UC Berkeley
UC Berkeley Previously Published Works

Title

Emerging technologies and radical collaboration to advance predictive understanding of watershed hydrobiogeochemistry

Permalink

https://escholarship.org/uc/item/1h45s0mk

Journal

Hydrological Processes, 34(15)

ISSN

0885-6087

Authors

Hubbard, Susan S
Varadharajan, Charuleka
Wu, Yuxin
et al.

Publication Date

2020-07-15

DOI

10.1002/hyp.13807

Peer reviewed

WILEY

Emerging technologies and radical collaboration to advance predictive understanding of watershed hydrobiogeochemistry

Susan S. Hubbard [iD] | Charuleka Varadharajan | Yuxin Wu |
Haruko Wainwright [iD] | Dipankar Dwivedi

Earth and Environmental Sciences Area, Berkeley Lab, Berkeley, California

Correspondence
Susan S. Hubbard, Earth and Environmental Sciences Area, Berkeley Lab, Berkeley, CA 94720.
Email: sshubbard@lbl.gov

1 | MOTIVATION

Increasing population and resource-intensive lifestyles are driving enhanced demands for clean water, food, and energy. In parallel, land-use change, climate change, and perturbations—including drought, floods, fires, and early snowmelt—are significantly reshaping interactions within watersheds throughout the world. While watersheds are the Earth's key functional unit for assessing and managing water resources, hydrological processes in watersheds also mediate biogeochemical interactions that support terrestrial life on Earth (Kaushal, Gold, Bernal, & Tank, 2018; National Research Council, 2012). Although society is dependent upon clean water availability, tractable prediction of watershed hydrobiogeochemical behavior, including watershed response to perturbations, remains a challenge. Central to the challenge are complex, multiscale interactions between plants, microorganisms, organic matter, minerals, dissolved constituents, and migrating fluids, which occur within and across bedrock-to-canopy compartments and along extensive lateral gradients of a watershed. Several recent community reports have synthesized formidable challenges associated with watershed science and technology (AGU, 2018; Blöschl et al., 2019). Here, we discuss emerging technologies and collaboration modes that are critical for developing generalizable insights about and predictive understanding of complex watershed hydrobiogeochemical behavior, which are important for underpinning optimized natural resource management.

Recent developments in field observatories and open-science principles provide foundational pillars for advancing predictive understanding of watershed hydrobiogeochemistry using emerging technologies. Field observatories have fostered crossdisciplinary collaboration and provided platforms for quantifying hydrological, biological, geological, geochemical, and atmospheric processes and their couplings (Bogena, White, Bour, Li, & Jensen, 2018). Observatory networks in the United States include the Critical Zone Observatories (Brantley et al., 2017), National Ecological Observatory Network (Loescher, Kelly, & Lea, 2017), the Long-Term Ecological Research Network (Hobbie, Carpenter, Grimm, Gosz, & Seastedt, 2003), and the Department of Energy (DOE) Watershed Network (U.S. DOE, 2019). Select international observatory networks include the German Terrestrial Environmental Observatories (Zacharias et al., 2011), the French OZCAR network (Gaillardet et al., 2018), and the Chinese observatories (Li et al., 2013). The observatories are complemented by long-term distributed measurement suites, such as the US Geological Survey stream discharge and concentration measurements (NASEM, 2018a) and the DOE AmeriFlux network carbon, water, and energy flux measurements (Novick et al., 2018). Open-science concepts (NASEM, 2018a), which have recently started to permeate watershed science, provide another foundational pillar. While the open-data FAIR (findable, accessible, interoperable, and reusable) principles (Wilkinson et al., 2016) are perhaps the most recognized aspect of open-science,

open-science concepts are also critical for generating and sharing data, knowledge, and models in a manner that promotes transferability and generalizability across watershed networks (U.S. DOE, 2019).

2 | EMERGING TECHNOLOGIES POISED TO ADVANCE WATERSHED HYDROBIOGEOCHEMISTRY

Several emerging technologies hold potential to greatly enhance predictive understanding of watershed hydrobiogeochemical behavior, including machine learning (ML) and artificial intelligence, exascale computing, 5G wireless communications, and cloud data storage and compute capacity. Deep learning using large neural networks with multilayered structures can identify abstract concepts about datasets (Schmidhuber, 2015); these methods have been useful in other fields to discover physical concepts from data. Several recent publications have also illustrated the potential of ML for advancing Earth sciences (Bergen, Johnson, Maarten, & Beroza, 2019; Nearing et al., 2020; Reichstein et al., 2019; Shen, 2018). Exascale computing will provide computing systems capable of at least a quintillion (or billion billion) calculations per second, representing a thousand-fold increase over the first petascale computers that came into operation about a decade ago. Teams of scientists are currently developing exascale-ready codes for several applications (Alexander et al., 2020), including for Earth sciences (Johansen et al., 2017). In addition, cloud platforms (such as Google Earth Engine and Amazon Web Services) are enabling an era of "Big Data" in the Earth sciences by providing on-demand storage, networking, and mid-range computing capacity. These developments are already spurring advances in geospatial data processing and bioinformatics (Yang, Huang, Li, Liu, & Hu, 2017). 5G technology refers to the fifth generation of digital wireless communication technologies, which will offer significantly enhanced data rates and low latency. 5G will also offer enhanced connectivity of massively parallel Internet of things devices—up to 1 million per square kilometer (Forbes, 2019).

Below, we provide examples of how emerging technologies are starting to be used to advance three key elements: watershed hydrobiogeochemical characterization, data management and informatics, and modeling.

