Earth’s Critical Zone and hydropedology: concepts, characteristics, and advances

H. Lin

The Pennsylvania State University, Department of Crop and Soil Sciences, 116 Agricultural Sciences and Industry Building, University Park, PA 16802, USA

Received: 31 March 2009 – Published in Hydrol. Earth Syst. Sci. Discuss.: 27 April 2009
Revised: 30 November 2009 – Accepted: 8 December 2009 – Published: 8 January 2010

Abstract. The Critical Zone (CZ) is a holistic framework for integrated studies of water with soil, rock, air, and biotic resources in the near-surface terrestrial environment. This most heterogeneous and complex region of the Earth ranges from the vegetation top to the aquifer bottom, with a highly variable thickness globally and a yet-to-be clearly defined lower boundary of active water cycle. Interfaces among different compartments in the CZ are critical, which provide fertile ground for interdisciplinary research. The reconciliation of coupled geological and biological cycles (vastly different in space and time scales) is essential to understanding the complexity and evolution of the CZ. Irreversible evolution, coupled cycling, interactive layers, and hierarchical heterogeneity are the characteristics of the CZ, suggesting that forcing, coupling, interfacing, and scaling are grand challenges for advancing CZ science. Hydropedology – the science of the behaviour and distribution of soil-water interactions in contact with mineral and biological materials in the CZ – is an important contributor to CZ study. The pedosphere is the foundation of the CZ, which represents a geomembrane across which water and solutes, as well as energy, gases, solids, and organisms are actively exchanged with the atmosphere, biosphere, hydrosphere, and lithosphere, thereby creating a life-sustaining environment. Hydropedology emphasizes in situ soils in the landscape setting, where distinct pedogenic features and soil-landscape relationships are essential to understanding interactive pedologic and hydrologic processes. Both CZ science and hydropedology embrace an evolutionary and holistic worldview, which offers stimulating opportunities through steps such as integrated systems approach, evolutionary mapping-monitoring-modeling framework, and fostering a global alliance. Our capability to predict the behaviour and evolution of the CZ in response to changing environment can be significantly improved if cross-site scientific comparisons, evolutionary treatment of organized complex systems, and deeper insights into the CZ can be made.

1 Introduction

“Our own civilization is now being tested in regard to its management of water as well as soil.” Daniel Hillel (1991)

The US National Research Council (NRC, 2001) recommended the integrated study of the “Critical Zone” (CZ) as one of the most compelling research areas in Earth sciences in the 21st century. This CZ is defined by the NRC (2001) as “a heterogeneous, near surface environment in which complex interactions involving rock, soil, water, air and living organisms regulate the natural habitat and determine availability of life sustaining resources.” This zone ranges from the top of the vegetation down to the bottom of the aquifer (Fig. 1). It encompasses the near-surface biosphere and atmosphere, the entire pedosphere, and the surface and near-surface portion of the hydrosphere and lithosphere (Fig. 2). The US National Science Foundation (NSF AC-ERE, 2005) also recommended a focus on water as a unifying theme for understanding complex environmental systems. Water-related research requires enhanced understanding of processes at environmental interfaces, approaches for integrating across scales, and improved coupling of biological and physical processes. Collectively, such an integrated, interdisciplinary, and multiscale effort will advance our ability to forecast and plan for changes and to address critical societal issues such as human safety, human health, economic prosperity, environmental quality, and sustainable development.
Fig. 1. Concepts of the Critical Zone, regolith, weathering profile, soil profile, and solum (modified from Schaetzl and Anderson, 2005). The unmodified/unweathered portion of the C horizon is labelled as D horizon (after Tandarich et al., 1994). All materials above fresh, unweathered bedrock (R horizon) are called regolith, which is equivalent to a broad definition of the soil. The Critical Zone is the broadest concept, going from the top of the tree to the bottom of the aquifer. Also illustrated on the right are (1) the abundance of microbes and insects in the soil, especially in the surface soil, and (2) fresh water seeping out of an aquifer and running over exposed bedrock (modified from anonymous sources).

Soil is at the central junction of the CZ, representing a geomembrance across which water and solutes, as well as energy, gases, solids, and organisms are actively exchanged with the atmosphere, biosphere, hydrosphere, and lithosphere, thereby creating a life-sustaining environment (Fig. 2). Water is the circulating force that drives many of these exchanges and is the major transport agent in the cycling of mass and energy in the CZ. Water flux into and through the soil and over the landscape is the essence of life, which resembles the way that blood circulates in a human body (Bouma, 2006). The interactions of soil and water are so intimate and complex that they cannot be effectively studied in a piecemeal manner; rather, they require a systems and multiscale approach. In this spirit, hydropedology has emerged in recent years as an intertwined branch of soil science and hydrology that addresses the interface between the pedosphere and the hydrosphere, with an emphasis on in situ soils in the landscape setting (Lin, 2003; Lin et al., 2006a).

The pedosphere is a unique, relatively immobile sphere that is easily impacted by human activities. In contrast to the other spheres of the Earth system, the pedosphere can neither quickly intermix (as the atmosphere does), nor rapidly move laterally along the landscape (as water does), nor clearly be separated into individual units and avoid undesirable environmental changes (as the biota can be and does), nor escape rapid human and biological perturbations (as is characteristic of the lithosphere). Therefore, each soil, as a relatively immovable and formed in situ natural body, is fated to react, endure, and record environmental changes by being transformed according to the interactions of climatic, biotic, and anthropogenic forcing, as conditioned by geologic and topographic setting, over geological and biological time scales. This makes the monitoring of soil change an excellent (albeit complex) environmental assessment, because every block of soil is a timed “memory” of the past and present biosphere-geosphere dynamics (Arnold et al., 1990).

This memory takes multiple forms, such as micromorphology, profile features, and various soil physical, chemical, and biological properties. Learning to “decode” soil features and their changes into environmental information is as valuable as reading the records of ice cores for atmospheric conditions and interpreting tree rings for eco-climatic dynamics.
Given the emerging interests in the CZ, hydropedology, and related efforts in establishing environmental observatories around the world, this article reviews and discusses growing opportunities for synergistically advancing soil science, hydrology, and geosciences. This discussion is timely because soil science programs worldwide have struggled to survive, the hydrology community is becoming united because of the current global concern for fresh water, and the geosciences community has embraced an expanded vision of its role and societal relevance. Specific objectives of this paper include: (1) a clarification of the basic ideas of the CZ and hydropedology, and their relations to relevant disciplines; (2) a suggestion of a broadened perspective for future soil science advancement; (3) a highlight of fundamental characteristics and research opportunities of the CZ and hydropedology; and (4) a discussion on possible ways for advancing CZ science and hydropedology. This paper focuses on physical and hydrologic aspects of the CZ, using hydropedology as an example (among many other disciplines involved in CZ studies) to illustrate the opportunity and the need for integrated, multiscale, and synergistic efforts. Coupling physical/hydrologic and biogeochemical/ecological processes in the CZ is an emerging topic, which has been recently reviewed by McClain et al. (2003), Lohse et al. (2009), and others.

2 Critical Zone science

2.1 History of the concept of the Critical Zone and its current meaning and utility

The generic term of “critical zone” first appeared a century ago in a German article by physical chemist D. E. Tsakalotos (1909), referring to the zone of a binary mixture of two fluids. In 1962, American mineralogist E. N. Cameron called a geological formation (the Bushveld Complex in South Africa) a “critical zone.” As of 15 March 2009, based on the Science Citation Index Expanded, a total of 314 publications...
have used the term “critical zone” to refer to wide-ranging things: from geological formation where precious metals (such as platinum and gold) can be mined (Wilhelm et al., 1997) to pipe corrosion in soil within ground water fluctuation zone (Decker et al., 2008); from the rhizosphere where soil and roots have close interaction (Ryan et al., 2001) to transitional zones in alluvial coastal plain rivers important for water resource management (Phillips and Slattery, 2008); from body ventricular slow conduction area with electrophysiological limitations (Elsherif et al., 1990) to local cold regions of ice-structure interaction where intense pressures occur over short time leading to ice failure (Johnston et al., 1998). The majority of these 314 papers were in the subject areas of geosciences (46%) and minerals/energy (41%). Only 15 of them used the term of “critical zone” in relation to soils, and 15 papers referred to water related issues.

