Designing a suite of measurements to understand the critical zone

Susan L. Brantley[1], R. A. DiBiase[1], T. A. Russo[1], Y. Shi[2], H. Lin[2], Kenneth J. Davis[3], M. Kaye[2], L. Hill[2], J. Kaye[2], D. M. Eissenstat[2], B. Hoagland[1], A. L. Dere[1,a], A. L. Neal[4], Kristen M. Brubaker[5], Dan K. Arthur[1]

[1]{Earth and Environmental Systems Institute, Department of Geosciences, Pennsylvania State University, PA, USA}

[2]{Department of Ecosystem Science and Management, Pennsylvania State University, PA, USA}

[3]{Earth and Environmental Systems Institute, Department of Meteorology, Pennsylvania State University, PA, USA}

[4]{Earth and Environmental Systems Institute, Pennsylvania State University, PA, USA}

[a]{now at: Department of Geography and Geology, University of Nebraska Omaha, Nebraska, USA}

[5] {Department of Environmental Studies, Hobart and William Smith Colleges, Geneva, NY, USA}

Correspondence to: Susan L. Brantley (sxb7@psu.edu)

Abstract

Many scientists have begun to refer to the earth surface environment from the upper canopy to the depths of bedrock as the critical zone (CZ). Identification of the CZ as an integral object worthy of study implicitly posits that the study of the whole earth surface will provide benefits that do not arise when studying the individual parts. To study the CZ, however, requires prioritizing among the measurements that can be made -- and we do not generally agree on the priorities. Currently, the Susquehanna Shale Hills Critical Zone Observatory (SSHCZO) is expanding from a small original focus area (0.08 km^2, Shale Hills catchment), to a larger watershed (164 km^2, Shavers Creek watershed) and is grappling with the prioritization. This effort is an expansion from a monolithologic first-order forested catchment to a watershed that encompasses several lithologies (shale, sandstone, limestone) and land use types (forest, agriculture). The goal of the project remains the same: to understand water, energy, gas, solute and sediment (WEGSS) fluxes that are occurring today in the context of the record of those fluxes over geologic time as recorded in soil profiles, the sedimentary record, and landscape morphology.

Given the small size of the Shale Hills catchment, the original design incorporated measurement of as many parameters as possible at high temporal and spatial density. In the larger Shavers Creek watershed, however, we must focus the measurements. We describe a strategy of data collection and modelling based on a geomorphological and land use framework that builds on the hillslope as the basic unit. Interpolation and extrapolation beyond specific sites relies on geophysical surveying, remote sensing, geomorphic analysis, the study of natural integrators such as streams, ground waters or air, and application of a suite of CZ models. We are hypothesizing that measurements of a few important variables at strategic locations within a geomorphological framework will allow development of predictive models of CZ behavior. In turn, the measurements and models will reveal how the larger watershed will respond to perturbations both now and into the future.

1 Introduction

The critical zone (CZ) is changing due to human impacts over large regions of the globe at rates that are geologically significant (Vitousek et al., 1997a; Vitousek et al., 1997b; Crutzen, 2002; Wilkinson and McElroy, 2007). To maintain a sustainable environment requires that we learn to project the future of the CZ. Models are therefore needed that accurately describe CZ processes and that can be used to project, or "earthcast," the future using scenarios of human behavior. At present we cannot earthcast all the properties of the CZ , but rather must model individual processes (Godderis and Brantley, 2014). Even so, many of our models are inadequate to make successful estimates of first-order CZ behavior today, let alone projections for tomorrow. For example, we cannot *a priori* estimate streamflow even if we know the climate conditions, soil properties, and vegetation in a given catchment, because of difficulties characterizing how much water is lost to evapotranspiration and to groundwater (Beven, 2011). Likewise, we cannot *a priori* estimate the depth or chemistry of regolith on a hillslope even if we know its lithology and tectonic and climatic history, because we do not adequately understand what controls the rates of regolith formation and transport (Amundson, 2004; Brantley and Lebedeva, 2011; Dietrich et al., 2003; Minasny et al., 2008). Perhaps even more unexpectedly, we often do not even agree upon which minimum measurements are needed to answer these questions at any location.

Such difficulties are largely due to two factors: i) we cannot adequately quantify spatial heterogeneities and temporal variations in the reservoirs and fluxes of water, energy, gas, solutes, and sediment (WEGSS); and ii) we do not adequately understand the interactions and feedbacks among chemical, physical, and biological processes in the CZ that control these fluxes. This latter problem reflects the fact that the CZ (Fig. 1) is characterized by tight coupling between chemical, physical, and biological processes which exert both positive and negative feedbacks on surface processes. Modelling the CZ is fraught with problems precisely because of these feedbacks and because the presence of thresholds means that extrapolation from sparse measurements is challenging (Chadwick and Chorover, 2001; Ewing et al., 2006).

However, the result of these couplings and feedbacks is that patterns of measureable properties emerge during evolution of Critical Zone systems that are repeated from site to site despite variations in environmental conditions. Such patterns include the distributions across landscapes or versus depth of such observables as regolith, fractures, bacterial species, or gas composition. Gradients in some important observable properties (e.g., surface slope, chemistry of water and regolith) emerge as indicators of the evolution of the CZ and reveal aspects of the underlying complex behavior (brown boxes, Fig. 1). For systems experiencing negative feedbacks, such gradients are thought to move toward steady-state conditions, i.e., gradients that remain constant over some interval of time.

In Fig. 1, some of these important gradients are arrayed from left to right to indicate the increasing length of time it takes for each gradient in general to achieve such a steady state. In other words, a steady-state soil gas depth profile might develop more rapidly than a steady-state regolith chemistry depth profile. Different disciplines tend to focus on different emergent properties (different gradients), and thus tend to emphasize processes operating at disparate timescales. However, CZ science is built upon the hypothesis that an investigation of the entire object – the CZ – across all timescales under transient and steady-state conditions (Fig. 1) will yield insights that disciplinary-specific investigations cannot. In turn, such integrative study and modelling should allow deeper understanding of the patterns that characterize the CZ.

Given that the mechanisms driving CZ change range from tectonic forcing over millions of years, to glacial-interglacial climate change over thousands of years, to the recent influence of humans on the landscape, building a model of the CZ is daunting and no single model has been developed. Instead, suites or cascade of simulation models have been used to address important processes over different timescales (e.g., Godderis and Brantley, 2014). To enable treatment using such a suite of models, each setting for CZ research including CZ observatories (CZOs (White et al., 2015)) must grapple with the necessity of measuring the processes at different timescales to understand the dynamics and evolution of the system.

At the Susquehanna Shale Hills CZO (SSHCZO), we have been investigating this challenge by studying the CZ in a 0.08 km^2 watershed located in central Pennsylvania (the Shale Hills catchment, Fig. 2). At the same time, we have been developing a suite of models that can be interconnected to address broad overarching CZ problems (Duffy et al., 2014; Table 1). The focus of the effort has been the small Shale Hills catchment which was established for hydrologic research in the 1970s (Lynch, 1976) and was expanded with other disciplinary studies as a CZO in 2007 as part of a network of CZOs in the U.S.A. The small spatial scale of Shale Hills allowed the development of a diverse but dense monitoring network that spans disciplines from meteorology to groundwater chemistry to landscape evolution (Fig. 2). Given the small size, we referred to our measurement paradigm as "measure everything, everywhere". For example, we inventoried all of the ~2000 trees with diameter greater than 20 cm at breast height, drilled 28 wells (up to 50 m deep), sampled soil porewaters at 13 locations at multiple depths approximately every other week during the non-snow covered seasons for more than a year, and measured soil moisture at 105 locations (Fig. 2).

The approach at Shale Hills has been to develop understanding incrementally by studying CZ systems of increasing complexity. The catchment itself is situated on a single lithology (shale), which simplified the boundary conditions for models with respect to initial chemical and physical conditions. We have monitored at ridgetops (where water and soil transport is approximately one-dimensional (1D)), along planar hillslopes (transects where such transport is essentially 2D), and within swales and the full catchment (where transport must be considered in full 3D). Where possible, these observations have then been paired with 1D, 2D and 3D model simulations. Using the conceptualization of "1D, 2D, and 3D" settings in the catchment has allowed measurements and modelling to proceed in a synergistic fashion: the reduction of complexity in 1D and 2D sites enabled development of models but also focused our sampling schemes. For example, our model conceptualizations of soil formation were developed first for ridgetops (1D) and then for planar (2D) hillslope systems and have been highly influenced by our soil chemistry measurements on ridgetops and planar hillslope catenas (Jin et al., 2010; Lebedeva and Brantley, 2013; West et al., 2013; Ma et al. 2013). In some cases modelling and measurement proceed hand in hand while in others, the modelling lags. For example, soil measurements have been collected in hillslopes characterized by convergent water and soil flow regimes, i.e., swales (Jin and Brantley, 2011) and soil observations have been collected across much of the catchment, but soil formation models for swales or the entire catchment still remain to be developed.

In contrast to the soil formation models that have targeted the 1D and 2D sites, our models of water flow have been developed for the entire catchment (e.g., Qu and Duffy, 2007). In fact, study of an entire catchment with a hydrologic model is sometimes more tractable than for smaller sub-systems because the large-scale study allows a continuum treatment whereas treatment of smaller scale sub-systems within the catchment might require measurements of the exact positions of heterogeneities such as fractures, faults, low-permeability zones, etc.

The goal of the SSHCZO project now is to grapple with some of these down- and up-scaling issues by expanding the CZO from Shale Hills to the encompassing 164 km^2 Shavers Creek watershed (Fig. 3). The expansion was designed to allow investigation of a broader range of lithologies (sandstone, calcareous shale, minor limestone) and land use (agriculture, managed forest, minor development), and to test models at larger spatial scales. To enable understanding of the larger watershed, we chose to analyze a suite of smaller subcatchments in detail, each of which were selected to be the largest that still drain a single rock unit or land-use type. This allows evaluation of how much of our understanding from Shale Hills is transferable to other lithologies with different initial conditions but with the same climate. Additionally, we are making targeted measurements of the mainstem of the stream in nested catchments of differing size within the larger watershed, in order to upscale our site-specific models to a relatively complex watershed.

Despite its small size, Shavers Creek contains much of the variability in CZ parameter space found within the Susquehanna River Basin and the Appalachian Valley and Ridge province in general. By measuring in detail paired catchments of similar size but different underlying conditions, along with targeted measurements in nested catchments of differing size, we aim to test theories of CZ evolution, parameterize models (Table 1) in different settings, and explore approaches toward upscaling across different size watersheds.

To understand the interaction of WEGSS fluxes in Shavers Creek and its smaller subcatchments, it is necessary to move beyond the paradigm of measuring “everything, everywhere” (Fig. 2) to an approach of measuring “only what is needed”. This phrasing, although simplistic, should resonate with any field scientist: the choice of measurement design is at the heart of any field project. But when we study the CZ as a whole, we are asking, how does one allocate resources to measure and model the dynamics and evolution of the entire CZ system? This paper describes our philosophy of measurement in the CZO; our previous paper describes the modelling approach (Duffy et al., 2014). Obviously, due to the wide range of CZ processes across environmental gradients (Fig. 1), the specifics of our proposed sampling design will differ from such designs at other sites. We nonetheless describe the philosophy behind our approach to stimulate focus on the broad question: how can we adequately and efficiently measure the entire CZ to best learn about its evolution and function? To exemplify our design, we also describe the first part of our expansion from Shale Hills to a sandstone subcatchment within Shavers Creek.

2 Connections between model development and field measurements

The suite of models shown in Table 1 is designed to develop understanding over the entire CZ as an integral object of study, i.e., one system. Field measurements are prioritized and driven by data needs for developing models (e.g., Table 1) and model development is dictated by observations in the field. Hand in hand with this system-level approach, researchers from different disciplines also bring discipline-specific hypotheses to their research that are related to disciplinary gaps in knowledge. Thus, disciplinary-level hypotheses also drive CZO research and sometimes these hypotheses feed directly into the overall CZ suite of models. Furthermore, because our understanding of the complicated suite of CZ processes is still in its infancy, both baseline measurements and curiosity-driven sample collection are still vital to determine the important processes. Throughout, models and observations are allowed to evolve to enable the two-way exchange of insights needed to maximize CZ science.

Given all the needs for data, the sampling plan which is implemented in a CZO must provide both measurements to test disciplinary hypotheses and observations necessary to bridge across disciplines. Additionally, certain measurements such as geophysical and remote sensing surveys, catchment-integrating stream measurements, and time-integrating analysis of alluvial and colluvial sediments can be made along with model simulations to upscale across space (from limited point or subregion measurements to the whole watershed) and time (from limited temporal measurements to geological timescales).

Perhaps the largest difficulty in *spatially* characterizing the CZ in any observatory is the assessment of the extremely heterogeneous land surface, ranging from the assessment of regolith and pore fluids down to bedrock to variations in land use. Because the mixing timescales of biota, regolith, and bedrock are relatively slow (compared to mixing of atmospheric and surface water reservoirs), the assessment of the spatial distribution of biota, regolith, and bedrock properties is both important and extremely challenging (Niu et al., 2014). On the other hand, rapid changes in the atmospheric reservoir make robust atmospheric measurements technically difficult. The hydrologic state is intermediate, exhibiting large spatial and temporal variability.

In recognition of these difficulties, the project started at Shale Hills precisely because it is a catchment almost 100% underlain by Rose Hill formation shale and strictly managed as forestland. Surface heterogeneities at Shale Hills were largely related to hillslope position, colluvium related to the Last Glacial Maximum (LGM), fracturing, differences in sedimentary layers, and relatively limited spatial variations in vegetation. To understand the CZ at the Shavers creek watershed, on the other hand, we must grapple with a more complex set of variations related to differences in lithology, land use, climate change, and landscape adjustment to changes in base level due to tectonics, eustasy or stream capture (Fig. 3). Here, the term base level refers to the reference level or elevation down to which the watershed is currently being eroded.

In recognition of the new complexities within Shavers creek, the sampling strategy was designed not to be random but rather to be stratified based on geological and geomorphological knowledge. An implicit hypothesis underlying this approach is the idea that sampling can be more limited for a stratified approach based on geological (especially geomorphological) knowledge. For example, a first-order observation about hillslope morphology in Shale Hills based in long-standing observations from hillslope geomorphology is the delineation between planar slopes and swales: the former experience largely 2D nonconvergent flow while the latter experience 3D convergent flow of water and soil. Where many randomly chosen soil pits might be necessary if the delineation of swales versus planar hillslopes was ignored, when representative pits are dug to investigate these features separately, the number of pits can be minimized.

Another aspect of our stratified sampling plan is to complement measurements at Shale Hills by targeted measurements in two new subcatchments of Shavers creek chosen to represent two of the new lithologies in the watershed. Once again the stratification of the sampling design is dictated by geological knowledge: bedrock geology is known to exert a first-order control on WEGSS fluxes in the CZ (e.g. Duvall et al., 2004; Williard et al., 2005). The first such new subcatchment is forested and underlain only by sandstone. The second subcatchment for targeted measurements is currently being identified on calcareous shale. This second subcatchment will also host several farms and will allow assessment of the effects of this land use on WEGSS fluxes.

To upscale from subcatchments to Shavers Creek, the targeted subcatchment data will be amplified by measurements of chemistry and streamflow along the mainstem of Shavers Creek as well as catchment-wide meteorological measurements (Fig. 3). The upscaling will rely on the small number of sites chosen for soil, vegetation, pore fluid, and soil gas measurements in each subcatchment. To extrapolate from and interpolate between these limited land surface measurements, models of landscape evolution (LE-PIHM), soil development (e.g., Regolith-RT-PIHM, WITCH), distribution of biota (BIOME4, CARAIB), C and N cycling (Flux-PIHM-BGC), sediment fluxes (PIHM-SED), solute fluxes (RT-Flux-PIHM, WITCH), soil gases (CARAIB), and energy and hydrologic fluxes (PIHM, Flux-PIHM) will be used. In effect, the plan is to substitute "everything everywhere" with measurements of "only what is needed" by using i) integrative measurements (geophysics, LIDAR, stream, atmosphere), and ii) models of the CZ. As a simple example, a regolith formation model is under development that will predict distributions of soil thickness on a given lithology under a set of boundary conditions. Since much of the water flowing through the upland catchments under study in the CZO flows as interflow through the soil and upper fractured zone (Sullivan et al., subm.), use of the regolith formation model will enable better predictions of the distribution of permeability. Of course, the models will be continually groundtruthed against pinpointed field measurements. With this approach, water fluxes in the subcatchments and in Shavers creek watershed itself will eventually be estimated.

For clarity in describing the measurements in each subcatchment that are needed for the models, we have given names to arrays of instruments (Table S1). The array of instruments in soil pits (1 m x 1 m x ~2 m deep) and in trees near the pits along a catena is referred to as "ground hydrological observation gear" (Ground HOG). The Ground HOG deployments also are the locations for assessments of vegetation across transects. Geophysical surveys and geomorphic analysis using LIDAR are conducted to interpolate between or extrapolate beyond the catenas.

In addition to Ground HOG, the energy, water, and carbon fluxes are measured using "tower hydrologic observation gear" (Tower HOG). Ground and Tower HOGs are in turn accompanied by measurements of stream flow, chemistry and temperature, groundwater levels and chemistry. As discussed above, these streams and ground waters provide natural spatial and temporal integrations over the watershed and therefore provide constraints on the 3D-upscaled models.

Data from Ground HOG and Tower HOG will be used to parameterize and constrain model-data comparison and data assimilation. In fact, the choice of targeted measurements are derived at least in part from an observational system simulation experiment (OSSE) completed for the Shale Hills catchment using the Flux-PIHM model (Table 1) (Shi et al., 2015b). The OSSE evaluates how well a given observational array describes the state variables that are targeted by Flux-PIHM. Specifically, this OSSE (Shi et al, 2015b) emphasized water and energy fluxes for the catchment.

Prior to the OSSE, a sensitivity analysis was performed (Shi et al. 2015a) to determine the six most influential model parameters that were needed to constrain and produce a successful simulation. We defined "successful simulation" as one that reproduced the temporal variations of the four land surface-hydrologic fluxes (stream discharge, sensible heat flux, latent heat flux, and canopy transpiration), and the three state variables (soil moisture, water table depth, and surface brightness temperature) (Table 1) with high correlation coefficients and small root mean square errors. Once the six most influential model parameters were determined -- porosity, van Genuchten alpha and beta, Zilitinkevich parameters, minimum stomatal resistance, and canopy water storage -- the OSSE was then performed.

The OSSE evaluated which of the fluxes and state variables were most important in constraining those model parameters. Shi et al. (2015b) found that the calibration coefficients for the most important model parameters were most sensitive to observations of i) stream discharge, ii) soil moisture, and iii) surface brightness temperature. (Alternately, instead of brightness temperature, measurements could focus on sensible and latent heat fluxes.) The OSSE has also been validated with assimilation of field observations at Shale Hills (Shi et al., 2015b).

On the basis of this OSSE, we are targeting measurement of stream discharge, soil moisture, and surface brightness temperature for each of the SSHCZO subcatchments on shale, sandstone, and calcareous shale. These measurements should allow us to reproduce subcatchment-averaged land-atmosphere fluxes and subsurface hydrology adequately. Once the three subcatchments are parameterized, the models will then be upscaled to the entire Shavers Creek watershed using information from LIDAR, SSURGO, geological maps, geophysical surveying, and land use.

Currently, the OSSE has only been used for assimilation of water and energy data but is being expanded to include biogeochemical variables. In other words, our ultimate aim is to complete an OSSE for C and N fluxes in each subcatchment. In the long run, we could also extend the OSSE to assimilate data for other solutes and for sediments.

Modeling results from Shale Hills indicated that an accurate simulation of the sub-catchment spatial patterns in soil moisture were achieved using a relatively limited set of hydrologic measurements made at a few points (Shi et al., 2015a). Specifically, we had to measure i) stream discharge at the outlet, ii) soil moisture at a few locations, and iii) groundwater levels at a few locations. The soil moisture (ii) and groundwater (ii) data used to calibrate the model were from 3 nearly co-located sites in the valley floor. These sites (referred to as RTHnet on Fig. 2) were the only sites with continuous data at the time of model calibration (COSMOS data were not yet available). The measurements were averaged across the three RTHnet sites (see data posted at http://criticalzone.org/shale-hills/data/dataset/3615/) to provide one calibration point in the model. Extending from this calibration point to the entire catchment was attempted using data from the SSURGO database (http://www.nrcs.usda.gov/wps/portal/nrcs/main/soils/survey/). However, because of the coarse spatial data available in SSURGO, this was not successful for the very small Shale Hills catchment. Therefore, porosity, horizontal and vertical saturated hydraulic conductivity, and the van Genuchten parameters α and β were separately measured for each soil series and then were averaged for the whole soil column for each soil series (Supplement Table S2). These soil core measurements for each soil series were used to constrain the shape of the soil water retention curve for each soil series in the model.

The result of this effort was that for the monolithologic 0.08 km^2 catchment of Shale Hills, five soil series were identified and soil properties measured (Lin et al., 2006). As we proceed with work on the new subcatchments, one of two approaches will be used. First, it is possible that relatively few soil moisture measurement locations are required in any given catchment, as long as we can obtain soil hydraulic properties for each soil series. Using the SSURGO soils database, such measurements could be made to parameterize the model. Alternately, spatially extensive soil moisture measurements based on COSMOS may be adequate to infer the variations in soil hydraulic properties on a series-by-series basis or based on geomorphological criteria. The overall plan is to use i) SSURGO, ii) geomorphological constraints, iii) COSMOS, and iv) soil moisture measurements along the catenas to parameterize Flux-PIHM.

To the extent possible, we parameterize these PIHM models with datasets and then evaluate the models with different datasets. The phrase "data assimilation" gets at the idea, however, that with more and more complex models, the data and the model output become harder to distinguish. For example, the output calculated for a given observable from a complex model may be more accurate than any individual measurement of that observable. As model output is used to parameterize other models, such data assimilation obscures the difference between model and data. Considered in a different way, data assimilation provides a means to combine the strengths of both *in situ* observations and numerical models. Data assimilation can thus provide optimal estimates of observable variables and parameters, taking into account both the uncertainties of model predictions and observations.

As new types of observations are provided, we first evaluate PIHM model output against the new observations prior to calibrations to see if the current calibration predicts the new data. This comparison is ongoing for the Garner Run subcatchment. If the prediction is poor, this yields insight into the capabilities of our model under new conditions. If we discover that even with a new calibration we cannot successfully predict the new observations, we will incorporate a new module that describes a new phenomenon in PIHM. For example, discrepancies between model output and preliminary observations at Garner Run has led us to hypothesize that the distribution of boulders on the land surface – a phenomenon not observed in the Shale Hills catchment – must be incorporated into the PIHM models. By tracking which parameters must be tuned and which processes must be added, we gain insights into both the model and system dynamics, and we learn which parameters must be observed if we want to apply our model to a new site or a new time period.

3 Implementation in the Garner Run subcatchment

These discussions about the design of a sampling strategy can best be explained through examples. In this section we introduce the Garner Run subcatchment, one of the two new focus subcatchments planned within the Shavers Creek watershed. To exemplify the approach, we describe the setting and some preliminary observations and measurements from soil pits, vegetation surveys, and water monitoring.

3.1 Geologic, geomorphic, and land use context of Garner Run

A central underlying hypothesis of SSHCZO work is that the use of geomorphological and land use analysis can inform sampling strategy so that measurements can be limited in number. Therefore, we start by describing the current knowledge of the geomorphological setting of the Garner Run subcatchment and land use. The subcatchment drains a synclinal valley underlain by the Silurian Tuscarora Formation between the NW-SE trending ridges of Tussey Mountain and Leading Ridge (Figs. 3-5). The Tuscarora Formation, which locally consists of nearly pure sandstone with minor interbedded shales, is the ridge-forming unit that caps the highest topography in Shavers Creek watershed. The hillslopes of both Tussey Mountain and Leading Ridge are nearly dip slopes, i.e., the roughly planar hillslopes parallel the bedding in the sandstone (Fig. 4, 5). Indeed, subtle bedding planes can be observed in LIDAR-derived elevation data (Fig. 6B). The strong lithologic control on landscape form is manifested clearly in the high-resolution (1 m) bare-earth LIDAR topography.

The hillslope morphology of the Garner Run subcatchment also contrasts strikingly in several ways from that of Shale Hills. Most notably, the sandstone hillslopes of Tussey Mountain and Leading Ridge are nearly planar in map-view: they have not been dissected with the streams and swales common in the shale topography of much of Shavers Creek (Fig. 6). Hillslopes underlain by the Tuscarora Formation are also nearly 10X longer (300-600 m) than those underlain by other geologic units within Shavers Creek, including shales. In Shale Hills, for example, hillslopes are 50-100 m in length (Fig. 7). Furthermore, the hillslopes at Garner Run are less steep (mean slope = 12°-17°) compared to those at Shale Hills (mean slope = 14°-21°), despite having significantly stronger underlying bedrock.

The observation of steeper hillslopes in Shale Hills versus Garner Run is particularly curious given that both subcatchments are presumed to have experienced similar histories of climate and tectonism. If the two landscapes were in a topographic steady-state with local erosion rate equal to the same regional rock uplift rate, we would expect that the sandstone would have evolved to generate steeper slopes. Thus the shallower slopes on the resistant sandstone contradicts the general idea that erosion and transport of more resistant bedrock that produces larger grain-size sediment generally requires steeper hillslopes.

Two issues may explain this apparent contradiction. First, while the morphology of the Shale Hills catchment bears little resemblance to the underlying structure of steeply-dipping shale beds, the topography of Garner Run is nearly entirely controlled by underlying Paleozoic structure (Fig. 4). Specifically, hillslope angles reflect dip-slopes rather than morphodynamic equilibrium. Second, as a headwater stream in the Shavers Creek watershed, Garner Run is isolated from the regional base level controls that influence downstream catchments such as Shale Hills (Fig. 6).

Specifically, analysis of stream longitudinal profiles on Garner Run and the mainstem of Shavers Creek reveals prominent knickpoints at elevations of ~320 m and 380 m, respectively (Fig. 7). Such breaks in channel slope geomorphically insulate the upper stream reaches from the mainstem of Shavers Creek and could be consistent with different rates of local river incision into bedrock in the upper and lower reaches (e.g., Whipple et al., 2013). Published cosmogenic nuclide-derived bedrock lowering rates ranging from 5-10 m/Myr from similar nearby watersheds (Miller et al., 2013; Portenga et al., 2013) may be a good estimate for rates in Garner Run upstream of the knickpoint (Fig. 7). These rates are indeed 3-4 times lower than bedrock lowering rates inferred for the Shale Hills catchment (20-40 m/Myr) (Ma et al., 2013; West et al., 2014; West et al., 2013), which lies downstream of the knickpoint on Shavers Creek.

The origin and genesis of these knickpoints is likely due to some combination of the following: regional base level adjustment on the Susquehanna River since the Neogene (3.5-15 Ma) due to epeirogenic uplift (Miller et al., 2013), stream capture and drainage reorganization (e.g. Willett et al., 2014), or temporal and spatial variations in bedrock exposure at the surface (e.g. Cook et al., 2009). Testing these competing controls will require additional direct measurements of bedrock lowering rates with cosmogenic nuclides at Garner Run, in addition to bedrock river incision models that can account for both variations in rock strength and temporal changes in relative base level.

In addition to variations in structure, lithology, and base level, Quaternary climate variations have left a strong imprint on the landscape of Garner Run and Shavers Creek in general. While the relict of the periglacial processes at Shale Hills are mostly observed in the subsurface colluvial stratigraphy (West et al., 2013), at Garner Run these processes have left behind boulder fields, solifluction lobes, and landslides observed at the land surface (Fig. 6). Such features are found throughout central Pennsylvania south of the limit of the LGM (last glacial maximum) (Gardner et al., 1991). These features document a major reorganization of the uppermost CZ by processes such as permafrost thaw. For example, the Leading Ridge hillslope (the southern hillslope defining the Garner Run subcatchment, Fig. 5) is characterized by a hummocky topography at the 5-10 m scale, with abundant partially vegetated boulder fields. The other side of the catchment -- Tussey Mountain hillslope -- is steeper at the top, has greater relief, retains evidence of past translational slides, and contains open, unvegetated boulder fields. At the foot of the Tussey Mountain hillslope is a strong slope break that demarcates a low-sloping region characterized by abundant solifluction lobes, which appear to have accumulated as a large, valley-filling deposit (Figs. 6, 7). Such features were either not as active or their evidence has been erased or buried at the Shale Hills subcatchment.

Many of these geomorphological features have controlled or been imprinted on CZ processes and human activities in Garner Run. For example, the modern flow pathways for surface and groundwater in Garner Run are significantly influenced by the forcing factors of tectonism, climate, and anthropogenic activity. Flow pathways are influenced i) by topography inherited from geologic events from 10^8 years before present, ii) by variations in soil grain size as dictated by periglacial processes operating 10^4 years ago, and iii) by modern land use over the last $10^2 – 10^3$ y.

In terms of land use, the influence of anthropogenic activity in the catchment is relatively minor and consistent with the surounding region. Neither Shale Hills nor Garner Run subcatchments show signs of having been plowed or farmed in row-crop agriculture, although some grazing may have occurred. The top of one of the ridges in Shale Hills appears to define a field edge. Both subcatchments were forested for at least 100 years. Based on historic aerial photographs, both watersheds contain intact, closed canopy forests in 1938 and show no sign of obvious stand level disturbance since that year.. In the mid 1800s, significant quantities of charcoal were made in this region to run several nearby iron furnaces. Given that charcoal hearths have been identified in the subcatchments from lidar, the subcatchments were probably cleared in the mid to late 1800s as most available wood was used for charcoal making. This land use was also often associated with fires.

This short analysis of the geomorphology and land use highlights the influence of the forcing mechanisms (tectonism, climate, anthropogenic activity) that operate over a wide range of timescales and yet influence modern CZ processes. The CZO efforts document the importance of providing geologic and geomorphic context for investigation of the CZ.

3.2 Water and energy flux measurements at Garner Run: Tower HOG

Surface energy balance measurements (eddy covariance (EC) measurements of sensible and latent heat fluxes or upwelling terrestrial radiation/skin temperature) are needed to constrain Flux-PIHM (Shi et al., 2014). Measurements of precipitation, atmospheric state and incoming radiation are needed as inputs to the model. These measurements provide the data needed to simulate the catchment hydrology that is critical to understanding today's WEGSS fluxes. In addition, these fluxes are drivers for millennial-timescale landscape evolution (Fig. 1).

Instrumentation for measurements of water and energy flux measurements are designed as part of the "tower hydrological observation gear" – referred to here as Tower HOG (Table 2, Table S1). While the ideal plan would locate Tower HOG within the Garner run watershed itself, the remote, rocky, heavily wooded terrain makes this too challenging. Therefore, precipitation will be measured near Garner Run on a road crossing Tussey Mountain that is also the site of a pre-existing communications tower (see Fig. 3). A disdrometer (LPM, Theis Clima GmbH) and weighing rain gauge have been in use at Shale Hills since 2009 and 2006 respectively to measure precipitation. To measure precipitation amount at Garner Run, we are installing the simpler instrument (Pluvio^2, OTT Hydromet weighing rain gauge). Measurements will be compared to the National Atmospheric Deposition Program measurements and samples of rainwater. According to the nearest NADP site, Garner Run receives 1006 mm/y precipitation with an average pH of 5.0 (Thomas et al., 2013).

EC and radiation instrumentation (Table 2) will also be implemented on the pre-existing communications tower on the Tussey Mountain ridgeline (Fig. 3). Although located out of the subcatchment, the measurement footprint for the tower will be sensitive to fluxes from forests representative of those in Garner Run. The complex terrain at Shale Hills and Garner Run make EC measurements difficult to interpret in stable micrometeorological conditions. Since the primary energy partitioning happens during the day when the atmosphere is typically unstable, daytime sensible and latent heat flux measurements are sufficient to constrain the hydrologic modeling system. Daytime carbon dioxide flux measurements will inform the biogeochemical modeling system.

3.3 Vegetation mapping

Vegetation impacts today's WEGSS fluxes and is known to have influenced regolith formation and sediment transport over geologic time. As we study subcatchments to understand budgets, we seek to learn enough about vegetation to extrapolate WEGSS fluxes to the Shavers creek watershed. As described below, we once again use the geomorphological framework to design the measurement strategy for vegetation. We also want to understand some of the biogeochemical controls on fluxes of nutrients such as nitrate out of Shavers creek. Ultimately, an OSSE will be run to compare measurements to model predictions as a way to determine the important parameters for predicting carbon and nitrogen fluxes. It may also be necessary to determine the effect of individual tree species on N flux (Williard et al., 2005).

As part of the geomorphological measurement strategy, we mapped the vegetation in Garner Run subcatchment across the Ground HOG catena (ridge top, midslope, and valley floor positions on one side of catchment and one midslope site on the other side, Fig. 5). The objective of the catena-based stratified sampling design was to measure spatial variability in vegetation, i.e., under the assumption that landscape position was an important control on vegetation. These measurements set the stage for planned re-measurements to understand temporal variability. For example, future assessments will quantify above-ground biomass, an important carbon pool. Variability in forest composition, standing biomass, and productivity across a watershed is generally related to gradients in biotic and abiotic resources such as soil chemistry or structure, water flux, and incoming solar energy. Therefore, the relatively restricted vegetation analysis design (Fig. 5) will be upscaled based on the team's developing knowledge of the distribution of soils across the watershed as well as LIDAR-based estimates of tree biomass and seasonal patterns of leaf area index and tree diameter growth. Given that we have not yet run an OSSE for C or N fluxes, our measurements of vegetation are relatively broad to enable such future analysis.

Vegetation measurements are important not only for C and N fluxes, but also for water flux. At Shale Hills, seasonal variation in tree transpiration has been estimated using tree sap flux sensors (Meinzer et al., 2013). While we sampled many different tree species in multiple locations at Shale Hills (Fig. 2), a more restricted number will be sampled at Garner Run. For example, sapflux sensors are planned for only the midslope positions of Ground HOG (Fig. 5). While eddy flux and soil moisture dynamics provide estimates of total transpiration and evaporation, sap flux provides direct estimates of tree transpiration that can constrain model predictions of transpiration. Collectively, these measures will help evaluate Flux-PIHM model processes. In addition, all approaches to measuring water fluxes are imperfect; errors can best be constrained when multiple approaches are used.

In addition to these sapflux measurements limited to midslope pits, vegetation has been sampled in linear transects parallel to the slope contour at each of the four soil pits

(LRRT, LRMS, LRVF, TMMS, Fig. 5, Section 3.4), i.e., at Leading Ridge ridge top, Leading Ridge midslope, Leading Ridge valley floor, Tussey Mountain midslope, respectively. Each vegetation transect was 10-m along the direction perpendicular to the valley axis and ~700-1400 m parallel to the valley axis.

Measurements along the transects yielded vegetation and forest floor cover data for 4.1 ha in the subcatchment (Table 3). The transects provide vegetation input data for land surface hydrologic models, and also evaluation data for a spatially-distributed biogeochemistry model (Flux-PIHM-BGC, Table 1). In the transected area, 2241 trees >10 cm diameter at breast height were measured, mapped, and permanently tagged. Understory vegetation composition was measured at 5 m intervals along transects and coarse woody debris was measured in 25 m planar transects parallel to the main transect, spaced every 100-m. Forest floor cover was classified as rock (typically boulder clasts from periglacial block fall), bare soil, or leaf litter every 1 m along each transect, and the dimensions (a, b, c axes) of the five largest exposed rocks was recorded every 25 m. Forest floor biomass was measured every 25 m along transects by removing the organic horizon from a 0.03 m^2 area for laboratory analysis: samples were dried, weighed and measured for carbon loss on ignition.

The transect observations document variations in vegetation along the catena (Table 3), as well as spatial variation in vegetation at each position. For example, mean tree basal area (BA; the ratio of the total cross-sectional area of stems to land surface area) in the LRRT transect is 25.3 $m^2\ ha^{-1}$ with measurements ranging from 0 to79 $m^2\ ha^{-1}$. The subcatchment contains a dry oak-heath community type (Fike 1999), primarily consisting of chestnut oak (*Q. montana*), red maple (*Acer rubrum*), black birch (*Betula lenta*), black gum (*Nyssa sylvatica*), and white pine (*Pinus strobus*) in the overstory, with a thick heath understory of mountain laurel (*Kalmia latifolia*), blueberry (*Vaccinium sp.*) and huckleberry (*Gaylussacia sp.*) species, and rhododendron (*Rhododendron maximum*) along Garner Run.

The transect work also highlighted a type of measurement that we had not needed for Shale Hills but which our models and observations are showing is important in the new subcatchment: the fraction of land surface covered by boulders. At LRRT, 16% of points sampled every meter fell on rock. Furthermore, rock coverage at some transect points was as high as 100% or as low as 0%. Vegetation and surface rockiness data from transects will be combined with a suite of ground and remotely sensed measurements from the watershed such as slope, curvature, aspect, solar radiation, and soil depth to model vegetation dynamics from environmental conditions and interpolate vegetation structure in areas of the watershed not directly sampled. Future re-measurements along transects will allow assessment of carbon uptake in vegetation, as well as changes in forest composition and structure.

Additional key vegetation parameters will be assessed at the soil pits described in Section 4.4 and Table S2. These additional measurements include root distributions, leaf area index (LAI, described in the next paragraph), litter fall, tree diameter growth and tree sap flux. Root distributions are being measured at all four soil pits in Garner Run using soil cores to assess the high length densities near the surface. Root distributions, combined with soil water depletion patterns, can allow estimation of depth of tree water use over the season. Depth of tree water use, an input parameter in the PIHM suite of models, is currently derived from a look-up table (http://www.ral.ucar.edu/research/land/technology/lsm/parameters/VEGPARM.TBL) to determine the rooting depth of each land cover type. We will explore whether the use of field-measured rooting depth as model input improves the modeling of water uptake. In addition, profile wall mapping is being used to analyze the architecture, mycorrhizal colonization, and anatomy of deep roots. By characterizing and understanding the controls on root traits along a hillslope, we will eventually be able to use such observations to inform both models of water cycling (Flux-PIHM) and regolith formation (RT-Flux-PIHM, see Table 1).

At weekly intervals in the spring and fall and monthly during the summer, LAI will be assessed with a Li-2200 plant canopy analyzer (LI-COR Inc., Lincoln, Nebraska USA). The Moderate Resolution Imaging Spectroradiometer (MODIS) also provides remotely-sensed 8-day composite LAI (Knyazikhin et al., 1999; Myneni et al., 2002). The MODIS LAI product, however, has a spatial resolution of 1 km^2, which cannot resolve the spatial structure in LAI within small watersheds. The product also has a notable bias compared to field measurements (e.g. Shi et al., 2013). The LAI field measurements will be used for detailed information on leaf phenology, which is an important driver for the modeling of water and carbon fluxes for land surface and hydrologic models (e.g., PIHM, Flux-PIHM (Table 1)), and provides calibration or evaluation data for biogeochemistry models like Flux-PIHM-BGC (Naithani et al., 2013; Shi et al., 2013).

Another important value we must estimate is net primary productivity (NPP). With NPP it is possible to constrain carbon and nutrient fluxes in vegetation stocks, which can be large components of the overall budgets. To estimate aboveground NPP, we will measure annual variation in trunk growth with dendrobands emplaced on examples of each of the six dominant tree species near each soil pit site. In addition, traps at each soil pit will collect litter fall for assessment. One of the key model outputs of Flux-PIHM-BGC is NPP, which can be evaluated using these measured data.

3.4 Soil pit measurements and Ground HOG instrumentation

3.4.1 Soil observations

The uplands of the Garner Run subcatchment land surface falls into one of three categories: i) fully soil mantled with few boulders emerging at the ground surface, ii) boulder-covered with tree canopy, and iii) boulder-covered without tree canopy. The coarse blocks of the Tuscarora sandstone range in diameter from ~10-200 cm, making it challenging to excavate large soil pits (Table 3). To assess the spatial heterogeneity of soils in the Garner Run subcatchment, we therefore focused efforts on four soil pits: three on the north-facing planar slope of Leading Ridge (LRRT, LRMS, LRVF) and one mid-slope pit on the south-facing slope of Tussey Mountain (TMMS) (Fig. 5). Three pits were dug by hand until deepening was impossible (LRRT, LRMS, and TMMS). The Leading Ridge Valley Floor (LRVF) pit was dug by hand and then deepened using a jackhammer until the inferred contact with intact bedrock was reached. The pits were excavated in the following soil series: TMMS, LRRT and LRMS (Hazleton-Dekalb association, very steep), and LRVF (Andover extremely stony loam, 0-8% slopes). This deployment of observations in soil pits along a catena, with an additional pit on the opposite valley wall, is here referred to as "Ground HOG" (ground hydrological observation gear) (Fig. 5, Supplement Fig. S1, S2) and is the result of our focus on a minimalist sampling design.

This design was informed by observations at Shale Hills and the new subcatchment and by modelling conceptualizations. As discussed earlier, the Shale Hills subcatchment upland land surface falls into one of two categories: hillslopes or swales. In contrast, we observed little evidence for swales in Garner Run. All four pits in the new subcatchment were therefore located on roughly planar or somewhat convex-up hillslopes (see below). The rationale for the positions of the pits is as follows. First, regolith formation at a ridge top is the simplest to understand and model (see, for example, Lebedeva et al., 2007; Lebedeva et al., 2010) because net flux of water is largely downward and net earth material flux is upward over geological time. We are now developing Regolith-RT-PIHM to simulate regolith development quantitatively for such 1D systems, using constraints from cosmogenic isotope analysis (Table 1). The next level of complexity is a convex-upward but otherwise planar hillslope. The intent for Regolith-RT-PIHM is that it will be able to model hillslopes as 2D systems (e.g., Lebedeva and Brantley, 2013). Soil pits along a convex-upward but otherwise planar hillslope such as those described for Shale Hills (Jin et al., 2010) can be used to parameterize both 1D and 2D models of regolith formation. Third, while both planar hillslopes and swales are important at Shale Hills (Graham and Lin, 2010 ; Jin et al., 2011; Thomas et al., 2013) the lack of swales at Garner Run allow focus on just one catena in the minimalist design. (In fact the lack of swales in the sandstone catchment is one of the observations that we hope we can eventually explain). Finally, the importance of aspect on soil development and WEGSS fluxes has been noted on shale at Shale Hills (Graham and Lin, 2011; Graham and Lin, 2010 ; Ma et al., 2011; West et al., 2014), as well as on sandstones in Pennsylvania o(Carter and Ciolkosz, 1991). For that reason, Ground HOG includes one pit on the northern side of the catchment (Fig. 5).

We will use numerical models to explore regolith formation and to extrapolate to other hillslopes within Shavers creek watershed. This highlights the importance of understanding the soil to the CZ effort. Soil provides a record of both transport of rock-derived material as well as fluxes of water over the period of pedogenesis. For example, the pits at Garner Run are characterized from land surface downward by a thin organic layer, a rocky layer, a leached layer characterized by sand-sized grains with few large clasts, a sandy mineral soil with a thin layer of accumulated organic and sesquioxide material, and a deeper clay-rich layer with larger interspersed rock fragments (Supplement Fig. S3, Supplement Table S2). Depth intervals of the soil every 10 cm and from basal rocks show variations in chemistry (Supplement Tables S3, S4), and are being analyzed for grain size, organic matter, and mineralogy.

These soil observations yield further clues to the history of the landscape. The Garner Run subcatchment has been mapped to lie on Lower Silurian Tuscarora sandstone (Flueckinger, 1969). Interpreted as reworked beach sediments (Cotter, 1982), this sandstone has been mildly metamorphosed to a highly indurated quartzite. Bulk compositions of five rocks collected from the bottom of the five Ground HOG pits were averaged to estimate composition of the protolith (Supplement Table S3). These samples contain >96 wt. % SiO_2, i.e., very similar to published Tuscarora compositions (Cotter, 1982). Minor titanium (Ti), generally present in sandstones in highly insoluble minerals, was present in the parent (Table S3) and at even higher concentration in soils (Table S4). This enrichment in soil could be due to several processes during weathering: for example, retention of Ti from the protolith, losses of elements other than Ti, or addition of Ti to the soil. If Ti in the soil was derived from protolith, loss or gain of other elements in the sandstone can be calculated from the mass transfer coefficient, τ_{ij}, where i is Ti and j is an element that was lost or gained (Anderson et al., 2002; Brimhall and Dietrich, 1987). Assuming Ti in soil was derived from the protolith, $\tau_{Ti,j}$ values = 0 within error for Al, Mg, and Fe, indicating they were neither added nor depleted compared to Ti. In contrast, $\tau_{Ti,K} > 0$, consistent with addition of K to the soil (Fig. 8). Error bars on many of the elements are very large because of the variability in the low concentrations of all elements except Si and O.

According to published arguments for this formation in this region, the thin and poorly developed ridgetop soil is likely residual (Ciolkosz et al., 1990). In contrast, soils on hillslopes likely developed not only from rock in place but also from colluvium (Fig. 5). Furthermore, previous researchers have pointed out that soils in central Pennsylvania commonly show a brown over red layering that may indicate two generations of weathering, i.e., a previously weathered red layer which was then covered by a colluvial layer that experienced additional weathering (the brown layer) (Hoover and Ciolkosz, 1988). Although the soils here did not show a strong brown over red color signature (Supplement Fig. S3), clay-rich soil at depth may document soil formation before the LGM (Table S2). The addition of K to the soils, even in the residual soils at the ridgetop (Supplement Fig. S3), is another complexity. K could have been added as exogenous dust inputs which were very important during and immediately after glacial periods (Ciolkosz et al., 1990). Alternately, K-containing clay particles could have percolated downward from weathering of the overlying units such as the Rose Hill shale before it was eroded away (Fig. 4). Such movement of fines downward from the Rose Hill have been observed at Shale Hills (Jin et al., 2010): such particles could have been added to the underlying Tuscarora and then retained in the soil. In that case, the assumed protolith composition could be erroneous, especially if Ti was added from the downward infiltrating fines. K enrichment could also be explained by shales within the Tuscarora formation itself (Flueckinger, 1969). If these interfingered shales were the protolith of the observed soils, this would mean that our estimated protolith composition was K-deficient. Thus, soil analysis (Fig. 8) leads to interesting hypotheses that will be investigated.

3.4.2 Ground HOG

The Ground HOG instrumentation enables the *in situ* measurement of soil moisture and temperature, as well as gas and pore-fluid compositions, all at multiple depths (Fig. 5, Supplement Fig. S2). Ground HOG complements the atmospheric measurements at Tower HOG (Section 3.2). Because Ground HOG sites are difficult to access, measurements were automated to the extent possible. However, the lack of access to electricity and the cost of automated sensors (for CO_2 for example) meant that a completely automated monitoring system was unfeasible as well. Therefore, our final approach (Supplement Fig. S2) included a few automated components recording a continuous time series of data, coupled with additional components to be monitored manually, but with lower temporal resolution.

In selecting depths for soil sampling we wanted to instrument the site so that results could be compared across all watersheds. Thus, we focused on a depth-based (as opposed to horizon-based) sampling scheme. In addition, we wanted to emphasize surface soils that have the highest water and biogeochemical flux rates. These layers also have the strongest influence on the atmospheric boundary layer. At the same time, we wanted to also document deep soil processes critical to understanding weathering and subsurface flowpaths. Thus, our final depth distribution included samples at 10, 20, and 40 cm from the top of the mineral soil (we used the top of the mineral soil as the depth reference because the O horizon depth varies greatly across the sites and among land-use types) and 20 cm above the bottom of the soil pit (coded “D-20”). At these four depths we installed from 1 to 4 component devices of the Ground HOG in each pit.

Automated soil moisture and temperature sensors (Hydra Probe, Stevens Water Monitoring Systems, Inc. Portland, OR) were emplaced to monitor at 10, 20 and 40 cm depths on the uphill face of each pit (Fig. S2). In addition, TDR waveguides (Jackson et al., 2000) for manual point estimates of soil moisture were installed at the same depths plus D-20 on the uphill pit face and the left and right pit faces (facing uphill). Wave guides are paired metal rods on a single cable that conduct a signal for time-domain reflectometry. The rods are 20 cm long and hand-made (Hoekstra and Delaney 1974, Topp et al 1980; Topp and Ferre 2002). We placed 12 (4 depths x 3 pit faces) in each pit. The automated sensors were emplaced at depths expected to have the most dynamic soil moisture. In contrast, the waveguides measure deep soil moisture where temporal variability is expected to be low. The use of waveguides added spatial replication at all depths (Fig. 5, Supplement Fig. S2).

Co-located with every soil moisture waveguide is an access tube to sample soil gas for measurements of the depth distribution of CO_2 and O_2 at a low temporal frequency. At 20 cm below the soil surface and 20 cm above the bottom of the uphill face of the pit, sensors are continuously measuring soil CO_2 (GP001 CO_2 probe, Forerunner Research, Canada) and O_2 (SO-110 Sensor, Apogee Instruments, Utah, USA) at the two midslope catena positions. We selected the midslope catenas for these sensors because they provide the best locations for contrasting north and south aspects. We placed one sensor at the D-20 location to document controls on acid and oxidative weathering near the bedrock interface. The second sensor is near the surface to monitor a zone of high biological CO_2 and O_2 processing. We did not install the sensors at the shallowest depth (10 cm) because we found that high diffusion and advection at shallower depths causes the gas concentrations at 10 cm to reflect atmospheric conditions, providing less information on soil biology (Jin et al., 2014; (Hasenmueller et al., 2015).

Lysimeters (Super Quartz, Prenart Equipment ApS, Denmark) have been emplaced to allow periodic manual sampling of soil pore water for chemical analysis at 20 cm and D-20 cm depths in all catena locations. The rationale for these depths is the same as described above for the automated CO_2 and O_2 sensors (they are co-located in the midslope pits). Overall, these Ground HOG measurements will parameterize the regolith formation models (Table 1) and will be used to test hypotheses linking hydrology, biotic production/consumption of soil gases, and weathering rates.

3.5 Upscaling from the pits to the catena using geophysics

To supplement the Ground HOG observations, we use geophysical and large-footprint methods to interpolate between and extrapolate beyond soil pits. For example, a cosmic-ray neutron detector (CR-1000B, Hydroinnova Inc.) has been emplaced to measure large-scale (~0.5 km radius) average soil moisture every 30 minutes. This COSMOS unit, already used in a variety of ecosystems (Zreda et al., 2013), will measure spatially averaged (3D) soil moisture content within the watershed. Data processing methods have been developed that accounts for various types of moisture storage (e.g. canopy storage, snow, water vapor (Franz et al., 2013; Zweck et al., 2013). The sensor has been installed near the LRVF (Leading Ridge valley floor) pit to provide spatially averaged moisture estimates across the valley.

The COSMOS fills in the gap between small-scale point measurements (Fig. 5) and large-scale satellite remote sensing. The footprint of COSMOS is optimal for hydrometeorological model calibration and validation at small watersheds. One sensor was installed at Shale Hills in 2011 and we are currently testing the COSMOS data with PIHM. We anticipate the results from both catchments will yield insights into the capabilities of cosmic-ray moisture sensing technology in steep terrain and will offer insights into the problem of upscaling soil moisture measurements.

Ground HOG measurements will be further complemented by geophysical mapping along the catenas, including ground penetrating radar (GPR) transects of subsurface structure. Electromagnetic induction (EMI) mapping of soil electric conductivity will similarly be used to measure soil spatial variations between pits. We plan repeated GPR and EMI surveys, in combination with terrain analysis using LIDAR topography, to identify subsurface hydrological features and soil distribution using published procedures (Zhu et al., 2010a, b). We will also field check regolith depths using augers, drills, etc. With repeated geophysical surveys over time (e.g., different seasons and/or before and after storm events), we can also explore temporal changes in heterogeneous soilscapes and subsurface hydrologic dynamics, as demonstrated at Shale Hills (Guo et al., 2014; Zhang et al., 2014).

Such geophysical mapping is necessary to link between and compare with soil-pit point measurements. For example, depth to bedrock along the catenas will be mapped using the geophysical surveys and compared to pit measurements (Fig. 5). These data can be used for upscaling biogeochemical patterns and processes. For example, we expect that soil depth and soil moisture exert the strongest controls on variation in soil gas concentrations, as observed in many places, including Shale Hills (Hasenmueller et al., 2015; Jin et al., 2014). Empirical relationships among these variables developed at Ground HOG points can be coupled with catchment scale soil moisture (from COSMOS) and soil depth (from GPR) data to upscale soil gas characteristics to the whole catchment.

To exemplify the utility of this approach, results from an investigation completed using a ground penetrating radar unit (TerraSIRch Subsurface Interface Radar System-3000) used to map the depth to bedrock in the Garner Run hillslope near the three major monitoring sites (LRVF, LRMS, LRRT) is shown in Fig. 9. Multiple GPR traverses were completed by pulling the antennae along the ground surface. A distance-calibrated survey wheel with encoder was bolted onto these antennae to provide greater control of signal pulse transmission and data collection. The survey wheel occasionally slipped in the challenging terrain, resulting in some errors. Relative elevation data were collected as described below along the traverse line to surface normalize the data.

A traverse line from near Garner Run to the summit was established that ascends Leading Ridge in a nominally west to east direction from ~494 to ~588 masl (Fig. 9). The dominant soils mapped along this traverse line (Supplement Table S2) include Andover, Albrights, Hazleton, and Dekalb. The very deep, poorly drained Andover and moderately well to somewhat poorly drained Albrights soils have been reported in general to have formed in colluvium derived from acid sandstone and shale on upland toe-slope and foot-slope positions. The moderately deep, excessively drained Dekalb and the deep and very deep, well-drained Hazleton soils formed on higher-lying slope positions in residuum weathered from acid sandstone. These soils have moderate potential for penetration with GPR.

The traverse line was cleared of debris but the ground surface remained highly irregular with numerous rock fragments and exposed tree roots that often halted the movement and caused poor coupling of the antennas with the ground. Flags were inserted in the ground at noticeable breaks in the topography along the traverse line. User marks were inserted on the radar records as the antenna passed these survey flags. Later, the elevations of these points were determined using an engineering level and stadia rod. The elevation data were entered into the radar data files and used to "surface normalize" or "terrain correct" the radar records.

In this preliminary investigation, the soil-bedrock interface was not easy to identify. This was attributed to poor antenna coupling with the ground surface in the challenging rocky terrain, noise in the radar records caused by rock fragments in the overlying soil, irregular and fractured bedrock surfaces, and varying degrees of hardness in both rock fragments and the underlying bedrock. These factors weakened the amplitude, consistency and continuity of reflections from the soil-bedrock interface. Nevertheless, preliminary results are described below.

Figure 9 shows two surface-normalized plots of data collected with the 400 MHz antenna as it was pulled from the summit of Leading Ridge to near Garner Run. Distance is measured from the summit area to near Garner Run. Differences in gross reflection patterns can be used to differentiate rock from soil, but the soil-bedrock interface is diffuse. We collected four repeated GPR transects using both 400 and 270 MHz antenna. Compared with the 400 MHz antenna, the lower resolution of the 270 MHz antenna smoothed out irregularities in the bedrock surface and reduced the noise from smaller, less extensive subsurface features, thus improving the interpretability of the soil-bedrock interface. Based on a total of 14,748 soil-depth measurements from ~400 m long GPR images along this traverse line, the interpreted depth to bedrock ranged from 0.58 to 2.42 m and averaged 1.37 m (Table 4, Fig. 9). Each entry in Table 4 indicates the frequency of depth to bedrock data collected with the 400 MHz antenna along a traverse line, grouped into four soil depth classes. The GPR-derived soil depths are reasonable compared to the values we estimated in the soil pits (Fig. 9, Table S2).

3.6 Hydrology: Groundwater measurements

Several methods are needed to characterize physical and chemical interactions of water with regolith and rock in a catchment. First, physical inputs and outputs to a catchment, including precipitation, interception, ET, soil infiltration, and groundwater discharge, must be understood. Often, groundwater flows are omitted from comprehensive hydrology-meteorology-vegetation models such as the Variable Infiltration Capacity (VIC) hydrologic model, or the Noah Land-Surface Model (LSM). However at Shale Hills, we have estimated that rough 50% of incoming water is evapotranspired and 5% reaches the regional groundwater table and returns to the stream as baseflow (Sullivan et al., subm.). At Garner Run, we also expect groundwater to play a significant role in streamflow and geochemical dynamics. For example, some researchers have found that drainage and runoff on sandstone catchments is controlled to great extent by bedrock (Hattanji and Onda, 2004), and specifically by flow through fractures in the upper meters of sandstone beneath the soil (Williams et al., 2010). In this section and the next section we focus on quantifying fluid flow and transport of solutes into surface water and groundwater. We aim to measure the relative magnitudes, timing, and spatial variability of these fluxes. We emphasize methodologies for measuring and characterizing groundwater and streamwater to identify subsurface flow paths of groundwater, and the drivers and controls on water-rock interactions.

In the spirit of "measuring only what is needed," well installation and solid earth sampling by coring will be reduced compared to Shale Hills. At Shale Hills, 28 wells were emplaced and then intermittently monitored (Fig. 2). In Garner run, deep samples (> 8m) have been extracted between Garner Run and Roaring Run from the Harrys Valley 1 well (HV1) drilled within the Garner Run catchment (see Fig. 3). Using a hand-held drill, three shallow wells will be installed and cores will be collected at the catena sites (Fig. 5) and additional monitoring wells will be installed along hillslopes and the valley floor. From these wells, we will also sample solid-phase chemistry and mineralogy.

All core samples will be analyzed for bulk chemistry and mineralogy to characterize the weathering reactions and protolith. Where possible, we will install groundwater monitoring wells in boreholes, with screened intervals spanning the water table. Monitoring at the wells will include hourly water level measurements using autonomous pressure loggers, hourly temperature measurements at two depths below the water table, and monthly water samples collected and analyzed for major ion chemistry. A pumping test will be conducted at the adjacent valley floor wells to measure aquifer storativity and hydraulic conductivity. Relative residence time of groundwater will be assessed from pathway analysis. If resources permit, SF_6 and chlorofluorocarbons (CFCs) will be measured in groundwater samples to assess residence time in the subsurface, as we have done for Shale Hills (Sullivan et al., in press).

Deep core samples and groundwater monitoring will provide a baseline understanding of the geologic and hydrologic system on the new sandstone lithology. Subsequent hypotheses about controls on weathering and hydrologic dynamics, as well as historical flow and solute fluxes, will be constrained by these observations at the catchment boundaries.

3.7 Hydrology: Streamwater flow and chemistry measurements

The Garner Run study reach is approximately 500 m long (Fig. 5) and consists of a rocky, often braided, channel. We have installed a flume at the downstream end of the reach to measure discharge. Stage is continuously monitored using a pressure transducer (Hobo U-20, Onset Computer Corp., Hyannis, MA). Surface water – groundwater (SW-GW) exchange characteristics have been measured using a short-term deployment of a fiber-optic distributed temperature sensor (FO-DTS), and two tracer injection tests. Stream chemistry, including dissolved oxygen (DO), pH, total dissolved solids (TDS), NO_3^-, SO_4^{2-}, Ca, K, Mg, Mn, Na, Fe, and Si, are measured biweekly or monthly in the field with handheld electrodes along the 500 m reach, or by grab sampling and laboratory analysis (inductively coupled plasma-atomic emission spectroscopy, organic carbon analyzer, and ion chromatography).

Stream chemistry is also explored using higher temporal resolution by using a s::can spectrometer and an autosampler during storm events (s::can, GmbH, Vienna, Austria). The s::can is an in-situ instrument capable of measuring such water quality parameters as pH, TDS, dissolved organic carbon (DOC), NO_3^-, DO, NH_4^+, K^+, and F^-. The chemistry and tracer test data will help quantify the flux of fluid and solutes through the subcatchment. The stream chemistry and discharge data will be combined with soil moisture, soil pore water chemistry, and groundwater data to estimate relative contributions to the stream, and underlying processes related to weathering in the near surface and aquifer.

Preliminary results from Garner Run indicate lower concentrations of Ca, Mg, and K compared to the stream discharging from Shale Hills. In addition, as expected, an initial constant injection tracer test at Garner Run revealed significant exchange with the subsurface during low-flow conditions (~0.004 m^3 s^{-1}). Tracer test and temperature results suggest that the stream is losing water along some sections of the 500 m experimental reach and gaining water in others. Both the FO-DTS and stream chemistry data indicate significant input of spring water at ~100 m downstream of the Ground HOG catena (Fig. 5), which is chemically distinct from the upstream surface water and local groundwater sampled from the deep HV1 well. The DTS time series data will be analyzed to identify locations and magnitudes of inputs to the stream, as well as characteristic responses to rainfall events. In combination with the tracer tests, DTS, and chemistry results, we will use well logs and LIDAR topography to explain the lithological and geomorphologic controls on the surface water – ground water (SW-GW) system.

To characterize the major controls and processes governing WEGSS fluxes through the entire Shavers Creek catchment, we are making strategic measurements across the watershed to represent variability: stream discharge, stream chemistry, lithology, and geomorphology. Specifically, stream discharge and chemistry are being monitored along the main stem of Shavers Creek (SCAL, SCBL, and SCO, Fig. 3). At each location we are monitoring stage continuously using pressure transducers (Hobo U-20, Onset Computer Corp., Hyannis, MA), and using periodic discharge measurements to construct stage-discharge rating curves. SW-GW exchange characteristics will be measured as the channel crosses varying lithologies using a series of tracer injection tests. Analyses of stream chemistry from the main stem of Shavers Creek provide a spatial integration of solute behavior from upstream lithologies and land-use types. Eventually, with data from the three subcatchments on shale, sandstone, and calcareous shale (Fig. 3), we will make estimates for non-monitored catchments and test up-scaled estimates of the processes observed in each small watershed.

Preliminary stream chemistry and discharge results indicate significant variability among the three monitoring locations along Shavers Creek (Fig. 10). We see declining concentrations with increasing discharge for Mg and Ca (not shown), and somewhat chemostatic behavior for Si, K, nitrate and others. In this context, chemostatic is used to refer to concentrations of a stream that vary little with discharge (Godsey et al., 2009). Concentrations of Si decrease downstream (a dilution trend), while concentrations increase for Mg and nitrate, presumably due to agricultural amendments in the lower half of Shavers Creek watershed where land use includes farmland. The variety of behaviors will be investigated with respect to land use and lithology changes through the catchment.

4 Model-data feedbacks

Throughout this paper we have described the two-way exchange of field-model insights needed to maximize the efficiency of CZ science. To understand the CZ requires models at all temporal and spatial scales. A measurement in most cases can be recorded as a number: the understanding that derives from that number requires a model. To the extent that models can be used to infer predictions about landscape behavior, field observations and measurements are necessary to provide data for calibration and testing.

The CZ approach of using models to cross from short to long timescales has an important major benefit. Investigations that target long timescales can tease out the effects of feedbacks and thresholds in complex systems that are difficult to discern in short-timescale studies. We thus use quantitative models to explore a vast range in both spatial *and* temporal scales. In this paper we emphasized our approach toward designing a CZO as a tool to understand the CZ as one integral system. We therefore emphasized only one modelling tool, the PIHM family of models. This cascade of models provides a quantitative way for different disciplines to interact about the CZ through the use of a shared suite of models. Our current conceptual understanding and our current computers do not allow us to produce one model that simulates the CZ at all timescales, hence the cascade of models (Table 1).

Such a suite of models is integral not simply for predicting landscape and ecosystem response, but also to building a heuristic understanding of individual CZ processes that may not be apparent from 1^{st}-order observations. Systems-level models are especially needed for proposing and testing hypotheses about feedbacks between climate, biota, and Earth surface and near-surface processes. Although not emphasized here, we have also cited publications throughout this paper that describe the many smaller scales or disciplinary-specific hypotheses and models that have been invoked to learn about individual CZ systems. For example, we point to our earlier observation of Ti and K enrichment in Garner Run soils (Fig. 8). We suggested several processes that could interact to explain data in Fig. 8, including preferential retention of some elements in protolith compared to others, depth variations in protolith composition, accumulation of fines from weathering formations above the current protolith, and dust additions. While first-order mass balance model calculations such as those implicit to Fig. 8 can be used to propose or test such hypotheses, use of Regolith-RT-PIHM (Table 1) or WITCH (Godderis et al., 2006) to model regolith formation are necessary to quantitatively test feasibility of such ideas. Better understanding of regolith formation will in turn inform the permeability distributions needed for hydrologic flow models in the CZO.

5 Conclusions: Measuring and modelling the CZ

Many environmental scientists worldwide are embracing the concept of the critical zone – the surface environment considered over all relevant timescales from the top of the vegetation canopy to the bottom of ground water. CZ science is built upon the hypothesis that an investigation of the entire object – the CZ – will yield insights that more disciplinary-specific investigations cannot. To understand the evolution and dynamics of the CZ, we are developing a suite of simulation models as shown in Table 1 (Duffy et al., 2014). These models are being parameterized based on measurements made at the Susquehanna Shale Hills Critical Zone Observatory (SSHCZO) which is currently expanding from less than 1 km^2 to 165 km^2.

In this paper we described an approach for assessing the CZ in the larger watershed. In effect, our measurement design is a hypothesis in answer to this question: if we want to understand the dynamics and evolution of the entire CZ, what measurements are needed and where should they be made? Our approach emphasizes upscaling from 1D to 2D to 3D using a catena paradigm for ground measurements that is extended with geological, geophysical, LIDAR, stream and meteorological measurements. Our dataset has very little sampling replication within each catchment and we have only designed for one catchment per parent material. This results from the tension between monitoring a core dataset over time (a geological or hydrological approach) versus the replication that is needed for spatial characterization (a soil science or ecological approach). Our spatial design was chosen based on the implicit assumption that implementation of ground- and tower-based measurements (Ground HOG, Tower HOG) in each subcatchment could be upscaled to the entire watershed by interpolation and extrapolation, as well as modelling (Table 1). For example, we are testing the hypothesis that fewer soil pits are needed because we are using a regolith formation model and geological knowledge to site the few pits that we dig. If we find that our limited digging of soil pits is not successful in characterizing the regolith adequately – if our models of regolith formation do not match observations or our models of water flow through regolith do not simulate observations – more pits can be dug or new approaches toward geophysical measurements can be refined. As we build understanding, regolith formation models will be used to extrapolate point measurements of soil thickness and porosity from catena observations to the broader Garner Run subcatchment and to other similar subcatchments in the Shavers creek watershed. In other words, the numerical models in Table 1 will be used to extend beyond the limited observations.

The sampling design described here is also being augmented with brief measurement campaigns outside the subcatchments and outside Shavers creek watershed as warranted. For example, while we will only monitor soil CO_2 continuously at a few catena positions and soil depths, we augment these high frequency data with spatially extensive but temporally limited measurements using manual soil gas samplers. Likewise, we are characterizing vegetation and surface soil properties at 3-5 additional catchments of each parent material type using the transect design initiated at Garner Run (Fig. 5). In general, these outside measurements will be discipline-specific excursions to understand a specific variable. This is a good example of targeted investigations that are not directly related to parameterization of the models in Table 1 for our CZO itself but are rather aimed at improving the process-based understanding that underlies models of CZ evolution. Another example is a set of measurements that are ongoing to investigate regolith formation and hillslope form in other catchments north of Shavers creek where the erosion rates differ. Such targeted investigations can also be compared to output from sensitivity tests where pertinent models are used to explore the effect of the targeted variables (Table 1). Measurements outside the CZO may therefore highlight problems in our limited sampling scheme or modelling approaches that must be improved.

As we improve our understanding of the behavior of components of the CZ, the point is to discover system-wide patterns and processes. Throughout, upscaling will remain a challenge. There is no comprehensive mathematical model of the critical zone, partly because it would be arduous to parameterize and perhaps more importantly because we do not yet understand all the interacting governing processes (Fig. 1). The research in Shavers Creek, and the work done at other CZ observatories (CZOs) around the world, is an attempt to develop a system-wide process model (or ensemble of models) and to identify the essential measurements required for parameterization. Of great interest are robust conceptual models that aid in understanding the CZ, but such conceptual understanding must also be encoded within complex numerical simulations that allow quantitative predictions for testing. Nonetheless, both conceptual and numerical models still include only a portion of the CZ. To really understand WEGSS fluxes quantitatively requires a model that successfully explains the dynamics between topography, groundwater levels, biota, atmospheric conditions, and regolith thickness – at present we are working mostly with conceptual relationships drawn between pairs of factors (Fig. 1).

In our efforts, new observations are tested against and incorporated into the PIHM models to explore the evolution of the CZ over time. In this endeavor, we can also ask, what does success look like? At a CZO, the point of data collection is to understand the CZ both at the scale of interest of the individual investigator and at the full spatial and temporal scale needed to project (earthcast) the CZ. Ultimately, success means that we gain deeper understanding of the system and can predict behavior in other places or with other datasets (e.g. tracers, water isotopes, etc.). Such testing is built in to our nested watershed approach (Fig. 3) and is also implicit to the design of the greater CZO network.

We can also imagine other indicators of success. For example, successful datasets will attract other researchers using other models. This in turn can lead to model-model inter-comparisons. If other models provide better simulations of the catchment, this will drive development of better models. One example of a model – model inter-comparison (RT-Flux-PIHM versus WITCH-Flux-PIHM, Table 1) has already driving new insights.

Another indicator of success is adoption by others of the strategies developed to study the CZ. Such strategies include design of a sampling paradigm for an individual CZO, design of a larger network of CZOs, development of suites of models, or approaches for data assimilation. While the CZO enterprise is still young, publications in the literature already attest to growth in use of the PIHM suite of models in other places (Kumar et al., 2013; Wang et al., 2013; Yu et al., 2015; Jepsen et al., 2015a; Jepsen et al., 2015b; Jepsen et al., in press.) and growth in use of the CZO concept worldwide (Banwart et al., 2012).

Author Contributions

S. L. Brantley, H. Lin, K. J. Davis, A. L. Neal, J. Kaye, and D. Eissenstat designed the study and wrote the sections on soil analysis, GPR, eddy correlation, GIS analysis, biogeochemical analysis, and tree root analysis respectively. R. DiBiase spearheaded the geomorphological treatment, and T. Russo and B. Hoagland led the ground and stream water research. Y. Shi performed the OSSE and worked on coupling models with PIHM. M. Kaye designed and led the vegetation surveys; L. Hill and J. Kaye designed and implemented the soil pit sensor research; A. Dere completed the soil descriptions; K. Brubaker contributed forest and land use descriptions. D. K. Arthur supervised overall data contributions.

Acknowledgements

Field help was provided by B. Forsythe, D. Pederson, and R. Davis. D. Arthur and J. Williams are acknowledged for data organization; A. Orr for photographing the soil pits. C. Duffy is acknowledged for leadership at Shale Hills and for PIHM. S. Macdonald is acknowledged for Garner Run soil geochemistry analysis. L. Li is acknowledged for implementation of the COSMOS at Garner Run, and E. Kirby and P. Bierman for planning the cosmogenic measurements. J. Doolittle from the USDA-NRCS is acknowledged for help with GPR data collection. This work was funded by NSF Critical Zone Observatory program grants to C. Duffy (EAR 07-25019) and SLB (EAR 12-39285, EAR 13-31726). Logistical support and data were provided by the NSF-supported Shale Hills Susquehanna Critical Zone Observatory. This research was conducted in Penn State's Stone Valley Forest, which is supported and managed by the Penn State's Forestland Management Office in the College of Agricultural Sciences. The Penn State Earth and Environmental Systems Institute and College of Agricultural Sciences are acknowledged for staff and funding support.

References

Amundson, R.: Soil formation, in: Treatise in Geochemistry: Surface and Ground Water, Weathering, and Soils, edited by: Drever, J. I., Elsevier Pergamon, Amsterdam, 1–35, 2004.

Anderson, S. P., Dietrich, W. E., and Brimhall, G. H.: Weathering profiles, mass balance analysis, and rates of solute loss: linkages between weathering and erosion in a small, steep catchment, Geol. Soc. Am. Bull., 114, 1143–1158, 2002.

Banwart, S., Menon, M., Bernasconi, S.M., Bloem, J., Blum, W.E.H., de Souza, D.M., Davidsdotir, B., Duffy, C., Lair, G.J., Kram, P., Lamacova, A., Lundin, L., Nikolaidis, N.P., Novak, M., Panagos, P., Ragnarsdottir, K.V., Reynolds, B., Robinson, D., Rousseva, S., de Ruiter, P., van Gaans, P., Weng, L., White, T., Zhang, B.: Soil processes and functions across an international network of Critical Zone Observatories: Introduction to experimental methods and initial results. Comptes Rendus Geoscience 344(11–12):758-772, 2012.

Bao, C., Li, L., Shi, Y., Sullivan, P. L., Duffy, C. J., and Brantley, S. L.: RT-Flux-PIHM: a hydrogeochemical model at the watershed scale, Water Resour. Res., submitted, 2015.

Berg, T. M., Edmunds,W. E., Geyer, A. R., and others: Geologic map of Pennsylvania, 2nd edn., Pennsylvania Geological Survey, 4th ser., Map 1, 3 sheets, scale 1 : 250,000, 1980.

Beven, K. J.: Rainfall–Runoff Modelling: The Primer, Wiley-Blackwell, 2011.

Brantley, S. L. and Lebedeva, M.: Learning to read the chemistry of regolith to understand the Critical Zone, Annu. Rev. Earth Pl. Sc., 39, 387–416, 2011.

Brimhall, G. and Dietrich, W. E.: Constitutive mass balance relations between chemical composition, volume, density, porosity, and strain in metasomatic hydrochemical systems: results on weathering and pedogenisis, Geochim. Cosmochim. Ac., 51, 567–587, 1987.

Carter, B. and Ciolkosz, E.: Slope gradient and aspect effects on soils developed form sandstone in Pennsylvania, Geoderma, 49, 199–213, 1991.

Chadwick, O. A. and Chorover, J.: The chemistry of pedogenic thresholds, Geoderma, 100, 321–353, 2001.

Ciolkosz, E. J., Carter, B., Hoover, M. T., Cronce, R., Waltman, W., and Dobos, R.: Genesis of soils and landscapes in the Ridge and Valley province of central Pennsylvania, Geomorphology, 3, 245–261, 1990.

Cook, K. L., Whipple, K. X., Heimsath, A. M., and Hanks, T. C.: Rapid incision of the Colorado River in Glen Canyon – insights from channel profiles, local incision rates, and modeling of lithologic controls, Earth Surf. Proc. Land., 34, 994–1010, 2009.

Cotter, E.: Tuscarora formation of Pennsylvania : guidebook, Society of Economic Paleontologists and Mineralogists, Eastern Section, 1982 field trip, Lewisburg, Pa., Dept. of Geology and Geophysics, Bucknell University, 1982.

Crutzen, P. J.: Geology of mankind, Nature, 415, 23–23, 2002.

Dietrich, W. E., Bellugi, D. G., Sklar, L. S., Stock, J. D., Heimsath, A. M., and Roering, J. J.: Geomorphic Transport Laws for Predicting Landscape Form and Dynamics, Dynamics, In Prediction, in: Geomorphology, Geophyscial Monograph, edited by: Wilcock, P. R., Iverson, R. M., 135, American Geophysical Union, 103–132, doi:10.1029/135GM09, 2003.

Duffy, C., Shi, Y., Davis, K., Slingerland, R., Li, L., Sullivan, P. L., Godderis, Y., and Brantley, S. L.: Designing a suite of models to explore critical zone function, Procedia Earth and Planetary Science, 10, 7–15, doi:10.1016/j.proeps.2014.08.003, 2014.

Duvall, A., Kirby, E., and Burbank, D.: Tectonic and lithologic controls on bedrock channel profiles and processes in coastal California. J. Geophys. Res.-Earth, 109, F03002 doi:10.1029/2003JF000086, 2004.

Ewing, S. A., Sutter, B., Owen, J., Nishiizumi, K., Sharp, W., Cliff, S. S., Perry, K., Dietrich, W., McKay, C. P., and Amundson, R.: A threshold in soil formation at Earth's arid-hyperarid transition, Geochim. Cosmochim. Ac., 70, 5293–5322, 2006.

Fike, J. Terrestrial and palustrine plant communities of Pennsylvania. Bureau of Forestry, PA. Department of Conservation and Natural Resources, 1999.

Flueckinger, L. A.: Report 176, Geology of a portion of the Allensville Quadrangle, Centre and Huntingdon Counties, Pennsylvania, Bureau of Topographic and Geologic Survey, Commonwealth of Pennsylvania State Planning Board, 1969.

Franz, T. E., Zreda, M., Rosolem, R., and Ferre, T. P. A.: A universal calibration function for determination of soil moisture with cosmic-ray neutrons, Hydrol. Earth Syst. Sci., 17, 453–460, doi:10.5194/hess-17-453-2013, 2013.

Gardner, T. W., Ritter, J. B., Shuman, C. A., Bell, J. C., Sasowsky, K. C., and Pinter, N.: A periglacial stratified slope deposit in the valley and ridge province of central Pennsylvania, USA: Sedimentology, stratigraphy, and geomorphic evolution, Permafrost Periglac., 2, 141–162, 1991.

Godderis, Y. and Brantley, S. L.: Earthcasting the future Critical Zone, Elementa, 1, 1–10, doi:10.12952/journal.elementa.000019, 2014.

Godderis, Y., Francois, L., Probst, A., Schott, J., Moncoulon, D., Labat, D., and Viville, D.: Modelling weathering processes at the catchment scale: the WITCH numerical model, Geochim. Cosmochim. Ac., 70, 1128–1147, 2006.

Godsey, S. E., Kirchner, J. W., and Clow, D. W.: Concentration-discharge relationships reflect chemostatic characteristics of US catchments, Hydrol. Process., 23, 1844–1864, doi:10.1002/hyp.7315, 2009.

Graham, C. and Lin, H. S.: Controls and frequency of preferential flow occurrence: a 175-event analysis, Vadose Zone J., 10, 816–831, 2011.

Graham, C. G. and Lin, H. S.: Spatial–temporal patterns of preferential flow occurrence in the Shale Hills CZO based on real-time soil moisture monitoring, Vadose Zone J., special issue, 10, 816–831, doi:10.2136/vzj2010.0119, 2010.

Guo, L., Chen, J., and Lin, H.: Subsurface lateral flow network on a hillslope revealed by timelapse ground penetrating radar, Water Resour. Res., 50, 21, 2014.

Hasenmueller, E., Jin, L., Stinchcomb, G., Lin, H., Brantley, S. L., and Kaye, J. P.: Topographic controls on the depth distribution of soil CO2 in a small temperate watershed, Appl. Geochem., 63, 58–69, doi:10.1016/j.apgeochem.2015.07.005, 2015.

Hattanji, T. and Onda, Y.: Coupling of runoff processes and sediment transport in mountainous watersheds underlain by different sedimentary rocks, Hydrol. Process., 18, 623–636, doi:10.1002/hyp.1262, 2004.

Hoekstra, P., and A. Delaney: Dielectric properties of soils at UHF and microwave frequencies, J. Geophys. Res., 79 (11), 1699–1708, doi:10.1029/JB079i011p01699, 1974.

Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., and Megown, K.: Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. Photogrammetric Engineering and Remote Sensing, v. 81 (5), 345-354, 2015.

Hoover, M. T. and Ciolkosz, E. J.: Colluvial soil parent material relationships in the Ridge and Valley physiographic province of Pennsylvania, Soil Sci., 145, 163–172, 1988.

Jackson, R.B., Anderson, L.J., Pockman, W.T.: Measuring water availability and uptake in ecosystem studies, Methods in Ecosystem Science, Springer New York, 199-214, doi:10.1007/978-1-4612-1224-9_14, 2000.

Jepsen, S.M., Harmon, T., Meadows, M., and Hunsaker, C.: Hydrogeologic influence on changes in snowmelt runoff with climate warming: Numerical experiments on a mid-elevation catchment in the Sierra Nevada, USA. in American Geophysical Union Annual Meeting (American Geophysical Union, San Francisco), https://agu.confex.com/agu/fm15/meetingapp.cgi/Paper/70885, 2015a.

Jepsen, S.M., Harmon, T.C., and Shi, Y.: Assessment of a watershed model calibrated to the shape of baseflow recession with and without evapotranspiration effects. Water Resources Research, in revision, 2015b.

Jepsen, S.M., Harmon, T.C., Meadows, M.W., and Hunsaker, C.T.: Hydrogeologic influence on changes in snowmelt runoff with climate warming: Numerical experiments on a mid-elevation catchment in the Sierra Nevada, U.S.A. Journal of Hydrology, in press.

Jin, L. and Brantley, S. L.: Soil chemistry and shale weathering on a hillslope influenced by convergent hydrologic flow regime at the Susquehanna/Shale Hills Critical Zone Observatory. Applied Geochemistry 26:S51–S56, doi:10.1016/j.apgeochem.2011.03.027, 2011.

Jin, L., Ravella, R., Ketchum, B., Bierman, P. R., Heaney, P., White, T. S., and Brantley, S. L.: Mineral weathering and elemental transport during hillslope evolution at the Susquehanna/Shale Hills Critical Zone Observatory, Geochim. Cosmochim. Ac., 74, 3669–3691, 2010.

Jin, L., Andrews, D. M., Holmes, G. H., Lin, H., and Brantley, S. L.: Opening the “black box”: water chemistry reveals hydrological controls on weathering in the Susquehanna Shale Hills Critical Zone Observatory, Vadose Zone J., 10, 928–942, doi:10.2136/vzj2010.0133, 2011.

Jin, L., Ogrinc, N., Yesavage, T., Hasenmueller, E. A., Ma, L., Sullivan, P. L., Kaye, J., Duffy, C., and Brantley, S. L.: The CO2 consumption potential during gray shale weathering:

insights from the evolution of carbon isotopes in the Susquehanna Shale Hills critical zone observatory, Geochim. Cosmochim. Ac., 142, 260–280, doi:10.1016/j.gca.2014.07.006, 2014.

Kaplan, J. O., Bigelow, N. H., Prentice, I. C., Harrison, S. P., Bartlein, P., Christensen, T. R., Cramer, W., Matveyeva, N. V., McGuire, A. D., Murray, D. F., Razzhivin, V. Y., Smith, B., Walker, A. D., Anderson, P. M., Andreev, A. A., Brubaker, L. B., Edwards, M. E., and Lozhkin, A. V.: Climate change and Arctic ecosystems: 2. Modeling, paleodata-model comparisons, amid future projections, J. Geophys. Res., 108, 8171, doi:10.1029/2002JD002559, 2003.

Knyazikhin, Y., Glassy, J., Privette, J. L., Tian, Y., Lotsch, A., Zhang, Y., Wang, Y., Morisette, J. T., Votava, P., Myneni, R. B., Nemani, R. R., and Running, S. W.: MODIS Leaf Area Index (LAI) and Fraction of Photosynthetically Active Radiation Absorbed by Vegetation (FPAR) Product (MOD15) Algorithm, Theoretical Basis Document, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA, 1999.

Kumar, M., Marks, D., Dozier, J., Reba, M., and Winstral, A.: Evaluation of distributed hydrologic impacts of temperature-index and energy-based snow models. Advances in Water Resources 56:77-89, 2013.

Lebedeva, M. and Brantley, S. L.: Exploring geochemical controls on weathering and erosion of convex hillslopes: beyond the empirical regolith production function, Earth Surf. Proc. Land., 38, 1793–1807, doi:10.1002/esp.3424, 2013.

Lebedeva, M. I., Fletcher, R. C., Balashov, V. N., and Brantley, S. L.: A reactive diffusion model describing transformation of bedrock to saprolite, Chem. Geol., 244, 624–645, 2007.

Lebedeva, M. I., Fletcher, R. C., and Brantley, S. L.: A mathematical model for steady-state regolith production at constant erosion rate, Earth Surf. Proc. Land., 35, 508–524, 2010.

Lin, H.S., Kogelmann, W., Walker, C., and Bruns, M.A.: Soil moisture patterns in a forested catchment: A hydropedological perspective. Geoderma 131:345-368, doi:10.1016/j.geoderma.2005.03.013, 2006.

Lynch, J. A.: Effects of Antecedent Soil Moisture on Storm Hydrographs, Doctor of Philosophy, Forestry, The Pennsylvania State University, University Park, PA, 1976.

Ma, L., Jin, L., and Brantley, S. L.: How mineralogy and slope aspect affect REE release and fractionation during shale weathering in the Susquehanna/Shale Hills Critical Zone Observatory, Chem. Geol., 290, 31–49, 2011.

Ma, L., Chabaux, F., West, N., Kirby, E., Jin, L., and Brantley, S. L.: Regolith production and transport in the Susquehanna Shale Hills Critical Zone Observatory, Part 1: Insights from U-series isotopes, J. Geophys. Res.-Earth, 118, 722–740, doi:10.1002/jgrf.20037, 2013.

Meinzer, F. C., Woodruff, D. R., Eissenstat, D. M., Lin, H. S., Adams, T. S., and McCulloh, K. A.: Above- and belowground controls on water use by trees of different wood types in an eastern US deciduous forest, Tree Physiol., 33, 345–356, 2013.

Merritts, D.J., Walter, R.C., Blair, A., Demitroff, M., Potter Jr., N., Alter, S., Markey, E., Guillorn, S., Gigliotti, S., and Studnicky, C.: LIDAR, orthoimagery, and field analysis of periglacial landforms and their cold climate signature, unglaciated Pennsylvania and Maryland: Abstract 329-1 presented at 2015 Annual Meeting, GSA, Baltimore, MD, 1–4 November, 47 (7), 831, 2015.

Miller, S. R., Sak, P. B., Kirby, E., and Bierman, P. R.: Neogene rejuvenation of central Appalachian topography: evidence for differential rock uplift from stream profiles and erosion rates, Earth Planet. Sc. Lett., 369–370, 13, 2013.

Minasny, B., McBratney, A. B., and Salvador-Blanes, S.: Quantitative models for pedogenesis – a review, Geoderma, 144, 140–157, 2008.

Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y.,Wang, Y., Song, X., Zhang, Y., Smith, G. R., Lotsch, A., Friedl, M., Morisette, J. T., Votava, P., Nemani, R. R., and Running, S. W.: Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data, Remote Sens. Environ., 83, 214–231, 2002.

Naithani, K. J., Baldwin, D., Gaines, K., Lin, H. S., and Eissenstat, D. M.: Spatial distribution of tree species governs the spatio-temporal interaction of leaf area index and soil moisture across a landscape, PloS ONE, 8, 12, doi:10.1371/journal.pone.0058704, 2013.

Niu, X., Williams, J. Z., Miller, D., Lehnert, K. A., Bills, B., and Brantley, S. L.: An ontology driven relational geochemical database for the Earth's Critical Zone: CZchemDB, J. Environ. Inf., 23, 10–23, doi:10.3808/jei.201400266, 2014.

Portenga, E. W., Bierman, P. R., Rizzo, D. M., and Rood, D. H.: Low rates of bedrock outcrop erosion in the central Appalachian Mountains inferred from in situ Be-10, Geol. Soc. Am. Bull., 125, 201–215, 2013.

Qu, Y. and Duffy, C. J.: A semi-discrete finite volume formulation for multiprocess watershed simulation, Water Resour. Res., 43, 1–18, 2007.

Shi, Y., Davis, K. J., Duffy, C. J., and Yu, X.: Development of a coupled land surface hydrologic model and evaluation at a critical zone observatory, J. Hydrometeorol., 14, 1401–1420, 2013.

Shi, Y., K. J. Davis, F. Zhang, C. J. Duffy, and Yu, X.: Parameter estimation of a physically-based land surface hydrologic model using the ensemble Kalman Filter: A synthetic experiment. Water Resources Research, 50, 706-724, doi:10.1002/2013WR014070, 2014.

Shi, Y., Baldwin, B. C., Davis, K. J., Yu, X., Duffy, C. J., and Lin, H.: Simulating high resolution soil moisture patterns in the Shale Hills catchment using a land surface hydrologic model, Hydrol. Process., doi:10.1002/hyp.10593, 2015a.

Shi, Y., Davis, K. J., Zhang, F., Duffy, C., J., and Yu, X.: Parameter estimation of a physically-based land surface hydrologic model using an ensemble Kalman filter: a multivariate real data experiment, Adv. Water Resour., 83, 421–427, doi:10.1016/j.advwatres.2015.06.009, 2015b.

Sullivan, P. L., Hynek, S., Gu, X., Singha, K., White, T. S.,West, N., Kim, H., Clarke, B., Kirby, E., Duffy, C., and Brantley, S. L.: Oxidative dissolution under the channel leads geomorphological evolution at the Shale Hills catchment, Am. J. Sci., submitted, 2015.

Tarboton, D. G.: A new method for the determination of flow directions and upslope areas in grid digital elevation models, Water Resour. Res., 33, 309–319, doi:10.1029/96WR03137, 1997.

Thomas, E., Lin, H., Duffy, C., Sullivan, P., Holmes, G. H., Brantley, S. L., and Jin, L.: Spatiotemporal patterns of water stable isotope compositions at the Shale Hills Critical Zone: linkages to subsurface hydrologic processes, Vadose Zone J., 12, doi:10.2136/vzj2013.01.0029, 2013.

Topp, G.C., Davis, J.L., and Annan, A.P.: Electromagnetic determination of soil water content: Measurement in coaxial transmission lines, Water Resour. Res., 16, 574-582, 1980.

Topp, G. & Ferre, P.-A.: Water content, in Method of soil analysis, Part 4, Physical methods, SSSA Book, vol.5 ed. J.H. Dane and G.C. Topp, Soil Sci. Soc. of Am., Madison, Wis., 417-446, 2002.

Vitousek, P. M., Aber, J. D., Horwath, R. W., Lubchenco, J., and Melillo, J. M.: Human alteration of the global nitrogen cycle, Science, 277, 494–499, 1997a.

Vitousek, P. M., Mooney, H. A., Lubchenco, J., and Melillo, J. M.: Human domination of Earth's ecosystems, Science, 277, 494–499, 1997b.

Wang, R., Kumar, M., and Marks, D.: Anomalous trend in soil evaporation in a semi-arid, snow-dominated watershed. Advances in Water Resources 57:32-40, 2013.

Warnant, P., Francois, L., Strivay, D., and Gerard, J.-C.: CARAIB: a global model of terrestrial biological productivity, Global Biogeochem. Cy., 8, 255–270, 1994.

West, N., Kirby, E., Bierman, P. R., Slingerland, R., Ma, L., Rood, D., and Brantley, S. L.: Regolith production and transport at the Susquehanna Shale Hills Critical Zone Observatory: Part 2 – Insights from meteoric 10Be. J. Geophys. Res.-Earth, 118, 1–20, doi:10.1002/jgrf.20121, 2013.

West, N., Kirby, E., Bierman, P. R., and Clarke, B. A.: Aspect-dependent variations in regolith creep revealed by meteoric 10Be, Geology, 42, 507–510, doi:10.1130/G35357.1, 2014.

Whipple, K. X., Dibiase, R. A., and Crosby, B. T.: Bedrock rivers, in: Treatise in Geomorphology, Fluvial Geomorphology, edited by: Wohl, E., Academic Press, San Diego, CA, 550–573, 2013.

White T., Brantley S., Banwart S., Chorover J., Dietrich W., Derry L., Lohse K., Anderson S., Aufdendkampe A., Bales R., Kumar P., Richter D., McDowell B.: The Role of Critical Zone Observatories in Critical Zone Science, in Developments in Earth Surface Processes 19, Principles and Dynamics of the Critical Zone (eds. Giardino, J., and Houser, C.), Elsevier, 15–78, 2015.

Wilkinson, B. H. and McElroy, B. J.: The impact of humans on continental erosion and sedimentation, Geol. Soc. Am. Bull., 119, 140–156, 2007.

Willett, S. D., McCoy, S. W., Perron, J. T., Goren, L., and Chen, C. Y.: Dynamic reorganization of river basins, Science, 343, 1–9, doi:10.1126/science.1248765, 2014.

Williams, J., Reynolds, R., Franzi, D., Romanowics, E., and Paillet, F.: Hydrogeology of the Potsdam sandstone in northern New York, Can. Water Resour. J., 35, 399–415, 2010.

Williard, K., Dewalle, D., and Edwards, P.: Influence of bedrock geology and tree species composition on stream nitrate concentrations in mid-Appalachian forested watersheds, Water Air Soil Poll., 160, 55–76, 2005.

Yu, X., Lamačová, A., Duffy, C., Krám, P., Hruška, J., White, T., and Bhatt, G.: Modelling long-term water yield effects of forest management in a Norway spruce forest. Hydrol. Sci. J.-J. Sci. Hydrol. 60(2):174-191, doi: 10.1080/02626667.2014.897406, 2015.

Zhang, J., Lin, H. S., and Doolittle, J.: Soil layering and preferential flow impacts on seasonal changes of GPR signals in two contrasting soils, Geoderma, 213, 560–569, doi:10.1016/j.geoderma.2013.08.035, 2014.

Zhu, Q., Lin, H., and Doolittle, J.: Repeated electromagnetic induction surveys for determining subsurface hydrologic dynamics in an agricultural landscape, Soil Sci. Soc. Am. J., 74, 12, 2010a.

Zhu, Q., Lin, H., and Doolittle, J.: Repeated electromagnetic induction surveys for improved soil mapping in an agricultural landscape, Soil Sci. Soc. Am. J., 74, 1763–1774, doi:10.2136/sssaj2010.0056, 2010b.

Zreda, M., Shuttleworth, W. J., Zeng, X., Zweck, C., Desilets, D., Franz, T., and Rosolem, R.: Corrigendum to “COSMOS: the COsmic-ray Soil Moisture Observing System” published in Hydrol. Earth Syst. Sci., 16, 4079–4099, 2012, Hydrol. Earth Syst. Sci., 17, 1065–1066, doi:10.5194/hess-17-1065-2013, 2013.

Zweck, C., Zreda, M., and Desilets, D.: Snow shielding factors for cosmogenic nuclide dating inferred from Monte Carlo neutron transport simulations, Earth Planet. Sc. Lett., 379, 64–71, 2013.

Table 1. Designing a suite of CZ models

	Modeling purpose	Model	Timescale of interest
Numerical models in use at SSHCZO	Topography (land scape evolution)	LE-PIHM	Days—millions of years
	Regolith composition and structure	Regolith-RT-PIHM, WITCH[a]	Hours—millions of years
	Distribution of biota	BIOME4[b], CARAIB[c], ED2	Days—centuries
	C and N pools and fluxes	Flux-PIHM-BGC	Days—decades
	Sediment fluxes	PIHM-SED	Hours—decades
	Solute chemistry and fluxes	RT-Flux-PIHM[d], WITCH	Hours—decades
	Soil CO_2 concentration and fluxes	CARAIB	Hours—decades
	Energy and hydrologic fluxes	PIHM[d], Flux-PIHM[f]	Hours—decades
Geological factor	Uplift rate, bedrock composition, bedrock physical properties, pre-existing geological factors such as glaciation		
External driver	Energy inputs, chemistry of wet and dry deposition, atmospheric composition, climate conditions, anthropogenic activities		

a: Godderis et al. (2006)

b: Kaplan et al. (2003)

c: Warnant et al. (1994)

d: Bao et al., 2015 (subm.)

e: Qu and Duffy (2007)

f: Shi et al. (2013)

Table 2. Measurements and instrumentation for Tower HOG system

Measurement	Manufacturer	Model	Collection frequency
[CO_2], [H_2O]	Li-cor	LI-7500A CO_2/H_2O analyzer	10Hz‡
3-D wind velocity, virtual temperature	Campbell Scientific	CSAT3 sonic anemometer	10Hz‡
Precipitation	OTT Hydromet	Pluvio^2 Weighing Rain Gauge	Every 10 min
T_{air}	Vaisala	HMP60 humidity and temperature probe	Every 30 min
Relative Humidity	Vaisala	HMP60 humidity and temperature probe	Every 30 min
Longwave Radiation*	Kipp & Zonen	CGR3 pyrgeometer	Every 30 min
Shortwave Radiation*	Kipp & Zonen	CMP3 pyranometer	Every 30 min
Snow depth†	Campbell Scientific	SR50A sonic ranging sensor	Every 30 min
Digital Imagery	Campbell Scientific	CC5MPX digital camera	Every 24 hr

* All four components of radiation (upwelling and downwelling (longwave and shortwave)) will only be measured at Shale Hills Tower HOG due to the location of the Garner Run Tower HOG. To model Garner Run we will use the Shale Hills data.

† Originally designed as part of tower system but will be deployed at LRVF Ground HOG location because the Garner Run tower will be located outside of the catchment.

‡The turbulent fluxes (sensible and latent heat) and the momentum flux are computed at 30 minute intervals via eddy covariance using these data collected at 10 Hz.

Table 3. Vegetation sampling in the Garner Run subcatchment

Site[1]	Sample area (ha)	Tree basal area (m^2 ha^{-1})	Tree density (trees ha^{-1})	Tree species richness (# species)	Dominant tree species (% basal area)	Forest floor cover (% rock)	Mean rock diameter (cm)	Organic horizon C (g m^{-2})
LRRT	1	25.3	607	9	*Quercus prinus* (44%) *Acer rubrum* (19%) *Pinus strobus* (19%) *Nyssa sylvatica* (12%)	16	29	1775
LRMS	1.4	25.1	610	12	*Betula lenta* (37%) *Quercus prinus* (21%) *Nyssa sylvatica* (15%) *Quercus rubra* (10%)	28	45	2208
LRVF	0.7	24.6	371	14	*Quercus rubra* (26%) *Betula lenta* (23%) *Quercus prinus* (20%) *Acer rubrum* (14%)	36	43	1122
TMMS	1	18.5	519	9	*Acer rubrum* (32%) *Betula lenta* (29%) *Nyssa sylvatica* (25%)	34	60	n/a

[1]LRRT: Leading Ridge ridge top, LRMS: Leading Ridge midslope, LRVF: Leading Ridge valley floor, TMMS: Tussey Mountain midslope. Measurements were made in linear belt transects 700 to 1400 m long and 10 m wide centered at each soil pit position (Fig. 5).

Table 4. Frequency distribution of bedrock depth measurements along GPR transect (Fig. 9)

Depth to bedrock	*Upper section*	*Lower section*
Shallow (< 0.5 m)	0.00	0.00
Moderately Deep (0.5 to 1 m)	0.26	0.04
Deep (1 to 1.5 m)	0.51	0.48
Very Deep (> 1.5 m)	0.24	0.48

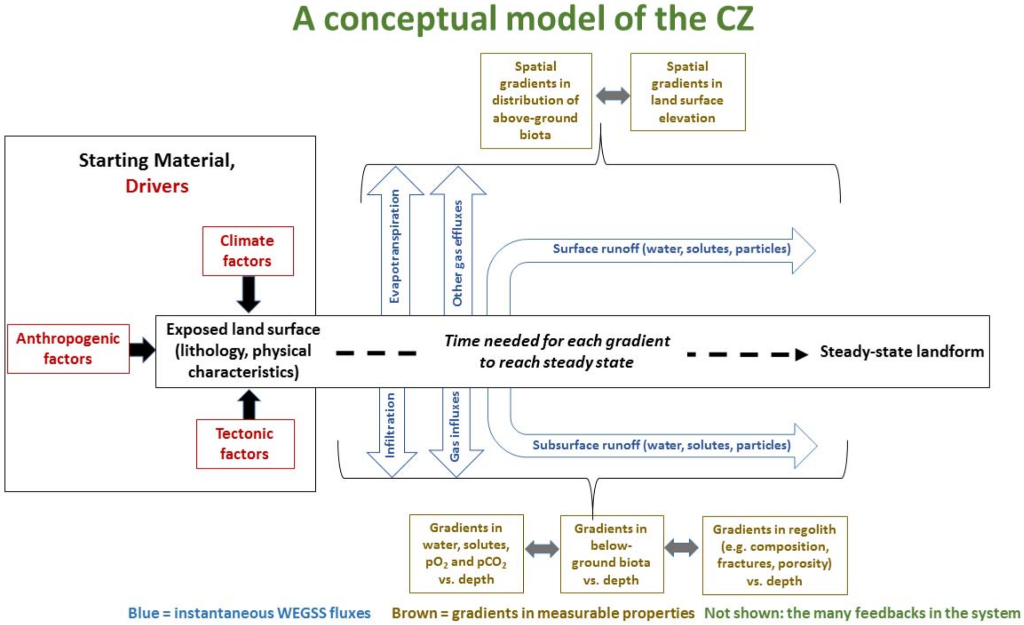

Figure 1. Critical zone science investigates the architecture, character, and dynamics of the earth surface from vegetation canopy to deep ground water at all time scales. As rock of a certain lithology and structural character is exposed at earth's surface due to uplift or erosion, climate-driven inputs transform it to regolith. This transformation, shown in the black box, is catalyzed by biota (a feedback which is not shown explicitly). Gradients of properties describing the CZ are shown in brown boxes. These gradients can become time-independent (steady state) due to the many feedbacks which are not shown. Boxes are placed from left to right to note the increasing duration of exposure time needed to achieve such steady states. For example, depth profiles of regolith composition can become constant when rate of erosion = rate of weathering advance in the presence of feedbacks related to pore water chemistry, soil gas composition, and grain size. The figure emphasizes that gradients to the left can achieve steady state quickly compared to properties to the right. Therefore properties to the left are often studied as if properties in boxes to the right are constant boundary conditions. However, over the longest timescales, all properties vary and can affect one another. The complexity of feedbacks (which are not shown for simplicity) can also create thresholds in system behavior. Red boxes indicate drivers and blue arrows are WEGSS fluxes (up arrows for above-ground and down arrows for below-ground).

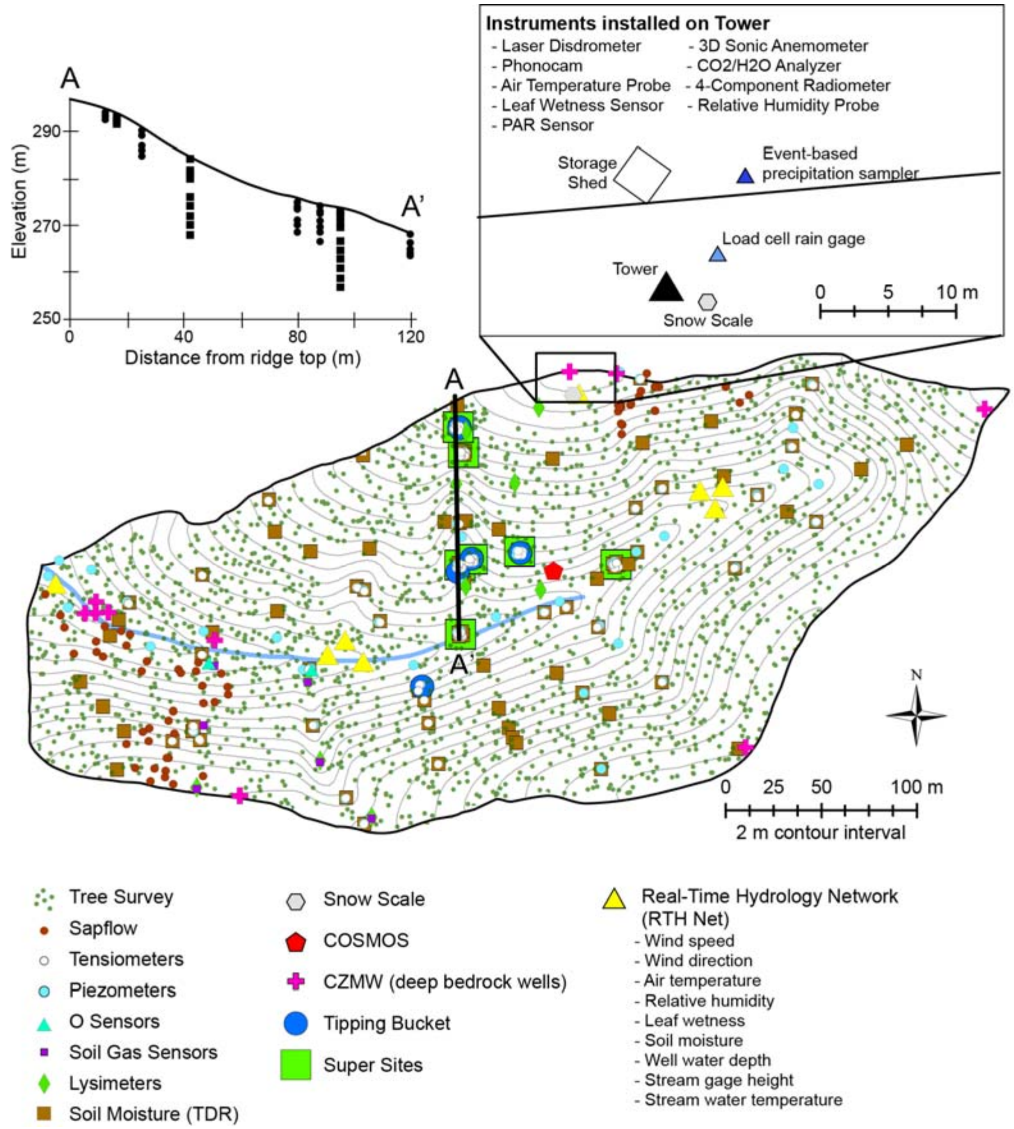

Figure 2. Mapped summary of the “everything, everywhere” sampling strategy at the Shale Hills subcatchment. Insets show soil moisture sensors (circles) and lysimeters (squares) along the transect shown on the map. Sensor and lysimeter depths are exaggerated five times compared to the land surface elevation. Second inset shows instrumentation deployed at the meteorological station on the northern ridge. Small green dots on the map are the trees that were surveyed and numbered: the subcatchment contains a dry oak-mixed hardwood community type (Fike, 1999) with an extremely diverse mix of hardwood and softwood species, including white oak (*Quercus alba*), sugar maple (*Acer saccharum*), pignut hickory (*Carya glabra*), eastern hemlock (*Tsuga canadensis*, and chestnut oak (*Q. montana*). The

sparse understory consists of American hop-hornbeam (*Ostrya virginiana*) and serviceberry (*Amelanchier sp*.). As we upscale the CZO to all of Shavers creek, many measurements will be eliminated in the Shale Hills subcatchment as we emphasize only a Ground HOG and Tower HOG deployment as described for the Garner Run subcatchment.

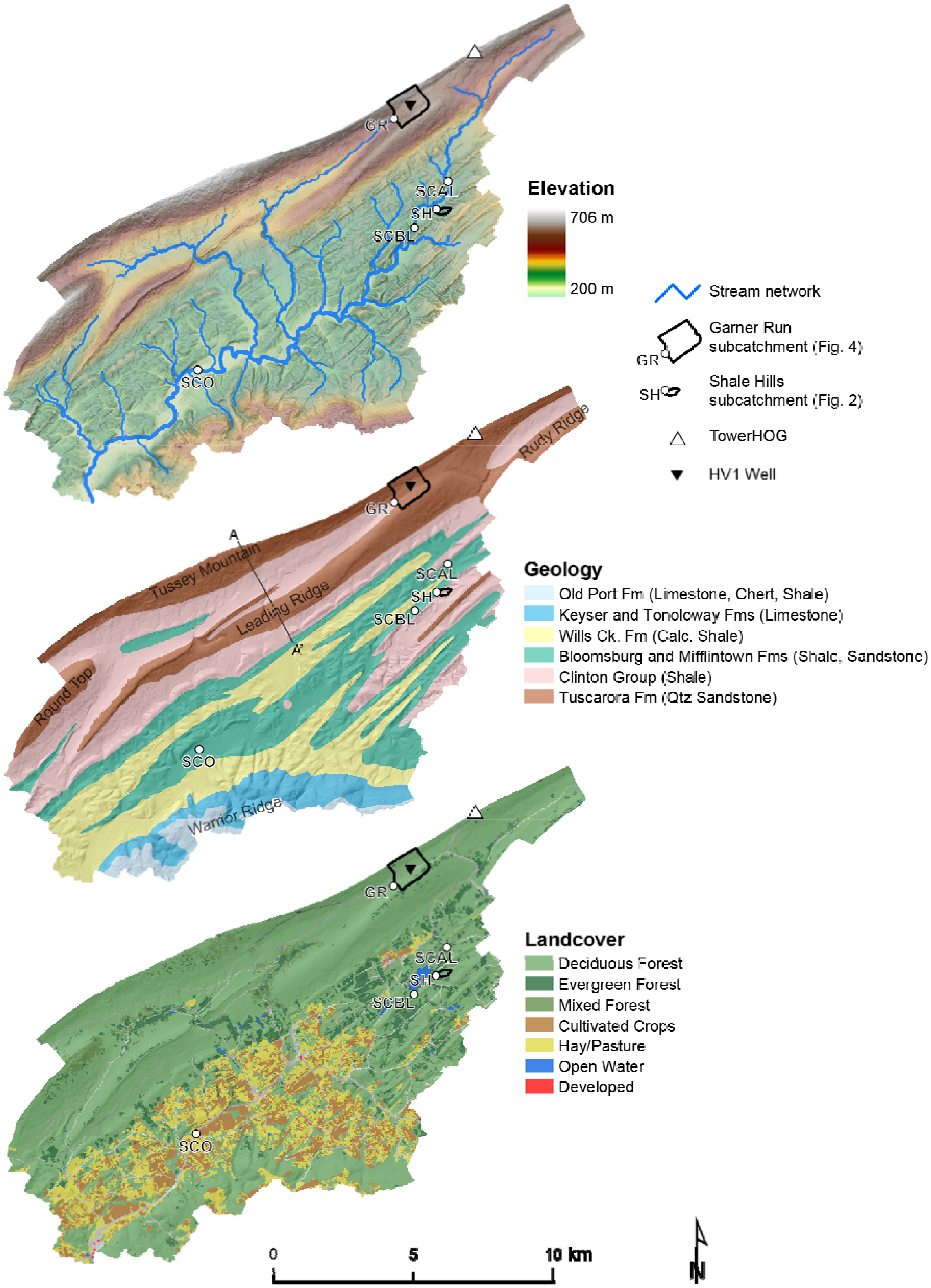
Elevation
706 m
200 m
Stream network
Garner Run subcatchment (Fig. 4)
Shale Hills subcatchment (Fig. 2)
TowerHOG
HV1 Well
GR
SH
SCAL
SCBL
SCO
Geology
Old Port Fm (Limestone, Chert, Shale)
Keyser and Tonoloway Fms (Limestone)
Wills Ck. Fm (Calc. Shale)
Bloomsburg and Mifflintown Fms (Shale, Sandstone)
Clinton Group (Shale)
Tuscarora Fm (Qtz Sandstone)
Tussey Mountain
Leading Ridge
Round Top
Rudy Ridge
Warrior Ridge
A
A'
Landcover
Deciduous Forest
Evergreen Forest
Mixed Forest
Cultivated Crops
Hay/Pasture
Open Water
Developed
0
5
10 km
N

Figure 3. Map of Shavers Creek Watershed, highlighting (a) topography derived from airborne LIDAR, (b) geology (Berg et al., 1980), and (c) land use (Homer et al., 2011). In moving from measure-everything-everywhere (our paradigm in the 8 ha Shale Hills catchment (SH) to measure-only-what-is-needed in the Shavers Creek Watershed (164 km^2)), we chose to investigate two new first-order sub-catchments: a forested sandstone site (along Garner Run, marked GR) and an agricultural calcareous shale site (to be determined). In addition, three sites on Shavers Creek have been chosen as stream discharge and chemistry monitoring sites (marked SCAL – Shavers Creek Above Lake, SCBL – Shavers Creek Below Lake, and SCO – Shavers Creek Outlet).

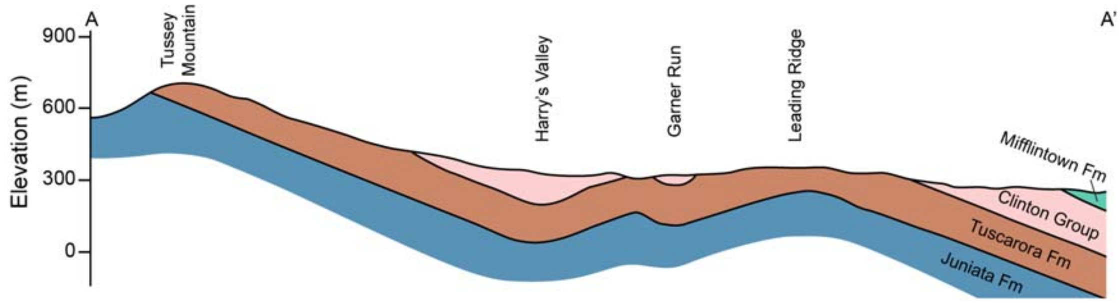

Figure 4. Geologic cross-section across Garner Run subcatchment reproduced from Flueckinger (1969). Map units include Mifflintown (Middle Silurian), Clinton group (including Rose Hill formation) and Tuscarora (Lower Silurian), and the Juniata (Upper Ordovician). Cross section position is downstream from the targeted subcatchment (see Fig. 3). The published map (Flueckinger, 1969) of the actual sub-catchment (not shown) shows no remaining Rose Hill formation outcrop in Garner Run subcatchment, i.e., the Tuscarora no longer outcrops upstream of this cross-section and Garner Run lies in the axis of Harry's valley. This cross-section from down valley of Garner Run subcatchment emphasizes that Rose Hill shale was originally present above the Tuscarora.

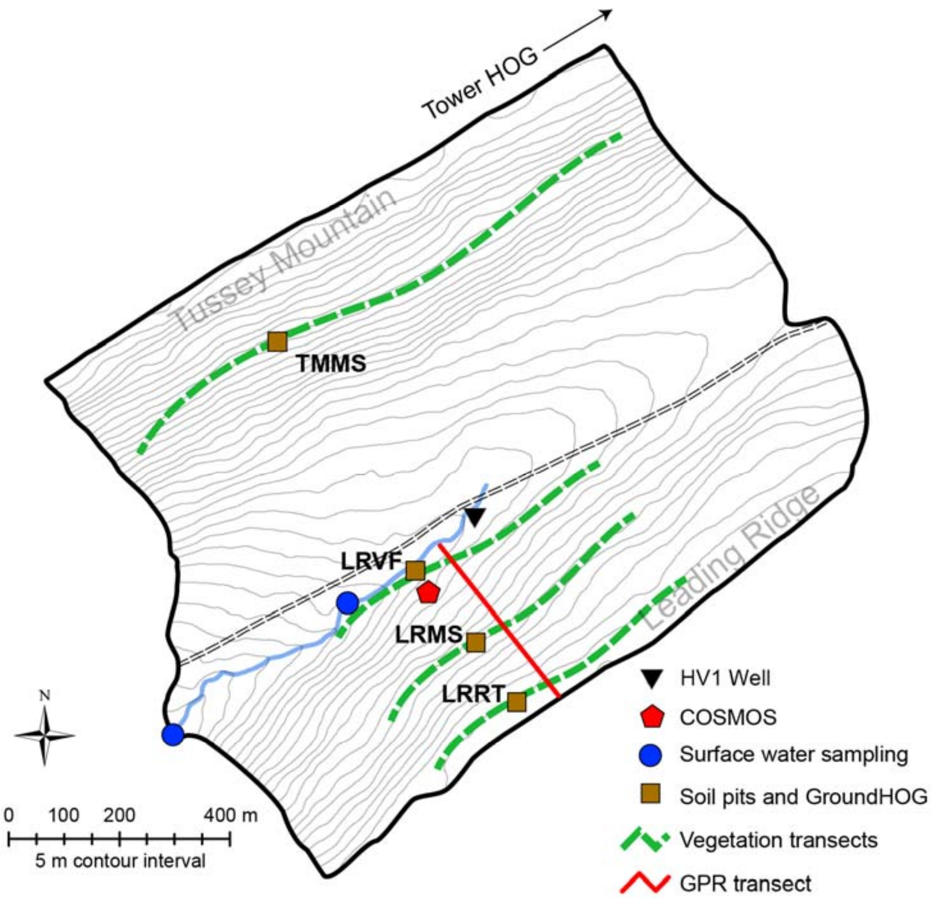

Figure 5. Map showing Garner Run subcatchment (blue line is the stream). Black dashed lines delineate Harry's Valley Road. The Harrys Valley well (HV1) is shown along with the location of the COSMOS unit and the outlet weir (blue dot to the southwest). The blue dot to the northeast indicates the approximate range of surface water sampling that is ongoing. Soil pits have been emplaced as shown, along with the Ground HOG deployment. Location of vegetation and GPR transects reported in this paper are also shown.

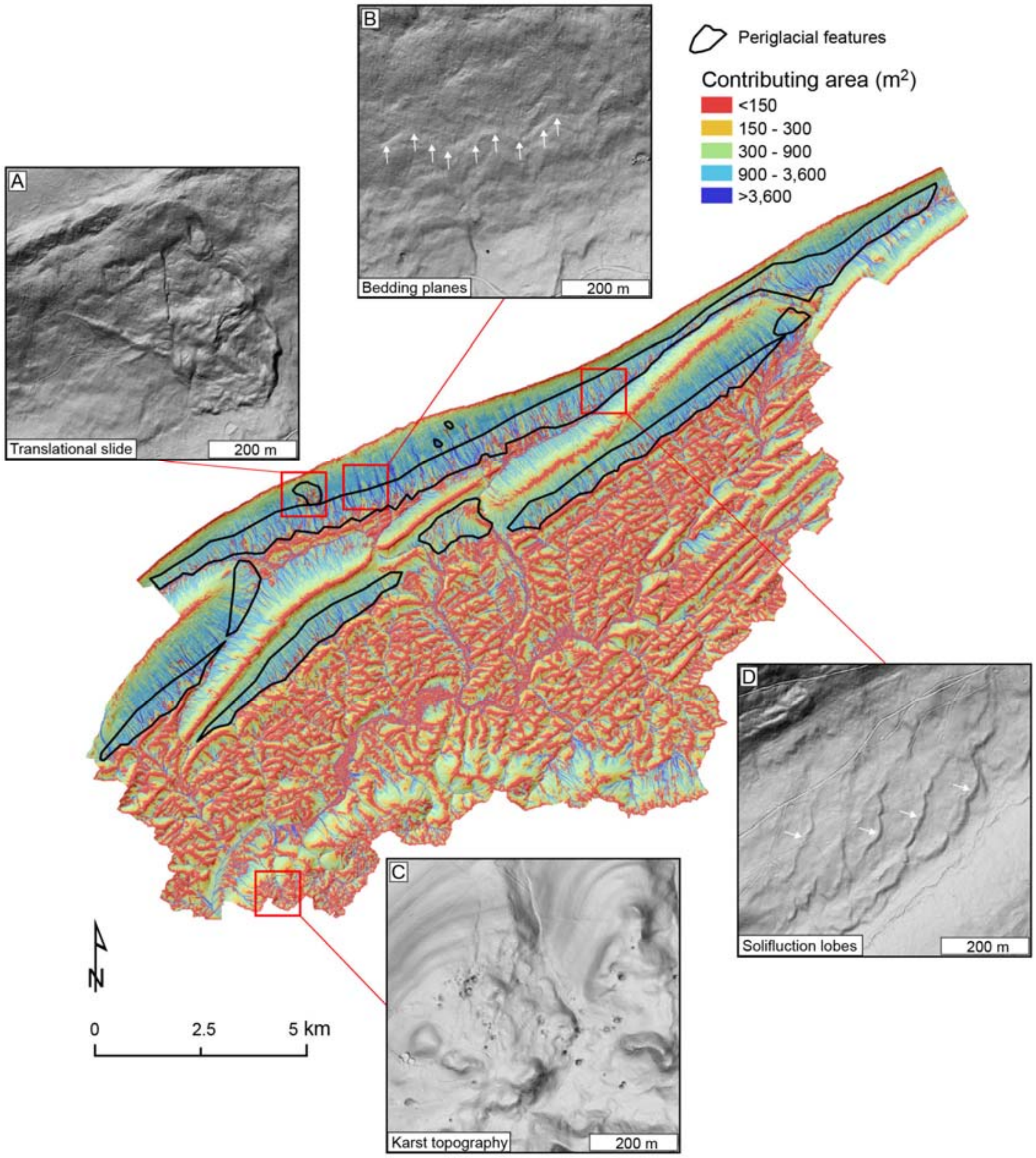

Figure 6. Map of bedrock and periglacial process controls on topography in Shavers Creek watershed. The contributing area was determined using the D-Infinity flow routing algorithm (Tarboton, 1997). The map highlights spatial variations in drainage density that correspond to sandstone (low drainage density and long hillslopes), shale (high drainage density and short hillslopes), and carbonate (intermediate drainage density and hillslope length) bedrock (see Figure 4). Black outlines correspond to periglacial features expressed in the 1 m LIDAR

topography, such as landslides (inset A) and solifluction lobes (inset D). Sandstone bedding planes (inset B) and limestone karst topography (C) are also prominent.

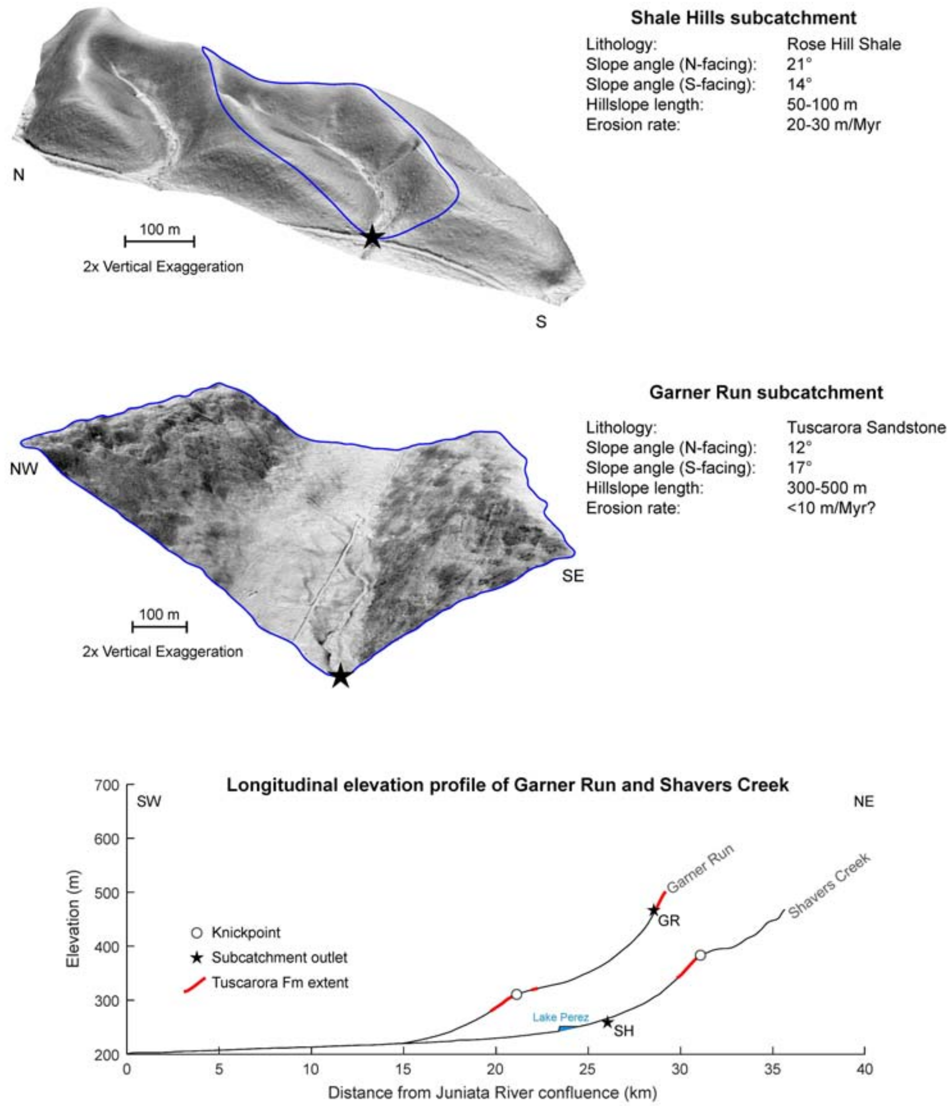

Figure 7. Perspective slopeshade maps (darker shades = steeper slopes) of Shale Hills (top panel) and Garner Run (middle panel) subcatchments, emphasizing differences in slope

asymmetry and hillslope length. Soil production and erosion rates for Shale Hills subcatchment were measured based on U-series isotopes and meteoric ^{10}Be concentrations in regolith respectively (Ma et al., 2013; West et al., 2013; 2014). Erosion rate for Garner Run subcatchment is estimated based on detrital 10Be concentrations from nearby sandstone catchments with similar relief (Miller et al., 2013). Bottom panel shows stream longitudinal profiles, highlighting the lithologic control on knickpoint locations. Note the location of the Shale Hills subcatchment (SH) downstream of the knickpoint on Shavers Creek and the location of the Garner Run subcatchment (GR) upstream of the knickpoint on Garner Run.

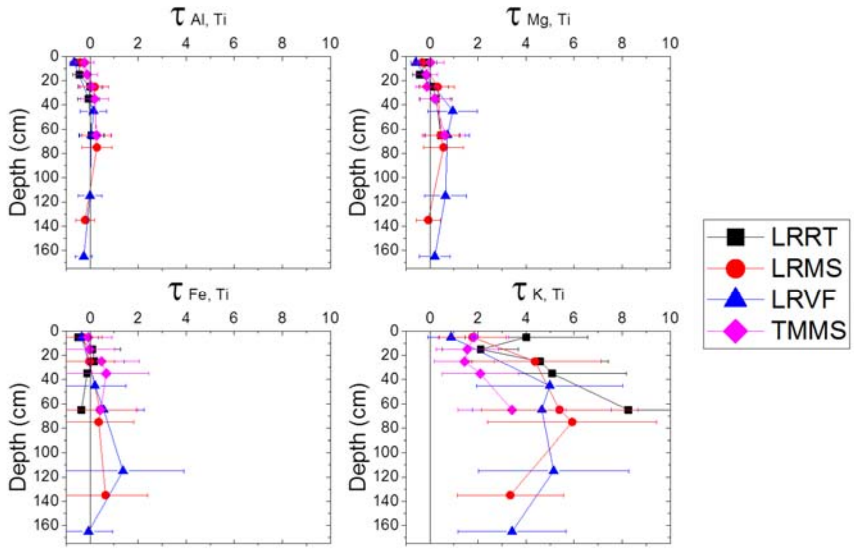

Figure 8. Plots of normalized concentration (τ) versus depth for soils analyzed from the four Garner Run sub-catchment soil pits (LRVF, LRMS, LRRT, TMMS). Y axis indicates the depth below the organic – mineral horizon interface. τ is the mass transfer coefficient determined using parent composition estimated as the average of 5 rocks (Supplement Table S3) from the bottom of several of the pits based on the assumption that Ti derives from protolith and is immobile. If parent is correctly estimated, τ = -1 when an element is 100% depleted, τ = 0 when no loss or gain has occurred, and is τ > 0 when the element has been added to the profile compared to Ti in the parent material.

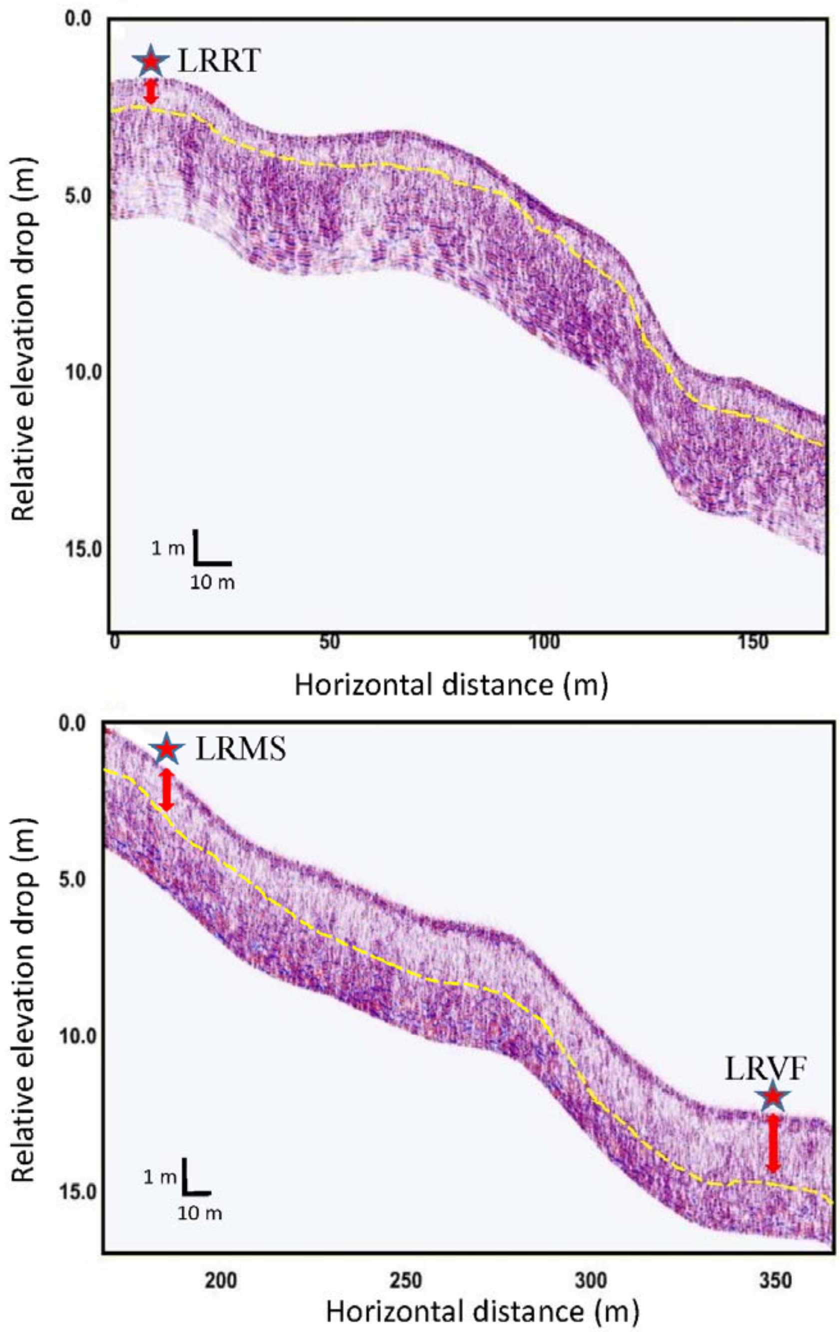

Figure 9. Ground Penetrating Radar (GPR) transect of the Leading Ridge catena, showing inferred location of bedrock-soil interface (yellow dashed curve). The three soil pits (LRRT,

LRMS, LRVF) are indicated by stars, with their observed depth to bedrock indicated by red arrow bar. LRRT and LRMS were dug by hand until refusal and LRVF was dug by hand and deepened with a jack hammer. GPR data are exaggerated by 10x in vertical dimension as compared to surface topography. Summary bedrock depths are tabulated in Table 4.

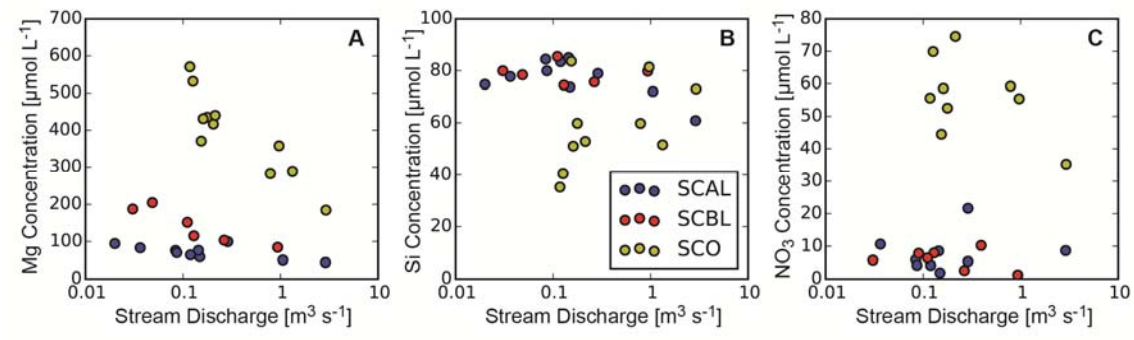

Figure 10. (A) Mg, (B) Si, and (C) nitrate dissolved concentrations (filtered at 0.45 µm) and stream discharge measured at three locations on Shavers Creek: Above Lake (SCAL, blue), Below Lake (SCBL, red), and the Outlet (SCO, yellow) as shown on Fig. 3.