2.1 | Watershed characterization

The desire to characterize and monitor watershed hydrobiogeochemical dynamics at increasing spatial and temporal resolutions has driven an explosion of observational technologies and platforms. For example, fiber-optic distributed sensors can now autonomously measure temperature and strain with very high spatiotemporal resolution in terrestrial and aquatic systems (Ajo-Franklin et al., 2017; Joe, Yun, Jo, Jun, & Min, 2018; Slater et al., 2010), and fiber-based approaches for sensing chemical and biological properties are in development (Ding et al., 2015; Lu, Thomas, & Hellevang, 2019). New sensing strategies are being tested to noninvasively monitor active plant-root functions in situ (Benjamin et al., 2020; Peruzzo et al., 2020) and to monitor nutrient fluxes (MacDonald, Levison, & Parker, 2017). Autonomous unmanned aerial vehicles (UAVs) equipped with various instruments can now sense previously difficult-to-reach environments in high-resolution. UAV approaches complement airborne and satellite imaging strategies (McCabe et al., 2017), all enabled through cloud storage. Remotely sensed, spatially distributed "data layers" of watershed compartments and associated properties now enable creation of 3D bedrock-to-canopy watershed "digital twins" (Wainwright et al., 2019). Networked sensing systems can now coincidentally and autonomously monitor bedrock-through-canopy fluxes (Dafflon et al., 2017), providing a "window" to remotely track fluxes across watershed compartments.

ML is starting to show promise for advancing watershed characterization using these and other diverse datasets (Ahmad, Kalra, & Stephen, 2010; Oroza, Zheng, Glaser, Tuia, & Bales, 2016). One promising approach has focused on using ML-based spatial clustering approaches for characterizing functional zones: parcels within a landscape that have unique distributions of properties relative to neighboring regions that influence how that zone functions from a hydrobiogeochemical perspective. The approach can include geomorphoric, topographic, vegetation, hydrogeologic, geochemical, and other properties that may influence behavior. For example, Hubbard et al. (2013) and Wainwright et al. (2015) used ML with geophysical and other datasets to identify functional zones in an Arctic tundra and to quantify the zone-based property suites important for carbon fluxes. Wainwright et al. (2019) used a suite of remotely sensed bedrock-through-canopy watershed data layers to identify zonation within a mountainous watershed. Current investigations are focused on quantifying how properties associated with distinct functional zones govern water and nitrogen export in response to snow dynamics and in turn contribute to the aggregated watershed concentration-discharge signature (Hubbard et al., 2018).

2.2 | Watershed data cyberinfrastructure and informatics

Management, sharing, and reuse of environmental data have greatly increased over the past decade in response to their increasing volume, diversity, and complexity (Blair et al., 2019; Rode et al., 2016). In parallel, advances in environmental data infrastructure, cloud computing, and ML have dramatically improved the ability to store and utilize diverse data for watershed science.

Data from research efforts are becoming available through open-data movements based on FAIR principles, which advocate the use of metadata and standards. Data from operational and experimental monitoring networks are now widely available through data systems such as the USGS National Water Information System, the US interagency water quality portal (Blodgett, Read, Lucido, Slawecki, & Young, 2016), and the European Union's Water Information System for Europe (Hering et al., 2010). Several repositories now enable

scientists to easily archive and publish data with essential metadata and provide easy-to-use watershed data access mechanisms, such as interactive portals and web service application programming interfaces. Examples in the US include the NSF-supported Hydroshare (Horsburgh et al., 2016) for hydrological data, and the DOE-supported ESS-DIVE (Varadharajan et al., 2019) for watershed hydrological and biogeochemical data from experimental and modeling research.

In parallel, cyberinfrastructure for synthesizing and processing these increasing amounts of data are being developed. For example, Varadharajan et al. (2019) recently demonstrated the utility of watershed-centric data infrastructure and tools using diverse hydrological, climate, geochemical, and biological data from a mountainous watershed in Colorado. The end-to-end data pipeline required development and use of several novel methods, including new software that allowed seamless, real-time integration of data stemming from a variety of sources with differing metadata, source formats, and vocabularies, through the use of a data broker (BASIN-3D). BASIN-3D, which also provides tools for semiautomated QA/QC, has enabled rapid synthesis of watershed time-series observations and interactive visualizations (Hubbard et al., 2018). The Consortium of Universities for the Advancement of Hydrologic Science hydrologic information system pulls time-series data from over 95 sources, which can be accessed via the Hydroclient interactive portal and the WaterOneFlow web services (Horsburgh et al., 2016). Other efforts, such as Pangeo (an ecosystem of open-source, interoperable, scalable tools), are enabling big data integration for HPC using cloud services (Eynard-Bontemps, Abernathey, Hamman, Ponte, & Rath, 2019).

The availability of these data and tools is enabling a new paradigm wherein data-driven methods are being used to probe hydrobiogeochemical scaling, similarity and function, as well as to generate and test new hypotheses (Peters-Lidard et al., 2017). Classical methods, such as statistical time-series analyses and concentration-discharge relationships, are being applied to large, regional data products to determine streamflow and water quality trends across catchments with different characteristics (Godsey, Hartmann, & Kirchner, 2019; Murphy & Sprague, 2019). Data mining and classification algorithms, such as ensemble decision trees and unsupervised clustering, can recognize patterns in coupled human–natural system behavior and watershed response to disturbance (Hamshaw, Dewoolkar, Schroth, Wemple, & Rizzo, 2018; Smith, Knight, & Fendorf, 2018). The application of mutual information theory and causal inference approaches is being used to derive explanatory relationships between environmental variables and hypothesis testing (Goodwell & Kumar, 2017; Nearing, Ruddell, Bennett, Prieto, & Gupta, 2020).