In 2001, the NRC specifically defined the “Critical Zone” (note capitalized letters). This CZ concept was suggested by a subgroup of the NRC Committee on Basic Research Opportunities in the Earth Sciences (L. Wilding, personal communication, 2006), which consisted of a sedimentologist (Gail Ashley), a pedologist (Larry Wilding), and a hydrologist (Stephen Burges). Since then, enthusiasm as well as skepticism have surfaced in scientific communities regarding the meaning and utility of this concept. An attempt is thus made here to clarify some related issues or concepts:

1. Many people thought that the CZ is (nearly) the same as soils. In reality, the CZ is much broader than soils (see Figs. 1 and 2). It is true that the CZ encompasses the entire pedosphere – the only sphere in the Earth system that is wholly included in the CZ. However, if the CZ is limited to just soils (we label this perspective as “inward” contraction; see Fig. 2), then it loses the integrating and unifying power of the CZ. Instead, a broadened perspective of soils is needed (this is labeled here as “outward” growth; see Fig. 2), thereby permitting an inclusive vision for future soil science. This “outward” growth perspective is consistent with the soil’s recognized “7+1” roles as depicted in Fig. 2 (Lin, 2005) and seven soil functions identified by the EU Soil Protection Strategy (Commission of the European Communities, 2006; Bouma, 2009). Soil scientists, while focusing on soil processes, can then see their knowledge in a much broader context.

2. Some have used the term CZ as synonymous with a common geological term “regolith” (defined as “the fragmental and unconsolidated rock material, whether residual or transported, that nearly everywhere forms the surface of the land and overlies the bedrock. It includes rock debris of all kinds – volcanic ash, glacial drift, alluvium, loess, vegetal accumulations, and soil;” Bates and Jackson, 1987). Regolith is synonymous with mantle or overburden, and is equivalent to the broader concept of soil that includes O-A-E-B-C horizons (the C horizon is often called saprolite; see Fig. 1). The classical narrower concept of soil has been driven by agriculture-centric conception of the soil as a medium for plant growth, which often only includes the A-B horizons (called solum, generally <1–2 m deep; see Fig. 1). The CZ, as defined by the NRC (2001), integrates regolith with above-regolith vegetation and below-regolith bedrock or sediments that interact with fluctuating ground water (Fig. 1).

3. Some people have questioned the utility of the CZ concept because of its imprecise lower boundary and highly variable thickness from place to place. On the contrary, this is precisely the benefit for using the CZ concept as it will force us to better define the variable and dynamic lower boundary of the active water cycle, especially under unsaturated flow conditions. Currently we do not know exactly where the active water flow ceases in the subsurface in different ecosystems and geographic regions, yet such an understanding is important as this demarcation influences the annual, decadal, and century hydrologic cycles. Hydrologic and biogeochemical models have often been forced to make assumptions about the lower boundary of the active water cycle (such as an impermeable bedrock or an artificial 2-m cut-off for soil depth). But the diffuse lower boundary of the CZ may extend to a kilometer or more beneath a fractured bedrock and the volume of water stored in this zone may be an order of magnitude larger than the combined volume of water in all rivers and lakes (NRC, 1991, 2001).

4. Many researchers believe that the CZ concept is useful because it is inherently process-oriented and is a unifying framework that accommodates the hydrologic cycle, the geochemical cycle, the carbon cycle, the nutrient cycle, gas exchange (major and trace gases), erosion and deposition, weathering (chemical and physical), lithification (diagenesis), soil formation and evolution (pedogenesis), life processes (macro- and microbial communities, including plants and animals), and human impacts (land use and management). The timescales included in the CZ concept range from seconds to eons and its spatial scales are enormous – from atomic to global (NRC, 2001). As indicated by the NRC (2001), the rapidly expanding needs for a sustainable society give special urgency to understanding the processes that operate within the complex CZ. Some pressing scientific issues involved include global climate change, terrestrial carbon cycle, life-water-mineral interactions, land-ocean interface, and earth history (NRC, 2001).
Fig. 3. Schematic of the Critical Zone (CZ) that is open to continuous energy and mass exchange with the surrounding environment, where the conservation of energy (E) and mass (M) and the accumulation of entropy (S) and information (I) are at work simultaneously, which dictate the evolutionary outcome and the functioning of the CZ. Irreversible evolution, coupled cycling, interactive layers, and hierarchical heterogeneity are the characteristics of the CZ. The reconciliation of geological and biological cycles vastly different in space and time is essential to the understanding of the complexity and dynamics of the CZ.

2.2 General characteristics of the Critical Zone

The CZ is perhaps “the most heterogeneous and complex region of the Earth” (NRC, 2001) and the soil has been recognized as “the most complicated biomaterials on the planet” (Young and Crawford, 2004; NRC, 2009). Despite such extreme complexity and dynamics, however, some general characteristics of the CZ can be identified (Fig. 3), which include:

1. *The CZ is evolutionary*. Changes in the CZ are generally irreversible and cumulative, which may be introduced by either slow, more gradual processes or extreme, abrupt changes. Increasingly, human alternations to the CZ have become pervasive and long-lasting.

2. *The CZ is a coupled system*. Cycles of geological and biological processes over vastly different spatial and temporal scales are a key to understanding the CZ, which involves the coupling of different spheres of the Earth system.

3. *The CZ has contrasting layers vertically*. The interfaces between belowground layers are critical controls of the CZ’s response and feedback times to aboveground changes such as climate and land use.

4. *The CZ is highly heterogeneous horizontally*. Ubiquitous heterogeneity in the CZ exhibit both hierarchical organizations (which exhibit gaps between conceptualized scales) and networks embedded in land mosaics (which provide efficient transfer of mass and energy across scales).

Each of the above four general characteristics is further elaborated in the following. Together, these can help forge a holistic and evolutionary worldview of the CZ, instead of the classical reductionistic and mechanistic worldview.

2.2.1 Irreversible evolution

Natural systems are open systems in terms of energy, because solar energy enters freely and heat energy goes back into space. Such thermodynamically open and dissipative systems have driven the evolution and functioning of the CZ...
(Lin, 2010b). As dictated by the second law of thermodynamics, nature has a tendency to go toward disorder, with spontaneous increase in entropy (which can be interpreted as an index of a system's disorder, according to Boltzmann, 1896). In the meantime, Prigogine's (1977) theory of dissipative structures suggests that, as energy dissipates, complex structures are formed through the system's self-organization (a process of attraction and repulsion in which the internal organization of a system increases in complexity without being guided by an outside source). A balancing act exists between these two opposing tendencies in nature: moving away from thermodynamic equilibrium (via organization) vs. slippage back towards it (through dissipation) (Ulanowicz, 2009).

Time irreversibility is essential to the evolution of the CZ (Lin, 2010b). Unlike space, time cannot be reversed because time always moves forward in one direction only. Many physical, chemical, and biological changes occurred in nature are also irreversible (Prigogine, 1977), because energy dissipation results in heat loss that cannot be totally converted back to work should the process be reversed. Consequently, the exact same conditions can not occur again in the natural world (Tiezzi, 2006). In particular, living creatures and ecosystems obey the laws of biological evolution: at any time they are different from what they were an instant before (Tiezzi, 2006). This is the essence of life. Time evolution also displays accumulative and memory effects. Indeed, life is only possible if it does not start from zero for every new generation; rather, mechanisms to store information gained, e.g., through genes, make the evolution of life possible (Jørgensen, 2006).

