Cyber-Innovated Watershed Research at the Shale Hills Critical Zone Observatory

Xuan Yu, Christopher Duffy, Yolanda Gil, Lorne Leonard, Gopal Bhatt, and Evan Thomas

***Abstract*—Cyberinfrastructure is enabling ever more integrative and transformative science. Technological advances in cyberinfrastructure have allowed deeper understanding of watershed hydrology by improved integration of data, information, and models. The synthesis of all sources of hydrologic variables (historical, real time, future scenarios, observed, and modeled) requires advanced data acquisition, data storage, data management, data integration, data mining, and data visualization. In this context, cyber-innovated hydrologic research was implemented to carry out watershed-based historical climate simulations at the Shale Hills Critical Zone Observatory. The simulations were based on the assimilation of data from a hydrologic monitoring network into a multiphysics hydrologic model (the Penn State Integrated Hydrology Model). We documented workflows for the model application and applied the model to short-time hyporheic exchange flow study and long-term climate scenario analysis. The effort reported herein demonstrates that advances in cyberscience allows innovative research that improves our ability to access and share data; to allow collective development of science hypotheses; and to support building models via team participation. We simplified communications between model developers and community scientists, software professionals, students, and decision makers, which in the long term will improve the utilization of hydrologic models for science and societal applications.**

***Index Terms*—Critical zone observatories (CZOs), cyberinfrastructure, data analytics, penn state integrated hydrologic model (PIHM), shale hills, watershed, web services.**

I. Introduction

WATERSHED modeling has become a fundamental tool for evaluating the quantity and quality of regional water resources. Spatially distributed watershed models make use of both geospatial information and observation systems to predict multiple hydrologic state variables necessary for assessing impacts of climate and land-use change or the response of extreme weather events [1]. Such models capture the experimental evidence generated by catchment scientists for the nonlinear behavior of coupled surface–subsurface systems. However, real-world applications of physics-based modeling require extensive observations from multistate sensors (soil moisture, groundwater level, streamflow, etc.), to characterize the space–time characteristics of the watershed, and an implicitly extensive analysis of the model parameter fields, calibration, and validation of the model results. Thus, it is still challenging to integrate models and data at appropriate scales for resolving watershed dynamics [2].

Another challenge of watershed models is the model reusability. Both development and applications of the model involve benchmark testing and real watershed validation. Often, the detailed modeling results serve a particular research project, with little interest (or funding) for openly available results or model annotation and documentation beyond project publication. Few scientists or engineers are trained in "best practices" for reusability of the model and model simulation results, which restricts the potential impacts of both. There is a clear demand to provide water managers and stakeholders efficient and simplified access to both models and data for assessing the nation's water resources [3].

Hydrological model data contribute not only to sustainable water resources management but also to water-related scientific applications. Earth and environmental sciences can also benefit from shared data and models. One example of such model data product is the National Land Data Assimilation System [4]. The data set has been providing easily accessible data of land surface forcing, energy, and water flux, which supports researchers in hydrology, ecology, and geology. Noticeably, collaborative science is becoming a *de facto* strategy for Earth science research (e.g., Critical Zone Observatory (CZO): http://criticalzone.org/national/; Long Term Ecological Research Network (LTER): http://www.lternet.edu/; National Ecological Observatory Network: http://www.neoninc.org/). In watershed hydrology, the most readily available data include streamflow, precipitation, soil moisture, groundwater table elevation, etc. However, they are often limited in space and/or time. The model-simulated fluxes, such as evapotranspiration (ET), recharge, and baseflow, are products valuable for testing hypotheses or future scenarios of change. It is also true that modeling other Earth-surface processes, such as sediment transport, solute transport, vegetation growth, nutrient redistribution, landscape evolution, etc., first requires detailed knowledge of the hydrologic regime. In many cases, the specific needs of understanding these processes, in terms of spatial and temporal resolution or scale, may differ. Earth scientists will need to rerun and redesign the hydrologic models to support their own research and hypotheses.

Manuscript received November 13, 2014; revised March 6, 2015, June 15, 2015, and August 15, 2015; accepted September 7, 2015. This work was supported by the National Science Foundation under Grant EAR-0725019, Grant EAR-1239285, Grant EAR-1331726, Grant ICER-1440323, Grant ICER-1343800, and Grant IIS-1344272. This work was also supported in part through instrumentation by the National Science Foundation under Grant OCI-0821527.

X. Yu was with the Department of Civil and Environmental Engineering, The Pennsylvania State University, University Park, PA 16802 USA. He is currently with the Department of Geological Sciences, University of Delaware, Newark, DE 19716 USA (e-mail: xuan@udel.edu).

C. Duffy, L. Leonard, G. Bhatt, and E. Thomas are with the Department of Civil and Environmental Engineering, The Pennsylvania State University, University Park, PA 16802 USA (e-mail: cxd11@psu.edu; lnl3@psu.edu; gxb913@psu.edu; emt175@psu.edu).

Y. Gil is with the Information Sciences Institute, University of Southern California, Marina del Rey, CA 90292 USA (e-mail: gil@isi.edu).

Digital Object Identifier 10.1109/JSYST.2015.2484219

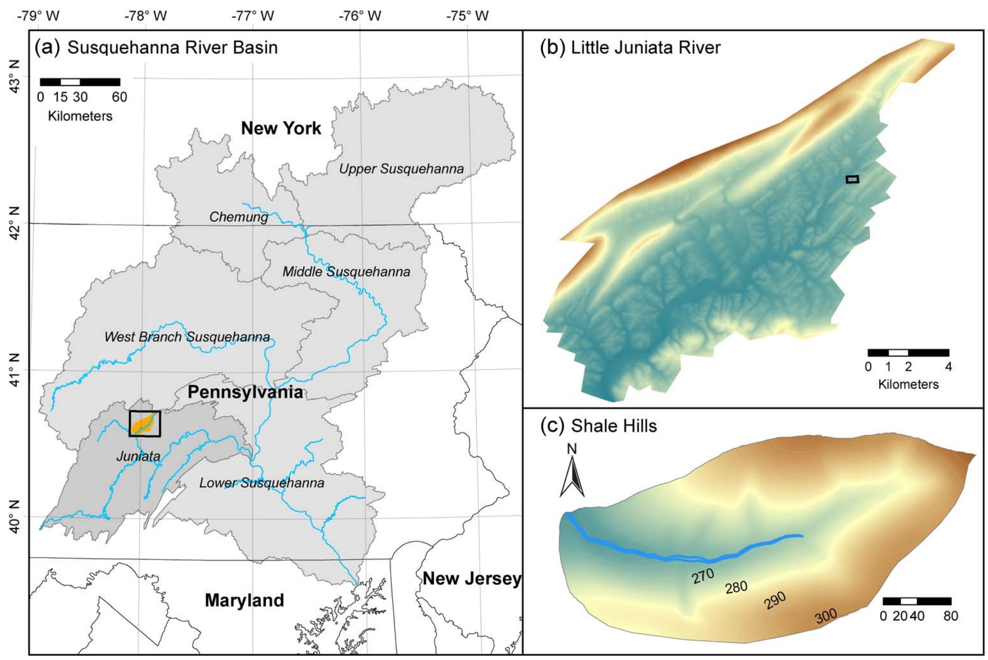


Fig. 1. Location of Shale Hills at the Susquehanna River Basin. The modeling study was focused on Shale Hills and then scaled up to Little Juniata River and further to Susquehanna River Basin.

It is fair to say that the participatory and collaborative nature of hydrologic models in "team science" is a major challenge [5].

In 2005, the U.S. National Science Foundation (NSF) created a new Office of Cyberinfrastructure (OCI). The OCI has been providing infrastructure for science and engineering research to enable integrative, transformative, and sustainable knowledge [6]. In 2011, the NSF initiated the "EarthCube" project to develop a common or shared cyberinfrastructure for geoscientists to improve and facilitate interdisciplinary research [7]. This progressive effort provides a challenging and stimulating opportunity for the development of domain scientists to implement state-of-the-art cyberinfrastructure resources that, in the past, could not or was not supported. Stewart *et al.* note that cyberinfrastructure consists of computational systems, data and information management, advanced instruments, visualization environments, and people, all linked together by software and advanced networks to improve scholarly productivity and enable knowledge breakthroughs and discoveries not otherwise possible [8]. It is timely to explicitly introduce advanced cyberinfrastructure in watershed hydrology by supporting data acquisition, data storage, model development, data management, data integration, data visualization, etc.