2.3 | Watershed hydrobiogeochemical modeling

While process-based, integrated numerical models have been used to predict watershed hydrological behavior (Fatichi et al., 2016; Maxwell et al., 2014; Troch, Carrillo, Sivapalan, Wagener, & Sawicz, 2013), a significant challenge remains to develop

computationally efficient capabilities that also incorporate reactive transport. Advancing a robust predictive understanding of watershed hydrobiogeochemical behavior requires numerical representation and coupling of hydrological and biogeochemical processes—from reaction (mm–cm) to watershed (km) scales and across bedrock-to-canopy compartments and terrestrial–aquatic interfaces (Bao, Li, Shi, & Duffy, 2017; Li, 2019; Li et al., 2017; Steefel, 2019; Troch et al., 2009). Here, we describe the potential for emerging technologies and strategies to meet this objective, including model interoperability, computational meshing strategies and architectures, and ML-based approaches.

There are tremendous opportunities to advance interoperability between models that solve for integrated (surface–subsurface) hydrology and reactive transport. While hydrology and reactive transport modeling have historically evolved along different paths, disparate communities have worked together in recent years to couple hydrologic and reactive transport models. Examples of models used for such coupling include PFLOTRAN (Hammond & Lichtner, 2010), advanced terrestrial simulator (ATS; Coon, Moulton, & Painter, 2016), ParFlow (Kollet & Maxwell, 2008), and CrunchFlow (Steefel et al., 2015). While these and other codes each offer certain strengths, none can currently simulate full multiscale, multiphysics watershed reactive transport (Dwivedi et al., 2016; Steefel et al., 2015). Recently, adoption of model interoperability approaches has enabled researchers to take advantage of select strengths of different codes to enhance predictive understanding (Heroux et al., 2020). For example, while ATS can simulate coupled surface–subsurface flow but not reactive transport, new interfaces now allow ATS to access the reaction engines of either the PFLOTRAN or CrunchFlow (Coon et al., 2016) to enable simulation of surface–subsurface hydrobiogeochemical processes in watersheds.

Following the hydrological modeling community's lead in considering the need for hyperresolution (0.1–1 km resolution; Bierkens et al., 2015), efforts are underway to advance the ability to similarly simulate watershed reactive transport in high resolution using leadership-class supercomputers. For example, Dwivedi, Arora, Steefel, Dafflon, and Versteeg (2018) used PFLOTRAN to numerically examine hot spots and hot moments influencing nitrogen cycling in a small watershed floodplain in 3D and with meter-scale spatial resolutions. This simulation required 250,000 CPU hours (or 24 wall clock hours) on the National Energy Research Scientific Computing Center supercomputer to simulate processes occurring over one water year. As exascale computers come online, the reaction calculations are expected to undergo a substantial speed up, rendering watershed-scale reactive transport and associated uncertainty quantification strategies not only feasible, but practical from a research perspective.

As exascale simulation capabilities are not expected to be accessible to the wider community and practitioners for many years, several strategies hold potential for reducing the computational burden required for simulating watershed reactive transport using current leadership-class computers. For example, approaches for adjusting the resolution in computational grids, such as including static and adaptive mesh refinement methods (AMR; Blayo & Debreu, 1999; Wang, Liu, & Kumar, 2018), offer potential for simulating small-scale reactive

hotspots and their influence on larger system behavior in a computationally efficient manner. The ability to employ variable resolution in mechanistic watershed reactive transport models, allowing codes to "telescope" into regions that are rapidly evolving, may provide a path forward for balancing accuracy and tractability associated with stimulating watershed reactive transport processes. AMR approaches have recently been developed to improve the efficiency of watershed hydrologic simulations (Wang et al., 2018) and to enable simulation of coupled hydrobiogeochemical processes (Özgen-Xian et al., 2020).

Recently, data-driven approaches have gained momentum in watershed modeling because of their computational efficiency and agility to incorporate diverse and multiscale data that are often difficult to incorporate into current process-based models (Shen, 2018). The value of ML for watershed modeling has been illustrated by applications focused on streamflow prediction (Kratzert, Klotz, Brenner, Schulz, & Herrnegger, 2018), early warning of droughts and floods (Mosavi, Ozturk, & Chau, 2018; Park, Im, Jang, & Rhee, 2016), groundwater level fluctuations (Müller et al., 2019), and chemical equilibrium calculations (Leal, Kulik, & Saar, 2017). However, as data-driven models are developed directly from observations, their effectiveness is limited when data are sparse. Data-driven modeling also does not provide insight about processes, which limits transferability of results.

A strategy for taking advantage of increasing data availability while honoring mechanistic process representation in a computationally efficient manner is hybrid modeling, also known as physics-based ML (Bergen et al., 2019; Reichstein et al., 2019). Hybrid-modeling strives to marry complementary aspects of mechanistic process-based models and ML, data mining, and genetic algorithms. Given the complexity of hydrobiogeochemical data and processes that occur across scales and compartments of a watershed, this strategy holds significant promise for advancing predictive understanding of watershed hydrobiogeochemical behavior. Nearing, Kratzert, et al. (2020) advocated the importance of integrating ML into hydrological workflows. We contend that ML, and in particular the hybrid modeling strategy, holds significant potential for advancing prediction of watershed hydrobiogeochemistry.

3 | LOOKING FORWARD

Early research is illustrating the promise of ML, 5G, computational strategies and architectures, and cloud-based technologies for improving watershed characterization, data handling, and modeling. We envision a future where the emerging technologies will be able to seamlessly unify sensing systems, data infrastructure, and computational tools to allow near real-time, autonomous communication, and feedback. To realize this vision, "codesign" strategies are needed whereby watershed sensing, data, and modeling systems are "born" to communicate with each other across multiple scales (Varadharajan et al., 2019). Enabled by the emerging technologies, codesign strategies hold potential for rapid synthesis and assimilation of increasingly diverse and autonomous data streams into data systems and models, and near real-time feedback

from models to observing systems, including instructions regarding what data should be collected where. We urge the community to work together to advance observation-data-modeling system codesign strategies, with an aim to improve watershed characterization and prediction within and across observatories, and eventually enable near real-time information for resource managers. We recognize that building and maintaining elaborate codesign strategies is challenging for individual research efforts, and that incorporation of technologies alone will not allow us to address the many existing watershed scientific questions. This brings us to a second recommendation.