Over time, evolution progresses towards more and more complex and organized systems (Tiezzi, 2006), and many ordered systems exhibit emergent properties (i.e., properties that a system possesses in addition to the sum of its components' properties, that is, the whole is more than the sum of its parts).

Dissipative structure, along with information stored and aesthetics (quality) accumulated over time, play a fundamental role in natural evolutionary processes (Tiezzi, 2006). Both energy and mass conservation and entropy and information accumulation are at work simultaneously in the evolution of the CZ (Fig. 3) (Lin, 2010b). Therefore, cross fertilization between evolutionary and conservative principles is essential to understanding physical vs. biological laws that govern the structure and function of the CZ (Lin, 2010b).

The CZ is also awash in events that occur once and never again (Lin, 2010b). Jørgensen et al. (2007) and Ulanowicz (2009) suggested that singular events in nature are not rare; rather they are legion and can occur everywhere all the time and at all scales. Such a common phenomenon is probably linked to the irreversible nature of time and dissipative processes described above. Furthermore, the duality of chance and choice is characteristic of evolution of natural systems (Monod, 1971; Tiezzi, 2006; Rinaldo et al., 2006). This duality leads to complex systems that disobey linear deterministic laws; instead, the combination of chance and selection involved in evolution gives rise to unexpected structures and events, the properties of which can be quite different from those of the underlying elementary laws and may take the form of abrupt transitions, a multiplicity of states, new patterns, or irregular and unpredictable outcomes, which is referred to as deterministic chaos (Nicolis, 1995; Tiezzi, 2006).

Increasingly, human impacts on the CZ have become fundamental (Richter, 2007; Richter and Mobley, 2009). Besides the forcing of tectonics, weathering, fluid transport, and biological activities as highlighted by the NRC (2001), the human forcing is increasingly recognized. Anthropocene is now recognized as “a new geological epoch in which humankind has emerged as a globally significant, and potentially intelligent, force capable of reshaping the face of the planet” (Clark et al., 2004). Soil thickness in the CZ, for example, illustrates the accelerated changes caused by humans: average rate of soil formation from bedrock weathering has been estimated by many as 1 mm per 1000 years or less, while the rate of soil erosion has reached an average rate of 10–20 mm per 1000 years in many areas around the globe because of human activities (McKenzie et al., 2004; Brantley et al., 2006). Such an imbalance between the slow production of new soils (which is generally infertile subsoil above fresh bedrock) and the accelerated lose of fertile topsoil has severe negative and irreversible consequences for the sustainability of the CZ.

2.2.2 Coupled cycling

In terms of physical matters – soil, water, air, and other resources – the Earth is essentially a closed system from a global perspective, at least at the biological timescale. The only exceptions are the slow escape of lightweight gases (such as hydrogen) from the atmosphere into space and the input of frequent but tiny meteors and cosmic and meteoric dusts (Christopherson, 2007). Thus, the Earth’s physical materials are finite (for all practical purposes) and hence tend to depend on cycling for continued use. This implies that, no matter how numerous and daring human technological reorganizations or transformations of matters may become, the material base of the CZ is fixed and limited.

The CZ has two overarching cycles that are vastly different in space and time scales: the geological cycle (dubbed “big” cycle) and the biological cycle (dubbed “small” cycle) (Fig. 4). It is the reconciliation of these two cycles that is critical to understanding the CZ’s complex evolution and dynamic functions. The geological cycle refers to the weathering of rocks, the erosion, transport, and deposition of weathered products that eventually end up in oceans as sediments, and then are lithified and uplifted back to the land by tectonic or volcanic activities. This cycle, occurring over the geological timescale and often a large area, is composed of three major sub-cycles – tectonic, lithologic, and hydrologic (Fig. 4). The tectonic cycle brings heat energy and new materials to
the surface and recycles materials, creating movement and deformation of the crust. The rock (or rock-soil) cycle produces igneous, metamorphic, and sedimentary rocks in the crust, including the soil cycle as rocks weather into soils and soils return to rocks over the geological timescale. At this geological timescale, the hydrologic cycle processes materials with the physical and chemical actions of water, ice, and wind.

The biological cycle refers to the production and consumption of food and energy in the ecosystem and the accumulation and decomposition of organic materials in soils. The flow of energy, the cycling of nutrients, and trophic relations determine the nature of an ecosystem. Compared to the geological cycle, the biological cycle occurs over a much shorter timescale and often a smaller area. Three principle sub-cycles involved are ecological, biogeochemical, and hydrologic cycles (Fig. 4). The ecological (or life) cycle generates biomass through producers such as microbes and plants, reaching consumers and eventually detritivores through the food chain. Soils support vast communities of microorganisms that decompose organic matter and re-circulate elements in the biosphere (see an illustration in Fig. 1). Anthropogenic impacts may be considered as part of the ecological cycle; alternatively, human activities can be elevated to a separate cycle because of their increasingly dominant impacts on the biological cycle. The biogeochemical (or elemental) cycle, combining biotic and abiotic processes, redistributes elements and materials through liquid (e.g., water), solid (e.g., sediments), and gas (e.g., air). At the biological timescale, the hydrologic cycle transports organic and inorganic materials and energy throughout the CZ.

Fluid transport is involved in both the geological and biological cycles, since water is the key conduit for mass and energy transfer. Life on Earth as we know it would be impossible without liquid water. The biosphere is ultimately what ties the major systems of the Earth together and drives them far out of thermodynamic equilibrium (Jacobson et al., 2000). The cyclic approach mimics the coupled nature of the CZ, e.g., biogeochemical cycle is inseparable from the hydrologic cycle, and together they determine the ecological cycle. However, traditional discipline-limited and individual component-based efforts have plagued our understanding of the coupled CZ system. It is therefore hoped that the emerging CZ science can significantly advance the holistic and evolutionary worldview of complex coupled systems such as the CZ.

The cyclic approach allows conceptual simplification of materials movement and their couplings to the environmental factors. Using the most basic description of the cyclic processes, mathematical models have been used to describe and predict the distribution of important elements of interest, such as water, sediment, carbon, nitrogen, and others (Jacobson et al., 2000). Many computer models, especially those of the global scale, use this approach. In many cases, the details of the distribution of an element within each of the reservoirs are disregarded, and for the most simplified calculations, the amounts of materials in each reservoir are assumed to be at steady-state. This allows an element budget to be defined for the entire cycle. Such a steady-state, budget-based approach, however, has limitations because it provides little or no insight into what goes on inside each reservoir or the nature of the fluxes between reservoirs (Jacobson et al., 2000). The average-based analysis also does not consider spatial and temporal variation, and therefore could give a false impression of certainty.

### 2.2.3 Interactive layers

The Earth is a vertically layered system: from the outer atmosphere down to the inner core, different layers of materials with vastly different thicknesses and characteristics are evident (Figs. 1 and 2). The CZ consists of the aboveground vegetation zone, the belowground root zone, the deeper vadose zone, and the saturated aquifer zone, with each layer having various sub-layers. This layering has a general trend of increasing density with depth (largely because of gravity, but with exceptions), and has the dampening effect of state variables with depth and an increase in distance to energy input at the soil surface. Therefore, there is an overall trend of increasing characteristic response time (i.e., time period needed to reach a quasi-steady state with the environment) to external perturbations from the atmosphere, down to the hydrosphere, biosphere, pedosphere, and further down to the lithosphere (Arnold et al., 1990). Conversely, the time period needed to feedback to climate and land use changes generally
decreases when moving up from the deep underground zone to the shallow subsurface zone to the surface zone and to the aboveground zone.

