B. EXECUTIVE SUMMARY

*The cornerstone of prediction is transferable knowledge. Our vision is to deliver knowledge, models, and data that are transferable across basins and needed to quantify, understand, and predict how watershed perturbations and biophysical settings combine to drive water and material transport into, perturbations within, and cumulative biogeochemical function of sediment-associated components of stream networks (**Fig. B.1**).* Over the next quadrennial period we will focus on spatiotemporal scaling of sediment-associated respiration (ERsed) in stream networks as influenced by perturbations and watershed system processes. Our intent is to study physical, chemical, and biological processes across the hillslope-to-stream continuum and use that systems knowledge to enable predictive understanding of ERsed spatiotemporal scaling. We focus on ERsed for three reasons: (*i*) We use it as a litmus test in that our ability to understand and predict spatiotemporal patterns of ERsed reflects the quality of our predictive understanding of integrated watershed processes, (*ii*) ERsed contribution to C cycling in stream networks ranges from being the dominant driver to being inconsequential, yet this variation has not been broadly quantified, understood, or predicted, and (*iii*) ERsed is likely sensitive to perturbations occurring within and beyond stream networks, but these influences remain largely unknown. Wildfire and the loss and gain of surface water (i.e., variable inundation) are two influential perturbations that occur across most of Earth's land surface and are increasingly common due to global change. In turn, these are our focal perturbations external to and within stream networks, respectively. We will use a hypothesis-driven approach to reveal how wildfire and variable inundation interactively influence ERsed through changes to hydrologic connectivity, organic matter chemistry, and biological communities. We will further reveal how climate and other biophysical features modulate these interactions. Our investigations of ERsed spatiotemporal scaling will use a theoretical framework based on properties of fractal distribution networks. This theory currently provides quantitative, but static predictions for how the cumulative function of stream networks scales with upstream watershed area; referred to here as 'ERsed allometry.' We will bring watershed system dynamics and perturbation impacts into this theory by integrating new understanding of time-varying processes across the hillslope-to-stream continuum into a hybrid watershed-scale modeling capability. This capability will be used test additional hypotheses and trigger follow-on data generation and modeling efforts, collectively focused on understanding how the temporal dynamics in ERsed allometry emerge from watershed system processes. Our work will be done within and across climatically divergent basins to enhance the transferability of resulting knowledge and models. By 'transferable' we mean rules or patterns that apply or exist across diverse biophysical conditions, or knowledge on how/why rules/patterns change across biophysical conditions. We will generate such knowledge to enhance predictive understanding via integration of diverse data and processes.

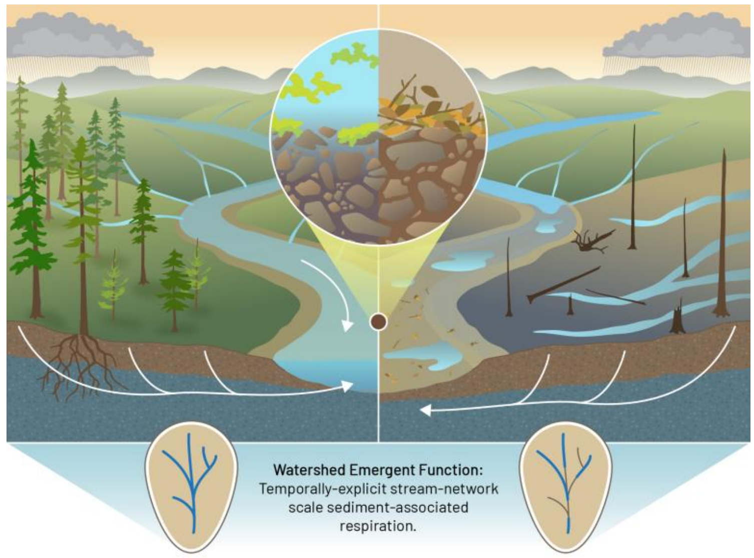


Figure B.1. The RC-SFA takes a systems approach to understand and predict hydro-biogeochemical function across divergent biophysical settings, including interactive influences of perturbations across the hillslope-to-stream continuum.

Table B.1. RC SFA Overview

Investigators	
PI James Stegen	**Co-PI** Xingyuan Chen; **Co-PI** Allison Myers-Pigg
Cross-Cut Leads	Etienne Fluet-Chouinard, Vanessa Garayburu-Caruso, Amy Goldman, Glenn Hammond, Peishi Jiang, Peter Regier, Jianqiu Zheng
Science Theme 1: Landscape Controls on Hydro-biogeochemistry	
Objective	Develop process-based understanding of the controls of watershed biophysical characteristics on the transport of OM to stream networks across climates.
Approach	Integrate the influences of vegetation, topography, subsurface physicochemical properties, and wildfire perturbations to predict export of water and DOC to streams.
Hypotheses	**H1.1:** Climate, vegetation composition, topographic slope, and subsurface hydrologic conductance are the key biophysical variables that drive cross-watershed variation in hydrologic connectivity and OM transport to stream networks. **H1.2:** Changes to hydrologic connectivity and transported OM following wildfire will be time-varying functions based on recovery dynamics of vegetation and soil hydraulic properties.
Science Theme 2: Impacts of Variable Inundation and Wildfire on ERsed from Molecular-to-Reach Scales	
Objective	Reveal the influences and underlying mechanisms of variable inundation and wildfires on spatial and temporal dynamics of ERsed.
Approach	Use a bottom-up approach to study how wildfire and variable inundation separately and collectively influence interactions among OM properties, microbial metabolic function, sediment-associated primary production, pathways of water delivery, and upstream to downstream hydrologic connectivity within stream networks.
Hypotheses	**H2.1:** The hottest moments of ERsed following re-inundation will emerge when water returns primarily via overland flow, at intermediate levels of stream network connectivity, and when local history of inundation leads to elevated OM thermodynamic favorability and higher diversity of expressed microbial genes linked to C cycling. **H2.2:** Wildfire-driven increases in overland flow will increase variability in ERsed due to increasing ERsed fueled by delivery of terrestrial OM and decreasing ERsed due to less light penetration to sediments. The magnitude of increase in ERsed due to inputs of wildfire-altered OM will be dependent on the diversity of C-cycling genes in sediment-associated microbial communities.
Science Theme 3: Process Integration and ERsed Allometry Dynamics	
Objective	Integrate watershed processes, perturbations, and molecular processes to understand and predict stream-network-scale ERsed as an emergent function of watershed systems.
Approach	Develop a new, hybrid watershed-scale hydro-biogeochemical modeling capability by integrating outcomes from ST-1 and ST-2 into. Update scaling theory that provides quantitative expectations for how cumulative stream network scale ERsed increases with upstream drainage area.
Hypotheses	**H3.1:** Across watersheds, increasing wildfire and variable inundation will lead to increasing role of activated control points, mediated by molecular properties, and increasing temporal variability in cumulative ERsed. **H3.2:** Temporal and between-watershed variation in ERsed spatial scaling will respond most strongly to spatiotemporal patterns of variable inundation at the stream-network scale, with secondary influences from other aspects of biophysical settings like vegetation and wildfires.
Outcome	
Deliver fundamental knowledge and interoperable models based on observations, experiments, and rigorous upscaling of watershed processes—through time and space—that govern emergent hydro-biogeochemical function within and across basins. Foster safe research culture by growing a diverse team.	

C. NARRATIVE

C.1 Background and Justification

C.1.1 Introduction

Watersheds are the foundational unit of Earth's terrestrial environments[1]. Inland waters flow through watersheds and release similar net amounts of C as the land and sea globally, yet estimates of these emissions are simple budgets and are highly uncertain [2,3]. A primary component of inland waters are stream channels that form networks beginning at the upper boundaries of watersheds that connect into larger streams and rivers[4] and coastal environments[5,6]. The biogeochemical function of stream networks is fundamentally linked to their hydrologic connections with hillslope environments. Accordingly, Battin et al. (2023)[7], recently emphasized that stream networks are Earth's largest biogeochemical nexus among land, atmosphere, and ocean environments. Stream networks are, thus, a major biogeochemical reactor within the Earth system.

Accounting for the role of stream networks in Earth system functions requires predictive understanding of these systems. This is challenging, as streams are impacted by many kinds of perturbations, ranging from gradual shifts in climate (e.g., loss of surface water) to episodic extreme events (e.g., wildfire) that can be directly or indirectly caused by humans. Perturbations within watersheds have effects that cascade downhill through overland and subsurface flow pathways[8]. These influences can shift hydrologic connectivity[9] and impact the timing, amount and characteristics of water and materials that reach stream channels[8,10–14]. Interactions between perturbations and the biophysical setting of watersheds (including physical, chemical, and biological attributes and processes) have compounding effects on the coupled hydro-biogeochemical function of these systems[15,16].

> We use 'perturbation' to indicate a temporal shift in system conditions leading to changes in properties of interest. Perturbation is used instead of 'disturbance' because shifts in system conditions are themselves a response to other drivers.

Current conceptual models must be expanded to accommodate the influence of perturbations on stream network hydro-biogeochemistry. Existing conceptual models emphasize different aspects such as hydrologic connectivity with hillslopes, biogeochemical processing in stream channels, and spatiotemporal scaling[17]. The River Continuum Concept [18] emphasizes down-network gradients in coupled hydro-biogeochemical functioning related to factors such as terrestrial inputs and within-stream primary production, shifting organic matter sources across stream orders from mostly terrestrial inputs toward a greater proportion of in stream sources of organic matter. The temporal dimension of how perturbations influence these gradients is captured by the Pulse-Shunt Concept (PSC), postulating that episodic events create 'pulses' that increase transport of materials downstream[19]. Several conceptual models emphasize processes within the water column of stream networks, while others emphasize streambed and hyporheic zone sediments as biogeochemical hot spots[15,20,21]. The river corridor concept takes a systems perspective and emphasizes influences from beyond stream channels and the tight coupling between flowing surface water and both surficial and subsurface sediments[22]. This physical coupling is driven by hydrologic exchange flows (HEFs) that occur when surface water moves into and through sediment pore channels[23]. HEFs heavily influence biogeochemical reactions associated with sediments and contribute to sediment components often being primary biogeochemical reactors of stream networks. Furthermore, recent work from our team indicates that CO_2 produced in stream ecosystems by respiration (ER) is primarily from aerobic microbial reactions associated with sediments, which we refer to as ERsed[24,25]. Here we consider ERsed to include aerobic respiration from all stream ecosystem components besides the water column, such as respiration from heterotrophic organisms in streambed and subsurface sediments, respiration from benthic primary producers, and respiration from aquatic plants rooted in sediments[25]. However, no conceptual models exist to describe spatiotemporal dynamics of ERsed within and across stream networks.

The river corridor concept further emphasizes that HEFs themselves and their biogeochemical influences are impacted by hydrologic connectivity with hillslopes, including the transport of water and other materials such as organic matter (OM). This highlights the need to consider impacts of different types of perturbations, some of which occur along hillslopes and some of which occur within stream channels. With respect to hillslope perturbations, Wymore et al.,[17] and Tank et al.,[26] recently emphasized the need to understand the impacts of wildfires on river corridor hydro-biogeochemistry due to their growing frequency and intensity, especially across the western United States. Relative to other perturbations, wildfires have unique influences mediated through a combination of physical (altered soil hydraulic properties)[27], biological (loss of vegetation and altered evapotranspiration[8]), and chemical (chemistry and quantity of organic matter transported to stream networks[28]) attributes. With respect to in-stream perturbations, there is also growing recognition of the need to understand biogeochemical influences of the loss and re-gain of surface water (i.e., variable inundation) in non-perennial streams. All stream networks across Earth include non-perennial streams and at least 60% of stream reaches go dry every year[29]. This dynamic leads to a broad range of ERsed responses, ranging from 'cold moments' (suppressed ERsed)[30] to 'hot moments' (elevated ERsed)[31]. The impacts of variable inundation on stream network biogeochemistry is, however, often neglected.

Table C.1. The RC-SFA integrates process understanding across a broad range of scales.

Scale	Definition
Batch	Zero-dimensional system focused on fundamental interactions between microbes and biogeochemical processes
Hillslope	Two-dimensional transect from upland to stream that includes vegetation, soil, and subsurface flow paths
Reach	Section of stream network with similar biophysical condition
Stream Network	All stream channels within a given watershed, that may hydrologically connect and disconnect throug time
Watershed	Physical, chemical, and biological processes from ridge lines to receiving waters in coupled surface-subsurface systems
Basin	Multi-watershed, including a wide range of biophysical conditions

Focusing on ERsed as a key and poorly constrained biogeochemical process within stream networks and taking a river corridor perspective, centered on hydrologic connectivity and HEFs, highlights three potential interactions between wildfire and variable inundation. First, wildfires alter the quantity of water flowing into stream networks and may thereby change the dynamics of variable inundation and the delivery of dissolved oxygen (DO) and OM that fuel ERsed[15,23,3231,33,34]. Second, wildfires can alter the pathways through which water travels to stream channels[35–37], which has been recently hypothesized to influence biogeochemical function (including ERsed) upon re-inundation of stream channels[38]. Third, wildfire and variable inundation both alter the concentrations of resources and the chemistry of OM in stream channels[3940,41], which may lead to unanticipated shifts in microbial metabolic pathways and ERsed rates. Wildfire and variable inundation, therefore, have the potential to exert strong controls over ERsed through physical, chemical, and biological mechanisms, but their influences have not been studied together.

Achieving predictive understanding requires knowledge of governing mechanisms and the ability to use that knowledge to scale influences of perturbations through time and space. There are emerging opportunities to do this by leveraging concepts from the metabolic theory of ecology (MTE). MTE views system function as an emergent property of fractal distribution networks that lead to characteristic increases in a functional property with size of the system. Such relationships are known as allometric scaling functions, and the MTE has formalized these relationships in systems ranging from organisms to ecosystems to cities[42]. The fractal nature of stream networks has been used as a parallel to MTE to develop theory for the allometric scaling of stream respiration. This theory predicts how total respiration within a stream network increases with the network's watershed drainage area. Currently the theory is limited to static, idealized conditions without perturbations. This presents an opportunity to extend the allometric scaling theory via incorporation of river corridor mechanisms (from hillslopes to stream channels) through which perturbations impact ERsed. Doing so will provide predictive understanding of how, why, and to what degree perturbations impact temporal variation in the allometric scaling of ERsed in stream networks embedded within river corridors that are further embedded within watershed systems.

Research Vision

The broad range of impacts caused by both wildfire and variable inundation, combined with increasing prevalence of both, highlights a pressing need to develop predictive understanding of their interactive influences over hydro-biogeochemical function. The resulting knowledge must be transferable across divergent biophysical settings to be of use across a broad range of Earth and environmental system science challenges. These considerations motivate us to propose the River Corridor Hydro-Biogeochemistry from Molecular to Multi-Basin Scales Science Focus Area (RC-SFA) as the next evolution of PNNL's watershed system science program. *The RC-SFA will achieve temporally and spatially explicit predictive understanding of how river corridor hydro-biogeochemical function is impacted by interactions among watershed biophysical settings and perturbations (**Fig. C.1.1**).*

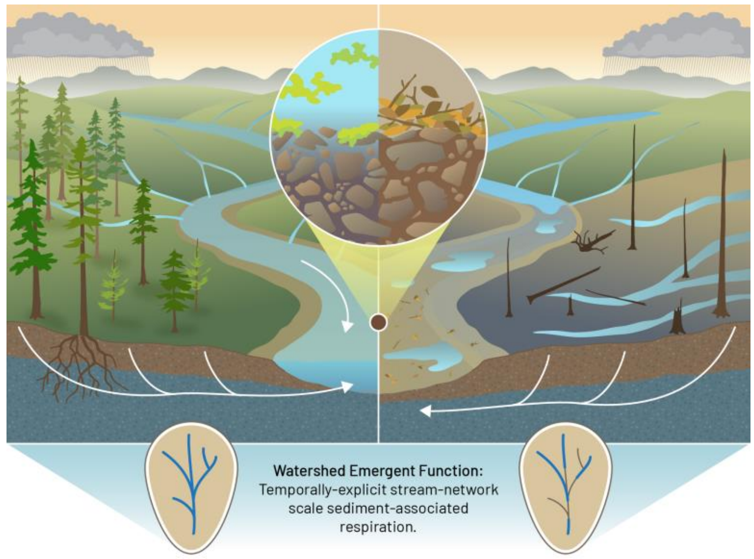


Figure C.1.1. The RC-SFA takes a systems approach to understand and predict hydro-biogeochemical function across divergent biophysical settings, including interactive influences of perturbations across the hillslope-to-stream continuum.

Scientific Grand Challenge: Develop predictive and transferable understanding of hydro-biogeochemical function of watershed systems—across biophysically diverse basins—based on mechanistic knowledge of how perturbations interact with biophysical settings to modulate the movement and transformation of materials from headwaters to the mainstem.

Overarching Science Question: How do watershed biophysical settings, wildfire, and variable inundation interactively govern hydro-biogeochemical function through space and time?

FY21-24	FY25-28	FY29-32
	Increasing transferability	
Hydrologic Exchange Flows	Watershed Biophysical Setting	Climate Resilience
Wildfire	Wildfire and Variable Inundation	Compounding Perturbations
Sediment/Stream Metabolism	Organic matter, microbes, and sediment respiration	Spatiotemporal Scaling of Redox Processes
Multi-Scale ModEx	AI-Guided ModEx	Autonomous ModEx
Foundation for Integrated River Corridor Model	Multi-Basin Emergent Function	Multi-Basin Emergent Function

Figure C.1.3. The RC-SFA's proposed science themes (FY25-28) build from the current quadrennial period (FY21-24) and lay a foundation for our longer-term vision that includes the subsequent quadrennial period (FY29-32).

Long-Term Research Roadmap: The long-term vision of the RC-SFA (**Fig. C.1.3**) is to deliver the knowledge, models, and data needed to quantify, understand, and predict how perturbations interact with biophysical settings to generate emergent hydro-biogeochemical functions of watershed systems. We will progressively increase the transferability of our science outcomes, which is foundational for achieving ESS's goal of predictive understanding. Our hypothesis-based approach will develop new process understanding embodied in multiscale quantitative models. Figure C.3 summarizes the key elements of scientific study in the current, proposed, and future quadrennial periods. Our long-term vision is well-suited to the SFA research model, as it requires a multidisciplinary team, access to state-of-art instrumentation and computational capabilities, and the ability to integrate field and laboratory experiments with numerical models to build quantitative understanding of complex system behavior. The RC-SFA approach in combination with Promoting Inclusive and Equitable Research (PIER) plan (**C.6**) implementation merges the stability and diversity of long-term capability stewardship and staff development at multi-program national laboratories, access to DOE user facilities, and the flexibility of university collaborations to fill gaps in expertise and provide educational opportunities to early career scientists.

Near-Term Research Plan: In support of our long-term vision, in the next four years we will expand the breadth of watershed processes considered in our work to include greater consideration of hillslope vegetation, watershed-level hydrologic connectivity, drivers and impacts of variable inundation, and spatiotemporal scaling of ERsed. These expanded components build from our continued emphasis on Ersed as influenced directly and indirectly by wildfire, HEFs, OM chemistry, and microbial communities. Research outcomes will provide a foundation for longer-term efforts focused on a broader suite of biogeochemical processes beyond Ersed (e.g., methanogenesis) as influenced by a broader suite of perturbations beyond wildfire and variable inundation (e.g., future climate change) (**Fig. C.1.3**).

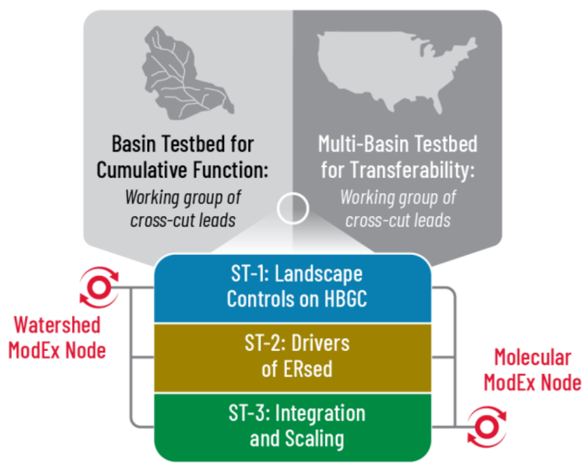


Figure C.1.4. The RC-SFA will generate transferable outcomes by working across multiple watersheds within a single basin testbed and additional watersheds across the CONUS as a multi-basin testbed. Project teams are organized into cross-cutting themes that work across both testbeds and jointly implement the project.

To pursue knowledge and models that are applicable across diverse biophysical settings (i.e., to achieve transferable outcomes), the RC-SFA will work within and across basins within CONUS (**Fig. C.1.4**). Basin-scale work will generate deep process understanding and development of models based on hillslope-to-stream integration. Multi-basin work will use CONUS as a transferability testbed focused on a batch, reach, and stream network scale processes. Within and across basins, we will continue to work heavily with the broader science community and other interested parties to facilitate evaluation of science questions and hypotheses while finding ways to provide value to those beyond our core team, as described in our PIER Plan (**C.6**). The project is further organized into three linked Science Themes (STs) that will address science questions through hypothesis-based investigations (**C.3**). The integration of the STs is in part facilitated by working groups of cross-cut leads (**C.4, C.7**). The specific questions and hypotheses to be pursued in the STs are motivated by existing knowledge gaps (**C.1.2**) and advances made by our team during the current quadrennial period (**C.2**).

We use 'interested parties' in place of 'stakeholder' to be inclusive of rightsholders and avoid implications of ownership.

C.1.2 Knowledge Gaps

In this section we describe key knowledge gaps that need to be resolved to understand how watershed perturbations and biophysical settings combine to drive ERsed (**Fig. C.1.6**). Below we summarize each knowledge gap at a high level and detail specific aspects that we plan to address. Outcomes from addressing these knowledge gaps will advance fundamental understanding and predictive capacity, thereby (1) enabling DOE to provide scientific leadership for informing inter-agency process-based watershed management and (2) providing a foundation for representing watershed hydro-biogeochemistry in the land component of Earth System Models (ESMs).

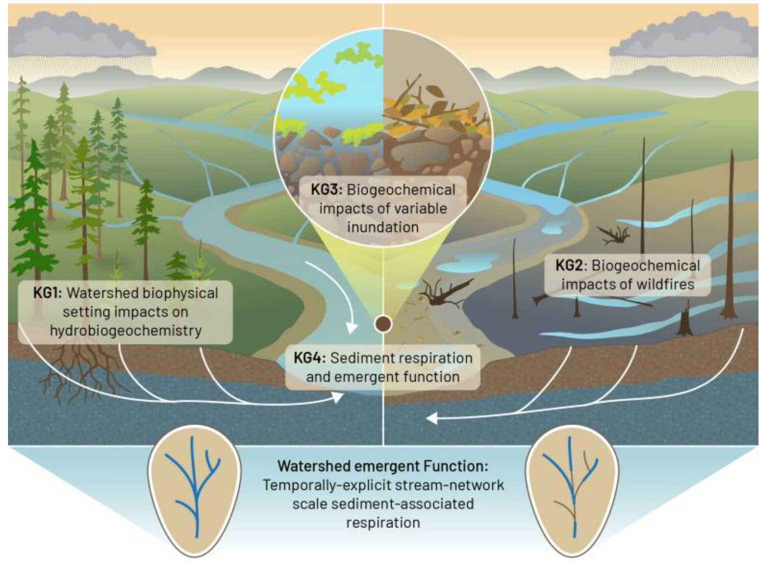


Figure C.1.6. The RC-SFA will resolve an integrated set of knowledge gaps that jointly reveal how watershed processes and perturbations outside of stream networks influence processes and perturbations within stream networks, and further how biogeochemical function emerges from that integrated suite of drivers and responses.

Knowledge Gap 1: Watershed biophysical setting impacts on hydro-biogeochemistry. Biophysical aspects of watersheds can directly and indirectly influence hydrologic and biogeochemical cycling, consequently altering the delivery of water, nutrients, and OM to the stream network. Our current knowledge is not sufficiently holistic to understand, quantify, and predict how interactions among watershed biophysical factors cascade through space and time to influence water and material transport/transformation through hillslopes and to stream networks as connected watershed sub-systems [9].

KG 1.1 (Hydrologic connectivity) – Hydrologic connectivity (**Fig. C.1.7**) describes the connections through which disparate regions on hillslopes are linked with streams and rivers via surface and subsurface water flow paths[43,44]. Spatiotemporal variation in how and how much water moves from hillslopes to stream networks influences surface water dynamics and velocity in streams, including variable inundation. In addition to water, hydrologic connectivity controls the transport of OM and nutrients from the land surface and subsurface to stream networks[9]. Our understanding of how hydrologic connectivity influences lateral and vertical export of terrestrial OM to inland waters in undisturbed watersheds is limited, especially at basin and multi-basin[3]. Lateral transport through the subsurface is influenced by subsurface hydrogeology (e.g., 3-dimensional permeability fields), but it's currently missing from models applied at basin to multi-basin scales. This undermines our capacity to correctly account for the hydrologic connectivity between hillslopes and stream networks. Resolving this gap is critical to improving estimates of material fluxes to inland waters in process-based models, which in turn can inform further field efforts[9,45].

KG 1.2 (Vegetation) – In combination with subsurface hydrogeology, vegetation is a primary regulator of hydrologic connectivity through its temporally and spatially varying uptake of soil water for transpiration,

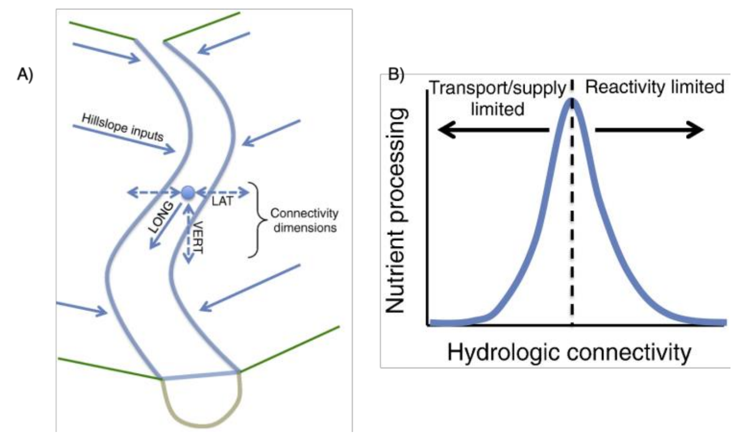


Figure C.1.7. A) Hillslope-to-stream hydrologic connectivity is driven by reactive transport over land and through subsurface soils, inputting OM and nutrients across the stream network. Connectivity is heavily influenced by subsurface hydrogeology and can be modified by hydrologic interactions across lateral, vertical, and longitudinal gradients and reactivity along the stream network. B) The relationship between hydrologic connectivity and biogeochemical cycling is modulated by supply to and reaction rates in the stream network. Modified from [9].

its redistribution of water within the soil profile, and its impacts on evaporation and surface runoff[8]. Transpiration returns most precipitation to the atmosphere[46], thus even small changes in transpiration have large impacts on water balance, hydrological connectivity, and variable inundation. Photosynthesis is the source of organic C to watersheds, and thus vegetation also influences the quantity and quality of OM. Theory regarding vegetation impacts on connectivity is mature but largely untested[8], leaving us with large predictive uncertainty. Connectivity through subsurface flow paths should vary predictably with vegetation traits, such as declining connectivity with greater rooting depths, leaf area, and hydraulic conductivity, among other traits[8]. Quantifying and understanding the direct role of vegetation on connectivity and OM chemistry is a key task required to improve models of watershed hydro-biogeochemistry. This leads to fundamental knowledge gaps of how vegetation influences connectivity and OM chemistry.