We submit that new modes of "radical" collaboration are needed to facilitate teams-of-teams to work together in a coordinated fashion across watershed networks using "Open Science by Design" principles (NASEM, 2018b; U.S. DOE, 2019). The collaborations could focus on addressing specific watershed science questions; advancing common measurements, models and codesign strategies; and discovering generalizable metrics and transferable insights. For example, collaborations working across diverse, globally distributed watersheds that span hydroclimatic and property gradients could investigate questions such as: how do different types of watersheds respond to different stressors, such as climate change, droughts, floods, wildfire, and land-use? How will multiple stressors impact sustainability of water, food, energy systems that rely on water? Can generalizable metrics of resilience be identified and tracked? What is the minimum but sufficient amount of information needed to predict watershed behavior at temporal and spatial scales critical for underpinning resource management decisions? In addition to working across gradients, international collaborations could use distributed watersheds to address more focused questions relevant to specific types of watersheds, such as pristine mountainous watersheds, contaminated watersheds, agriculture-dominated watersheds, or urban watersheds. Figure 1 illustrates how investigations carried out across geographically distributed mountainous or agriculture-dominated watersheds could be useful for exploring the influence of various stressors on the functioning of these systems, and the associated impacts on water supply and water quality, power generation, agricultural productivity, and other societal benefits.

Systematic incorporation of emerging technologies and adoption of radical modes of collaboration require substantial coordination, resources and commitment to overcome technical, social, and organizational barriers. We are encouraged by the many recent efforts focused on advancing collaborations and tools across watershed communities, observatories, and government agencies (U.S. DOE, 2020; NSF critical zone collaborative network opportunity; U.S. DOE, in prep.). We are also encouraged by a new generation of watershed scientists who embrace open watershed science philosophies and approaches to advance connectivity across scales, sites, disciplines, and nations (Arora et al., 2019; Wymore et al., 2017). As resource managers struggle to make increasingly difficult decisions in the coming decades, we envision that the emerging technologies and radical collaborations described here will mobilize the scientific enterprise toward providing actionable information over space and time scales useful for such decisions (Figure 1).
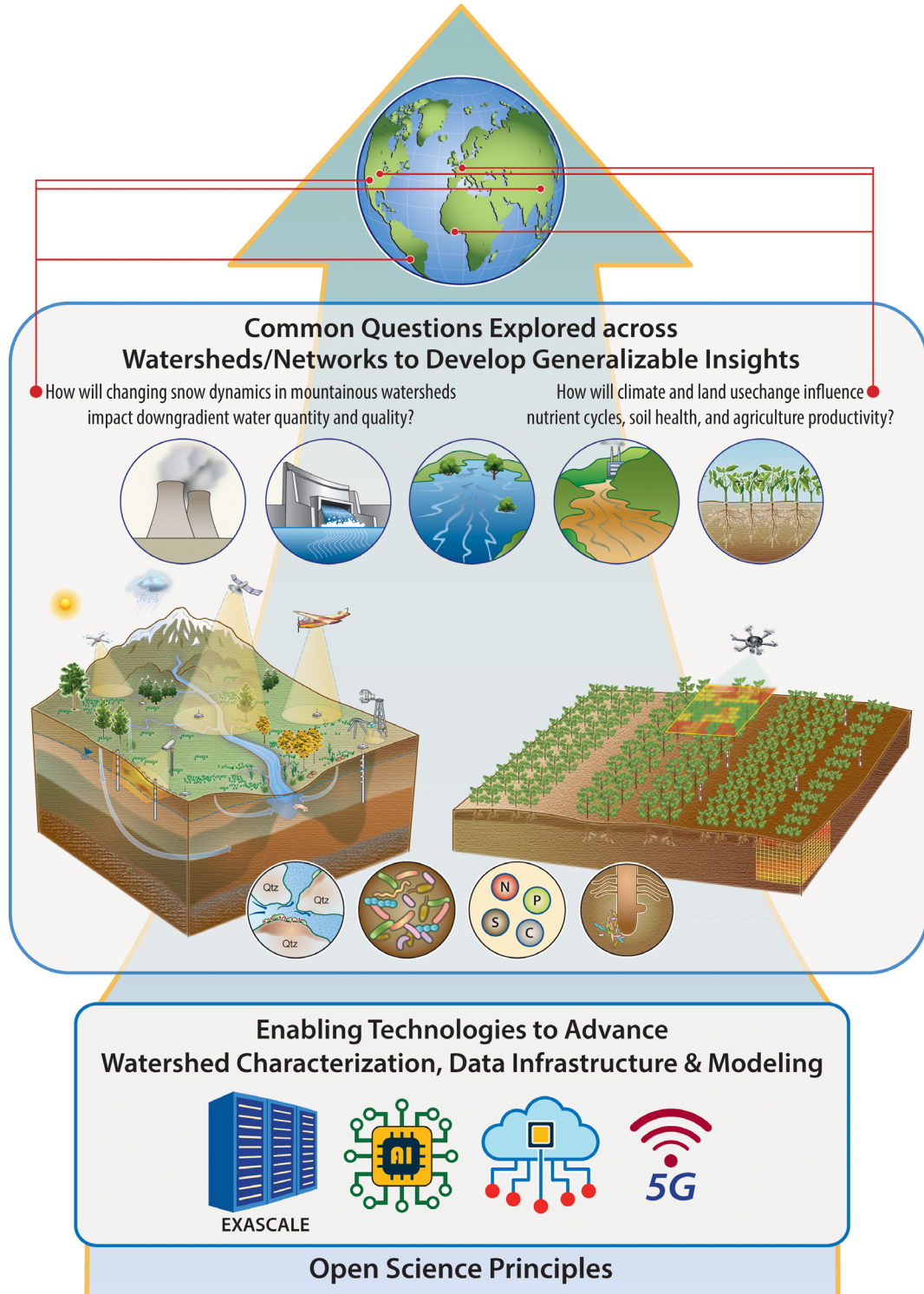
FIGURE 1 Enabling technologies and new modes of collaboration, together with open science principles and watershed networks, hold potential to transform our ability to address complex scientific questions and to develop generalizable insights and predictive understanding of watershed hydrobiogeochemical behavior critical for resource management