Key interfaces in the CZ include the land surface-atmosphere interface, the soil-vegetation interface, the vadose zone-ground water interface, the surface water-ground water interface, the soil-stream interface, and the soil-bedrock interface. These interfaces are unique and important controls of the landscape-soil-water-ecosystem-climate relationships. For example, fluctuating ground water has a significant influence on the nature and properties of the lower part of soil profiles (Jenny, 1941: Narasimhan, 2005). Changing ground water levels also has effects on the energy balance at the soil surface and the susceptibility of a region to drought (Maxwell and Kollet, 2008). The NRC (2004) highlighted the importance of ground water fluxes across interfaces when estimating recharge and discharge to aquifers. However, many challenges remain in understanding and measuring dynamic interchanges among the water reservoirs of atmosphere, surface, and subsurface, especially with the subsurface. Sophocleous (2002) reviewed recent advances in ground water and surface water interactions, and pointed out that most models today are not well equipped to deal with local phenomena related to flow near domain boundaries (i.e., interfaces), and yet the upper few centimeters of sediments beneath nearly all surface water bodies (i.e., hyporheic zone) have a profound effect on the chemistry of water interchanges. Pielke et al. (1998) showed convincing evidence that terrestrial ecosystems significantly interact with atmospheric processes, and thus they are as important as changes in atmospheric dynamics and composition, ocean circulation, ice sheet extent, and orbit perturbations in the studies of past and future climate change.

Within a soil profile, important interfaces also exist, including the soil horizons interface, the soil-root interface, the macropore-matrix interface, the microbe-aggregate interface, the water-air interface, and the soil-water table interface, all of which are places where important interactions occur. For instance, soil chemical reactions generally occur at various interfaces, including the equilibrium and kinetic processes of dissolution/precipitation, adsorption/desorption, oxidation/reduction, and polymerization/biodegradation. Another example is illumination that leads to various coatings on soil interfaces, such as clay films, carbonates, or redox features at the surfaces of soil particles or aggregates. Interfaces between soil layers often slow down vertical water percolation and promote lateral flow, especially in sloping landscape with an underlying water-restricting layer. Interfaces between soil horizons with different textures or structures also often trigger preferential flow, and thus can impact the scaling of flow, transport, and reaction processes. Such interfaces are also important in defining at what depth a soil profile will begin to saturate and at what time of a year, giving rise to various runoff patterns in catchments (Lin et al., 2008b).

2.2.4 Hierarchical heterogeneity

Ubiquitous heterogeneity and scale bridging underlie nearly all of the CZ studies. Heterogeneity here differs from randomness: the former is associated with order while the later is linked to disorder (Lin, 2010a, b). Currently no single universal theory has emerged that is ideal for spatial aggregation (or upscaling), disaggregation (or downscaling), and temporal inference (or prediction) of diverse processes occurred in the CZ. Despite this, certain causes of heterogeneity in the CZ can be identified, and possible ways of bridging scales can be further explored.

Heterogeneity is expected based on Prigogine’s (1977) theory, i.e., a dissipative system in nature is characterized by the spontaneous appearance of symmetry breaking (leading to anisotropy) and the formation of complex structures (leading to heterogeneity). Soil-forming theory also offers a useful framework for understanding diverse mechanisms leading to subsurface heterogeneity, where systematic (ordered) variation can be identified by environmental controls (such as climate, organisms, landform, and parent material), whereas random (disordered) variation is stochastic. Differentiation of systematic vs. random variations allows a focus on the portion of the CZ variability that can be related to known causes, because patterns of heterogeneity are the diagrams of processes (Bell, 1999).

Patterns in the landscape, however, are often unclear because they comprise many layers of elements, with each element having its own heterogeneity. These elements are often intricately woven together due to the interactions of different processes occurred in the CZ. Bell (1999) suggested that it is possible to consider landscapes as complexes of networks and mosaics. The networks are patterns of linear-oriented features, such as the meandering and branching systems that run through and between the elements that produce the mosaics. Mosaic patterns can be found over a wide range of spatial scales (e.g., from the submicroscopic soil matrix to the entire pedosphere). Mosaics arise because of uneven and dynamic energy inputs into the open system of the CZ, leading to structural and compositional heterogeneity at all scales. The mosaic patterns can be determined by mechanisms characteristic of underlying processes, including (Bell, 1999):

- **Inherent processes**: the substrate heterogeneity beneath the land mosaic is dependent on the processes of geology, geomorphology, pedology, and hydrology, interacting with the climate and biota;
- **Extrinsic processes**: the effect of natural disturbances (such as fires, hurricanes, insect pests, and diseases) to the biota that colonize and grow on the variable substrate;
- **Anthropomorphic processes**: human activities ranging from land use/land cover changes to modifications of landform, urbanization, and interference with the climate and biota.
Pedon at a local point is considered here as the basic scale \( i \) of observation or monitoring, and \( i \pm 1 \ldots 4 \) indicate arbitrary labels of larger (+) or smaller (−) scales. Note gaps exist between scales. (b) Three types of systems with respect to different methods of thinking (modified from Weinberg, 1975). Note the wide gap between the two extremes, where systems of medium numbers are too complex for analysis and too organized for statistics.

There is an apparent hierarchy in the heterogeneity of soils and the CZ (Fig. 5). Patterns may emerge at a large scale (i.e., emergent properties) from the complex interactions of a large number of different elements at a smaller scale. The dominant process and controlling factor may also change in a hierarchical manner as scale changes. Therefore hierarchical frameworks have been utilized by geoscientists, soil scientists, and hydrologists as a means for organizing natural systems from the pore scale up to the global scale (Fig. 5). However, a quantitative and operational hierarchy that can be integrated into models of scaling, flow, and rate processes remains a major research challenge today, where gaps exist in between hierarchical levels (Fig. 5). The present inability to adequately characterize subsurface heterogeneity exacerbates the scaling problem and leads to significant uncertainties in data interpretations and model predictions for the CZ across scales (Sophocleous, 2002).

Von Bertalanffy (1968) indicated that there is a general lack of appropriate scientific means for dealing with systems between extremes – the systems of so-called medium numbers (Fig. 5). Weinberg (1975) described three types of systems that require different thinking:

1. Systems of “organized simplicity” (Region I in Fig. 5): this is the region of machines, or mechanisms. This region has small populations with a great deal of structure, which could be treated analytically;

2. Systems of “unorganized complexity” (Region II in Fig. 5): this is the region of large populations, where statistical mechanics applies. This region is complex, but yet sufficiently random so that it is sufficiently regular to be studied statistically. Randomness (or structureless) is the property that makes statistics work here;

3. Systems of “organized complexity” (Region III in Fig. 5): this is the region too complex for analysis and too organized for statistics. This is the region of systems with medium numbers, within which the two classical methods (analytical and statistical treatments) essentially fail.

As remote sensing techniques for monitoring large-area phenomena and in situ sensors for point-based measurements continue to advance, bridging multiple scales from points to watersheds and to the entire globe becomes essential. While advanced imaging, spectroscopic, and other technologies become widely used at the pore or even molecular levels, and computer modeling coupled with remote sensing and other techniques are increasingly used at the global level, the intermediate scale (e.g., from the pedon to the hillslope and catchment scales; Fig. 5) becomes a significant gap that is often neglected and has only limited and often rudimentary tools and techniques for investigating the most complex and dynamic subsurface. This intermediate scale is also
essentially the region of medium number systems that urgently needs a new way of thinking and a significantly improved scientific treatment (Lin, 2010b).

3 Hydropedology

3.1 Fundamentals of hydropedology

Hydrogeosciences have encountered a new intellectual paradigm that emphasizes connections between the hydrosphere and other components of the Earth system. While hydrogeology, hydrometeorology, and ecohydrology are now well recognized, an important missing piece of the puzzle is the interface between the hydrosphere and the pedosphere. Hydropedology addresses this interface, and seeks to answer the following two basic questions (Lin et al., 2008a, b):

1. How do soil architecture and the distribution of soils over the landscape exert a first-order control on hydrologic processes and associated biogeochemical and ecological dynamics across spatio-temporal scales?

2. How does landscape water and the associated transport of energy, sediment, chemicals, and biomat-erials by flowing water influence soil genesis, evolution, variability, and functions?

Landscape water here encompasses the source, storage, availability, flux, pathway, residence time, and spatio-temporal distribution of water in the variably unsaturated or saturated near-surface environment (Lin et al., 2006a). While source, storage, availability, and flux of water have been studied extensively in the past, attention to flow pathways (especially flow networks), residence time (age of water), and spatio-temporal pattern of flow dynamics (and its underlying organizing principle) has been much limited (Lin et al., 2006a; McDonnell et al., 2007).

Fundamental scientific issues of hydropedology, at this point of its development, may be considered under the following four headings (Lin et al., 2005):

1. Soil structure and horizonation vs. in situ water flow and chemical transport: Hydropedology emphasizes soil architecture (solid + pore + their interfaces) and its quantitative links to preferential flow across scales;

2. Soil catena and distribution pattern vs. water movement over the landscape: Hydropedology focuses on quantitative relationships between soils and their surrounding landscape and the impacts of such relationships on hydrologic (and related biogeochemical/ecological) processes;

3. Soil morphology and pedogenesis vs. soil hydrology and soil change: Hydropedology emphasizes quantitative soil hydromorphology as a signature of soil hydrology and the soil as valuable records of environmental change over time;

4. Soil functions and maps vs. “carriers” of soil quality and soil-landscape heterogeneity: Hydropedology promotes quantitative delineation of functional soil units in the landscape as well as precision soil-landscape mapping for diverse applications.

Lin et al. (2005) have provided a comprehensive review on these fundamental issues. In the following, only a brief update is provided to exemplify the contributions from hydropedology to the needed integrated, interdisciplinary, and multiscale studies of the CZ.

3.1.1 Soil structure and horizonation: their impacts on preferential flow

Natural soils generally contain multiple horizons, with each horizon structured to various degrees (Fig. 6). Soil architecture is the entirety of how the soil is structured, which encompasses at least three parts: (1) solid components, including soil matrix (represented by soil texture and soil microfabric) and soil aggregation (represented by the type, quantity, and size distribution of peds, and aggregate stability); (2) pore space, including the size distribution, connectivity, tortuosity, density, and morphology of various pores; and (3) interfaces between solid components and the pore space, such as coatings on peds or pores, and the macro pore-matrix, soil-root, microbe-aggregate, and horizon interfaces.

Because of heterogeneous soil architecture, variability in energy and mass inputs to soils, diversity in biological activities, and nonlinear dynamics of hydrologic processes, preferential flow can occur in practically all natural soils and landscapes (Fig. 6) (Lin, 2010a). As Clothier et al. (2008) summarized, preferential flow can occur spatially at the pore scale of spatial order 10⁻⁵ m, at the core scale (10⁻¹ m), in pedons (10⁰ m), down hillslopes (10¹–10³ m), through catchments (10⁴–10⁵ m), and across large regions of ≥10⁶ m. Time-wise, preferential flow can operate during fluid flows at the temporal order of 10⁰–10¹ s, during hydrological events 10⁰–10³ h, throughout seasonal changes 10⁰ year, and across inter-annual variations of 10¹ years.

Based on three theoretical considerations and numerous published evidence, Lin (2010a) has attempted to justify the likely universality of preferential flow in natural soils — meaning that the potential for preferential flow occurrence is everywhere in nature, although the actual occurrence of preferential flow depends on local conditions (Lin and Zhou, 2008). Lin (2010a) also showed that networks are abundant in soils, such as root branching networks, mycorrhizal mycelial networks, animal borrowing networks, crack and fissure networks, artificial subsurface drainage networks, and pore networks between soil particles and aggregates. These networks provide preferential flow conduits, which in return reinforce or modify the existing networks.
3.1.2 Soil catena and distribution: their controls on landscape hydrology

There appear two general categories of soil distribution patterns that can be differentiated in terms of their main controls at different scales: (1) At the hillslope and landscape scales, soil patterns are heterogeneous mainly due to factors that vary over short distances such as topography and parent materials (i.e., the site factors of soil formation). Such a soil pattern is often referred to as a “soilscape”, i.e., the pedologic portion of a landscape (Hole, 1985; Buol et al., 2003), which includes catenae and other more localized soil distribution patterns (such as gleization and Histosols); and (2) At the regional and global scales, zonal soil patterns can be expressed by a gradual change in soils over large areas, resulting from climatic and vegetative gradients (i.e., the flux factors of soil formation). These physiographically-oriented soil patterns are recognized in the US as Major Land Resources Areas (which are defined as geographically associated land resource units that are characterized by a particular pattern of soils, water, climate, and land use; USDA-NRCS, 2006).

Catenae in different climatic and physiographic regions may exhibit markedly different relationships between soil and hydrologic properties (e.g., Schaetzl and Anderson, 2005; Lin et al., 2006b; Wilcox et al., 2007; Lin et al., 2008b). Drainage condition, water table depth, and fluxes of water, solutes, and sediments typically differ in soils along a catena (Fig. 7). However, contrasting hydrology and soil morphology (e.g., Fig. 7a vs. b) may occur depending on the local controls of surface or subsurface topography. Figure 7a can be explained by the classical catenary model where surface topography controls hydrological regimes; whereas in Fig. 7b the topography of the underlying weathered rock substrates with low-permeability causes subsurface distribution patterns of soil and hydrology (Coventry, 1982). The soil-bedrock interface and bedrock topography have now been recognized as important to subsurface stormflow in many hillslopes (e.g., Freer et al., 2002; Tromp Van Meerveld and McDonnell, 2006; Lin et al., 2006b). Sommer and Schlichting (1997) suggested that an integration of three different approaches to study soil distribution over the landscape – geomorphic/stratigraphic, hydrologic, and pedologic approaches – is the most promising way forward.
Soil macro- and micro-morphology have long been used to infer soil moisture regimes, hydraulic and biogeochemical properties, and landscape processes (Lilly and Lin, 2004; Lin et al., 2008b). In particular, water-dominated pedogenesis leads to so-called soil hydromorphism (Figs. 6–8) – a result of permanent or temporary state of soil saturation in the soil associated with conditions of reduction (predominantly the accumulation or loss of Fe, Mn, S, or C compounds) (USDA-NRCS, 1998). The spatial relationships of redox depletions and redox concentrations may be used to interpret the pattern of water and air movement in soils (Veperaskas, 1992). The presence of organic matter and a suitable temperature and pH are generally required for hydromorphism to occur. Such biochemical processes might offer implications for finding possible clues to life on Mars – if Martian soil hydromorphism or paleo-hydromorphism (i.e., formed in ancient condition) can be found. This is because biological activity is often involved in soil hydromorphism on the Earth, while diagenetic hydromorphism is also considerably accelerated by microorganisms.

Hydromology has been suggested as a possible integrating factor of soil formation and a main driving force of soil dynamic changes (Lin et al., 2005). This is because all of the five natural soil-forming factors affect and are affected by hydrology. Water from precipitation is a primary requisite for parent material weathering and soil development. To reach a highly developed stage, sufficient amount of water must not only enter the profile and participate in weathering reactions, but also percolate through the profile and translocate weathering products (such as solutes and clays). Therefore, the characteristics of a soil profile may reflect the total amount water that has passed through over time, which has potentials in quantifying pedogenesis. For example, Fig. 8a illustrates a general sequence of soil development from a weakly-developed Entisol to a highly-weathered Ultisol under well-drained condition: as soil age increases, soil thickness increases as weathering progresses, but saturated hydraulic conductivity in the subsoil (e.g., B horizon) would first increase and then decrease after reaching a certain developmental stage (e.g., an argillic horizon in an Ultisol may turn into an aquitard, impeding further vertical percolation of water). Under poorly-drained condition, soil genesis and morphology would be quite different, as illustrated in Fig. 8b, which can be quantitatively related to seasonal water table fluctuation and frequency of soil saturation.

3.1.4 Soil functions and maps: their connections to soil hydraulic properties

Soil quality is the capacity of a soil to function within ecosystem boundaries to sustain biological productivity, maintain environmental quality, and promote plant and animal health (SSSA, 2009). Soil carbon content is widely recognized as a major factor in the overall quality and functions of soils, but human land use/management has significant impacts on soil carbon and overall soil quality. The concepts of “genoform” (for genetically defined soil series) and “phenoform” (for soil types resulting from a particular form of management in a given genoform) (Droogers and Bouma, 1997) offers an interesting means to incorporate management effects into hydopedologic characterizations of soil functions, which can enhance pedotransfer functions that involve soil series and land use as carriers of soil hydraulic properties (Lin, 2003). Soil map is a common way to portray soil heterogeneity and to provide soil input parameters to models. However, virtually every delineation of a soil map unit includes other soil components or miscellaneous areas that are not identified in the name of a map unit because of limitations in map scale and other factors (Soil Survey Division Staff, 1993). Quantification of map unit purity for different scales of soil maps is a needed area of improvement in modern soil surveys.
3.2 Characteristics of hydropedology and its link to Critical Zone science

Two general characteristics of hydropedology are suggested below, which link to the two basic questions of hydropedology as described in Sect. 3.1.

First, hydropedology emphasizes in situ soils in the landscape, where distinct pedogenic features (e.g., aggregation, horizonation, and redox features) and soil-landscape relationships (e.g., catena, soil distribution patterns, and soil map units) are essential to understanding interactive pedologic and hydrologic processes (Fig. 6). Developing quantitative relationships between complex natural soil architecture and soil hydrologic functions across scales is an important research area of hydropedology. Three related key aspects are noted here:

(Arnold and Wilding, 1991). Precision soil mapping is of increasing demand for site-specific applications of soils information (such as precision agriculture, landscape hydrology, and urban development). Modern soil maps also need improvements for functional characterizations rather than just for general land use planning as done in the past. Soil maps can no longer be static documents; rather, derivative and dynamic maps, tailored for a specific function or purpose, must be generated and updated on a regular basis. Making connections between hydropedology and digital soil mapping is an exciting research area, which can improve the connection between spatial soil mapping and process-based modeling.

Fig. 8. (a) A general sequence of soil development from a weakly-developed Entisol (left) to a highly-weathered Ultisol (right) under well-drained conditions. The graduate formation of various soil horizons and the deepening of soil profile through time depend on the weathering rate of the underlying bedrock (R horizon), the accumulation rate of organic matter (O and A horizons), and the percolation rate of water through the soil profile. Further development from Alfisol to Ultisol leads to distinct eluviation of clay, iron, and aluminium oxides, leaving behind a light color and coarse texture E horizon and forming a high clay accumulation Bt horizon that often becomes an aquitard. (b) Two soil profiles with typical hydromorphic features and their related monitoring data. On the left is a hydric soil from the Mid-Atlantic Coastal Plain in Maryland that is related to both organic matter accumulation and depletion/segregation of iron oxyhydroxides. The yearly water table fluctuation below the soil surface and the corresponding redox potential (Eh values) at 1 m depth are illustrated (from Rabenhorst et al., 1998). On the right is a hydric soil from the Coastal Plain of North Carolina, where the amount of redox depletions (formed by Fe reduction) increases with depth, and corresponds well to the average number of saturation events per year. Saturation event is calculated as \((T - 1) + (D/21)\), where \(T\) is the number of times of saturation with a period of \(>21\) days in a year, and \(D\) is the longest period of saturation in a year (courtesy of M. Vepraskas of North Carolina State Univ.)
Hydropedology calls for a new era of soils research that is based on soil architecture (rather than solely on soil texture) so that the prediction of flow and reaction pathways, patterns, and residence times can be made more realistically. This requires innovative techniques for improved quantification of soil architecture at different scales, especially in situ noninvasively, and then linking such soil architectural parameters to field-measured soil hydraulic properties;

- Hydropedology considers the soil in the real world as a “living” entity in the landscape, not a “dead” material. As Kubiena (1938) pointed out, a crushed or pulverized sample of the soil is related to the soil formed by nature like a pile of debris is to a demolished building. Lin (2007) also suggested that a crushed soil sample is as akin to a natural soil profile as a package of ground beef is to a living cow;

- Hydropedology attempts to link the form and function of a soil system across scales (Lin et al., 2006a), rather than mapping soils without considering soil functions or modeling soils without incorporating soil architecture and soil-landscape patterns.

Jenny (1941) noted, “The goal of soil geographer is the assemblage of soil knowledge in the form of a map. In contrast, the goal of the “functionalist” is the assemblage of soil knowledge in the form of a curve or an equation... Clearly, it is the union of the geographic and the functional method that provides the most effective means of pedological research.” Such a union of soil maps and soil functions is what hydropedology hopes to accomplish, in quantitative ways.

Second, hydropedology deals with the variably unsaturated or saturated zone in the near-surface environment, including the shallow root zone, deeper vadose zone, temporally-saturated soil zone, capillary fringe, wetlands, and subaqueous soils (soils formed in sediment found in shallow permanently flooded environments such as in an estuary) (Demas and Rabenhorst, 2001). Three related key aspects are noted here:

- Hydrology has the potential to be an integrating factor for quantifying soil formation and evolution as well as for understanding soil changes (Lin et al., 2005), hence a focus on water can provide an enhanced means for quantifying dynamic soil functions;

- New ways of characterizing and mapping soils could be improved by linking to hydrology, such as delineating hydropedologic functional units as soil-landscape units with similar pedologic and hydrologic functions (Lin et al., 2008b);

- The interpretation and quantification of soils as historical records of environmental changes could be improved if hydrologic data are considered simultaneously. This has been demonstrated in some palesol and palehydrology studies (e.g., Ashley and Driese, 2000).

Hydropedology is closely linked to CZ science because of the key roles that soil and water play together in the CZ’s evolution and functioning. Specifically, these links include the following:

- The interrelationships between hydropedology and eco-hydrology, and how they influence soil moisture, ground water recharge, ecological health and diversity, and environmental quality in general (e.g., Young et al., 2007; Li et al., 2009);

- The integration of hydropedology and hydrogeology, which can provide a more holistic view and prediction of subsurface flow and transport from the ground surface all the way down to the aquifer (e.g., Lin, 2003);

- The linkage between hydropedology and hydrometeorology, which includes issues related to soil moisture and global climate change, soil carbon sequestration, greenhouse gas emission from soils, and remote sensing of soil climate (e.g., Lam et al., 2007);

- The coupling of hydropedology and biogeochemistry, including the identification of hot spots and hot moments of biogeochemical cycles in different ecosystems (e.g., McClain et al., 2003; Lohse et al., 2009);

- The study of palesols and palehydrology, which shows valuable historical records of the past environment and ancient landscape-soil-water-ecosystem-climate relationships (e.g., Ashley and Driese, 2000);

- The connection between hydropedology and land use planning, because how natural soils “throb” upon precipitation inputs under various climates offers clues as to “what” can best be done and “where” with the lowest risks and the greatest opportunities for land use and management (e.g., Bouma, 2006).

4 Opportunities for advancing Critical Zone science and hydropedology

Given the growing interests in CZ science (e.g., Brantley et al., 2007; Richter and Mobley, 2009) and hydropedology (e.g., Lin et al., 2006a, 2008a; Li et al., 2009), it is beneficial at this early stage of their developments to discuss some possible ways forward. In this section, three approaches are discussed, which are based on recent efforts in Critical Zone Observatories (CZO’s) in the US (Anderson et al., 2008) and other related efforts around the world (Richter and Mobley, 2009). Key approaches or steps that need to be taken...
or strengthened in order to achieve significant advancements in our understanding and prediction of the CZ are explored, with a focus on those aspects germane to hydropedology. No attempt, however, is made here to exclusively consider all possible approaches or steps. The examples given below are meant to be illustrative. Such a discussion hopefully can stimulate more synergistic efforts in the community.

### 4.1 Integrated systems approach

Most natural systems are irreducible, meaning that the systems properties cannot be revealed by a reduction to some observations of the behavior of their components. Instead, it is necessary to observe the entire system to capture its behavior because everything is dependent on everything else by direct or indirect linkages, and emergent properties exist at the systems level that are beyond the individual components’ properties. Tiezzi (2006) suggested the need for a new theory to describe ecosystem’s behavior, because the laws that describe the system may be qualitatively different from those that govern its individual units. Reductionist’s approach does not work well for revealing the complexity in soils and the CZ because of evolutionary nature, cumulative effects, and constant interactions and feedback loops among individual units. Toffler (1984) remarked, “One of the most highly developed skills in contemporary Western civilization is dissection: the split-up of problems into their smallest possible components. We are good at it. So good, we often forget to put the pieces back together again . . . We say ceteris paribus—all other things being equal. In this way we can ignore the complex interactions between our problem and the rest of the universe.”

The four general characteristics of the CZ discussed in Sect. 2.2 offer a holistic framework for a systems-based understanding of the CZ (Fig. 3). Lin (2010b) further discussed the use of non-equilibrium thermodynamics, general systems theory, and complexity science to help develop an evolutionary and holistic worldview of the soil and the CZ (which goes beyond the classical mechanistic and reductionistic worldview). In particular, Lin (2010b) emphasized the accumulation of entropy and information in the evolution of the CZ, in addition to the conservation of energy and mass (Fig. 3). Entropy, as a core thermodynamic variable, offers a possible new perspective to understand the interactions between soil systems and their surrounding environment. Entropy has the intrinsic properties of time irreversibility as well as quality and information that other thermodynamic functions lack (Tiezzi, 2006; Jørgensen et al., 2007; Ulanowicz, 2009). “Conservation without evolution is death. Evolution without conservation is madness.” These words of Bateson (1979) underline a fundamental characteristic of complex natural systems such as soils and the CZ.

Integration of disciplinary research is a key to future progress in CZ science. Besides hydropedology, many other disciplines are involved in CZ studies (e.g., Brantley et al., 2006, 2007; Richter and Mobley, 2009; Li et al., 2009), as alluded to in Figs. 1 to 5. The CZ’s general characteristics, as summarized in Fig. 3, imply the following four grand challenges for advancing CZ science that demand an integrated systems approach:

1. **Forcing:** Traditional geosciences have viewed humans and life processes as external drivers rather than as intrinsic parts of the whole system. However, it is these forces that have driven the CZ far away from equilibrium, and provided the CZ with chances and choices in its evolutionary process (Lin, 2010b). The hydrologic community is now keenly interested in human impacts and biological forcing in the hydrologic cycle (CUAHSI, 2007); the geochemical community is tackling weathering that includes anthropogenic forcing (Anderson et al., 2004); the soil science community is embracing land use/management impacts on soils as a new frontier (Richter, 2007).

2. **Coupling:** Biogeochemical and ecological cycles are tightly coupled with the hydrologic cycle and the soil reservoir. Reconciliation of the geological “big” cycle and the biological “small” cycle is essential for understanding and predicting the CZ (Fig. 4). The classical average budget-based cyclic approach needs to be complemented by more process-based dynamic approach to quantify enormous variability across space and time.

3. **Interfacing:** The presence of interfaces between two or more compartments is an important element of the CZ (Fig. 2). The role of these interfaces in understanding the interactions and feedbacks between compartments and their regulatory impacts on the whole CZ deserves more attention in defining research questions for the CZ (H. Vereecken, personal communication, 2009). Key interfaces in the CZ provide fertile ground for transformative research, where interactions among physical, chemical, biological, and anthropogenic processes prevail, leading to possible “hot spots” and “hot moments” in the CZ (McClain et al., 2003).

4. **Scaling:** Scale bridging remains a huge challenge in nearly all hydrologic, pedologic, biogeochemical, and ecological studies. It is highly desirable to clarify how dominant processes and their controls change with spatial and temporal scales, and to explore quantitative means to transfer knowledge from microscopic (e.g., molecular and pore), to mesoscopic (e.g., pedon and catena), macroscopic (e.g., watershed and regional), and to megascopic (e.g., continental and global) levels. The intermediate scale associated with medium number systems (Fig. 5) deserves particular attention, which appears to be the most challenging but also the most easily neglected.
4.2 Evolutionary mapping-monitoring-modeling framework

An iterative loop of mapping, monitoring, and modeling (3M) can provide a possible integrated and evolutionary approach to address the complexity and dynamics in the CZ (Fig. 9), perhaps as a good starting point to deal with many uncertainties involved. Such an evolutionary approach allows the development of adaptive strategy and the refinement of models and monitoring networks as knowledge and database are accumulated.

Monitoring is essential to record temporal dynamics and coupled cycles in the CZ. It also provides model inputs, validates model outputs, and supplies ground truthing for remote sensing. Considering the multi-phase nature of the soil system (gaseous, liquid, solid, and biotic phases), it is inadequate to determine soil change by only one characteristic, because each soil phase and property has its own characteristic response time. Some soil changes are inherently long-term, undetectable in a short period of time, but irreversible and threshold-like in the long-term. This evolutionary process presents significant challenges to designing and implementing scientific monitoring program for the CZ. However, ecosystem services, biodiversity, sustainable land management, and global environmental changes all depend on the evolution of the soil. Therefore, long-term monitoring of soil changes is necessary.

Models are useful tools for quantitative assessment, knowledge integration, and scenario prediction. Models can also guide the design and site selection for monitoring and ground truthing. When integrated with real-time monitoring and spatially-distributed mapping, models can better address temporal trends and spatial patterns. However, the hydrology community is awaiting a conceptual breakthrough that can go beyond the classical small-scale physics (Beven, 2006; Kirchner, 2006; McDonnell et al., 2007; Lin, 2010a, b). For example, subsurface flow networks are often embedded in land mosaics, which may be conceptualized to provide an alternative to modeling coupled hydrologic and biogeochemical processes, where internal network structures can govern vertical and lateral preferential flow dynamics and threshold-like hydrologic response (e.g., Tromp Van Meerveld and McDonnell, 2006; Lehmann et al., 2007; Lin et al., 2008b). Such an approach, however, would hinge on technological breakthrough in characterizing subsurface networks (Lin, 2010a).
The value of mapping in the study of the CZ should not be overlooked. Ground-based monitoring generally collects point-based data, while modeling often attempts to cover large areas. A key to connect these two is the fabric of the subsurface over the landscape that should be mapped at appropriate spatial and temporal resolution so that meaningful extrapolation and upscaling of point-based data could be made. Mapping also provides a means of diagnosing and stratifying the landscape for determining optimal location and number of monitoring sites and for designing meaningful model experiments. Mapping is also a prerequisite for spatially-distributed modeling. While geostatistics provides powerful interpolative tools after a dataset has been gathered on a particular area, it is not a powerful extrapolative tool. Geostatistical functions should be derived from landscape stratified units (such as soil type, geology, land use) and not indiscriminately across a broad landscape without prior partitioning of the causes of variability. In addition, we can use the knowledge of soil formation and the known interrelations of soils within landscapes to establish a mapping strategy that is supported by a soil-landscape modeling.

Inherent heterogeneity of spatial structures across scales remains a major challenge in understanding the CZ. All processes within the subsurface are bound to this structural framework, which is typically unknown or hard to quantify with currently available technologies (Vogel and Roth, 2003; Lin, 2010a, b). This is a fundamental difference compared to atmospheric monitoring where the heterogeneity of the system can be explored at one sensor location. But the signals of two sensors at nearby locations in many soils may be completely uncorrelated. This is why we need the mapping of the subsurface heterogeneity. Then, and only then, can the point-based monitoring provide the required observations to develop and improve predictive potential of process-based models (H.-J. Vogel, personal communication, 2009). The solid land is unlike the atmosphere and the ocean (which have much greater intermix and can be modeled as a continuous fluid); rather, the land poses hierarchical heterogeneities with controlling structures that are different at various scales, which dictate the pathways and patterns of flow and reaction. Flow networks and their temporal changes are as important as flow rates in the three-dimensional heterogeneous landscape.

Significant technological advancements are needed in many aspects of CZ science and hydropedology, such as:

- The need for new and improved sensors that are reliable in all seasons and robust for long-term monitoring. In addition, many CZ properties can not yet be monitored in situ with a real-time sensor;

- The need for advanced subsurface characterization technologies. Currently, the spatial resolution, or temporal frequency, or depth penetration of many geophysical and remote sensing technologies remain limited for high-resolution investigations of belowground CZ;

- The need for online databases that are open to the community and are comprehensive with visualization and analysis toolboxes for extracting useful information out of these databases effectively.

### 4.3 Fostering a Global Alliance

With growing interest in international scientific communities to establish various environmental observatory networks and to address “big” science questions, a synergistic effort to foster a global alliance for monitoring, mapping, and modeling of the CZ is desirable. Long-term monitoring, along with precision spatial mapping and process-oriented modeling, of the CZ across scales and geographic regions (Fig. 9) can serve many purposes of societal importance. Optimization of whole systems for multiple benefits rather than one benefit permit synergistic outcomes and would be more cost-effective in the long-run. An integrated network for observing, modeling, and sustaining the Earth’s CZ, as a whole (rather than individual compartments of the CZ) is still in its early stages of development, but it is clear that it will require inputs from many basic and applied disciplines. No one team or organization can do that alone, and a diversity of funding sources supporting a heterogeneous mixture of overlapping programs is probably the best formula for long-term stability of observatory networks (Keeling, 2008). Therefore, a global alliance is suggested here.

In 2007, the US National Science Foundation funded three Critical Zone Observatories (CZO) (Fig. 9) “that will operate at the watershed scale and that will significantly advance our understanding of the integration and coupling of Earth surface processes as mediated by the presence and flux of fresh water” (http://www.nsf.gov/pubs/2006/nsf06588/nsf06588.pdf). In 2009, three additional CZOs are selected for funding (Fig. 9). In the meantime, German Helmholtz Association has established four TERRrestrial ENvironmental Observatories (TERENO) in 2008–09 to investigate the consequences of global change for terrestrial ecosystems and the associated socioeconomic implications (http://www.tereno.net/). In 2009, the EU 7th Framework Programme (http://cordis.europa.eu/fp7/) has selected a project for funding four CZOs in EU (S. Banwart, personal communication, 2009). Similar efforts are also being pursued in some other countries around the globe. Therefore, now is the right time to foster a global alliance for studying the CZ, which may be best realized in a coordinated way to maximize the benefit for global environmental research. Furthermore, interests have also emerged in the community to forge independently conceived observatories into a network from which broader understanding – larger spatial scales, cross-site comparisons, and deeper insights – may be gained (Anderson et al., 2008). Together, our capability to predict the behavior and evolution of the CZ in response to changing environment can be improved significantly – if such a global alliance can be fostered effectively.
 Besides spatial networking of CZOs and alike, the long-term nature of monitoring is also critical. In this time of accelerating global change, continuous CZ observations are essential as they have the potential to open our eyes for unexpected but relevant developments and processes. For example, the famous “Keeling Curve” of long-term CO$_2$ data has demonstrated the value of continuous recording of a seemingly routine atmospheric measurement, which turned out to be a vital sign of the Earth’s climate and led to the first alert to the world about the anthropogenic contribution to the “greenhouse effect” and global warming (Keeling, 2008). Similarly, the long-term study at the Hubbard Brook Experimental Forest has demonstrated undiminishing scientific returns of routine water and ecological measurements that led to the discovery of “acid rain” in North America (Likens et al., 1972; Likens and Bormann, 1974). Long-term recording of the health of the soil – the CZ’s foundation, through monitoring its “blood pressure” (soil water potential), temperature, respiration, carbon content, and other potentially key signs of global land change, is essential to the sustainability of ecosystem services, continued food production, lasting air and water quality, and a balanced growth of human society with natural systems.

5 Summary and conclusion

The emerging interests in the CZ and the establishment of CZOs and alike in different parts of the world provide an excellent opportunity to advance the understanding and management of the most complex and heterogeneous biomaterials on the Earth surface. Some of the key approaches or steps for achieving significant advancements may include (but not limited to):

- Integrated systems approach that cross-fertilizes the principle of energy and mass conservation with the principle of entropy and information accumulation in the evolutionary, multiscale, and interdisciplinary studies of the complex CZ;

- Evolutionary mapping-monitoring-modeling framework that allows the development of adaptive strategy and systematic characterization of subsurface heterogeneity across scales;

- Formation of a global alliance that monitors, maps, and models the CZ in a coordinated manner to facilitate cross-site scientific comparisons, evolutionary treatment of complex systems, and deeper insights into the CZ.

Water-soil interaction is a key to understanding the Earth’s CZ. However, while the soil is at the central junction of various interacting compartments in the CZ, the CZ as a whole is much broader than just soils. A broadened perspective is therefore needed to propel soil science forward and to better integrate its knowledge base with other bio- and geosciences.

Hydropedology is an important contributor to CZ science through addressing the interface between the pedosphere and the hydrosphere. Hydropedology is concerned with how the subsurface heterogeneity develops and evolves, how soil architecture impacts flow and transport in the field, how soil distribution pattern influences hillslope and watershed hydrology, and how the hydrologic cycle feedbacks to pedogenesis and soil functions. The fundamental questions of hydropedology call for quantitative connection between complex soil architecture and soil hydrologic functions across scales and an improved understanding of how hydrology may be used to quantify soil formation and evolution. The hydrologic cycle, together with human activities, have become prominent forces of changes in the soil and the CZ (at least at the human time scale), and thus deserve elevated attention in the integrated studies of the CZ.

Acknowledgements. The author acknowledges the support from the US National Science Foundation through the Shale Hills Critical Zone Observatory grant (EAR-0725019). Valuable comments on an earlier version of this manuscript were provided by Johan Bouma, David Chittleborough, Daniel Fritton, Chris Graham, and Jack Watson. John Maximilian Köhne is thanked for translating the paper of D. E. Tsakalotos. Peer review comments provided by anonymous reviewers (one not so anonymous, Harry Vereecken) and the handling editor Hans-Jörg Vogel helped improve the quality of this paper.

Edited by: H.-J. Vogel

References


H. Lin: Earth’s Critical Zone and hydopedology


USDA-NRCS: Field indicators of hydric soils in the United States (Version 4.0), USDA-NRCS, Ft. Worth, TX, USA, 1998.