In this paper, we use the Susquehanna Shale Hills Critical Zone Observatory (SSHCZO) as a testbed to demonstrate the diverse cyber-innovated watershed hydrology for interdisciplinary research, and further explore how interdisciplinary research is benefiting from the advanced cyberinfrastructure. Specifically, we first compare historical hydrological research at the testbed and the current cyber-innovated hydrological development. We then demonstrate that how current advances integrate the observed watershed data and model simulation. Such cyber-based integration can facilitate the understanding of external collaborators and be reused for other studies. Finally, we provide specific examples of interdisciplinary research based on the shared model and data.

II. Historical Research at Shale Hills

The Shale Hills Watershed, with an area of 0.08 km^2, is entirely forested with an ephemeral first-order stream in the uplands of the Juniata River watershed, which is the second largest tributary of the Susquehanna River (see Fig. 1). The research history of Shale Hills can be traced back to 1958, when it was paired with a neighboring watershed, i.e., Leading Ridge Watershed, to understand the water yield from forested and managed watersheds [9]–[14]. Extensive observations were made for streamflow, weather, water quality, nutrients, and atmospheric deposition. In 1974, a controlled irrigation experiment was conducted at the Shale Hills Watershed [12]. The watershed was implemented with a spray irrigation network to precisely control the amount of artificial rainfall over the entire watershed. From July to September 1974, a series of six equal artificial rainfall events (0.64 cm/h for 6 h) was applied to the entire watershed. During the experiment, a spatial array of 40 groundwater level and soil moisture sites was measured daily. The streamflow was recoded at a 15-min interval. The data were used for studies by forest hydrologists to resolve the role of antecedent moisture in runoff peak flows within a forest canopy. Only part of the data set is available at an unmaintained website [14]. In 2007, a CZO was established at Shale Hills with the goal of developing integrated, extensive, and accessible Earth science data sets for research. Since then, a wide range of data have been maintained at the SSHCZO website by a team of data management specialists. In 2014, the NSF initiated another research program at Shale Hills (NSF IIS-1344272), particularly focusing on the development of cyberinfrastructure, to provide new collaborations across diverse scientific communities and to share and normalize data to solve scientific problems through an open framework.

Given the historical and modern experimental data, the Shale Hills Watershed is interesting as a hydrological model testbed for decadal change. The early research on antecedent soil moisture and storm flow involved a regression model [12]. The model was built based on the correlation analysis between antecedent soil moisture and baseflow and storm flow in the experiment in 1974. Later, the Penn State Integrated Hydrologic Model (PIHM), a physics-based fully distributed model, was developed and implemented at Shale Hills [15], which initially was used to explain the antecedent soil moisture effects on storm hydrographs from a physical perspective. As improved and new environmental data sets became available, new model data processing toolkits emerged [16], which took advantage of both the historical and the new experimental research [17]. A recent modeling development study focused on coupling land surface processes (energy and vegetation dynamics) in an extended hydrological modeling system. Flux-PIHM, which is the coupled hydrologic and land surface model, improves the energy balance at land surface and integrates physical constraints to surface heat fluxes and subsurface water movement [18].

III. Current Multiphysics Approach

The multiphysics approach used in the watershed modeling code requires intensive data and computation resources, which can of course be benefited by advanced cyberinfrastructure. Our goal is to use cyberinfrastructure to facilitate the PIHM application, which involves data acquisition, data management, data integration, data sharing, and data visualization (see Fig. 2).

A. Data Acquisition

The data acquisition at SSHCZO includes both local observational data collection and national geospatial and data harvesting.

We have designed and built a basic system based on wireless sensor network technology for low-power wireless support of sensor nodes for large arrays of multistate digital sensing. The sensors include pressure, moisture, water level, wind, temperature, electrical conductance, relative humidity, infrared skin temperature, and acoustic snow depth sensors. The network is fully integrated with standard Campbell Scientific data loggers, with two-way web access and sensor control, which provides the real-time monitoring of the watershed. Table I listed the hydrologic data collected at SSHCZO.

Under separate funding, national watershed data services were developed to support web-based acquisition of Essential Terrestrial Variables, which are basic infrastructures for environmental models (HydroTerre). HydroTerre represents the fundamental national data necessary to run high-resolution catchment models anywhere in the USA [19], [20]. This data acquisition service is available to scientists, students, and other research organizations at the catchment scale.

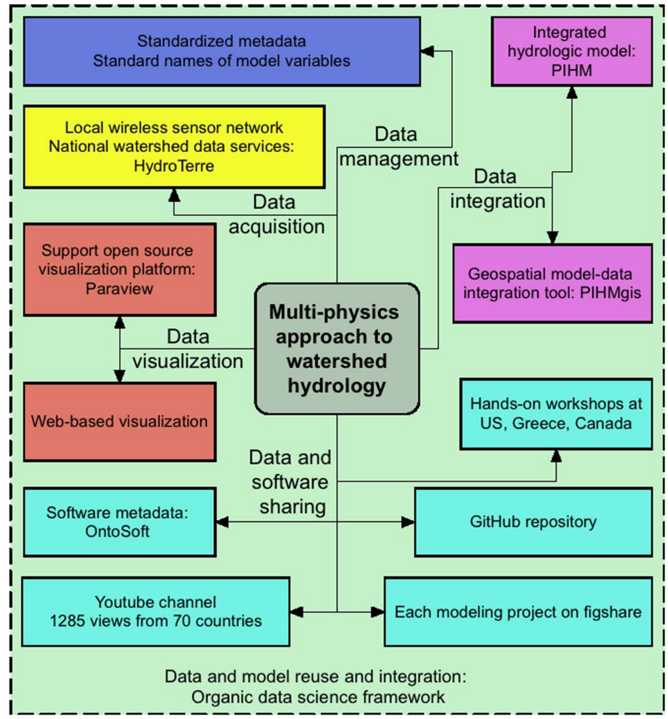


Fig. 2. Cyberinfrastructure functions of PIHM development. The links are listed in Table II in the Appendix.

B. Data Management

The 1974 irrigation experiment data were original preserved on punch cards and digitalized for PIHM application [15]. The hydrologic observations were mapped to a standard name database, which is maintained by the Community Surface Dynamics Modeling System (CSDMS). The standard name database is available through CSDMS variables, process models, data sets, and their associated variables [21].

C. Data Integration

To integrate the growing observational data at Shale Hills, PIHM has been developed to meet the new modeling requirements, where the hydrology is coupled with ecosystem, geochemical, and geomorphic processes. The PIHM model itself is “tightly-coupled” with PIHMgis [16], which is an open-source Geographical Information System (GIS) tool designed for PIHM. The PIHMgis provides the interface linking national and observatory digital data sets (terrain, forcing, and parameters) with functions for domain decomposition and mesh generation, and parameters initialization. Such data integration tool provides users a well-organized layout of the hydrologic data.