ORCID

Susan S. Hubbard [ID] https://orcid.org/0000-0003-2966-5631
Haruko Wainwright [ID] https://orcid.org/0000-0002-2140-6072

REFERENCES

AGU, Hydrology Newsletter, Dec. (2018). https://connect.agu.org/hydrology/newsletter

Ahmad, S., Kalra, A., & Stephen, H. (2010). Estimating soil moisture using remote sensing data: A machine learning approach. *Advances in Water Resources*, 33(1), 69–80. https://doi.org/10.1016/j.advwatres.2009.10.008

Ajo-Franklin, J., Dou, S., Daley, T., Freifeld, B., Robertson, M., Ulrich, C., ... & Wagner, A. (2017). Time-lapse surface wave monitoring of permafrost thaw using distributed acoustic sensing and a permanent automated seismic source. *SEG Technical Program Expanded Abstracts 2017* (pp. 5223–5227). Society of Exploration Geophysicists, SEG International Exposition and 87th Annual Meeting.

Alexander, F., Almgren, A., Bell, J., Bhattacharjee, A., Chen, J., Colella, P., ... Dubey, A. (2020). Exascale applications: Skin in the game. *Philosophical Transactions of the Royal Society A*, 378(2166), 20190056. https://doi.org/10.1098/rsta.2019.0056

Arora, B., Ireson, A., Bouteiller, C. L., Hector, B., Ali, G., Stegen, J., ... & Groh, J. (2019). Toward an International Critical Zone Network-of-Networks for the Next Generation Through Shared Science, Tools, Data, and Philosophy. Workshop SCIWS19, American Geophysical Union Fall Meeting, San Francisco 2019.

Bao, C., Li, L., Shi, Y., & Duffy, C. (2017). Understanding watershed hydrogeochemistry: 1. Development of RT-Flux-PIHM. *Water Resources Research*, 53(3), 2328–2345. https://doi.org/10.1002/2016WR018934

Benjamin, M., Peruzzo, L., Boaga, J., Cenni, N., Schmutz, M., Wu, Y., ... Cassiani, G. (2020). Time-lapse monitoring of root water uptake using electrical resistivity tomography and mise-à-la-masse: A vineyard infiltration experiment. *The Soil*, 6(1), 95–114.

Bergen, K. J., Johnson, P. A., Maarten, V., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid Earth geoscience. *Science*, 363(6433), eaau0323. https://doi.org/10.1126/science.aau0323

Bierkens, M. F., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., ... Flörke, M. (2015). Hyper-resolution global hydrological modelling: What is next? "Everywhere and locally relevant". *Hydrological Processes*, 29(2), 310–320. https://doi.org/10.1002/hyp.10391

Blair, G., Henrys, P., Leeson, A., Watkins, J., Eastoe, E., Jarvis, S., & Young, P. (2019). Data science of the natural environment: A research roadmap. *Frontiers in Environmental Science*, 7, 121. https://doi.org/10.3389/fenvs.2019.00121

Blayo, E., & Debreu, L. (1999). Adaptive mesh refinement for finite-difference ocean models: First experiments. *Journal of Physical Oceanography*, 29(6), 1239–1250. https://doi.org/10.1175/1520-0485(1999)029<1239:AMRFFD>2.0.CO;2

Blodgett, D., Read, E., Lucido, J., Slawecki, T., & Young, D. (2016). An analysis of water data systems to inform the open water data initiative. *JAWRA Journal of the American Water Resources Association*, 52(4), 845–858. https://doi.org/10.1111/1752-1688.12417

Blöschl, G., Bierkens, M. F., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., ... Stumpp, C. (2019). Twenty-three unsolved problems in hydrology (UPH)–a community perspective. *Hydrological Sciences Journal*, 64(10), 1141–1158. https://doi.org/10.1080/02626667.2019.1620507

Bogena, H. R., White, T., Bour, O., Li, X., & Jensen, K. H. (2018). Toward better understanding of terrestrial processes through long-term hydrological observatories. *Vadose Zone Journal*, 17(1), 1–10. https://doi.org/10.2136/vzj2018.10.0194

Brantley, S. L., McDowell, W. H., Dietrich, W. E., White, T. S., Kumar, P., Anderson, S. P., ... Grant, G. (2017). Designing a network of critical zone observatories to explore the living skin of the terrestrial Earth. *Earth Surface Dynamics*, 5, 841–860. https://doi.org/10.5194/esurf-5-841-2017

Coon, E. T., Moulton, J. D., & Painter, S. L. (2016). Managing complexity in simulations of land surface and near-surface processes. *Environmental Modelling & Software*, 78, 134–149. https://doi.org/10.1016/j.envsoft.2015.12.017

Dafflon, B., Oktem, R., Peterson, J., Ulrich, C., Tran, A. P., Romanovsky, V., & Hubbard, S. (2017). Coincident above and below-ground autonomous monitoring strategy: Development and use to monitor Arctic ecosystem freeze thaw dynamics. *Journal of Geophysical Research: Biogeosciences*, 122, 1321–1342. https://doi.org/10.1002/2016jg003724

Ding, Z., Du, Y., Liu, T., Liu, K., Feng, B., & Jiang, J. (2015). Distributed optical fiber current sensor based on magnetostriction in OFDR. *IEEE Photonics Technology Letters*, 27(19), 2055–2058.