KG 1.3 (Perturbations) – Increasing rates, intensity, and spread of event-based perturbations (e.g., wildfire) along with chronic environmental change (e.g., warming) could increase or decrease hydrologic connectivity between hillslopes and stream networks, with follow on impacts to variable inundation and OM delivery[8]. Myriad perturbations affect watershed systems, but wildfire is unique in directly impacting both hydrologic connectivity via changes to soil hydrologic conductivity and OM source chemistry via heat-induced transformations. Further, the influence of wildfire on stream network biogeochemistry has increased over recent decades[47]. Wildfires are also globally ubiquitous, yet despite their increasing frequency and severity in many regions, knowledge gaps remain in terms of the mechanisms through which wildfires influence hydrologic connectivity, variable inundation, and OM delivery to stream networks[48]. There are two major challenges to resolving these knowledge gaps: first, wildfires create spatially heterogenous mosaics of alterations to terrestrial landscapes[49], and second, process-based understanding of the alterations to transpiration and evaporation following wildfire on watershed hydrologic responses is inadequate. Furthermore, following wildfire there are temporal dynamics in hydrologic connectivity, variable inundation, and OM transport as watershed systems recover[50]. Mechanisms driving these dynamics are poorly understood, and predictive understanding of these dynamics that applies across divergent biophysical settings is lacking.

Knowledge Gap 2: Biogeochemical impacts of wildfires in the stream network. Wildfires are a terrestrial disturbance that alter watershed biophysical properties which in turn influence in-stream hydro-biogeochemistry [51,52]. Current knowledge is insufficient to quantify, understand, and predict how wildfire impacts cascade through space and time to influence ERsed due to processes discussed in KG-1.3, such as: (*i*) changes in the chemistry and quantity of material transported from watershed hillslopes into stream channels and (*ii*) changes in hydrologic connectivity and influence of velocity and flow paths returning water to stream channels.

KG 2.1 (ERsed post-fire) – Variability in whole stream respiration, and thus ERsed, can be influenced by biological, physical, and chemical features and processes [53], all of which are altered by wildfire [48]. Few studies have estimated whole-stream respiration after fire, and those that have shown highly variable influences of wildfires on respiration (including increases, decreases, and no change; **Section C.2.2.4**) [41,54–56]. Wildfire-induced changes in physical properties of soil (as summarized in KG 1), water chemistry (e.g., pH, turbidity), and OM quantity and quality (in the dissolved and particulate phases) may interact with each other and lead to non-linear shifts in ERsed, including increased or more variable respiration rates post-fire [41,54,55]. For example, changes to fine sediment delivery with wildfire have previously been linked to shifts in whole stream respiration[57]. Altered material transport to stream networks after wildfire can persist for years, altering water velocity and water delivery mechanisms that influence ERsed through scouring, light availability, and nutrient/OM delivery [41,54]. For example, Robinson et al. (2005)[41] found elevated ERsed rates for 20 years following fire, related to loosely bound OM in sediments. The mechanisms and transferability of this observation have not been revealed. There is, therefore, a lack of transferable understanding on how ERsed responds to wildfire and what mechanisms underlie those responses (**Fig. C.1.8**).

KG 2.2 (OM chemistry post-fire) – Wildfires can significantly alter the composition OM transported to streams [58–60](**C.2.2.2**). As wildfires alter terrestrial landscapes to greater extents, shifts in OM chemistry follows. Collectively, this fire-altered pool of OM is called pyrogenic OM (PyOM). First, depolymerization of terrestrial OM macromolecules occur, and with increasing thermal transformations, the OM then becomes recondensed and more polyaromatic[59,61–63]. As wildfires heterogeneously impact the landscape, they produce PyOM with varying chemistries. PyOM spans a range of thermodynamic properties which can influence PyOM's potential bioavailability in the environment[64]. Aqueous incubations of PyOM suggest large changes in PyOM chemistry due to microbial processing [65], thus PyOM has the potential to alter the relationships between OM and ERsed post-fire. Together, this suggests that understanding how wildfires influence ERsed requires knowledge of OM composition and input post-fire. Therefore, specific PyOM compositional shifts that occur due to an interaction between wildfire and landscape type [59] may influence the mechanisms behind when and where there are hotspots of ERsed in wildfire-impacted systems. Building upon knowledge that streams receive and actively process PyOM[66,67], the linkage between these changes, how they are modulated by watershed biophysical settings, and the resulting impacts on ERsed is unexplored and remains a knowledge gap.

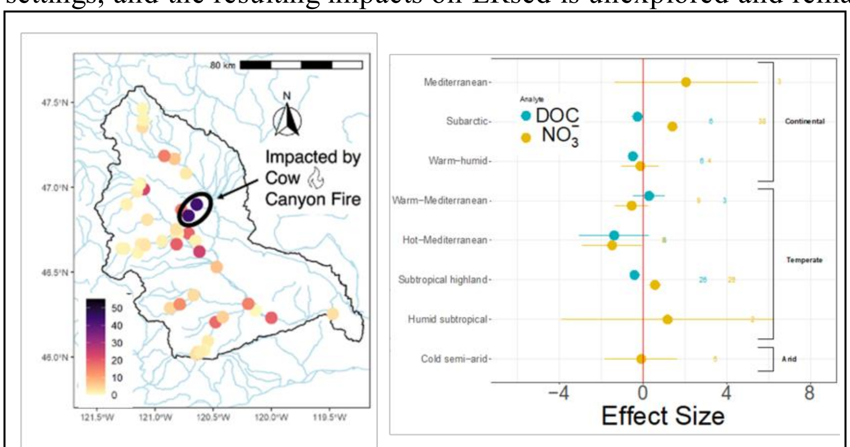


Figure C.1.8: Batch-scale estimates of ERsed from the Yakima River Basin suggest that wildfire may lead to ERsed hot spots (Left). The transferability of this result is unknown. A meta-analysis indicates cross-biome variability in impacts of wildfire on concentrations of C and N in streams (Right). The mechanisms underlying this variation these responses are poorly understood.

KG 2.3 (Microbial communities post-fire) – Wildfires can alter the composition and expressed metabolism of microbial communities in the terrestrial landscape[68]. Current understanding of shifts in microbial composition and metabolic processes after wildfire is based primarily on soil systems, with limited studies in the river corridor (though see Bistarelli et al. [67]). Direct linkages between in-stream OM chemistry, microbial enzymatic activity, and biofilm processing of aromatic PyOM was observed in streams where PyOM was experimentally introduced[67]. Thus, the microbial processing underpinning the degree of transformations of PyOM and fueling ERsed are likely important yet unknown contributions to stream metabolism in fire-impacted systems.

Knowledge Gap 3: Biogeochemical impacts of variable inundation. Current knowledge is insufficient to quantify, understand, and predict spatial and temporal variation in the impacts of the loss and regain of surface water (i.e., variable inundation) in stream networks on (*i*) ERsed in terms of absolute rates, fractional contributions, and spatiotemporal scaling, (*ii*) the role of chemistry of sediment-associated OM, and (*iii*) the composition and metabolic expression of sediment-associated microbial communities.

KG 3.1 (ERsed following re-inundation) – There are no published studies that have estimated field-scale ERsed following re-inundation of stream reaches. A small number of studies have estimated whole-stream respiration following re-inundation[69,70], but none have parsed out the relative contributions of ERsed. We anticipate that ERsed following re-inundation will vary markedly across stream networks and through time within stream networks. These expectations are based on three key observations. First, lab-estimated ERsed following re-inundation vary by at least three orders-of-magnitude across CONUS (**Fig. C.1.9**). This variation is likely to translate to reach-scale rates as it appears to be linked to field-scale parameters that vary significantly across streams, such as sediment texture. Second, dry conditions inhibit benthic primary production and it takes time for those organisms to reestablish once flow has returned [14]. Given the strong influence of benthic primary production on ERsed (**C.2.2.4**), the time course of recovery for primary production will likely lead to strong temporal variation in ERsed. Third, a recent conceptual model posits that biogeochemical rates following re-inundation are modulated by the pathways through which water returns to dry stream channels [38]. Given that there are no field-scale estimates of ERsed following re-inundation in stream networks, there are also no estimates for the quantitative influences of variable inundation on temporal variation in the spatial scaling of ERsed. There are initial estimates for how much total CO_2 flux could be contributed by non-perennial streams [71], but these estimates are not mechanistic and do not specifically parse out the role of ERsed. Other recent scaling efforts aim to quantify how stream-network-scale cumulative biogeochemical rates increase with upstream drainage area, which is known as allometric scaling [72]. The associated theory assumes temporally static conditions and idealized stream networks [72]. Our recent work has started to relax these assumptions (**C.2.2.7**) and while variable inundation will likely drive significant temporal dynamics in allometric scaling of ERsed, such influences remain unknown.

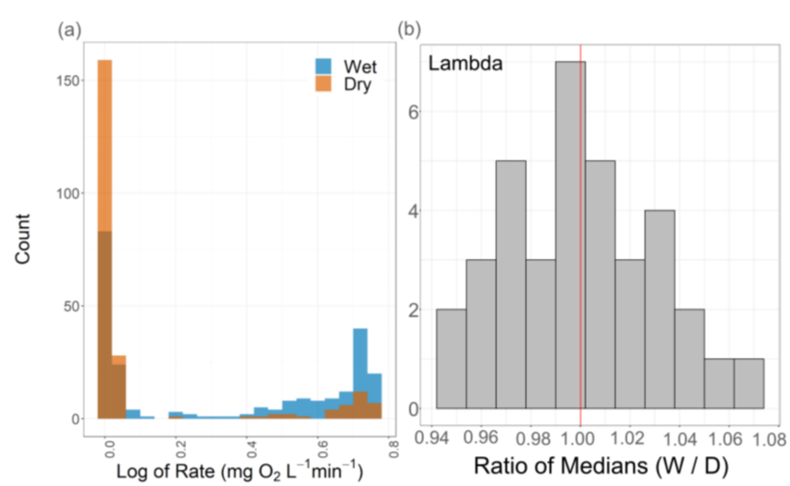


Figure C.1.9: (a) CONUS-wide batch-scale ERsed in variably inundated (dry) and consistently inundated (wet) experimental treatments. (b) Distribution of site-specific effect size for impacts of variable inundation on OM thermodynamics, as measured by lambda as a proxy for efficiency of conversion to microbial biomass.

KG 3.2 (OM chemistry following re-inundation) – Variable inundation of streambed sediments can significantly alter OM chemistry/composition and thermodynamics (cite). Upon re-inundation, terrestrial OM inputs and the abundance of more labile high molecular weight compounds increases[73–75]. Changes in OM composition are thus a function inundation history (i.e., drying duration and frequency) where intensifying drying conditions show an increase in humic-like compounds and a decrease of protein-like compounds[76]. Re-inundation after dry periods impact OM chemistry and riverine biogeochemical processes more profoundly than other hydrologic events[73,74,77]. Additionally, a significant body of literature further shows that OM chemistry/composition and thermodynamics strongly influence ERsed (**C.2.2.5, C.2.2.6**). Studies linking OM thermodynamics to respiration have primarily been done with sediments from perennial streams[30,78–81]. The same thermodynamic properties can be altered by drying and re-inundation of sediments [30], but the qualitative direction and quantitative magnitude of these influences vary markedly across CONUS stream systems (**Fig. C.1.9**). The degree of influence can also depend non-linearly on the history of drying and re-inundation [30]. The largest studies-to-date have been done under controlled laboratory conditions with simple drying/inundation experimental treatments. It is unclear how these lab-based outcomes relate to field conditions. In turn, there is insufficient data to elucidate linkages between *in situ* histories of inundation and OM thermodynamics. Another knowledge gap for which there is little information is the degree to which *in situ* variable inundation alters the link between OM thermodynamics and ERsed.

KG 3.3 (Microbial communities following re-inundation) –Variable inundation can alter the composition of active microbial communities[82]. There is evidence from a lab study that drying/re-inundation dynamics over the time scale of weeks does not strongly alter which microbes are present but does significantly alter which microbes are metabolically active [30]. However, microbial composition and metabolism might shift as a function of inundation history (i.e., drying duration and frequency) and further influence ERsed. Field experiments show significant differences in sediment microbial communities across three hydrological conditions (i.e., inundated, two months of being dry and one month after re-inundation[82]). Additionally, lab experiments show that less intense and shorter drying periods lead to an ERsed pulse and a more rapid and drastic change in sediment microbial composition and diversity, while more intense and prolonged drying periods show a slower change in microbial composition and a lag in ERsed for the first 24 hours after flow resumption[83,84]. Across non-perennial streams, we expect marked variation in which genes are expressed by microbes even if metabolic potential does not vary strongly. More poorly understood is how a change in gene expression—due to variable inundation—impacts ERsed, potentially in coordination with changes to OM chemistry and thermodynamics. Until the development of the Genome Resolved Open Watersheds (GROW) database (**C.2.2.6**), the metabolic potential and expressed metabolisms of microbial communities from stream systems were very poorly characterized and understood [85,86]. Potential and expressed metabolisms of microbial communities are often assayed via shotgun metagenomic (metaG) and metatranscriptomic (metaT) sequencing [87]. With respect to these two data types, stream systems were, prior to GROW, one of the least sampled/characterized ecosystem types on Earth[85]. GROW is the planet's largest database of metaG and metaT data for stream/river systems [86], but it currently lacks data from non-perennial streams. The lack of metaG/metaT data from non-perennial streambed sediments precludes the development of transferable understanding for the influence of variable inundation on microbial communities in these systems.

Knowledge Gap 4: Drivers of ERsed and perturbation impacts to ERsed allometry dynamics. Current knowledge is insufficient to quantify, understand, and predict spatial and temporal variation in how watershed biophysical settings interact with perturbations beyond and within stream networks to collectively influence ERsed. This is true with respect to fractional and absolute rates, underlying processes, and temporal dynamics of spatial scaling.

KG 4.1 (Rates and drivers) – There is significant cross-stream variation in ERsed rates and the fractional contribution of ERsed to whole-stream-system respiration (**C.2.2.4**). Incorporating our recent work in the YRB with estimates from previous literature, the fractional contributions from sediment range from 3-100% of whole-stream-respiration. This broad range may be due, in part, to more than 1000-fold variation across CONUS in the rates of batch-scale ERsed (**Fig. C.1.10**). The batch-scale ERsed estimates come from laboratory incubations using WHONDRS-collected fine-grain (<2mm) sediments from over 200 locations (**C.2.2.5**). While machine learning models have achieved 80% predictive skill across CONUS (**Fig. C.1.10**), mechanistic understanding and associated models have lagged. There is also a relative paucity of field-scale estimates, which leads to poor understanding for how rates from fine-grain sediments and their underlying drivers contribute to reach-scale ERsed. Our recent basin-scale work indicated strong influences of sediment-associated gross primary production and physical attributes related HEFs (**C.2.2.4**). The multi-basin transferability of those results has not been evaluated. In turn, temporal and spatial variations in reach-scale ERsed are poorly characterized and understood. Furthermore, most studies of ERsed are focused on processes occurring within stream networks and do not explicitly consider indirect influences of hillslope processes or perturbations external to stream networks. This is a major knowledge gap as perturbations may alter the primary drivers of ERsed. For example, increased turbidity following wildfire may indirectly cause ERsed to be limited by low light availability to benthic primary producers, instead of being limited by HEFs (as in Son et al. 2022). The processes that drive ERsed ultimately drive cumulative, network-scale ERsed and associated allometric scaling.

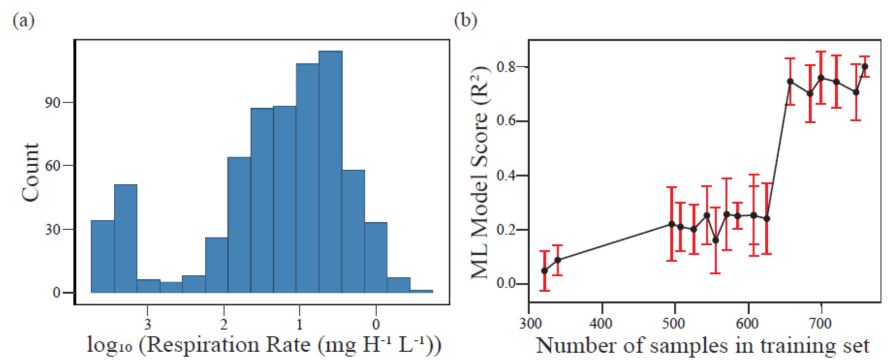


Figure C.1.10. (a) Distribution of respiration rates across CONUS. (b) Machine learning (ML) model performance across ModEx iterations as additional samples were collected.

KG 4.2 (ERsed allometric spatial scaling) – Understanding variation in the function of stream networks requires quantitative approaches that allow comparative analyses across biophysical settings. A powerful approach is to quantify how cumulative network-scale biogeochemical rates increase with upstream drainage area[72,88]. This can be conceptualized as an allometric spatial scaling function, which has conceptual alignment with the metabolic theory of ecology (MTE)[42]. MTE views organismal metabolism as an emergent property of fractal distribution networks such as circulatory systems of animals and fractal branching structures of tree canopies[89]. This is the foundation for allometric scaling laws in biology, whereby metabolic rate increases to the 3/4 power of body size[90]. Stream networks are also fractal distribution networks[91] and taking an allometric approach allows for the integration of physics-based biological theory into watershed systems science[72]. This approach further links observed ERsed allometric scaling to robust theory, as opposed to simply characterizing a statistical scaling function. For example, downstream gradients in per-area reach-scale ERsed should lead to steeper allometric scaling slopes[72]. Observations that align with this expectation provide support for the underlying theory and can be used to expand upon river corridor conceptual models summarized in **C.1.1**. Quantitative aspects of ERsed allometry are, however, only beginning to be examined (**C.2.2.7**). The primary factors governing among-watershed variation in ERsed allometric scaling parameters (slope, intercept, R^2) are poorly known. Our recent study indicates that HEFs influence scaling parameters[88], but the underlying point-scale model predictions do not align with inverse modeling estimates of ERsed[25] (**C.2.2.4**). This leads to significant uncertainty in the estimated allometries. Furthermore, despite dominant influences of primary production and HEFs on ERsed[24,25] (**C.2.2.4**), no study has yet considered both system attributes when estimating ERsed allometry. We also lack mechanistic models that can operate at watershed scales and represent the physical, chemical, and biological processes and properties that govern allometric scaling of ERsed.

KG 4.3 (ERsed allometry dynamics) – With respect to ERsed allometry, the current state-of-science is based on a theoretical and mechanistic foundation in which governing variables do not change in time[72,88]. While a useful place to start, there is a need to integrate time-varying components of watershed systems (e.g., event-based perturbations) into our understanding of ERsed allometry. Largely lacking are foundational knowledge, data, models needed to understand how temporally dynamic and spatially variable perturbations such as wildfire and variable inundation lead to ERsed allometry dynamics. To our knowledge, no study has tried to quantify or predict ERsed allometry dynamics. As a corollary, no studies have attempted to mechanistically understand or predict how ERsed allometry dynamics emerge from the integration of watershed biophysical settings and interacting perturbations within and outside of stream networks.

C.1.3 Mission Alignment

The overarching objective of the ESS program is to advance "…an integrated, robust, and scale-aware predictive understanding of terrestrial systems and their interdependent biological, chemical, ecological, hydrological, and physical processes." Our research plan is strongly aligned with this objective, with an emphasis on developing predictive, transferable understanding of the hydro-biogeochemistry of watershed systems from reaction to among-basin scales. We will observe, experimentally study, and numerically model watershed systems using an approach that iterates between model-guided data generation and data-guided model advancement (i.e., iterative ModEx). Our research is focused directly on the priority research objectives of ESS as indicated below (paraphrased ESS objectives in italics; related SFA research emphases in plain text):

ESS Priority 1: *Quantify how biotic and abiotic processes control the mobility of contaminants, nutrients, and key biogeochemical elements.* We will elucidate how the biological (e.g., microbial gene expression), chemical (e.g., OM chemistry), and physical (e.g., topographic gradient) environment of watersheds control microbial transformations of OM within and across environmentally diverse stream networks.

ESS Priority 2: *Quantify and predict how hydrology drives fine-scale biogeochemical processes in surface-subsurface systems.* We have a strong emphasis on hydrology and hydrologic connectivity across surface-subsurface systems as controlling factors over all aspects of our biogeochemical investigations, spanning reaction to multi-basin scales.

ESS Priority 3: *Translate biogeochemical behavior across molecular to watershed scales to predict flows of water, nutrients, and contaminants.* We will deliver quantitative and mechanistic understanding of the time-integrated watershed-scale ERsed, as predicted by process-based models that explicitly couple molecular-scale biogeochemical transformations to watershed-scale hydrology.

ESS Priority 4: *Identify, quantify, and predict watershed responses to natural and anthropogenic perturbations.* We will reveal how influences of perturbations driven by natural factors and indirectly via anthropogenic factors (e.g., climate change) cascade through space and time to modify watershed hydro-biogeochemical function, including network-scale cumulative ERsed as an emergent property.

ESS Priority 5: *Translate predictive understanding of watershed system function and evolution into near- and long-term environmental and energy strategies.* Our research has implications for watershed and stream network management, including understanding and being able to predict impacts of ubiquitous and increasingly intense perturbations on integrated hydro-biogeochemical functions that influence water quality and quantity.

We actively engage with ESS program activities by partnering with the IDEAS-Watersheds and ExaSheds projects, participating in ESS Cyberinfrastructure Working Groups, and collaborating with other ESS investments via WHONDRS. We will continue active collaborations across BER-supported facilities and resources, including EMSL, JGI, KBase, NMDC, and ESS-DIVE. We will also continue to leverage intersection points of mutual benefit with other ESS-funded SFAs focused on watershed systems.

Our research is also well-aligned with the objectives of the larger EESSD program, which emphasizes the interplay among hydrology, biogeochemistry, and the function of human and Earth systems. The Earth and Environmental Systems Modeling program is developing innovative modeling capabilities for the integrated Earth system, such as the Energy Exascale Earth System Model (E3SM). Our research uses several codes related to E3SM, and we will fill a key gap in the land component of that model, namely the representation of watershed and stream network hydro-biogeochemical processes and how they are impacted by several globally relevant perturbations. Our SFA science advisory board (**C.7**) includes E3SM scientists to enable connections and synergy.

C.1.4 Key Outcomes

Watersheds are the foundational unit of terrestrial systems [1] and within watersheds, stream networks couple surface and subsurface processes and act as Earth's central nexus linking atmosphere, land, and sea [7]. Diverse natural and human processes influence emergent hydro-biogeochemical function of these systems. Changes in perturbation regimes in terms of their frequency, intensity, and duration alter function and lead to a critical need for transferable scientific understanding embodied in models capable of predicting how perturbations modify hydro-biogeochemical dynamics. The SFA's grand challenge, long-term vision, and near-term plans are designed to address these needs.

We will emphasize the following key themes for which new knowledge is critically needed:

- Fundamental controls of watershed biophysical setting on stream network hydro-biogeochemistry, as mediated by hydrologic connectivity and material inputs to stream networks, with emphasis on vegetation, soil/sediment physical properties, climate, and geomorphology.
- Within-stream-network hydro-biogeochemical impacts of watershed perturbations that occur outside stream networks, with emphasis on how impacts of wildfire and associated vegetation dynamics are modulated by urban/agricultural land use and climate-related variables (e.g., vapor pressure deficit) that are part of watershed biophysical settings.
- Biogeochemical impacts of variable inundation as a hydrologic perturbation that occurs within stream networks, with emphasis on how the history of inundation influences sediment-associated OM chemistry, and microbial community gene expression.
- Integrated influences of watershed biophysical setting, perturbations, and molecular properties on spatial and temporal variation in ERsed, including absolute rates, fractional contributions to whole-stream respiration, and time-integrated spatial scaling of cumulative ERsed.

The cornerstone of prediction is transferable knowledge, and in context of ESS this means knowledge that applies across biophysically diverse watershed systems. Key to achieving transferability is engagement with a broad group of researchers working across biophysically divergent environments and open sharing of ideas, data, knowledge, and models. It is also vital to work in a Coordinated way (per ICON) to design studies and data for interoperability and synthesis. The SFA has demonstrated leadership in establishing new paradigms for community interactions via ICON-based participatory science via WHONDRS and sophisticated FAIR data publishing/management approaches. The proposed research leverages our PIER plan to build on that momentum to establish new and stronger collaborations among researchers, groups, institutions, and agencies working across watersheds within the CONUS. This allows for a nimble and highly distributed approach for evaluating the degree to which place-based outcomes are consistent across biophysically divergent watersheds/basins. We cannot rely exclusively on highly distributed approaches, however, as not all science questions can be evaluated using the relatively simple methods that are inherently required when working across numerous CONUS watersheds. A central tenet of our approach is, thus, to merge intensive place-based approaches, lighter-touch distributed science, and ModEx to achieve transferable outcomes.

An important form of transferability is the embodiment of diverse data types and robust process understanding in data-driven and mechanistic quantitative models. Current models do not integrate watershed biophysical setting, watershed perturbations that are external to stream networks, and perturbations that occur within stream networks to achieve predictive understanding of emergent function. One reason is that the state-of-knowledge is insufficient to robustly develop such models. We will address both knowledge and model limitations by providing new understanding and next-generation multiscale modeling capabilities that can ultimately be used to predict watershed function under contemporary and future environmental conditions related to multiple types of perturbations. Outcomes from our work will be useful across a broad range of interested parties because our data, knowledge, and models will be mechanistic enough to be relevant for place-based challenges and at spatial scales large enough to be relevant to Earth-system scale challenges. Mutually benefit of our outcomes will be further enhanced by engaging with interested parties to ask if they are useful and if they can be improved, aligning with Networked (per ICON). For example, our science goals require development of mechanistic models to estimate influences of variable inundation on time-integrated stream-network-scale ERsed. To do that we must be able to simulate where and when streams go dry and re-inundate. Other groups can use such models to predict where and when fish habitat is lost entirely or becomes fragmented, even though that is not our objective. That same model can be used to estimate time-integrated spatial scaling at watershed to basin scales of ERsed based on watershed features for which there are abundant spatially continuous data (e.g., vegetation composition, topography, soil properties, wildfire history, land use). In turn, large-scale models can use our outcomes to benchmark simulations and/or provide sub-grid parameterizations of watershed hydro-biogeochemical function.

C.1.5 SFA Team Uniquely Qualified

The SFA team (**C.7, G**) is uniquely qualified to generate fundamental knowledge of processes that govern watershed function and to translate that knowledge into predictive hydro-biogeochemical models applicable across environmentally diverse basins. Our team has demonstrated expertise (**C.2**) needed for successful measurement of *in situ* conditions (e.g., reach-scale ERsed), experimental manipulation (e.g., variable inundation in batch reactors), modeling across scales (e.g., molecular OM thermodynamics to basin-scale hydro-biogeochemistry), and their integration (e.g., AI-guided sampling with iterative ModEx across CONUS). We have simultaneously led WHONDRS, gone through 18 full ModEx cycles in 1.5 years, and developed the YRB into a basin-scale testbed. The result is our team having unique expertise in merging highly distributed participatory science, intensive place-based approaches, and both data generation and modeling components of ModEx. Bringing together this collection of skills is vital for enabling transferable science outcomes applicable across biophysically divergent watersheds and basins. In addition, our team participates in IDEAS and ExaSheds, providing opportunities to exchange knowledge and models between the SFA and these other EESSD investments. External collaborators (**H**) were selected for their ability to deliver high-impact science related to hillslope-to-stream network integrated modeling, watershed perturbations, and ERsed in context of variable inundation. The team benefits from collaborations with EMSL, JGI, KBase, NMDC, and ESS-DIVE (**H.4**) to leverage cutting-edge approaches needed to reveal how molecular-scale attributes lead to emergent biogeochemical function in context of variable hydro-biogeochemical conditions. As described in our PIER Plan (**C.6**), we equip our team with infrastructure, training, mentoring and professional development to leverage the team's qualifications into sustainable action, which relies on maintaining a culture of trust and belonging.