D. Data and Software Sharing

Data sharing includes distribution of both data and model because hydrologic data are usually tightly coupled with model simulation. An important element of watershed hydrology research at Shale Hills is the community science and team research activities and the concept of “community models” for

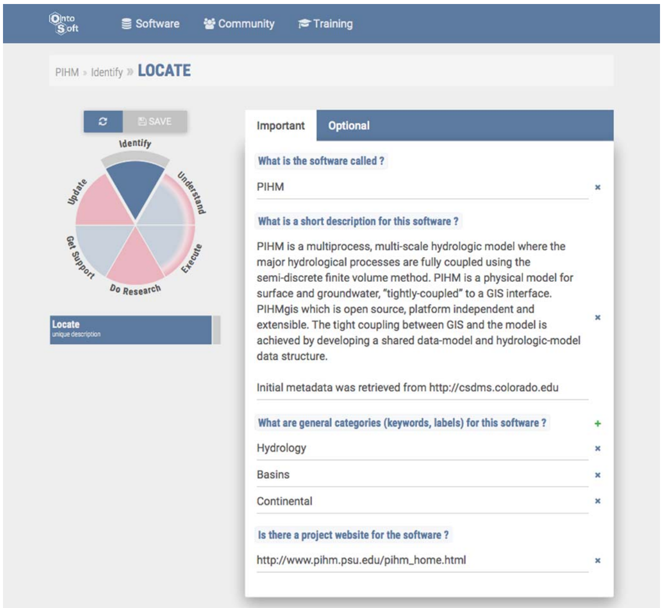


Fig. 3. Metadata for the PIHM software, captured in the OntoSoft portal.

prediction of environmental variables. PIHM has been maintained as an open and extensible numerical platform available on the PIHM group website (www.pihm.psu.edu) and on the SourceForge website. The PIHM team has made a serious effort to update and make PIHM freely available. PIHM workshops were organized, in the United States, Greece, and Canada, and open to researchers. An informal group of consultants supported e-mail communications about the problems in PIHM development, implementation, and applications. In addition, there are many other explorations on the potential practices to promote the utility of PIHM. For example, PIHM tutorials on YouTube have been viewed 1285 times. The task-oriented online collaboration tool was developed to endeavor research between multiple communities. PIHM is also now distributed on GitHub for the source code version control and development.

We also started to document and upload data sets on figshare, which is an online digital repository where researchers can preserve and share their research outputs.

Repositories, such as figshare and GitHub, can assign Digital Object Identifiers (DOIs) to data sets and software versions, respectively, along with a form of citation in papers. This enables proper credit to the software authors, as well as detailed specification in support of reproducibility.

In addition, we used the OntoSoft portal to describe the PIHM software; hence, it is easier for others to understand and reuse (http://www.ontosoft.org/portal). OntoSoft relies in an ontology to capture scientific software metadata [22]. Fig. 3 shows a snapshot of the OntoSoft portal with a portion of the description for PIHM. The circular icon is used to indicate which metadata is still missing. The metadata is exported as an XML file and HTML; hence, it can be linked from the PIHM GitHub site.

OntoSoft also enables feature-based comparisons of different scientific software with similar function. Fig. 4 compares

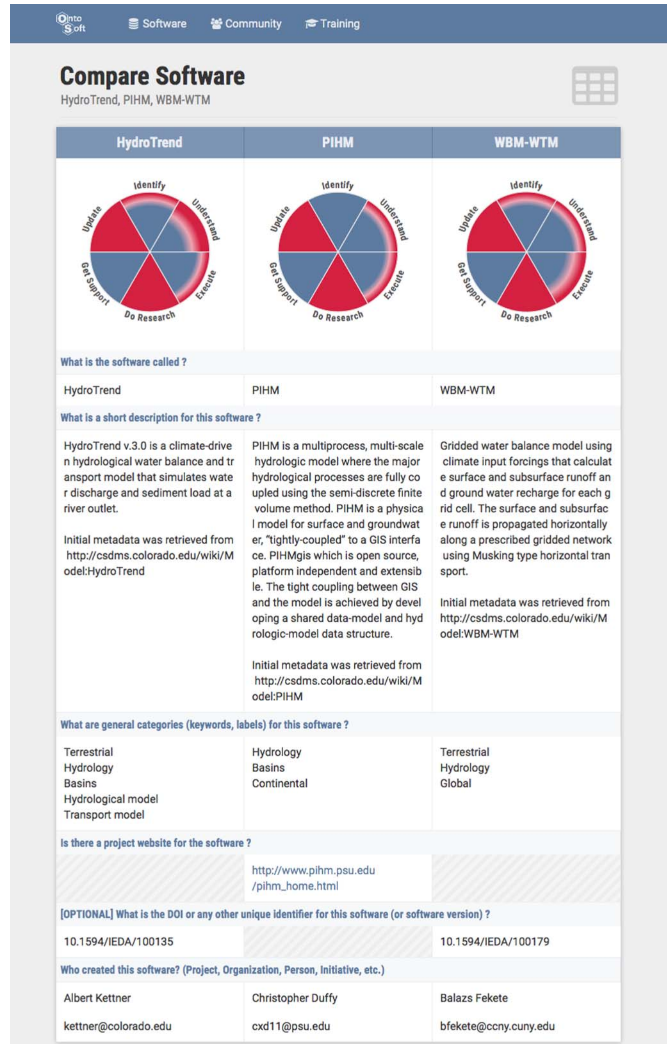


Fig. 4. Comparing scientific software with similar function using the OntoSoft portal.

software for hydrological modeling, which is written in C and released under a GNU General Public License 2.0.

E. Data Visualization

PIHM simulates spatially distributed hydrologic variables (see Table I), which require efficient geospatial data visualization. The output format of the PIHM is only plain text. Currently, the team is developing tools for the standard output format for ParaView, R packages for data analytics and presentation, and web-based interactive visualization tools (http://www.pihm.psu.edu/lysina/forest.html).

F. Data and Model Reuse and Integration

A major challenge in watershed research is the reuse of models, for novel purposes, and the integration of models, particularly across disciplines. One such project is a joint effort with limnologists at the University of Wisconsin, where we are integrating analytical frameworks from two communities, i.e., hydrology and isotope modeling in CZOs and

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I
MONITORING DATA AND HYDROLOGICAL MODEL INTEGRATION

Sensor	Variables	Hydrologic processes	Model parameters	Calibration group*
HOBO water level data logger, v notch weir	streamflow	Surface water flow, water balance	River Mannings roughness, Evapotranspiration parameters	EG, SG
Druck pressure transducers	Water table	Subsurface flow, recharge	Hydraulic conductivity in matrix and macropore	EG
Eddy flux tower	Water loss	Evapotranspiration	Evapotranspiration parameters	SG
DT-100 liquid water isotope analyzer	Water stable isotope	Transport	Hydraulic conductivity in matrix and macropore	EG
Snow scale	Snow water equivalent	Snow melt	Melt factor	EG
Sapflow	Transpiration rate	Transpiration	Minimum canopy resistance	SG
Time-domain reflectometry instrument system	Soil moisture	Infiltration	Soil water retention characteristics	EG

*EG means event-scale group, and the parameters control the flooding processes. SG means seasonal time scale group, and the parameters control the seasonal energy processes.

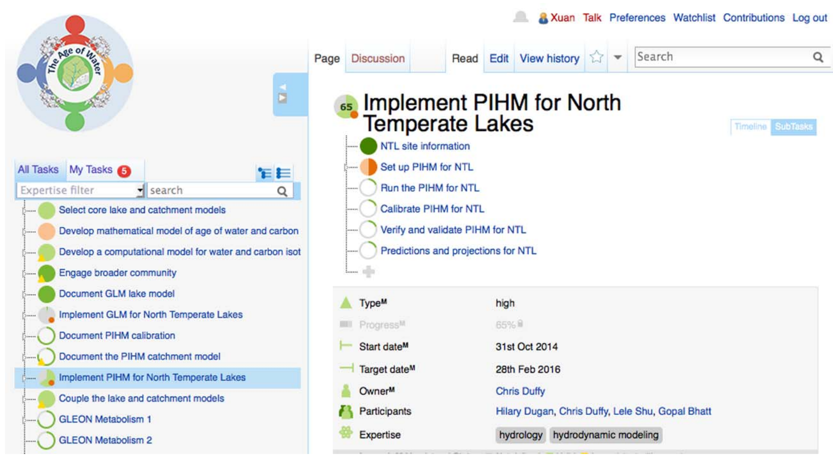


Fig. 5. Using PIHM to study the age of water in a lake-catchment ecosystem, using the Organic Data Science framework for collaboration with limnologists.

hydrodynamic water quality modeling from the Global Lake Ecological Observatory Network (GLEON), to quantify water and material fluxes from two research sites: the Shales Hills CZO and the GLEON member site, North Temperate Lakes LTER. As water age and the associated flowpaths are identified, scientists will use that information to infer the sources of organic carbon to lake-catchment ecosystems, their fluxes from the landscape to lakes, the fates as storage, conversion or export, and understanding of the uncertainties surrounding these quantities (see Table II in the Appendix).