Dwivedi, D., Arora, B., Steefel, C. I., Dafflon, B., & Versteeg, R. (2018). Hot spots and hot moments of nitrogen in a riparian corridor. *Water Resources Research*, 54(1), 205–222. https://doi.org/10.1002/2017WR022346

Dwivedi, D., Steefel, C., Siirila-Woodburn, E., Kallemov, B., Moulton, J. D., Kikinzon, E., ... & Maxwell, R. (2016). Testing code interoperability and productivity on modeling integrated surface subsurface water flow and biogeochemical cycling in the hyporheic zone – IDEAS use case 1. Annual joint investigators meeting of the Department of Energy's Office of Biological and Environmental Research (BER)At: Potomac, Maryland. https://doi.org/10.13140/RG.2.1.2515.5441

Eynard-Bontemps, G., Abernathey, R., Hamman, J., Ponte, A., & Rath, W. (2019). The PANGEO big data ecosystem and its use at CNES. *Proceedings of the 2019 Conference on Big Data from Space (BiDS'2019), EUR 29660 EN* (pp. 49–52). Luxembourg: Publications Office of the European Union.

Fatichi, S., Vivoni, E., Ogden, F. L., Ivanov, V. Y., Mirus, B., Gochis, D., ... Tarboton, D. (2016). An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. *Journal of Hydrology*, 537, 45–60. https://doi.org/10.1016/j.jhydrol.2016.03.026

Forbes Magazine (2019). Retrieved from https://www.forbes.com/sites/forbestechcouncil/2019/11/18/the-5g-iot-revolution-is-coming-heres-what-to-expect/#7ce7f33b6abf.

Gaillardet, J., Braud, I., Hankard, F., Anquetin, S., Bour, O., Dorfliger, N., ... Gourcy, L. (2018). OZCAR: The French network of critical zone observatories. *Vadose Zone Journal*, 17(1), 1–24.

Godsey, S. E., Hartmann, J., & Kirchner, J. W. (2019). Catchment chemostasis revisited: Water quality responds differently to variations in weather and climate. *Hydrological Processes*, 33(24), 3056–3069.

Goodwell, A. E., & Kumar, P. (2017). Temporal Information Partitioning Networks (TIPNets): A process network approach to infer ecohydrologic shifts. *Water Resources Research*, 53(7), 5899–5919. https://doi.org/10.1002/2016wr020218

Hammond, G. E., & Lichtner, P. C. (2010). Field-scale model for the natural attenuation of uranium at the Hanford 300 Area using high-performance computing. *Water Resources Research*, 46(9), 1–31. https://doi.org/10.1029/2009WR008819

Hamshaw, S. D., Dewoolkar, M. M., Schroth, A. W., Wemple, B. C., & Rizzo, D. M. (2018). A new machine-learning approach for classifying hysteresis in suspended-sediment discharge relationships using high-frequency monitoring data. *Water Resources Research*, *54*, 4040–4058. https://doi.org/10.1029/2017WR022238

Hering, D., Borja, A., Carstensen, J., Carvalho, L., Elliott, M., Feld, C. K., ... Solheim, A. L. (2010). The European Water Framework Directive at the age of 10: A critical review of the achievements with recommendations for the future. *Science of the Total Environment*, *408*(19), 4007–4019.

Heroux, M. A., Carter, J., Thakur, R., McInnes, L., Ahrens, J., Munson, T., ... & Vetter, J. S. (2020). *ECP software technology capability assessment report* (No. ORNL/TM-2020/1457). Oak Ridge National Lab. (ORNL), Oak Ridge, TN (United States). Retrieved from https://info.ornl.gov/sites/publications/Files/Pub137158.pdf

Hobbie, J. E., Carpenter, S. R., Grimm, N. B., Gosz, J. R., & Seastedt, T. R. (2003). The US long term ecological research program. *Bioscience*, *53*(1), 21–32.

Horsburgh, J. S., Morsy, M. M., Castronova, A. M., Goodall, J. L., Gan, T., Yi, H., Stealey M. J. & Tarboton, D. G. (2016). Hydroshare: Sharing diverse environmental data types and models as social objects with application to the hydrology domain. *Journal of the American Water Resources Association (JAWRA)*, *52*(4), 873–889. https://doi.org/10.1111/1752-1688.12363

Hubbard, S. S., Gangodagamage, C., Dafflon, B., Wainwright, H., Peterson, J., Gusmeroli, A., ... Tweedie, C. (2013). Quantifying and relating land-surface and subsurface variability in permafrost environments using LiDAR and surface geophysical datasets. *Hydrogeology Journal*, *21*(1), 149–169. https://doi.org/10.1007/s10040-012-0939-y

Hubbard, S. S., Williams, K. H., Agarwal, D., Banfield, J., Beller, H., Bouskill, N., ... Falco, N. (2018). The East River, Colorado, Watershed: A mountainous community testbed for improving predictive understanding of multiscale hydrological–biogeochemical dynamics. *Vadose Zone Journal*, *17*(1), 1–25. https://doi.org/10.2136/vzj2018.03.0061

Joe, H. E., Yun, H., Jo, S. H., Jun, M. B., & Min, B. K. (2018). A review on optical fiber sensors for environmental monitoring. *International Journal of Precision Engineering and Manufacturing-Green Technology*, *5*(1), 173–191. https://doi.org/10.1007/s40684-018-0017-6

Johansen, H., Rodgers, A., Petersson, N. A., McCallen, D., Sjogreen, B., & Miah, M. (2017). Toward exascale earthquake ground motion simulations for near-fault engineering analysis. *Computing in Science & Engineering*, *19*(5), 27–37. https://doi.org/10.1109/MCSE.2017.3421558

Kaushal, S. S., Gold, A. J., Bernal, S., & Tank, J. L. (2018). Diverse water quality responses to extreme climate events: An introduction. *Biogeochemistry*, *141*(3), 273–279. https://doi.org/10.1007/s10533-018-0527-x

Kollet, S. J., & Maxwell, R. M. (2008). Capturing the influence of groundwater dynamics on land surface processes using an integrated, distributed watershed model. *Water Resources Research*, *44*(2), 1–18. https://doi.org/10.1029/2007WR006004

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences*, *22*(11), 6005–6022.

Leal, A. M., Kulik, D. A., & Saar, M. O. (2017). Ultra-fast reactive transport simulations when chemical reactions meet machine learning: Chemical equilibrium, 1–25. *arXiv*, preprint arXiv:1708.04825. https://arxiv.org/pdf/1708.04825.pdf

Li, L. (2019). Watershed reactive transport. *Reviews in Mineralogy and Geochemistry*, *85*(1), 381–418. https://doi.org/10.2138/rmg.2018.85.13

Li, L., Maher, K., Navarre-Sitchler, A., Druhan, J., Meile, C., Lawrence, C., ... Jin, L. (2017). Expanding the role of reactive transport models in critical zone processes. *Earth-Science Reviews*, *165*, 280–301. https://doi.org/10.1016/j.earscirev.2016.09.001

Li, X., Cheng, G., Liu, S., Xiao, Q., Ma, M., Jin, R., ... Wen, J. (2013). Heihe watershed allied telemetry experimental research (HiWATER): Scientific objectives and experimental design. *Bulletin of the American Meteorological Society*, *94*(8), 1145–1160. https://doi.org/10.1175/BAMS-D-12-00154.1

Loescher, H. W., Kelly, E. F., & Lea, R. (2017). National Ecological Observatory Network: Beginnings, programmatic and scientific challenges, and ecological forecasting. In A. Chabbi & H. W. Loescher (Eds.), *Terrestrial ecosystem research infrastructures*. Boca Raton, FL: CRC Press.

Lu, X., Thomas, P. J., & Hellevang, J. O. (2019). A review of methods for fiber-optic distributed chemical sensing. *Sensors*, *19*(13), 2876.

MacDonald, G., Levison, J., & Parker, B. (2017). On methods for in-well nitrate monitoring using optical sensors. *Groundwater Monitoring & Remediation*, *37*(4), 60–70. https://doi.org/10.1111/gwmr.12248

Maxwell, R. M., Putti, M., Meyerhoff, S., Delfs, J. O., Ferguson, I. M., Ivanov, V., ... Lopez, S. (2014). Surface-subsurface model intercomparison: A first set of benchmark results to diagnose integrated hydrology and feedbacks. *Water Resources Research*, *50*(2), 1531–1549. https://doi.org/10.1002/2013wr013725

McCabe, M. F., Rodell, M., Alsdorf, D. E., Miralles, D. G., Uijlenhoet, R., Wagner, W., ... Shi, J. (2017). The future of Earth observation in hydrology. *Hydrology and Earth System Sciences*, *21*(7), 3879–3914.

Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. *Water*, *10*(11), 1536. https://doi.org/10.3390/w10111536

Müller, J., Park, J., Sahu, R., Varadharajan, C., Arora, B., Faybishenko, B., & Agarwal, D. (2019). Surrogate optimization of deep neural networks for groundwater predictions. *Journal of Global Optimization*, 1–25. https://doi.org/10.1007/s10898-020-00912-0

Murphy, J., & Sprague, L. (2019). Water-quality trends in US rivers: Exploring effects from streamflow trends and changes in watershed management. *Science of the Total Environment*, *656*, 645–658.

NASEM, National Academies of Sciences, Engineering, and Medicine. (2018a). *Future water priorities for the nation: Directions for the US geological survey water mission area*. Washington, DC: The National Academies Press.

NASEM, National Academies of Sciences, Engineering, and Medicine. (2018b). *Open science by design: Realizing a vision for 21st century research*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25116

National Research Council. (2012). *Challenges and opportunities in the hydrologic sciences*. Washington, DC: National Academies Press.

Nearing, G. S., Kratzert, F., Alden Keefe Sampson, C. S., Pelissier, D. K., Frame, J. M., & Gupta, H. V. (2020). What role does hydrological science play in the age of machine learning. *Water Resources Research*.

Nearing, G. S., Ruddell, B. L., Bennett, A. R., Prieto, C., & V. Gupta, H. (2020). Does information theory provide a new paradigm for earth science? Hypothesis testing. *Water Resources Research*, *56*(2): e2019WR024918.

Novick, K. A., Biederman, J. A., Desai, A. R., Litvak, M. E., Moore, D. J., Scott, R. L., & Torn, M. S. (2018). The AmeriFlux network: A coalition of the willing. *Agricultural and Forest Meteorology*, *249*, 444–456.

Oroza, C. A., Zheng, Z., Glaser, S. D., Tuia, D., & Bales, R. C. (2016). Optimizing embedded sensor network design for catchment-scale snow-depth estimation using LiDAR and machine learning. *Water Resources Research*, *52*(10), 8174–8189.

Özgen-Xian, I., Kesserwani, G., Caviedes-Voullième, D., Molins, S., Xu, Z., Dwivedi, D., ... Steefel, C. I. (2020). Wavelet-based local mesh refinement for rainfall–runoff simulations. *Journal of Hydroinformatics*, jh2020198. https://doi.org/10.2166/hydro.2020.198

Park, S., Im, J., Jang, E., & Rhee, J. (2016). Drought assessment and monitoring through blending of multi-sensor indices using machine learning approaches for different climate regions. *Agricultural and Forest Meteorology*, *216*, 157–169. https://doi.org/10.1016/j.agrformet.2015.10.011

Peruzzo, L., Chou, C., Wu, Y., Schmutz, M., Mary, B., Wagner, F. M., ... Ma, X. (2020). Imaging of plant current pathways for non-invasive root phenotyping using a newly developed electrical current source density

approach. *Plant and Soil*, 1–18. https://doi.org/10.1007/s11104-020-04529-w

Peters-Lidard, C. D., Clark, M., Samaniego, L., Verhoest, N. E., Van Emmerik, T., Uijlenhoet, R., ... Woods, R. (2017). Scaling, similarity, and the fourth paradigm for hydrology. *Hydrology and Earth System Sciences*, *21*(7), 3701–3713.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., & Carvalhais, N. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204.