Conceptual advances, model developments, and capability innovation by the SFA team provide a foundation for pursuing a process-based reaction- to among-basin modeling framework. This places the team in a unique position to make rapid progress toward understanding processes governing watershed and stream network hydro-biogeochemistry, to use that knowledge to develop a new paradigm in mechanistic modeling, and to generate mechanistic and quantitative understanding of emergent watershed function. The team will integrate our capabilities using an iterative ModEx approach applied at reaction, reach, basin, and multi-basin scales to produce transferable knowledge embodied in predictive models that can inform watershed management and ultimately the representation of watershed hydro-biogeochemistry in Earth system models.

C.2 Research Progress for FY2021-2024

C.2.1 Progress Summary

The RC-SFA continues to lead-by-example[92,93] and help the ESS community towards (1) increasingly sophisticated use of ICON principles via leadership of the ICON Science Cooperative; (2) robust FAIR data management via heavy use of ESS-DIVE community reporting formats and deeply documented ESS-DIVE data packages (DPs), in alignment with ICON's Coordinated and Open principles (**C.5**); (3) highly-distributed participatory science via leadership of the WHONDRS consortium, in alignment with ICON's Networked principle; (4) truly iterative ModEx based on the 'hypothesis handshake' whereby modelers provide untested hypotheses to data generators and data generators provide new hypotheses to modelers, in alignment with ICON's Integrated principle; and (5) heavy engagement with BER-led workshops and meetings such as AI4ESP and the ESS PI meeting. Data packages (DPs) submitted by this SFA account for approximately 40% of all ESS-DIVE DPs that use the ESS-DIVE community reporting formats and almost 70% of all ESS-DIVE DPs using international generic sample number (IGSN) unique identifiers. These noted activities are critical for expanding DOE's leadership in environmental system science in general and watershed system science in particular. They also provide a foundation based on mechanistic understanding and predictive models to enable cooperation with other agencies to develop a national science capability to support integrated watershed management and operations.

Our team has a significant publication record (Appendix 1), with 123 peer-reviewed papers published in the current four-year period and 26 additional manuscripts in review. Some of these publications are led by researchers beyond our core team that reach out to leverage our broad range of expertise and capabilities. Further, our publications are in high-quality journals; five are in DOE-designated high-impact journals (Science, Nature Series, PNAS), 35 are in ISI-designated top-10 journals, 69 are in first-quartile journals, and 98% are in journals with impact factors above the median in their field.

Research proposed for FY25–28 builds on conceptual and capability progress during the current four-year period, as summarized by the following progress highlights. They are organized in parallel to our research plan (**C.3**), from hillslope-scale processes and perturbations, to stream channel molecular processes and ERsed at batch to reach scales, and ending with allometric scaling of ERsed across basins (**Table C.2**). Advances in each highlight motivated us towards specific knowledge gaps and/or technical approaches, which in turn motivated our Research Plan (**C.3**) questions, hypotheses, and methods tied to specific Science Theme (ST) Activities (e.g., 1A, 1B, etc.).

Table C.2. Highlights. Overview of progress highlight outcomes, which knowledge gaps (KG) they motivated, and which proposed Science Theme (ST) Activities they most influenced in terms of questions, hypotheses, and/or methods.

Progress Highlight	Summary of Outcomes	Associated KG (C.1.3)	Associated ST Activity (C.3)
Section C.2.2.1	Linked watershed biophysical properties, watershed perturbations, and hillslope-to-stream hydrologic connectivity.	*KG 1.1, KG 1.2*	ST-1A, ST-1B

Section C.2.2.2	Quantified wildfire impacts on stream network HBGC via influences of burn area and severity on OM chemistry and nutrient concentrations.	*KG 1.3, KG 2.2*	ST-1A, ST-1B, ST-2A
Section C.2.2.3	Advanced mechanistic modeling from molecular to watershed scales by incorporating thermodynamic properties of OM and impacts of wildfire on soil physics into reactive transport models.	KG 1.3, KG 2.2	ST-1A, ST-1B, ST-2A
Section C.2.2.4	Revealed that nearly all respiration in YRB streams is from ERsed, that benthic primary producers drive ERsed, and that wildfire perturbations do not significantly change whole stream respiration.	*KG 2.1, KG 4.1*	ST-2B, ST-3A, ST-3B
Section C.2.2.5	Developed predictive models of batch-scale ERsed across diverse CONUS streams via 18 AI-guided ModEx cycles, and provided a mechanistic foundation for hybrid ML models of ERsed.	*KG 4.1*	ST-2A
Section C.2.2.6	Discovered CONUS-scale transferability of OM bioavailability, spatial scaling of OM biochemical diversity, and microbial community gene expression profiles.	*KG 4.1, KG 4.2*	ST-2A, ST-3A, ST-3B
Section C.2.2.7	Showed that a dominant influence of HEFs on reach-scale ERsed is transferable across basins, via a new River Corridor Model, and used model outputs to guide field data generation and reveal factors influencing ERsed allometry.	*KG 4.1, KG 4.2*	ST-2B, ST-3A, ST-3B
KG 1.1: Hydrologic connectivity. KG 1.2: Vegetation. KG 1.3: Perturbations. KG 2.1: ERsed post-fire. KG 2.2: OM chemistry post-fire. KG 2.3: Microbial communities post-fire. KG 3.1: ERsed following re-inundation. KG 3.2: OM chemistry following re-inundation. KG 3.3: Microbial communities following re-inundation. KG 4.1: Rates and drivers. KG 4.2: Allometric spatial scaling. KG 4.3: ERsed allometry dynamics.			
ST-1A: Watershed Hydrologic Connectivity. ST-1B: Transport and Transformation of DOM from Upland to Streams. ST-2A: Interaction of PyOM and Variable Inundation on ERsed. ST-2B: Influence of Hydrologic Flow Paths on ERsed. ST-3A: Model Development and ERsed Allometry Dynamics in the YRB. ST-3B: ERsed Allometry Dynamics Across Basins.			

C.2.2 Progress Highlights

C.2.2.1 Climate conditions significantly impact post-fire ecosystem recovery

Challenge: Perturbations such as wildfire, insect outbreaks and droughts cause significant but poorly understood impacts on water and C cycling, posing an unpredictable threat with increasing frequency, spread, and severity of disturbances. McDowell et al.[8] identified key challenges limiting our predictive understanding of hydrologic responses to perturbations. The most critical of these is our limited understanding of coupling of streamflow to transpiration, which results in part from the amount, variation, and regulation of hydraulic connectivity between the pools providing for streamflow versus transpiration. Perturbations alter watershed characteristics like vegetation succession, leaf area index, rooting depth, etc., which, in turn, impact the hydraulic connectivity and transpiration-streamflow coupling. This challenge is greatly exacerbated by rising CO_2 and vapor pressure deficit. Therefore, it is crucial to comprehend post-perturbation recovery processes in watersheds to elaborate our understanding of perturbation-impacted ecohydrological processes in context of changing climate.

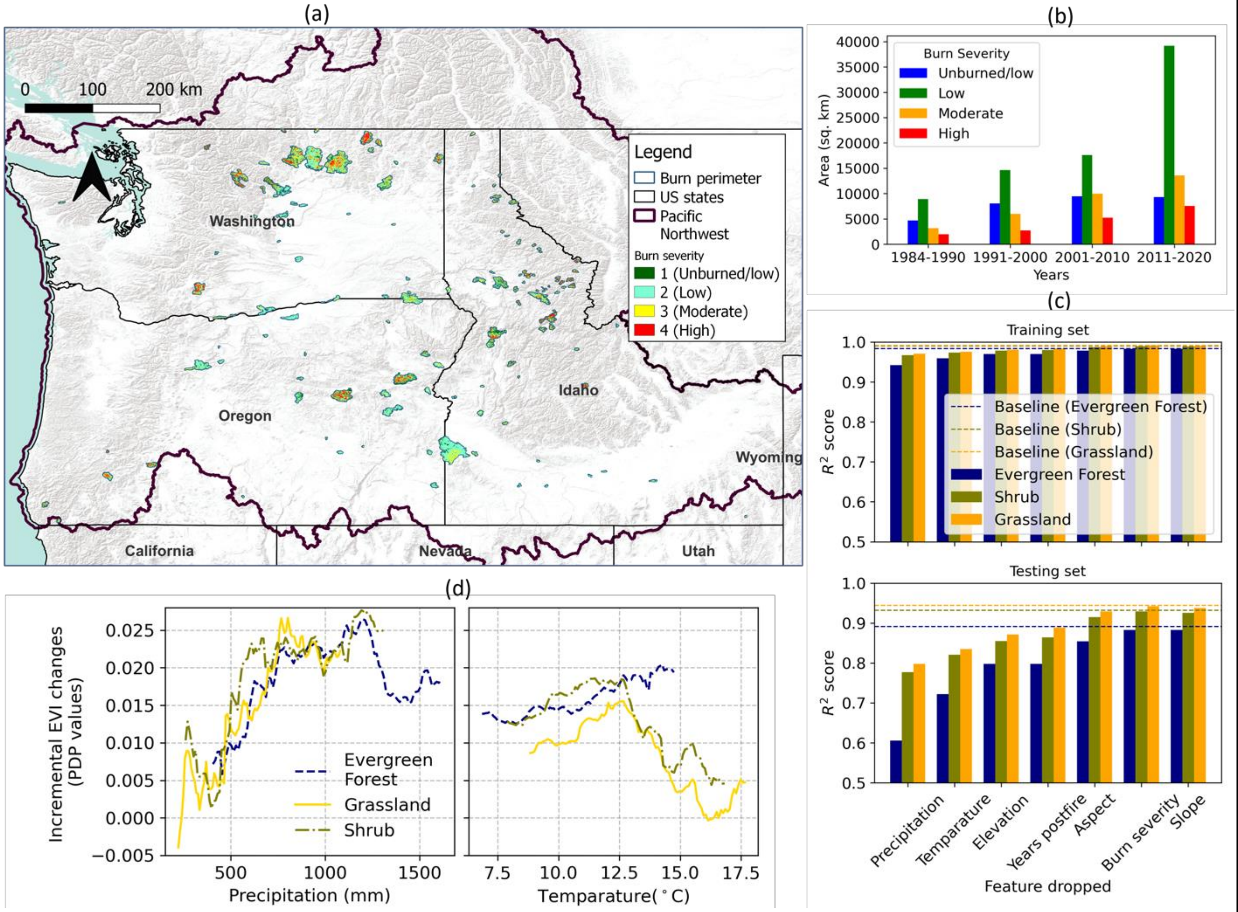


Figure C.2.1 (a) Burn areas within the study region during 2015, (b) historical burn area in the Pacific Northwest between 1984-2020, (c) variable importance scores, and (d) partial dependent plots (PDP) for the two most important features.

Approach and Results: To understand vegetation recovery following wildfire, Zahura et al.[94] investigated climate, topography and fire impacts on growth responses in burn areas of WA, OR and ID during 2015 (**Fig. C.2.1**). Enhanced vegetation index (EVI) was used as a proxy of vegetation biomass recovery after wildfire. Machine learning (ML) was used to quantify the nonlinear relationships of incremental EVI recovery with input features including annual precipitation and annual median maximum daily temperatures during post-fire years (2017-2022), elevation, slope, aspect, burn severity and years elapsed after fire. Separate models were trained for evergreen forest, shrubs and grasslands, to assess recovery differences across land cover types.

The ML model predicted incremental EVI changes with high accuracy, indicating that the model can be further used to analyze the influence of various features on the target variable. Feature importance analysis revealed precipitation and temperature were most influential on EVI recovery. Among the topographic features, elevation was the most important. Partial dependence plots (PDP) for both single and paired input features were examined to assess the variation in EVI corresponding to changes in individual input feature (e.g., precipitation, temperature, etc.) and combined impacts of changes in paired features (e.g., precipitation-temperature pair). PDPs revealed that increasing precipitation was associated with increasing incremental EVI changes for all land cover types. Post-wildfire years with precipitation below 700 mm showed negative incremental EVI values, suggesting vegetation mortality. Evergreen forest showed increasing EVI with increasing temperature. Conversely, shrubs and grassland showed rising trends in EVI until reaching 12.5°C, above which a declining trend in EVI was observed. The most influential topographic feature, elevation, revealed that incremental EVI values showed decreasing trend for elevation above 1500 m for evergreen forests and shrubs. Across all land cover types, a higher incremental change in EVI value was associated with higher burn severity, presumably due to greater resource availability for rapid re-growth.

Significance and Impact: Our remote sensing-based analyses indicated several variables (e.g., precipitation, burn severity) linked to post-wildfire vegetation recovery. This helps understand ecohydrological responses to wildfire, but also motivates us to pursue more mechanistic studies that enable prediction. Such studies will be pursued in ST-1, with an emphasis on how vegetation recovery post-wildfire influences overland and subsurface hydrologic connectivity with stream networks. Context dependent results from this work, tied to prevailing climate and vegetation, further motivate us to study how climate modulates recovery dynamics, with particular emphasis on the continuum from arid to mesic conditions across all three STs.

C2.2.2 Wildfire influences on stream network hydro-biogeochemistry are related to watershed biophysical properties, burn severity, and OM chemistry

Challenge: Wildfires impact river corridor hydro-biogeochemical function by altering the availability of OM and nutrients within the landscape and the hydrological mechanisms responsible for their delivery to stream networks. A shift in material input to stream networks after wildfire can have cascading impacts on C and N stoichiometry that may further impact stream respiration. The complex spatial and temporal heterogeneity in watershed characteristics and the extent to which they are impacted by wildfire (e.g., burn extent, burn severity) remains a key point of uncertainty in our ability to understand and predict influences of wildfires on watershed hydro-biogeochemistry. To address this knowledge gap, we coupled a series of meta-analytic, field-based, and modeling approaches to identify key watershed scale linkages to stream network responses after wildfire.

Approach and Results: To identify the influence of wildfires on stream network hydro-biogeochemistry across space and time, we collated studies across broad spatial scales to assess the influence of climate, time since wildfire, and watershed burn extent as relevant for explaining stream hydrology after wildfire[95]. This meta-analysis, spanning 3 biomes and 160 watersheds representative of the western United States, identified a persistent shift in stream water chemistry (higher nitrate, lower organic C) for at least 5 years following wildfire. This indicates that longer-term ecosystem recovery trajectories should be considered for biogeochemical impacts of wildfire. Conversely, the net influence of climate and percentage of watershed burn area on water chemistry exhibited high variability, demonstrating complexity in upscaling localized watershed scale processes that couple an array of spatially and temporally explicit drivers of material export to stream networks.

To better understand localized spatial and temporal drivers in stream water quality responses to wildfire that may influence how such drivers impact stream network responses at a watershed scale, we explored the effects of the 2020 Holiday Farm Fire near Eugene, Oregon. This wildfire burned ~629 km^2 of predominately hemlock forest over a range of low-high severities across the Mackenzie River basin. We used a Spatial Stream Network (SSN) model that was built from ~100 sites above, within, and below the burn perimeter with repeat measurements across multiple seasonal wetness conditions[96]. The SSN indicated the wildfire impacts were primarily masked by the variability of site-level landscape characteristics across the watershed. However, as the basin rewet, burn severity did influence observed DOC concentrations, with a decrease in DOC with increasing burn severity. Additionally, we found that groundwater dominated transport during drier seasons were less likely to carry immediate short term fire signals. In the same watershed, when high temporal resolution water samples were collected from five sub-watersheds during the first major storm flush after wildfire, a systematic shift in the chemistry of burned materials delivered to the individual streams were directly linked with burn severity[58]. Streams in catchments burned at higher severity were enriched with N containing aromatic compounds compared to those that burned at lower severity.

Recent evidence suggests that the transport of fire-impacted compounds (otherwise termed pyrogenic OM or PyOM) are important for biogeochemical processing in stream networks. Estimating the bioavailability of PyOM has remained elusive. Our team addressed this gap by using a substrate-explicit model to assess the potential bioavailability of diverse PyOM compounds (**Fig. C.2.2**). Model outputs indicate a wide array of potential PyOM impacts on stream network biogeochemical function, indicating that PyOM may be an underappreciated driver of stream network biogeochemistry[64].

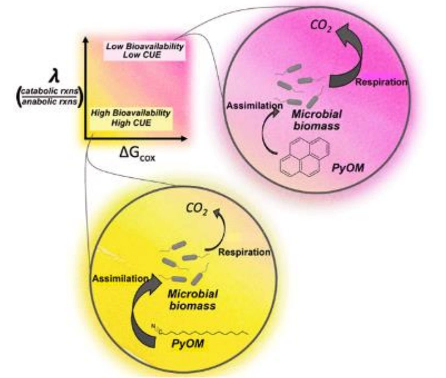


Figure C.2.2 This conceptual diagram of PyOM highlights variable molecular composition, which can influence its ability to be used by microbes contributing to ERsed.

Significance and Impact: Collectively, this work advanced our understanding of wildfire impacts on stream network chemistry and function. Developing a predictive understanding of wildfire impacts has been a challenge for the research community due to the complex array of spatial and temporal drivers that impact stream water chemistry after fire. We observed that stream water responses within a single burn perimeter were modulated by localized hydrologic processes. This motivates us to emphasize the impacts of wildfire on the transport of OM to stream networks via changes in hydrologic flow paths. ST-1 will study how wildfire interacts with biophysical setting to change OM transport, while ST-2 and ST-3 will examine the resulting influences on batch, reach, and network scale ERsed.

C.2.2.3 **Integrating OM** measurements into watershed hydro-biogeochemical models

Challenge: Watersheds play significant roles in modulating C and N cycling and removal of excess nutrients. However, the connections between OM measurements and biogeochemical reaction network models from batch to watershed scales are missing. To fill the gap, we incorporated of molecular-level characterization of OM from field measurements into watershed hydro-biogeochemical models to improve the modeling ability in capturing water quality and quantity signatures following perturbations. The connections between OM measurements and biogeochemical reaction network models from batch to watershed scales are missing, however.

Approach and Results: We developed an integrated modeling approach that establishes a connection between distributed OM chemistry measurements and batch reactions, and extends seamlessly to watershed-scale modeling (**Fig. C.2.3**). We developed a workflow that connects organic carbon speciation informed by high resolution Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) into the PFLOTRAN reaction sandbox[90.]. The Jupyter Notebook-based workflow ingests high-resolution Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) data, calculates the distribution of lambda values (analogous to C use efficiency), generates an aerobic respiration-based reaction network, completes a biogeochemical simulation in the PFLOTRAN batch reactor, and performs parameter estimation when observed dissolved oxygen is available . We successfully used the workflow[97] to parameterize aerobic respiration reactions to various types of experimental incubations, which is critical for upscaling molecular-scale charaeterization at larger scales.

Further, to account for the role of non-oxygenic electron acceptors in regulating OM turnover and the fate of C, we generalized the aerobic respiration-based C speciation by incorporating both detailed OM chemistry and electron acceptors other than O_2[98]. We demonstrated the effectiveness of the proposed generalized approach on experimental data. The results show how key microbial growth parameters such as C use efficiency and reaction rates vary across different electron-accepting processes and highlights the importance of representing other electron acceptors in watershed-scale respiration process.

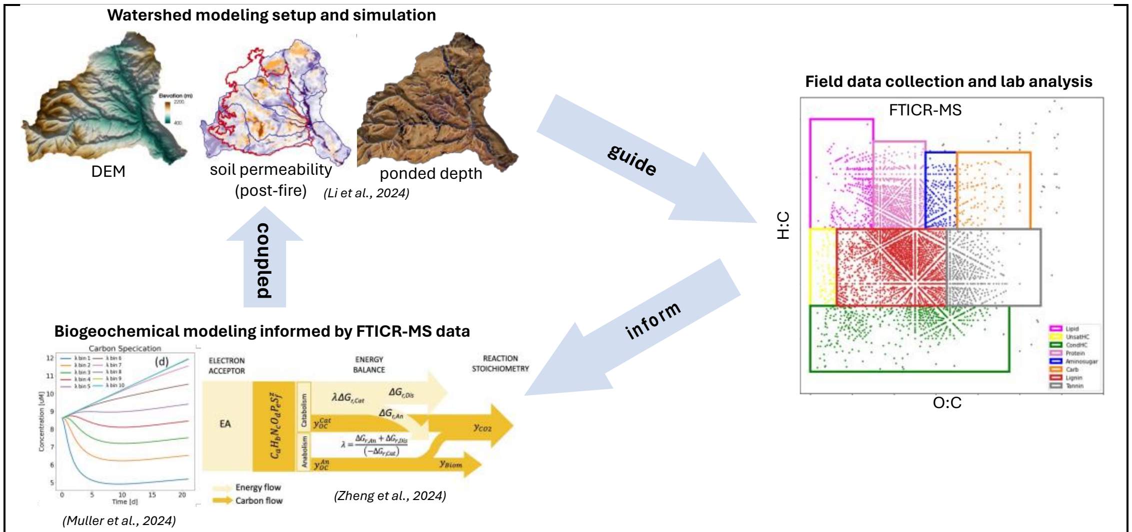


Figure C.2.3. Iteratively coupling a watershed model with a reaction network informed by FTICR-MS data derived from field samples of OM chemistry.

We then developed an integrated watershed model that couples the formulated PFLOTRAN batch reaction with the distributed flow and transport model ATS. We employed the model to study the impact of wildfire disturbance on watershed hydrological response by implementing a new fire module to modify the soil properties in fire scars[99]. We used the watershed model to evaluate the downstream runoff to wildfires across multiple watersheds in Pacific Northwest, such as Naches River Watershed. Our results underscore the critical roles of burn severity and post-fire precipitation in increasing peak flow discharge.

Significance and Impacts: Our work paved the way for future analysis on the wildfire impacts on watershed hydro-biogeochemistry processes. The mechanistic representation of hydrological processes (e.g., subsurface lateral flows) in ATS is crucial to accurately capture solute transport. Aided by the proposed workflow, we can delineate C and N cycling from field measurements and simulate the subsequent reactive transport processes across watersheds through ATS-PFLOTRAN. We also found that the mechanistic representation of hydrological processes (e.g., subsurface lateral flows) in ATS is crucial to accurately capture solute transport. To further study perturbation impacts on watershed biogeochemistry, future work will incorporate the generalized C speciation method into the workflow and enable characterization of spatiotemporal variation of reaction networks in watershed models. Other key processes, such as benthic primary production and microbial gene expression, will also be incorporated into watershed models to enhance understanding of ERsed. Further integration will leverage additional advances our team has contributed to, such as the omics-to-reactive-transport pipeline that links microbial genomics to PFLOTRAN[100].

C.2.2.4 Hypothesis-based ModEx reveals drivers of ERsed across the YRB

Challenge: Stream networks emit significant amounts of CO_2 to the atmosphere, both from abiotic and biotic processes, yet the biotic CO_2 contributions from different stream components (e.g., benthic, hyporheic, water column) are variable and poorly understood at basin scales. The relative importance of different components on stream network respiration over space and time, including the importance of watershed perturbations on respiration, is significantly understudied. We filled these knowledge gaps by using basin-scale, hypothesis-based ModEx (**Fig. C.2.4**) to increase mechanistic understanding of biogeochemical processes within and between stream network components and extend these mechanisms into conceptual models that include the role of perturbations.

Approach and Results: We used numerical simulation to generate hypotheses regarding key drivers of reach-scale ERsed across the YRB, and used these hypotheses to guide field observations. First, our newly developed River Corridor Model was used estimate ERsed from the hyporheic zone[24]. Our results showed that HEFs were the main driver of spatial variation in ERsed. These results guided the design of a basin-scale field study that evaluated if spatial variation in model-predicted hyporheic zone ERsed explained the spatial variation of field observed ERsed. We found no correlation between the predicted and observed rates; thus, we explored additional mechanisms by studying the contributions from water column processes and sediment-associated components (i.e., benthic algae, submerged macrophytes, and shallow hyporheic zones). While water column respiration was very small[101], ERsed represented most of the whole stream respiration and was primarily explained by benthic gross primary productivity and channel slope[25]. This indicated that sediment-associated processes, excluding deep hyporheic zone respiration, are primary drivers of spatial variation in stream network respiration throughout the YRB.

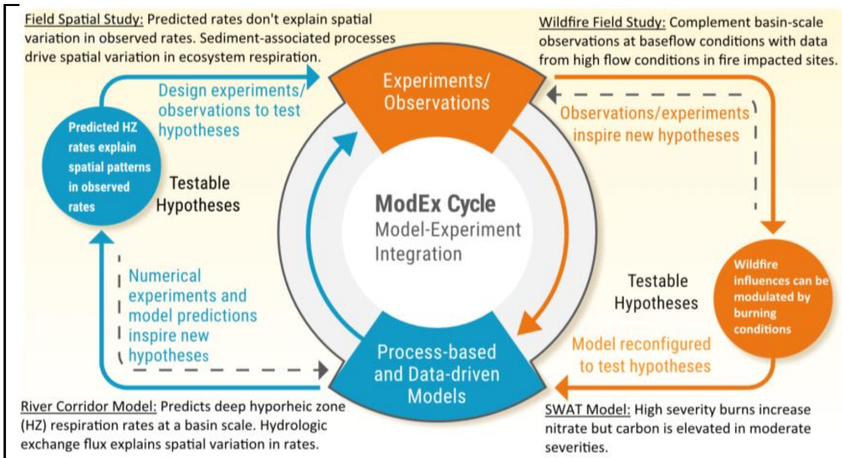


Figure C.2.4: Conceptual diagram of hypothesis-based ModEx. We use observations to generate hypotheses that can be tested via models. We used model predictions to inspire hypotheses that can be tested with new data can be used to guide further model developments.

We carried out a second ModEx cycle to investigate the effects of wildfire on stream respiration across the YRB. We hypothesized that wildfires influence stream respiration only during high hydrologic connectivity to the terrestrial landscape and carried out a study that complemented basin-scale field observations at baseflow conditions with observations collected during high flow conditions[102]. In a parallel effort, we coupled knowledge from field observations with predictive watershed models to evaluate the changes in water quality after fire by manipulating the watershed area burned and the burn severity in a series of counterfactual simulations leveraging a recently developed fire module of the SWAT model[103]. Model results showed that high burn severity led to the greatest increase in nitrate and moderate burn severity led to the greatest increase in DOC concentrations across all areas burned[104]. This highlights that there are important relationships between watershed characteristics, including burn parameters, and stream network hydro-biogeochemistry, that could be further used to inform predictive understanding of the influence of wildfires on ERsed.

Significance and Impact: Our work uncovered important mechanisms modulating stream respiration across the YRB, with and without wildfire perturbations. The coupling between modeling and observations enabled testing of specific hypothesis that couldn't have been tested with either method in isolation. Results from this work motivate us to extend conceptual and physical models of ERsed to include physical sediment properties that influence shallow HEFs, primary productivity, hydrologic connectivity and wildfire effects linked to burned area and burn severity. This will be done via integrated efforts across all three STs. ST-1 will emphasize wildfire impacts on hydrologic connectivity. ST-2 will emphasize perturbation impacts to ERsed. ST-3 will link ST-1 and ST-2 outcomes to understand drivers of ERsed allometry dynamics.

