Computational Challenges for Mechanistic Modeling of Terrestrial Environments

Workshop Report (Draft v5: final)

Workshop held March 26-27, 2014.
DOE Headquarters, Germantown, Maryland

Organizing Committee

David Moulton
Los Alamos National Laboratory

Tim Scheibe
Pacific Northwest National Laboratory

Carl Steefel
Lawrence Berkeley National Laboratory

Members

Mike Heroux
Sandia National Laboratory

Stephen Long
University of Illinois, Urbana Champaign

Jonathan Lynch
Pennsylvania State University

Lois Curfman McInnes
Argonne National Laboratory

Peter Thornton
Oak Ridge National Laboratory

Writing Team

David Moulton
Los Alamos National Laboratory

Carl Steefel
Lawrence Berkeley National Laboratory

Tim Scheibe
Pacific Northwest National Laboratory

Scott Painter
Oak Ridge National Laboratory

Jonathan Lynch
Pennsylvania State University

Stephen Long
University of Illinois, Urbana Champaign

Mike Heroux
Sandia National Laboratories

Lois Curfman McInnes
Argonne National Laboratory

Eoin Brodie
Lawrence Berkeley National Laboratory

Dave Weston
Oak Ridge National Laboratory

Organizers

Climate and Environmental Sciences Division

David Lesmes
David.Lesmes@science.doe.gov

Paul Bayer
Paul.Bayer@science.doe.gov

Table of Contents

1. Executive Summary iii
2. Introduction 1
3. Motivation / Grand Challenges 6
3.1 Plant/Soil Systems 6
3.2 Biogeochemical Cycling 8
3.3 Watershed Hydrology and Ecosystem Dynamics 10
4. Research Opportunities 13
4.1 Models of the Coupled Plant-Soil System (Virtual Plant-Soil System) 13
4.2 Virtual Plot 18
4.3 Virtual Watersheds 20
5. Software Infrastructure for Virtual Ecosystems 23
5.1. Managing complexity in process-rich applications 23
5.2. Disruptive changes in computer hardware and software 25
5.3. Interdisciplinary Teams and Training 26
5.4. Software Lifecycles and Business Models 27
5.5. Leveraging new programs in software productivity 28
6. Concluding Remarks 30
Acknowledgements 32
Appendices 33
Appendix A. Agenda 33
Appendix B. Workshop Participants 35
Appendix C. References 36
Appendix D. Acroynms and Abbreviations 40

1. Executive Summary

The mission of the U.S. Department of Energy's Climate and Environmental Sciences Division within the Office of Science's Office of Biological and Environmental Research – 'to advance a robust predictive understanding of Earth's climate and environmental systems and to inform the development of sustainable solutions to the Nation's energy and environmental challenges' – is a formidable challenge that requires quantification of stocks and controls on states, fluxes and residence times of water, carbon, and other key elements through all components of the terrestrial system, including vegetation, soils, the deep vadose zone, groundwater, and surface water. To achieve this level of predictive understanding, a new generation of multiscale, multiphysics models is needed for terrestrial systems, models that take into account process couplings and feedbacks between the various "pools" (vegetation, soils, subsurface aquifers, surface waters) across wide ranges of spatial and temporal scales.

To explore the potential of a new generation of multiscale, multiphysics models to revolutionize our understanding of terrestrial ecosystem dynamics, the Office of Biological and Environmental Research (BER) held the *Computational Challenges for Mechanistic Modeling of Terrestrial Environments* workshop on March 26-27, 2014 at DOE Headquarters in Germantown, Maryland. That workshop brought together 28 researchers with diverse expertise, including environmental scientists, ecologists, plant scientists and computational scientists. Through a combination of invited talks, break-out sessions and report-back discussions, the workshop attendees identified the challenges and research opportunities for developing a more seamless and continuous framework for mechanistic modeling of terrestrial environments extending from bedrock to the boundary layer and from single-plant systems to crops to watersheds and river basins.

Three broad scientific challenges were identified. At the scale of individual plants, a more robust predictive capability requires greater focus on integrating, at the whole organism level, the rapidly developing mechanistic understanding of plant growth, form, function, and interactions with the surrounding biotic and abiotic environment in the soil and rhizosphere. At the scale of catchments and watersheds, the central challenge is to develop high-resolution watershed-scale models of the hydrologic, carbon, and nutrient cycles with tractable representations of the integrated vegetation-hydrologic-biogeochemical systems. Cutting across the disparate spatial and temporal scales is the challenge of improving the models for biogeochemical cycling, including more realistic descriptions of microbiological communities and their function, and better representations of the effects of small-scale processes and heterogeneities at field scales. These three individual challenges are interlinked, emphasizing the multiscale, interdisciplinary nature of the overarching challenge.

Three research opportunities were identified: a *virtual plant-soil* model that combines mechanistic models of plant growth, form, and function with high-resolution models of hydrological and biogeochemical processes in the surrounding soil and rhizosphere; a *virtual plot* model that combines multiple virtual plants or parameterizations of individual plants with detailed representations of belowground processes; and a *virtual watershed* that tracks fluxes and storage of water, energy, carbon, and nutrients on the surface and in a three-dimensional

subsurface, taking into account the effects of small-scale processes and heterogeneity and with more sophisticated and comprehensive treatments of hydrologic, elemental, and nutrient cycling in subsurface and surface waters mediated by vegetation and exchanges between the atmosphere and land surface.

Just as progress in experimental science depend on technological advances in instrumentation, so progress in computational science require advances in hardware, software, and algorithms. A new generation of multiscale, multiphysics models for terrestrial systems will require careful attention to software design, productivity tools, and programming models. Recent trends in hardware design raise uncertainty about programming models and the real performance of application codes. Moreover, software development tools and programming models have not kept pace with those changes in hardware, creating significant uncertainty for domain and computational scientists. This confluence of disruptive trends in computer hardware with the drive toward predictive multiscale simulations is putting immense pressure on the scientific community to find new ways to maintain its scientific productivity. However, those challenges also create opportunities, which led the Office of Advanced Scientific Computing Research (ASCR) to convene an interdisciplinary workshop on Software Productivity for Extreme-Scale Science in January 2014. An important finding of that productivity workshop was that significant improvements in development practices and interoperability of libraries could underpin a shift towards a more agile collection of high-quality composable components, ultimately maintaining or even enhancing productivity. This paradigm shift to an agile collection of interacting components, which was coined a “software ecosystem”, expresses the need to go beyond the modularity of traditional multiphysics codes to a higher-level of interoperability.

To meet BER’s programmatic goals as outlined in several recent documents (workshop reports, strategy document, and advisory reports), it will be necessary to adopt this paradigm shift within the BER scientific community. This presents an exciting opportunity to establish a new community approach to modeling and simulation in which multidisciplinary domain scientists work closely with computational scientists to develop interoperable modules that can be assembled in flexible configurations within a common framework, making it possible to simulate the multiscale structure and function of a variety of terrestrial environments. In our view, adopting this new approach is necessary to overcome the challenges associated with increasing model complexity and the disruptive effects of new computer architectures. The result will be a significant improvement in the scientific productivity of a BER research portfolio that is increasingly focused on predictive simulation tools as an integrative science outcome.

2. Introduction

The stated mission of the U.S. Department of Energy's Climate and Environmental Sciences Division within the Office of Science is 'to advance a robust predictive understanding of Earth's climate and environmental systems and to inform the development of sustainable solutions to the Nation's energy and environmental challenges'. This is a formidable challenge, however, which requires quantification of stocks and controls on states, fluxes and residence times of water, carbon, and other key elements through all components of the terrestrial system, including vegetation, soils, the deep vadose zone, groundwater, and surface water. To achieve this level of predictive understanding, a new generation of multiscale, multiphysics models are needed for terrestrial systems, models that take into account process couplings and feedbacks between the various "pools" (vegetation, soils, subsurface aquifers, surface waters). In addition, these models must be able to consider multiple spatial scales given that the region of interest is potentially large (we may want to understand the dynamics of an entire river system), while the critical interfaces within the larger watershed often require fine scale resolution of spatial gradients to calculate water, nutrient, and elemental fluxes correctly. Multiple timescales typically need to be resolved as well, since ultimately our interest is to be able to project ecosystem dynamics over longer climate-related time scales while resolving some variables like temperature (especially in soils and vegetation) at time scales as short as the diurnal.

From prehistoric times, humans have observed their environment and developed knowledge of correlative and/or cause-and-effect relationships. Over time, this developed into the first pillar of science, experimental observation, in which humans manipulated conditions and observed outcomes to gain empirical knowledge. As the scientific method matured, a second pillar of science developed, in which humans generalized their empirical knowledge in terms of mathematical models and theories. Recently, a third pillar has emerged (e.g., Oden and Ghattas, 2014) in which computational numerical methods are used to simulate complex interacting processes in a manner that integrates theory and experimental observation, often with a level of process detail that is far beyond what can be achieved with purely analytical mathematical treatments (Steefel et al., 2005).

Numerical modeling extends classical scientific approaches in a number of ways. Numerical simulations can be used to explore system behaviors in situations where experiments and observations may be too controversial (e.g., drug discovery), too hazardous (e.g., natural disasters), prohibited by law or treaty (e.g., nuclear detonations), difficult to instrument (e.g. subsurface, ocean or space environments), expensive (e.g., high-energy physics), or requiring too much time to observe (e.g., climate change) (Keyes, 2012). Scientific computation can reconstruct past events and utilize inverse methods to infer the past causes of observed phenomena, and can extend experimental results to other scenarios in lieu of additional experiments (Oden and Ghattas, 2014). It can also be used to evaluate the consequences of both parametric uncertainty (e.g., uncertainty quantification and/or sensitivity analysis) and model

inadequacy (e.g., multimodel assessment). When used appropriately, modeling can be used to extend fundamental scientific discoveries to larger scale natural systems, for example, those characterizing the Earth's crust (Steefel et al., 2005).

The use of modeling to augment the understanding produced by the traditional empirical and theoretical perspectives is particularly valuable for untangling complex interactions in environmental systems (e.g. Larsen et al. 2014), with perhaps the greatest potential being in advancing our understanding of the various interacting processes—hydrology, biogeochemistry, microbiology, vegetation dynamics—that operate at the Earth's surface. Collectively, those *critical zone* processes mediate the interaction between the Earth's surface and the atmosphere, thus controlling a wide range of environmental services and processes of interest to society including water supply and quality, and the terrestrial carbon cycle and its effect on atmospheric CO_2. For example, the quality and redox status of groundwater aquifers, and the flux of constituents from these aquifers into surface waters (rivers and lakes), and in some situations back into the atmosphere, are the result of a complex set of interactions between hydrologically induced nutrient fluxes, reactive mineral phases in the subsurface, and the resident microbial communities. The subsurface in turn is connected to the atmosphere through the relatively thin veneer we call soil, with exchanges (both water and gas) mediated in large part by vegetation. Vegetation has a large impact on hydrology at the watershed scale through transpiration, while the hydrology itself influences or controls which vegetation type dominates, or whether plants survive at all. All of these complex process couplings and feedbacks operate at multiple scales ranging from the watershed to the individual plant and its root zone (Figure 1).

Although steady progress has been made in mechanistic modeling and simulation of key terrestrial processes in isolation, the potential for using numerical simulations to couple those processes and provide a more complete picture of the critical zone functioning has not been realized. This is especially significant in the context of predictive simulation of performance of new bioenergy crops and effects of future climate change because it renders calibration-based approaches questionable and provides great impetus for mechanistic representations. Part of the challenge is that coupled physical, chemical and biological processes take place in an environment that is highly heterogeneous across a wide range of spatial scales, and may give rise to emergent phenomena that are not readily predicted on the basis of concatenation of individual processes alone. For example, incorporating subgrid mechanisms and properties in a way that adequately accounts for local mixing of reacting agents (or lack thereof) is a major challenge in reactive transport modeling (Steefel et al, 2005). Our current understanding of mechanisms that determine interscale relationships, and theory for how numerical models should best represent the poorly resolved multiscale processes, is still rudimentary. In addition, technical challenges associated with managing model complexity in simulations that combine many process models, and looming disruptive changes in computer hardware combine to make those advances in simulation far from inevitable.

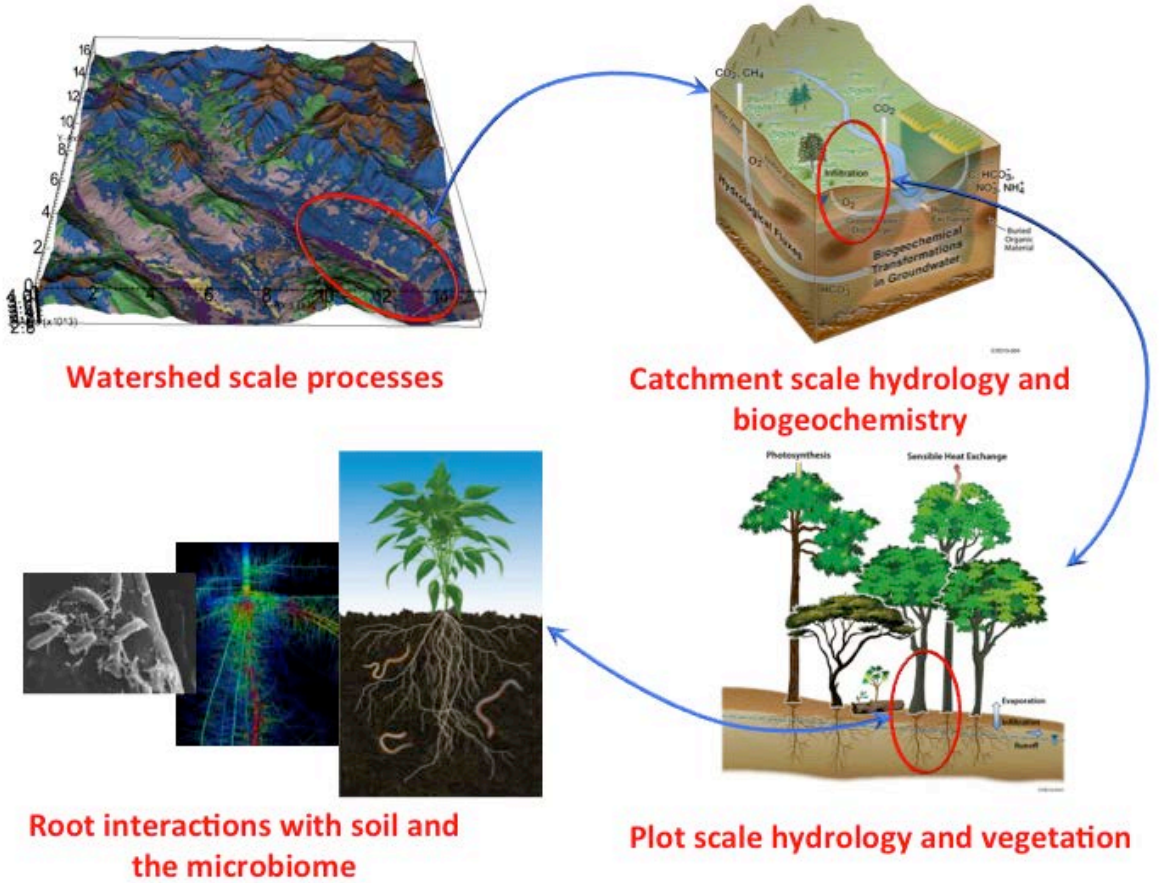

Figure 1: The large range of scales in terrestrial ecosystems, from the watershed, to the individual hydrologic catchment, to the plot scale, and finally to the individual plant/tree or root system scale, call for a new terrestrial modeling framework that can account for the feedbacks between surface and subsurface hydrology, vegetation, soils, and the atmosphere. The Grand Challenge is to capture the larger system behavior while representing accurately the smaller scale fluxes, processes, and system features in highly heterogeneous terrestrial systems.

