveground vegetation carbon and soil mineral nitrogen patterns predicted by Flux-PIHM-BGC without the nitrogen transport module. The simulated average soil carbon pattern is similar to the aboveground vegetation carbon and is not shown. When the nitrogen transport module is removed, the simulated spatial patterns of carbon and nitrogen stocks (Fig. 7e and f) are almost the opposite of the patterns simulated by the full Flux-PIHM-BGC model (Fig. 7a and d). The predicted watershed average aboveground vegetation carbon (8.21 kg C m^{-2}) in this simulation is about 7% lower than the full Flux-PIHM-BGC simulation. The Flux-PIHM-BGC model with the simple nitrogen leaching scheme fails to capture the observed

macro spatial pattern of aboveground carbon at the Shale Hills watershed (Fig. 6). Although simulated aboveground vegetation carbon and soil mineral nitrogen are still higher on the south-facing slope than the north-facing slope, they are lower in the swales and near the stream than on the planar slopes. The model grids in the swales and near the stream have both high water inflow and high water outflow, thus higher nitrogen leaching rates compared with the grids on the planar slopes. When the nitrogen transport module is replaced by the simple nitrogen leaching scheme, the nitrogen inflow from upslope is ignored, which leads to lower soil mineral nitrogen due to strong nitrogen leaching.

Flux-PIHM-BGC simulated spatial patterns of leaf area index and leaf onset date are depicted in Fig. 8a and b. In Biome-BGC, leaves of deciduous trees grow until the beginning of leaf senescence, thus the LAI reaches its maximum at the end of the growing season. The LAI pattern presented in Fig. 8a is for 21 October, 2010, on which day the watershed average LAI reaches its maximum. The leaf onset date pattern (Fig. 8b) is an average over three years. Because the BGC module's carbon allocation scheme uses a set of fixed fractions for all plant structures (Table 1), the spatial pattern of LAI (Fig. 8a) is consistent with the spatial pattern of vegetation carbon (Fig. 7a), with higher LAI in the swales and near the stream than on the planar slopes, and higher LAI on the south-facing slope than the north-facing slope. The spatial pattern of leaf onset date (Fig. 8b) is determined by the spatial pattern of soil temperature (Fig. 4d) because leaf onset is controlled by the degree-day summation of soil temperature. The earliest leaf onset and latest leaf onset in different areas of the watershed are as far as 13 days apart. When averaged over the three simulation years, the earliest leaf onset is 21 April on the warmest south-facing planar slope, and the latest is 4 May in the coolest stream mouth. Such spatial variability in

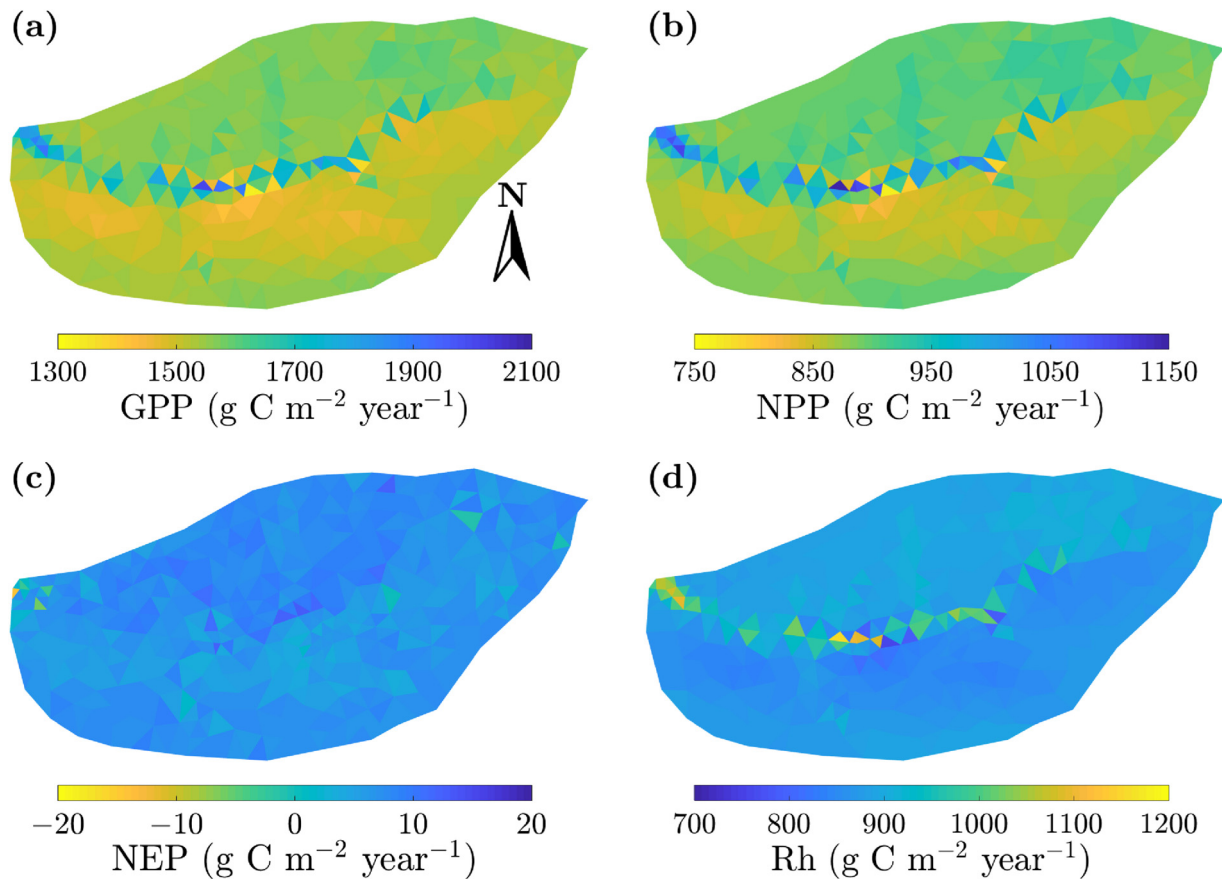


Fig. 5. Average gross primary production (GPP), net primary production (NPP), net ecosystem production (NEP), and heterotrophic respiration (Rh) patterns at Shale Hills simulated by Flux-PIHM-BGC.

maximum LAI and phenology will produce different evapotranspiration patterns and affect land surface and hydrologic conditions.

To demonstrate the effect of distributed LAI, we perform a Flux-PIHM simulation (not coupled to the BGC module) driven by the Flux-PIHM-BGC predicted watershed average LAI to be compared with the Flux-PIHM-BGC simulation. The Flux-PIHM simulation uses the same land surface and hydrologic initial conditions as the Flux-PIHM-BGC simulation. Flux-PIHM and Flux-PIHM-BGC have exactly the same formulations for land surface and hydrologic simulation. The only difference is that Flux-PIHM is driven by spatially uniform LAI, which is the LAI predicted by Flux-PIHM-BGC averaged over the whole watershed, and Flux-PIHM-BGC has spatially distributed diagnostic LAI predicted by the biogeochemistry module. The spatial patterns of transpiration on 27 April, 2010 predicted by both models are presented in Fig. 8c and d to illustrate the differences in spatially distributed LAI and uniform LAI. The Flux-PIHM simulation using spatially uniform LAI yields a more homogeneous transpiration pattern (Fig. 8d), whereas the Flux-PIHM-BGC simulation produces a stronger contrast between the north- and south-facing slopes (Fig. 8c). The largest difference appears near the watershed outlet. In the Flux-PIHM prediction, the watershed outlet has the highest transpiration rate in the watershed. In the Flux-

PIHM-BGC prediction, however, the watershed outlet has almost zero transpiration because leaf onset has not begun yet. Because both models have the same watershed average LAI at each time step, the watershed-integrated transpiration does not differ much.

To identify the important abiotic driving factors of vegetation carbon distribution at the Shale Hills watershed, we plot annual average aboveground vegetation carbon at each model grid as functions of available root zone soil moisture fraction, top layer soil temperature, incoming solar radiation, and soil mineral nitrogen (Fig. 9a–d). Root zone soil moisture fraction, soil temperature, solar radiation and soil mineral nitrogen represent the impacts of water, energy, and nutrient availability on vegetation carbon. A random forest of 200 regression trees is used to estimate the importance of the abiotic factors for the prediction of vegetation carbon distribution. The out-of-bag (OOB) permuted predictor errors are used as the estimates of predictor importance (Breiman, 2001), and are presented in Fig. 9e. As shown in Fig. 9d, the distribution of aboveground vegetation carbon has a strong linear relationship with soil mineral nitrogen. Soil mineral nitrogen is suggested as the most important driver of vegetation carbon distribution by the random forest analysis. Soil moisture, soil temperature and solar radiation are about equally important in determining the spatial

Table 2

Annual total precipitation, average air temperature, Flux-PIHM-BGC simulated average solar radiation, GPP, NPP, NEP, and Rh.

Year	Precipitation	Air temperature	Solar radiation	GPP	NPP	NEP	Rh
–	(m)	(°C)	(kWh m ⁻² day ⁻¹)	(g C m ⁻² year ⁻¹)			
2009	1.05	9.46	3.43	1530.9	909.1	52.3	856.7
2010	0.90	10.62	3.72	1624.1	922.1	44.1	877.9
2011	1.32	10.50	3.35	1496.4	845.6	–76.0	921.6

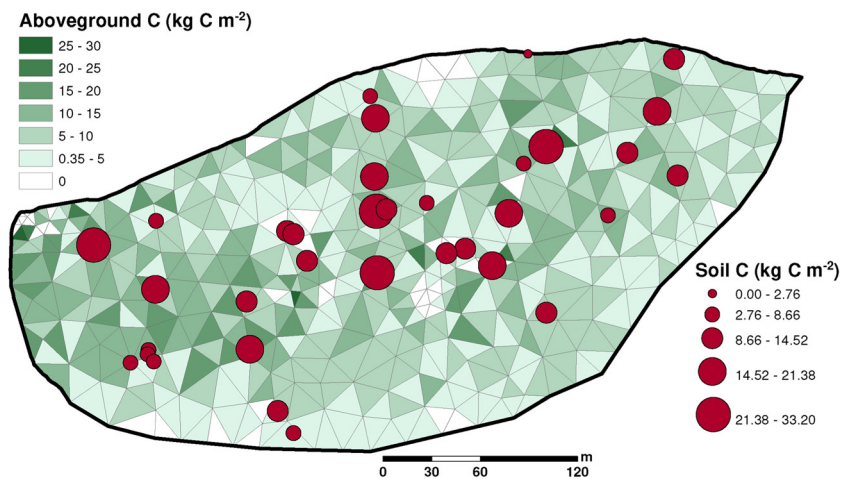


Fig. 6. Observed spatial patterns of aboveground carbon [Smith et al. \(2017\)](#) and soil organic carbon ([Andrews et al., 2011](#)) at the Shale Hills watershed. The observed aboveground carbon density is presented for every Flux-PIHM-BGC grid using different shades of green, and the red data points represent soil organic carbon storage. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

pattern of aboveground vegetation biomass ([Fig. 9e](#)), with soil moisture showing slightly higher importance. Soil moisture and solar radiation show some linearity with aboveground carbon in many areas of the watershed ([Fig. 9a](#) and [c](#)). The relationship between aboveground carbon and soil temperature is not clear.

5. Discussion and conclusions

This paper presents a coupled land surface hydrologic terrestrial biogeochemistry modeling system with a nitrogen transport module, Flux-PIHM-BGC, based on a high fidelity and high spatial resolution land surface hydrologic model. Flux-PIHM-BGC adds the simulation of carbon and nitrogen cycles to Flux-PIHM. Owing to the nitrogen transport module and topographic radiation module, Flux-PIHM-BGC is a truly spatially distributed terrestrial biogeochemistry modeling system. Flux-PIHM-BGC is able to simulate the complex interaction between water, energy, nutrient, and carbon in time and space, which

cannot be achieved without the coupling of the hydrologic, land surface, and biogeochemical model components.

Model results show that carbon stocks and fluxes are strongly affected by topography and landscape positions. When averaged over the entire simulation period, the areas near the stream and in the swales, which are concave and wetter, yield higher productivity (GPP, NPP, and NEP), higher heterotrophic respiration ([Fig. 5](#)), more vegetation, soil, and litter carbon, and more soil mineral nitrogen storage ([Fig. 7](#)). The south-facing slope, which has higher incoming solar radiation, has higher GPP and NPP, lower NEP, higher heterotrophic respiration, more vegetation, soil, and litter carbon, and more soil mineral nitrogen storage. Flux-PIHM-BGC is able to capture the observed macro patterns of aboveground vegetation carbon and soil carbon at the Shale Hills watershed. The ability of Flux-PIHM-BGC to predict the differences in biomass between the north- and south-facing slopes is critical because vegetation cover is a key variable in influencing the asymmetry of critical zone processes on equator- and pole-facing slopes ([Pelletier](#)

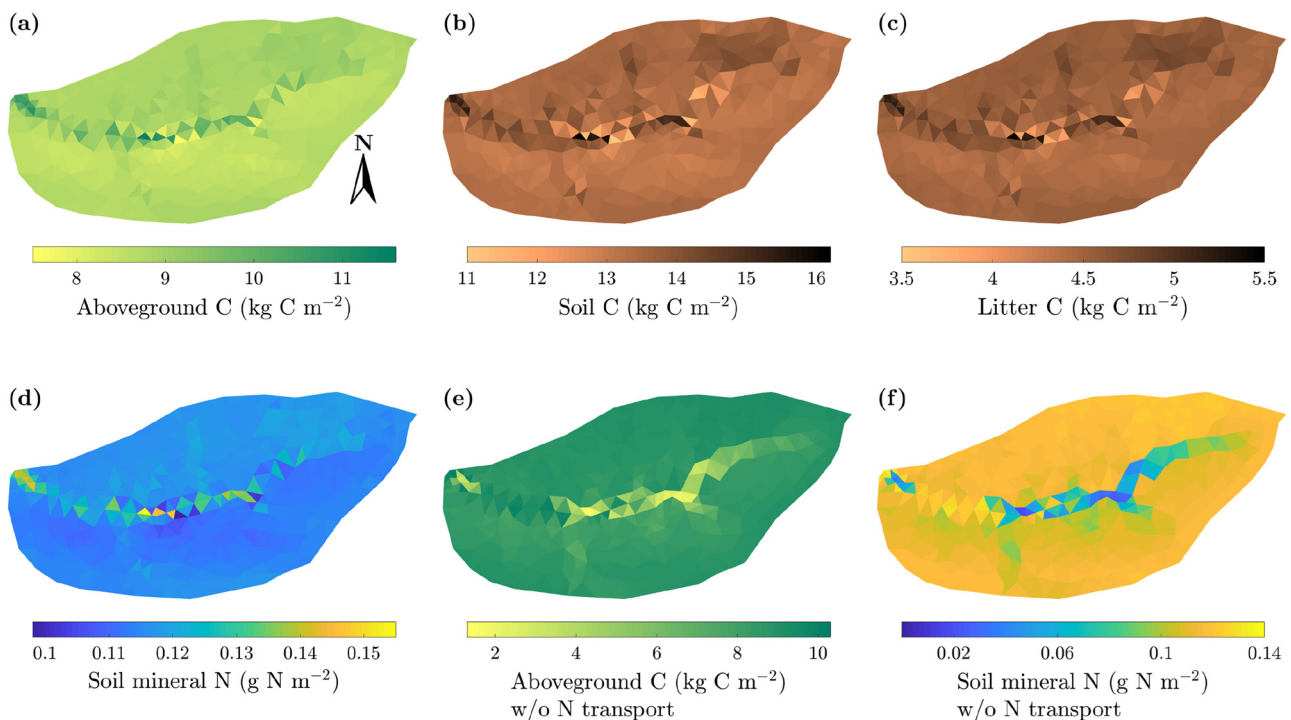


Fig. 7. Average (a) aboveground vegetation carbon, (b) soil carbon, (c) litter carbon, and (d) soil mineral nitrogen spatial patterns at Shale Hills simulated by Flux-PIHM-BGC. Average (e) aboveground vegetation carbon and (f) soil mineral nitrogen spatial patterns simulated by Flux-PIHM-BGC without the nitrogen transport module are also presented.

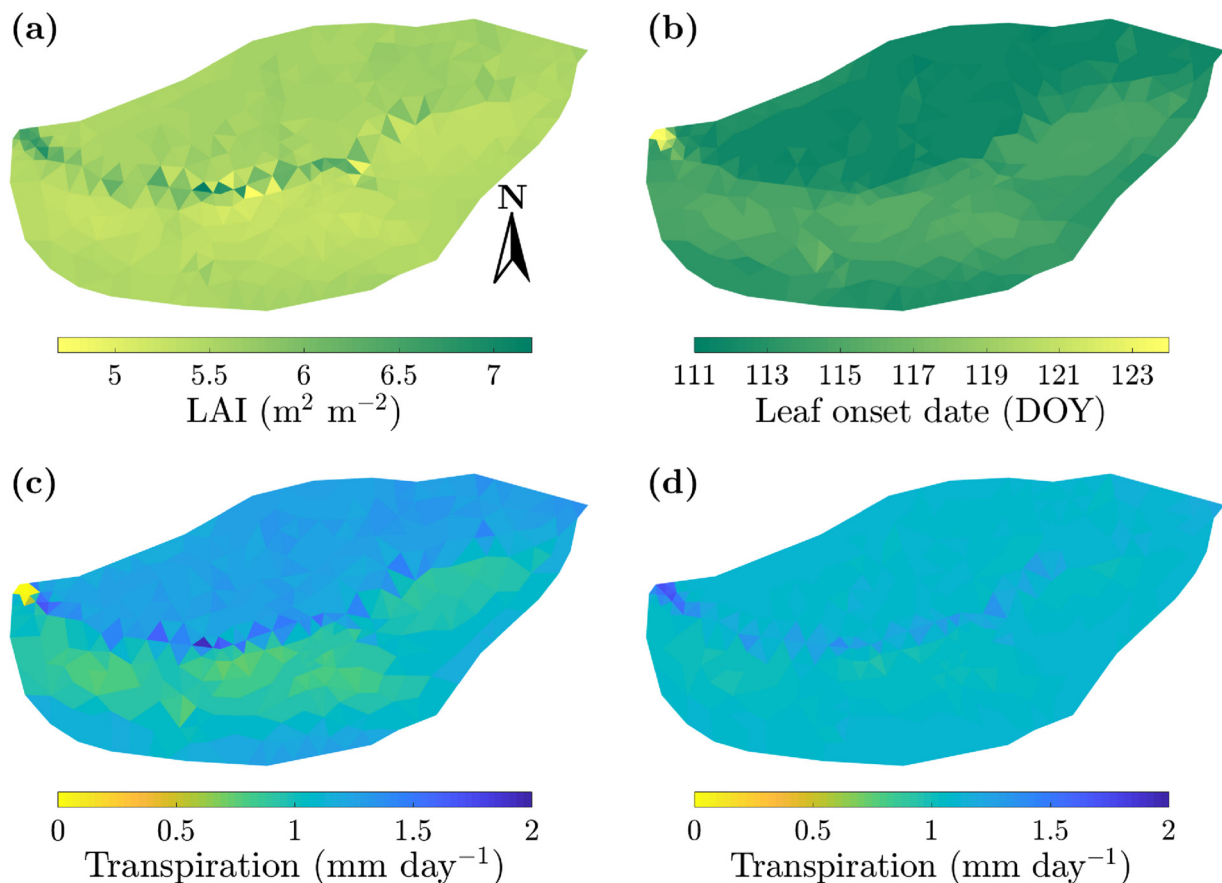


Fig. 8. Spatial patterns of (a) leaf area index on 21 October, 2010 and (b) average leaf onset date over the simulation period at Shale Hills, as simulated by Flux-PIHM-BGC. Spatial patterns of transpiration on 27 April, 2010 predicted by Flux-PIHM-BGC and Flux-PIHM (using spatially uniform leaf area index) are presented in (c) and (d).

et al., 2018).

The predicted watershed average aboveground carbon and soil carbon are close to the observed values. It should be noted that the ecophysiological parameters are not optimized in this study. Ecological observations including aboveground vegetation carbon (Smith et al., 2017), litter carbon (Smith et al., 2017), soil organic carbon (Andrews et al., 2011), soil efflux, and net ecosystem exchange (Shi et al., 2013) would enable future studies to calibrate the ecophysiological parameters. The data assimilation system, which has been developed for Flux-PIHM for parameter and state estimation (Shi et al., 2014b, 2015b), could be used to optimize Flux-PIHM-BGC model parameters.

Although the model is able to capture the macro patterns and watershed average values of carbon stocks, it significantly underestimates the spatial variability in carbon stocks. Fig. 6 illustrates a high degree of spatial variability in above and below ground carbon stocks that is not captured by our modeling system. We note that the spatial resolutions of our observations are fine enough to capture the relatively random processes of individual tree mortality and recruitment. Our modeling system is not designed to simulate these stochastic processes. It is also possible that the spatial variability in the observations reflects random error in the measurements. Quantification of the measurement error at these resolutions would be valuable, but is not currently available. In either case, it is unlikely that our modeling system will capture these variations.

There are additional processes that might explain some of this variability, but are not captured properly in our modeling system. Some ecosystem processes important for simulating soil microbial respiration, e.g., microbial enzyme processes (Zhang et al., 2014), are not included in Flux-PIHM-BGC. Smith et al. (2017) found litter of some of the tree

species at Shale Hills moved horizontally after falling from the canopy. Leaf litter transport downslope or by wind could affect the distribution of vegetation carbon and soil carbon at the Shale Hills watershed. The horizontal movement of litter, however, is not simulated by Flux-PIHM-BGC. This may be one reason that the observed carbon stock in the swales and near the valley floor, which can act as traps of litter, can have much greater soil carbon than simulated in Flux-PIHM-BGC. In addition, differences in canopy heights are not taken into account when calculating topographic solar radiation. This could further underestimate the canopy productivity on the valley floor because the trees grow taller on the valley floor than on the ridge and planar slopes (Gaines et al., 2015), and should receive more solar radiation than predicted. These limitations may lead to the underestimation of the spatial variability in carbon stocks, and highlight the areas of potential future improvements in the modeling system.

The coupled land surface hydrologic terrestrial biogeochemistry modeling system provides an important means to study the interaction among water, energy, carbon, and nitrogen cycles that cannot be achieved without the coupling of these model components. Model results show that hot spots of water, vegetation carbon, soil carbon, litter carbon, and soil mineral nitrogen are collocated in the watershed (Fig. 7), which reveals the interaction among these variables. Because of the topographic gradient and differences in soil hydraulic properties, water and nitrogen accumulate near the stream and in the swales, producing favorable conditions for plant growth. Higher plant productivity leads to more vegetation carbon storage, as well as more litter fall, i.e., more litter carbon storage. When decomposed, litter carbon enters into the soil and becomes soil organic carbon. Higher soil organic carbon and higher decomposition rate (due to the higher soil moisture)

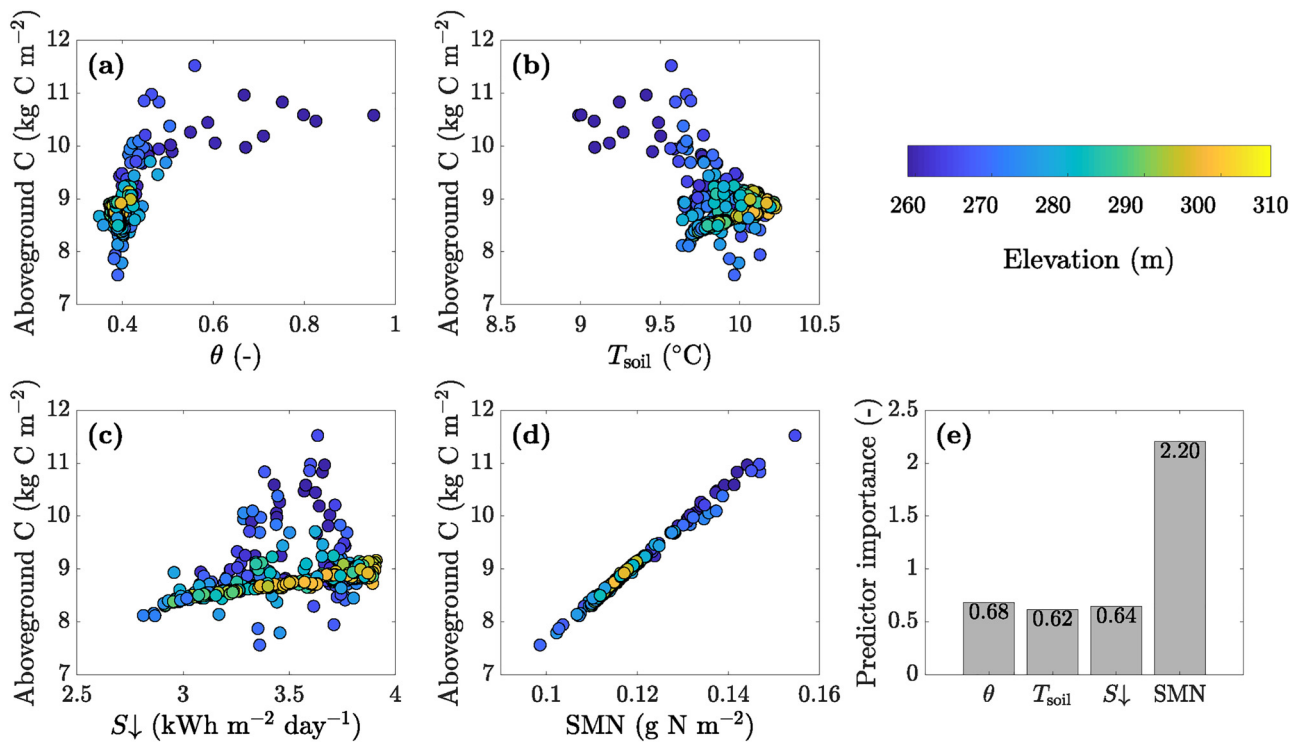


Fig. 9. Average aboveground vegetation carbon versus (a) available root zone soil moisture fraction, (b) 5-cm soil temperature, (c) incoming solar radiation, and (d) soil mineral nitrogen, as predicted by Flux-PIHM-BGC. Each point on the figure represents one model grid. The colors represent surface elevation to indicate landscape positions. The random forest estimated importance of these abiotic variables in predicting aboveground vegetation carbon distribution is presented in (e). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

yield more nitrogen mineralization, which provides more nitrogen for plant growth. Therefore, swales and the areas near the stream become hot spots of carbon and nitrogen storage, owing to the positive feedback among water, vegetation carbon, soil carbon, litter carbon, and soil mineral nitrogen.

Soil mineral nitrogen is ranked as the most important driver of aboveground carbon (Fig. 9e), and the predicted vegetation carbon and soil mineral nitrogen exhibit a strong linear relationship in the watershed (Fig. 9d). It suggests tree growth at the Shale Hills watershed is nitrogen limited. The comparison between the full Flux-PIHM-BGC simulation and the simulation without nitrogen transport (Fig. 7) also suggests that the distribution of carbon stocks is primarily controlled by the movement of nitrogen. Soil mineral nitrogen transport is strongly affected by topographically driven subsurface water flux. The second important driver of vegetation carbon, soil moisture, is affected by landscape position, slope, and aspect. Therefore, the model results clearly demonstrate the impact of topography on watershed carbon and nitrogen processes.

When the nitrogen transport module (Eq. (5)) is replaced by the simple nitrogen leaching scheme (Eq. (9)), the model fails to capture the macro pattern of carbon stocks, and the predicted patterns are almost the opposite of the patterns predicted by the full model, which has the nitrogen transport module. This simulation yields lower soil mineral nitrogen storage in the watershed. As a result, the carbon stocks are also lower than the full Flux-PIHM-BGC run, because the carbon stocks are strongly correlated with soil mineral nitrogen storage (Fig. 9). These results indicate that spatially distributed biogeochemical modeling, including lateral transport of nitrogen, is necessary to reproduce the spatial structure of the carbon cycle within low order watersheds with strong topographic gradients. Although a number of recent biogeochemical models have spatially distributed hydrologic components to improve the representation of hydrologic processes including lateral water flow, nutrient transport with lateral water flow is not accounted for. Results from this study show that models without grid-to-grid

nitrogen transport simulation may not be sufficient in resolving the spatial structure in carbon and nitrogen stocks, even when lateral water flow is simulated.

Land surface and hydrologic simulations usually assume that LAI only varies by land cover type and only responds to the change of season without inter-annual variability. Remotely-sensed LAI products (e.g., Knyazikhin et al., 1999; Myneni et al., 2002) can be used to drive land surface hydrologic models, but their spatial resolutions are limited. The simulation of carbon and nitrogen in Flux-PIHM-BGC provides spatially-distributed diagnostic LAI to the land surface and hydrologic simulation (Fig. 8), which means the coupled modeling system is able to simulate the response of vegetation to changing hydrologic and land surface conditions in different parts of the watershed. This makes the coupled land surface hydrologic biogeochemistry modeling system an ideal tool for long term simulation of watershed response to changing climate conditions.

Acknowledgments

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research, under Award Number DE-SC0012003, and work facilitated by NSF Critical Zone Observatory program grants to CJD (EAR 07-25019) and SLB (EAR 12-39285, EAR 13-31726). This research was conducted in Penn State's Stone Valley Forest, which is supported and managed by the Penn State's Forestland Management Office in the College of Agricultural Sciences. We thank Dr. Henry Lin, Lauren Smith, Dr. Danielle Andrews, and Dr. Christopher Duffy for providing soil survey data, model hydrologic calibration data, aboveground vegetation data, and soil organic carbon data. The NLDAS-2 forcing data used in this study were acquired as part of the mission of NASA's Earth Science Division and archived and distributed by the Goddard Earth Sciences (GES) Data and Information Services Center (DISC). The data used are listed in the references, tables and CZO data repository at

<http://criticalzone.org/shale-hills/data/datasets/>.

References

- Andrews, D.M., Lin, H., Zhu, Q., Jin, L., Brantley, S.L., 2011. Hot spots and hot moments of dissolved organic carbon export and soil organic carbon storage in the Shale Hills catchment. *Vadose Zone J.* 10 (3), 943.
- Bailey, R.T., Gates, T.K., Romero, E.C., 2015. Assessing the effectiveness of land and water management practices on nonpoint source nitrate levels in an alluvial stream-aquifer system. *J. Contam. Hydrol.* 179 (3), 102–115.
- Bailey, R.T., Morway, E.D., Niswonger, R.G., Gates, T.K., 2013. Modeling variably saturated multispecies reactive groundwater solute transport with MODFLOW-UZF and RT3D. *Groundwater* 51 (5), 752–761.
- Baldwin, D.C., 2011. Catchment-scale soil water retention characteristics and delineation of hypopedological functional units in the Shale Hills Catchment (M.S. thesis). The Pennsylvania State University.
- Bao, C., Li, L., Shi, Y., Duffy, C., 2017. Understanding watershed hydrogeochemistry: 1. Development of RT-Flux-PIHM. *Water Resour. Res.* 53, 2328–2345.
- Benda, L., Hassan, M.A., Church, M., May, C.L., 2005. Geomorphology of steep-land headwaters: the transition from hillslopes to channels. *J. Am. Water Resour. Assoc.* 41 (4), 835–851.
- Bodman, R.W., Rayner, P.J., Karoly, D.J., 2013. Uncertainty in temperature projections reduced using carbon cycle and climate observations. *Nat. Climate Change* 3 (8), 725–729.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32.
- Camporese, M., Paniconi, C., Putti, M., Orlandini, S., 2010. Surface–subsurface flow modeling with path-based runoff routing, boundary condition-based coupling, and assimilation of multisource observation data. *Water Resour. Res.* 46 (2), W02512.
- Chen, F., Dudgeon, J., 2001. Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: model implementation and sensitivity. *Mon. Weather Rev.* 129 (4), 569–585.
- Clark, M.P., Fan, Y., Lawrence, D.M., Adam, J.C., Bolster, D., Gochis, D.J., Hooper, R.P., Kumar, M., Leung, L.R., Mackay, D.S., Maxwell, R.M., Shen, C., Swenson, S.C., Zeng, X., 2015. Improving the representation of hydrologic processes in earth system models. *Water Resour. Res.* 51 (8), 5929–5956.
- Cosby, B.J., Hornberger, G.M., Clapp, R.B., Ginn, T.R., 1984. A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. *Water Resour. Res.* 20 (6), 682–690.
- Dozier, J., Frew, J., 1990. Rapid calculation of terrain parameters for radiation modeling from digital elevation data. *IEEE Trans. Geosci. Remote Sens.* 28 (5), 963–969.
- Ek, M.B., Mitchell, K.E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., Tarpley, J.D., 2003. Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.* 108 (D22), 8851.
- Farquhar, G.D., von Caemmerer, S., Berry, J.A., 1980. A biochemical model of photosynthetic CO₂ assimilation in leaves of C₃ species. *Planta* 149 (1), 78–90.
- Fatichi, S., Ivanov, V.Y., Caporali, E., 2012. A mechanistic ecohydrological model to investigate complex interactions in cold and warm water-controlled environments: 1. Theoretical framework and plot-scale analysis. *J. Adv. Model. Earth Syst.* 4 (5), M05002.
- Fatichi, S., Pappas, C., Ivanov, V.Y., 2016. Modeling plant-water interactions: an ecohydrological overview from the cell to the global scale. *Wiley Interdiscip. Rev. Water* 3 (3), 327–368.
- Fisher, J.B., Huntzinger, D.N., Schwalm, C.R., Sitch, S., 2014. Modeling the terrestrial biosphere. *Annu. Rev. Environ. Resour.* 39 (1), 91–123.
- Friedlingstein, P., Meinshausen, M., Arora, V.K., Jones, C.D., Anav, A., Liddicoat, S.K., Knutti, R., 2014. Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *J. Climate* 27 (2), 511–526.
- Gaines, K.P., Stanley, J.W., Meinzer, F.C., McCulloh, K.A., Woodruff, D.R., Chen, W., Adams, T.S., Lin, H., Eissenstat, D.M., 2015. Reliance on shallow soil water in a mixed-hardwood forest in central Pennsylvania. *Tree Physiol.* 36 (4), 444–458.
- Guo, Q., 2010. Susquehanna Shale Hills Critical Zone Observatory: Leaf off survey. Houghton, R.A., 2005. Aboveground forest biomass and the global carbon balance. *Glob. Change Biol.* 11 (6), 945–958.
- Ivanov, V.Y., Bras, R.L., Vivoni, E.R., 2008. Vegetation-hydrology dynamics in complex terrain of semiarid areas: 1. A mechanistic approach to modeling dynamic feedbacks. *Water Resour. Res.* 44 (3), W03429.
- Jarvis, P.G., 1976. The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field. *Philos. Trans. R. Soc. B: Biol. Sci.* 273 (927), 593–610.
- Jenkins, J.C., Chojnacki, D.C., Heath, L.S., Birdsey, R.A., 2003. National-scale biomass estimators for United States tree species. *For. Sci.* 49 (1), 12–35.
- Jobbágy, E.G., Jackson, R.B., 2000. The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecol. Appl.* 10 (2), 423–436.
- Ju, W., Chen, J.M., Black, T.A., Barr, A.G., McCaughey, H., Roulet, N.T., 2006. Hydrological effects on carbon cycles of Canada's forests and wetlands. *Tellus B* 58 (1), 16–30.
- Knyazikhin, Y., Glassy, J., Privette, J.L., Tian, Y., Lott, A., Zhang, Y., Wang, Y., Morisette, J.T., Votava, P., Myneni, R.B., Nemani, R.R., Running, S.W., 1999. MODIS leaf area index (LAI) and fraction of photosynthetically active radiation absorbed by vegetation (FPAR) product (MOD15) algorithm theoretical basis document. Tech. rep. NASA Goddard Space Flight Center, Greenbelt, Maryland.
- Kollet, S.J., Maxwell, R.M., 2006. Integrated surface-groundwater flow modeling: a free-surface overland flow boundary condition in a parallel groundwater flow model. *Adv. Water Resour.* 29 (7), 945–958.
- 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), W02402.
- Koven, C.D., Riley, W.J., Subin, Z.M., Tang, J.Y., Torn, M.S., Collins, W.D., Bonan, G.B., Lawrence, D.M., Swenson, S.C., 2013. The effect of vertically resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4. *Biogeosciences* 10 (11), 7109–7131.
- Lin, H., 2006. Temporal stability of soil moisture spatial pattern and subsurface preferential flow pathways in the Shale Hills catchment. *Vadose Zone J.* 5 (1), 317–340.
- Lin, H., Zhou, X., 2007. Evidence of subsurface preferential flow using soil hydrologic monitoring in the Shale Hills catchment. *Eur. J. Soil Sci.* 59 (1), 34–49.
- Lohse, K.A., Brooks, P.D., McIntosh, J.C., Meixner, T., Huxman, T.E., 2009. Interactions between biogeochemistry and hydrologic systems. *Annu. Rev. Environ. Resour.* 34 (1), 65–96.
- Maxwell, R.M., Condon, L.E., Kollet, S.J., 2015. A high-resolution simulation of groundwater and surface water over most of the continental US with the integrated hydrologic model ParFlow v3. *Geosci. Model Dev.* 8 (3), 923–937.
- Myneni, R.B., Hoffman, S., Knyazikhin, Y., Privette, J.L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G.R., Lott, A., Friedl, M., Morisette, J.T., Votava, P., Nemani, R.R., Running, S.W., 2002. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sens. Environ.* 83 (1–2), 214–231.
- Niu, G.-Y., Paniconi, C., Troch, P.A., Scott, R.L., Durcik, M., Zeng, X., Huxman, T., Goodrich, D.C., 2014. An integrated modelling framework of catchment-scale ecohydrological processes: 1. Model description and tests over an energy-limited watershed. *Ecohydrology* 7 (2), 427–439.
- Niu, G.-Y., Yang, Z.-L., Mitchell, K.E., Chen, F., Ek, M.B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., Xia, Y., 2011. The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *J. Geophys. Res.* 116 (D12), D12109.
- Oleson, K.W., Lawrence, D.M., Bonan, G.B., Flanner, M.G., Kluzek, E., Lawrence, P.J., Levis, S., Swenson, S.C., Thornton, P.E., Dai, A., Decker, M., Dickinson, R., Feddes, J., Heald, C.L., Lamarque, J.-f., Mahowald, N., Niu, G.-Y., Qian, T., Randerson, J., Running, S., Sakaguchi, K., Slater, A., Stöckli, R., Wang, A., Yang, Z.-L., Zeng, X., 2

- Smith, L.A., Eissenstat, D.M., Kaye, M.W., 2017. Variability in aboveground carbon driven by slope aspect and curvature in an eastern deciduous forest, USA. *Can. J. For. Res.* 47 (2), 149–158.
- Tague, C.L., Band, L.E., 2004. RHESys: Regional Hydro-Ecologic Simulation System – an object-oriented approach to spatially distributed modeling of carbon, water, and nutrient cycling. *Earth Interact.* 8 (19), 1–42.
- Thornton, P.E., Law, B.E., Gholz, H.L., Clark, K.L., Falge, E., Ellsworth, D.S., Goldstein, A.H., Monson, R.K., Hollinger, D., Falk, M., Chen, J., Sparks, J.P., 2002. Modeling and measuring the effects of disturbance history and climate on carbon and water budgets in evergreen needleleaf forests. *Agric. For. Meteorol.* 113 (1–4), 185–222.
- Thornton, P.E., Rosenbloom, N.A., 2005. Ecosystem model spin-up: estimating steady state conditions in a coupled terrestrial carbon and nitrogen cycle model. *Ecol. Model.* 189 (1–2), 25–48.
- Tian, Y.Q., Davies-Colley, R.J., Gong, P., Thorrold, B.W., 2001. Estimating solar radiation on slopes of arbitrary aspect. *Agric. For. Meteorol.* 109 (1), 67–74.
- Tucker, G.E., Lancaster, S.T., Gasparini, N.M., Bras, R.L., Rybarczyk, S.M., 2001. An object-oriented framework for distributed hydrologic and geomorphic modeling using triangulated irregular networks. *Comput. Geosci.* 27 (8), 959–973.
- van Genuchten, M.T., 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Soc. Am. J.* 44 (5), 892–898.
- White, M.A., Thornton, P.E., Running, S.W., 1997. A continental phenology model for monitoring vegetation responses to interannual climatic variability. *Glob. Biogeochem. Cycles* 11 (2), 217–234.
- White, M.A., Thornton, P.E., Running, S.W., Nemani, R.R., 2000. Parameterization and sensitivity analysis of the BIOME-BGC terrestrial ecosystem model: net primary production controls. *Earth Interact.* 4 (3), 1–85.
- Wubbels, J.K., 2010. Tree species distribution in relation to stem hydraulic traits and soil moisture in a mixed hardwood forest in central Pennsylvania (Master thesis). The Pennsylvania State University.
- Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Lettenmaier, D., Koren, V., Duan, Q., Mo, K., Fan, Y., Mocko, D., 2012. Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *J. Geophys. Res.* 117, D03109.
- Yu, X., Duffy, C.J., Baldwin, D.C., Lin, H., 2014. The role of macropores and multi-resolution soil survey datasets for distributed surface–subsurface flow modeling. *J. Hydrol.* 516, 97–106.
- Yu, X., Duffy, C.J., Kaye, J., Bhatt, G., Shi, Y., 2015. Watershed reanalysis of water and carbon cycle models at a Critical Zone Observatory. In: Lakshmi, V., Alsdorf, D., Anderson, M., Biancamaria, S., Cosh, M., Entin, J., Huffman, G., Kustas, W., van Oevelen, P., Painter, T., Parajka, J., Rodell, M., Rüdiger, C. (Eds.), *Remote Sensing of the Terrestrial Water Cycle*. John Wiley & Sons, Inc, Hoboken, NJ, pp. 493–509 (Chapter 31).
- Zhang, X., Niu, G.-Y., Elshall, A.S., Ye, M., Barron-Gafford, G.A., Pavao-Zuckerman, M., 2014. Assessing five evolving microbial enzyme models against field measurements from a semiarid savannah – what are the mechanisms of soil respiration pulses? *Geophys. Res. Lett.* 41 (18), 6428–6434.
- Zhang, Y., Slingerland, R., Duffy, C., 2016. Fully-coupled hydrologic processes for modeling landscape evolution. *Environ. Model. Softw.* 82, 89–107.