C.2.2.5 Beyond data assimilation: A hypothesis-driven ModEx approach to predictive modeling of ERsed across CONUS

Challenge: The hyporheic zone can contribute up to 100% of stream respiration, but this contribution varies widely and isn't effectively captured in current models. The hyporheic zone's variable permeability and diverse exchange mechanisms result in heterogeneous conditions, posing a significant challenge for modeling due to the complexity of coupled physical and biogeochemical processes. Studies often focus on specific aspects (e.g., physical, chemical, or biological) of hyporheic zone dynamics, limiting comprehensive understanding and scaling efforts due to data and knowledge constraints.

Approach and Results: We took a hypothesis-driven approach to address the challenge. Starting with conceptual models built from known theory and data patterns, we identified key hypotheses we may test with available data. Using a continental-scale dataset from WHONDRS, we first tested the prediction that batch-scale ERsed will decrease with increases in the number of unique organic molecules (i.e., OM molecular richness). Our continental-scale analyses rejected the hypothesis of a direct limitation of respiration by OM molecular richness. Rather, we found that DOC concentration imposes a primary constraint over ERsed, with secondary influences of OM richness[81]. ERsed was related to DOC concentration as a non-linear constraint boundary, with most respiration rates falling below the boundary (**Fig. C.2.5-A**). This suggests additional controls driving ERsed below its potential maximum. In turn, we conducted numerical experiments to explore the control of ERsed, spanning chemical, physical, and biological processes (**Fig. C.2.5-B**). These potential controls were explored using a simple thermodynamic-based kinetic model, where OM thermodynamics, microbial biomass and substrate accessibility were included as key drivers. Integrating numerical experiments and observational data, we identified that the OM molecules fueling respiration are less than 1% of the measured DOC[105]. These results motivate deeper evaluation of molecular mechanisms governing ERsed.

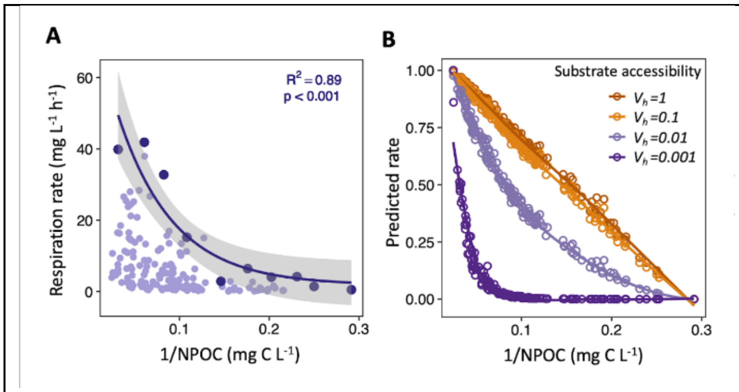


Figure C.2.5. (A) Across CONUS ERsed is linked to DOC via a constraint boundary, which motivated (B) numerical experiments to generate hypotheses for why most samples fall below the constraint boundary. The resulting model-generated hypotheses are being evaluated with additional data from biophysically diverse stream networks across CONUS.

Additional ModEx work across CONUS provides further motivation to dive deeper into governing mechanisms tied to molecular properties. Starting with the same data used for the studied summarized above, we partnered with industry to develop ML models of batch-scale ERsed. Outcomes were used to guide additional data collection across CONUS, enabled by the WHONDRS consortium. The predictive skill of the ML models improved to 80% across 18 AI-guided ModEx loops. These efforts added data from 110 new field sites, and the predictive capacity of the ML models stabilized at ~80%. While this is impressive for a CONUS-scale model of a highly variable process, we observed limitations in extrapolating to new field locations. This highlights the need to move beyond the process-free ML models to achieve predictive understanding and models that are transferable across all stream networks. Further motivation to focus specifically on OM chemistry is that a feature permutation importance analysis identified ultra-high resolution OM chemistry data as one of the key variables explaining CONUS-wide variation in batch-scale ERsed[106]. In turn, the RC-SFA is now generating molecular data for the samples collected across the 18 ModEx cycles to enable a hybrid (i.e., process-informed) ML approach.

Significance and Impact: Our hypothesis-driven approach, based on known theory and data patterns, has enabled a novel way integrating models and observations. The continental-scale dataset from the

WHONDRS consortium allowed us to challenge and refine exiting hypotheses with numerical experiments, and eventually build predictive models that transcend biophysically divergent stream networks and offer a path forward based on the integration of deep process understanding with highly scalable AI/ML methods. The outcomes motivate us to continue focusing on how OM chemistry influences ERsed and to evaluate how these influences scale from batch to reach scales in ST-2. We will also examine the degree to which OM chemistry influences ERsed allometry (i.e., at the stream network scale) in ST-3.

C.2.2.6 Reconciling variation and interconnections in stream network OM degradation and microbial community activity across scales

Challenge: Substantial amounts of dissolved organic matter (DOM) are transported throughout stream networks. During transport, DOM biotically and abiotically interacts with the surrounding ecosystem and significantly impacts biogeochemical cycles. While we understand some fundamental interactions between DOM and the ecosystem, we do not understand specific interactions between many ecosystem components, the impact of these interactions at different spatial scales, and the characteristics of DOM which significantly drive these interactions. To address these knowledge gaps and reveal DOM-ecosystem interactions, we used high-resolution DOM characterization, microbial sequencing, and novel data science approaches to investigate DOM and microbial patterns across spatial scales (**Fig. C.2.6**).

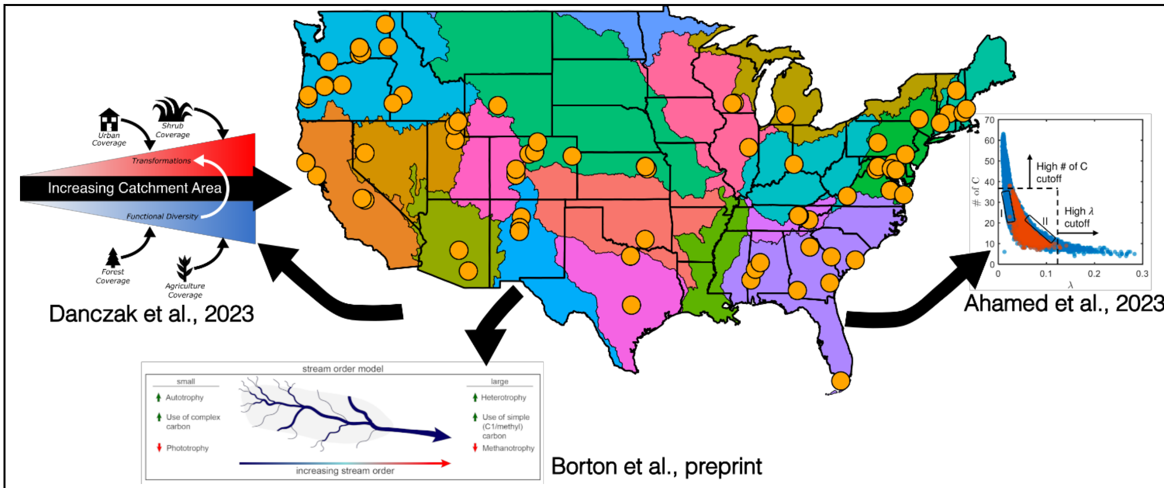


Figure C.2.6: Sampling sites across CONUS and major conceptual outcomes from corresponding manuscripts. These studies revealed patterns and mechanistic linkages related to DOM, microbes, and ERsed that are transferable across biophysically divergent basins. Colors on the map are major CONUS basins.

Approach and Results: We performed field-based studies across the CONUS leveraging ultrahigh-resolution DOM characterization and demonstrated that watershed size (e.g., stream order) and land cover (e.g., forested, urban, etc.). were significant factors affecting DOM composition. Within the YRB, we collected samples from 6 sites every week from April to May 2021 and observed that DOM functional diversity and the diversity of potential biochemical transformations increased linearly with watershed drainage area. We demonstrated that the spatial scaling of DOM might be transferrable by further connecting these patterns to land cover[107]. To evaluate this transferability, we identified putative biochemical transformations in 42 sites from 5 different basins across CONUS. Results indicated that, while the YRB spatial scaling was not transferable across all basins, there is a transferable link between biochemical transformations and that ratio of reaction rate to residence time[108].

We expanded our investigation using an ML approach to identify bioavailable DOM characteristics related to batch-scale ERsed across biophysically divergent watersheds across CONUS[80]. This method indicated that thermodynamic favorability, the number of C atoms, and C:N ratios were important DOM properties that predict ERsed. Moreover, this revealed that a subset of molecular formulas might be considered bioavailable and be more likely to be consumed by microorganisms. Putative connections between DOM and microbial activity were assessed by combining the ultrahigh-resolution DOM characterization data with shotgun metagenomic and metatranscriptomic sequencing for samples collected across the CONUS[86]. We uncovered 2093 unique metagenome assembled genomes (MAGs) belonging to 27 phyla and diverse functional clades. Unlike the DOM which became more functionally diverse as watersheds increased in size, microbial expression became less functionally diverse and dominated by clades capable of using simple C compounds. This is reflected in DOM relationships whereby heterotrophs with expressed polymeric degradation capability were negatively related to the relative abundance of highly aromatic organic compounds.

Significance and Impact: These studies revealed that DOM exhibits multi-basin transferable patterns related to physical properties of watershed systems. When compared to existing conceptual models like the River Continuum Concept (RCC) which predicts that DOM will decrease in diversity as stream size increases, our emerging conceptual model indicates that DOM diversity is controlled by the balance between internal thermodynamics and external watershed properties. In contrast to these DOM-based patterns, the microbial data more closely mirrors the RCC where microbial function becomes more focused on simple C compound degradation with increasing stream size. These contrasting results highlight divergent rules governing DOM chemistry and microbial communities within stream networks but emphasizes emergent connections that result in multi-basin transferable linkages between DOM chemistry and microbially driven ERsed. These results give us confidence that we will continue to discover broadly transferable patterns and linkages across OM chemistry, microbial ‘omics, and ERsed. These facets will be studied within ST-2, including how perturbations influence detailed molecular properties and the connections between those properties and ERsed.

C.2.2.7 Multi-basin modeling, regional transferability, and ERsed allometry

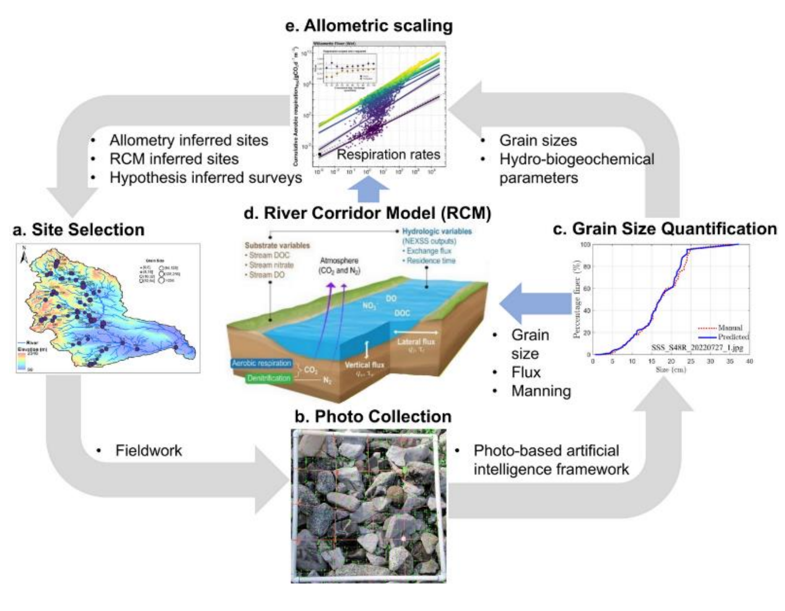


Figure C.2.7. ModEx loop including initial site selection guided by the river corridor model (a), photo collection and AI grain-size and hydro-biogeochemical parameter estimation (b,c), incorporating grain size and flux data in the river corridor model to derive cumulative ERsed (d), and assessing ERsed allometry (e). The ModEx loop will continue into the proposed quadrennial period to guide data generation and model development.

Challenge: While lab, field, and numerical studies have improved our understanding of hyporheic zone (HZ) biogeochemical processes, scaling these processes up to the reach-scale and beyond or transferring this knowledge from one basin to another remains a challenge due to modeling, observation, and scaling theory challenges. In modeling, we lack a basin-wide model capability to determine key factors controlling the spatial variability of microbe-driven reactions. In observation, we lack the capability to measure key HZ hydro-geological properties such as grain sizes that influence hydro-biogeochemical parameters at the basin scale. These combined challenges limit our capacity to evaluate the transferability of model-derived hypotheses that link reach-scale ERsed to hydro-geological properties.

Approach and Results: To help resolve these challenges we initiated a ModEx loop (**Fig. C.2.7**) by developing a C-N coupled river corridor model (RCM) to quantify HZ aerobic and anaerobic respiration[24]. Key factors controlling their spatial variability within the Columbia River Basin were identified using ML. Model results show that CO_2 emissions from ERsed are spatially heterogenous and vary with sub-basin dry-wet conditions, stream/river size, and O_2 availability. Specifically, with abundant O_2, wetter sub-basins showed higher CO_2 emissions than drier ones and medium-sized rivers generated the highest CO_2 emissions. Without O_2, reaches in agricultural areas generated relatively high CO_2 emissions. The ML model, trained from the RCM outputs, found that most of the spatial variation of CO_2 emissions were explained by HEFs that were themselves influenced by streambed grain size.

Due to the fundamental control of grain size on HEFs and CO_2 emissions, a photo-driven, artificial intelligence (AI)-enabled, and theory-based model framework was subsequently developed to quantify grain sizes and hydro-biogeochemical parameters from field photos[109]. The AI model was trained with ~12,000 grain labels representing 9 typical stream environments using You Only Look Once (YOLO), a state-of-the-art computer vision AI, with a Nash–Sutcliffe efficiency of 0.98 and overall relative error of 6.7%. We then applied the AI to extract the quantities, distributions, and uncertainties of streambed grain sizes from 1,999 photos collected from 66 sites, and used these grain sizes to estimate HZ hydro-biogeochemical parameters. The results show that distributions of grain size, Manning coefficient, Darcy-Weisbach friction, interstitial velocity magnitude, and nitrate uptake velocity, follow log-normal, normal, positively skewed, near log-normal, and negatively skewed distributions, respectively. The photo-derived grain sizes were compared to datasets from USGS, NEXSS, and other ML-derived products. The results show that USGS and ML-based grain size products are less than 6 mm on average, but NEXSS and YOLO grain size estimates vary between 20-25 mm, which is more consistent with visual observation[110].

We further assessed if ERsed scales allometrically with watershed area—as predicted by theory. The allometric scaling was based on reach-resolved whole-stream-network predictions of ERsed from the RCM in two biophysically divergent basins[88]. The results revealed a cross-basin transferable link between allometric scaling parameters and HEFs. In both basins, scaling slopes became steeper (super-linear) and much tighter (higher R^2) with increasing catchment-level HEFs.

Significance and Impact: This work guided us to emphasize characterizing and understanding spatial variation in sediment grain size as a master variable influencing ERsed at reach to stream network to multi-basin scales. The new modeling and measurement capabilities enable quickly estimating grain size, HEFs, and ERsed from field photos that are simple to collect across diverse stream networks and can be automatically analyzed via ML. ST-2 will use these field and ML methods to scale from batch to reach. In addition, the RCM is the foundation from which an 'enhanced RCM' will be developed in ST-3 as an integrative modeling capability to be used by ST-3 to understanding ERsed allometry dynamics.

C.3 Research Plan

To resolve our science questions, address our grand challenge (**C.1.3**), and fill knowledge gaps (**C.1.4**), we developed three integrated Science Themes (STs) with underlying Research Activities (**Fig. C.3.1**). All STs will use a hypothesis-based approach and are connected to each other through integration pathways, whereby knowledge, data, and/or models from one ST provide inputs to other STs (**Figs. C.3.1, C.3.2**). Such leveraging and integration are hallmarks of SFA science and increase science impact.

Over the past four years, the RC-SFA has been structured into research campaigns (RCs) that each emphasize integration across RCs and feedbacks between models and experimentation (i.e., ModEx). We continually evolve towards deeper integration and increasingly sophisticated use of ModEx. As part of this evolution, the proposed STs move beyond RCs that have had single science leads, whereby each ST will be co-led by members of the SFA leadership team that come from different backgrounds and disciplines. This co-leadership structure helps ensure that the RC-SFA science outcomes are more than the sum of the parts.

PIER and ICON are central elements - How we do science is as important as the hypotheses we test and the technical methods we use. We value team science built on a foundation of an inclusive, equitable, and safe environment. Our PIER plan (**C.6**) details new and continuing project-specific activities we will use to enhance research impact and support new and existing team members. Psychological safety is a cornerstone of innovation. We are committed to a team culture that engenders psychological safety (e.g., by following our team co-created code of conduct) that will lead to innovations that will continually improve upon our proposed research plans.

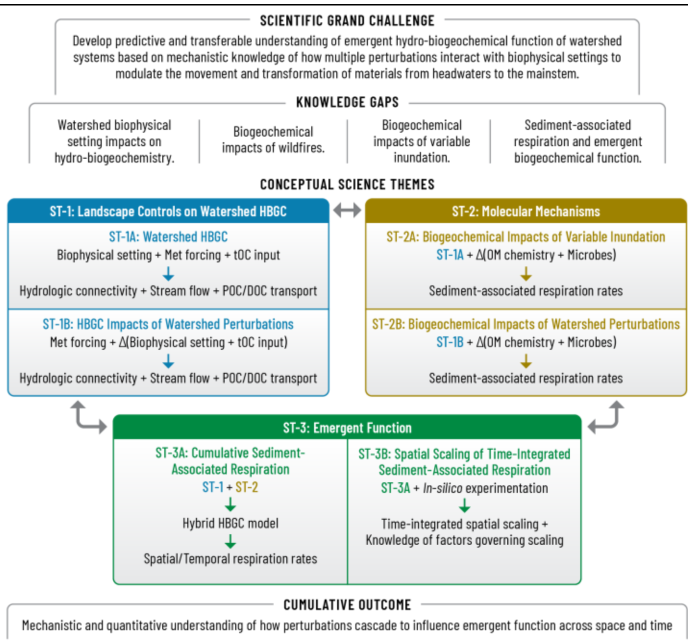


Figure C.3.1: The SFA will implement hypothesis-driven research in three science themes (STs) that address knowledge gaps tied to our grand challenge. Each ST has two Research Activities (e.g., ST-1A). In accordance with the Integrated principle of ICON, we will deliver scientific innovation by coupling knowledge, data, and models across disciplines and scales that span physical, chemical, and biological components of watersheds, from hillslopes to stream networks.

As part of our PIER plan (**C.6**) we will use ICON science principles[1,111] to guide design and implementation of our science plan. ICON provides a framework to achieve transferable and mutually beneficial science outcomes by being intentionally ***I****ntegrated* across disciplines and scales to accelerate understanding of systems, ***C****oordinated* with consistent methods to enable interoperability, ***O****pen* throughout the research lifecycle to increase long-term value and accessibility, and ***N****etworked* across those interested in and impacted by our science to understand needs and co-develop opportunities for mutual benefit. The Open component of ICON includes the Findable, Accessible, Interoperable, and Reusable (FAIR) data principles [112], and complements FAIR by sharing ideas, tools, and data from the early stages of ideation to the final stages of manuscript publication.

In accordance with the ***I****ntegrated* component of ICON, the SFA will include ModEx-oriented teams of modelers and experimentalists; these teams are focused on co-mentoring and complement the cross-cutting ModEx implementation teams. Members of the co-mentoring teams will help each other towards mutual understanding of perspectives, needs, and challenges associated with modeling, data generation, and their iterative feedback. The ModEx implementation teams will guide iterative ModEx and be responsible for ***C****oordinated* use of methods that maximize interoperability. Data and code interoperability will be achieved through mutual agreement on protocols to be used consistently across STs. Such decisions will be based on cross-ST communication and ***O****pen* engagement with the broader science community (e.g., **C.3.2**) to align with technical approaches used outside the SFA. Two ModEx implementation teams are envisioned. One will focus on molecular-to-reach scale processes and the other will focus on reach to watershed scale processes. The establishment of ModEx implementation teams is an evolution of the 'Multi-Scale ModEx' integrating activity that was part of the RC-SFA over the past four years, which contributed to several advances we build from (**C.2**).

The ***N**etworked* component of ICON is also foundational to advancing our science and developing our team, while finding and delivering mutual benefit to interested parties beyond our core team. To achieve mutual benefit requires understanding the needs of interested parties and adapting to them while staying true to our science vision and objectives. To this end, all STs will network with a range of interested parties, including other ESS-funded SFAs, projects funded by other agencies, regional groups, and the global science community. These interactions will enhance our capacity to evaluate transferability of place-based outcomes and improve technical innovations (e.g., by enabling multi-watershed data generation to test model predictions across CONUS, via the WHONDRS consortium). This will also facilitate long-term relationship building with diverse researchers.

Model-experiment (ModEx) iteration within and across basins - We will advance science by merging model-experiment (ModEx) iteration, intensive place-based research aimed at deep mechanistic understanding, and widely distributed efforts aimed at transferability. Across all three STs we take an approach to ModEx in which conceptual and mechanistic understanding guides model development, model predictions guide data generation, and interpretation of generated data guides further model development. In our approach to ModEx, hypotheses are the primary currency that drives iterative feedback between models and experiments/data generation.

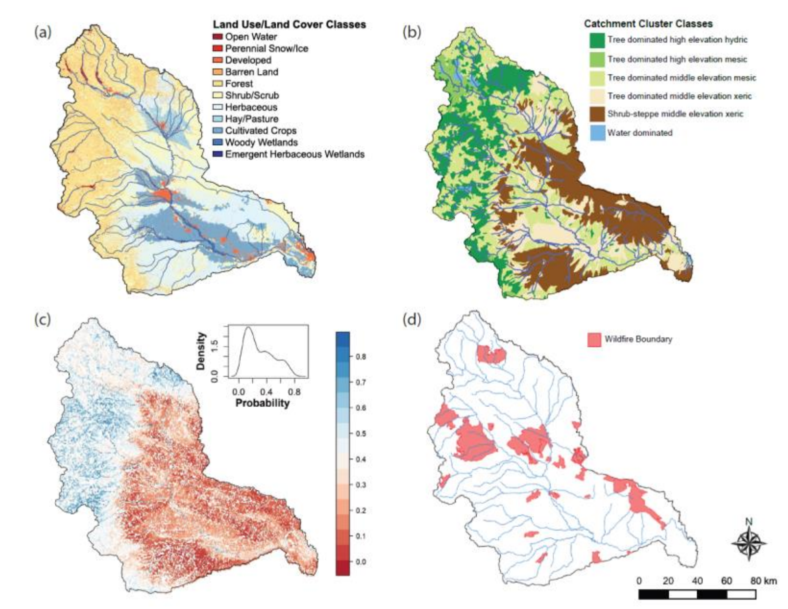


Figure C.1.5. The Yakima River Basin will be the single-basin testbed for the RC-SFA because it is diverse in terms of (a) land cover, (b) biophysical characteristics, (c) variable inundation regimes, and (d) wildfire perturbations.

One reason we take a ModEx approach is it can help efficiently generate transferable understanding across diverse environments. Efficiency is key as we propose to work across watersheds in a single-basin testbed and across watersheds in basins across CONUS (used here as a multi-basin testbed). All STs will use the Yakima River Basin (YRB) as the single-basin testbed. The YRB in south-central Washington State (**Fig. C.1.5**) will be our basin testbed because it is (*i*) large enough at ~16,000 km^2 to contain several extensive stream-networks, (*ii*) it is environmentally diverse spanning multiple ecoregions and all major biophysical classifications of the larger Columbia River Basin[101], which itself spans the entire northwest region of CONUS, (*iii*) has ERsed rates that are representative of rates from across CONUS[25], (*iv*) has stream networks that span a large gradient in the fraction of reaches that experience variable inundation[113], (*v*) has recently experienced and will likely continue to experience large wildfires across multiple biophysical classes[101,114], (*vi*) includes numerous heavily-invested parties/groups ranging from indigenous nations to local, state, and federal agencies, and (*vii*) is practically feasible due to its physical proximity to PNNL's main campus. For fieldwork in the YRB we will work with regional partners (e.g., Confederated Tribes and Bands of the Yakama Nation) with elements of co-design to ensure mutual benefit (**C.6**). Field site selection across the YRB will be driven by model-refined hypotheses and associated predictions.

We will use watersheds across multiple basins to evaluate transferability of YRB outcomes and to build a foundation of knowledge, data, and models to enable advancement towards our longer-term vision (**Fig. C.1.3**). As in the YRB, study sites across CONUS will be selected by model-refined hypotheses and associated predictions. Our multi-basin work will lean heavily on our experience with the Worldwide Hydrobiogeochemistry Observation Network for Dynamic River Systems (WHONDRS consortium), which currently spans over 400 field sites and institutions across CONUS. This highlights the power of working with participatory research networks like WHONDRS to enable nimble, model-guided data generation and iterative CONUS-scale ModEx.

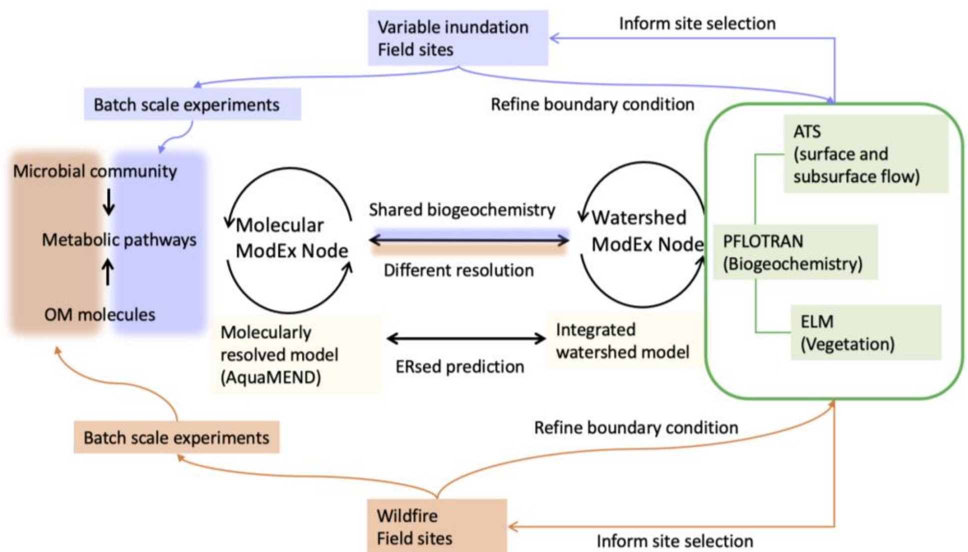


Figure C.3.2: The RC-SFA will advance predictive understanding via an integrated research program based on feedbacks spanning molecular to watershed processes/scales and ModEx-based iteration across a diverse collection of models and data types. The three Science Themes (STs) are intermingled across the project, with ST-2 falling more to the left half of the diagram and ST-1 and ST-3 falling more to the right half. The STs are not visually distinguished here because the boundaries are meant to be somewhat fluid. This is a purposeful project design element to facilitate sustained cross-ST integration.

Science Theme summaries – Below are short summaries of each ST, with an emphasis on cross-ST integration.

ST-1: Landscape Controls on Hydro-Biogeochemistry - ST-1 will develop process-based understanding linking watershed biophysical characteristics to hydrologic connectivity, with an emphasis on the influences of climate, vegetation, topography, subsurface properties, and wildfire perturbations. ST-1 will focus on the hillslope-to-stream continuum (**Fig. C.3.X, horizontal axis**). and generate new knowledge and spatiotemporal predictions of variation in water flow paths from land to stream reaches, streamflow dynamics that lead to variable inundation, and quantity and composition of OM transported from hillslopes to streams with and without wildfire perturbation. These outputs will be used for hypothesis evaluation in ST-2 and model-development followed by hypothesis evaluation in ST-3. ST-1 will work primarily in the YRB with the potential for some efforts in highly instrumented watersheds across the Pacific Northwest (**Fig. C.3.X, vertical axis**).

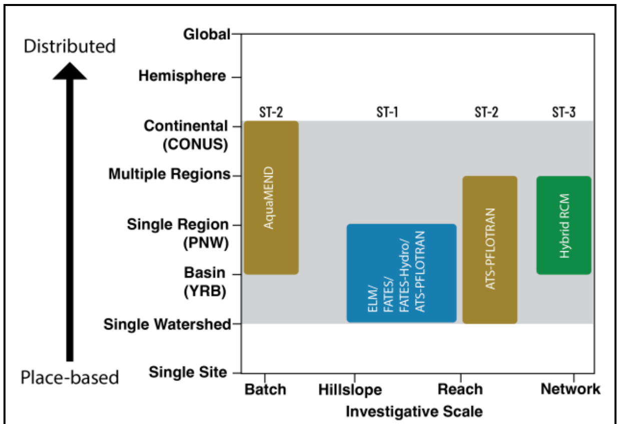


Figure C.3.X: Science Themes (STs) of the RC-SFA investigate batch to network scales (horizontal) at watershed to CONUS spatial extents (vertical). The RC-SFA occupies a significant portion (grey box) of the continuum from single site place-based science to maximally distributed global science.

ST-2: Drivers of ERsed - ST-2 will reveal the influences and underlying mechanisms of variable inundation and wildfires on spatial and temporal dynamics of batch to reach scale ERsed. ST-2 complements ST-1 and ST-3 with a bottom-up approach to understand the molecular mechanisms controlling ERsed in context of dynamic ecosystem properties. ST-2 will study how wildfire and variable inundation jointly influence interactions among OM chemistry (e.g., thermodynamics), microbial metabolic function, sediment-associated primary production, and pathways of water delivery. ST-1 outputs (e.g., flow pathways) will be used to evaluate ST-2 hypotheses and will also guide site selection for ST-2 field-based data generation. Batch to reach scale models developed in ST-2 will be integrated with ST-1 models within ST-3. ST-2 is further linked to ST-1 by providing remote sensing data to iteratively improve ATS predictions of variable inundation. ST-2 will hand this approach off to ST-3 in last two years of the project. ST-2 will work within the YRB and across biophysically divergent CONUS watersheds (**Fig. C.3.X, vertical axis**).

Science Theme 3 (ST-3): Integration and Scaling - ST-3 will integrate knowledge and models from ST-1 and ST-2 into a new watershed-scale hydro-biogeochemical model and use this model to understand and predict ERsed allometry dynamics. The resulting model will account for the combined influences of watershed biophysical setting, wildfire, variable inundation, and molecular processes on spatial and temporal variation in ERsed. Model predictions of ERsed will be used to characterize ERsed allometry at the stream network scale (**Fig. C.3.X, horizontal axis**). The model will be used to run counterfactual *in silico* experiments for hypothesis evaluation. Initial hypothesis evaluation will focus on understanding the individual and interactive effects of variable inundation and wildfire on ERsed allometry dynamics. Hypothesis evaluation outcomes will guide ST-3 field deployments to test a follow-on hypothesis focused on how climate modulates the influences of perturbations on ERsed allometry dynamics. This work will be placed in context of and used to advance recently developed scaling theory that provides quantitative expectations for ERsed allometry. ST-3 will work within the YRB and across CONUS watersheds arrayed along an arid-to-mesic continuum (**Fig. C.3.X, vertical axis**).

An adaptive approach – All three STs include model-guided data generation following hypothesis evaluation. The practical "where, when, and what" details of data generation must, therefore, adapt as hypotheses are evaluated. This means we cannot know all the details *a priori*. Any project that conducts iterative ModEx faces this challenge. Our solution is to summarize technical plans for different potential outcomes or use an approach based on dynamically changing field sites across ModEx iterations (**e.g., C.2.2.5**). The ST descriptions (**C.3.1-C.3.3**) each have one or two figures that overview ModEx steps either across alternative outcomes (from hypothesis evaluation) or across iterative model-data-model loops. Each Activity within the STs has a slightly different approach to these ModEx steps based on technical needs.

C.3.1 Science Theme 1 (ST-1): Landscape Controls on Hydro-Biogeochemistry (Co-Leads: Xingyuan Chen and Allison Myers-Pigg)

ST-1 Objectives

- Impact of Climate, Biophysical Settings, and perturbations on Hydrologic Connectivity: Generate process-based knowledge of drivers and mechanisms of hydrologic connectivity.
- OM Transport and Transformation: Generate process-based knowledge of how hydrologic connectivity impacts OM transport and transformation and how wildfires change their spatial and temporal dynamics.

ST-1 has several connection points with ST-2 and ST-3 (**Fig. C.3.3**), builds from advances made during the SFA's past quadrennial period (**C.2**), and leverages our unique expertise in hypothesis-driven ModEx to generate transferable science outcomes (**C.2.2.5 to C.2.2.7**). ST- 1 will build from previous work linking molecular-level OM chemistry characterization with watershed hydro-biogeochemical models (**C.2.2.3**), understanding ecohydrologic responses to disturbances (**C.2.2.1**), and the interactions between wildfire and watershed biophysical influences on stream OM content and chemistry (**C2.2.2**). Additionally, ST-1 benefits from our team's leadership and engagement in community ideation and synthesis workshops (e.g., Hydro-ML workshop) and model development efforts (e.g., ExaSheds project). Our collaborations (e.g., Steve Good) in these science areas have also led to advances that include early career scientists from the PNNL-OSU Distinguished Graduate Research Program (DGRP) (**C.6**).

Table C.3.1 ST-1 Science Questions and Hypotheses evaluated at watershed scale.

Question	Hypothesis
1.A Across YRB watersheds, how do climate, vegetation, subsurface hydrogeology and wildfires control the hydrologic connectivity and flow intermittency?	(1A.1) With increasing aridity in time or space, the relative contribution of subsurface transport to streamflow increases due to higher evapotranspiration, but further increases in aridity will break subsurface connectivity and lead to variable inundation. (1A.2) Temporal variations in subsurface connectivity and inundation will increase following wildfire due to loss of vegetation, reduced topsoil permeability, and more variable evapotranspiration.
1.B Across YRB watersheds, how important are the subsurface transport pathways to the transport and transformation of DOM, and how do wildfires impact DOM export under different climate conditions?	(1B.1) Increased contributions of vertical and lateral subsurface transport pathways enhance biotic and abiotic transformations of DOM, resulting in less temporally variable quantity and chemistry of OM exported to the streams. (1B.2) Following severe wildfires in more arid systems there will be less subsurface transport and, in turn, more temporal variability in the content and chemistry of OM delivered to streams; this effect will be larger and longer-lived due to slow ecosystem recovery, relative to low severity fires in mesic systems.

ST-1 Approach and Activities

ST-1 will use empirical datasets as well as numerical experiments using models to test the role of drivers and mechanisms in regulating connectivity of water in **ST-1A** and the export of OM in **ST-1B**. ST1 research will address the questions and hypotheses listed in **Table C.3.1**. In **ST-1A** we will test hypotheses **H1A.1 and H1A.2** regarding the connectivity of water in response to environmental drivers, and in **ST-1B** we will test hypotheses **H1B.1 and H1B.2** by integrating the connectivity results with reactive transport modeling to investigate how biophysical conditions and connectivity changes influence the quantity and composition of OM exported from uplands to streams. We will quantify connectivity as overland or subsurface flows to the streams in space and time. Model simulations will turn on or off the connectivity drivers (e.g. vegetation ET, wildfires, vegetation recovery) to assess their relative impacts on connectivity of water and export of OM and their underlying mechanisms.

The process-based models that will be coupled and applied in ST-1 will leverage the watershed modeling ecosystem in the IDEAS-Watersheds, including the coupling of ATS with PFLOTRAN, and with both ELM-FATES and ELM-FATES-Hydro, along with the interfaces between these models (shown in Figure

C.3.4). These models represent cutting edge representations of surface and subsurface flow and transport (ATS), subsurface reactive transport (PFLOTRAN), and vegetation dynamics and land surface processes with and without plant hydraulics (ELM, ELM-FATES and ELM-FATES-Hydro). Land surface models, including ELM, ELM-FATES (no hydraulics) and ELM-FATES-Hydro will first be run independently and inter-compared to understand the importance of vegetation and vegetation recovery post perturbations. We will then couple land surface models with PFLOTRAN and ATS-PFLOTRAN to provide coupled hydrologic and reactive transport processes from the land surface to surface and subsurface compartments.

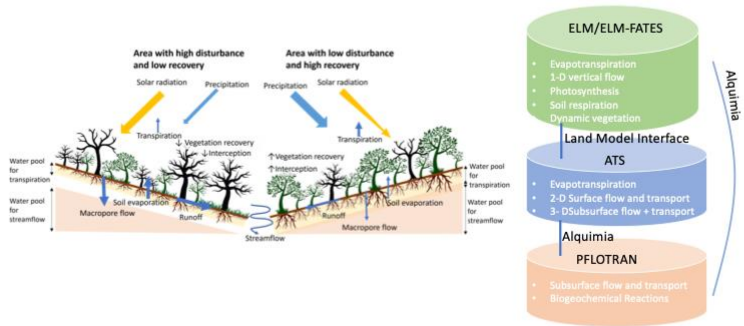


Figure C.3.4: Model couplings necessary (right) to represent processes in each compartment of hillslope-stream modeling efforts conducted in ST-1 (left).

We will evaluate model predictions at select sites where key responses have been measured (e.g., river discharges, variable inundation, OM composition, etc), starting from YRB, where we have performed extensive field measurements and ATS or ATS-PFLOTRAN simulations in the current performance period. A comprehensive list of data that is available within YRB is provided in Table C.3.2. We will make new empirical measurements of OM chemistry, water isotopes, and other variables as guided by our iterative ModEx approach. We will use paired intact and burned watershed design, coupled with our numerical experiments and machine learning, to test their relative impacts of wildfires. We will then extend these models to other watersheds across the CONUS. We will prioritize sites where key responses have been measured, such as stream flows, water isotopes, lateral flow or OM composition. These include sites within existing networks with Open data, such as the Critical Zone Collaborative Network, the Long-Term Ecological Research Network, National Ecological Observatory Network (NEON) and Ameriflux sites.

Table C. Existing data types from studies across the Yakima River Basin (YRB). All sites from each study also have associated observational metadata. Abbreviations are defined as follows: non-purgeable organic carbon (NPOC); total dissolved nitrogen (TN); total suspended solids (TSS); dissolved organic carbon (DIC); fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS), Iron II (Fe); specific conductance (SpC); total dissolved solids (TDS); temperature (temp.); dissolved oxygen (DO); dissolved oxygen percent saturation (DO sat.); Chlorophyll A (ChlA).

Study	No of sites	Data types	Data Package DOI
Temporal 2020-2022 *(Evans Canyon Fire)*	2	**Sample**: NPOC, TN, TSS **Sensor**: SpC, pH, TDS, water temp, water pressure, DO, turbidity	doi:10.15485/1991624

Temporal 2021-2022	6	**Sample**: NPOC, TN, DIC, TSS, Ions, FTICR-MS **Sensor**: barometric pressure, air temp., water temp., SpC, turbidity, pH, TDS, DO, DO sat., water pressure	doi:10.15485/1898912 doi:10.15485/1892054
Temporal 2022-2024	6	**Sample**: NPOC, TN, DIC, TSS, Ions, FTICR-MS **Sensor**: barometric pressure, air temp., water temp., SpC, turbidity, pH, TDS, DO, DO sat., water pressure	Planned for late 2024
Spatial 2021	47	**Sample**: NPOC, TN, DIC, TSS, Ions, FTICR-MS **Sensor**: barometric pressure, air temp., water temp., SpC, turbidity, pH, DO, DO sat., TDS **Other**: water column respiration	doi:10.15485/1898914 doi:10.15485/1892052 doi:10.15485/2283171
Spatial 2022	48	**Sample**: NPOC, TN, TSS, Fe, respiration, grain size, decay rate **Sensor**: barometric pressure, air temp., water temp., water pressure, SpC, turbidity, pH, ChlA, DO, DO sat. **Other**: water column respiration, sediment respiration (ERsed), total ecosystem respiration, depth, width, d50	doi:10.15485/1969566 doi:10.15485/1923689 doi:10.15485/1987520 doi:10.15485/1972232
Spatial 2023 *(Schneider Springs Fire)*	17	**Sample**: NPOC, TN, TSS **Sensor**: barometric pressure, air temp., water temp., water pressure, SpC, turbidity, pH, ChlA, DO, DO sat., velocity **Other**: velocity, depth, width, water column respiration	doi:10.15485/2205676

Key Contributors: Xingyuan Chen, Allison Myers-Pigg, Peishi Jiang, Yi Xiao, Nate McDowell, Kevin Bladon, Heping Liu

Activity ST-1A. Watershed Hydrologic connectivity

Approach: To evaluate Question/Hypothesis 1A.1 and 1A.2 (**Table C.3.1**) and address our first objective (Impact of Climate, Biophysical Settings and Perturbations on Hydrologic Connectivity), we will take the steps illustrated in Figure C.3.5.

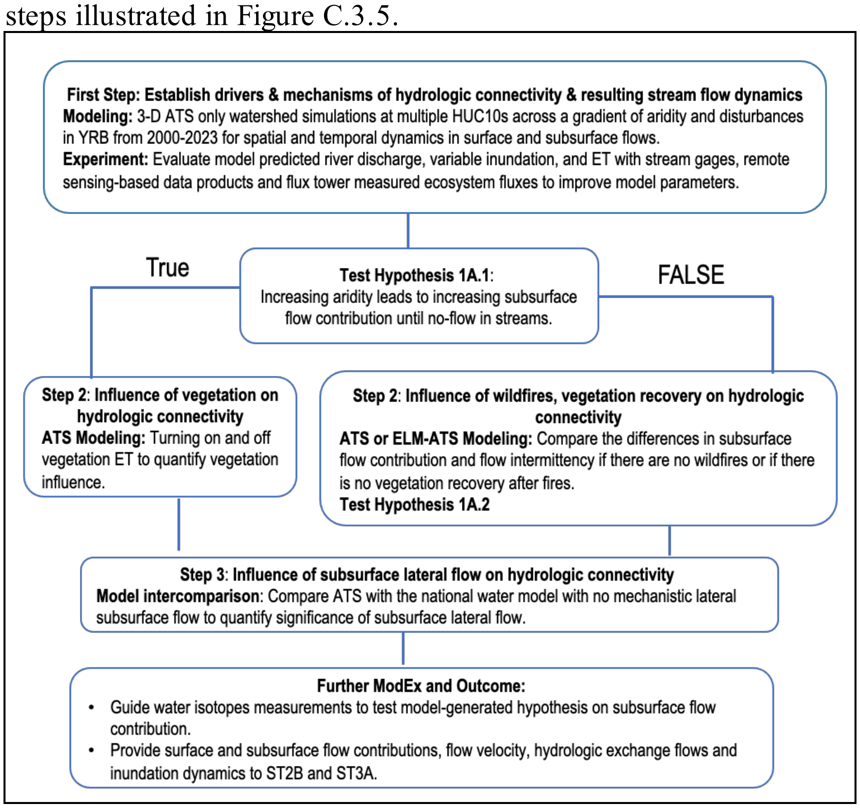


Figure C.3.5 Primary steps that will be taken to test Hypotheses 1A.1 and 1A.2, and address Question 1A.

__Step 1__: We will first establish drivers and mechanisms of hydrologic connectivity and resulting stream flow dynamics using 3-D ATS-based watershed simulations across a gradient of aridity and disturbances in YRB. We will start from the Naches River watershed and Satus River watershed, which include multiple HUC10 watersheds covering a wide range of aridity and disturbances as shown in Figure C 3.6. We will perform simulations from 2000-2023 to obtain spatial and temporal dynamics in surface and subsurface flows for all these watersheds. We will use Daymet as the meteorological forcing to drive our models while using remote sensing-based LAI from MODIS to parameterize the changes in vegetation above-ground biomass. We will use spatially-distributed databases such as the digital elevation map (DEM) from USGS, soil properties from SSURGO, and geologic layers from GLYMPS to set up the topography and subsurface hydrogeologic properties. The spatial resolution of the ATS model will be 50-100 m along the river and gradually coarsened to 1000 m to more upland areas. Hourly observations of channel water depth, discharge, and overland and subsurface fluxes to the river reaches will be generated for all the watersheds. For watersheds impacted by wildfires, we will incorporate biophysical changes induced by wildfires, through loss of transpiration, changes in surface energy balance, and fire-induced changes in soil hydraulic properties, following our previous work by Li et al. (submitted to Water Resources Research).

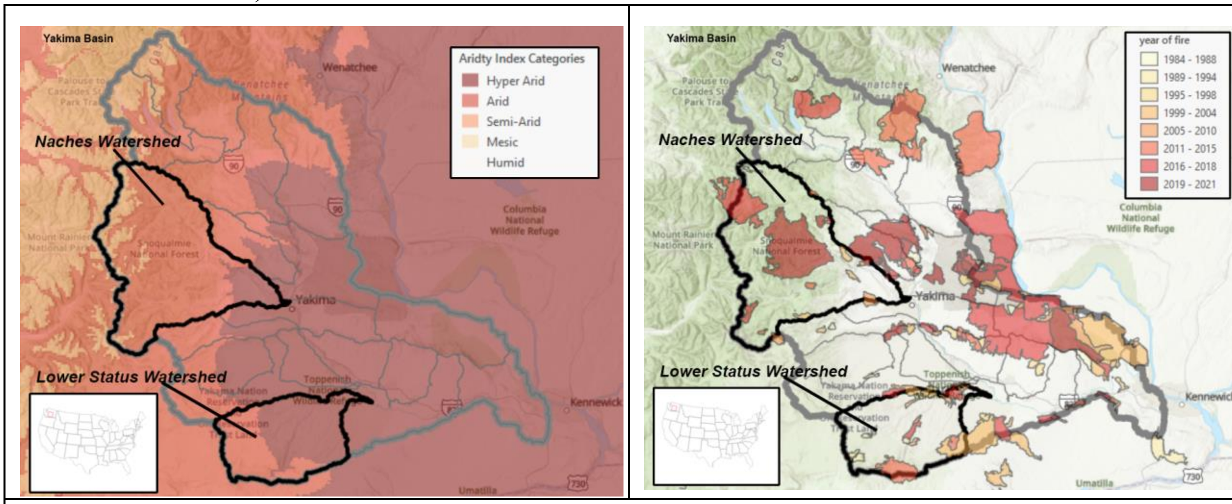


Figure C.3.6 Maps of Naches and Satus Watersheds in YRB, with the HUC10 boundaries, the aridity indices (left), and fire scars (right).

Model simulations will be confronted against observations from multiple sources. We will evaluate the ATS simulation on discharge and inundation with stream gages and remote sensing based variable inundation statistics. Any mismatch between simulated and observed water fluxes would potentially improve the estimation of soil hydraulic properties, initialized by public data source, through a recently developed knowledge-guided deep learning method (Jiang et al., 2023). Further, we will evaluate the evapotranspiration simulation with measurements from AmeriFlux tower sites or from MODIS data products to estimate canopy physiological parameters in ATS.

Variable inundation dynamics across stream networks and at field sites are essential to evaluate hypotheses in ST-2 and ST-3. To enhance model performance, we will evaluate ATS model predictions against near-daily multispectral imagery from Planetscope at 3m resolution. This optical imagery will be supplemented with commercial radar backscatter imagery at sub-meter resolution to capture hydrological conditions in the river corridor through cloud cover. All imagery will contribute to a classification of hydrological intermittency per stream reach. These data will also be used to characterize riparian vegetation and inundation dynamics in stream reaches wider than 3m. A range of signal processing and classification methods will be tested to detect sub-pixel inundation dynamics. These detection methods will be benchmarked against *in situ* measurements (**Fig. C.3.13**). For streams narrower than 3m, we will acquire commercial Synthetic Aperture Radar (SAR) imagery the Umbra Space company. The sub-meter resolution of this SAR imagery and capacity to penetrate cloud cover is a promising approach for detecting variable inundation across stream networks within the YRB and across CONUS.

After the ATS models are evaluated and improved, we will quantify the contribution of subsurface flow to overall stream discharge over time and space, calculate the no-flow days, and characterize variable inundation characteristics. To test **Hypothesis 1A.1,** we will develop ML models to relate subsurface flow contributions, variable inundation features with climate as represented by the aridity index, land cover, subsurface hydrogeology, disturbances and baseflow contribution to rank relative importance of these controlling factors. We will also perform partial dependence plots (PDPs) to examine whether there are detectable trends between each controlling factor with resulting subsurface flow contribution and flow intermittency, thus testing **Hypothesis 1A.1**.

Step 2: To understand the role of vegetation plays in hydrologic connectivity, we will run ATS simulations in Naches and Satus Creek watersheds with evapotranspiration turned off and compare the connectivity and flow intermittency results with the baseline simulations. For those watersheds impacted by wildfires, we will also perform ATS simulations assuming no vegetation recovery in fire scars and compare the results against ATS simulations with and without vegetation ET to parse the effects of vegetation recovery post wildfires. The differences in those scenario simulations will allow us to examine the interplay between climate, land cover, subsurface hydrogeology and disturbances, guiding design of field sampling effort to evaluate model predicted flow pathways. We can test **hypothesis H1A.2** by comparing connectivity changes immediately post-fire between with and without wildfires scenarios as a function of the relative amount of watershed LAI loss. Furthermore, the connectivity differences between the scenarios with and without vegetation recovery in subsequent years under different climate conditions will reveal the importance of vegetation succession in fire-impacted watersheds. Upon the availability of coupled ELM-ATS capability from the COMPASS-GLM project, we will use ELM-ATS to perform the scenario simulations for better mechanistic representations of vegetation ET than ATS only simulations.

Step 3: To understand the importance of explicitly accounting for the lateral flows in subsurface flow in predicting variable inundation dynamics, we will perform a model intercomparison between ATS, ELM-ATS (when available), and the National Water Model. We will evaluate the spatial and temporal differences in subsurface soil moisture, baseflow contribution to stream discharge. Water isotopes will be used to evaluate the predicted contribution of subsurface flow from different models.**Activity ST-1B. Transport and Transformation of DOM from Upland to Streams**

Approach: To evaluate Q/H 1.2 and address our second objective (Impacts of Hydrologic Connectivity and Watershed Perturbations on DOM Export), we will take the steps illustrated in Figure C.3.10.

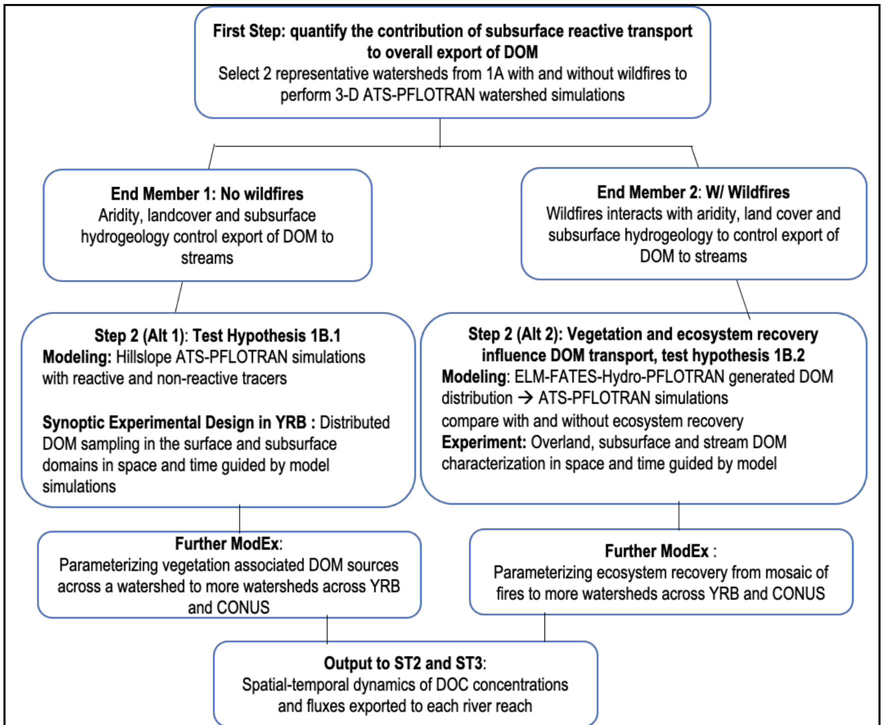


Figure C.3.7

Step 1: We will first perform ATS-PFLOTRAN simulations in a paired HUC12 catchments in Naches or Satus Creek Watersheds based on ST1A variable inundation dynamics. The ATS setup will be same as in ST1A, while PFLOTRAN will simulate the reactions and transport of carbon pools (CWD, litter, DOM) and gases (O_2, CO_2) in the subsurface, with productions rates of CWD and litter from ELM as boundary conditions. The reaction network implemented will include decomposition of CWD and litterfall, DOM respiration, DOM fermentation, and dissolved oxygen consumption.

The carbon budget in fire-impacted watersheds is subject to substantial alterations to accommodate the fire-caused physical and chemical transformations in soil and vegetation. Fires constitute a dual impact on carbon budget: they expedite vertical carbon transference from land to the atmosphere via biomass combustion, and simultaneously generate substantial amount of pyrogenic carbon (PyC) within terrestrial ecosystems. To account for the fire-derived impacts of PyC to the ecosystems, we will represent the fates of PyC and PyC-caused elevated carbon source as the form of dissolved organic carbon (DOC) in wildfire-impacted watersheds. Prior to wildfire events, the terrestrial sources of DOC mainly come from the rooting zones in forested areas, as well as the leaching from soil carbon pool. After wildfires, the PyC transport and its transformation into DOC brings an extra process to govern carbon cycling. To quantify the impact of PyC in ATS-PFLOTRAN, we will use the Monitoring Trends in Burn Severity (MTBS) and the Burned Area Emergency Response (BAER) datasets to parameterize the pre- and post-fire DOC sources in wildfire-impacted terrestrial areas.

The initial spatial and temporal patterns in DOC export to each reach across the Naches and Satus Creek watersheds will be used to select representative hillslopes within those watersheds to run more detailed scenario simulations using multiple models to understand

<u>Step 2</u>: On the selected hillslopes, we will predict how differing hillslope-stream hydrologic connectivity impacts organic carbon (OC) transport from hillslopes to stream networks by employing hillslope-scale 2D transect modeling using ATS-PFLOTRAN, ELM-PFLOTRAN, and fully coupled ELM-ATS-PFLOTRAN (**Fig. C.3.4**). We anticipate that changes in vegetation, physical processes and hydrologic parameters will lead to change in OC fluxes to stream channels, which we will test with a series of *in silico* model experiments with increasing process-based representation as we increase model couplings (**Fig. C.3.4**). Compared with ELM, ELM-PFLTORAN has a more mechanistic representation to both vertical and lateral flows, thus it will enhance our understanding to the OM transport in the subsurface at watershed scale and its interaction with vegetation and hydrological exchanges. Comparisons to Parflow and ATS will provide guidance on the critical processes that must be represented below and above ground to accurately predict connectivity, with further knowledge gained about the role of vegetation hydraulics through comparison of FATES and FATES-Hydro. These simulations will be run over time before and after disturbances to allow hypothesis testing regarding the mechanisms by which disturbance severity and vegetation recovery influence connectivity. These simulations will allow the testing of Hypotheses 1B.1 and 1B.2.

Scenarios will be designed to understand the combined impact of various biophysical settings – climate (VPD and precipitation), landcover, and subsurface hydrogeology – on hydrologic connectivity and OM transport and transformation, which will culminate in model outputs of predicted OM fluxes (kg OC km^{-2} watershed area $year^{-1}$). In the coupled ELM-PFLOTRAN model, we will use ELM model to simulate photosynthesis, allocation of carbon to growth, respiration and background carbon turnover such as litterfall and mortality. Water stress on gas exchange will be predicted from soil moisture simulated by PFLOTRAN. Reactions in ELM-PFLOTRAN will match what we will use in 3D ATS-PFLOTRAN simulations.

We will assess the variable flow path contributions (surface, distal subsurface, and near subsurface) to dissolved organic matter inputs to the stream network at two reaches, building upon the extensive body of literature that leverages hydrographic mixing models to address hydrologic flow paths to the river corridor (reviewed in Klaus and McDonnell, 2013), which is commonly done for water pathways but less commonly employed for tracking DOM flow path dynamics (Saiers et al., 2021). We will build upon literature that performs inversions of these mixing models to assess DOM source water composition and concentrations (Saiers et al., 2021; Miller et al., 2017). This approach is complementary to assessment of hydrologic connectivity and DOM sources via concentration-discharge relationships (Evans and Davies, 1998), including DOM chemistries (Wagner et al., 2019).

To collect the necessary datasets for DOM source mixing models, in-stream sensors will be deployed equipped with specific conductivity and fluorescence dissolved organic matter (FDOM) measured at high-frequency (sub-hourly), in collaboration with ST-2B. FDOM data will be corrected for effects of stream water temperature, turbidity and light absorbance of DOM using standard methodology (Downing et al., 2012) and YRB specific turbidity corrections based on relationships between turbidity and light attenuation, as these relationships are dependent on particle size and composition (Regier et al., 2020). Sensor deployments will be coupled with lower temporal resolution synoptic sampling over different portions of the hydrograph for water isotopes and organic matter composition (FDOM and high-resolution mass spectrometry) co-located with deployed sensors, and from end members (e.g., precipitation, the hyporheic zone, soil leachates, and vegetation leachates), which will be used in the development and verification of the mixing models.
The continuous measurements on DOC chemistry will facilitate characterizing subsurface reaction network used by ELM-PFLOTRAN and ATS-PFLOTRAN. For each measurement location, we will perform 2D transect reactive transport modeling and evaluate the DOC simulation against the

measurements. We will estimate the parameters of the reaction network, including the rate coefficient of DOC production that is assumed uniform across the 2D transect and linearly controls the DOC generation with other environmental factors (e.g., soil moisture and leaf area index). Mutual information-based sensitivity analysis will be first performed to identify the dominant parameters, followed by a deep learning-based inverse mapping to estimate the sensitive reaction parameters from DOC measurements. The parameter estimation of these photosynthesis and stomatal conductance processes in ELM will be performed by using a differentiable land surface model with and without wildfire impacts.

To test H1-B.1, DOM compositional and source information will be linked with hillslope models. These approaches will also be used to test H1-B.2 will be employed in fire-impacted streams across aridity gradient, namely across the Naches and Satus Creek Watersheds.

Further ModEx: We will evaluate the need to instrument any hillslopes in YRB through model-guided experimental design. Beyond the YRB, we will run FATES and FATES-Hydro at paired burned and unburned watersheds across CONUS with sufficient streamflow records pre and post wildfires. The CONUS-wide datasets will include streamflow, DOM, plant and soil hydraulic and photosynthetic traits from public databases, MODIS LAI and burn severity, SMAP vegetation optical depth for soil and canopy water contents, and GEDI Lidar for stand biomass. A subset of watersheds that have intensive monitoring of other hydrologic and vegetation processes, such as evapotranspiration, growth, and soil water content, and will be used for evaluation if the models get the right result (streamflow, sometimes with DOM) for the right reason. These key processes will be assessed in relation to subsurface characteristics, climate and CO2 change (all watersheds will have at least 20 years of records), and disturbance.

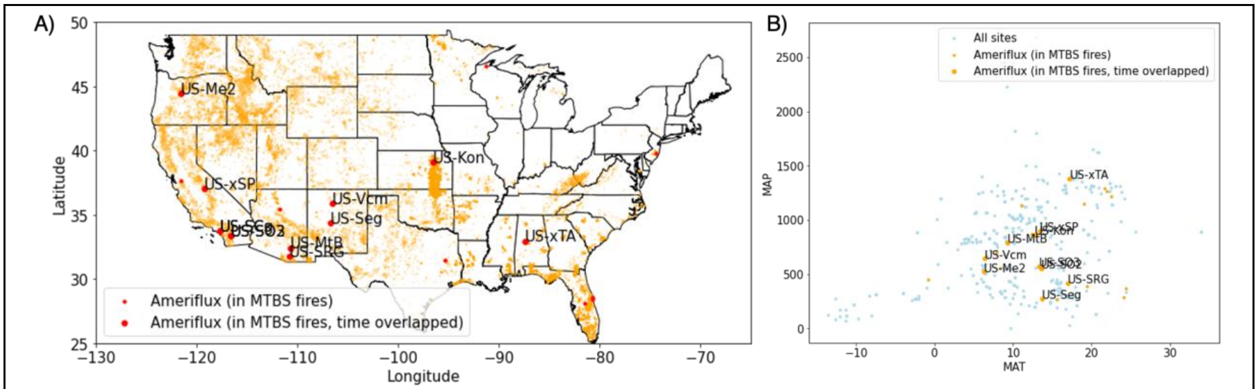


Figure C.3.8. A) Locations of AmeriFlux sites affected by fires. B) Wildfire-impacted AmeriFlux sites span a range of climate conditions illustrated by mean annual precipitation (MAP) and mean annual temperature (MAT).

ST-1 Outcomes:

ST-1 will produce key hydrologic and OM transport process-based knowledge and characterizations necessary for ST-1 and ST-3 ModEx loops. ST-1A will produce hillslope and watershed scale predictions on hydrologic flow paths, OM transport and streamflow patterns necessary to ST-2A efforts. ST-1B will include parallel characterizations under the influence of wildfires, including predictions on where sites may sit upon the ecohydrological recovery trajectory spectrum, necessary for ST-2B efforts. Process-based understanding and observations of hydrologic flow paths and OM fluxes with and without wildfire perturbation is essential for ST-3 science objectives and scaling from reach-to-network scale objectives in ST-3. At the highest level, ST-1 will provide critical information on reach-scale water delivery pathways and OM fluxes as summarized in the output boxes in **Figures C.3.5 and C.3.7.** We anticipate that ST-1 will uncover that specific water delivery pathways and OM fluxes will be dependent on knowable mechanistic relationships with biophysical properties observable via remote sensing, allowing for computationally viable surrogate model inputs necessary for the hybrid model developed in ST-3A. The associated knowledge, data, and models from ST-1 will be essential for ST-3 to represent the watershed biophysical underpinnings necessary to robustly reveal time-varying allometric scaling of ERsed. ST-1 will further inform site selection in ST-2, and provide a basis for the watershed model development in ST-3.

C.3.2 Science Theme 2 (ST-2): Impacts of Variable Inundation and Wildfire on ERsed from Molecular-to-Reach Scales (Co-Leads: Allison Myers-Pigg and James Stegen)

ST-2 Objectives

- Interaction of PyOM and Variable Inundation on ERsed: Combine models, data, and conceptual advances from ST-1B with reaction-scale modeling and laboratory observations within and across basins to understand and predict how fire and variable inundation induced changes in the quality, quantity and delivery of OM fuel ERsed.
- Impact of Flow Path, Velocity and Sediment Texture on ERsed: Combine field sampling data and modeling efforts to generate statistical distribution of ERsed post-inundation across a range of hydrologic connectivity mechanisms and reach scale conditions to understand the mechanistic processes controlling ERsed hotspots at the reach scale.

ST-2 has several connection points with ST-1 and ST-3 (**Fig. C.3.6**), building from several advances made during the SFA's past quadrennial period (**C.2**). In ST-2 we will leverage our unique expertise in integrating distributed participatory science, intensive place-based approaches and hypothesis-driven ModEx to generate transferable science outcomes (**C.2.2.5 to C.2.2.7**). ST-2 will specifically build from our previous work on characterizing and predicting ERsed rates across CONUS (**C.2.2.5**), identifying generalizable trends in OM chemistry and microbial genomics (**C.2.2.6**), understanding wildfire impacts on OM chemistry (**C.2.2.2**), and the deepening of hypothesis-based ModEx approaches used across the project (**C.2.2.7**). Additionally, ST-2 benefits from our team's leadership and engagement in community ideation and synthesis workshops and manuscripts, through active engagement with the Dry Rivers Research Coordination Network and leading a wildfire-focused workshop. Our formal (e.g., Dr. Kevin Bladon; **Apx 3**) and informal (e.g., Dr. Sasha Wagner) collaborations in these science areas have also led to advances that include early career scientists from the DOE-Office of Science Graduate Student Research (SCGSR) and PNNL-OSU DGRP programs, which we will build upon via our PIER plan (**C.6**). We also formally collaborate with Dr. Bob Hall, who is a global expert on estimation of reach-scale ERsed, and informally with Dr. Sarah Roley, an expert in nutrient dynamics and whole-stream metabolism. We further have regular discussions with the associated research community to share experiences and collectively advance technical strategies. We are actively engaged with teams working on solving these technical challenges, including several researchers associated with the National Science Foundation (NSF)-funded AIMS project led by Dr. Amy Burgin (an informal collaborator).

Table C.3.2 ST-2 Questions and Associated Hypotheses evaluated at batch and reach scales.

Question	Hypothesis
2.A Across YRB and CONUS watersheds, how do variable inundation and PyOM interactively influence ERsed via changes to microbial communities and OM chemistry?	Variable inundation increases the expression of microbial genes related to the degradation/use of less bioavailable OM (e.g., more aromatic OM). This selects for microbes in non-perennial systems that can efficiently use aromatic and/or less thermodynamically favorable components of PyOM. This results in a more positive influence of PyOM on ERsed in variable inundated sediments, relative to consistently inundated sediments.

2.B Across YRB reaches, what is the optimal combination of flow path, flow velocity, inundation history, and sediment texture that maximize ERsed post-inundation?	ERsed following re-inundation is maximized during intermediate flow velocity, intermediate durations of dry conditions, and dominance of overland flow in locations with sediment textures that support algal growth. Higher turbidity for a given flow velocity following wildfire will result in the ERsed maximum being at lower velocity in wildfire impacted reaches.

ST-2 Approach and Activities

To evaluate hypotheses 2.A and 2.B, ST-2 will work from batch to reach scales. Batch scale refers to laboratory microcosms or bottle experiments, which are conducted across multiple experimental conditions, these will be the primary focus of ST-2A. Reach scale efforts are conducted at stream and river field sites across the YRB and will be a major focus of ST-2B. Modeling and data generation in ST-2 will integrate molecular scale processes focused on OM chemistry and microbial metabolic function as variables regulating ERsed, and examining how these interact at the reach scale with physical attributes such as flow path, inundation history and flow velocity. As in ST-1, for YRB elements, ST-2 will work with regional partners (e.g., Confederated Tribes and Bands of the Yakama Nation) with elements of co-design to ensure mutual benefit in alignment with ICON. For CONUS elements, ST-2 will take a mutually-beneficial participatory research approach via collaboration with the WHONDRS consortium (**C.6**).

Batch-scale efforts will focus on identifying how wildfire and variable inundation interact through alterations to OM chemistry and microbial communities that, in turn, influence ERsed. Reach-scale efforts will link ST-3 outcomes focused variable inundation dynamics with our recent ModEx-based advances that bring together process-based models and field data generation guided by model-based hypotheses (**C.2**).

Field estimates of ERsed will be conducted primarily via *in situ* dissolved O_2 time (DO) series [116]. Estimating ERsed combines single-station DO sensors for whole stream respiration (ERtot) with *in situ* dark bottle incubations to isolate water column respiration (ERwc). Respiration directly and indirectly associated with sediments (ERsed) is found the difference between ERtot and ERwc:

$$ERsed = ERtot - ERwc$$

Key Contributors: Allison Myers-Pigg, James Stegen, Jianqiu Zheng, Vanessa Garayburu-Caruso, Bob Danczak, Bob Hall, Yunxiang Chen, Peter Regier, Etienne Fluet-Chouinard, Amy Goldman

Activity ST-2A. Interaction of PyOM and Variable Inundation on ERsed

Approach: To evaluate Q/H 2.A (**Table C.3.2**) and address our first objective (Interaction of of PyOM and Variable Inundation on ERsed), we will follow the approach laid out in Figure H2.1.

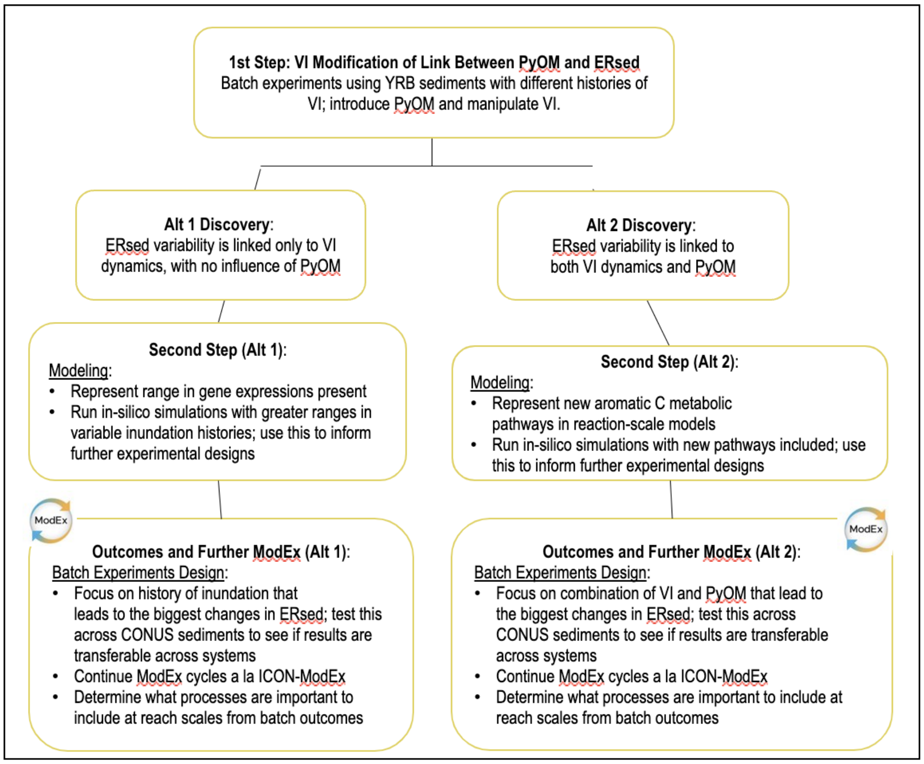
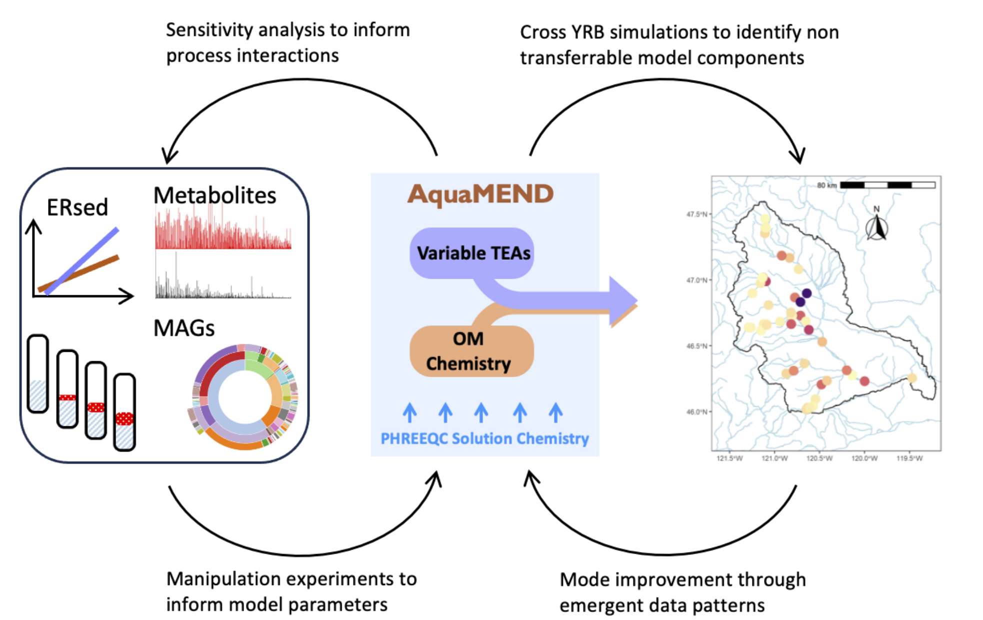


Figure H2.1. First, second, and further steps of the ModEx process associated with testing Hypothesis 2.A

Hypothesis 2.A evaluation in the YRB (**1st ModEx step; Fig. H2.1**): We will first conduct laboratory batch-scale experiments focused on the interaction between PyOM and ERsed that will use sediments collected from the YRB. These experiments will start with four treatment types: 1) Fully Inundated with No PyOM, 2) Fully Inundated with PyOM, 3) Variably Inundated with No PyOM, and 4) Variably Inundated with PyOM. In each, oxygen drawdown will be measured to estimate ERsed in analogous ways to previous WHONDRS and ICON-ModEx experiments, for interoperability. We will examine OM chemistry via targeted and non-targeted approaches (benzene polycarboxlic acids as a proxy of aromatic PyOM and FTICR-MS high-resolution mass spectrometry), coupling this with information on gene expression via metagenomics and meta transcriptomics. These datasets are vital to understanding what OM chemistry and microbial variables are changing in relation to ERsed. The variability in ERsed across treatments will be correlated to shifts in OM chemistry and microbial gene expression. We anticipate that there will be greater polyphenolic degradation genes expressed in batches that contain PyOM. We will assess this with the CAMPER database, which contains curated annotations for profiling microbial polyphenol metabolic potential[117]. We expect that expression of genes associated with aromatic degradation (e.g., catechol and protocatechuate dioxygenases/decarboxylases) will be positively related to ERsed in variably inundated systems but have no relationship to ERsed in fully inundated systems. Concurrently, PyOM signatures will decrease in variably inundated systems, as these molecules are utilized by microbial communities (e.g., NOSC will increase, aromaticity will decrease).

Include figure of AquaMEND simulations with aromatic OM

Hypothesis 2.A evaluation *in silico* experimentation (**2nd ModEx step; Fig. H2.1**): Following these experimental results, *in silico* experiments will be performed with microbially explicit 0-D modeling with AquaMEND, developed in Zheng et al. [98]. Specifically, we will incorporate the relative contribution of each pathways to tune the model to modify their relative expressions with numerical experiments. The thermodynamic model [98] predicts microbial function and ERsed via mechanistic constraints imposed by physical accessibility and thermodynamic favorability of OM. The simulations will generate new insights into what mechanisms are mediating the cascade of influences from variable inundation to OM chemistry to microbial gene expression to ERsed. The thermodynamic model [98] predicts microbial function and ERsed via mechanistic constraints imposed by physical accessibility and thermodynamic favorability of OM.



Modex approach figure needs to be updated for ST-2A objectives and outcomes ie. YRB picture, and aquaMEND modeling

Hypothesis 2.A evaluation across CONUS (**Further ModEx; Fig. H2.1**): Results from the *in silico* experiments will be used to guide follow-on batch scale experiments and knowledge gained will also inform reach scale efforts in ST-2B. We will expand up batch experiments with wider ranges of wet and dry starting conditions and manipulations (8 treatment types): 1) Historically Dry Sediment with No PyOM re-wetted, 2) Historically Wet Sediment with No PyOM dried and rewetted, 3) Historically Dry Sediment with PyOM re-wetted, 4) Historically Wet Sediment with PyOM dried and rewetted, 5) Recently Dry Sediment with No PyOM re-wetted, 6) Recently Wet Sediment with No PyOM dried and rewetted, 7) Recently Dry Sediment with PyOM re-wetted, 8) Recently Wet Sediment with PyOM dried and rewetted. To enhance transferability of our YRB batch scale experiments, we will assess impacts of OM chemistry on microbial functions using sediment samples collected across CONUS (via WHONDRS) and a similar experimental design as explained above for the YRB. The exact sites targeted in the CONUS efforts will depend on the results from 1st ModEx step, but sediments collected will span divergent wildfire impacts, variable inundation histories (i.e., number of days a system has been wetted or dried), and sediment textures. Exact sites will be focused on variable inundation history and /or the degree of wildfire-impacts, depending on the outcomes from the previous steps (**Fig. H2.1**). To assess generalizability of the importances of these mechanistic results across systems, we will run *in silico* experiments across a broader range of sites and sediment textures. [ADD IN MORE DETAILS ON HOW RESULTS WOULD GUIDE NEXT MODEX STEPS 3-5 sentences]

We will derive conceptual understanding from the comparison of new data to model predictions and formulate that new understanding into updated hypotheses to guide further modeling studies. For example, knowledge gained from the YRB and CONUS batch scale experiments will guide microbially explicit modeling development relevant for ST-3 objectives and together will guide knowledge gained on what microbially mediated OM processing reaction networks are important moderators of ERsed, which can be tested at the reach (ST-2B) and network scales (ST-3).

Activity ST-2B. Influence of Hydrologic Connectivity and Reach Scale Dynamics on ERsed

Approach: Reach-scale efforts in ST-2B will focus on evaluating H2.B, related to how variation in water delivery pathways, flow velocity, inundation history and sediment texture collectively influence mechanisms supporting ERsed (such as optimal sediment texture to support algae growth) following re-inundation, and how wildfires modulate these dynamics. To evaluate Q/H 2.B (**Table C.3.2**) and address our second objective (Impact of Flow Path, Velocity, and Sediment Texture on ERsed), we will follow the approach laid out in Figure H2.2.

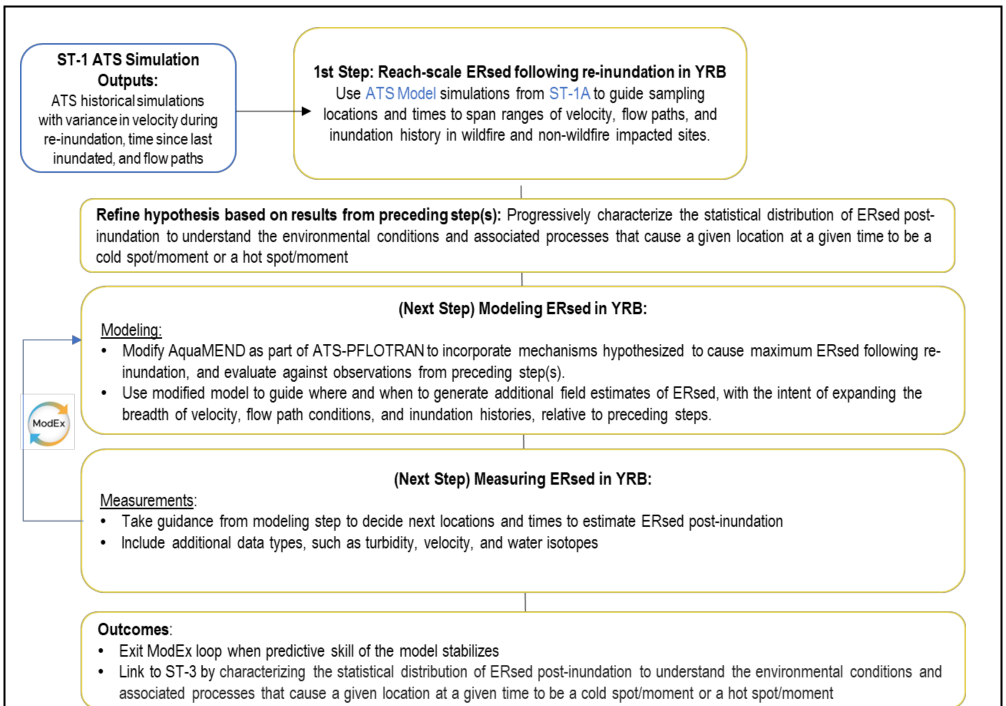


Figure H2.2. First, second, and further steps of the ModEx process associated with testing Hypothesis 2.B

Hypothesis 2.B evaluation in the YRB (**1st ModEx step; Fig. H2.2**): Reach scale field site selections will be guided by ATS hydrological simulations conducted in ST-1. We will first select ten reaches within the YRB that (1) that array across flow path, flow velocity, inundation history, and sediment texture and (2) where the simulations have the largest uncertainty in ATS simulations.

To select the first ten reaches for field data collection of ERsed, we will need to first gather information across the YRB to assess locations where flow path, inundation history, flow velocity, and sediment texture are varied.

Flow path and velocity: To complete reach field site selections, we will group all the river reaches from ST-1 watershed-scale ATS modeling informed by the remote sensing of variable inundation, into clusters. K-means clustering will be performed using local features including locations, wildfire history (percent drainage area burned), soil hydraulic properties (SSURGO), land cover type (NLDC), mean annual temperature (Daymet), mean annual precipitation (Daymet), and mean annual streamflow (ATS). Then, in each cluster, we will pick the sampling reach whose exchange fluxes exhibits the largest uncertainty or information. The amount of information will be calculated by using the Shannon's entropy of HEFs outputted from the ATS simulation.

Inundation history: To better assess inundation history, we will acquire SAR images for identification of variable inundation in YRB reaches. The tasking schedule and framing of commercial satellite imagery will be informed by ATS simulations to capture wetting/drying transitions. Initial tests have shown limitations of remote sensing at the 3-5m resolution over narrow and vegetated stream channels (**Fig. C.3.14**). Our mitigation strategy for this is the inclusion of sub-meter SAR imagery (**Fig. C.3.14**). However, the capacity of commercial SAR platforms to detect meter-scale stream channels remains uncertain. Moreover, the tasking schedule of satellite image acquisitions informed by low-latency ATS predictions will require imputation with novel data streams (e.g., a regional climate model or interpolated weather station data) in between Daymet data releases. Differences in meteorological forcings can impact ATS outputs [115]. [1-2 sentences about mitigation strategy for the forcings] Iteration between ATS simulations and remote sensing data will be conducted in close collaboration with ST-1. We will pursue the capacity to autonomously task Umbra satellites using automatic analysis of ATS model predictions. This will enable semi-autonomous ModEx between ATS and satellite imagery, essential to informing reaches to target sampling upon resumption of flow to test H2.B. Resulting ATS predictions will also be integral to the enhanced RCM in ST-3A and provide spatiotemporally resolved information on stream network hydrologic connectivity and reach-resolved inundation histories.

Figure C.3.14: Comparison of commercial remote sensing images. Optical from Planetscope (left) and SAR from Umbra (right) over the same location. Umbra's sub-meter resolution and cloud-penetration captures narrow canals undetected by Planetscope.

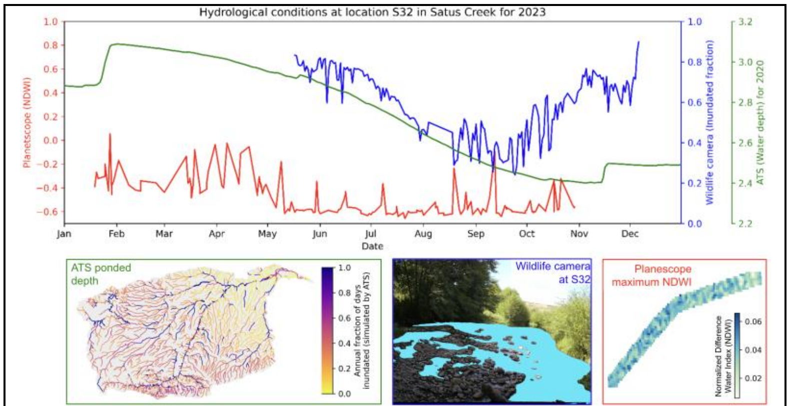


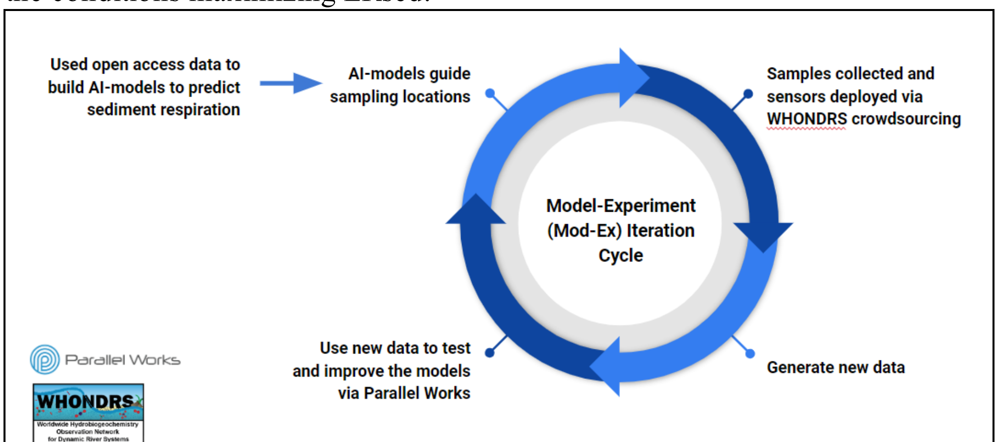
Figure C.3.13: (top) Comparison of hydrological conditions at location S32 in Status Creek over 2023 as captured by: ponded water depth from ATS (bottom-left), inundated stream fraction from *in situ* wildlife camera (bottom-center) and NDWI from Planetscope imagery at 3m resolution (bottom-right). The comparison highlights the challenge of monitoring variable inundation in small streams. Sub-meter resolution satellite imagery is expected to improve monitoring at remote locations.

Sediment texture: We will use ML and multi-source multi-resolution photos to assess sediment texture at the reach scale. ML models will be trained using field photos and ERsed measured at batch (ST-2A) and reach scales (ST-2B). We will collect photos from wildlife cameras, smartphones, drones, and remote sensing at YRB. Spatial resolution varies across these approaches from ~1 mm to ~10m. Batch scale ERsed is measured on fine-grain sediment (<2mm) and the fraction of *in situ* streambed composed of fine-grain sediment will be determined from the photos. The fractional contribution of fine-grain sediments will be combined with batch-based ERsed to partially constrain ML-predicted reach-scale ERsed. These constraints will be merged photo-based training of ML models using multi-step and end-to-end neural networks. In the multi-step approach, we will first train a ML to detect and quantify biophysical features such as grain size (**C.2.2.7**) and algal cover. This will use object detection and segmentation AI. We will combine photo-derived features with other non-photo data to train an additional ML model for reach-scale ERsed, constrained by batch-scale ERsed. In the end-to-end neural networks approach, the model directly learns the relationship between input photos and ERsed using a vision transformer (ViT). The important features that control ERsed can be determined from the ViT attention-maps generated by this method. The ML models developed in ST-2B will further be used to correlate remote sensing data with features in the photos to facilitate ST-3 network-scale predictions of ERsed in response to variable inundation.

Site Deployments: At each selected site, we will co-deploy DO, pressure and turbidity sensors in those reaches when the intermittent streams are flowing in areas of slope breaks and estimate metabolism parameters, before flow ceases. These measurements will facilitate an accurate estimate of gas exchange that will be used to estimate metabolism as soon as flow resumes, to test H2.B. Areas of slope breaks allow for spots with turbulent flow as a river dries into non-flowing pools. Additionally, we will measure water column respiration and water depth when the river is flowing, as soon as possible upon re-inundation, and once flow stabilizes to better estimate ERsed after re-inundation. This will help to mitigate the technical challenges associated with unsteady flow and low primary productivity upon re-inundation which hinders model-based estimates of metabolism and gas exchange [116,118]. This approach has been developed in consultation with the broader non-perennial stream research community. In highly turbid systems, such as those resuming flow after a wildfire, we will develop relationships between depth, turbidity and metabolism that will help us understand ERsed during high flow events. In such cases we will use gas exchange estimated from similar discharge conditions under low turbidity (i.e., when ERsed is expected to be faster, which makes gas exchange estimation more reliable).

In conjunction with the sensor deployments, we will determine the degree of transferability between our batch scale ERsed analyses in ST-2A and the reach scale by collecting sediment grab samples to analyze for OM chemistries and metabolic gene expressions. We will track expression patterns for aromatic degradation genes identified in our CAMPER analysis from field collected sediments. We will use the existing GROWdb to help resolve species and metabolisms predominantly involved in PyOM degradation across reaches. We will couple these species-resolved observations to ERsed patterns and DOM molecular formula characteristics (e.g., aromaticity, NOSC). This will allow for direct assessment of batch to reach transferability of mechanisms in ST-2A and assist in further ModEx refinement of batch scale experiments in ST-2A. These field data will also be incorporated into the GROWdb to increase inclusion of disturbed systems into the database.

ERsed will also be linked to site-scale information on key properties that may be influenced by wildfires that may impact ERsed, such as streambed grain sizes. The reach scale efforts across the first ten reaches in YRB will provide the datasets required to make a first assessment of the ranges of velocity, flow paths, inundation history and sediment texture that could maximize rates of ERsed across fire and non-fire impacted sites. The relative importance of these influencing factors will be assessed via [some sort of statistical distribution test, open to ideas here] and will guide further ModEx steps to continue evaluating the conditions maximizing ERsed.



Modex approach figure needs to be updated for ST-2B objectives and outcomes

<u>Hypothesis 2.B Further Evaluation of Conditions Maximizing ERsed (**Subsequent ModEx steps; Fig. H2.2**)</u>: We will progressively increase our characterizations of the statistical distribution of ERsed across ranges of velocity, flow paths, inundation history and sediment textures by selecting ten new reaches each ModEx cycle for synoptic deployments during different times of the year, based on iterative further model developments and predictions of new places that ERsed is likely to be maximized. We envision that these deployments will be dynamic in space and time, targeting specific hydrographic conditions that embody different hydrologic flow paths and velocities (e.g., rewetting season, dry season, etc). Sensor deployments will leverage sensor packages equipped with DO, turbidity, and pressure sensors; this will be coupled with targeted sampling of ERwc, water isotopes, OM chemistry, sediment texture, DOC and nutrient concentrations, and MetaG/MetaT , to mechanistically link ERsed with hypothesized in stream controller variables (turbidity, water sources, microbially mediated processing of OM, and sediment texture).

[3-5 sentences on how field/ reach scale measurements from above will be iteratively used in ATS-PFLOTRAN via AquaMEND] Reaction networks from ST-2.A will be simplified via sensitivity analysis to reduce complexity of reaction network to be built into PFLOTRAN sandbox.

We will re-group all the river reaches from ST-1 watershed-scale ATS modeling informed by the remote sensing of variable inundation in a new *k*-means clustering using updated local features of influence on ERsed based on the first round of reach scale efforts, such as sediment textures, runoff ratios, and turbidity and repeat this re-clustering after each ModEx cycle. The ModEx cycles consist of new field sites instrumented and sampled as above in ST-2B 1st ModEx section, providing additional datasets required to refine assessments of the ranges of velocity, flow paths, inundation history, and sediment texture that could maximize rates of ERsed across fire and non-fire impacted sites. The relative importance of these influencing factors will continue to be assessed via their statistical distributions. These updated efforts will guide further model refinements, until predictive power has stabilized.

[paragraph on how this actually would be done similar to ICON-MODEX].

ST-2 Outcomes: ST-2 will merge generated knowledge, models, and data into a summary outcome that maps ERsed across the axes of water delivery pathway, stream network connectivity, and local perturbation history. We envision that specific combinations of values across those three axes will lead to ERsed that is of large enough magnitude to represent an activated ecosystem control point (i.e., a very hot biogeochemical spot/moment with heavy influence over network scale cumulative ERsed). Whatever the outcome, the associated knowledge, data, and models will be essential for ST-3 to robustly reveal time-varying allometric scaling of ERsed. ST-2A will link OM chemistries and microbial processes with batch ERsed rates from variably inundated systems across YRB and CONUS. ST-2B will provide complementary characterization and understanding of ERsed, but at reach scales where physical watershed properties and water delivery mechanisms may interact with molecular mechanisms. Thus, ST-2A and ST-3B will jointly reveal molecular mechanisms underpinning observed ERsed, in context of variable inundation and wildfire, and represent that knowledge in predictive models. Observations and mechanistic understanding of temporally-resolved ERsed following perturbations are fundamental for achieving ST-3 science objectives, including quantification and understanding of temporal variation in ERsed allometric scaling. ST-2 will also provide quantitative insights on the relative influences of variable inundation and wildfire perturbations on the patterns of and processes controlling ERsed. For example, ST-2 will examine potentially divergent recovery trajectories of benthic primary production contributions to ERsed following re-inundation events and with high turbidity following wildfire (ST-2B).

C.3.3 Science Theme 3 (ST-3): Process Integration and ERsed Allometry Dynamics (Co-Leads: James Stegen and Xingyuan Chen)

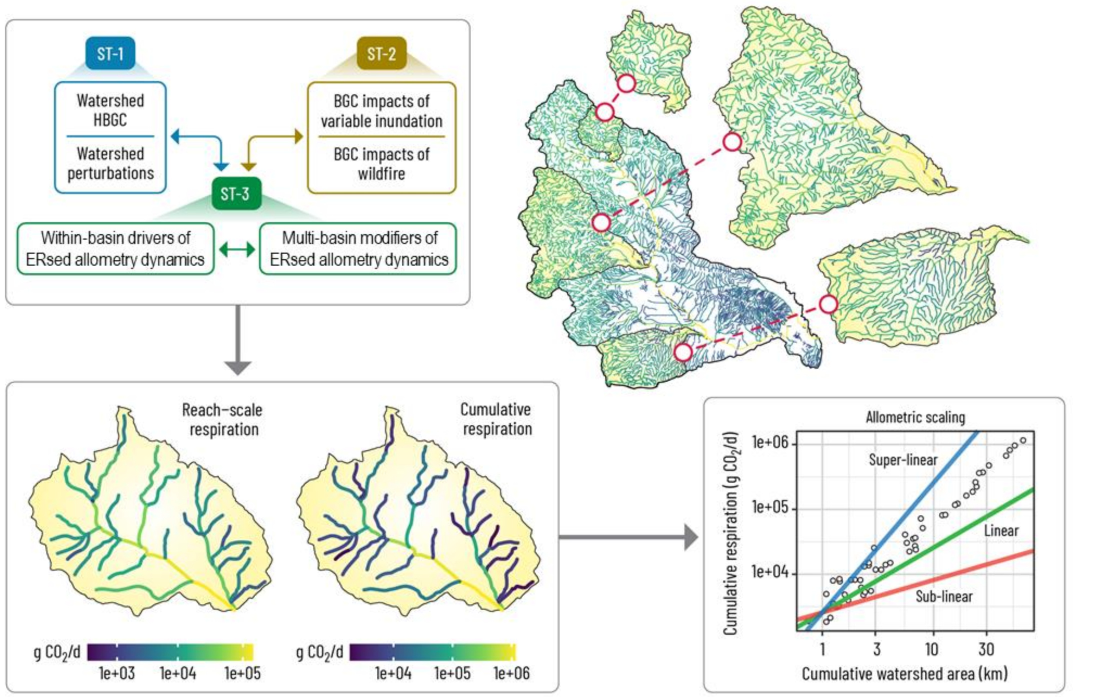


Figure C.3.9: ST-3A will integrate ST-1 and ST-2 outcomes (upper left) into a hybrid model that will make time-varying predictions of ERsed at each reach within a given watershed (lower left). Those predictions will be made across watersheds within the YRB (upper right) and across basins (not shown). ST-3A and ST-3B will use this model and its predictions to test hypotheses related to drivers of ERsed allometry (lower right).

ST-3 Objectives

- Process Integration: Combine models, data, and conceptual advances from ST-1 and ST-2 into a hybrid watershed-scale hydro-biogeochemical modeling capability, referred to as the 'enhanced RCM.'
- ERsed Allometry Dynamics: Use the enhanced RCM along with field data generation to understand the relative influences of different processes and perturbations on temporal variation in the spatial scaling of ERsed (i.e., allometry dynamics).

ST-3 builds from several outcomes across the past quadrennial period (**C.2**), such as the inference that OM chemistry (**C.2.2.6**), benthic primary production, and HEFs are primary controlling factors over ERsed (**C.2.2.4**). Our recent work to characterize and understand whole-stream-network allometric scaling of ERsed (**Fig. C.3.9**; **C.2.2.7**) and the incorporation of OM chemistry into ATS-PFLOTRAN (**C.2.2.3**) are additional advances that ST-3 builds from. ST-3 also benefits from informal collaborations (e.g., Dr. Wil Wollheim) and advances being made via other ESS-funded efforts including other SFAs (e.g., ORNL-led coupling of ELM and ATS) and IDEAS-Watersheds (e.g., integration of genomics into reactive transport codes).

Table C.3.3 ST-3 Science Questions and Hypotheses evaluated at the stream network scale.

Question	Hypothesis
3.A Across YRB watersheds, to what degree do variable inundation, wildfire, and molecular properties individually and interactively influence ERsed allometry dynamics, in context of constraints imposed by HEFs?	HEFs will govern mean slope and intercept of the ERsed allometry, but allometry dynamics will be governed primarily by wildfire impacts on water delivery pathways that lead to re-inundation. At the stream network scale, influences of molecular properties will be overwhelmed by physical setting.
3.B Across CONUS watersheds with divergent biophysical conditions, to what degree do the primary drivers of ERsed	The individual and/or interactive influences of variable inundation and wildfire over ERsed allometry dynamics will

allometry change and are there systematic shifts in the primary drivers across the arid to mesic continuum?	be strongest in watersheds with weaker HEFs, but their influences will diminish moving from arid to mesic climates.

ST-3 Approach and Activities

ST-3 has two activities to jointly address the two Q/H pairs in **Table C.3.3** and the objectives summarized above. Hypothesis evaluation in ST-3 will start with *in silico* experiments using an enhanced version of our previously developed RCM (**C.2.2.7**). The enhancements will integrate components from ST-1 focused on hydrological connectivity across the hillslope-to-stream continuum and components from ST-2 focused on batch-to-reach scale processes in stream networks. ST-3A will use the resulting model to test Q/H 3.1 in the YRB. ST-3B will take guidance from ST-3A outcomes to select locations for reach-scale ERsed estimation across basins spanning an arid-to-mesic continuum. Evaluation of hypothesis 3.1 will also guide efforts across ST-1 and ST-2 aimed at enhancing ATS based on iterating between ATS and remotely sensed and *in situ* data. ST-3B will evaluate Q/H 3.2 via *in silico* experiments after integrating ST-3A data and conceptual outcomes with the enhanced RCM across basins.

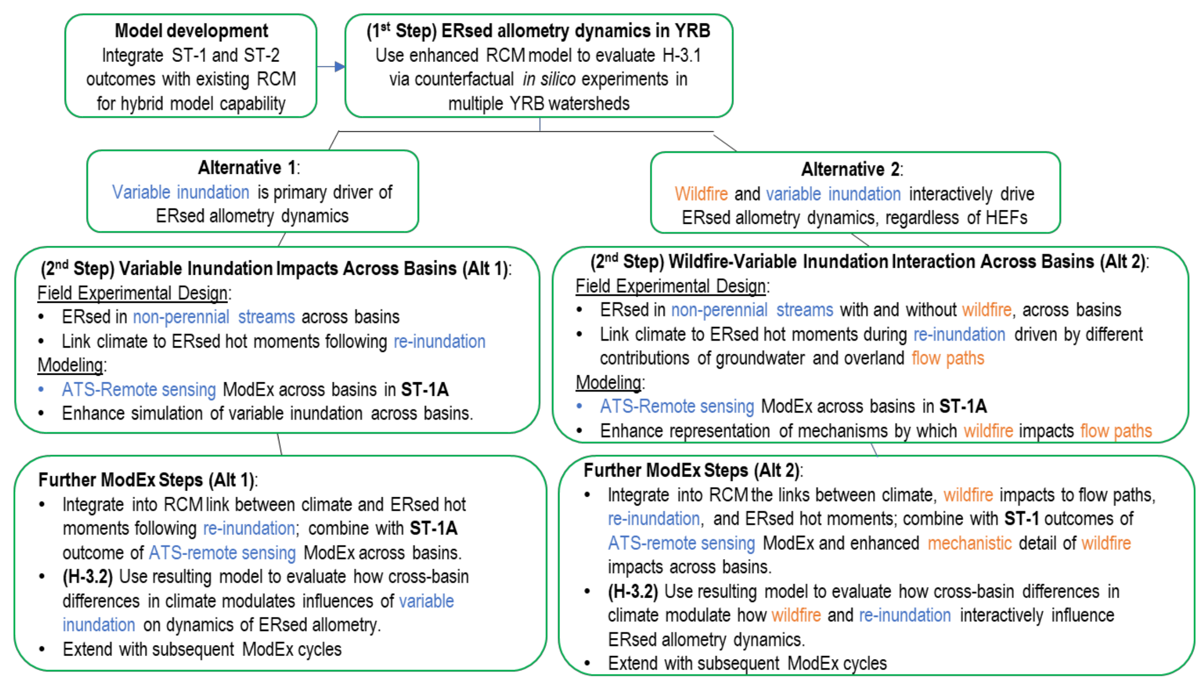


Figure H3.1. First, second, and further steps of the ModEx process associated with testing Hypothesis 3.1

Key Contributors: James Stegen, Xingyuan Chen, Jianqiu Zheng, Peter Regier, Bob Hall, Vanessa Garayburu-Caruso, Etienne Fluet-Chouinard

Activity ST-3A. Model Development and ERsed Allometry Dynamics in the YRB

Approach: Activity ST-3A addresses our first objective (Process Integration) and second objective (ERsed Allometry Dynamics) across watersheds in the YRB, and evaluates hypothesis 3.1.

Model development: To test both ST-3 hypotheses, the current RCM (Son et al. 2022) needs to include additional processes beyond HEFs and subsurface respiration. The RCM is a hybrid model that integrates both process-based and data-driven components. When applied to watersheds in the YRB and across CONUS, the enhanced model will enable quantification of the spatial and temporal variability of ERsed allometry as influenced by watershed biophysical settings, perturbations, and molecular processes. Model development will integrate a hierarchy of process-based simulation models targeting different watershed ecosystem components (e.g., sediment, vegetation). This integration will facilitate mechanistic understanding of how interactions among biophysical setting, wildfire, and variable

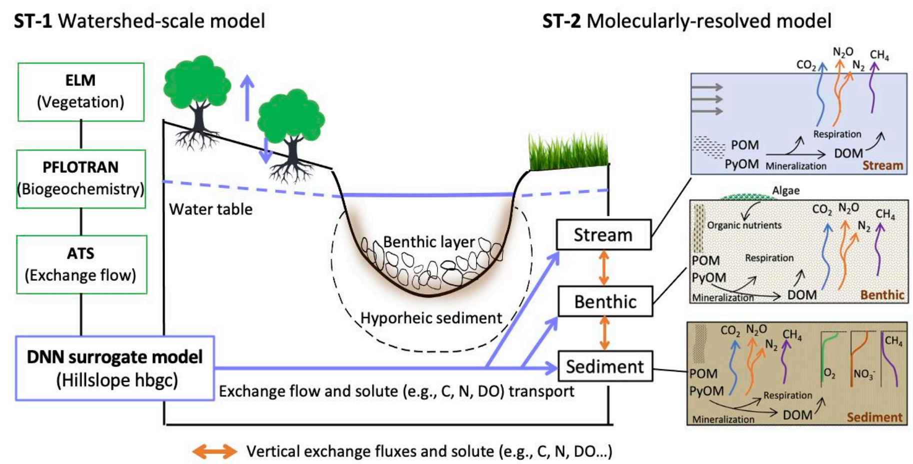


Figure C.3.10: To mechanistically link hillslope-to-stream processes and perturbations to ERsed allometry dynamics across stream networks we will enhance the RCM built by our team (**C.2.2.7**). We will integrate ST-1 outputs related to hydrologic connectivity and OM export and ST-2 outputs related to patterns and processes underlying batch-to-reach scale ERsed.

inundation contribute to spatiotemporal patterns of reach-to-network scale ERsed. ST-3A model development will further enable the expansion of molecular mechanisms identified in ST-2 to the stream network scale by linking molecular-scale models from ST-2 (e.g., AquaMEND) with watershed modeling from ST-1 (e.g., ATS-PFLOTRAN). This will enable evaluation of the changing influence of molecular mechanisms on ERsed along the continuum from batch to reach to stream network scales.

The batch-to-reach scale modeling efforts from ST-2 will translate mechanisms into simplified process representations to inform integrated RCM development. ST-2 will deliver three sets of reaction networks to ST-3 that capture biogeochemical transformations within the three compartments: stream water column, benthic layer, and hyporheic zone. The water column reaction network will emphasize influences of DOM content and chemistry, which will be routed to the stream via models developed in ST-1. To simplify DOM chemistry from many thousands of molecules to a tractable number of features we will use molecular properties (e.g., DOM molecule size and C/N ratio) with our previously developed thermodynamic model (Zheng et al. 2024; **C.2.2.3**). Benthic layer biogeochemistry will consider influences of both heterotrophic and autotrophic organisms, including algal biofilm respiration, growth, and delivery of DOM to fuel heterotrophic respiration. Hyporheic zone biogeochemical transformations will emphasize DOM chemistry and electron acceptors, including the depletion of oxygen as a factor limiting ERsed. These reaction networks will be further implemented within ATS-PFLOTRAN to incorporate dynamics and influences of variable inundation and wildfire, with an emphasis on changes in DOM chemistry and its release from POM as well as influences of subsurface vs. overland flow paths.

The ST-2 outcomes will be merged with ST-1 outcomes to enable the enhanced RCM to predict daily ERsed at each NHDPlus reach segment of each studied watershed. The model domain is defined in the region beneath the riverbed, from the top benthic cell to the hyporheic zone cell. ST-1 watershed models will provide the exchange flow and solute transport information, with reaction networks from ST-2 molecularly-resolved modeling. The integrated RCM will be operated in a simplified batch mode and a more complex column-based mode to delineate reaction profiles across the two cells. In batch mode, the reactions developed in ST-2 will be solved by using two ordinary differential equations for the benthic and hyporheic zone cells, separately. The vertical flow and transport process will be explicitly simulated in the column-based mode through one-dimensional reactive transport that discretizes the combined benthic and hyporheic zone column into multiple layers, using ATS-PFLOTRAN for flow and transport.

To reduce the computational cost of running a 3D watershed model, we will develop a deep learning-based surrogate model for ATS-PFLOTRAN to predict the flow and transport conditions at a given reach. This development will leverage the existing ATS-PFLOTRAN simulations in ST-1 as the training data. The surrogate model will take the atmospheric forcings and watershed hydrogeological features as the inputs and predict streamflow, HEFs, and solute concentrations at a daily timestep. To find the best surrogate, we will first develop the surrogates using various deep learning models at both point (using MLP and LSTM) and watershed scales (using transformer and graph neural networks). Then, we will use both transfer learning and meta learning strategies to perform the emulation at a new watershed.

<u>Hypothesis 3.1 evaluation in the YRB (**1st ModEx step; Fig. H3.1**)</u>: Initial quantification of ERsed allometry across contrasting basins indicated transferable relationships between scaling parameters and HEFs (Guerroro et al. In prep; **Section C.2.2.7**). However, this analysis did not incorporate temporal dynamics or the roles of variable inundation, wildfire, or molecular processes. We hypothesize that these perturbations and processes will govern the ERsed allometry dynamics (**Hypothesis 3.1**). For initial hypothesis evaluation we will use *in silico* counterfactual experiments (**Fig. C.3.11**). These experiments will be designed to isolate and quantify the individual and combined impacts of wildfire, variable inundation, and molecular processes on ERsed allometry dynamics. We will first do this work using watersheds within the YRB (**Fig. C.3.9)**; this represents the '1st step' of the ModEx process focused on evaluation of hypothesis 3.1, and the outcomes will guide follow-on data generation and modeling as the '2nd step' of ModEx in ST-3B, ST-1, and ST-2 (**Fig. H3.1**).

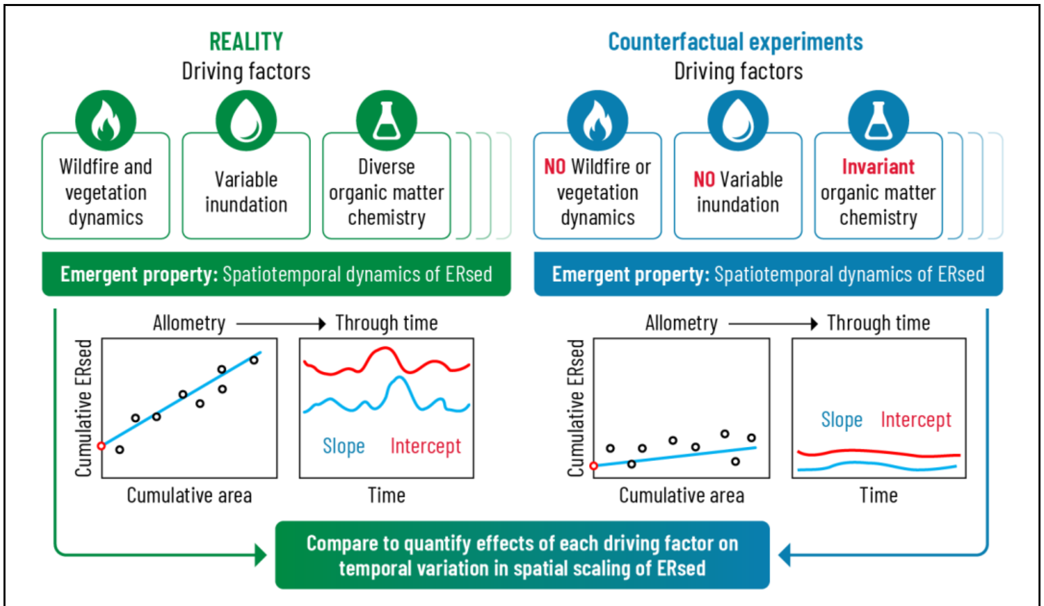


Figure C.3.11: Counterfactual simulation experiments will quantify the influence of specific factors (e.g., variable inundation) by comparing what actually happened (in terms of allometric scaling parameters) to what would have happened (to allometric scaling parameters) if a given factor did not occur (e.g., no variable inundation). We will use counterfactual simulations for each factor of interest (i.e., all the elements of the hypotheses 3.1 and 3.2)

Reality case: For the *in silico* experiments, we will use the ST-3A model (**Fig. C.3.10**) as a digital representation of reality that includes our best understanding of spatiotemporal variation in biological, physical, and chemical processes. This model will be used to predict ERsed at each reach at each point in time in the stream network of a given YRB watershed (**Fig. C.3.9**). We will aim for daily temporal resolution of these ERsed predictions and run the model for at least one year, following the spin up period. The resulting spatiotemporal predictions will be used to quantify allometric scaling of ERsed at each point in time (**Fig. C.3.9**) and the temporal variation in associated scaling parameters (**Fig. C.3.11 left panels**) over the simulated year.

Counterfactual case: Next, each of the three primary factors identified in **Q3.1** will be removed from the model, individually and in combination (**Fig. C.3.Integration**). This will reveal what would have happened (to ERsed allometry dynamics) in the absence of wildfire, absence of variable inundation, and without spatiotemporal variation in molecular processes. Other elements of the RCM such as HEFs and benthic primary producer dynamics are a core part of the RCM and will not be directly manipulated in the experiments; they will be allowed to vary due to streamflow dynamics in both the reality and counterfactual experiments.

ERsed allometry dynamics from the counterfactual experiments will be compared to the reality case to quantitatively test hypothesis 3.1 (**Fig. C.3.11**). The experimental 'treatment' that leads to the largest change in ERsed allometry dynamics would be considered the primary driver of those dynamics. This will be done within each watershed studied in the YRB. The identity of the primary driver may be consistent across watersheds or may vary. In addition, we may observe non-additive interactive influences. Hypothesis 3.1 proposes that such an interaction is most likely between wildfire and variable inundation. The signature of this interaction would be a quantitative shift in ERsed allometry dynamics (when wildfire and variable inundation are 'turned off') that deviates from the quantitative shift expected by summing the individual influences of wildfire and variable inundation (**Fig. C.3.Interaction**).

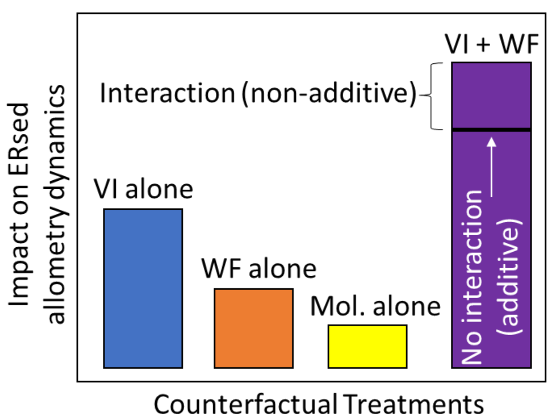


Figure (C.3.Interaction): Conceptual outcomes of counterfactual *in silico* experiments that would be consistent with ERsed allometry dynamics influenced by an interaction between variable inundation (VI) and wildfire (WF) and minimal influences of molecular (Mol.) processes. When VI and WF are both turned off in the model, the change in ERsed allometry dynamics is greater than the summed individual influences of VI and WF, indicating a non-additive interaction (*Alt 2* in Fig. H3.1.). If VI + WF effect was additive, no interaction would be indicated and VI would be inferred as the primary driver (*Alt 2* in Fig. H3.1.). Different outcomes trigger different follow-on data generation and modeling efforts (Fig. H3.1).

Activity ST-3B. ERsed Allometry Dynamics Across Basins

Approach: Activity ST-3B addresses our second objective (ERsed Allometry Dynamics) at the multi-basin scale (i.e., beyond the YRB), and evaluates hypothesis 3.2.

<u>Outcome-triggered field data generation and modeling (**2nd ModEx step**; **Fig. H3.1**)</u>: The outcomes of our *in silico* counterfactual model experiments in ST-3A will guide us to focus more on some processes and less on others and trigger a combination of data generation and modeling across basins. To illustrate how ST-3A modeling will guide our ST-3B data generation approach, we focus on two possible model outcomes: variable inundation is the primary driver of ERsed allometry dynamics (*Alt 1*) or wildfire and variable inundation as interactive drivers (*Alt 2*). In our two examples, Alt1 and Alt2 see a strong influence of variable inundation over ERsed allometry dynamics, indicating very fast rates of ERsed (i.e., hot moments) following re-inundation in YRB streams based on ST-3A in silico experiments. Hypothesis 3.2 requires evaluating the transferability of this result across basins, where we anticipate that climate-driven changes in vegetation

and hydrology will modulate the link between ERsed, variable inundation, and wildfire (e.g., by influencing how dry sediments become in the absence of surface water or influencing streamflow dynamics post-fire). To test both alternatives, ST-3B will deploy sensor networks to measure ERsed following re-inundation in non-perennial streams across basins that collectively span an arid to mesic climate continuum. *In situ* data generation will include multi-parameter sensors to estimate whole ecosystem respiration via dissolved oxygen, key respiration controls (turbidity, OM and chlorophyll), pressure sensors to capture variability in atmospheric and hydrologic conditions, and dark bottle incubations to measure water-column respiration. Collectively, these data-streams will estimate reach-scale ERsed in several nested streams across basins (see detailed methods for estimating reach-scale ERsed in **C.3.2)**. We have identified USGS gages that span broad climates on non-perennial streams and in burned watersheds (**Figs. C.3.xx, C.3.xx**), and our field data collection efforts leverage existing data on current and historic streamflow dynamics (i.e., when does re-inundation occur) by co-locating our deployments with active discharge and water quality sensors when possible. Using this approach, we will focus sensor deployments in watersheds across the arid-to-mesic continuum.

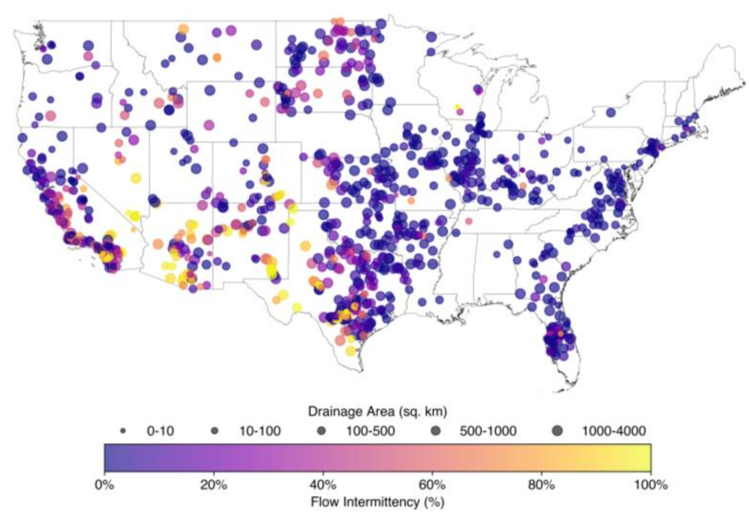


Figure C.3.7: Location of active USGS gages on non-perennial streams draining watersheds that are feasible for setting up the ATS model (in ST-3). Colors indicate the frequency of recorded no-flow conditions. ST-2 will collaborate with ST-3 to select a biophysically diverse collection of associated watersheds as sediment sources for ST-2 experiments, field sites for ST-2 data generation, and ST-3 ATS efforts.

In the *Alt 1* case (variable inundation is primary), we would instrument 40 non-perennial streams across 5 watersheds spanning arid to mesic climates, with each watershed in a different CONUS basin. Within each watershed, 8 sites will be nested along flow paths that simultaneously cover relevant biophysical gradients like elevation and land-cover (e.g., Guerroro et al., in prep), and span stream orders to capture cumulative respiration signatures for quantifying allometric scaling behavior. In the *Alt 2* case (wildfire + variable inundation interaction) we would still instrument 40 non-perennial streams spread across climates, but we would focus on 4 pairs of burned/unburned watersheds. The *Alt 2* case will also utilize preliminary ATS simulations to identify reaches most likely to receive OM transported out of burned hillslopes, and physically similar reaches downstream of unburned hillslopes/catchment. In both cases we would instrument 5 nested locations within each watershed guided by NHDPlus stream designations (perennial/non-perennial) and preliminary ATS simulations to identify reaches across each watershed that are likely to go dry and re-inundate during the deployment period. Deployments will last at least 1 month at each location and target times in which re-inundation is expected (both cases) and high hydrologic connectivity is expected from burned hillslope (*Alt 2* case only). To enable these cross-basin efforts we would work with the WHONDRS consortium and leverage existing DO sensor deployments via USGS, NEON, and the StreamPulse network as much as possible.

Resulting data from the fieldwork described above would reveal the degree to which our understanding of ERsed hot moment linkages to re-inundation and/or wildfire in the YRB transfers across basins/climates. If the YRB result is not transferable, we would learn what biophysical factors (e.g., climate) may suppress the ERsed responses to variable inundation and wildfire. Either way, the results would provide new conceptual understanding, and the data would be needed to run simulation experiments using the enhanced RCM across climatically divergent watersheds to evaluate hypothesis 3.2. The ERsed data alone would not be sufficient, however, for this hypothesis evaluation. In both alternative cases, we would need the capacity to simulate variable inundation across the climatically divergent watersheds where we are estimating ERsed. As such, both alternatives would trigger (1) preliminary ATS simulations in all watersheds where we would estimate ERsed to guide field deployments and (2) further improvement of ATS's ability to simulate variable inundation via iteration with remote sensing and ground measurements in each watershed, as will be done initially in the YRB (see ST-1 for details). The *Alt 2* case would indicate that wildfire-induced changes to water flow paths significantly modifies how ERsed responds to re-inundation. In turn, in addition to the ATS work, *Alt 2* would trigger further ModEx-based model development in ST-1 focused on ensuring that the RCM model can accurately capture wildfire influences on flow paths so that the interaction between wildfire and variable inundation is properly simulated across climatically divergent watersheds/basins.

Both outputs are needed to test hypotheses and drive biogeochemical reaction networks in the RCM model and enable hypothesis evaluation in ST-3B. We recognized that doing research across the YRB to CONUS will pose significant challenges, but through the approaches described above these risks can be mitigated. For example, our team has successfully worked across the YRB as well as across CONUS through participatory science approaches during the last quadrennial period. Such approaches lend themselves to highly distributed ground-truthing of the remote sensing data (e.g., via participatory deployments of wildlife cameras at select locations).

<u>Hypothesis 3.2 evaluation across basins (**Further ModEx steps; Fig. H3.1**)</u>: For cross-basin hypothesis evaluation the workflow will be similar for both alternative cases. We will integrate the new data and model developments described above into the enhanced RCM so it can be run in the watersheds used for new ERsed data generation. In turn, the same counterfactual approach used in the YRB will be used across climatically divergent watersheds. Outcomes will be used to evaluate how climate modulates the link between ERsed allometry dynamics and variable inundation and/or wildfire. Quantitative results like those conceptualized in **Fig. C.3.Interaction** will be compared across watersheds. Cross-watershed variation in the results from the counterfactual *in silico* experiments will provide a direct evaluation of hypothesis 3.2. A consistent profile of influences across watersheds would indicate that the YRB outcomes are transferable, whereby changes in biophysical setting (e.g., climate) does not the degree to which different factors influence ERsed allometry dynamics. This would reject hypothesis 3.2, but would nonetheless indicate a highly transferable outcome. If we find that the profile of influences (as summarized in **Fig. C.3.Interaction**) changes systematically across the arid-to-mesic continuum, this would be consistent with hypothesis 3.2 and indicate transferability in terms of a climate-based modifier of ERsed allometry dynamics. Lastly, if the profile of influences varies across watersheds, but not in any systematic way across biophysical conditions, this would indicate there is some underlying process that is not yet understood, measured, or identified. Such an outcome would motivate follow on efforts to reveal the mechanistic driver.

ST-3 Outcomes:

ST-3 is the overarching integrator of knowledge, data, and models across the RC SFA. The result will be quantitative, mechanistic, and predictive understanding of ERsed allometry dynamics as influenced by watershed system processes and perturbations that are shifting because of global change. New understanding from ST-3 will be transferable, applying equally within and across biophysically divergent basins. ST-3 outcomes will also enable the diagnosis of why and when some catchments within a given watershed have outsized influences on cumulative ERsed, and why/when others do not. Such inferences are key to guiding future ESS research efforts towards places, times, and processes with high leverage over watershed and Earth system function. Further, ST-3 will reveal the scales across which molecular processes need to be considered to achieve predictive understanding, and at which scales they can be simplified, which is a persistent challenge in ESS science. The enhanced RCM represents another major ST-3 outcome that will be useful for deepening our understanding of and capacity to predict the future of watershed systems. Like all our models, the RCM will be openly available for integration of additional processes and use for additional locations and purposes. More broadly, ST-3's mechanistic integration of wildfire, variable inundation, biophysical features, and molecular processes across environmentally divergent basins is unprecedented. These system components have never been integrated even in a single field site, let alone across basins. ST-3 will bring them together in what amounts to a sea change in our collectively capacity to achieve transferable, integrative environmental system science.

C.3.4 Potential Risks and Alternative Approaches

The RC SFA identifies potential scientific/technical risks, reviews cross-SFA interdependencies, considers options, and develops graded approaches to mitigate those risks. The PI Team monitors project progress and directs project activities consistent with the project management approach (**C.4.2**).

Our review of interdependencies during preparation of this plan revealed low to moderate levels of risk associated with proposed research scope. STs provide key information to each other, reflecting strategic project interdependencies (**Fig. C.3.1**). These exchanges are generally low risk, however, and focus on the handoff of model-generated outcomes to guide further data generation and modeling work. Although our analysis and modeling approaches are state-of-the-art, we have demonstrated the feasibility of many of them in the past quadrennial period and are confident in their successful application and our team's increasing refinement of our ModEx approaches. These interdependencies will be monitored regularly by the PI Team and Cross-Cut Leads (**Fig. C.4.1**), allowing us to make any necessary adjustments in a responsive and timely manner.

C.3.5 Project Timelines and Milestones

The ST descriptions above are defined at the level of activities (major components of each theme). The research timeline (**Fig. C.NN**) shows the planned timing of each of the activities.

Figure C.NN. Structure and timing of the three STs and their activities.

To provide a mechanism for evaluating progress towards project objectives we have established annual milestones for each ST (**Table C.5**). These milestones map directly to outcomes of the activities defined above. Progress towards each milestone will be evaluated quarterly and reported on an annual basis.

Table C.5 Project Milestones

FY	Milestones
ST-1 XXX	

2025	Establish drivers and mechanisms of hydrologic connectivity and resulting stream flow dynamics and DOM export
2026	Perform distributed DOM and water isotopes sampling in surface and subsurface domains in space and time across the YRB, guided by model simulations. Provide model predicted flow paths across YRB watersheds to ST-2B to enable hypothesis evaluation and to ST-3A to facilitate development of flow components of the enhanced RCM.
2027	Quantify the influences of wildfires and ecosystem recovery on DOM transport. Provide model predicted DOM transport across YRB watersheds to ST-2B to enable hypothesis evaluation, and to ST-3A to facilitate development of DOM transport of the enhanced RCM.
2028	Continue to perform scenario simulations manipulating drivers of vegetation dynamics, water connectivity and fire impacts across CONUS watersheds studied in ST-3; provide ATS predictions of variable inundation in these CONUS watersheds to facilitate use of the enhanced RCM in ST-3.
ST-2 XXX	
2025	Conduct batch scale experiments with variable inundation and PyOM using YRB sediments. Identify key gene expressions and OM chemistries. Install sensors in Satus Creek and Naches River watersheds, with ST-1. Acquire SAR images for remote identification of variable inundation and provide findings to ST-1. Upon flow resumption, estimate reach-scale ERsed across YRB to test Q/H 2.B.
2026	Run in-silico AquaMEND simulations, incorporating knowledge gained in FY25 from field and laboratory efforts, to guide planning of additional sediment sample collection and fieldwork. Incorporate remote sensing of variable inundation into ATS-PFLOTRAN simulations work with ST-1 to generate new predictions of ERsed across greater ranges in velocity, flow paths, time since last inundation, and sediment texture, to guide planning of additional sediment sample collection and fieldwork. Begin planning CONUS WHONDRS efforts for sediment collection (ST-2A). Plan and execute new synoptic sites with sensor deployment locations in YRB (ST-2B), and execute synoptic field sampling of ERsed upon resumption of flow to test Q/H 2.B.
2027	Conduct batch scale experiments across CONUS sediments selected for range in flow, velocity, time since last inundation, and sediment texture from fire and non-fire impacted sites to further test Q/H 2.A. Iterate among batch experiments and in-silico AquaMEND simulations to inform ST-3.
2028	Continue ModEx loops iterating among batch experiments and in-silico AquaMEND simulations across CONUS sediments to inform important molecular reactions fueling ERsed. Exit ModEx loop when predictive skill of the model stabilizes.
ST-3 XXX	
2025	Initiate development of the enhanced RCM with a focus on setting up infrastructure for integrating ST-1 and ST-2 outcomes that will be delivered starting in FY26.
2026	Use FY25 outcomes from ST-1 and ST-2 to formalize the enhanced RCM thereby addressing ST-3's 'Process Integration' objective. Update all data products and cyberinfrastructure needed to rapidly select field deployment locations following hypothesis evaluation in FY27.
2027	Use enhanced RCM to test hypothesis 3.A via counterfactual in silico experiments. Use hypothesis evaluation outcomes to trigger field deployments in watersheds across CONUS basins and to trigger further process refinements in enhanced RCM.
2028	Use outcomes of field deployments to incorporate the modulating influences of climate within enhanced RCM and use the updated model to test hypothesis 3.B, thereby addressing ST-3's 'ERsed Allometry Dynamics' objective. and set the stage for the subsequent quadrennial period.
Promoting Inclusion and Equitable Research (PIER) Plan	
2025-2028	PIER Plan milestones are described in Section C.6 and span the period of performance. Creation of SFA governance documents (i.e., code of conduct) will occur within FY25.

C.4 Management and Team Integration

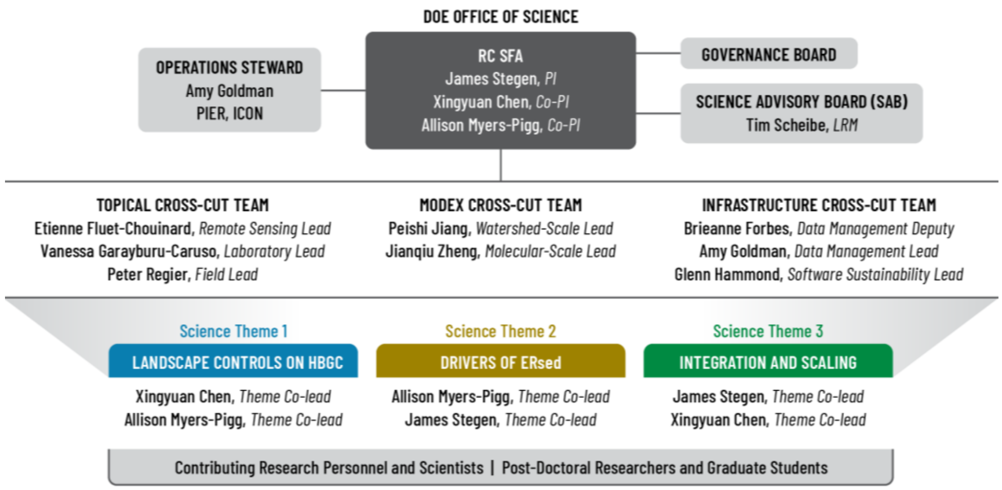


Figure C.4.1. RC SFA's organizational structure assures that we execute our mission using a formal system of effective communication, planning, resource allocation, evaluation, and realignment that is institution agnostic.

The RC SFA will continue to advance its scientific mission through proactive management, informed by frequent team feedback, that fosters a cohesive, collaborative, and inclusive culture of engagement. Our management approach emphasizes team science with regular, open, and effective communication. It is supported by our PIER Plan (**C.6**) that emphasizes a psychologically and physically safe, equitable, and inclusive environment co-created with the team and supported by infrastructure, training, mentoring, and professional development. Our research activities are designed with the broader research community and interested parties to maximize the impact of the knowledge we create and mitigate risk. We also mitigate risk by identifying contingency plans for high-risk tasks, regularly seeking input from the science advisory board (SAB) and broader community, and implementing an adaptive management approach to allow for adjustments in response to emerging risks. We will continue developing early career scientists and creating opportunities for enduring leadership and diverse partnerships (**C.6**).

Key and contributing staff (**Fig C.4.1, Section C.7**) are drawn from across PNNL's research directorates and are selected for their expertise to fulfill SFA objectives in a supportive, inclusive team science culture. Together with our collaborators and DOE user facilities, SFA team members provide the diverse expertise and capabilities needed to address the multidisciplinary scientific vision of the SFA. The leadership team mentors and provides opportunities for SFA team members (including early career staff) to lead research efforts and pursue professional development (**C.6**). Cross-cut leads and, as needed, sub-task leads report to the ST co-leads and update progress on a weekly to bi-weekly basis.

C.4.1 Organizational and Leadership Structure

The SFA organizational structure is shown in Figure C.4.1, including principal investigator (PI) Stegen, co-PI Chen, and co-PI Myers-Pigg. Our science leadership PI team coordinates interdisciplinary research through a tightly woven, synergistic structure, with information flowing across science themes (STs). The leadership team is supported by an SAB (Scheibe chair), a PIER and ICON focused operations steward (Goldman), and cross-cut leads that work across the STs (**Fig. C.4.1**). The RC SFA is organized into three STs, each co-led by members of the PI Team. The PI Team co-designed the STs together with input from the broader RC SFA team to ensure integration and alignment with the SFA vision and objectives. Each set of ST co-leads is responsible for managing a specific scope with associated schedule, budget, staff, and deliverables (i.e., publications, data, analyses, or modeling products needed by other team members). This leadership structure differs from that used in FY21–FY24 with the goal of increasing integration and communication across the project. The STs are further linked through the use of three cross-cutting teams: topical (i.e., remote sensing, laboratory, field), ModEx (i.e., watershed-scale, molecular scale), and infrastructure (i.e., data management, software sustainability). The roles and responsibilities for the ST co-leads and the cross-cut leads will be defined in the project's governance documents (**C.6**) and informed by team feedback. Many of the cross-cut leads are early career researchers, and these roles will help amplify their voices to guide the RC SFA direction.

The Science Advisory Board is composed of diverse, nationally recognized experts in multi-basin and river corridor science, wildfire, and stream intermittency, including scientists with experience leading complex field and modeling programs and with expertise in the Yakima River Basin (**C.6**). They provide preliminary review and feedback at significant stages of the RC SFA program, and we convene with them formally once per year and informally approximately twice per year.

Institutional Oversight. PNNL is the lead institution for the RC SFA. As the PI, Stegen works with the co-PIs to track progress toward overall project objectives, ensures that scientific and operational risks are aligned with PNNL/DOE policies, and oversees subcontracts with performing institutions. Each partner institution has an institutional lead who will be responsible for managing subcontract responsibilities, including performance and staffing issues, and will engage with the PI to address any performance or staff issues. The institutional leads are also responsible for managing adherence to the PIER code of conduct and engaging with the operations steward as needed (**C.6**). The institutional leads are Bladon (OSU), Hall (UMT), and Liu (WSU). They will work with the PI and a PNNL project coordinator (Sandra Carrasco) to manage subcontracts and resolve any organizational compliance issues.

C.4.2 Management Approach and Team Engagement

Regular coordination among members of the PI Team enables effective tracking of progress, ensures that inter-activity and inter-ST products are delivered when needed, avoids diversion of research effort to poorly aligned activities, and facilitates problem solving. The RC SFA uses a range of tools to provide information to team members in all project roles. Project controls are implemented using Microsoft Teams and Microsoft Planner software provided by PNNL, which facilitates team-oriented planning, open communication, accountability, and tracking. We use Microsoft Teams for working group discussions, to plan field efforts, and for announcements. Microsoft SharePoint, linked to Microsoft Teams, and Confluence are repositories for meeting agendas, minutes, project photos, presentations, and posters, as well as a document tracking publication progress, allowing all team members to see all relevant activities and engage across working groups. We facilitate frequent updates and discussion through virtual weekly whole team meetings to encourage engagement across our highly spatially distributed team (**Table C.1**, **C.6**).

Table C.1. Team meetings. Team integration is promoted through a purposeful structure and frequency of communications.

Meeting and Frequency	Mode	Participants	Purpose
All-Hands Meeting, annual	P	All team members	Review annual progress and refine plans; team building.
PIER and Team Science Meeting, every 5 weeks	V	All team members	Learn and discuss PIER, DEIA, and team science topics (**C.6**)
Literature Meeting, every 5 weeks	V	All team members	Review recent publications relevant to SFA.
ST (ST-1, 2, 3) Meetings, rotate every 5 weeks	V	All team members	Technical integration and progress reporting.
Activity Meetings, as determined by leads	V/H	Activity leads and key contributors	ST activity coordination and progress reporting.
ST and Cross-Cut Leads Team Meeting, monthly	V/H	PI Team and cross-cut leads	Review progress and address budget/staffing/research priority issues
SFA Leadership Team Standup Meeting, biweekly	V	PI Team; ops steward; LRM	Monitor progress, integration, and project management (e.g., budget/staffing/ performance).
1:1 Leadership/team mentoring meetings, at least 2/yr per team member	V/H/P	PI Team and all team members (1:1)	Team member mentoring, communication, and connection (**C.6**)
1:1 Mentor/mentee meetings, as determined by pair	V/H/P	Mentor and mentee pairs	Team member mentoring, communication, and connection (**C.6**)
Science Advisory Board Meeting, at least 1/yr formally	V	PI Team, SAB chair, members	Review SFA progress, discuss directions, and provide feedback

DEIA = diversity, equity, inclusion, accessibility. H = hybrid. Ops = operations. P = in person. PI = principal investigator. PIER = promoting inclusive and equitable research. SAB = science advisory board. ST = science theme. V = Virtual.

C.4.3 Staffing and Performance Indicators

Resource and personnel allocations within the SFA are annually determined by the PI Team and operations steward in consultation with cross-cut and activity leads and revised as needed throughout the year. Changes to staff funding and/or responsibilities can be made by the PI in response to performance, revisions to milestones, and/or evolving research findings. SFA performance expectations are described in the SFA governance documents described in the PIER Plan (**C.6**), incorporated into annual staff performance goals, and evaluated against negotiated program metrics (such as key milestones, publications).

C.4.4 Scientific Directions

The evolution of the SFA is influenced by multiple factors, including (1) changes to BER/ESS research directions and goals, (2) new capabilities via laboratory investments or new user facility capabilities, (3) transformative SFA research findings, and (4) new staff. The LRM facilitates project evolution based on changes in BER goals via guidance to the PI Team. Concepts for new science directions are solicited from team members at the weekly virtual and annual in person all-hands meetings. