Hillslope Hydrology in Global Change Research and Earth System Modeling


1Department of Earth and Planetary Sciences, State University of New Jersey Rutgers, New Brunswick, NJ, USA, 2National Center for Atmospheric Research, Boulder, CO, USA, 3Department of Environmental Sciences, University of Virginia, Charlottesville, VA, USA, 4Earth and Environmental Systems Institute, Department of Geosciences, Pennsylvania State University, University Park, PA, USA, 5Department of Geology and Geophysics, University of Utah, Salt Lake City, UT, USA, 6Department of Earth and Planetary Science, University of California, Berkeley, CA, USA, 7Department of Geosciences, Boise State University, Boise, ID, USA, 8College of Earth Ocean and Atmospheric Sciences, Oregon State University, Corvallis, OR, USA, 9Department of Environmental Systems Science, ETH Zürich, Zürich, Switzerland, 10Department of Geography, The State University of New York at Buffalo, Buffalo, NY, USA, 11School of Environment and Sustainability, University of Saskatchewan, Saskatoon, Saskatchewan, Canada, 12U.S. Geological Survey, Princeton, NJ, USA, 13Department of Geography and Atmospheric Science, University of Kansas, Lawrence, KS, USA, 14Bren School of Environmental Science & Management, University of California, Santa Barbara, CA, USA, 15Department of Environmental Sciences, University of California, Riverside, CA, USA, 16Department of Civil and Environmental Engineering, Duke University, Durham, NC, USA, 17Department Earth and Environmental Sciences, Universities of Freiburg, Freiburg, Germany, 18Department of Hydrology & Atmospheric Sciences, University of Arizona, Tucson, AZ, USA, 19Department of Geosciences, University of Arizona, Tucson, AZ, USA, 20Laboratory of Hydrology and Water Management, Ghent University, Ghent, Belgium, 21Department of Civil Engineering, University of Bristol, Bristol, UK, 22Department of Civil and Environmental Engineering, Northeastern University, Boston, MA, USA, 23Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, West Lafayette, IN, USA, 24Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, WA, USA, 25Department of Civil, Environmental and Architectural Engineering, University of Colorado Boulder, Boulder, CO, USA, 26Department of Biological and Agricultural Engineering, Texas A&M University, College Station, TX, USA, 27Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, USA, 28School of Engineering, University of California, Merced, Fresno, CA, USA, 29USDA Forest Service, PSW Research Station, Fresno, CA, USA, 30Department of Civil and Environmental Engineering, Pennsylvania State University, University Park, PA, USA, 31Department of Catchment and Urban Hydrology, Deltares, Delft, The Netherlands, 32Department of Geological Sciences and Engineering, University of Nevada, Reno, Reno, NV, USA, 33Institute of Industrial Science, The University of Tokyo, Tokyo, Japan

Abstract Earth System Models (ESMs) are essential tools for understanding and predicting global change, but they cannot explicitly resolve hillslope-scale terrain structures that fundamentally organize water, energy, and biogeochemical stores and fluxes at subgrid scales. Here we bring together hydrologists, Critical Zone scientists, and ESM developers, to explore how hillslope structures may modulate ESM grid-level water, energy, and biogeochemical fluxes. In contrast to the one-dimensional (1-D), 2- to 3-m deep, and free-draining soil hydrology in most ESM land models, we hypothesize that 3-D, lateral ridge-to-valley flow through shallow and deep paths and insolation contrasts between sunny and shady slopes are the top two globally quantifiable organizers of water and energy (and vegetation) within an ESM grid cell. We hypothesize that these two processes are likely to impact ESM predictions where (and when) water and/or energy are limiting. We further hypothesize that, if implemented in ESM land models, these processes will increase simulated continental water storage and residence time, buffering terrestrial ecosystems against seasonal and interannual droughts. We explore efficient ways to capture these mechanisms in ESMs and identify critical knowledge gaps preventing us from scaling up hillslope to global processes. One such gap is our extremely limited knowledge of the subsurface, where water is stored (supporting vegetation) and released to stream baseflow (supporting aquatic ecosystems). We conclude with a set of organizing
hypotheses and a call for global syntheses activities and model experiments to assess the impact of hillslope hydrology on global change predictions.

**Plain Language Summary** Hillslopes are key landscape features that organize water availability on land. Valley bottoms are wetter than hilltops, and sun-facing slopes are warmer and drier than shaded ones. This hydrologic organization leads to systematic differences in soil and vegetation between valleys and hilltops, and between sunny and shady slopes. Although these patterns are fundamental to understanding the structures and functions of water and terrestrial ecosystems, they are too fine grained to be represented in global-scale Earth System Models. Here we bring together Critical Zone scientists who study the interplay of vegetation, the porous upper layer of the continental crust from vegetation to bedrock, and moisture dynamics deep into the weathered bedrock underlying hillslopes and Earth System Model scientists who develop global models, to ask: Do hillslope-scale processes matter to predicting global change? The answers will help scientists understand where and why hillslopes matter, and to better predict how terrestrial ecosystems, including societies, may affect and be affected by our rapidly changing planet.

### 1. Hydrology in Earth System Models

Earth System Models (ESMs) are essential tools for understanding and predicting global environmental change. They must simulate the multitude of interactions within and among the atmosphere, the land, and the oceans through physical, chemical, and biogeochemical pathways. Designed for global coverage and century-long simulations of fast atmospheric motion, ESMs are built on large model grid cells (0.2–2° latitude-longitude, ~20–200 km), which cannot explicitly resolve finer-scaled but fundamental processes in the atmosphere, the oceans, and the land alike. On land, as ESMs incorporate ecosystem and biogeochemical processes, their ability to predict terrestrial water stores and fluxes becomes increasingly important because of the ubiquitous role that water plays in regulating biogeochemistry and sustaining ecosystems and societies. As ESMs continue to move toward increasingly mechanistic process representation and increasing breadth of the processes that are represented, they will require their modeled hydrology to advance along with their other aspects. Many land model processes depend on the details of the hydrologic states, for example, vegetation water stress and plant distribution, organic matter decomposition, methane production and emissions, fire ignition and spread, soil thermodynamics, nutrient leaching, dissolved organic and inorganic carbon export, and irrigation. In addition, as large-scale models begin to sharpen their focus on societally relevant issues such as water and land management and water security, the importance of getting runoff and streamflow correct (for the right reasons) also grows.

Hydrologic processes are traditionally studied at hillslope-to-catchment scales (tens of meters to kilometers). At this scale, topographic gradients from ridges to valleys drive water, sediment, and biogeochemical fluxes down the hillslopes and out of the catchments. And at this scale, the sunny and shady sides of a ridge exhibit contrasting radiative, hydrologic, and ecological conditions. Field instruments are traditionally aligned with, and hillslope- and catchment-scale models are built to capture, such fundamental gradients across the landscape (e.g., Angermann et al., 2017; Band, 1991, 1993; Band et al., 1993, 1991; Bergstrom et al., 2016; Duffy, 1996; Dunne & Black, 1970; Ebel et al., 2007; Hwang et al., 2009; McDonnell, 1990; McGuire & McDonnell, 2010; Montgomery et al., 1997; Nippgen et al., 2015; Staudinger et al., 2017; Wilson & Dietrich, 1987). Recently, Critical Zone (CZ) science—the study of the structure, function, and evolution of the soil and underlying regolith (the weathered mantle overlying fresh bedrock) that support terrestrial life (e.g., Brantley et al., 2006; Brantley, McDowell, et al., 2017)—has catalyzed an integrative, system-level approach toward understanding the coevolution of the landscape and terrestrial life, bringing together diverse perspectives from the broader Earth Surface Processes community.

In this new CZ science, hillslopes and catchments remain as the scales of instrumentation, conceptualization, and modeling, for they continue to be recognized as fundamental organizers of water, energy, and biogeochemical states and fluxes across the landscape (Brantley, Lebedeva, et al., 2017; Rempe & Dietrich, 2018).

Over the past six decades or so, hillslope and catchment research has produced prodigious field observations, theories, and models, at individual research sites and across organized research networks including the long-term ecological research network and the US Department of Agriculture Agricultural Research Service and Forest Service experimental watersheds (e.g., Stringer et al., 2016). The nascent international network
Of Critical Zone Observatories (CZOs) is laying the foundation for cross-site syntheses (e.g., Baatz et al., 2018; Brantley, McDowell, et al., 2017; Paola et al., 2006), making possible a global assessment of the most salient structures and functions of hillslope processes that may matter to large-scale fluxes in ESMs. Thus, catchment and CZ scientists, as a community, can offer deep insights into the lay of the land, and we hope to tap into their collective wisdom so that we can identify the most critical processes to implement in ESMs.

At the planetary scale, ESMs embody our best understanding of Earth system interactions, and the Intergovernmental Panel on Climate Change studies of past and future climate depend on ESMs. In the early years, hydrology was portrayed as one-dimensional (1-D, vertical) infiltration into, and evapotranspiration (ET) from, the surficial soils, neglecting the 3-D (vertical + lateral) water movement on and below the land surface. As ESMs include subgrid terrestrial vegetation dynamics (e.g., Fisher et al., 2018) and human impacts such as large-scale irrigation (e.g., Leng et al., 2013), a land hydrology with comparable subgrid complexity has become justifiable. In this synthesis, we bring together the hydrology-CZ community and the ESM community to achieve the following objectives: (1) to identify the sub-ESM grid-scale processes that are fundamental organizers of hydrologic stores and fluxes across the landscape that may matter to ESMs, (2) to explore simple mechanistic ways to implement them in ESMs, and (3) to formulate a set of testable hypotheses and future global synthesis efforts to test them.

This synthesis builds onto an earlier paper by Clark et al. (2015) outlining challenges and opportunities for improving land hydrology in ESMs. Clark et al. (2015) articulated the scientific motivations and reviewed hydrologic process representations in current ESM land models. They highlighted the gaps between ESM hydrology and current process understanding in the hydrology and CZ communities and identified key opportunities to advance ESM hydrology. These opportunities include (1) representing hillslope-scale lateral hydrologic convergence and two-way surface-groundwater exchange, (2) a comprehensive benchmarking system using hydrologic and CZ observations, and (3) stronger interactions between the CZ and ESM communities (exemplified by continental-scale hydrologic modeling uniting the hydrology and ESM communities; Archfield et al., 2015). Subsequent workshops and proof-of-concept testing using the Community Land Model further shaped the discussions. These discussions converged on the need to (1) pool the diverse views in the hydrology and CZ community, (2) set priorities among the multitude of processes viewed as important, and (3) recommend to the ESM community what hydrologists and CZ scientists consider as the most salient structures and functions of sub-ESM grid processes. Through CUAHSI (https://www.cuahsi.org/) we initiated a call for white paper contributions, followed by a series of webinars and meetings at the annual AGU (American Geophysical Union) conference. Here we report the outcome of these discussions.