The complex suite of resources, including data sets, computer models, computing resources, or technological staff, must be coordinated and directed toward a common goal. We are using the Organic Data Science framework as a structured environment that can handle this complexity [23], [24]. By documenting the scientific progress, unresolved tasks that must be undertaken are made clear, as a reminder not only to the principal investigators but also to new members who want to contribute. The wiki provides a legacy of documentation and a trail of how results were obtained. Fig. 5 illustrates the use of the Organic Data Science framework to document the tasks involved in setting up PIHM as the catchment model.

IV. APPLICATION

PIHM is a physics-based and spatially distributed hydrologic model (available online at http://www.pihm.psu.edu/). It simulates the terrestrial water cycle, including interception, throughfall, infiltration, recharge, evaporation, transpiration, overland flow, unsaturated soil water, groundwater flow, and channel routing, in a fully coupled scheme [15]. ET is calculated using the Penman–Monteith approach adapted from Noah_LSM [25]. Overland flow is described in 2-D diffusive wave simplification of St. Venant equations. Movement of moisture in the unsaturated zones is assumed to be vertical, which is modeled using the Richards equation. The model assumes that each subsurface layer can have both unsaturated and saturated storage components. The recharge to and from the water table couples the unsaturated and saturated zones to simulate the variably saturated subsurface processes. Channel routing is modeled using 1-D estimation of St. Venant equations, with PIHM again using a diffusive wave approximation. For saturated groundwater flow, the 2-D Dupuit approximation is applied. Spatially, the modeling domain is decomposed into Delaunay triangles. This triangular mesh allows users to resolve spatial data over the watershed, and this can be constrained by

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

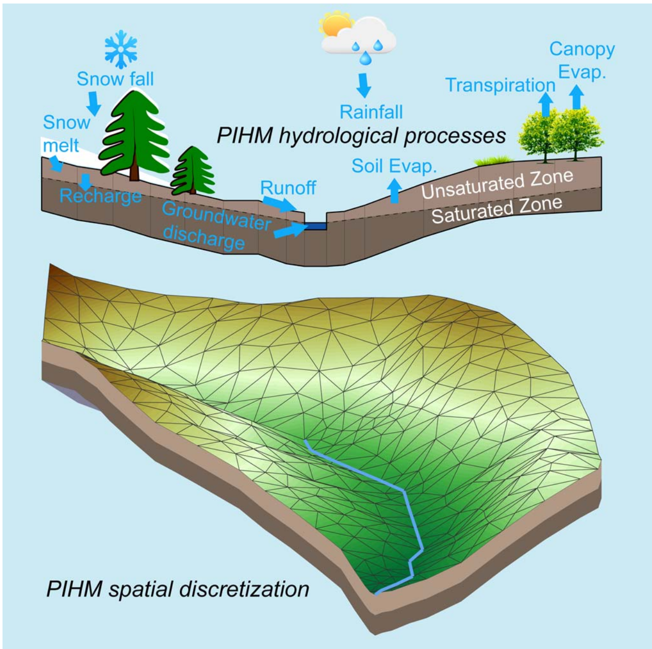


Fig. 6. PIHM structure and processes. The upper subfigure shows the hydrological processes of PIHM at a cross section of a watershed. The lower subfigure shows the spatial structure of PIHM. Blue lines represent the stream channels, and triangles represent the catchment domain.

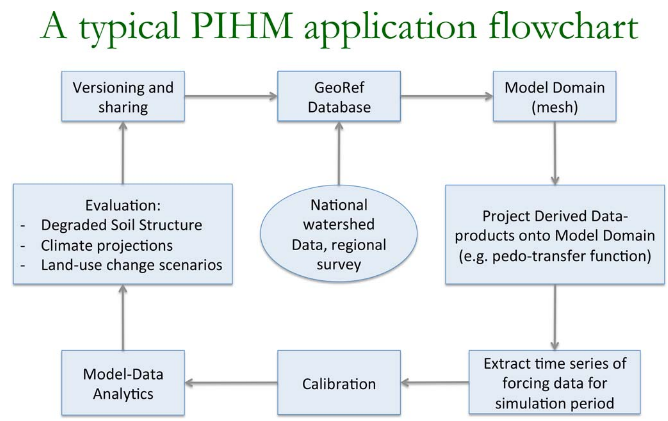


Fig. 7. PIHM application workflow.

point or vector data (e.g., stream gauge, wells, soil maps, and land cover) and the watershed boundary conditions [26]. The model resolves hydrological processes for land surface energy, overland flow, channel routing, and subsurface flow, which are governed by partial differential equations (PDEs) (see Fig. 6). The PDE system is discretized on the triangular mesh and projected prism from canopy to bedrock. PIHM uses a semidiscrete finite-volume formulation for solving the system of coupled PDEs, resulting in a system of ordinary differential equations representing all processes within the prismatic control volume. The main equations of PIHM are listed in Table III in the Appendix. On each prismatic control volume, the original hydrological processes can be easily improved, and new processes can also be integrated into this system. The flexible approach of coupling multiscale hydrological processes makes it adaptable for integrated hydrological simulation of a wide range of interests.

Watershed models are very data intensive, and PIHM simulation requires a wide range of geospatial/geotemporal data to parameterize the physical properties of the watershed. The workflow of PIHM application is presented in Fig. 7. Usually, these data are obtained from national geospatial database products and/or regional surveys. For fast processing geospatial data, a GIS and hydrologic model user interface, i.e., PIHMgis, was developed [16], as mentioned earlier. PIHMgis provides functionalities for watershed delineation, domain decomposition, parameter assignment, simulation, visualization, and analyses. The forcing of PIHM is the meteorological time series, including precipitation, temperature, wind speed, relative humidity, solar radiation, etc. The input file of forcing allows a flexible format for timestamp and total duration, harmonizing the disparate sources of meteorological data required (e.g., sampling rates and resolutions). At this step, PIHM is ready for the initial simulation and calibration. For calibrating parameters in real settings, a partition calibration strategy has been developed to optimize the parameters of PIHM [27]. The partition calibration strategy is based on the two driving forces of hydrologic processes in PIHM: energy from and gravity. The energy-driven processes are evaporation and transpiration, which operate in seasonal to annual time scales, while flood events are largely controlled by gravity. A natural separation in the parameters based on an event-scale group (EG) and a seasonal time scale group (SG) is carried. The Covariance Matrix Adaptation Evolution Strategy [28] is used first to optimize the EG parameters. Then, the SG parameters are sequentially resulting in an efficient and fast global water balance. A typical application is model calibration, reconstruction of the historical hydrologic conditions, followed by a projection of future conditions, all of which are made available to analyze management scenarios, specific scientific hypothesis testing, or other purposes. Finally, documenting and versioning model instances are important steps in reusability and adaptability of the code.

The recent monitoring network at Shale Hills is shown in Fig. 8. Monitoring devices include precipitation observations for amount, intensity, and types, at a 10-min resolution. A network of 17 groundwater wells was installed in the valley bottom and in swales where shallow groundwater was observed periodically. Additional deep-water wells were installed along the ridge top to monitor deep groundwater dynamics. Suction cup lysimeters were installed in swales and on planar hillslopes and sampled biweekly. Tensiometers and soil moisture probes were installed throughout the catchment and equipped with real-time loggers to monitor soil moisture dynamics. Additionally, a passive cosmic-ray sensor COSMOS [29] probe was installed in the center of the catchment to monitor soil moisture dynamics hourly, as well as a double V-notch weir at the catchment outlet to monitor streamflow at high and low flow conditions [9] both in real time.

To setup the PIHM simulation at Shale Hills, a 0.5-m-resolution digital elevation model was flown and processed to represent the surface topography in the model. Geophysics tools were used to map bedrock depth to estimate the thickness from regolith. Detailed tree survey data were used for the land cover

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

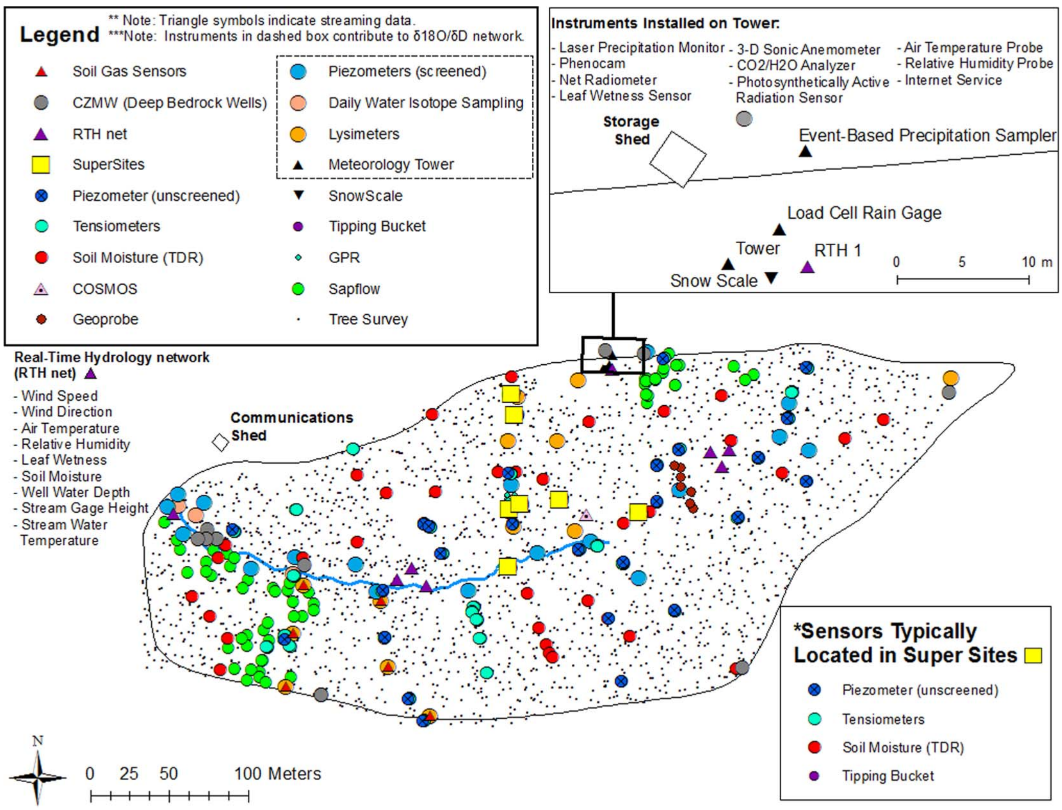


Fig. 8. Monitoring network at Shale Hills. (Information of the sensors is obtained at http://criticalzone.org/shale-hills/.)

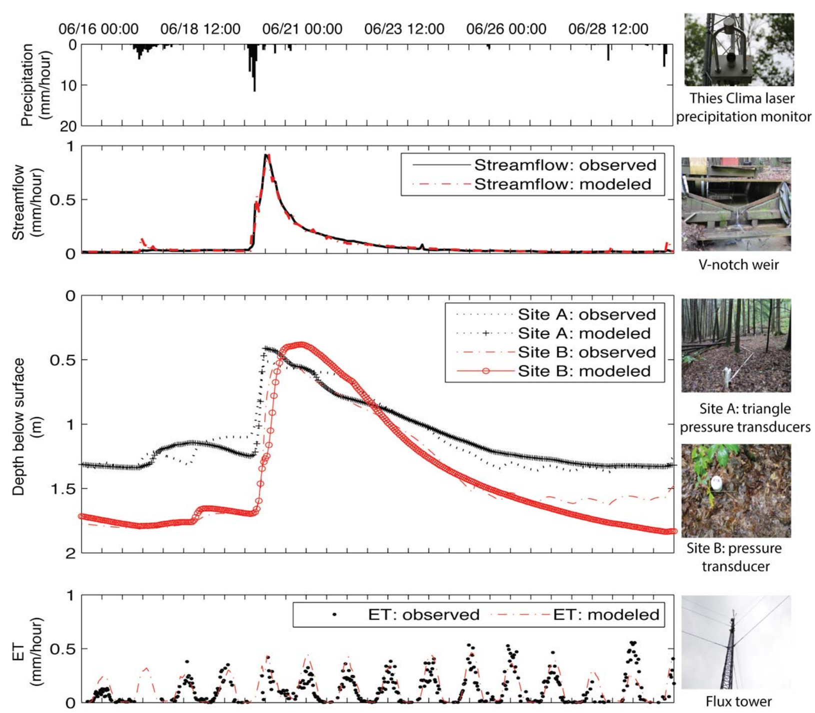


Fig. 9. Calibration result in 2009. The precipitation is monitored by the Thies CLIMA Laser Precipitation Monitor at the weather station. The streamflow is monitored by the notch. The groundwater depth at site A is observed by a Druck pressure transducer CS420-L. The groundwater depth at site B is observed by a 0.5-m Odyssey capacitance water level recorder. The latent heat flux is measured with a LI-COR LI-7500 CO2/H2O Analyzer and then is converted into ET.

classification and parameterization [30]. The soil survey data were used for the soil mapping and parameterization [31]. The forcing data for PIHM include basic meteorological variables observed at the weather station. Table I lists the field data and corresponding hydrologic processes, which were used for the model parameter estimation. The calibration of EG parameters was carried out on the Penn State CyberSTAR: A Scalable Terascale Advanced Resource. The EG calibration targeted

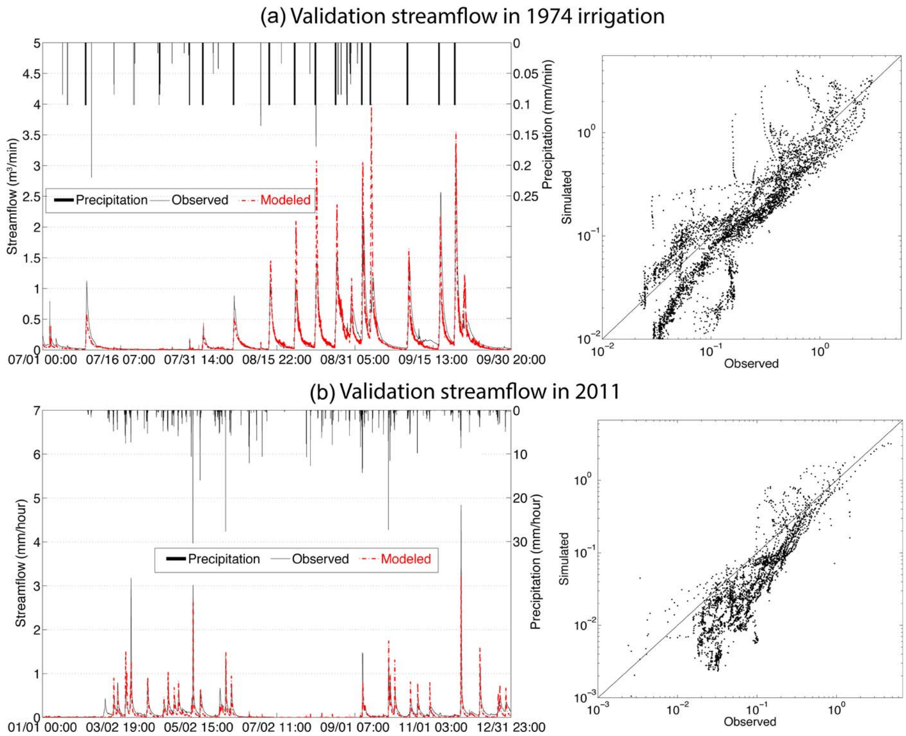


Fig. 10. Validation of the rainfall runoff responses in (a) 1974 and (b) 2011.

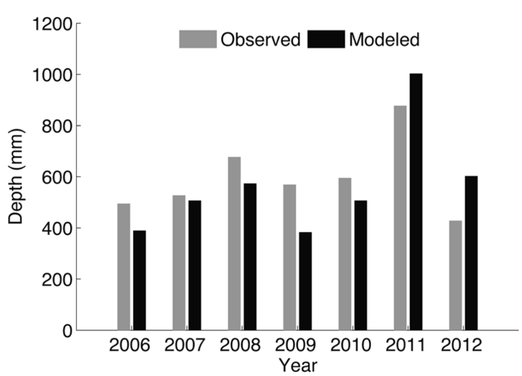


Fig. 11. Validation of annual streamflow.

short-term land surface fluxes, while the SG calibration targeted the seasonal water budget fluxes and states [27]. Observed streamflow during 2009 was the calibration period. Fig. 9 shows the rainfall runoff event used for EG calibration. Model parameters were validated with observation periods in 1974 and 2011. The modeled and observed streamflows were in good agreement (see Fig. 10). The comparison between modeled and observed annual streamflow (see Fig. 11) demonstrates that the PIHM simulation captured the long-term hydrological regime and short-term event dynamics.