Fortunately, those challenges create an exciting opportunity to establish a new community approach to modeling and simulation, where hydrologists, earth and environmental scientists, ecologists, plant scientists, and microbiologists work together with computational scientists to develop interoperable modules that can be be assembled in flexible configurations within a common framework to simulate the multiscale structure and function of a variety of terrestrial environments.

The Environmental System Science (ESS) program within Climate and Environmental Sciences Division (CESD), which encompasses the Terrestrial Ecosystems Science and Subsurface Biogeochemical Research (SBR) programs, has the stated goal of "Advancing a ***robust predictive understanding*** of terrestrial ecosystems from 'bedrock to atmosphere' and from global to molecular scales using an iterative approach to model-drive experimentation and

observation." (emphasis added). The program's science strategy is intended to address predictive modeling of "systems of systems", in which system complexity and diversity increases with the scale of consideration (Figure 2). The focus on robust predictive understanding emphasizes the need for incorporating mechanistic understanding from small scales (molecular, pore, plant, plot) into predictions at larger scales (watershed, basin, global). Figure 2 shows major BER field research efforts that span the range of scales of interest, and serve as both motivation for and potential use cases for development of a new community modeling framework. In contrast to this vision, the current modeling landscape comprises the application of many distinct and unconnected models, each typically at a single scale, informed by limited data describing complex systems with multiscale heterogeneity. It has been noted by a recent BER Advisory Committee report that the "…fragmentation of science, technologies, and predictive capabilities among and within disciplines and the focus on studying mostly individual, scale-based system components…leads to fundamental uncertainties about how coupled subsystems interact with each other and respond to environmental changes across different space and time scales. The lack of sufficient science-based capabilities to predict these interactions and responses hinders the ability to sustainably manage and mitigate energy and environmental problems" (BERAC, 2013).

To explore the vision of a multiscale, multiphysics community framework for modeling terrestrial systems relevant to bioenergy applications and climate science, the Office of Biological and Environmental Research (BER) held the *Computational Challenges for Mechanistic Modeling of Terrestrial Environments* workshop on March 26-27, 2014 at DOE Headquarters in Germantown, Maryland. That workshop brought together 28 researchers with diverse expertise, including hydrologists, environmental scientists, ecologists, microbiologists, plant scientists and computational scientists. A combination of invited talks on key fields, as well as lightening talks on specific research topics or simulation codes, was used to set the stage for three parallel breakout sessions. The breakouts were organized with common themes: (1) Science Challenges and Opportunities, (2) Multiscale Frameworks for Mechanistic Modeling, and (3) Prioritization of Research Needs. Through the workshop sessions and report-back discussions, the workshop participants identified the challenges and research opportunities for developing a more seamless and continuous framework for mechanistic modeling of terrestrial environments extending from bedrock to the boundary layer and from single plant systems to crops to watersheds and river basins. This report summarizes the workshop findings.

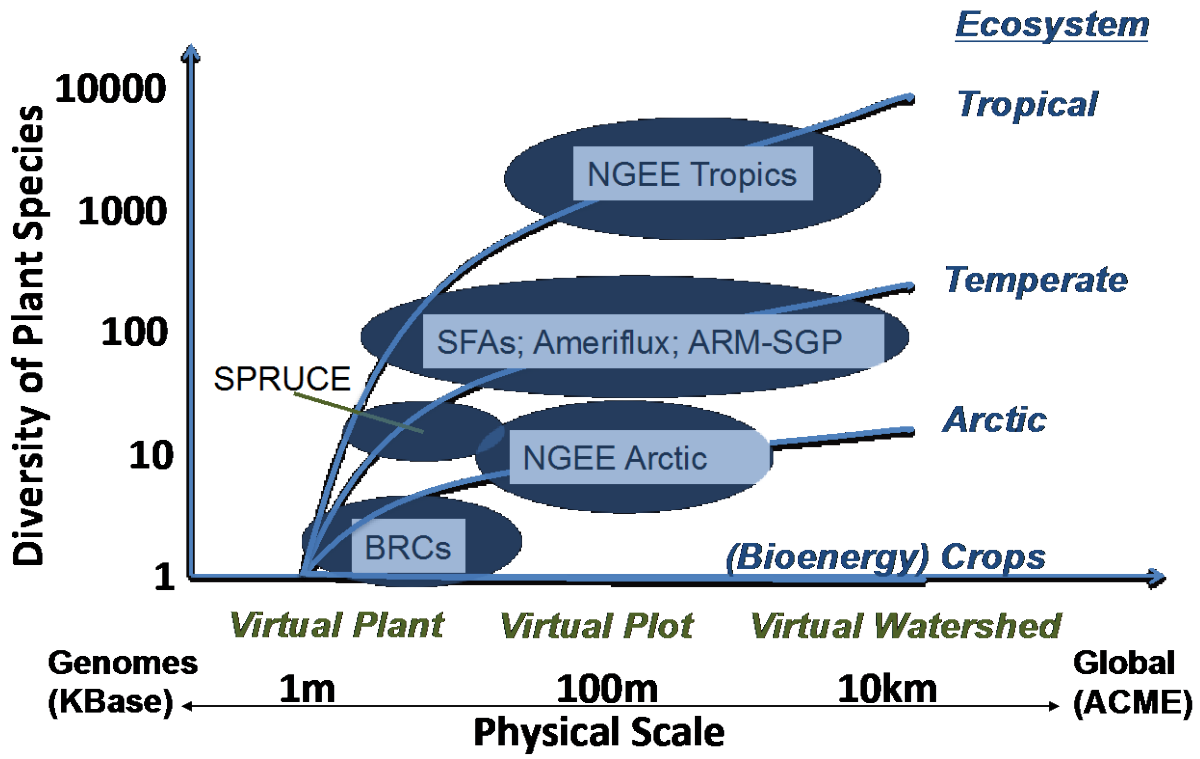

Figure 2: BER field research activities extend across a wide range of ecosystems and scales (plants to plots to watersheds). These well-characterized and highly instrumented field "laboratories" enable multidisciplinary teams of scientists to test and advance fundamental understanding of key processes and interactions in the context of the overall system behavior. The research conducted at these sites is informed by and informs BER research efforts at both smaller scales (e.g., KBase:Systems Biology Knowledge Base) and larger scales (e.g., ACME: Accelerated Climate Modeling for Energy). The need for a more seamless modeling framework that enables the interoperability of multiscale-multiphysics modeling components, has been identified as a priority by several previous BER workshops and reports. The field research activities in this figure are supported by the following BER projects: BRCs = Bioenergy Research Centers; NGEE = Next-Generation Ecosystem Experiment; SFAs = Scientific Focus Areas (national laboratory research projects); SPRUCE = Spruce and Peatland Responses Under Climatic and Environmental Change; ARM-SGP = Atmospheric Radiation Measurement - Southern Great Plains site; Ameriflux = American flux measurement network.

3. Motivation / Grand Challenges

3.1 Plant/Soil Systems

The 21st century challenge of sustaining more than 10 billion people with a declining natural resource base and a changing climate creates a great need for crops and crop production systems with greater productivity, resource efficiency, resilience, and stress tolerance. For most of human history the improvement of food crops has relied on empirical selection for obvious phenotypic traits such as yield and quality. While our contemporary food crops represent centuries of selection and a huge investment over the past 50 years, crops for sustainable bioenergy production are barely different from the wild plants from which they were selected. Scientific advances have now made it possible to develop plants with entirely new traits or with novel combinations of traits by direct manipulation of the plant genome.

An overarching challenge is the development of predictive understanding of the plant phenome including several processes that are essential to food and bioenergy crop performance, which is needed to guide manipulation of the plant genome to meet productivity and efficiency objectives. Underlying this challenge are several critical scientific questions:

- What controls tolerance to environmental stresses including drought, heat, and low soil fertility?
- How are plant resources partitioned to diverse processes and organs in time and space, including respiration, allocation to roots and rhizosphere, and reproduction/yield?
- How do root traits and processes control the acquisition of water and nutrients, and rhizodeposition of C?
- How does the plant microbiome, including rhizosphere communities, mycorrhizal symbioses, and endophytes contribute to plant performance?

As the primary component of terrestrial ecosystems, plants also mediate the exchange of mass and energy between the Earth's surface and the atmosphere. A scientific challenge that is equally important to the crop development challenge is to better understand the role of vegetation in the Earth system and how it will change in a changing climate. The representation of vegetation in current land-surface and watershed models is greatly simplified and based largely on empirical and phenomenological information. They therefore fail to match the mechanistic strength of models of physicochemical processes in the soil and atmosphere, yet represent a key link between these two parts of the Earth System. Plant feedstocks (biofuels) are a promising alternative to fossil fuels, but the development of a bioenergy-focused agriculture will lead to significant changes in the plant-soil-microbe ecosystem, with further complexities introduced by climate change and variability (BER, 2013). Although empirical and semi-empirical models may be reliably calibrated in current climates and current ecosystem/crop compositions, the lack of a mechanistic basis introduces great uncertainty in projecting plant response to a changing crop composition and climatic environment.

The use of empirical representations of vegetation in land-surface and watershed models and the optimization of crops and crop systems have not taken full advantage of significant advances in mechanistic understanding of plant function, growth, and form. Among the improved mechanistic models that have been developed (see Figure 3) are models for the complete photosynthetic process from biochemistry (Zhu et al., 2013, Wang et al., 2014) to the plant canopy (Song et al., 2013), growth regulator fluxes in development (Steinacher et al., 2012, Bennett et al., 2014), shoot patterning (Domagalska & Leyser, 2011), flowering (Song et al., 2012), root structural and functional dynamics (Lynch et al, 1997, Lynch, 2013, Dyson et al., 2014), and linkage to gene regulatory networks (Hill et al., 2013, Chew et al., 2014).

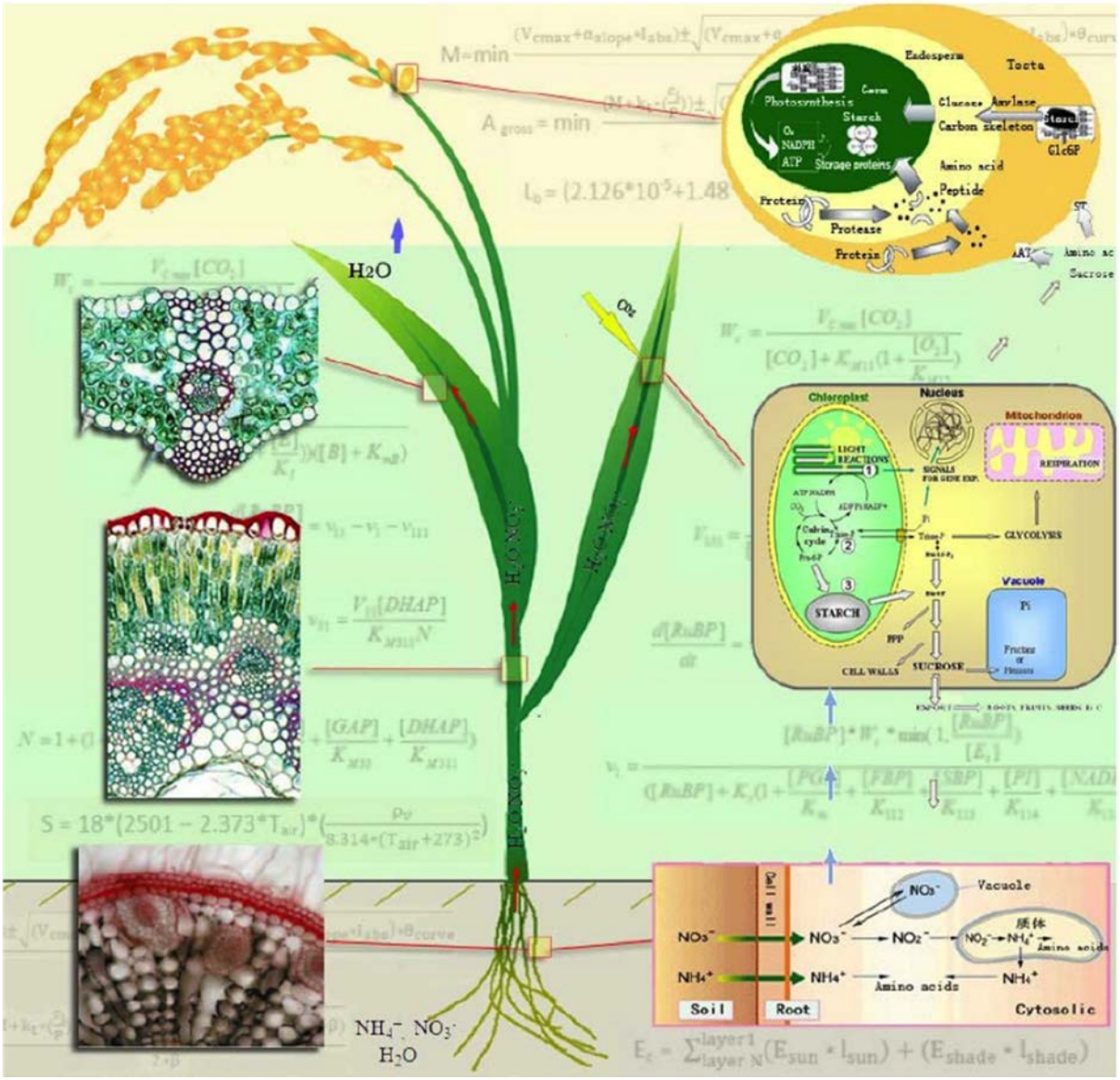

Figure 3: Components of a mechanistic rice production model building from the biochemistry and biophysics of photosynthesis, compartmentation of processes between cell and tissue types, through to use of photosynthate and nutrient in root and grain development (Zhu *et al.*, 2011). The complex process interactions between the plants, the soil and its resident microbial communities, and the atmosphere indicate the need for a new modeling framework at this critical interface.

Emerging efforts to develop mechanistic models have largely been confined to single laboratories or consortia, and often rely on heritage code. Effort is thus dispersed and often duplicated. The resulting disparity in algorithms and software design modalities and software development practices creates formidable barriers to integration. Moreover, mechanistic models of plant function at molecular and cellular scales have not been successfully integrated at the whole organism scale, much less at the level of plant interaction with microbes and the rhizosphere. The challenge, then, is to effectively integrate the rapidly developing mechanistic models at distinct scales into a virtual plant model, that then has the capability of linking with models of the abiotic and biotic environment. Not only will this provide a much stronger plant component to bioenergy crop and natural ecosystems, but will also allow the application of optimization algorithms to identify the more resource efficient ideotypes to guide breeding of emerging sustainable bioenergy crops (e.g. Drewry et al. 2014).

3.2 Biogeochemical Cycling

Soils and subsoils are the largest repository of organic carbon in the terrestrial environment, with the top three meters estimated to contain about 2344 Pg of C, more than the atmosphere and plants combined (Jobbágy and Jackson, 2000). Subsoils, although they contain lower concentrations than surface soils, nevertheless also comprise a significant pool of organic carbon. Changes in the water cycle and climate will significantly impact existing soil and subsoil pools, carbon dynamics, and carbon and elemental cycling, but it remains an open and pressing question what the nature of that influence will be as a function of ecosystem type, and in turn how those changes will affect ecosystem and crop system function and productivity, water distribution and fluxes, and fluxes of greenhouse gases to the atmosphere.

An overarching challenge is predicting carbon cycle responses to changing climate and the role of belowground processes in modulating climate change impacts on ecosystems and bioenergy crop systems. Underlying this challenge are several critical scientific questions:

- How will soil carbon stores and carbon fluxes between soils and the atmosphere and surface water bodies change in a changing climate?
- How do soil and rhizosphere processes including microbial processes influence plant function, productivity, and dynamics, and will those controls change in a changing climate?
- How will land use changes, long-term climate trends, and changes in frequency, duration and magnitude of extreme climate events affect nutrient export from watersheds?
- How do microbial community dynamics change as a result of climate stresses and watershed-wide biogeochemical cycling, and to what extent can we use metagenomics and functional response models (soil community, litter community, root-associated community) to improve the representations of these dynamics?
- How do macropores and associated preferential flow created by the effects of roots, bioturbation, freeze/thaw cycles, swelling/shrinking clays, and other mechanical processes affect carbon and nutrient transport?

Subsurface flow and reactive transport models have developed a fairly high level of sophistication, particularly at pore, laboratory, and local field scales, but connecting process understanding across these scales and to larger scales (watershed to regional and global scales) remains a significant gap (Figure 4). Broadly stated, the critical technical challenges are associated with how to best measure and model the ecosystem-scale manifestations of complex coupled processes that are controlled by highly heterogeneous local environmental conditions: 1) analysis of the degree of complexity / parsimony that is needed for robust predictive understanding, 2) development of specific algorithms for upscaling or coupling nested model scales, and 3) assimilation of diverse data and validation against observations across a range of scales.

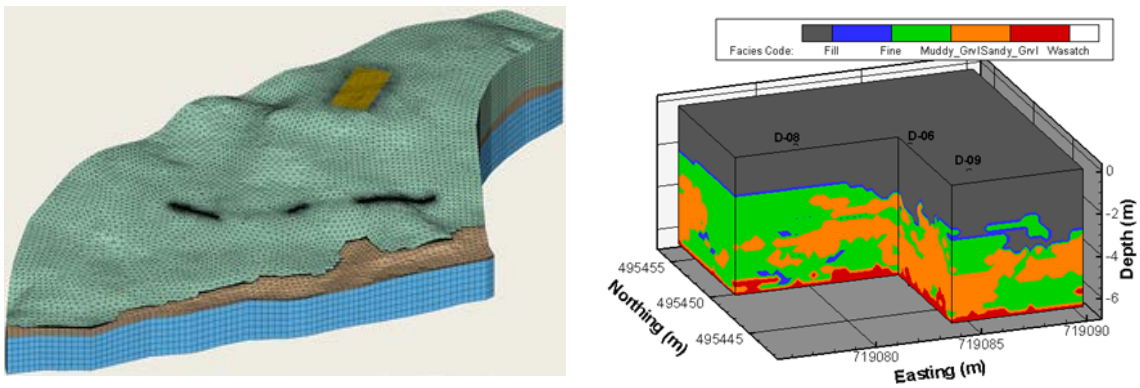

Figure 4: Left: Unstructured grid used to simulate groundwater flow at the F-Area Seepage Basins in Savannah River is able to capture both topography and stratigraphy at very high resolution (figure courtesy Terry Miller, LANL). Right: Distribution of reactive facies in a section of Rifle alluvial floodplain aquifer is used to simulate uranium migration during a biostimulation experiment (Yabusaki et al, 2010).

Moreover, while continuum subsurface flow and reactive transport models have reached a new level of maturity (Steefel et al, 2014), the incorporation of realistic descriptions of microbiological communities and their function into these models is still only in its first stages (Bouskill et al, 2013). Treatments of the role of microbes in mediating subsurface reaction rates have typically assumed steady-state conditions, or microbial communities were represented simply as non-competing communities that grew according to a fixed growth yield. The dynamic treatment of microbial community composition and function, including the partitioning between catabolic (energy) and anabolic (biomass growth/enzyme synthesis) pathways, leads to complex biogeochemical reaction networks in which electron acceptors, donors, and the microbial community itself need to be considered as evolving quantities. The challenge now is to incorporate the diversity of microbial communities known to exist in the subsurface, and to relate community composition to function (reaction rates) by developing physiological models tied to environmental conditions (Figure 5).

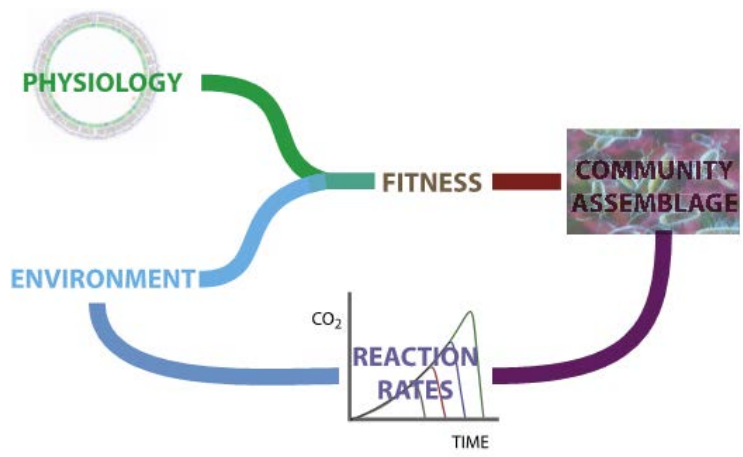

Figure 5: Schematic figure showing coupling between the local geochemical environment and microbial physiology responsible for determining the community assemblage. The coupled nature of the microbial processes and the local biogeochemistry suggests the need for a coupled modeling approach (Figure courtesy Nicholas Bouskill, LBNL, adapted from Mick Follows, http://darwinproject.mit.edu/).

In addition to the need to include more sophisticated treatments of the interactions and couplings between microbial community composition/function and geochemistry, a number of computational issues cross-cut the above process modeling challenges. Model testing and validation requires development of linkages to experimental systems and existing databases for establishment of community benchmark problems. Quantification of model uncertainty and how uncertainties propagate across scales is a fundamental element of developing the mechanistic underpinning for predictive models and requires significant attention. Efficient and reliable simulation of systems of the complexity described above requires development of tools for automated hierarchical and unit testing, specification of interfaces among components (and associated data models), data management and provenance capture, model setup and execution, and analysis and visualization of complex model outputs.

3.3 Watershed Hydrology and Ecosystem Dynamics

Watersheds and their associated ecosystems provide a variety of critical climate regulating and hydrologic services including biosequestration of carbon, regulation of surface energy balances, supply of water for household use and to support energy production, mitigation of extreme precipitation events, and protection of water quality. The hydrologic behavior of a particular watershed depends not only on the physical features of topography and subsurface permeability, which require appropriate representation and spatial resolution in models (Wood et al., 2011), but also on the nature and state of the associated ecosystem.For example, plant root uptake and resulting transpiration is a dominant component of the water cycle in many watershed ecosystems (Brutseart, 1988; Voepel et al. 2011). Terrestrial ecosystem composition, health, and climate regulation function depend, in turn, on available water. The water cycle and ecosystem

function are thus tightly coupled and must be considered as an integrated, multiscale system if the terrestrial carbon cycle and the impact of human activity and climate change on watershed services are to be better understood.

The scientific community has limited ability to project watershed response under novel climate conditions using current watershed models because many important processes associated with plant-soil hydrologic interactions are represented with highly parameterized (reduced complexity) models at the watershed scale (e.g. Jarvis, 1989). Although parameterized representations of coupled plant-soil hydrology can be reliably calibrated in current climate conditions, the lack of a strong mechanistic foundation brings into question their predictive capability in future climate conditions that lack a present-day analog ("no analog future", Williams, 2007). More generally, calibrated models may not provide accurate predictions of system response under any conditions that differ from those the model was calibrated to (e.g., modified land use and crop patterns, changes in groundwater pumping rates or river flow regulation patterns). Although the enormous complexity and ubiquitous multiscale heterogeneity of natural systems will likely limit the degree of mechanistic detail that can be realistically accommodated in watershed-scale models, making calibration of effective parameters necessary, increased mechanistic detail in the process representations may still result in a more robust predictive capability by guiding selection of model structure and otherwise providing insights to constrain calibrated watershed scale models.

One limitation of reduced complexity models calibrated in current climates is that they may not adequately represent plant transpiration in future drought conditions when flow from dry soil to the plant roots becomes the rate-limiting process. A second concern is that models based on calibrated empirical parameters implicitly presume that ecosystem composition is static. Without a more mechanistic representation, changes in ecosystem composition caused by changing climate will be difficult to include in projections of watershed response. A third limitation is that the current generation of models has limited capability to address geochemical cycling at the watershed scale. Improved models of geochemical transport and reaction at the watershed scale, including interaction between surface and subsurface, are needed to better understand how ecosystem function and productivity and fluxes of greenhouse gases to the atmosphere will change in a changing climate. Such improvements are also needed to better understand nutrient export from watersheds and the resulting eutrophication of downstream surface water bodies. Finally, watershed models generally lack the capability to represent surface water temperatures, which is important both for its control of geochemical cycling and energy balance and for its impact on downstream energy production.

To meet BER's goal of a robust predictive capability for terrestrial systems, high-resolution watershed-scale models with robust representations of the integrated vegetation-hydrologic-geochemical systems are thus needed. Very high-resolution models have been identified as a key research need in the global land-surface modeling context (Wood et al. 2011), and similar research needs exist for individual watersheds. Such a capability must go beyond the 3-D

variably saturated subsurface flow models coupled to surface flow based on high resolution digital elevation maps (Figure 6) to include nutrient transport and associated biogeochemical reactions, thermal processes, and more robust mechanistic models of plant interaction with the environment, vegetation dynamics, and vegetation diversity. Because terrestrial plant productivity is almost universally limited by edaphic constraints, including suboptimal availability of water and nutrients, soil acidity, alkalinity, and physical barriers to root growth, better representation of the soil/plant interface is a key consideration. This is intrinsically a multiscale system because the natural scales for representing plant-soil interactions are much smaller than a typical grid cell in a whole-watershed model. It is also intrinsically a multiphysics system involving many coupled processes of varying complexity. It is important to note that the level of complexity in this multiphysics, multiscale challenge is unprecedented in terrestrial systems modeling, which has typically coupled a small number of process models in predefined coupling modes. The central challenges are to 1) develop tractable watershed-scale representations or simulation approaches for smaller scale processes such as plant growth and function, and mass transfer between soil and roots, including hydraulic redistribution and nutrient uptake; and 2) effectively manage the unprecedented level of model complexity associated with coupled high-fidelity models of hydrology, biogeochemical cycling, vegetation interaction with the soil, and ecosystem dynamics.

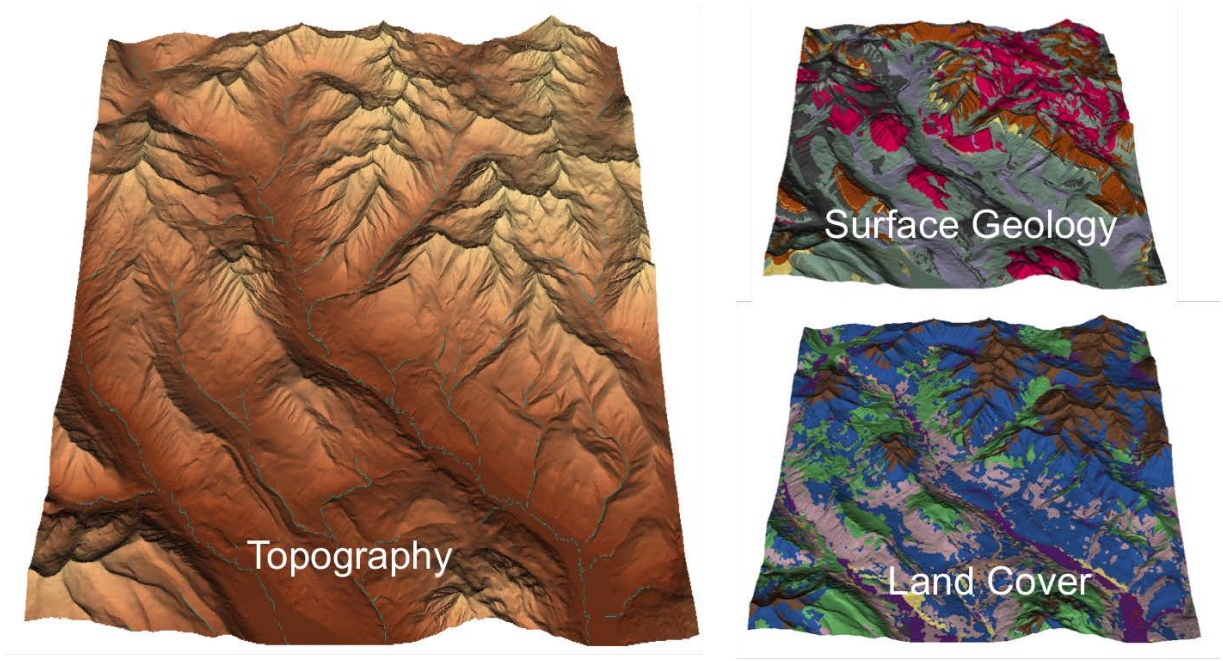

Figure 6: Topography based on Digital Elevation Maps (DEM) from the East River Catchment, Gunnison County Colorado, with rock type (surface geology) and land cover (vegetation) shown at similar resolution. A typical analysis might begin with the high-resolution topography used to develop coupled surface-subsurface hydrologic flow models. The flow models are then expanded by incorporating soil and vegetation distributions, which then interact with the atmosphere (figures courtesy Christine Pribulick and Reed Maxwell, Colorado School of Mines).

4. Research Opportunities

Advances in process-level understanding across multiple disciplines, software design, and computing hardware performance create the opportunity for a transformative multiscale, multiphysics simulation framework that uses high-resolution, high-fidelity representations of plants and plant/soil interaction at the small scale to inform watershed-scale representations of hydrology, ecosystem dynamics, carbon and nutrient cycling. It is envisioned that these models will be informed at all scales by plant genomics (highlighted in the see side bar on Page 14) and microbiome genomics (expanded in the side bar on Page 15).

Three characteristic spatial scales are identified for the simulation framework. At the smallest scales, a framework that couples a representation of an entire plant including roots with high-resolution representations of processes in the surrounding soil could be used to study interactions of plants with the rhizosphere and surrounding soil (Virtual Plant). The next level of complexity would include multiple virtual plants in virtual soil for the purposes of better understanding interactions among plants (Virtual Plot). Finally, results of the Virtual Plant and Virtual Plot models would inform next-generation watershed models (Virtual Watershed).

Cutting across the three spatial scales are two science issues: incorporating microbiome genomic information into reactive transport models to better represent soil microbial controls on carbon and nutrient cycling, and how to transfer information across scales. Discussions of those two issues are integrated into the three scale-focused subsections below. Two cross-cutting technical issues - software design to help manage model complexity in models that combine large numbers of process representations, and how to design for future architectures for computing hardware - are discussed in Section 5.

4.1 Models of the Coupled Plant-Soil System (Virtual Plant-Soil System)

Advances in plant genomics, modeling of plant growth and function, computing power, and software design create the opportunity to build a virtual representation of plant growth, development, productivity and interactions that is faithful to the current understanding of the underlying mechanisms. In other words, the opportunity exists to grow entire plants and study their interactions with the environment *in silico*. It is envisioned that the model would go beyond a virtual plant representation to also include a mechanistic representation of transport, biogeochemical, and microbial processes in the surrounding soil. The mechanistic nature of such a Virtual Plant-Soil (VPS) system would allow integration of gene function to cellular, organismic, and ecosystem scales, thereby leveraging DOE's efforts in plant and microbiome genomics. The higher fidelity representations enabled by a VPS framework would represent a reliable approach to prediction beyond the empirical experience base but also provide a better basis for interpretation of new observations.

A virtual plant-soil system framework would represent an important transformational shift in the development of mathematical models of plants and vegetation, from disparate and largely empirical models to an integrated community of mechanistic models that interact to capture emergent properties of plant systems across a range of scales, from the genome to plant

communities. It is envisioned that the VPS would provide a sustainable and flexibly configured framework to integrate models of key processes, providing a shared community platform to facilitate incorporation, testing and curation of new process models, with the goal of continually improving models of the plant-soil system. No one monolithic model could capture the relevant biology across the range of spatiotemporal scales needed. Instead, this framework is envisioned to be a collection of process models that are implemented as interoperable components interacting with each other as needed to resolve specific research questions. Realizing such a framework will require interoperable libraries, toolkits and frameworks, data and software standards (Section 5); most importantly, it must become an organized community of scholars skilled in the relevant domain expertise as well as in the coordination and curation of modern software ecosystems.

A VPS framework would have multiple applications. For example, it could be used to more reliably predict plant response to a changing climate, which will help improve representations of the terrestrial carbon cycle in Earth System Models (ESMs). It would also help improve the representation of ecosystems in watershed models (see Section 4.2), therein improving the identification and mitigation of potential climate-induced threats to water supply and water quality. This new capability would also be useful for interpreting remote sensing and other data streams. For example, spectroscopic remote sensing (otherwise known as “hyperspectral” remote sensing) of plants and regions has become routine, yet the connection between molecular mechanisms affecting leaf and canopy properties and spectroscopic remote sensing is poorly understood. Representation of the full photosynthetic process in plant canopies would allow prediction of spectral changes in different environmental conditions, thus informing remote sensing of the key spectral shifts that can be expected and understood in terms of mechanism at the plant level. This framework would also be an invaluable tool in elucidating the rhizosphere, the zone of soil around roots, that is central to understanding and managing the C cycle in soil and nutrient transformations that determine water quality as well as ecosystem productivity.

A key opportunity for the virtual plant-soil system concept is to enable the efficient integration of models with data streams from plant sequencing, genomics and high-throughput phenotyping programs. This approach could provide a novel and powerful analytical framework for understanding gene function at the organismic scale in diverse phenotypic and environmental contexts. Indeed, one important potential application of this framework would be to improve understanding of how genetic variation translates to system behavior at greater scales. This application highlights the need for the underlying models to be sufficiently mechanistic to accurately capture the emergent properties of the plant-soil system.

A VPS framework could have both predictive and heuristic utility. When existing models are sufficiently robust, as is the case for example with canopy photosynthesis or plant hydraulics, this framework can be employed to address broader research questions, and as an element of models at greater scale. In cases where the relevant biology is poorly understood, as is the case for example with plant soil interactions, rhizosphere dynamics, or multicellular signaling networks, this framework can function as a heuristic research tool to explore poorly understood processes and guide empirical research.

SIDE NOTE:
The Virtual-Plant Soil Framework: Integrating Phenotype Data with Genomic Predictions

An important aspect of the Virtual Plant-Soil (VPS) system is the transition from largely empirical models to a group of coupled mechanistic models that capture complex system behavior across spatial and temporal scales. At the organismic level, this virtual system will provide a much-needed framework for investigating gene and pathway function, as well as *in silico* simulations for novel root and shoot phenotypes and environmental combinations useful for bioenergy feedstock design under future climate scenarios. The VSP also will enable assessment of the vast array of genetic diversity present within and among plant populations and the consequences of unique allelic combinations on functional phenotypes. Such a strategy will provide a critical intermediate step in understanding genotype to phenotype associations by providing possible mechanisms.

A similar strategy, especially for dominate keystone species in more natural ecosystems, may become possible as DNA sequencing, whole-plant physiology, phenotype trait databases, and modeling efforts continue to develop. *Populus*, for example, is a recognized bioenergy feedstock, yet it exists in vast ranges as a keystone genus and species throughout North America, contributing substantially to regional carbon and nutrient cycling. Currently, a major limitation in predicting terrestrial carbon fluxes is inaccurate representation of plant photosynthesis and growth in land models due to uncertainty in parameter values for these characteristics. The VPS framework (Figure 7) presents a unique opportunity to reduce this uncertainty by using genomic predictions for keystone species and extension to additional species using community-based stored and measured phenotype values (e.g., LeafWeb, www.leafweb.ornl.gov; TRY, www.try-db.org). The scaling of nutrient and carbon fluxes from organismic to plot and regional levels, remains a critical challenge. However, using sophisticated sequencing data to assess genetic variation—and the predicted consequences of that variation on functional phenotypes—through the Virtual Plant-Soil system modeling framework, holds great potential.

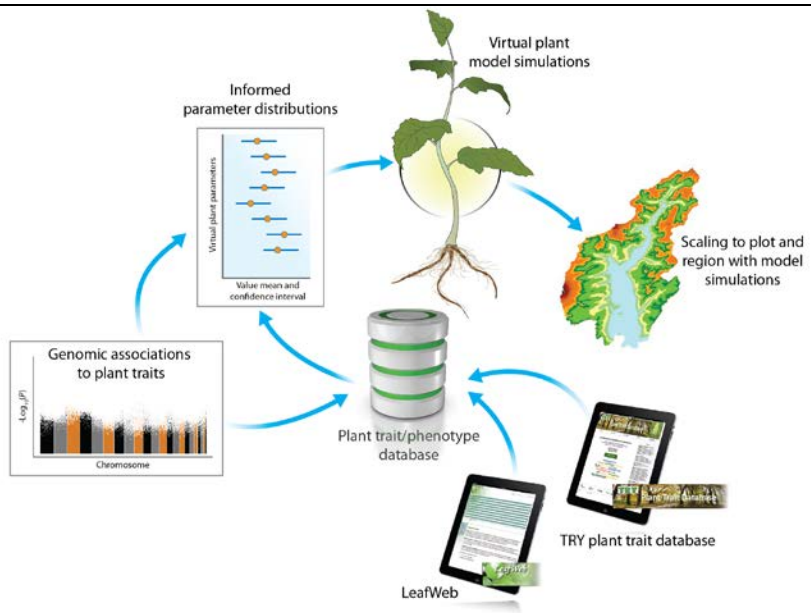

Figure 7: Schematic illustrating the integration of community-based phenotype/trait measurements and genomic predictions to constrain the uncertainty surrounding virtual plant model parameters.

SIDE NOTE: Microbial Genomic and Metagenomic Information

An important opportunity is associated with the coupling of microbial community composition and function in the plant-soil system. Plants exert a strong selection on the growth and activity of soil microorganisms, through factors such as rhizodeposition and litter composition. In turn, microorganisms play a key role in the mobilization or fixation of nutrients that regulate plant productivity and species distributions, and are key regulators of the fate of carbon in soils. The intrinsic linkage of these two system components in nature should be reflected in their treatment in modeling frameworks. Yet significant challenges remain. The complexity of soil microbiota is vast and key ecosystem/biogeochemical functions are associated with microorganisms that are currently unknown to biology. However with recent technological advances these organisms can be identified and profiled in terms of their metabolic potential using genome-resolved metagenomics, combined with functional information. The microbial genome properties can inform key model parameters such as growth rate and temperature adaption; genomic information from selected sites can be catalogued to build thermodynamic biogeochemical-microbial reaction networks (Figure 8). In time, through resources such as the DOE Systems Biology KnowledgeBase (KBase), the assimilation of microbial information from multiple ecosystems could lead to generalizable distributions of microbial functional traits and their genomic linkages. This would enable the innovative treatment of the microbial communities as fully dynamic components of the coupled plant-soil system that ultimately exert an important control on carbon and nutrient transformation, the speciation of redox sensitive constituents, and their sequestration within, or mobilization from, the soil, aquifer or watershed systems. Complementary ecological modeling approaches can add value through characterization of ecological factors that govern microbial community composition, namely 1) selection (growth constrained by environmental conditions); 2) dispersal (movement or transport of organisms); and 3) drift (random or stochastic growth and decline of populations). Recent work (e.g., Stegen et al., 2013) has defined new ways of distinguishing the degree of control these ecological factors have on a given community and relating them to hydrologic and biogeochemical processes. Quantitative ecological analysis provides tools for mapping ecological factors based on observed measurements, and then relating them to model properties such as geological facies type.

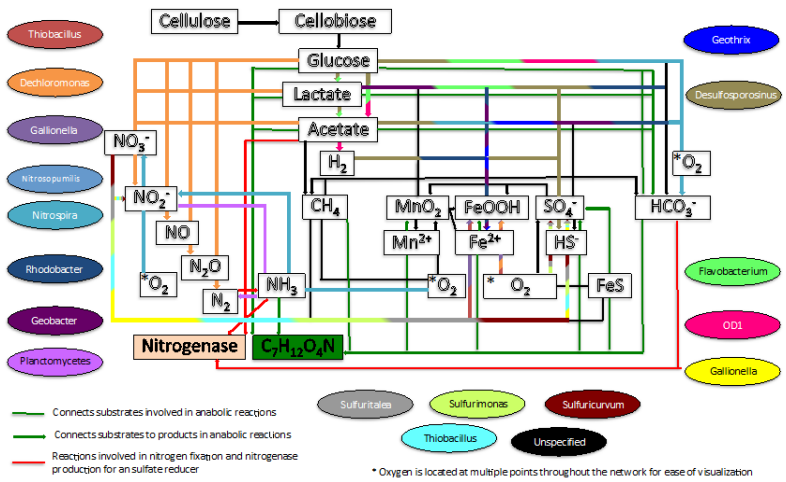

Figure 8: Schematic showing microbially mediated biogeochemical reaction network in the Rifle alluvial floodplain aquifer in western Colorado, with a range of solid and dissolved organic carbon compounds reacting with terminal electron-accepting components. Colored ovals refer to the microbe responsible for mediating portions of the network. (Figure courtesy Eric King and Eoin Brodie, LBNL).

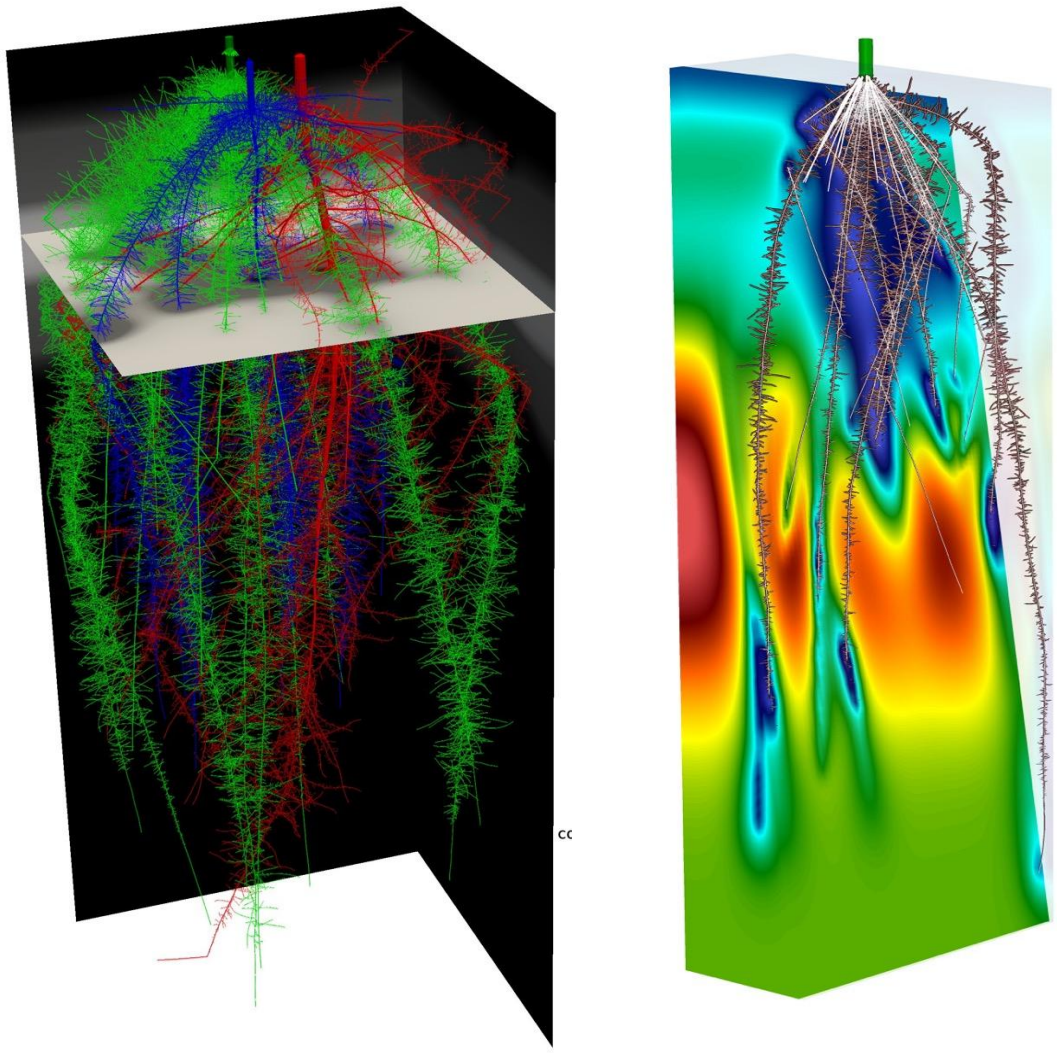

Figure 9: A functional-structural plant model, SimRoot, was used to visualize the root architecture of maize, bean, and squash as they competed for a pulse of nitrate in a leaching front. (For more details, see Postma and Lynch, 2012).

A potential application of a VPS framework to biofuel crops is the development of ideotypes or ideal phenotypes to guide crop breeding. Modeling is uniquely valuable in ideotype development because it permits the evaluation of many potential phenotypes in many different environments, including phenotypes and environments that do not yet exist in nature, such as future climate scenarios. It permits the analysis of how distinct traits and the genes that control them interact to affect crop performance at the level of plants, fields, and watersheds. The complexity of these systems and the large number of phenotypic permutations of interest exceed the capabilities of empirical research, and indeed require evolutionary optimization algorithms or other approaches to accommodate very large decision spaces.

The virtual plant-soil system has the potential to be a breakthrough discovery platform for understanding plants as hierarchies of inter-related systems at multiple scales. It would be an unparalleled tool for the analysis of how genetic variation affects the properties of whole plants interacting with their environment; how plant traits are integrated to form functional phenotypes; how specific phenotypes are suited to target environments; and how processes modeled at smaller scales (e.g. the scale of soil pores) and larger scales (e.g. the scales of stands and watersheds) interact with plant function.

4.2 Virtual Plot

A virtual plot would extend the virtual plant concept to include assemblages of individuals with explicitly resolved interactions between each individual and its aboveground and belowground environments. Such a virtual plot concept would help bridge the gap between the scale of an individual plant and that of a crop, hillslope or watershed. Multiple options exist for virtual plot models depending on the intended use, the spatial scale of interest, and the degree of parameterization of dynamics within individual plants and microbial groups.

One option for the virtual plot concept, applicable at the low end of the scale range, would be to include multiple instances of the virtual plant model within the same soil region. This approach would be most applicable to agricultural systems such as bioenergy crops, in which there is low diversity of plant types but important variability in other system states such as soil moisture content or organic matter distribution. Such a representation of the virtual plot could highly resolve three-dimensional soil processes in a manner that couples non-isothermal flow and microbiome genomics-informed reactive transport models with multiple instances of the virtual plant. Bringing together mechanistically resolved plant dynamics based on fundamental physical and metabolic constraints with explicitly resolved soil processes would provide an unprecedented opportunity to advance understanding of belowground interactions among individuals. It is expected that such a capability would be invaluable for developing tractable representations of tradeoff surfaces associated with different physical, metabolic, and life history strategies that are grounded in well-resolved theory and observations.

Another option that would be applicable at the upper end of the virtual plot scale range (thousands of square meters, and/or timescales associated with plant successional dynamics) would be to include more competing individuals with less detail in the process representations and without explicitly resolving spatial location (Figure 10) using simple emergent principles and properties as for water acquisition in Couvreur et al. (2012, 2014). Such an approach would represent an end member of the ecosystem demography class of models with cohorts coinciding with individuals. It is envisioned that this model would be coupled with an appropriately parameterized representation of belowground processes and would be used to advance understanding of the complex ecological dynamics and interactions that occur at the scale of a grid cell in a watershed model, taking into account the stochastic dynamics that arise among the individuals.

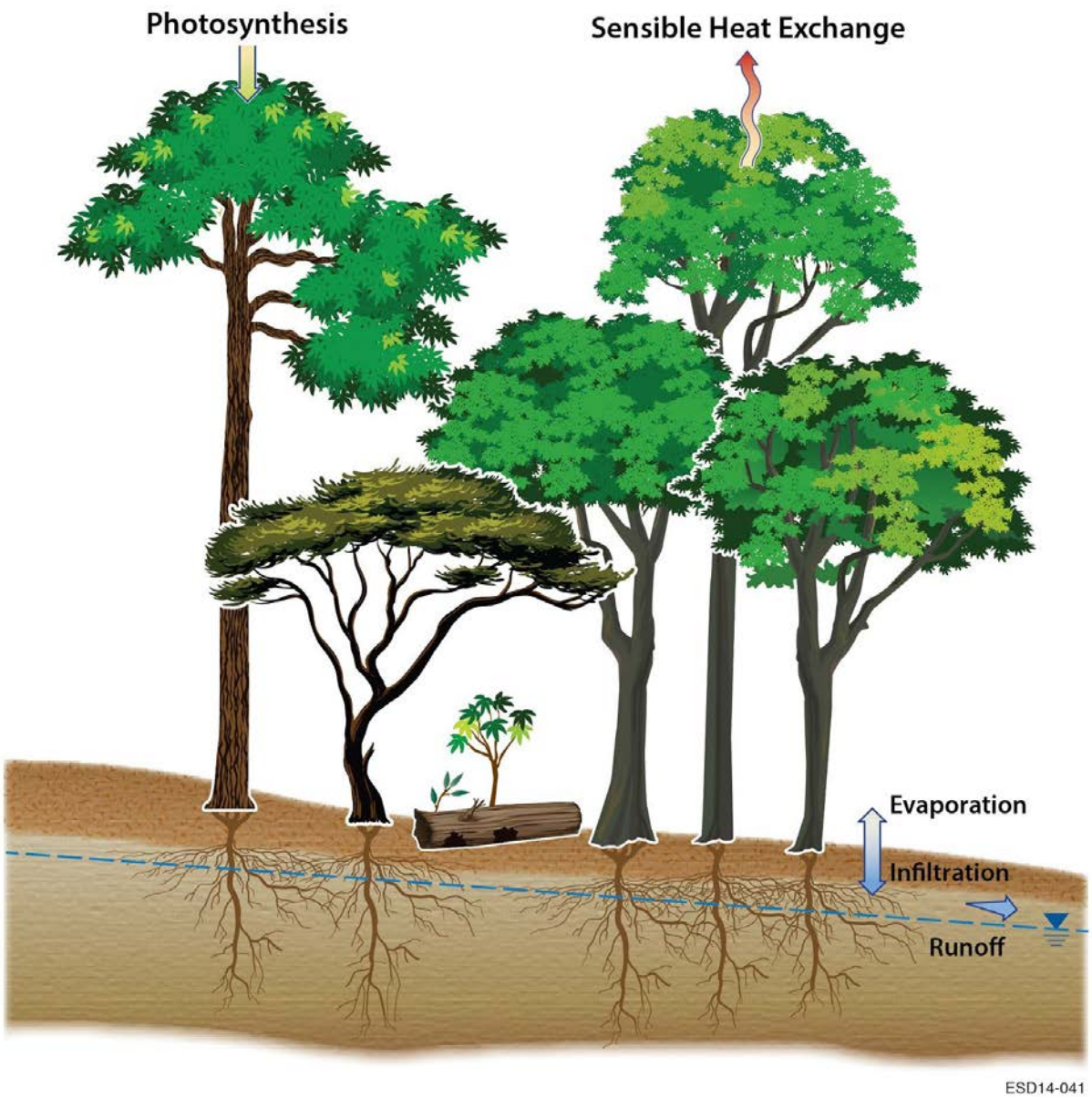

Figure 10: Plot scale interactions between vegetation, soil, local-scale hydrology, and the atmosphere (Figure from NGEE Tropics observations whitepaper; https://drive.google.com/file/d/0Bz9Zv4YIp0kPa3BqS2V1U1JiYTQ).

The virtual plot model could be used, for example, to address a key uncertainty accompanying global warming is the degree to which changes to the stores of carbon in terrestrial ecosystems will amplify or mitigate such changes. Models continue to predict large uncertainties in the magnitude of these feedback processes (Friedlingstein et al., 2006; 2014). Much of this uncertainty arises from the unknown response of plant and soil processes as ecosystems respond to shifting biome boundaries and the emergence of novel climates. Current Earth system models represent plant diversity via a small number of plant functional types (PFTs), defined with static traits that encompass both categorical and continuous plant properties, and only allow changes to

plant traits to occur through shifting PFT boundaries. However, plasticity of continuous traits may occur at multiple levels of ecosystem structure, from individuals to intra- and inter-species compositional changes. For example a large fraction of potential C losses arise from tropical forests warming beyond the thermal optimum for photosynthetic uptake, and the ability for plants to acclimate to such changes is not well constrained. One possibility is that the distributions of plant and microbial traits within a given community may shift to favor individuals whose traits and strategies are more competitive in the new environment, thus giving rise to a continuously optimizing ecosystem response to climate change. However key uncertainties for this are (a) the degree to which such traits can and will optimize under a changing climate, (b) the predictability of such shifts given the complexity of ecological interactions, the lack of understanding of the tradeoffs associated with these shifts, and the multi-faceted nature of global change, (c) and the timescale required for these shifts to occur (Scheiter et al, 2012). A particular challenge, when expanding the range of resolved plant traits, is to understand the tradeoffs that underlie the diversity of plant communities; such tradeoffs will reflect both physical limits, such as mass conservation or fundamental metabolic rates, as well as more emergent relationships by which only certain trait combinations are competitive—and therefore occur—in a given environment. Categorical traits, such as tree versus shrub form, may be better represented by the current PFT paradigm, but even in this case, model predictions of PFT shifts should be based on fundamental process representation rather than empirical bioclimatic rules as are currently prescribed in global models. A virtual plot model, which is mechanistic and modular in its formulation, could be used to run millions of in-silico experiments with varying conceptual formulations of the system structure and function and with varying parameter values to robustly assess the ecosystem response to a changing climate. As the range of environmental conditions that can be explored in ecosystem experiments and manipulations is very limited, and all ecosystem experiments are site specific, the use of mechanistic models to support the design, interpretation and extrapolation of field experiments is vital (e.g., Figure 2).

4.3 Virtual Watersheds

Although the challenges outlined in Section 3.2 are significant, research addressing individual process understanding is maturing rapidly. The confluence of this enabling research with anticipated advances in computing hardware, software design, and computational algorithms presents an unprecedented opportunity for a comprehensive community watershed simulation framework. Such a modeling framework is envisioned to be applicable from the hillslope to watershed and river-basin scales and be driven by weather and other environmental inputs on subdiurnal time scales over periods of years to decades. The starting point for such a capability would likely be 3-D variably saturated subsurface flow models coupled to surface flow models. The modeling system would extend this surface/subsurface capability to track fluxes and storage of water, energy, carbon, and nutrients on the surface and in the subsurface flows in three dimensions. Outputs from the watershed scale would be most useful if structured to pass fluxes of each conserved quantity to larger scales as atmospheric, surface, or subsurface flows to adjacent reservoirs (including other instances of the watershed model). Such a framework would

enable the evaluation of system behavior across a range of spatial and temporal contexts, including future climate and management scenarios.

As a next-generation capability, a key component would be the integration of dynamic vegetation models with the thermal hydrogeochemical model. This is necessary to track the effects of changing climate and disturbances due to fires, floods, insects, windthrow, and timber harvest on the hydrology, carbon, and nutrient cycles. Effective coupling between hydrogeochemical system and ecosystem representation requires attention to the plant-soil interface. A tractable coupling across these two scales is achievable with a flexible software design that enables coupling with finer scale mechanistic models of individual plants, the plant-soil interface, the rhizosphere, and surrounding microbial environments. The framework will be of most value if it presents developers and users with options for implementing coupling strategies that range from subgrid models that are informed by fine-scale mechanistic models through to explicitly coupled multiscale models using nested and/or adaptive meshes.

Figure 11: Stream meanders in the East River catchment, Gunnison County, Colorado. One hypothesis is that hyporheic zone flow through organic C rich sediments between meanders may have a large impact on the integrated carbon cycling in the river basin. This suggests the need to capture hydrologic and biogeochemical gradients at the fine scale (<1 m) and upscale these to the larger kilometer scale length of the river. This is clearly a computational challenge that will require a new terrestrial modeling software framework (Figure courtesy R. Kaltschmidt, LBNL [http://photos.lbl.gov/viewphoto.php?imageId=10591067]).

Significant improvements in the watershed-scale representation of the carbon and nutrient cycle are anticipated to be achievable by placing reactive transport models in a multiscale framework that is able to more accurately represent fluxes at the larger scales taking into consideration smaller scale heterogeneity of natural surface and subsurface materials. Three general approaches are suggested to achieve high fidelity upscaling of reactive transport processes: 1) high resolution deterministic modeling that captures gradients of small to large scale features directly, perhaps with the use of adaptive mesh refinement (AMR) methods, 2) stochastic approaches based on multiple runs at finer scales to generate probability distribution functions for use in larger scale simulations, and 3) reduced order methods (Pau et al., 2013). A multiscale framework would make it possible to compare observations and modeling results at multiple scales within a single aquifer or watershed system. To test the different approaches, catchments with smaller scale features that could potentially impact larger scale fluxes are ideal. An example might be the stream meander system at the East River site in Gunnison County, Colorado, where hyporheic zone flow through organic rich sediments over the meter scale may have a large impact on the integrated carbon cycling in the river system (Figure 11).

A multiphysics, multiscale watershed model that includes more mechanistic representation of the associated ecosystem will likely require a combination of directly measurable input (i.e. high-resolution topography, weather) and parameter estimation to infer model input parameters from spatially and temporally integrated measurements (eddy covariance flux, runoff, remote sensing products). Coordinated and co-located observations of weather, biophysical, geochemical, soil and subsurface processes are clearly needed, as are proven and well-documented procedures to perform the model/data integration. Significant opportunities exist for leveraging multiple experimental and sensor research efforts (e.g., NEON, and Critical Zone Observatories), and coordinated geospatial/temporal “Big Data” efforts from multiple agencies (satellites, lidar, reanalysis). Moreover, it is envisioned that the modeling framework will interface with existing tools that facilitate parameter estimation.

5. Software Infrastructure for Virtual Ecosystems

5.1. Managing complexity in process-rich applications

To realize the exciting potential of the virtual systems that are highlighted in Section 4, these systems must be able to leverage the existing pool of mechanistic and parameterized models as components of new models that target specific scientific questions. In addition, scientists need to be able to add new models to this collection of components as scientific understanding is gained. This ability to select models from a collection, independent of their original simulation code, and to configure and couple them in specific ways at runtime for the purpose of answering specific questions is the motivation behind referring to this new paradigm as a *software ecosystem of interoperable components*. This interoperability is critical to sharing models across application

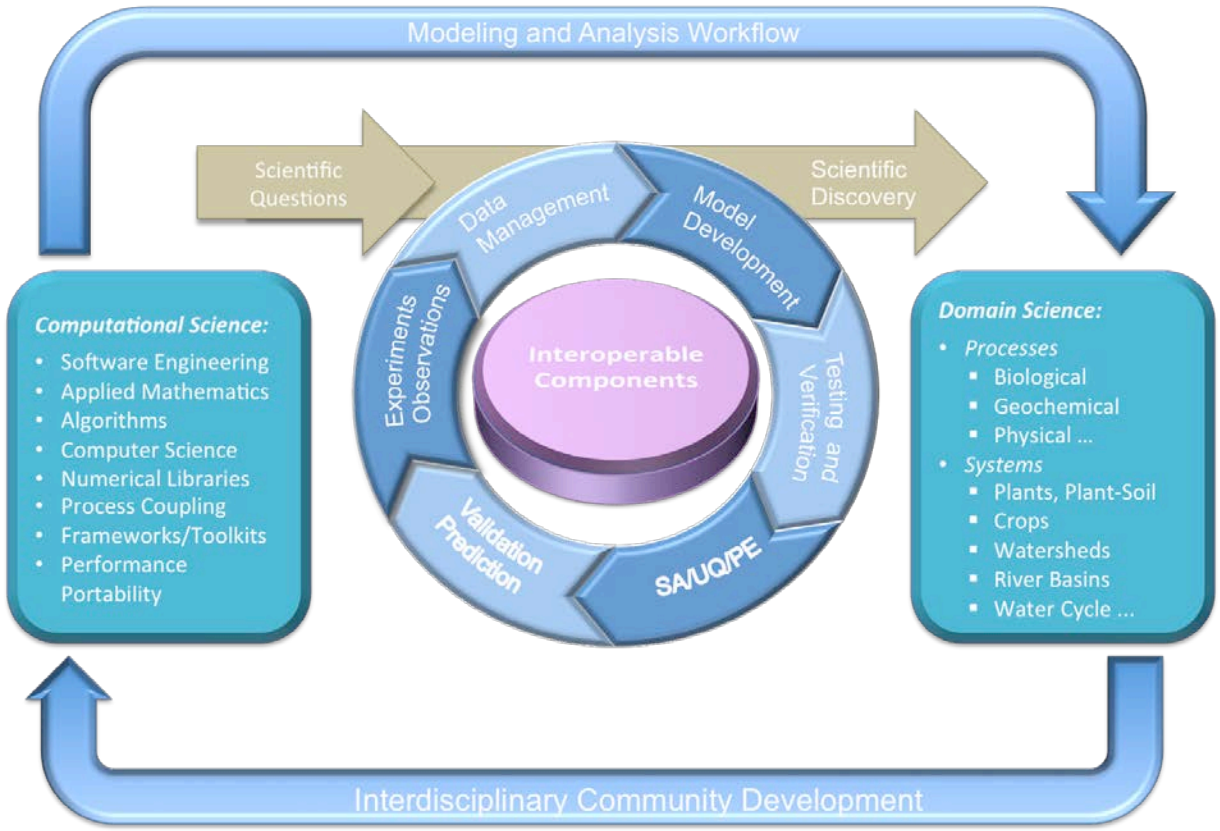

Figure 12: This schematic shows the integrated software ecosystem that is needed to realize the potential of the virtual systems envisioned in Section 4. Here, productivity of the modern scientific workflow (center ring) is enhanced because the critical phases of model development, simulation, and analysis leverage expertise and capability from the interdisciplinary community. Model development, leverages interoperable components generated by multiple projects, as well as contributing new model components to this pool within a flexible framework. Similarly, significant gains in the efficiency of the analysis phase, which includes Sensitivity Analysis (SA), Uncertainty Quantification (UQ), and Parameter Estimation (PE), are realized through more flexible and modular designs that enable efficient collaboration between the computational science and domain science communities.

teams, and for teams to benefit from the advances in numerical libraries and advanced computer architectures as well. This software ecosystem represents a critical step in realizing the BER Virtual Laboratory (BERAC, 2013), and is shown schematically in Figure 12.

To develop and support this collection of interoperable components, it is necessary to move beyond the design of traditional multiphysics simulation codes where all the data dependencies and model coupling strategies are enumerated and managed explicitly. Instead, more abstract lightweight multiphysics frameworks are needed that provide a uniform application programming interface (API) for process models, and can use dynamic model registration and data management approaches. Foundational work using directed acyclic graphs (DAGs) to characterize and manage variable dependencies was done by Notz et al. (2012), and has been applied in a number of applications. A simple DAG for an algebraic expression of water content is shown in Figure 13 (right). Recently, the use of DAGs has been combined with a graph-based representation of the model and its couplings (Coon et al., 2014) to provide a more accessible abstraction that can be used by scientists to create and interact with these models. This graph-based representation is referred to as a process-kernel tree, and an example for a surface/subsurface thermal-hydrology simulation that includes coupling to a surface energy balance model is shown in Figure 13 (left).

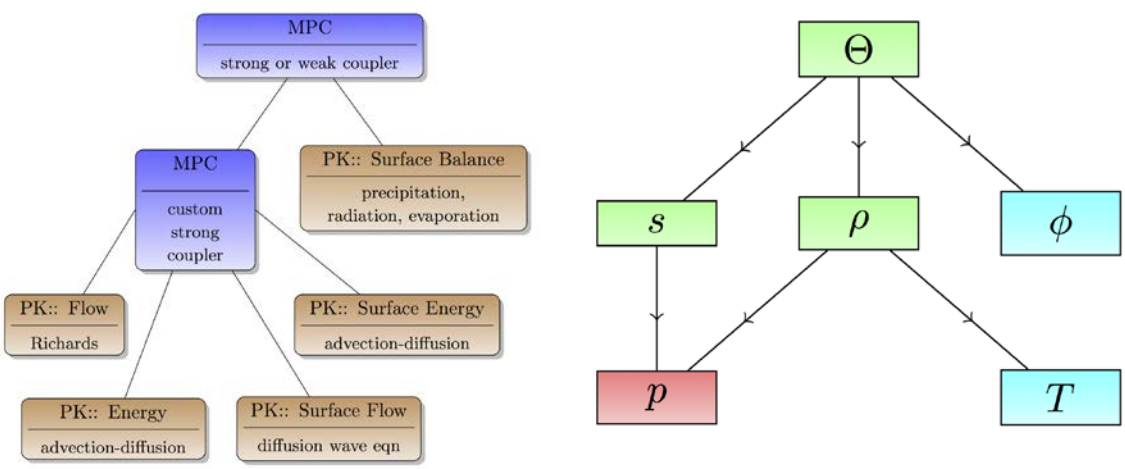

Figure 13: An abstraction of the process coupling hierarchy as a process tree is shown in the left panel, with strong and weak couplers shown in purple, and the model components (process kernels) shown in brown. This hierarchical representation, when combined with the directed acyclic graph (DAG) that describes the dependencies of model variables, is one approach to flexible runtime management of model configuration. An example of a DAG for the algebraic equation governing water content is shown in the right panel, with the primary variable shown in red, independent variables in blue, and secondary variables in green (Coon et al, 2014).

The benefits of this software ecosystem go beyond the flexibility of composing new complex multiscale models, playing a critical role in their maintainability and performance on new computer architectures. Specifically, by supporting standardized interfaces to enable collaborators outside the original application team to use the components, significantly more testing and verification is required. This additional testing improves reproducibility of scientific results and provides critical support to the increasingly frequent refactoring that will be needed to perform efficiently on new computer architectures. In addition, these more flexible and modular designs make it easier to expand the use of numerical libraries, and hence, aid model developers in minimizing the impact of changing hardware on the code they maintain.

Finally, these virtual systems must support the integration of an increasing amount of observational and simulation data from a variety of sources and across a range of scales. Recent work in the area of data management for multiphysics applications (Slattery et al., 2013) suggests that the abstractions used in this software ecosystem can be expanded to support this flexible integration of models and data. While not discussed extensively here, data management is recognized as a critical element of the software ecosystem. Other initiatives within BER are developing new community approaches to data management that must be integrated within the overall modeling and analysis workflow as denoted in Figure 12. In particular, the recent BERAC report on the BER Virtual Laboratory (BERAC, 2013) outlines a vision for a unified data management framework that is currently being implemented by BER's program manager for Climate Information and Data Management.

5.2. Disruptive changes in computer hardware and software

Based on results of the Gordon Bell Prize competitions, computational hardware performance delivered to real applications has increased by a factor of more than one million (1.35×10^6) in the two decades between 1988 and 2008, and improvements in algorithms and software design have provided an increase in computational power comparable in magnitude to the gains from hardware advances (Keyes, 2012).

However, a looming challenge is that recent trends in computing hardware design create significant uncertainty about programming models and the real performance of application codes. Those trends include large numbers of slower, simpler "cores"; less and generally slower memory per core; heterogeneous systems mixing multi-core, many-core, and GPU; uncertainty about the design and a general lack of tools. Moreover, software development tools, programming models and application architectures have not kept pace with these changes in hardware, creating significant uncertainty for domain and computational scientists. This is a shocking departure from the preceding period of approximately 20 years during which stable programming models, increasing single-processor speed, and MPI-based multiprocessor applications for scalability essentially guaranteed new science could be explored with incremental changes to existing codes. Thus, this confluence of disruptive trends in computer hardware with the drive toward predictive multiscale simulations is putting immense pressure on the scientific community to find new ways to maintain scientific productivity.

The need for new approaches to software development led ASCR to convene a workshop on Software Productivity for Extreme-Scale Science. The accompanying report (US DOE, 2014) describes in more detail the challenges and opportunities facing the community. An important finding of this productivity workshop was that significant improvements in development practices and interoperability of libraries could underpin a shift toward a more agile, high-quality ecosystem of composable components, ultimately maintaining or even enhancing productivity. Moreover, this paradigm shift at the library level could provide a catalyst for a similar shift in applications to an ecosystem of interoperable components developed and used by interdisciplinary teams. In many ways this BER workshop acknowledges the importance of this paradigm shift, and examines it in the context of a growing need for predictive multiscale simulation in terrestrial ecosystems.

5.3. Interdisciplinary Teams and Training

The potential impact of these disruptive changes is increasingly appreciated in the Terrestrial Ecosystem community because these applications are inherently process rich, and hence, their development and investigation naturally pushes beyond traditional multiphysics approaches. In particular, a wide range of mechanistic models are being actively developed for key components of plants, crops, and watersheds, and are evolving quickly as data and understanding begin to emerge. This creates an earnest need for flexible testing and exploration of model characteristics, predictive skill, and coupling, and contrasts the more traditional multiphysics simulations that implemented a small number of possible couplings for a small set of processes. The potential scientific gain from such an agile multiscale modeling framework is undeniable. Moreover, the multi-faceted challenges of delivering such a capability is well-suited to the interdisciplinary teams that are necessitated by the changing architectures.

Nevertheless, the transition to interdisciplinary teams is a very challenging proposition. Historically, a monolithic software base has been used very successfully, and migration away from this insulated environment carries with it both risk and opportunity. However, over the last ~15 years some codes have started to embrace toolkits, frameworks, and libraries as a means to provide MPI-based parallel capabilities, and accelerate development. For example, PFLOTRAN used the Portable, Extensible Toolkit for Scientific Computation (PETSc), to migrate a serial flow and reactive transport code to an MPI-based parallel environment. Similarly, Amanzi, a flow and transport code developed for the ASCEM (Advanced Simulation Capability for Environmental Management) focused on a modular and object-oriented high-level design leveraging Trilinos. These efforts have been very successful and are complementary in many respects, but despite significant efforts, the lack of interoperability between the underlying toolkits and libraries makes it difficult or impossible to share capabilities between them. In addition, more work is required in the design of interfaces and abstractions that will enable scientists to engage effectively in their areas of expertise, and remain both sheltered from and valuable to the other disciplines involved in the project.

Finally communication across interdisciplinary teams is very challenging because team members

have limited or no common vocabulary and experience. Fortunately, the shift to a software ecosystem of interoperable components will help drive the required shift to more thorough testing of components and their integration. This testing can provide a focal point that draws interdisciplinary teams together, allowing everyone to understand and appreciate the big picture, while contributing to the testing, analysis, and documentation of their components or library services.

5.4. Software Lifecycles and Business Models

Another key ingredient of successful interdisciplinary teams is formalizing both the lifecycle of the application and the software development methodologies used by the team. Software engineering is the methodical process for creating and deploying software products. Although, it is immature compared to other engineering disciplines, it is providing increasing value to software development organizations in many disciplines. Granted, the history of formal software engineering methods in computational science is not uniformly positive. Early methods, developed for business software projects, were naively applied with little adaptation to scientific projects. However, in recent years, the software engineering community has identified, described and promoted practices that can be incrementally adopted by software development teams. Test-driven development, sprints (focused development efforts of specific features) and iterative, incremental development are a few of the practices that have proven value for many different software projects (Feathers, 2004; McConnell, 2004). In addition, multiphase lifecycle models that manage the expectations and practices of a project from research to deployment (Feathers, 2004) and a broad collection of freely available, high quality software tools all provide new opportunities for software engineering to positively impact scientific software projects.

Although these changes in software engineering practices are very positive, they do highlight another important technical and cultural challenge. Specifically, the development of documentation and tests for research codes, and even some production codes is generally weak. Tests are often written during capability development, but they are not necessarily documented beyond the research paper, and are often not maintained as current development shifts to other capabilities. This is always a concern with portability and reproducibility, but the trends in hardware and uncertainty of programming models suggest testing must be sufficient to support nearly continuous refactoring of some parts of the code. It is critical that the community raise the expectation of testing and documentation, and automate many of the associated tasks to ensure that the impact on productivity is positive. The positive impact of testing on productivity has the additional benefit that it underpins an evaluation of business models for this new software ecosystem.

All software projects have a business model. Historically, within much of the scientific community this business model is based on competing for research funding to produce new modeling and simulation capabilities. In this case the business model is primarily implicit, with the software design, development and maintenance performed as a required activity to provide these capabilities. How funding is disbursed for these software projects is almost entirely a local

decision, and accountability for project decisions is seldom a concern, as long as the specific research objectives for the project are obtained. Governance is typically done via the leadership of the project principal investigator.

Although, this traditional business model has been successful in the past, it naturally evolves software and capabilities in a manner that optimizes the competitive position of a small team or principal investigator to obtain additional research funding, as opposed to optimizing scientific contributions to the community. Consequently, it discourages the use of numerical libraries or having external dependencies in the code, and encourages the development of one-off capabilities, which incur disproportionate maintenance costs in the future. In addition, it places no explicit value on reproducibility of simulation results beyond publishing journal articles. Thus, it cannot be sustained through this period of intense change in both computer hardware and software. Instead, careful evaluation and development of a more explicit business model is needed to foster a healthy and vibrant community that supports the software ecosystem envisioned here. This new model must recognize and support contributions of high-quality software to the community, in order to balance the important roles of collaboration and competition. This balance is critical to enabling multiple loosely-coupled teams, to effectively share capabilities across projects, and for the funding agencies to establish a more holistic approach to modeling and simulation across their portfolios. Cross-project working groups or leadership teams may be the best way to begin exploring this critical issue.

5.5. Leveraging new programs in software productivity

The preceding discussions of both the scientific drivers and the pending crisis in software productivity point to a variety of opportunities that arise through the engagement of an interdisciplinary community. In fact, a growing consensus among scientists across the relevant disciplines is that these combined challenges cannot be handled by a single group, but require the formation and training of this broader community. Thus, in response to the Software Productivity for Extreme-Scale Science workshop and report (US DOE, 2014), a Funding Opportunity Announcement (FOA) jointly supported by ASCR and BER was made in late May 2014, and after completing the review process, an award was made. This new multi-lab project, titled *Interoperable Design of Extreme-scale Application Software* (IDEAS), has a strong focus on software development methodologies, and significant improvements to the interoperability of important open-source libraries and toolkits (e.g., PETSc and Trilinos). In addition, its phased development is driven by Use Cases that have been developed in close collaboration with projects in the Terrestrial Ecosystem Science (TES) and Subsurface Biogeochemical Research (SBR) programs.

It is anticipated that IDEAS will provide critical enabling technology to help support the transition of BER applications to an open-source, community supported, ecosystem of interoperable components. Specifically, drawing on expertise in computational science, and in

collaboration with geoscientists in the BER driven Use Cases, this project will:

- Explore formal development of Lifecycle plans for a range of codes critical to BER.
- Develop and support higher levels of interoperability between libraries.
- Develop, document and promote modern software methodologies.
- Prototype frameworks and tools for hierarchical automated testing.
- Develop formal approaches for componentization of important capabilities.
- Develop design patterns and abstractions suitable for task-centric architectures.
- Co-develop a flexible and extensible open framework and interface APIs for the community.

These activities provide critical support for developing the virtual systems highlighted in Section 4, within an ecosystem of interoperable components. In particular, development of an open framework that standardizes APIs for both specialized libraries such as biogeochemistry, and more generic components (e.g., flow) enables the effective use of an ecosystem of interoperable components. Existing models can be componentized using this API, or necessary interface layers can be written. This capability extends the established multiphysics approaches that *Flow and Reactive Transport Models* pioneered 20 years ago, and can extend the useful life of many of their components. But this is really just the beginning. There are significant research opportunities that will emerge in this area as the needs of the applications are explored, the integration of data is addressed, and the workflows are enhanced and automated.

Finally, it is important to reiterate that this significant transformation of both the community and its software is an evolutionary process. To be successful, a phased approach is needed that clearly serves the needs of the community, and enhances its scientific productivity. For example, significant improvements in software engineering methodologies, testing and documentation, increased flexibility in coupling, and componentization are ideal targets for the first phase (~2 years). Looking out further, enabling new science through an ecosystem that supports more dynamic multiscale and multiphysics coupling is possible in the second phase (5 years), including support for stochastic and probabilistic representations. In addition, this phase could include more aggressive refactoring for performance, and performance engineering, as well as important integration of data for UQ, SA or PE studies. Underlying library components and toolkits will have made significant improvements in portability and performance during this phase as well. Finally, in the last phase (10 years) the framework and interoperability of its components would mature, supporting more complex iterative workflows involving multiscale models, a variety of national and local data sources and real-time data streams, a mix of cloud based and exa-scale simulations, and enhanced visualization-driven analytics.

6. Concluding Remarks

Several recent reports developed by the Office of Biological and Environmental Research have emphasized the need for improved multiscale representations of coupled and heterogeneous process mechanisms in computational simulations to build predictive understanding.

A 2010 workshop on "Complex Systems Science for Subsurface Fate and Transport" (BER, 2010) noted "predictive models provide the context for knowledge integration… State-of-science understanding codified in models can provide a basis for testing hypothesis, guiding experiment design, integrating scientific knowledge… and translating this information to support informed decision making…" The participants identified three high-impact research opportunities, all of which motivate the research directions defined in this report: 1) Understand fundamental subsurface process coupling; 2) Identify and quantify scale transitions in hierarchical subsurface systems; and 3) Understand integrated system behavior. A 2012 workshop on "Community Modeling and Long-Term Predictions of the Integrated Water Cycle" (BER, 2012a) identified as one of three Science Grand Challenges the need for "Modeling the multi-scale atmospheric and terrestrial processes and their interactions." A 2013 workshop on "Research for Sustainable Bioenergy: Linking Genomic and Ecosystem Sciences"(BER, 2013) highlights multiscale modeling as one of four key research opportunities, and points out that "the opportunity to develop multiscale, mechanistic models is expanding as… process-level functional understanding of genomic and phenomic differences among plants and their microbiomes improves."

The first of five goals outlined in the Climate and Environmental Sciences Division Strategic Plan (BER, 2012b) is to "synthesize new process knowledge and innovative computational methods advancing next-generation, integration models of the human-Earth system." Similarly, the Genomic Science Strategic Plan (BER, 2014) defines one of five program objectives as "develop the knowledgebase, computational infrastructure, and modeling capabilities to advance predictive understanding and manipulation of biological systems."

Finally, a BER Advisory Committee report on "BER Virtual Laboratory: Innovative Framework for Biological and Environmental Grand Challenges" (BERAC, 2013) calls for development of a Cyberinfrastructure, Analytics, Simulation, and Knowledge Discovery (CASK) framework that "would provide the computational infrastructure needed to integrate disparate and multiscale measurements, theory, and process understanding into predictive models."

It is clearly evident that the broad range of BER scientists, program managers, and advisors that contributed to these documents have recognized the importance of the concepts discussed in this workshop to the future directions of BER science.

This workshop report details a vision for development of advanced simulation frameworks crosscutting three scales of critical importance to terrestrial system function: 1) the Virtual Plant-Soil System; 2) the Virtual Plot; and 3) the Virtual Watershed. Each virtual system will provide

improved predictive understanding of terrestrial ecosystems at their respective scales, with higher-fidelity models informing improved process representations and parameterizations at larger scales. The complexity of the envisioned process models and their multiscale coupled interactions requires a new computational paradigm that moves away from traditional monolithic code development practices. The new approach is based on community development of interoperable component process models linked by advanced process coupling algorithms and standardized component interface specifications. It incorporates advanced software engineering practices, including rigorous component testing and documentation that will improve reliability and reproducibility of scientific results. We perceive that only through adoption of such a shift by the BER community of scientists, and through close collaboration with the Office of Advanced Scientific Computing Research such as has been initiated through the IDEAS project, will it be possible to overcome the disruptive effects of new computer architectures, thereby maximizing the scientific productivity of a BER research portfolio that is increasingly focused on predictive simulation tools as an integrative science outcome.

Acknowledgements

The workshop organizers and co-leads thank all of the scientists who energetically participated in the workshop discussions and generously contributed their time and ideas to this important activity for the Office of Biological and Environmental Research. We especially thank Gary Geernaert, Director of CESD for his support and interest, the speakers that assisted with the background session (Eoin Brodie, and Stephen Long) and the lightning talks (Lois Curfman McInnis, Glenn Hammond, Scott Painter, Xinguang Zhu, Chris Duffy), as well as all of the break-out leads, scribes and report-back presenters (Shawn Serbin, Chris Duffy, David Weston, Amilcare Porporato, Anthony Bishopp, Velentin Couvreur, Scott Denning, Elena Shevliakova, Darren Drewry, Glenn Hammond, Eoin Brodie, Gretchen Miller, Scott Painter, David Bernholdt). In addition, we thank Peter Thornton and Chris Duffy for providing input to the writing team on data integration and modeling at watershed scales, and for Scott Denning's thoughtful contributions to business models and governance. Also, we thank Charlie Koven for helpful discussion and contributions on vegetation modeling and representation at the plot scale. We extend our sincere appreciation to David Thomassen for his thoughtful review and detailed comments on early drafts; his feedback significantly improved the report. Finally, we are thankful to Andrew Flatness, Nver Mekerdijian (BER), Keri Cagle, and Deneise Terry (ORISE), for organizing the workshop logistics and supporting the workshop participants during the meeting.

Appendices

Appendix A. Agenda

Computational Challenges for Mechanistic Modeling of Terrestrial Environments
March 26-27, 2014.

Day1: Wednesday, March 26, 2014.

7:30 am – 7:45 am	Transport from Hotel to DOE Headquarters
7:45 am – 8:15 am	DOE Security Check
8:15 am – 8:30 am	Coffee and snacks in A-410
8:30 – 8:45	**Welcome: Background and Goals for Workshop** *David Lesmes, BER – CESD*
8:45 – 9:00	**Multiscale-Multiphysics Modeling: Science Challenges** *Tim Scheibe, PNNL*
9:00 – 9:15	**"Software Crisis" = Opportunity!** (Multiscale-Multiphysics Modeling: Computational Challenges) *David Moulton, LANL; Lois Curfman McInnes, ANL; Mike Heroux, SNL*
9:15 – 9:30	**Genome Based Trait Models: Microbes to Plants to Ecosystems** *Eoin Brodie, LBNL and David Weston, ORNL*
9:30 – 9:45	**Reactive Transport Models (RTMs): Pores to Plots to Watersheds** *Carl Steefel, LBNL*
9:45 – 10:00	**Plant Models: Plant Tissues to Whole Plants to Crops** *Stephen Long, UIUC and Jonathan Lynch, PSU*
10:00 – 10:15	**Terrestrial Ecosystem Models (RTM-CLM): Plants to Plots to Watersheds and Beyond….** *Peter Thornton, ORNL*
10:15 – 10:30	**Open Discussion Q&A**
10:30 – 10:45	**Introduction to Breakout #1**
10:45 – 11:00	**BREAK**
11:00 – 1:45 pm	**Breakout #1: Science Challenges and Opportunities**

Lunch break of 45 min. is at the groups discretion, but splitting into two one hour sessions is likely the best option.

Session 1A: Multiscale Modeling of Coupled Plant-Soil Systems
(Mechanistic Models at Whole Plant to Crop Scales)
Session 1B: Multiscale Modeling of Terrestrial Ecosystems
(Mechanistic Models at Plant to Plot to Watershed Scales)

1:45 – 2:15	**Reports from Breakout #1: Sessions A and B**
2:15 – 2:30	**BREAK**
2:30 – 2:45	**Software Engineering Practices for Community Code Development** *Mike Heroux, SNL; Lois Curfman McInnes, ANL; David Moulton, LANL*

2:45 – 3:15	**Lightning Talks about Code Development followed by Open Q&A** *Lois Curfman McInnes, ANL* - PETSc *Glenn Hammond, SNL* – PFLOTRAN (built on PETSc) *Scott Painter, LANL* – Arctic Terrestrial Simulator (built on Trilinos) *Xinguang Zhu*, Shanghai, e-Photosynthesis (built with Matlab…?) *Chris Duffy*, PSU, *PIHMS: The model-data nexus* (SUNDIALS/QGIS)
3:15 – 3:30	**Introduction to Breakout #2**
3:30 – 5:30	**Breakout #2: Multiscale Frameworks for Mechanistic Modeling: Design Requirements, Governance, Implementation, Business Models, etc…** (CS team spread across the three breakouts)

Session 2A: Multiscale Modeling of Coupled Plant-Soil Systems
Session 2B: Multiscale Modeling of Terrestrial Ecosystems
Session 2C: Multiscale Modeling of Reactive Transport

5:30 – 6:00	**Reports from Breakout #2: Sessions A, B and C**

Day 2: Thursday, March 27

7:30 am – 7:45 am	Transport from Hotel to DOE Headquarters
7:45 am – 8:15 am	DOE Security Check
8:15 am – 8:30 am	Coffee and snacks in A-410
08:30 – 09:00	**Open Discussion Q&A**
09:00 – 09:15	**Introduction to Breakout #3: Prioritization of Research Needs**
09:15 – 09:30	**BREAK**
09:30 – 11:30	**Breakout #3: Prioritization of Research Needs**

Session 3A: Multiscale Modeling of Coupled Plant-Soil Systems
Session 3B: Multiscale Modeling of Terrestrial Ecosystems
Session 3C: Multiscale-Multiphysics Libraries, Couplers, Workflow, etc.

11:30 – 12:15 pm	**Reports from Breakout #3: Sessions 3A to 3C**
12:15 – 12:30	**Concluding Remarks - Adjourn**
12:30 – 01:30	Lunch in DOE Cafeteria
01:30 – 04:30	Writing Team Begins Drafting Final Report

Appendix B. Workshop Participants

Organizing Committee

Co-Leads

David Moulton
Los Alamos National Laboratory

Tim Scheibe
Pacific Northwest National Laboratory

Carl Steefel
Lawrence Berkeley National Laboratory

Members

Mike Heroux
Sandia National Laboratory

Stephen Long
University of Illinois, Urbana Champaign

Jonathan Lynch
Pennsylvania State University

Lois Curfman McInnes
Argonne National Laboratory

Peter Thornton
Oak Ridge National Laboratory

Participants

David Bernholdt
Oak Ridge National Laboratory

Michael Ek
National Oceanic and Atmospheric Administration

Lawren Sack
University of California, Los Angeles

Anthony Bishopp
University of Nottingham, UK

Glenn Hammond
Sandia National Laboratory

Shawn Serbin
University of Wisconsin / Brookhaven National Laboratory

Eoin Brodie
Lawrence Berkeley National Laboratory

Kerstin Kleese Van Dam
Pacific Northwest National Laboratory

Elena Shevliakova
National Oceanic and Atmospheric Administration

Valentine Couvreur
University of California, Davis

Paul Moorcroft
Harvard University

Dali Wang
Oak Ridge National Laboratory

Scott Denning
Colorado State University

Gretchen Miller
Texas A&M

Dave Weston
Oak Ridge National Laboratory

Darren Drewry
Jet Propulsion Laboratory - NASA

Scott Painter
Oak Ridge National Laboratory

John Wu
Lawrence Berkeley National Laboratory

Chris Duffy
Pennsylvania State University

Amilcare Porporato
Duke University

Xinguang Zhu
Shanghai Institutes for Biological Sciences

Appendix C. References

Ainsworth, E. A., A. D. B. Leakey, D. R. Ort, S. P. Long, FACE-ing the facts: inconsistencies and interdependence among field, chamber and modeling studies of elevated [CO2] impacts on crop yield and food supply. *New Phytologist* **179**, 5 (2008).

Bennett, T., G. Hines, O. Leyser, Canalization: what the flux? *Trends in Genetics* **30**, 41 (Feb, 2014).

BER, DOE Workshop on Community Modeling and Long-Term Predictions of the Integrated Water Cycle, Report from the September 2012 Workshop, DOE/SC-0155, U. S. Department of Energy, Office of Science, Biological and Environmental Research, Washington, D. C. (2012a).

BER, Climate and Environmental Sciences Division Strategic Plan, DOE/SC-0151, U. S. Department of Energy, Office of Science, Biological and Environmental Research, Washington, D. C. (2012b).

BER, Research for Sustainable Bioenergy: Linking Genomic and Ecosystem Sciences, DOE/SC-0167, U. S. Department of Energy, Office of Science, Biological and Environmental Research, Washington, D. C. (2013).

BER, DOE Genomic Sciences Program: Mission-Drive Systems Biology, 2014 Strategic Plan, U. S. Department of Energy, Office of Science, Biological and Environmental Research, Washington, D. C. (2014).

BERAC, BER Virtual Laboratory: Innovative Framework for Biological and Environmental Grand Challenges, *DOE/SC-0156*. U. S. Department of Energy, Biological and Environmental Research Advisory Committee, Washington, D.C. (2013).

Brutsaert, W., The Parameterization Of Regional Evaporation - Some Directions And Strategies. Journal of Hydrology **102**, 409-426 (1988).

Cheq, Y. H. *et al.*, Multiscale digital Arabidopsis predicts individual organ and whole-organism growth. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E4127 (Sep, 2014).

Coon, E. T., J. D. Moulton, and S. Painter, Managing complexity in simulations of land surface and near- surface processes. Technical Report LA-UR 14-25386, Los Alamos National Laboratory, 2014. *submitted to Environmental Modelling & Software* (2014).

Couvreur, V., J. Vanderborght and M. Javaux, A simple three-dimensional macroscopic root water uptake model based on the hydraulic architecture approach, *Hydrol. Earth Syst. Sc.*, 16, 2957-2971 (2012).

Couvreur, V., J. Vanderborght, L. Beff and M. Javaux, Horizontal soil water potential heterogeneity: Simplifying approaches for crop water dynamics models, *Hydrol. Earth Syst. Sc.*, 18, 1723-1743 (2014).

Domagalska, M. A, O. Leyser, Signal integration in the control of shoot branching. *Nature Reviews Molecular Cell Biology* **12**, 211 (Apr, 2011).

Drewry D.T., Kumar P. & Long S.P. Simultaneous improvement in productivity, water use, and albedo through crop structural modification. *Global Change Biology* **20**, 1955-1967. (2014)

Dyson, R. J. *et al.*, Mechanical modelling quantifies the functional importance of outer tissue layers during root elongation and bending. *New Phytologist* **202**, 1212 (Jun, 2014).

Feathers, M., *Working Effectively with Legacy Code*, Prentice Hall, 2004.

Hill, K. *et al.*, Root Systems Biology: Integrative Modeling across Scales, from Gene Regulatory Networks to the Rhizosphere. *Plant Physiology* **163**, 1487 (Dec, 2013).

Jarvis, N. J., A simple empirical model of root water uptake, J. Hydrol. **107**, 57–72 (1989).

Keyes, D., High performance computing in science and engineering: The tree and the fruit, presentation at the Networking and Information Technology Research and Development (NITRD) Symposium, available online at www.cra.org/ccc/files/docs/nitrdsymposium/keyes.pdf, (Feb. 16, 2012).

Kollet, S. J., & Maxwell, R. M. Capturing the influence of groundwater dynamics on land surface processes using an integrated, distributed watershed model. *Water Resources Research*, **44** (2008).

Lynch, J. P., Nielsen, K. L., Davis, R. D., & Jablokow, A. G. SimRoot: modelling and visualization of root systems. *Plant and Soil* **188**, 139-151, (1997).

Lynch, J. P., Steep, cheap and deep: an ideotype to optimize water and N acquisition by maize root systems. *Annals of Botany* **112**, 347 (Jul, 2013).

McConnell, S., *Code Complete: A Practical Handbook of Software Construction*, Microsoft Press, 2004.

Oden, T. and O. Ghattas, Computational science: The “Third Pillar” of science, TAMEST blog article, http://blog.tamest.org/computational-science-the-third-pillar-of-science/, The Academy of Medicine, Engineering & Science of Texas, University of Texas at Austin, 2014.

Pau, G. S. H., Zhang, Y., & Finsterle, S. Reduced order models for many-query subsurface flow applications. *Computational Geosciences* **17**, 705-721, (2013).

Postma, J. A., & Lynch, J. P. Complementarity in root architecture for nutrient uptake in ancient maize/bean and maize/bean/squash polycultures.*Annals of botany* **110**, 521-534, (2012).

Scheibe, T.D., E.M. Murphy, X. Chen, A.K. Rice. K.C. Carroll, B.J. Palmer, A.M. Tartakovsky, I. Battiato, B.D. Wood. An analysis platform for multiscale hydrologic modeling with emphasis on hybrid multiscale methods. Groundwater DOI: 10.1111/gwat.12179 (2014).

Song, Q., G. Zhang, X.-G. Zhu, Optimal crop canopy architecture to maximise canopy photosynthetic CO2 uptake under elevated CO2 - a theoretical study using a mechanistic model of canopy photosynthesis. *Functional Plant Biology* **40**, 109 (2013, 2013).

Song, Y. H., R. W. Smith, B. J. To, A. J. Millar, T. Imaizumi, FKF1 Conveys Timing Information for CONSTANS Stabilization in Photoperiodic Flowering. *Science* **336**, 1045 (May,

2012).

Steefel, C.I, D. DePaolo, and P.C. Lichtner, Reactive transport modeling: An essential tool and a new research approach for the Earth sciences, *Earth and Planetary Science Letters* **240**: 539-558 (2005).

Steefel, C.I., Appelo, C.A.J., Arora, B., Jacques, D., Kalbacher, T., Kolditz, O., Lagneau, V., Lichtner, P.C., Mayer, K.U., Meeussen, J.C.L., Molins, S., Moulton, D., Shao, H., Šimůnek, J., Spycher, N., Yabusaki, S.B., Yeh, G.T. Reactive transport codes for subsurface environmental simulation, *Computational Geosciences*, DOI: 10.1007/s10596-014-9443-x (2014).

Stegen, J. C., X. J. Lin, J. K. Fredrickson, X. Y. Chen, D. W. Kennedy, C. J. Murray, M. L. Rockhold, and A. Konopka, Quantifying community assembly processes and identifying features that impose them, ISME Journal, 7(11): 2069-1079, DOI:10.1038/ismej.2013.93 (2013).

Steinacher, A., O. Leyser, R. H. Clayton, A computational model of auxin and pH dynamics in a single plant cell. *Journal of Theoretical Biology* **296**, 84 (Mar, 2012).

U.S. DOE, Extreme-Scale Scientific Application Software Productivity [pdf], report prepared for the Office of Advanced Scientific Computing Research (ASCR) Workshop on *Software Productivity for Extreme-Scale Science*, Hilton Rockville Rockville, MD, January 13-14, 2014.

U.S. DOE, Complex Systems Science for Subsurface Fate and Transport: Report from the August 2009 Workshop, DOE/SC-0123, U.S. Department of Energy Office of Science http://www.science.doe.gov/ober/BER_workshops.html (2009).

Voepel, H.; Ruddell, B.; Schumer, R.; Troch, P. A.; Brooks, P. D.; Neal, A.; Durcik, M.; Sivapalan, M., Quantifying the role of climate and landscape characteristics on hydrologic partitioning and vegetation response. Water Resources Research, **47** (2011).

Wang, Y., S. P. Long, X. G. Zhu, Elements Required for an Efficient NADP-Malic Enzyme Type C4 Photosynthesis. *Plant Physiology* **164**, 2231 (Apr, 2014).

Williams, J. W.; Jackson, S. T., Novel climates, no-analog communities, and ecological surprises. Frontiers in Ecology and the Environment **5**, 475-482 (2007).

Wood, E. F., Roundy, J. K., Troy, T. J., Van Beek, L. P. H., Bierkens, M. F., Blyth, E., et al., Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water.*Water Resources Research* **47** (2011).

Zhu, X.G., Y. Wang, D. R. Ort, S. P. Long, e-photosynthesis: a comprehensive dynamic mechanistic model of C3 photosynthesis: from light capture to sucrose synthesis. *Plant Cell and Environment* **36**, 1711 (Sep, 2013).

Zhu, X. *et al.*, The next generation models for crops and agro-ecosystems. *Science China-Information Sciences* **54**, 589 (Mar, 2011).

Appendix D. Acroynms and Abbreviations

ACME	Accelerated Climate Modeling for Energy
Ameriflux	American flux measurement network
AMR	Adaptive Mesh Refinement
ARM-SGP	Atmospheric Radiation Measurement - Southern Great Plains site
ASCEM	Advanced Simulation Capability for Environmental Management
ASCR	Office of Advanced Scientific Computing Research
BER	Office of Biological and Environmental Research
BERAC	Biological and Environmental Research Advisory Committee
BRC	Bioenergy Research Centers
CASK	Cyberinfrastructure, Analytics, Simulation, and Knowledge Discovery
CESD	Climate and Environmental Sciences Division
C	Carbon
CO_2	Carbon Dioxide
DAG	Directed Acyclic Graph
EES	Environmental System Science
ESM	Earth System Model
FOA	Funding Opportunity Announcement
IDEAS	Interoperable Design of Extreme-scale Application Software
KBase	Systems Biology Knowledge Base
MPI	Message Passing Interface
NGEE	Next Generation Ecosystem Experiment
NEON	NSF National Ecological Observatory Network
NSF	National Science Foundation
PE	Parameter Estimation

PETSc	Parallel Extensible Toolkit for Scientific Computation
SA	Sensitivity Analysis
SBR	Subsurface Biogeochemical Research
SFA	Scientific Focus Areas (national laboratory research projects)
SPRUCE	Spruce and Peatland Responses Under Climatic and Environmental Change
TES	Terrestrial Ecosystem Science
Trilinos	Enabling technologies within an object-oriented software framework
UQ	Uncertainty Quantification
VPS	Virtual Plant-Soil system