As reviewed in detail by Clark et al. (2015), early ESM land models described land hydrology as 1-D fluxes, partitioning canopy throughfall into surface runoff versus infiltration then partitioning infiltration into soil storage supporting ET versus deep drainage loss to river baseflow. In this 1-D construct, a model grid cell was a large flat slab with no local topographic gradients, that is, the hydrologic condition was uniform within a grid cell (~20–200 km wide). This approach neglected certain processes that are important at the landscape scale—specifically, the lateral flow from ridges to valleys and the topographic controls on solar insolation. In addition, the large model slab was only a few meters thick and freely drained; drainage water was placed in the model “river storage” instantaneously, neglecting drainage impediment due to shallow water tables in lowland valleys, and the delayed discharge into streams via lateral flow in the subsurface. The consequence of this 1-D free-draining model construct, for example, in the Amazon, was fast drainage loss during wet periods (Miguez-Macho & Fan, 2012a), reduced terrestrial water storage in dry periods (Pokhrel et al., 2013), and shutting down of ET and headwater streamflow in the dry season (Miguez-Macho & Fan, 2012b), contrary to flux tower and streamflow observations.

The past decades have seen several advances in the representation of land hydrology in ESMs. One example is the use of the variable source area concept (e.g., Dunne & Black, 1970; Hewlett & Hibbert, 1967) and the application of TOPMODEL (Beven & Kirkby, 1979) to divide large model cells into zones of varying saturation (more in section 4), capturing localized surface runoff production and the disproportionate impact of riparian transpiration (Famiglietti & Wood, 1994). The second example is the redistribution of snow based on terrain characteristics (e.g., Liston, 2004). The third is treating individual grid cells as mosaics of different surface properties, such as land use and cover types, further refined by dividing vegetated patches into plant functional types (PFTs), for example, forests versus grasslands (Koster & Suarez, 1992). However, the
structures and functions of the terrain within a grid cell remained largely absent; there is no explicit and
dynamic lateral flow from uplands to lowlands (except by Subin et al., 2014) nor are sunny and shady slopes
distinguished from one another. The modeled distribution of PFTs (e.g., grassland vs. forest) is generally
(with some exceptions, e.g., the ORCHIDEE Land Model, https://orchidee.ipsl.fr/about-orchidee/) not asso-
ciated with local hydrologic conditions, even though these conditions vary systematically from uplands to
lowlands and from sunny to shady slopes, which may explain the patchwork of different PFTs under any
given climate. The common disassociation of PFTs from their water-energy environment makes it difficult
for models to predict that forest ET can continue over part of the landscape, despite water-stressed condi-
tions on average (Miguez-Macho & Fan, 2012b). At a fundamental level, topographic complexity alters the
amount and timing of energy and plant available water across the landscape. The resulting spatial variability
in energy-water coupling is reflected in the covarying spatial structure of vegetation, soil and regolith devel-
opment, biogeochemical fluxes, and other aspects of the CZ. Recent advances in the mechanistic under-
standing of these interactions call for a renewed evaluation of how to best represent land hydromorphy in ESMs.

From the perspective of hillslope and catchment hydrologists and CZ scientists, there are many aspects of
ESM hydromorphy that are considered unrealistic and in need of major revisions or expansions (as reviewed
in Clark et al., 2015). CZ science is revealing increasing complexities and evermore nuanced insights, partic-
ularly regarding moisture availability in the weathered bedrock beneath soil (Rempe & Dietrich, 2018;
Salve et al., 2012), subsurface preferential flow paths with multiple perched lateral flows (the fill and spill
concept; Tromp-van Meerveld & McDonnell, 2006), and the intermittent hydrologic connectivity that drives
plant water use and geochemical releases from catchments (e.g., Brantley, Lebedeva, et al., 2017; Godsey
et al., 2009; Hopp & McDonnell, 2009; Lanni et al., 2011; McDonnell, 2014; Rahman & Rosolem, 2017).
Indeed, attempts have recently been made to represent the rapid recharge to the deep store via rock fractures
in global models (Hartmann et al., 2015, 2017; Vrettas & Fung, 2015, 2017). However, the state of knowledge
and the lack of global subsurface data do not yet allow us to test the importance of these emerging properties.
And given the increasing complexity of ESMs and the associated computation burden, it is prudent to focus
on the first-order and well-understood processes. For example, that water flows downhill and that the sunny
slopes are warmer and drier than shady slopes have been long and universally acknowledged. Incorporating
these processes in ESMs—through characterization and estimation of their grid-scale functions—may now
be possible, given the global availability of hillslope-resolving topographic data. Therefore, our focus will be
on processes that are obvious and quantifiable and that can potentially impact ESM predictions in water-
stressed places and times.

These considerations will guide the presentation below. We seek to identify salient structures and processes
that (a) shape the abiotic and biotic landscape; (b) have the potential to alter large-scale water, energy and
biogeochemical fluxes between the land and the atmosphere; (c) can be described by basic physical princi-
ples with a few measurable parameters; and (d) can be tested by comparing model outcomes against obser-
vations. We pose the following questions to the two communities jointly.

1. How does the terrain (topography, lithology, and geomorphic history) organize the regolith (soil and
weathered bedrock) and vegetation and the stores and fluxes of water, energy, and carbon across the
landscape? (section 2)
2. How does this terrain influence vary throughout the world under the wide range of climate-terrain com-
binations? Where do we expect the greatest impact on ESM water, energy and carbon fluxes? (section 3)
3. How do we divide an ESM grid cell into functional units that best reflect terrain controls? How do we
represent the connectivity among the units? And how do we parameterize the models from observable
variables? (section 4)
4. What are the testable hypotheses regarding the terrain-driven CZ structures and functions in the context
of ESM predictions and global change research? What are the immediate and future efforts needed to test
these hypotheses? (section 4.2)

The purpose of this synthesis is twofold. On one hand, we strive to distill the best process understanding
from the broader CZ community to improve ESM realism. ESMs are powerful tools for synthesizing and test-
ing first-order Earth system-level interactions, and they are the only tools for the society to foresee the future.
The CZ community has the knowledge and hence the obligation to contribute to ESM advancement. On the
other hand, as future ESMs become able to capture processes at scales meaningful to CZ scientists, the CZ
community will gain a powerful tool to explore coupled climate-CZ coevolution, the notion of a dynamic CZ regulated by, and regulating, global change (Gaillardet et al., 2018). Our emphasis here on connecting site-based science to global predictions resonates with the decadelong effort in the hydrology community to seek and organize hydrologic similarities with the aim of predicting unobserved places (McDonnell & Woods, 2004; Wagener et al., 2007). The CZ-ESM collaboration can offer new motivation and momentum for ongoing inquiries, expose new knowledge gaps, and generate new hypotheses. Hence, the intended audience of this paper includes both the broader hydrology and CZ community, who observe and decipher how nature works at the inherent scales, and ESM developers and users, who must abstract the first-order interactions to make global predictions.

We emphasize the impact of present-day terrain structure on land hydrology, neglecting hydrologic regulation of weathering and erosion that shape the landscape at geologic time scales. Although these long-term processes are a central focus of CZ science, the design time scales of ESMs (seasonal to decadal) dictate that we first assume a static landform, which controls the short-term dynamics of water storage and flow. However, to characterize the depth profile of soil/regolith permeability and porosity in hilly terrain—which are critical to water storage and flow but not directly observable—we must depend on CZ knowledge of the processes and geomorphic history that have created the landscape we see today. Therefore, our synthesis will probe into the depth structure of the CZ along hillslope profiles but will be limited to what they look like now, to help parameterize water storage and transmission in the model soil/regolith.

We focus on vegetation and its associated ET flux for the following reasons. First, natural vegetation selects, adapts to, exploits, and thus expresses the integrated water-energy-nutrient environment. By observing vegetation, which is readily observable at individual, stand, ecosystem, and global scales, one can infer the critical resources that are limiting (e.g., Meinzer, 1927). Second, vegetation exerts a strong influence on CZ evolution through myriad pathways (e.g., Brantley, Eissenstat, et al., 2017; Roering et al., 2010; Sullivan et al., 2016); thus, understanding landform-vegetation relations is a key CZ inquiry. Third and critically, plant photosynthesis and ET are the primary pathways for land-atmosphere exchange of water, energy, and carbon, and hence, understanding the structure and function of vegetation is of central interest to ESMs and global carbon cycle research.

2. Terrain Organization of Vegetation and ET via Water and Energy

It is well known that vegetation patterns follow the topography in mountainous terrain (e.g., Schimper, 1903; von Humboldt, 1807). Figure 1b reproduces the famous tableau of von Humboldt describing the vertical vegetation zones encountered on his ascent of the Andes in 1799–1803. The sheer elevation range, from sea level to over 4,000 m, gives rise to a steep climate gradient that underlies the striking vegetation gradient described by von Humboldt. The entire vertical range occurs within a horizontal distance of only ~1° longitude (~100 km at the equator), the dimension of a typical ESM grid cell (Figure 1c). Clearly, in such regions of the world, the many PFTs within an ESM grid cell would have little meaning without their relation to the elevation range they occupy.

In flatter parts of the world (e.g., the central Amazon; Figure 1d), the elevation range and climate gradient are subdued. Local gradients dominate the relief, and a regular motif of plateaus, side slopes, and stream valleys ensues. The climate here does not exhibit striking vertical zones, and the local hydrologic gradient emerges as a key differentiator for vegetation (e.g., Day & Monk, 1974; Hales et al., 2009; Hoylman et al., 2018; Moulatlet et al., 2014; Schietti et al., 2014; Whittaker, 1956).

As ESM land model grid cells approach a fraction of a degree (1/8–1/5° or ~12.5–20 km), the vertical climate zonation will further decrease. At the 20-km grid scale (e.g., Figure 1e), it may be reasonable to neglect vertical climate zonation in moderate-to-low relief terrain and to represent it in steep terrain through subgrid atmospheric downscaling. In the discussions below, we assume that the regional climate is uniform across an ESM grid cell, and under this uniform climate, we examine how the local terrain and hydrology may organize the vegetation. We focus on two basic hillslope structures: the gravity-driven down-valley convergence of surface and subsurface flow and the systematic difference in insolation between sunny and shady slopes.
2.1. Down-Valley Drainage

Here the term drainage includes both the vertical percolation at any position on the landscape and the lateral, ridge-to-valley convergence of surface and groundwater, both driven by gravity. Hill-to-valley convergence creates drier hills and wetter valleys, resulting in deep or absent water tables under the hills, and shallow water tables under the valleys. This hill-to-valley convergence facilitates efficient vertical drainage in uplands and impedes drainage in valleys, thus forming riparian wetlands. In places and times with water stress, a wetter valley can support higher plant productivity than one would expect from the climate alone. Figure 2a shows a landscape in the southwestern United States under a semiarid climate. The valley-hill hydrologic contrast translates directly into a vegetation contrast. Here the PFTs (aspen forest vs. desert shrub/grass), duly differentiated in ESMs, correlate well with drainage positions. Without this inherent coupling and assuming uniform soil water status across an ESM grid cell, it would be difficult to predict the continued transpiration from the aspen forest through the entire growing season. This ridge-to-valley subsidy also occurs in seasonally dry climates where abundant wet season precipitation recharges the regolith and groundwater (Goulden et al., 2012; Schwantes et al., 2018; Swetnam et al., 2017; Tai et al., 2017). Groundwater flow converges toward the valleys with a delay, and continues into the dry season, because groundwater moves slowly. Gallery forests line the stream corridors, such as in the Mediterranean climate of central California (Figure 2b) and the monsoonal climate of eastern Africa (Figure 2c). The latter landscape lies in a transition zone between rainforests and open savanna, which can coexist owing to the effects of topographic relief (Kim & Eltahir, 2004). In the landscapes represented by Figures 2a–2c, it is aridity, at least seasonally, that makes hill-to-valley drainage relevant to vegetation patterns.
In humid and low-relief regions where water is in excess, lateral drainage is also important but for different reasons. Here regional drainage is impeded, resulting in waterlogged soils and oxygen stress for plants. The slightly elevated hills can improve local drainage and alleviate waterlogging. In the vast seasonal wetlands of the Pantanal (Dubs, 1992; Figure 2d), the low hills support islands of forests above the poorly drained...
In the white sand forests of Rio Negro in Venezuela (Figure 2e), the slightly elevated sandy hills are excessively drained and support low Bana forests only, while the lowlands are perennially waterlogged and support palms only, and the intermediate positions are occupied by high Caatinga forests (Coomes & Grubb, 1996). The mosaic appearance of the boreal forest in northern Denmark (Figure 2f) reflects drainage, and in this cold region, drainage also affects soil temperature and growing season length, with waterlogged soils warming up later in the springtime. A positive feedback reinforces the terrain-vegetation association; denser vegetation and higher ET on higher terrain further lowers the water table, improving drainage and nutrient conditions and enabling higher productivity and higher ET, and so forth (Sullivan et al., 2016).