V. Interdisciplinary Research Implications

A. Hyporheic Zone Hydrological Processes

Hyporheic zone (HZ) dynamics is a current topic of a range of researchers in hydrology, biogeochemistry, and ecology to examine the complex ecohydrological and biogeochemical processes at the interface between groundwater and surface water [32]. A general definition of HZ is a region beneath and adjacent to a streambed, where there is mixing of shallow groundwater and surface water. A PIHM simulation of the calibrated model examined the response of rainfall events on hyporheic exchange flow (HEF). The heaviest storm in the year 2009 occurred on October 24th. The groundwater flow direction is shown in Fig. 12. The left panel is the relative dry condition that existed before the precipitation event. We observe that the HEF exchanges surface and groundwater in a dynamic and spatially variable way according to the topographic features of the watershed and geometry of the stream channel. Note that the right panel in Fig. 12 represents wet conditions during the precipitation event and the stream is mainly recharging the aquifer. Current research is assessing the hydrologic, topographic, and weather regimes that impact the HEF.

B. Climate Change Impacts

The possible effects of climate change on hydrology were investigated by creating historical and future climate scenarios based on the output of one global climate model from phase 3 of the Coupled Model Intercomparison Project (CMIP3) [33]. Because differences among climate models account for much of the spread in future climate projections, it is preferred to use multiple climate models when projecting the impact of future climate change. However, computational resource limitations in running PIHM forced us to select a single model for the hydrologic impact assessment. The historical scenario in this study is based on years 1979–1998 from the twentieth-century experiment (20C3M), and the future scenario is based on years 2046–2065 from the Special Report on Emissions Scenario (SRES) A2. The scenario's climate forcing showed

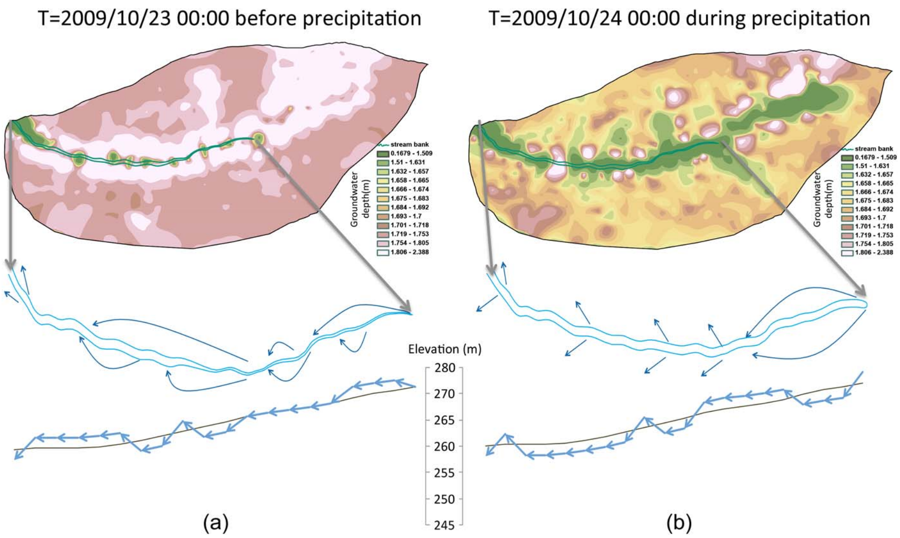


Fig. 12. Hyporheic exchange flow variation before and during the precipitation event on October 24, 2009. The top subplot shows the simulated spatial distribution of groundwater. The middle subplot shows the simulated flow direction around the stream. The bottom subplot shows the simulated flow direction across the riverbed. (a) Dry watershed: Pre-event groundwater table depth and HEF path. Horizontally, HEF vaies (gaining/loosing) with stream reaches. Vertically, some reaches have no exchange with the river beds. (b) Wet watershed: During-event groundwater table depth and HEF path. Horizontally, the stream is mainly recharging the equifer; Vertically, all the reaches have dynamic HEF with the river beds.

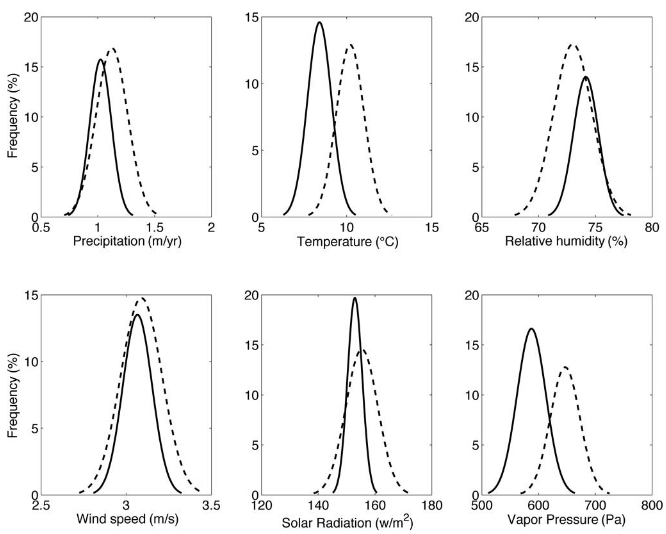


Fig. 13. Frequency variation of meteorological forcing in historical scenario (solid line) and future scenario (dash line).

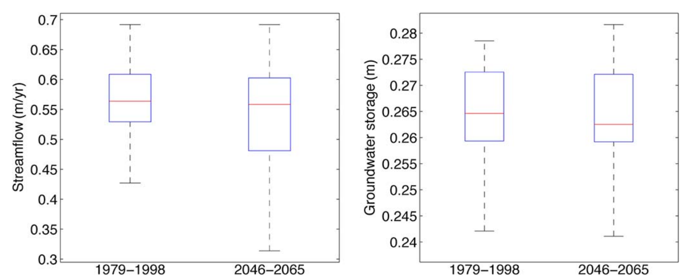


Fig. 14. Hydrological responses of climate change at SSHCZO. The subplots show the simulation annual streamflow and average groundwater storage variation during history (1979–1998) and future (2046–2065) scenarios.

that it would be warm and wet in the middle of this century in Pennsylvania (see Fig. 13). Due to opposite impacts of rising precipitation and temperature, the model simulated hydrological response had different results according to which impact is stronger. The PIHM simulation result here shows modest decrease in average streamflow and groundwater table, and significant increase in the variance or extreme hydrological conditions under the Intergovernmental Panel on Climate Change (IPCC) future scenarios (see Fig. 14). Present work is scaling up PIHM to the whole Susquehanna River Basin to assess the larger scale hydrologic response to climate change.

VI. Summary

In this paper, we have initiated a prototype of cyber-innovated watershed hydrology and explored the impact of such technologies on a real watershed system. This paper demonstrates the widespread and pervasive use of computing technologies and cyberinfrastructure in carrying out this research and making the research reusable. We demonstrate how cyber-innovated watershed hydrology serves as a foundation for interdisciplinary research at the SSHCZO. This paper documents a workflow utilizing real-time processing of sensor data, to community development of models, open scientific data products, and support of individual research hypotheses. Cyber-innovated watershed hydrology is capable of reducing the burden of data and model management and integration. It facilitates data collection and model development and makes hydrologic analyses accessible to research teams of ecosystem and geosciences communities.

Clearly, watershed models and modelers will benefit from the continued improvement and implementation of modern cyberinfrastructure, and serve as a comprehensive toolkit for understanding of hydrologic cycle, to improve our capability for testing hypotheses, and to support team-science.

One should not underestimate the long-term impacts of cyber-innovated Earth system research, for seamless integration of data and models, as well as promotion of model–data reliability, reusability, and preservation, on the promotion of scientific knowledge and technical innovation.

TABLE II
LINKS FOR THE CYBERINFRASTRUCTURE OF PIHM

Appendix I Links for the cyberinfrastructure of PIHM

Theme	Link	Content
Original website	http://www.pihm.psu.edu/	The website was used to provide the source code, documents, and examples of PIHM
Data server	http://www.hydroterre.psu.edu/	The website is providing national watershed data for distributed hydrologic modeling including PIHM
PIHM wiki	http://cataract.cee.psu.edu/PIHM/	The website is used for community driven development of PIHM
PIHM @ Github	https://github.com/pihmadmin	Source code, collaborative coding
PIHM @ figshare	http://dx.doi.org/10.6084/m9.figshare.1328521 http://dx.doi.org/10.6084/m9.figshare.1506789	Input files of PIHM
PIHM @ YouTube	https://www.youtube.com/PIHMgis	Tutorials
PIHM @ CSDMS Standard Names	http://csdms.colorado.edu/wiki/CSN_Examples	Standard names for the modeling community
PIHM @ OntoSoft	http://www.ontosoft.org/portal/#browse/Software-s4ru6v7tr0hc	Structured metadata to describe the PIHM software
PIHM @ Organic Data Science Framework	http://www.organicdatascience.org/ageofwater/	On-line collaborative tasks and workflows

TABLE III
MAIN EQUATIONS OF PIHM

Appendix II Main equations of PIHM.

Process	Governing equation/model	Original governing equations	Semi-discrete form *
Interception	Bucket model	$\frac{dh}{dt}=P-E_c-P_t$	$\frac{dh_{0I}}{dt}=P_v-E_c-P_t$
Snowmelt	Temperature index model	$\frac{dh}{dt}=P-E_{snow}-\Delta w$	$\frac{dh_{0S}}{dt}=P-E_{snow}-\Delta w$
Evapotranspiration	Penman-Monteith approach	$ET_0=\frac{\Delta(R_n-G)+\rho_a C_p\frac{(\varepsilon_s-\varepsilon_a)}{r_a}}{\Delta+\gamma(1+\frac{r_s}{r_a})}$	$ET_0=\frac{\Delta(R_n-G)+\rho_a C_p\frac{(\varepsilon_s-\varepsilon_a)}{r_a}}{\Delta+\gamma(1+\frac{r_s}{r_a})}$
Overland flow	St. Venant equation	$\frac{\partial h}{\partial t}+\frac{\partial(uh)}{\partial x}+\frac{\partial(vh)}{\partial y}=q$	$\frac{dh_1}{dt}=p_n-q^+-e+\sum_{j=1}^{3}q_j^s$
Unsaturated flow	Richards equation	$C(\Psi)\frac{\partial\Psi}{\partial t}=\nabla\cdot K(\Psi)\nabla(\Psi+Z)$	$\theta_s\frac{dh_2}{dt}=q^+-q^0$
Groundwater flow	Richards equation		$\theta_s\frac{dh_3}{dt}=q^0+\sum_{j=1}^{3}q_j^g$
Channel flow	St. Venant equation	$\frac{\partial h}{\partial t}+\frac{\partial(uh)}{\partial x}=q$	$\frac{dh_{4,5}}{dt}=p-e+\sum_{j=1}^{2}(q_l^s+q_l^g)+q_{in}^c-q_{out}^c$

* Notation: h_{0I} is the vegetation interception storage, P_v is the total precipitation, E_c is the evaporation from canopy interception. h_{0S} is the snow water equivalent storage, P is the solid precipitation water equivalent, Δw is snow-melting rate. Δ is the slope of the saturation vapor pressure-temperature relationship, R_n is net radiation at the vegetation surface, G is soil heat flux density, $\varepsilon_s-\varepsilon_a$ represents the air vapor pressure deficit, and ρ_a is the air density, C_p is specific heat of the air, γ is the psychometric constant, r_s and r_a are the surface and aerodynamic resistances. h_1 is the shallow water depth above the ground surface, p_n , q^+ and e are throughfall, infiltration, and evaporation, respectively, q_j^s is the normalized lateral flow rate from element *i* to its neighbor *j*. θ_s is the moisture content, h_2 is the unsaturated storage depth, h_3 is the groundwater depth, q^0 is flux between unsaturated-saturated zone, q_j^g is the normalized lateral groundwater flow rate from element i to its neighbor j. $h_{4,5}$ is depth of water in the channel or beneath the channel, , q_l^s and q_l^g are the lateral surface flow and groundwater interaction with the channel respectively from each side of the channel or beneath the channel, the upstream and downstream flow for each channel segment or beneath the channel are q_{in}^c and q_{out}^c respectively.

APPENDIX

See Tables II and III.

ACKNOWLEDGMENT

The authors would like to thank the associate editor Prof. N.-B. Chang and the reviewers for the attention and time they took to review the manuscript. Their insightful and constructive comments and questions were very useful in improving the earlier drafts.

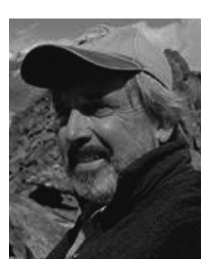
REFERENCES

[1] R. M. Maxwell *et al.*, "Surface-subsurface model intercomparison: A first set of benchmark results to diagnose integrated hydrology and feedbacks," *Water Resour. Res.*, vol. 50, no. 2, pp. 1531–1549, Feb. 2014.
[2] J. W. Kirchner, "Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology," *Water Resour. Res.*, vol. 42, no. 3, Mar. 2006, Art. ID. W03S04.
[3] C. Duffy, L. Leonard, G. Bhatt, X. Yu, and L. Giles, "Watershed reanalysis: Towards a national strategy for model-data integration," in *Proc. IEEE 7th Int. Conf. eScienceW*, 2011, pp. 61–65.
[4] NASA, LDAS|Land Data Assimilation Systems—NLDAS-2 Forcing Data Description/Information. (Accessed Mar. 4 2015). [Online]. Available: http://ldas.gsfc.nasa.gov/nldas/NLDAS2forcing.php
[5] S. Ahalt *et al.*, "Water science software institute: An open source engagement process," in *Proc. 5th Int. Workshop Softw. Eng. Comput. Sci. Eng.*, 2013, pp. 40–47.
[6] National Science Foundation, "Cyberinfrastructure vision for 21st century discovery," NSF Cyberinfrastructure Council, Arlington, VA, USA, Tech. Rep. NSF 07-28, 2007. [Online]. Available: http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf
[7] Y. Gil, M. Chan, B. Gomez, and B. Caron, "EarthCube: Past, present, and future," Nat. Sci. Found., Arlington, VA, USA, EarthCube Proj. Rep. EC-2014-3, Dec. 2014.
[8] C. A. Stewart *et al.*, "What is cyberinfrastructure," in *Proc. 38th Annu. ACM SIGUCCS Fall Conf.*, 2010, pp. 37–44.
[9] W. L. Nutter, "Determination of the head-discharge relationship for a sharp-crested compound weir and a sharp-crested parabolic weir," M.S. thesis, Department of Forest Hydrology, The Pennsylvania State Univ., University Park, PA, USA, 1964.
[10] J. A. Lynch, W. E. Sopper, E. S. Corbett, and D. W. Aurand, "Effects of management practices on the quality and quantity: The Penn. State Experimental watersheds," in *Proc. West Virginia Municipal Watershed Manage. Symp.*, Upper Darby, PA, USA, 1973, pp. 32–46.
[11] K. N. Eshleman, "A linear model of the effects of disturbance on dissolved nitrogen leakage from forested watersheds," *Water Resour. Res.*, vol. 36, no. 11, pp. 3325–3335, Nov. 2000.
[12] J. A. Lynch, "Effects of antecedent soil moisture on storm hydrographs," Ph.D. dissertation, School of Forestry Resources, The Pennsylvania State Univ., University Park, PA, USA, 1976.
[13] USFS, A Study of Forest Influences on Streamflow in the Ridge and Valley Province of Pennsylvania, 2015. (Accessed Mar. 4, 2015). [Online]. Available: http://www.fs.fed.us/ne/global/ltedb/catalogs/cat80.html
[14] C. J. Duffy, "Shale Hills Watershed Experiment," 2015. (Accessed Mar. 4, 2015). [Online]. Available: http://cataract.cee.psu.edu/shalehills
[15] Y. Qu and C. J. Duffy, "A semidiscrete finite volume formulation for multiprocess watershed simulation," *Water Resour. Res.*, vol. 43, no. 8, Aug. 2007, Art. ID. W08419.
[16] G. Bhatt, M. Kumar, and C. J. Duffy, "A tightly coupled GIS and distributed hydrologic modeling framework," *Environ. Model. Softw.*, vol. 62, pp. 70–84, Dec. 2014.
[17] W. Li, "Implementing the Shale Hills watershed model in application of PIHM," M.S. thesis, Dept. Civil Eng., Penn State Univ., State College, PA, USA, 2010.
[18] Y. Shi, K. J. Davis, C. J. Duffy, and X. Yu, "Development of a coupled land surface hydrologic model and evaluation at a critical zone observatory," *J. Hydrometeorol.*, vol. 14, no. 5, pp. 1401–1420, Oct. 2013.
[19] L. Leonard and C. Duffy, "Essential terrestrial variable data workflows for distributed water resources modeling," *Environ. Model. Softw.*, vol. 50, pp. 85–96, Dec. 2013.
[20] L. Leonard and C. Duffy, "Automating data-model workflows at a level 12 HUC scale: Watershed modeling in a distributed computing environment," *Environ. Model. Softw.*, vol. 61, pp. 174–190, Nov. 2014.
[21] S. D. Peckham, "The CSDMS standard names: Cross-domain naming conventions for describing process models, data sets and their associated variables," in *Proc. 7th Int. Congr. Environ. Model. Softw.*, San Diego, CA, USA, 2014.
[22] Y. Gil, V. Ratnakar, and D. Garijo, "OntoSoft: Capturing scientific software metadata," in *Proc. 8th ACM Int. Conf. Knowl. Capture*, Palisades, NY, USA, 2015, pp. 1–4.
[23] Y. Gil *et al.*, "Supporting open collaboration in science through explicit and linked semantic description of processes," in *Proc. 12th ESWC*, Portoroz, Slovenia, 2015, pp. 1–15.
[24] Y. Gil *et al.*, "A task-centered framework for computationally-grounded science collaborations," in *Proc. 11th IEEE Int. Conf. eScience*, Munich, Germany, 2015, pp. 1–10.
[25] F. Chen and J. Dudhia, "Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system—Part I: Model implementation and sensitivity," *Mon. Weather Rev.*, vol. 129, no. 4, pp. 569–585, Apr. 2001.
[26] M. Kumar, "Toward a hydrologic modeling system," Ph.D. dissertation, Dept. Civil Environ. Eng., Penn State Univ., State College, PA, USA, 2009.
[27] X. Yu, G. Bhatt, C. Duffy, and Y. Shi, "Parameterization for distributed watershed modeling using national data and evolutionary algorithm," *Comput. Geosci.*, vol. 58, pp. 80–90, Aug. 2013.
[28] N. Hansen and A. Ostermeier, "Completely derandomized self-adaptation in evolution strategies," *Evol. Comput.*. vol. 9, no. 2, pp. 159–195, Jun. 2001.
[29] M. Zreda *et al.*, "COSMOS: The COsmic-ray Soil Moisture Observing System," *Hydrol. Earth Syst. Sci.*, vol. 16, no. 11, pp. 4079–4099, Nov. 2012.
[30] D. Eissenstat and M. Kaye, "CZO Dataset: Shale Hills—Vegetation (2008–2012)—Tree Survey," 2012. Retrieved Jun. 12, 2015. [Online]. Available: http://criticalzone.org/shale-hills/data/dataset/2648/
[31] H. Lin, "Temporal stability of soil moisture spatial pattern and subsurface preferential flow pathways in the Shale Hills catchment," *Vadose Zone J.*, vol. 5, no. 1, pp. 317–340, Jan. 2005.
[32] S. Krause *et al.*, "Interdisciplinary perspectives on processes in the hyporheic zone," *Ecohydrology*, vol. 4, no. 4, pp. 481–499, Jul. 2011.
[33] G. A. Meehl *et al.*, "The WCRP CMIP3 multimodel dataset: A new era in climate change research," *Bull. Amer. Meteorol. Soc.*, vol. 88, no. 9, pp. 1383–1394, Sep. 2007.

Xuan Yu received the B.E. degree in water resources engineering from the China University of Geosciences, Beijing, China; in 2006, the M.E. degree from the China Institute of Water Resources and Hydropower Research, Beijing, China, in 2009; and the Ph.D. degree in civil and environmental engineering from The Pennsylvania State University, University Park, PA, USA.

He is currently a Postdoctoral Researcher with the Department Geological Sciences, University of Delaware, Newark, DE, USA.

His research interest includes watershed models, groundwater-surface water interaction. His current project is about the potential impacts of sea-level rise on contaminant migration in urban coastal zones.



Christopher Duffy is a Professor in water resources engineering with the Department of Civil and Environmental Engineering, The Pennsylvania State University, University Park, PA, USA. He held faculty appointments with Utah State University, Logan, UT, USA, in 1981–1989 and visiting appointments with Los Alamos National Laboratory in 1998–1999; Cornell University, Ithaca, NY, USA, in 1987–1988; the Ècole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, in 2006–2007. He was also a Visiting Scientist at the University of Bonn, Bonn, Germany, in 2015. He and his team have focused on developing the spatially distributed and physics-based computational code PIHM (The Penn State Integrated Hydrologic Model) for multiscale and multiprocess applications (http://www.pihm.psu.edu/) and an online national data service for access to geospatial watershed data (www.hydroterre.psu.edu) anywhere in the continental U.S.

He was a Senior Fellow with the Smithsonian Institution while in residence at the Smithsonian Environmental Research Center, in 2007 and a Visiting Senior Fellow at the University of Bristol, U.K., for 2014–2016.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Yolanda Gil received the M.S. and Ph. D. degrees in computer science from Carnegie Mellon University, Pittsburgh, PA, USA.

She is the Director of Knowledge Technologies and the Associate Division Director at the Information Sciences Institute of the University of Southern California, Marina del Rey, CA, USA, where she is also a Research Professor in computer science. Her research interests include intelligent user interfaces, knowledge-rich problem solving, and the semantic web. A recent focus area is collaborative analysis of scientific data through semantic workflows.

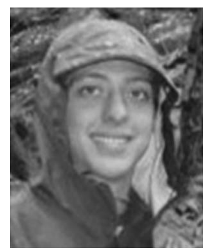
Dr. Gil is an elected Fellow of the American Association of Artificial Intelligence. She initiated and chaired the World Wide Web Consortium (W3C) Provenance Group that led to a community standard in this area. She is the Chair of the Association for Computing Machinery's Special Interest Group on Artificial Intelligence.

Lorne Leonard received undergraduate degrees from The University of Western Australia, Perth, Australia; the master's degree from The University of Illinois at Champaign-Urbana, Urbana, IL, USA; and the Ph.D. degree in civil and environmental engineering from The Pennsylvania State University, University Park, PA, USA.

He is currently with the Department of Civil and Environmental Engineering, The Pennsylvania State University. His research interests include large-scale hydrology, high-performance scientific computing, visualization, and geographic information sciences.

Gopal Bhatt received the Ph.D. degree in civil and environmental engineering from The Pennsylvania State University (Penn State), University Park, PA, USA.

He is a Research Associate with the Penn State Institutes of Energy and the Environment. He is currently working with the Chesapeake Bay Program Office, Annapolis, MD, under a U.S. Environmental Protection Agency Cooperative Agreement with Penn State. His research interest is the development and refinement of Bay Watershed Model operation, calibration, analysis, verification, and uncertainty tools and carrying out modeling team support and outreach. His research expertise and interests include distributed hydrologic model, geographic information system, computational science, numerical methods, climate change, wetland ecohydrology, coastal watershed hydrology, and modeling residence times of the watershed.



Evan Thomas received the M.S. degree in civil and environmental engineering from The Pennsylvania State University, University Park, PA, USA.

He is currently an Environmental Engineer with Amec Foster Wheeler, Madison, WI, USA. His research interests included stable water isotope hydrology and time-series analysis. He is currently involved in large contaminated sediment remediation projects.