Rode, M., Wade, A. J., Cohen, M. J., Hensley, R. T., Bowes, M. J., Kirchner, J. W., ... Skeffington, R. A. (2016). Sensors in the stream: The high-frequency wave of the present. *Environmental Science & Technology*, *50*(19), 10297–10307. https://doi.org/10.1021/acs.est.6b02155

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Shen, C. (2018). A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research*, *54*(11), 8558–8593. https://doi.org/10.1029/2018WR022643

Slater, L. D., Ntarlagiannis, D., Day-Lewis, F. D., Mwakanyamale, K., Versteeg, R. J., Ward, A., ... Lane, J. W., Jr. (2010). Use of electrical imaging and distributed temperature sensing methods to characterize surface water–groundwater exchange regulating uranium transport at the Hanford 300 Area, Washington. *Water Resources Research*, *46*(10), 1–13. https://doi.org/10.1029/2010WR009110

Smith, R., Knight, R., & Fendorf, S. (2018). Overpumping leads to California groundwater arsenic threat. *Nature Communications*, *9*(1), 1–6.

Steefel, C. I. (2019). Reactive transport at the crossroads. *Reviews in Mineralogy and Geochemistry*, *85*(1), 1–26. https://doi.org/10.2138/rmg.2019.85.1

Steefel, C. I., Appelo, C. A. J., Arora, B., Jacques, D., Kalbacher, T., Kolditz, O., ... Molins, S. (2015). Reactive transport codes for subsurface environmental simulation. *Computational Geosciences*, *19*(3), 445–478. https://doi.org/10.1007/s10596-014-9443-x

Troch, P. A., Carrillo, G., Sivapalan, M., Wagener, T., & Sawicz, K. (2013). Climate-vegetation-soil interactions and long-term hydrologic partitioning: Signatures of catchment co-evolution. *Hydrology and Earth System Sciences*, *17*(6), 2209–2217. https://doi.org/10.5194/hess-17-2209-2013

Troch, P. A., Martinez, G. F., Pauwels, V. R., Durcik, M., Sivapalan, M., Harman, C., ... Huxman, T. (2009). Climate and vegetation water use efficiency at catchment scales. *Hydrological Processes: An International Journal*, *23*(16), 2409–2414 https://doi.org/10.1002/hyp.7358

U.S. DOE, U.S. Department of Energy, (2019). *Open watershed science by design: Leveraging distributed research networks to understand watershed systems workshop report* (DOE/SC-0200). U.S. Department of Energy Office of Science. doesbr.org/openwatersheds/

U.S. DOE, U.S. Department of Energy, (2020). *Integrated hydro-terrestrial modeling: Development of a national capability*, report (in preparation) of an interagency workshop hosted by the National Science Foundation, Alexandria, VA, September 4–6, 2019.

Varadharajan, C., Agarwal, D. A., Brown, W., Burrus, M., Carroll, R. W., Christianson, D. S., ... Henderson, A. (2019). Challenges in building an end-to-end system for acquisition, management, and integration of diverse data from sensor networks in watersheds: Lessons from a Mountainous Community Observatory in East River, Colorado. *IEEE Access*, *7*, 182796–182813. https://doi.org/10.1109/access.2019.2957793

Wainwright, H. M., Dafflon, B., Smith, L. J., Hahn, M. S., Curtis, J. B., Wu, Y., ... Hubbard, S. S. (2015). Identifying multiscale zonation and assessing the relative importance of polygon geomorphology on carbon fluxes in an Arctic tundra ecosystem. *Journal of Geophysical Research: Biogeosciences*, *120*(4), 788–808. https://doi.org/10.1002/2014jg002799

Wainwright, H. M., Falco, N., Dafflon, B., Chadwick, K. D., Maher, K., Breckheimer, I., ... & Devadoss, J. (2019). Toward watershed characterization: Investigation of above-belowground interactions with hyperspectral remote sensing and near-surface geophysical measurements within the Upper Colorado Basin. American Geophysical Union Fall Meeting, San Francisco B21A-07.

Wang, D., Liu, Y., & Kumar, M. (2018). Using nested discretization for a detailed yet computationally efficient simulation of local hydrology in a distributed hydrologic model. *Scientific Reports*, *8*(1), 1–13. https://doi.org/10.1038/s41598-018-24122-7

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... Bouwman, J. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

Wymore, A. S., West, N. R., Maher, K., Sullivan, P. L., Harpold, A., Karwan, D., Marshall J. A., Perdrial J., Rempe D. M. & Ma, L. (2017). Growing new generations of critical zone scientists. *Earth Surface Processes and Landforms*, *42*(14), 2498–2502. https://doi.org/10.1002/esp.4196.

Yang, C., Huang, Q., Li, Z., Liu, K., & Hu, F. (2017). Big Data and cloud computing: Innovation opportunities and challenges. *International Journal of Digital Earth*, *10*(1), 13–53.

Zacharias, S., Bogena, H., Samaniego, L., Mauder, M., Fuß, R., Pütz, T., ... Bens, O. (2011). A network of terrestrial environmental observatories in Germany. *Vadose Zone Journal*, *10*(3), 955–973. https://doi.org/10.2136/vzj2010.0139