In the landscapes represented by Figures 2d–2f, it is too much water, and the associated soil oxygen stress, that makes lateral drainage important to vegetation patterns.

### 2.2. Slope Aspect Difference

Another widely acknowledged consequence of topographic relief is that it creates variations in local solar angle and thus the intensity of solar radiation received at the surface. In places and times with energy limitation, warmer slopes can support longer growing seasons. Conversely, in places and times with water stress, cooler slopes can support higher plant productivity. Figure 3 gives examples in the Northern Hemisphere (north toward the right in all images) where south facing slopes are warmer and drier as articulated by the vegetation (e.g., Ping et al., 2005; Bennie et al., 2006; Newman et al., 2014; Brooks et al., 2015; Smith et al., 2017; and a synthesis by Pelletier et al., 2018). In arid western Texas (Figure 3a), and seasonally arid California (Figure 3b) and Idaho (Figure 3c), the cooler north facing slopes support a moister habitat and mesic vegetation. At a site in Israel, north facing slopes harbor higher vegetation biomass with lower susceptibility to insect herbivory (Auslander et al., 2003). In regions that depend on winter snowpacks for warm-season water supply, snow persists longer into the summer on north facing slopes (Figure 3d), delaying water release into the dry season. In landscapes represented by Figures 3a–3d, it is aridity, at least seasonally, that makes topographic aspect relevant to vegetation patterns.

In the energy-limited high latitudes of the Yukon (Figure 3e), tree line position and species richness and turnover are best explained by slope aspect, because soils thaw more deeply on the warmer south facing slopes (Dearborn & Danby, 2017). That tree line position and ecotonal plant community composition vary systematically across hillslope aspects is a long and well-established concept (e.g., Elliott & Cowell, 2015; Elliott & Kipfmueller, 2010; Johnson, 1848). Near Fairbanks, Alaska (Figure 3f), south facing slopes with well-developed silt loam soils support mature white spruce that “should be rated among the most rapidly growing and valuable forest stands of interior Alaska” (Krause et al., 1959), while on the north facing side of the ridge, “a poorly drained half-bog soil supports a stand of dwarfed black spruce with an extremely slow growth rate.” Krause et al. (1959) also wrote that “none of the alterations of the vegetative cover is comparable to the radical and lasting changes imposed by micro-climatic influences on slopes of different exposures, especially those of northern and southern aspects.” At such high latitudes, the low sun angle amplifies even low relief, affecting permafrost thawing, land drainage and aeration, and vegetation phenology (e.g., Endalamaw et al., 2017; Ping et al., 2005). In landscapes represented by Figures 3e and 3f, it is energy stress, at least seasonally, that makes topographic aspect relevant to vegetation patterns.

The generalizations presented above are complicated by the intersections between aspect-driven energy availability and gravity-driven lateral drainage. In warmer locations or on sunny slopes, valleys harbor greater biomass (Perdrial et al., 2018; Swetnam et al., 2017), but in cooler locations and/or shady slopes, valleys support lower biomass (Zapata-Rios et al., 2016). In more mesic locations, the observed systematic increase in biomass downslope has been attributed to nutrient subsidies carried by the lateral hydrologic subsidy (Shi et al., 2018; Weintraub et al., 2017). By mechanistically representing these fundamental drivers and their interactions, ESMs can offer a powerful tool to elucidate the complex patterns of vegetation distribution across the globe.

### 2.3. Soil/Regolith Thickness Along Topographic Gradients

A critical parameter in ESM land models is the so-called “soil depth,” intended to include sufficient soil water storage to support ET during the dry intervals between precipitation events. Early models included the top 2 m of the land surface, thought to represent the depth containing the bulk of the root biomass in woody plants (e.g., Sellers, Randall, et al., 1996; Sellers, Tucker, et al., 1996). Such model soil depths were
found to be inadequate to explain the high dry-season ET in the Amazon, where deeply weathered tropical soils extend far deeper and roots were seen at 11-m depth (Nepstad et al., 1994). ESM simulations of ET and regional climate are highly sensitive to soil depth in the Amazon and elsewhere (Baker et al., 2008; Harper et al., 2010; Kleidon & Heimann, 2000; Mankin et al., 2017; Markewitz et al., 2010; Milly & Dunne, 1994), yet it is a poorly constrained parameter. The choice has been somewhat ad hoc, depending on convention and influenced by conflicting reports of rooting depths of particular plants in particular places; the report of 11-m deep roots by Nepstad et al. (1994), for example, motivated model studies with 10-m soil depth (e.g., Baker

**Figure 3.** Contrasting energy-water-vegetation conditions between northern and southern exposure (north to the right in all images), in (a) water-limited environment of Texas (Guadalupe Peak in Guadalupe Mountains National Park, http://thecommonmilkweed.blogspot.com/2009/10/guadalupe-peak-hike.html), (b) seasonally water-limited California (GoogleEarth), and (c) southwest Idaho near Anderson Ranch Reservoir (https://upload.wikimedia.org/wikipedia/commons/d/d8/Effects_of_aspect_on_vegetation_SW_Idaho.JPG). (d) Higher up in Idaho near Mount Cramer (GoogleEarth), snow persists into the summer on north facing slopes. In energy-limited environments, (e) tree line elevation and plant community composition differ across hillslope aspects in southwest Yukon (photo by Ryan Danby, Queen’s University, Canada; used with permission) and (f) interior Alaska near Fairbanks (GoogleEarth).
et al., 2008; Lee et al., 2005). Yet the question remains: what is the right model soil depth? How deep is deep enough in different parts of the world? What is the scientific basis for defining model soil depth, and how can it be quantified across the diverse environments of the world?

The struggle to define the right soil depth by ESM scientists in many ways parallels the struggle to define the thickness of the CZ by CZ scientists (see a recent attempt by Riebe et al., 2017). Both communities have in mind the layer of Earth surface material that is porous and permeable, that envelops the circulation of meteoric water, that supports terrestrial life, and that is physically, chemically, and biologically altered by water and life. In this regard, knowledge of CZ depth and structure, one of the central inquiries of CZ science, is directly relevant to clarifying, conceptualizing, and quantifying the soil depth in ESMs.

Because of the difficulties in seeing the subsurface, our knowledge of the depth and structure of the CZ is very limited. Figure 4 offers a few “windows” into the dark belowground, showing cross sections of the
CZ along ridge-valley transects at several U.S. CZOs, obtained from coring (Figure 4a, core locations given) and a variety of subsurface geophysical inversion tools, particularly seismic velocity that indicates rock density, degree of fracturing, and/or porosity development. Some tentative, ESM-relevant remarks can be made across these sites, which are limited in sample size but cover a range of climates, lithologies, plant biomes and landforms, and thus can at least prompt some fruitful synthesis questions.

First, the porous/permeable and hydrologically active layer is far thicker than the 2- to 3-m depth commonly assumed in ESM land models. At the Eel River CZO in California (Figure 4a; Rempe & Dietrich, 2018), the traditionally defined soil is only 0.5-0.75 m thick, but the underlying weathered or fractured bedrock is also hydrologically active and is >20 m thick. Roots were found in fractures at 16-m depth tapping the “rock moisture” (exchangeable water stored in the unsaturated zone in the weathered bedrock), which held more than one fourth of the annual precipitation, sustaining the Douglas fir forest through the dry summer. The rock moisture zone is also the primary conduit for lateral groundwater drainage from hills to valleys that sustains stream baseflow at this site. At the Southern Sierra CZO, also in California (Figure 4b; Holbrook et al., 2014), about one third of annual ET comes from water storage below 1-m depth in the deeper regolith; this storage is recharged during the winter and is available for tree use during the summer and fall (Bales et al., 2011; Klos et al., 2018). The importance of weathered and fractured rocks in supporting ET is widely recognized in mountainous terrain with a dry season (Arkley, 1981; Bales et al., 2011; Goulden et al., 2012; Graham et al., 2010; Johnson et al., 2018; Miller et al., 2010; Rempe & Dietrich, 2018; Salve et al., 2012). Thus, in hilly terrain where the “soil” is shallow, and in biomes where plants depend on deep moisture in the dry season, ESM soil depth must include the weathered or fractured rocks to account for the observed dry-season plant survival.

Second, in general, the vertical change in porosity/permeability with depth can be gradual, particularly in warmer and wetter environments where weathering fronts are deep (Figure 4f, lower two rows). This gradual change—in physical, chemical, and biological markers alike—makes defining the base of the CZ difficult (Riebe et al., 2017), particularly in areas of the world underlain by deep sediments with a deep, long-distance regional groundwater flow system (Schaller & Fan, 2009). From the hydrologic perspective, it is difficult to define the base of groundwater flow without deep profiles of material properties. Given that deeper processes are generally associated with longer time scales and given that the slow processes can be important to ESM objectives (e.g., deeper flows sustain baseflow and riparian ecosystems in drought months and years), a sharp cutoff depth can be difficult to delineate across the wide range of hydrogeologic conditions of the world. Instead, an approach that quantifies the vertical gradients in subsurface properties may be more meaningful, as elaborated later in section 4 on parameterizing the model lower boundary.

Third, there is evidence that the CZ thickness and structure may correlate with topography, which to a large extent drives water and sediment fluxes, and weathering and erosion rates that shape CZ thickness. Moving from the valley bottom to the ridge top, the porous-permeable layer thickens at the California sites (Figures 4a and 4b), but the shape of the fresh bedrock boundary differs (convex vs. concave up). Rempe and Dietrich (2014) propose that the vertical advance of the weathering front is controlled by the rate of drainage from the fresh bedrock, and they present an analytical model to predict the convex shape of the hilltop transition to fresh bedrock, as well as the consequent upslope thickening of the weathered profile. At the Boulder Creek CZO in Colorado (Figures 4c and 4d for two subcatchments), the soil/regolith is thicker under the valleys in the upper catchments (Figure 4c; sections 1 and 2 and Figure 4d, section 1), because recent river incision has not reached this far upstream and the present regolith profiles reflect a past climate (Befus et al., 2011). The cross section at the Shale Hills CZO in Pennsylvania indicates no systematic trend (Figure 4e), although the latest data (not shown) indicate that the regolith is thicker under the ridgelines.

Figure 4f reveals the possible interplay between the local topography and the regional tectonic stress field in initiating CZ development. The high correlations among the rock failure potential (left column), the stress field (middle column), and the seismic velocity (right column, reflecting rock density) are striking across the three sites representing very different lithologies, climates, and histories (Boulder Creek CZO in Colorado, Calhoun CZO in North Carolina, and Pond Branch in Maryland). These observations raise the possibility of predicting the CZ depth, or at least the hard-rock threshold, based on our knowledge of
tectonic stress fields and the readily available high-resolution maps of surface topography (Slim et al., 2015; Riebe et al., 2017 and their Hypothesis 1).

Across hillslope aspects, there is a difference between south and north facing slopes at Boulder Creek CZO (Figures 4c and 4d), with substantially thicker soil and regolith on the north facing slopes (left side of the figures). Stronger frost action is proposed as a key factor (Anderson et al., 2013; Riebe et al., 2017 and their Hypothesis 2). The weathered zone also seems thicker on the north facing slope at Shale Hill CZO (Figure 4e) despite a steeper slope, which is suggested to reflect the relic periglacial landscape from the last ice age; a modeling study here suggests that ~100 mm more water passes annually through the north facing than the south facing slopes. At the Reynolds Creek CZO in southwestern Idaho, Geroy et al. (2011) wrote, “Soils on the north aspect retain as much as 25% more water at any given soil water pressure than samples from the south aspect slope. Soil porosity, soil organic matter and silt content were all greater on the north aspect, and each contributed to greater soil water retention. These results, along with the observation that soils on north-aspect slopes tend to be deeper, indicate that north-aspect slopes can store more water from the wet winter months into the dry summer in this region, an observation with potential implications on ecological function and landscape evolution.” But in studies in the Yukon (Dearborn & Danby, 2017) and Alaska (Krause et al., 1959), the opposite seems true, with south facing slopes supporting more deeply thawed, drained, and developed soils, which in turn support larger vegetation and higher productivity. Thus, it appears that, where water is most limiting, the north facing slopes are hydrologically and ecologically more active, with water and life penetrating deeper into the substrate, both critical agents of weathering and CZ deepening, but where energy is most limiting, the south facing slopes are hydrologically and ecologically more active. This is a hypothesis to be tested globally (section 4.2).

The above discussions only touch upon a few of the many nodes and threads of an enormously complex system of process interactions that shapes the structure of the modern-day CZ. But they raise the possibility of simplification, extrapolation, and global prediction, albeit crudely and tentatively. A key question is, what do we know now that will allow us to improve upon the current ESM soil depth prescriptions?

2.4. Plant Rooting Depth Along Drainage Gradients

The link between the moisture stored in the CZ and the aboveground vegetation is the plant root system. As the pioneer root ecologist Weaver (1919) stated, “one cannot understand why a particular plant is found in a particular place without knowing its root system”. Among the many root traits, the rooting depth is perhaps the closest proxy for the depth of the CZ that is exploited and altered by vegetation. Here we ask the question, are there systematic changes in plant rooting depth along drainage and aspect gradients? Because roots are difficult to observe, we have only a very limited view of their morphologies and functions. Existing observations seem to suggest that hillslope hydrology is an important driver of plant rooting depth (Fan et al., 2017). Figure 5A plots 2020 rooting depth observations of >1,000 species, on a log scale, against several abiotic and biotic factors. At a given mean annual precipitation (a) or within a given soil texture class (b), rooting depths vary over orders of magnitude. Although larger and evergreen plants (d) tend to develop deeper roots, rooting depths vary widely within individual growth forms. Of the 30 best sampled genera, species in arid and semiarid climates tend to develop deeper roots, but the wide range underscores the large plasticity of root response to local conditions.

These local conditions include bedrock depth (c) and water table depth (f). The strong correlation with bedrock and hardpan depth (c) hints at the importance of knowing the CZ structure and physical barriers to root penetration. But more than half of the sampled roots penetrated below the barriers, either as a result of ambiguous reporting of bedrock depth or as a testimony to plants’ persistence in securing resources. Since water is a primary resource, the strong correlation with water table depth (f) is unsurprising. As conceptualized in Figure 5B, the moisture profile at any location may vary systematically from the hill to the valley. The profile is wetted above by intermittent infiltration and wetted below by steady capillary rise from the water table. In between there may exist a dry gap of varying duration, which narrows toward the valley. Roots of plants on the ridge (position 1) cannot cross the dry gap and are limited to shallow infiltration depths (Bucci et al., 2009; Cannon, 1911). At the lower position 2, roots may sense the capillary rise, and dimorphic roots develop, with a shallow cluster using rain and a deep cluster using groundwater in dry seasons (Dawson & Pate, 1996; Howard, 1925; Kimber, 1974). At position 3, infiltration meets capillary rise and water is not limiting. At position 4, seasonal waterlogging limits roots to the oxygenated soils above the water.
table, and shallow or aerial roots are common in lowland forests (Pavlis & Jeník, 2000; Smith, 1972; Stone & Kalisz, 1991). At position 5, permanent waterlogging selects wetland species that are insensitive to the water table.

These observations provide evidence that plant rooting depths covary with CZ structure and hydrology and that the aboveground organization of vegetation by hillslope hydrology is facilitated by belowground root-water relations. These relations are the biophysical link between CZ hydrology and land plants. By resolving the fundamental gradients in CZ hydrology (including soil moisture, rock moisture, and the water table...
depth) across the landscape, ESM land models can avoid prescribing plant rooting depth and instead let the CZ moisture profile and plant water demand dictate the necessary depth of root growth (Fan et al., 2017).

3. Global Significance of Terrain Influence on Vegetation

The above highlights of the well-known terrain influence on vegetation through drainage and aspect and the less well-known terrain influence on soil/regolith depth and structure (which in turn determine the water storage capacity for ET) are based on anecdotal, location-specific evidence (Figures 1–4). The next questions are as follows: where and when, across the diverse and dynamic environments of the globe, do we expect that these terrain influences will matter to ESM predictions of large-scale water, energy, and biogeochemical fluxes? Do such areas add up to a significant portion of the global land area? And will the hillslope-scale structures, however, deterministic and predictable, simply average out over an ESM grid cell and hence matter little to global predictions?

Regions of strong topographic relief are obvious candidates for strong terrain influence, but strong relief is unnecessary under some conditions and insufficient under others. For example, if an ESM grid cell with strong relief is located in the ever-wet tropics, with no water or energy limitations on photosynthesis, then down-valley drainage will have little impact on either uplands or lowlands, both receiving abundant and frequent rain to sustain ET and both well drained to maintain aerated soils. Its tropical latitude will minimize the aspect difference as well. Thus, topographic relief will be most important under certain combinations of climate and terrain that create water shortage or waterlogging or energy stress. Figure 6 illustrates some of the end member climate-terrain combinations where (a) down-valley drainage and (b) aspect difference may matter. The example of the ever-wet tropics with high local relief (assuming no substantial climate zonation) would map onto position 1 in Figure 6a.

As postulated in Figure 6a, down-valley drainage may matter the most in a seasonally dry climate with moderate to high relief (positions 4 and 5). Examples are Figures 2b and 2c. The wet season ensures water surplus in part of the year, and the relief drives upland surplus toward the lowlands. Depending on slope length and gradient, the flow will take time, so that the spatial carryover from uplands to lowlands implies a temporal carryover from wet to dry seasons, supplying ET in lowlands and dry seasons. In such places, although the portion of the landscape (or grid cell) remaining green in the dry season may be small, the continued survival of valley forests may have far-reaching ecological and societal consequences. Although a similar argument can be made regarding positions 7 and 8 (arid climate) in Figure 6a, without a strong wet season there is little surplus for spatial-temporal carryover and the importance of down-valley drainage is diminished. However, a wetter riparian corridor can exist further down the regional gradient along larger rivers fed by long-distance groundwater convergence, discussed further in section 4 on regional-scale hydrologic connectivity.

Down-valley convergence also matters in sites under ever-wet climate with low relief (position 3, Figure 6a). Here the problem is too much water, and the elevated mounds improve local drainage and support higher productivity. Examples are Figures 2e and 2f. The same occurs in a climate with seasonal flooding (position 6, e.g., Figure 2d). If captured in ESMs, the improved drainage of the low hills can support islands of larger vegetation that would otherwise be impossible.

As shown in Figure 6b, aspect variations are highest in the high latitudes, especially in moderate- to high-relief landforms (positions 1 to 2), which are found to be the strongest factors driving plant distribution (Dearborn & Danby, 2017, Figure 3e; Krause et al., 1959, Figure 3f). At the high latitudes, even low relief (position 3) can be amplified by the low Sun angle. Such predictable variations in insolation, if captured in ESMs, will result in longer snow ice and soil-groundwater residence times in the shady portion of the landscape, affecting the period of snowmelt, permafrost freeze/thaw, vegetation phenology, and ET.

Figure 6 suggests that terrain influence on vegetation matters the most where vegetation is under stress (e.g., in water, energy, or soil oxygen), at least seasonally. Such places often mark the transitions of biomes, or ecotones, where terrain-driven local variations in the stressors may create different plant habitats. The position of the tree line is an example; the tree line elevation and the associated transitional plant community composition are markedly different on south versus north facing slopes (Dearborn & Danby, 2017; Elliott & Cowell, 2015; Elliott & Kipfmueller, 2010; Johnson, 1848; Krause et al., 1959). Another example is the forest-savanna transition marked by increasing seasonal drought; here the valleys support forests and hills
support grasslands (Bucci et al., 2008; Cole, 1992; Dubs, 1992; Furley, 1992; Grogan & Galvão, 2006; Jirka et al., 2007; Ratter, 1992; Rossatto et al., 2012; Thompson et al., 1992; Villalobos-Vega, 2010). All the examples in Figures 2 and 3 are places where water or energy is limiting (too little or too much). This is a hypothesis to be tested as discussed in section 4.2.

To assess the global extent of terrain influence (do they add up to a substantial area of the land surface?), we may outline and exclude areas where it matters little (positions 1 and 9 in Figure 6a and positions 6 to 9 in Figure 6b), as attempted in Figure 7b. Regarding drainage, position 1 (ever wet, high relief, well watered, and well drained) is excluded in turquoise (Af climate in Figure 7a). Position 9 (arid climate and low relief) is

Figure 6. Schematic diagram of climate-terrain combinations where down-valley drainage (a) and difference in slope aspect exposure to the Sun (b) can potentially influence Earth System Model grid-level energy/water/carbon fluxes. ET = evapotranspiration.
excluded in red. It can be hypothesized that over the rest of the world, lateral drainage may differentiate hydrologic and plant conditions within an ESM grid cell. Regarding aspect, positions 7 to 9 in Figure 6b (tropical latitudes) are excluded in orange and position 6 (midlatitude and low relief) in yellow. It can be hypothesized that over the rest of the world, slope aspects may differentiate energy, water, and plant conditions within a grid cell. This qualitative assessment suggests that over the majority of the land surface, drainage and aspect potentially play a role in creating systematic and predictable variations in water and energy states, leading to systematic and predictable variations in ecosystem types and functions within an ESM grid cell. This is a hypothesis to be tested as discussed in section 4.2.

A key question is how these subgrid structures affect grid-mean states and fluxes communicated to the atmosphere in ESMs; that is, does the subgrid variation average out over a grid cell? The answer requires quantitative analyses accounting for the nonlinear processes linking water and energy availability to ET fluxes. Such an attempt was recently made by Rouholahnejad-Freund and Kirchner (2017). They assessed the impact of subgrid heterogeneity and lateral drainage on grid-mean ET, based on the Budyko model.
(Mezentsev, 1955; Turc, 1954) of mean annual ET as a nonlinear function of mean annual precipitation (P) and potential evapotranspiration (PET). Since P and PET (input variables) vary spatially across an ESM grid cell, there are two ways to compute the spatially averaged ET (output): averaging the input or averaging the output. The difference, termed the “heterogeneity bias,” turned out to be always positive, that is, ET is always overestimated by averaging the heterogeneous climatic inputs rather than averaging the heterogeneous outputs (which is what the real-world atmosphere does, at the scale of ESM grid cells). This finding is relevant to the aspect effect discussed above, whereby PET can vary substantially within an ESM grid cell on sunny versus shady slopes. Most ET models include substantial nonlinearities and will similarly lead to heterogeneity bias, although different models may lead to either overestimates or underestimates, depending on whether the nonlinearity is downward curving or upward curving. Rouholahnejad-Freund and Kirchner (2017) also assessed the effect of lateral drainage by adding upland surplus (P-PET) onto lowland P. They found that grid cell average ET is enhanced only where the upland has a surplus and the lowland has a deficit. However, the annual-mean Budyko model does not include seasonal variations in P and PET and the storage from wet to dry seasons, so the significance of the down-valley drainage is probably underestimated. Further global analyses and modeling that resolve seasonal and interannual storage dynamics are urgently needed, as discussed in section 4.2 on testable hypotheses and future tasks.

4. Representing Hillslope Hydrology in ESMs

Here we address the following questions. How should we divide a 20- to 200-km wide ESM grid cell into functional units that best reflect our understanding of terrain control on vegetation dynamics? How should we represent the connectivity among these units? And how should we parameterize the models based on physical principles and globally available observations? We explore how to represent the hydrologic structures and functions with conceptual clarity and parametric and computational efficiency. ESMs cannot and need not resolve every hillslope but can succinctly capture the effects of ridge-valley gradients and sunny-shady exposures.

We define a hillslope as the strip of land connecting a ridge line to its nearest stream in the valley (e.g., at the scales in Figures 2–4). Because ridges and valleys can be defined at nested scales (e.g., Montgomery & Dietrich, 1988) and because wetted channel networks are dynamic (e.g., Godsey & Kirchner, 2014; Marani et al., 2001), an inherent scale ambiguity is unavoidable. But for practical purposes, the degree to which a ridge and a stream can be resolved depends on the resolution of the digital terrain data, globally available at 1-arc second (~30 m) grids. In the discussions below, we assume that this grid resolution is sufficient to reveal the main terrain and vegetation structures within an ESM grid cell. We also assume that, with ESM grid scales approaching 20 km in the near future, the vertical climate zones can be neglected in characterizing land hydrology but can be captured as fine-scaled atmospheric forcing via atmospheric downscaling in steep terrain.

4.1. Subdividing an ESM Grid Cell Into Hydrologic Functional Units

4.1.1. Implicit and Equilibrium Hydrologic Partitioning

As reviewed in Clark et al. (2015), many attempts have been made to represent landscape heterogeneity within large ESM grid cells, from deterministic differentiation to statistical characterization of varying land surface properties. One particularly relevant effort is the application of TOPMODEL (Beven & Kirkby, 1979) toward partitioning a grid cell into drainage units. TOPMODEL uses a digital elevation model (DEM) to derive a topographic wetness index (TWI), the ratio of the logarithm of the contributing area above a unit contour length to the local terrain slope, reflecting a balance between topographic convergence (inflow) and local drainage (outflow). A map of TWI across a catchment, with high values in convergent and low-gradient valleys, and low values on divergent and high-gradient hilltops, elegantly and powerfully captures the essence of topography-driven hillslope and catchment hydrology. A large ESM grid cell can be partitioned into TWI zones, reflecting varying likelihoods of water deficit or excess, with infiltration and ET calculated on each zone separately. Various TOPMODEL-based approaches have been implemented in large-scale models (e.g., Clark & Gedney, 2008; Famiglietti & Wood, 1994; Koster et al., 2000; Niu et al., 2005; Walko et al., 2000) as reviewed in detail by Clark et al. (2015). What is missing in this approach is the explicit and dynamic treatment of surface and groundwater convergence (Beven, 1997; Beven & Freer, 2001). TWI is applied instantaneously to redistribute the grid cell water balance across the zones. Doing
so assumes that hydrologic equilibrium is achieved within a model time step (minutes in ESMs), without explicit calculations of the variable flow rates from ridges to valleys, the intermittent connectivity between them, and the delayed delivery from uplands to lowlands.

### 4.1.2. Explicit and Dynamic Hydrologic Partitioning

Several approaches have been used to partition large model grid cells into distinct drainage zones and to explicitly and dynamically route the flow down the gradient (see review in Clark et al., 2015). One simple approach is to divide an ESM grid cell into elevation bands (e.g., Nijssen et al., 1997). Over each band, standard 1-D (vertical) fluxes are computed, and the surplus is routed from high to low bands, both above and belowground. This approach has the ability to represent vertical climate zones in high-relief terrain, and it is particularly useful for modeling snow dynamics (see review by Clark et al., 2011). Because each elevation band contains both hillslope and channel elements, this approach requires tracking both hillslope and channel dynamics within each band, as well as routing surface water and groundwater among the bands. An approach that separates hillslope and channel processes can improve the conceptual clarity and computational efficiency.

This can be achieved by using the concept of “equivalent” or “representative” hillslopes (Ajami et al., 2016; Hazenberg et al., 2015), where hillslope and channel processes are treated separately. Using high-resolution DEMs, channels are defined first, and the hillslopes are delineated by connecting channels to ridgelines. The numerous and complex hillslope forms in an ESM grid cell are collapsed into a few theoretical hillslope types, such as convergent, uniform, and divergent slopes (Fan & Bras, 1998); convex versus concave (in plan or profile) forms (Troch et al., 2003), and headwater versus side slopes (Ajami et al., 2016; Hazenberg et al., 2015). Each representative hillslope type consists of multiple “columns” of varying widths and elevations, along which water is routed from high to low columns, and only the lowest column feeds into the streams. This approach separates hillslope processes from channel processes, and the theoretical abstractions allow analytical insights (Troch et al., 2003), but in practice it can be difficult to reduce the highly complex terrain into a few theoretical hillslope forms.

Recognizing this challenge, Chaney et al. (2018) introduced a statistical approach to group the hillslopes into natural clusters. Within an ESM grid cell, along the channel network, hillslopes are traced upward, each with attributes including geometry, aspect, soil, and vegetation. A cluster analysis is then used to reduce the large number of computed hillslopes into a (user specified) number of characteristic forms. In doing so, many of the covarying features of the landscape are implicitly considered in grouping the hillslopes (see also Flügel, 1995; Newman et al., 2014). These characteristic hillslopes have been represented in ESMs implicitly (Milly et al., 2014) and explicitly (Chaney et al., 2018; Subin et al., 2014). In the explicit approach, the characteristic hillslopes are discretized into multiple height bands above nearest drainage, each band further subdivided into different land cover soil types.

Recently, the concept of HAND (Height Above Nearest Drainage), first proposed by Nobre et al. (2011), has emerged as a powerful way to divide an ESM grid cell into distinct drainage zones according to the drainage position of each DEM pixel along the ridge-valley transect. The DEM elevation of a pixel is referenced to the global mean sea level (Figure 8a), an example in central Amazon, whereas the HAND value of the pixel is the elevation referenced to its nearest stream channel following surface flow directions (Figure 8b), thereby removing the regional topography and retaining only hillslope-scale topography. All stream pixels in a grid box have a HAND value of zero, although they occur at different elevations. The next HAND level represents the riparian zone immediately above and surrounding the streams, and so forth. Figure 8c shows a histogram of the HAND values from Figure 8b, which separates, without overlap, the distinct drainage zones in the grid box. In the central Amazon, these distinct positions are related to the water table depth (Figure 8d) and hence HAND can substitute for water table depth as a powerful delineator of plant community composition and species turnover, also shown in Figure 8d (Schietti et al., 2014).

In the context of this discussion, HAND offers a simple way to divide an ESM grid cell into zones with linear drainage relations, so that the higher HAND zone drains only into the successively lower zone via hillslope flow, and only the lowest zone interacts with the streams. An example is shown in Figure 8e, where a watershed is divided into five HAND zones modeled as five steps (Figure 8f), each with its own elevation and lateral distance above the stream, its own surface area, and its own contact width with the adjacent zones. This approach is similar to the representative hillslope concepts above but without fitting the
complex terrain into a set of hillslope types. Instead, a watershed or grid cell is represented by a single (giant) slope which can be very wide, and the area of the zones adds up to the total grid cell area, facilitating water and energy budget closure. Hillslope aspect calculations can be readily performed at the DEM pixel level, and each HAND band can be further segregated into multiple aspects, similar to a stadium with tiers of seats facing different directions. HAND has been used to map continental-scale flood inundation (Liu et al., 2018; Zheng et al., 2018), and it is increasingly employed as an ecosystem niche indicator (Moulatlet et al., 2014; Schietti et al., 2014). Global high-resolution DEM data have been processed to derive HAND values (Yamazaki et al., 2017) which can be readily applied.

Figure 8. (a) Elevation above mean sea level in the Cuieiras catchment, central Amazon, (b) height above nearest drainage (HAND) over the same area, (c) frequency of HAND values for waterlogged, ecotone, and upland vegetation (all from Nobre et al., 2011), (d) schematic of a HAND profile in Reserva Ducke where plant composition changes with HAND and depth to water table (Schietti et al., 2014), (e) an example to illustrate HAND bins above a channel (light blue line) in a catchment in Idaho, and (f) the HAND bins represented in models of four zones with different elevation, area, and interface width. DEM = digital elevation model.
4.2. The Depth and Nature of the Lower Boundary

ESMs need to specify the model soil depth and the hydrologic conditions at this depth. Here we explore how to constrain this variable across the globe based on our knowledge of the depth structure of the CZ and globally available information such as climate, tectonic stress, lithology, topography, soil, and vegetation.

Past ESM land models, with their 1-D (vertical) construct, sought to represent the water storage in the plant rooting zone, assumed to be ~2 m for woody plants and ~1 m for herbaceous plants (e.g., Sellers, Randall, et al., 1996; Sellers, Tucker, et al., 1996). It was further assumed that soil drains freely, constrained by the hydraulic conductivity at the water content at the base of the soil column. Agricultural soil surveys provided particle size information to compute hydraulic properties via pedotransfer functions (e.g., Clapp & Hornberger, 1978; van Genuchten, 1980; although these functions, based on midlatitude soils, cannot represent the deep weathered tropical clay which “holds on to moisture like clay but drains like sand”; Tomasella et al., 2000). Such shallow and freely drained model soil columns have been shown to hold insufficient storage for continued ET in the dry season in the Amazon and other regions of the world with a strongly seasonal climate (e.g., Baker et al., 2008; Brunke et al., 2016; Fan et al., 2017; Kleidon & Heimann, 1998; Kuppel et al., 2017; Miguéz-Macho & Fan, 2012b; Milly & Shmakin, 2002; Nepstad et al., 1994), prompting several inverse-modeling studies to estimate the “effective” root zone depth necessary to support satellite-observed leaf area, based on seasonal precipitation and atmospheric ET demand (e.g., Fan et al., 2017; Kleidon & Heimann, 1998; Wang-Erlandsson et al., 2016; Yang et al., 2016). These inverse estimates place integrated (atmosphere, vegetation, and soil) plant water constraints on the necessary model soil depth, but the persisting assumption that the land drains freely requires unrealistically deep soil column where the water table is within the 2- to 3-m soil column and plant rooting depth is restricted by the shallow water table (Figure 5), such as in wetlands and river valleys. Such estimated deep soil column certainly meets ET demand, but it requires large carbon allocation toward root growth, biasing ESM carbon budget calculations.

As we advance ESM land hydrology beyond the 1-D construct, and as ESMs attempt to represent groundwater and river dynamics and the ecosystems they support, the depth sufficient for ET may not be sufficient to encompass the zone of lateral groundwater flow that sustains stream baseflow and regulates regional aquifer systems. The zone of significant flow can occur below the rooting depth in the uplands, especially in the tropics where it can be tens to hundreds of meters deep (Ollier & Pain, 1996). Buss et al. (2013) report fractured flow at least 37 m deep in the volcanic terrain at the Luquillo CZO in Puerto Rico and 20 m below the local stream, exporting water out of the catchment through the subsurface flow paths that bypass local stream outlets. On the windward side of a volcano in Hawaii, Goodfellow et al. (2014) report weathering of basalts to >40-m depth. This depth is even greater in large sedimentary basins that host the major aquifer systems of the world, such as the High Plains in the United States where groundwater is recharged from the streams draining the Rockies hundreds of kilometers away and where the aquifers supply the vast irrigated agriculture that fills the nation’s “breadbasket” (Weeks et al., 1988; Winter et al., 1998). It seems that knowledge of the depth of significant porosity and permeability, or, equivalently, the depth of the CZ, is the most meaningful way to constrain ESM soil depth.

The depth to the fresh bedrock has been estimated globally (e.g., Hengl et al., 2017; Pelletier et al., 2016; Xu & Liu, 2017) integrating various observations and models. These products for the first time offer a terrain-based, versus the earlier climate plus biomass based, global estimate of the depth to negligible fluid storage and motion and thus an immediately useful product to constrain ESM soil depth. However, there remains the conceptual difficulty of drawing a physical line, above which there is fluid circulation and below which there is not. While this may occur in landscapes underlain by fresh bedrock, the transition from the high porosity/permeability surficial materials to the “tighter” deeper materials (mostly due to compaction in sedimentary basins) is usually fuzzy and transitional (Figures 4b–4d and 4f). This is more than a point of view; it also has physical implications. The decrease in porosity and permeability with depth results in stratified groundwater flow rates and residence times, generally faster near the surface and slower at increasing depths. The slow deep flow imparts temporal persistence to the deeper moisture and baseflow to streams. In places with strong dry seasons and interannual droughts, it is the deep flow, which may be tens of meters deep, that delivers the water surplus from prior wet months and years to maintain valley groundwater and streamflow in dry months and years (Hodnett et al., 1997a, 1997b; Ogden et al., 2013). In particular, this deep
long-distance groundwater flow is the main source for streams and wetlands in arid regions (e.g., Jobbagy et al., 2011; Schaller & Fan, 2009), harboring ecosystem refugia with disproportionate ecological importance in an otherwise hostile and barren landscape (Fan, 2015; McLaughlin et al., 2017). Therefore, where deep hydrologic memory is needed to support ecosystems through droughts, it is necessary to account for the stratified flow system that results from the gradual reduction in porosity-permeability with depth. Thus, instead of framing the problem as one of the definitive depths, it may be useful to frame it as one of the changes in permeability and porosity with depth, with a characteristic vertical scale.

In general, porosity and permeability of continental crusts decrease with depth, and over depths of kilometers, the decrease appears linear in competent rocks such as sandstone and exponential in less competent rocks such as shale and mudstone (e.g., Kuang & Jiao, 2014; Manning & Ingebritsen, 1999; Saar & Manga, 2004; Shmonov et al., 2003; Stober, 2011). At depths of meters to tens of meters that are more relevant to ESM time scales, it is widely observed and modeled that porosity and permeability decrease exponentially with depth (Beven & Kirkby, 1979; Cardenas & Jiang, 2010; Decharme et al., 2006; Fan et al., 2013; Heimsath et al., 1997; Jiang et al., 2009, 2010; Milly et al., 2014; Sakata & Ikeda, 2013; Wang et al., 2011). An exponential function with monotonic decrease with depth does not represent the sharp transitions or local reversals in porosity and permeability that are frequently found in sedimentary rocks, but it preserves the overall trend while integrating fluid flow over depth, circumventing the difficult problem of parameterizing fine-scaled vertical heterogeneity. The exponential function also has other advantages. It represents the stratified flow system and preserves its dynamics across a range of temporal scales. By doing so, it also captures the negative feedbacks between the hydraulic head and the flow rate, whereby the lower head shifts the zone of flow to deeper and slower paths. This further reduces flow rate, contributing to the commonly observed long tails in streamflow recession curves (Eltahir & Yeh, 1999) and to the wide range of temporal scaling observed in chemical and isotopic tracers (Jasechko et al., 2016; Kirchner & Neal, 2013; Kirchner et al., 2001). Practically, it reduces model sensitivity to the choice of soil/regolith depth, and it allows for analytical integration of groundwater flow as a 2-D problem, reducing computation and storage given the large demands on both in global ESM simulations.

If the problem is framed as one of the changes with depth, then our knowledge of lithology, topography, soil, vegetation, and climate history can be used to constrain the rate of decrease or the e-folding depth in the case of the exponential function. Given that hillslope-resolving topography is readily available globally, the e-folding depth has been formulated as a function of terrain slope (Fan et al., 2013), because the steepness of the land directly influences infiltration and weathering that thicken the CZ, versus surface runoff and erosion that thin the CZ (Gilbert, 1909; Hooke, 2000). This approach was adopted in recent continental-scale studies (e.g., Gleeson et al., 2016; Martinez et al., 2016; Maxwell & Condon, 2016; Zhang et al., 2016). Going forward, multiple terrain attributes need to be considered besides steepness (e.g., Tesfa et al., 2009). Given the fundamental importance of rock lithology (e.g., Bazilevskaya et al., 2015; Lebedeva & Brantley, 2017), tectonic stress (e.g., St. Clair et al., 2015), uplift and erosion rates (e.g., Rempe & Dietrich, 2014), the climate (e.g., Anderson et al., 2013; Goodfellow et al., 2014), climate history (e.g., Buss et al., 2013), and vegetation (e.g., Brantley, Eissenstat, et al., 2017; Roering et al., 2010) as controls on CZ structure, a comprehensive approach considering these factors is needed. A useful question is whether existing global estimates of CZ depth, e.g., Pelletier et al. (2016) and Xu and Liu (2017), can inform global estimates of how porosity/permeability change with depth.

4.3. Lateral Hydrologic Connectivity Within an ESM Grid Cell

4.3.1. Lateral Flow From High to Low Drainage Zones

By dividing an ESM grid cell into distinct drainage zones (e.g., Figure 8f), a traditional 1-D soil hydrology model can be applied over each zone first, and the surplus, at each time step, can be routed down the drainage gradient toward the channel along surface and subsurface paths. The simplest way to compute lateral flow along a hillslope is the kinematic wave approximation, where the terrain slope substitutes for surface water slope for routing surface runoff and streamflow and for water table slope for driving lateral groundwater flow. Beven (1981) reviewed the validity of this approach for groundwater flow, suggesting that it is best used in steep terrain with shallow bedrock where a perched water table nearly parallels the land surface. For global modeling representing the full range of hydrologic settings, the kinematic wave method must be used with caution.
Darcy's law is a fundamental physical law describing fluid flow in porous media driven by the hydraulic potential. Darcy's law has the same physical foundation as Ohm's law of electric currents driven by electric potentials, Dalton's law of evaporation driven by vapor pressure differences, and Fick's law of heat conduction driven by thermal gradients and molecular diffusion driven by concentration gradients. In Darcy's law, because the flow rate depends on the head, and the resulting head depends on the flow rate, a negative feedback ensues that accelerates drainage at times of high water tables and decelerates drainage at times of low water tables, leading to the characteristic asymptotic relaxation of the water table following a recharge pulse. Together with the decrease in porosity and permeability with depth, this negative feedback further prolongs deep soil water storage and hill-to-valley transfer into the dry season.

This negative feedback cannot be achieved with the kinematic wave approximation where the driving force (terrain slope) remains constant regardless of the water table slope. Because the land is always steeper than the water table slope, particularly in drier regions, using the land slope to drive flow depletes the water storage too soon, undermining a critical aspect of hill-valley feedback: the slow and delayed release of groundwater during the dry season. As the water table is explicitly tracked in some ESM land models (e.g., Community Land Model), Darcy's law can be readily applied. With an exponential decrease in porosity-permeability downward, Darcy flow can be integrated analytically (e.g., Beven & Kirkby, 1979; Fan et al., 2007), efficiently capturing both negative feedbacks.

4.3.2. Hillslope-Channel Exchange

The exchange between groundwater and streams can proceed in both directions (Winter et al., 1998). As illustrated in Figure 9a, across the hydrologic landscape from continental divides to the coastal oceans, a river can lose its water to the sediments or rock fractures below (Figure 9b, losing stream), or gain from the higher water table at the foot of the hillslopes (Figure 9c, gaining stream), all depending on the stream-groundwater level difference. This gradient can alternate in space along the same stream, such as from headwaters to lower reaches, and can also alternate in time, such as from rainy to dry periods. To represent this dynamic exchange across space and time, it is necessary to formulate the problem as potential-driven flow described by Darcy's law, as in the U.S. Geological Survey's MODFLOW model (Harbaugh et al., 2000). The barrier to the exchange, termed river hydraulic conductance, depends on the surface-groundwater contact area (river width X length in a grid cell, neglecting depth) and the riverbed permeability, all observable quantities, albeit poorly quantified over large areas.

Two additional feedback mechanisms are captured by the MODFLOW approach through the hydraulic barrier (river hydraulic conductance). First, riparian groundwater cannot instantaneously discharge to streams, limiting groundwater loss and prolonging the supply. Second, it is well known that, in humid climates, the stream channel network expands as the water table rises and intercepts a larger land area (higher river conductance) and contracts as it falls (e.g., Eltahir & Yeh, 1999; Godsey & Kirchner, 2014; Jenсен et al., 2009; Marani et al., 2001). In dry and seasonally dry climates, this may be expressed as dry stream reaches (shorter active channels) in dry seasons and drought years (Lovill et al., 2018). The longer/wider streams accelerate groundwater discharge when the water table is high, and as the water table falls, the network shrinks and decelerates groundwater discharge, preserving storage and prolonging the recession. All of these self-regulating mechanistic “valves” in groundwater-surface water exchange act as delay mechanisms, so that the wet-season surplus stays in the system longer than it would otherwise.

In the lower floodplains of large rivers, an important mode of groundwater-river exchange is floodwater infiltration into bank and floodplain sediments, illustrated in Figure 9d. At stage A, the river is gaining, but at stages B and C, it is losing and filling bank and floodplain sediments. This bank and floodplain storage dampens peak flow and releases stored water to rivers and floodplain lakes long after the passing of flood waves. Floodplain storage and delay has been documented in the Amazon (e.g., Bonnet et al., 2008; Borma et al., 2009; Cullmann et al., 2006; Lesack, 1995; Lesack & Melack, 1995; Miguez-Macho & Fan, 2012a; Richey et al., 2011) where up to 30% of the river flow is estimated to have passed through the floodplain (Richey et al., 1989, 1989). In drier climates with a deeper water table and larger water storage capacity in the unsaturated sediments, river leakage is the primary groundwater recharge, as documented in the largest inland floodplains in Africa. For example, in the Okavango, 80–90% of the seasonal floodwater infiltrates, recharging groundwater and sustaining vast and vibrant wetland ecosystems in an arid climate (e.g., Bauer et al., 2006; McCarthy, 2006). In the Sudd, where the Nile River tops its banks annually,
flooding raises the groundwater as evidenced by the large seasonal cycle in water table depth with little local rainfall (e.g., Mohamed et al., 2006).

The above modes of groundwater-surface water exchange can be readily captured in ESMs via Darcy’s law (the MODFLOW approach) without user decisions and parameter tuning, such as calibrating a “delay parameter” in moving groundwater to rivers (e.g., Decharme et al., 2010). Using Darcy’s law for surface water-groundwater exchange without parameter tuning, Pokhrel et al. (2013) demonstrate that seasonal
water storage in the Amazon is closer to GRACE water storage change estimates, with peaks shifting later and wet-season storage lasting longer, compared to the free-draining and one-way, instant placement of soil drainage into the rivers. The groundwater-to-river delay varies in space and time and across scales. It is a natural outcome of potential-driven flow operating under dynamic head gradient and resistance, which no single parameter can fully encompass.

4.4. Lateral Hydrologic Connectivity Among ESM Grid Cells

Aboveground, river flow is routinely routed among grid cells in ESM land models following river flow directions. In most of these models, as the river passes through a grid cell it collects the drainage loss from the 1-D and freely drained model soil column, but it does not lose water to the grid cell. As discussed above, river leakage loss to the sediment is known to support the vast and seasonally dynamic wetland ecosystems in inland basins such as the Okavango, the Sudd, and the Pantanal, and it is a large term in the river water budget in these settings. Thus, allowing leakage loss is equally important to allowing seepage gain as the rivers pass through a grid cell.

River leakage loss in the headwaters (Figure 9e) also drives a deeper groundwater circulation. Stream bed infiltration reaches the water table and flows toward regional discharge zones along deeper and longer paths, resurfacing at larger channels down gradient, which tend to lie below the regional water table (Shen et al., 2016). As illustrated in Figure 9f, groundwater flow can occur at multiple scales (Tóth, 1963), with shallow and short “local flow” to local streams, deeper and longer “intermediate flow” to higher-order streams down gradient, and even deeper and longer “regional flow” to remote discharge zones such as coastal wetlands (Figure 9g). The impact of the deeper flow on the basin water budget varies greatly, from >90% of local P-ET that leaves the basin through the subsurface path ( exporting groundwater, Figure 9f) at one end to streamflow eightfolds greater than its local P-ET (importing groundwater) on the other, among the 1,555 U.S. basins analyzed by Schaller and Fan (2009). This long-distance groundwater transfer to downgradient ecosystems, termed regional groundwater subsidy (Jobbagy et al., 2011), is particularly important in drier landscapes (Figure 9e, lower panel); here the headwater streams lose water to the regional water table and that water only resurfaces at the lowest end of the regional gradient. Deep and long-distance groundwater flow can also be important in humid landscapes that favor deep weathering; Genereux and Jordan (2006), Genereux et al. (2005, 2002) report significant intercatchment groundwater transfer in the humid lowland forests of Costa Rica, and Buss et al. (2013) report deep fractured flow 20 m below local stream beds at the Luquillo CZO in Puerto Rico, stating that “not all the water in the watershed is discharged to the stream.”

The importance of groundwater flow among grid cells is a function of scale (Krakauer et al., 2014); larger basins are more self-contained, with nearly all P-ET that leaves the basin passing the river outlet, but small low-order streams collect only the shallow flows, with a large portion of the water budget leaving the catchment via deeper paths bypassing the streams. For example, these deep flow paths are estimated to be 88% of the annual budget of a 13 km² forested catchment in southern Amazon (Neu et al., 2011). As ESMs adopt finer grids, the significance of among-grid-cell groundwater flow will increase. It is not difficult to calculate this flux given the mean water table of the grid cells. This flux is likely to be slow because it occurs in the deeper, less permeable rocks and sediments. But it is steady, reflecting and regulating the long-term dynamics of storage at interannual, decadal, or century scales. Spatially, this long-distance groundwater flow connects headwater losing streams with lower-reach gaining streams (Figure 9e), facilitating water budget closure within and among ESM grid cells.

Groundwater discharged at the lower end of long regional gradients is responsible for the belt of freshwater wetlands along the Atlantic and Gulf coasts (Figure 9g), fed by countless springs of aged groundwater (reviewed in Schaller & Fan, 2009; Fan & Miguez-Macho, 2011). Groundwater emerges before reaching the coastline because the sea level is the ultimate baseline for continental drainage. As the sea level fell more than 100 m at the last glacial maximum, coastal wetlands moved offshore on the exposed continental shelves, and as the sea level rose, groundwater was backed up and wetlands moved inland (Faure et al., 2002). If ESMs are to predict, instead of prescribing, the major wetlands of the world, then long-distance groundwater flow among grid cells, ultimately tied to the global sea-level, would be a key mechanism necessary to predict the spatial distribution and temporal dynamics of wetlands in response to major shifts in the Earth’s climate system.
4.5. A Land-Based Land-Model Grid System

The discussions so far conformed to the standard rectangular atmospheric grids of ESMs. Surface and shallow subsurface drainage are organized by a hierarchy of hydrologic units (hillslopes, low-order catchments, and higher-order basins), and subsurface flow is controlled by the geologic and sediment structures. Neither conform to the Cartesian coordinates convenient for mathematical manipulation. In the earlier 1-D land model construct, this fact hardly matters. But as we begin to recast the problem into 2-D or 3-D and account for lateral flow and deep flow, it would be natural to adopt a land-based grid system. Unlike the atmosphere and the oceans, landscape structures are permanent over ESM time scales, making them a natural framework for defining land grids.

Aboveground, a “grid cell” can be a catchment of certain ordered streams, which is self-contained in surface and shallow subsurface convergence, facilitating mass balance closure and correct flow routing among the grid cells. With rectangular grids, a point near the edge of a grid cell may drain into the neighbor cell, or the stream in one cell can collect runoff from hillslopes in neighbor cells, a problem alleviated by using catchments as grid units. Catchment delineation is available globally (e.g., Lehner & Grill, 2013; Verdin, 2017), and this idea has been applied in several large-scale models (e.g., Beighley et al., 2009, 2011; Goteti et al., 2008; Koster et al., 2000; Yamazaki et al., 2011, 2009), in which the land-atmosphere flux exchange is computed on overlapping areas of the two grid systems. Besides conceptual and computation clarity, a catchment-based grid system will also yield ESM simulations of hydrologic conditions that are readily testable using the large number of streamflow observations around the world. The results will also be more meaningful to water resource managers whose jurisdictions are often defined by watershed boundaries.

Belowground, the water table gradient sets the fluid in motion, but the flow path can be distorted by the geologic structure (e.g., Bense et al., 2013; Fan et al., 2007; Hartmann, 2016; Meerveld & Weiler, 2008; Rempe & Dietrich, 2014). Groundwater basins are often incongruent with river basins (e.g., Schaller & Fan, 2009), and topography alone offers incomplete guidance, particularly in karst terrain (Ford & Williams, 2007; Worthington et al., 2016) where the cave and conduit network can reach >500 km (e.g., the Mammoth Cave System in Kentucky; Groves & Meiman, 2005) through which groundwater moves as fast as rivers. Recent work including karstic structures significantly altered the modeled hydrologic partitioning (Hartmann et al., 2015, 2017; Longenecker et al., 2017; Rahman & Rosolem, 2017). The karst example illustrates that many of the improvements in large-scale models would be futile without acknowledging such first-order geologic controls on groundwater flow. Global lithologic maps differentiating geologic provinces and rock types (Hartmann & Moosdorf, 2012), with detailed karst maps (Chen et al., 2017), and global estimates of permeability based on rock types (Gleeson et al., 2014; Huscroft et al., 2018) already exist. It is a logical step to “overlay” the geologic units onto the surface drainage units to jointly define a land-based land grid system that acknowledges the natural hydrologic plumbing network above and below the land surface.

5. Conclusions and Testable Hypotheses

In this synthesis paper, we attempted to answer the following questions: (1) What are the first-order structures and functions of hydrologic processes that organize water and energy across the landscape? (2) Where in the world do these structures and functions matter, and do they manifest themselves over large ESM grid cells, or do they simply average out? (3) How can we efficiently represent these structures and functions in ESMs, so that we can begin to test their large-scale significance? (4) What are the testable hypotheses regarding CZ structures and functions in the context of ESM predictions and global change research?

Regarding question (1), we conclude that among the myriad hydrologic processes across the wide spectrum of spatial-temporal scales, two terrain-induced hydrologic structures—the lateral drainage from hills to valleys and the aspect difference between sunny and shady slopes—are first-order controls on water and energy availability across the landscape. Regarding question (3), the implicit “representative hillslope” concept and the application of potential-driven flow formulations can allow us to capture these first-order hillslope structures and the mechanistic feedbacks that regulate seasonal to decadal-scale dynamics.

However, regarding question (2), we do not yet know the answer. Our knowledge of hillslope and catchment hydrology leads us to hypothesize that, in water and energy limited places and times, hillslope-scale
organization of water and energy can make a difference in ESM predictions. Below, we outline two sets of testable hypotheses in order to address questions (2) and (4). The first set will need to be tested through cross-network and cross-site synthesis efforts on the part of the hydrology, ecology, and CZ science communities and the second set to be tested through regional and global ESM modeling experiments.

H1. Ridge-to-valley drainage gradient and slope aspect difference are first-order organizers of CZ depth structure, water storage and flux, and vegetation across the landscape, under given climatic, lithologic, and tectonic regimes and histories. To guide focused synthesis activities, we expand this hypothesis into the following.

H1a. The depth and the porosity-permeability structure of the CZ, and thus the capacity for water storage and transmission, vary systematically from ridges to valleys and from sunny to shady slopes. The lack of a global view of CZ structure along drainage and aspect gradients remains the largest knowledge gap, which translates to large uncertainties in the capacity of the land to store and transmit water. Greater investment into “seeing” the subsurface and cross-network and cross-site syntheses is needed to provide the ESM community with critical guidance on how to parameterize ESM model soil depth and properties.

H1b. In regions with seasonal water shortage, a greater valley storage can support larger plants, higher productivity, and greater water use, and the hill-valley vegetation contrast is particularly pronounced during the dry season. An expression of this contrast is the mosaic of vegetation in the transition from tropical forest to savanna, with trees in valleys and grassland on ridges. The main hydrologic mechanism is down-valley lateral flow whereby upland surplus subsidizes lowland deficit (spatial carryover). Slow subsurface flow delays the delivery to valley ecosystems, arriving long after rain or snowmelt events (temporal carryover). Through such carryovers, drainage positions organize plant available water and physiological water use traits across space and time. A second mechanism for higher valley water availability is the larger water storage capacity due to thicker soils and sediments in valleys, in some places of the world, which is yet to be confirmed by testing H1a above.

H1c. In lowland areas with water excess (waterlogging), better drained hills support larger plants and higher productivity. Two feedback mechanisms further enhance the plant-drainage association: high productivity and water use lowers the water table, further improving drainage; in cold regions, better drained hills warm up earlier and use water earlier, further improving drainage and soil thermal status.

H1d. In regions with at least seasonal water shortage, shady slopes support larger plant forms and higher biomass, and the vegetation contrast across hillslope aspects is particularly sharp in the dry season (as grasses and herbs enter dormancy and as trees restrict water use by stomatal and root response). The transition between forests and grasslands in water-limited midlatitudes takes the form of a mosaic, with trees on shady slopes and grasslands on sunny slopes. A key hydrologic mechanism is the lower insolation and thus lower ET demand on shady slopes, reducing hydraulic failure and plant mortality. Another key mechanism is the higher water supply on shady slopes, due to their deeper soils and regolith which have greater water holding capacity, which is yet to be confirmed by testing H1a above.

H1e. In energy-limited regions, sunny slopes are warmer and support larger PFTs, and the contrast is particularly sharp at the beginning and the end of the growing season. The altitude of the tree line varies accordingly (higher on sunny slopes and lower on shady slopes). In the high latitudes, aspect differences are amplified by the low Sun angle, and the latitude of the transition between taiga and tundra vegetation varies accordingly. One mechanism is the longer growing day and growing season on sunny slopes. But another mechanism is in the subsurface: the thicker soil and regolith development, and thus the thicker thawed and drained depth on sunny slopes, which is yet to be confirmed by testing H1a above.

To test Hypothesis 1 above, cross-network and cross-site syntheses on current understanding of process controls on soil and regolith depth (for example, Pelletier et al., 2018) are needed. Such syntheses efforts, on where-how-which processes interact to create the modern-day CZ structure, will provide the core knowledge to extrapolate across the globe, utilizing globally available information on topography, uplift rates, tectonic stress fields (e.g., Heidbach et al., 2016), bedrock geology, climate, vegetation, and other relevant factors. Field observational campaigns are needed to map the CZ structure at endmember places and to test
theories of CZ evolution. Such global extrapolation is necessary for constraining ESM soil depth and parameters that are critically important to water storage and flow and thus to vegetation dynamics. Such an activity is synergetic with, and can reinvigorate, past hydrologic synthesis efforts to translate knowledge from observed to unobserved sites (e.g., Blöschl, 2006; Wagener et al., 2010). But here we call for more emphasis to be given to seeing the dark subsurface, which is difficult to observe but holds the key to understanding nearly everything we observe above the ground concerning the spatial structure and temporal dynamics of water and plants.

In addition, to test Hypothesis 1, global-scale, high-resolution and joint terrain and climate and vegetation analyses are needed. Global coverage and hillslope-resolving terrain data exist already, and global high-resolution vegetation data are available at seasonal to decadal time scales. A comprehensive mapping of vegetation distribution with regard to drainage position (as shown in Figure 2) and slope aspect (Figure 3), under different end member combinations of terrain and climate of the world (as shown in Figure 6), can now be readily performed, which can provide an unprecedented global perspective on the explanatory power of terrain structure on vegetation distribution.

H2. Implementing hillslope drainage and aspect effects will alter ESM predictions of water, energy, and biogeochemical fluxes from the land to the atmosphere in resource-limited places and times, through several direct and indirect mechanisms expanded below.

H2a. In water-limited regions, larger PFTs are associated with topographic valleys and shady hillslope aspects. In lowland regions with waterlogging and oxygen stress, larger PFTs are associated with topographic highs. In energy-limited regions, larger PFTs are associated with sunny hillslope aspects. By linking PFTs with the drainage and aspect they occupy, ESM land models will likely simulate higher plant productivity across the landscape, and the productivity is likely to be less sensitive to water and energy limitations. This is particularly so in places with seasonal water or energy limitations, because the favorable seasons will allow large growth and resource needs to be met in the stressed seasons.

H2b. By implementing ridge-to-valley groundwater convergence, which takes days to months, instead of instant drainage (as in the early 1-D land models), and by allowing snow and rain to stay on the shady slopes longer, the net effect will be the lengthened residence times of precipitation on land. Except for the ever-wet places of the world, this implies greater water availability for vegetation in the dry intervals between precipitation or melt events. Where water is limiting to vegetation, this implies higher ET flux to the atmosphere and lower river runoff to the oceans.

H2c. The temporal delay inherent in the hill-to-valley delivery is accentuated by several “self-preserving” mechanisms. By using Darcy’s law, by allowing stratified and deep flow (vs. a sharp cutoff depth), and by allowing two-way surface-groundwater exchange driven by the hydraulic potential and regulated by a dynamic river-groundwater connectivity, ESM land models will have longer hydrologic memory and persistence, to the benefit of the vegetation and aquatic ecosystems in water stressed seasons.

H2d. By extending model soil depth to include the weathered bedrock, informed by the knowledge of CZ structure, ESM land models will simulate a larger terrestrial water storage capacity, higher amplitude of seasonal and interannual change as seen by the GRACE satellite, and longer residence times of water on land. The deeper “model soil”, and a better articulated hillslope hydrology, will allow the extension of plant rooting depth to respond spatially and dynamically to deeper moisture, further modulating plant distribution and ET along hillslope hydrologic gradients.

H2e. By differentiating uplands from lowlands and sunny from shady slopes, energy fluxes to the atmosphere can also be altered through changes in surface albedo due to changes in vegetation, even if the grid-mean latent heat flux is unaltered. The positive albedo feedbacks associated with snow further enhance snow persistence on shady slopes.

H2f. Wetlands are fundamentally land drainage features, and poor drainage occurs in all climates. By accounting for lateral flow within an ESM grid cell, ESM land models can mechanistically predict the location and dynamics of local wetlands. By allowing leakage loss from large rivers and long-distance groundwater flow among ESM grid cells that is ultimately linked to the sea level, ESM land models can mechanistically predict large regional wetlands.

H2g. A land-based land grid system, defined by catchments for surface and shallow subsurface connectivity, and geologic structures for deeper subsurface connectivity, offers conceptual and computational
advantages. Such a system will render the hydrology in ESM land models more efficient, accurate, and testable at the grid cell level (catchments) with widely available streamflow observations. It will also make ESM predictions more useful for water resource managers.

H2h. By differentiating uplands from lowlands and sunny from shady slopes within an ESM grid cell, output from ESMs can deliver internally consistent, hillslope-scale hydrologic and ecosystem forecasts under future global change. This will expand the power and reach of Intergovernmental Panel on Climate Change future assessments to directly inform local policies and management decisions, bypassing the step of impact analyses through climate downscaling using different methods and assumptions.

To test Hypothesis 2 above, model experiments are needed to quantify the sensitivities of ESM simulations to the inclusion of hillslope hydrology. High-resolution, regional to continental-scale, off-line and coupled land-atmosphere model simulations have played a critical role in revealing how the finer hydrologic structures can impact soil moisture, ET, wetland distribution, and land-atmosphere fluxes and regional climate (e.g., Maxwell et al., 2007; Maxwell & Kollet, 2008; Miguez-Macho & Fan, 2012a, 2012b). They will continue to play a critical role in testing some of the hypotheses posed above. However, these high-resolution models are still far from resolving the hillslope-scale structures emphasized here; there are no comparable high-resolution data sets to support them (Beven & Cloke, 2012), and they are computationally demanding and hence not yet feasible to be directly implemented in global and fully coupled ESM simulations. Therefore, model experiments directly using ESMs are necessary. These experiments can involve a hierarchy of simulations of increasing complexity in terms of how the hillslopes are represented, how covariations of PFTs with topography are represented, and complexity of represented processes (e.g., carbon cycle and ecosystem demography). Additional experiments can include land-only versus coupled as well as transient historical simulations into the future. A systematic series of ESM “sensitivity” experiments, adding one hillslope structure and one covariation in soil regolith and PFTs at a time, will be useful toward testing each of the subhypothesis posed above. We hold the view that a new generation of land hydrology models that represent the fundamental scales and processes, and that are consistent in complexity and scale with the next generation of vegetation models tracking PFTs and the demographic structures within them (Fisher et al., 2018), have the potential to elevate ESMs to a new level of realism in predicting ecosystem responses and feedbacks to global change.

More than ever, we are aware of humanity’s power in shaping the future trajectory of this planet (Steffen et al., 2018). More than ever, we appreciate the residence and flow of fresh water on land for sustaining human societies and the natural systems on which humans depend. More than ever, we recognize the intricate linkages and feedbacks between the biotic and abiotic worlds, nearly all enabled by the common currency of flowing water. And more than ever, we have the necessity to foresee the future, via ESMs, so that actions can be taken to steer the planet on a sustainable path. With the dawn of a new century in AGU science and service to society, we link two communities across scales in the Earth Science, to translate what is learned from intensely measured field sites into Earth system level interactions that shape the trajectory of this planet.

Acknowledgments
This community synthesis project is supported by the National Science Foundation directly through an INSPIRE grant (NSF-EAR-1528298) and indirectly through CUAHSI cooperative agreement (NSF-EAR-0733521). We thank the CUAHSI leadership (Rick Hooper and Jerad Bales) and staff and NCAR leadership for their support and the many members of the hydrology, CZ, and ESM community who contributed to the discussions at the workshops and webinars. We thank Caio Reis Costa Mattos for making the maps in Figure 1. Finally, we thank Elene Blythe, Dennis Lettenmaier, Keith Lucy, and two anonymous reviewers for their constructive comments which substantially improved the manuscript. There are no new data or models in this study.

References


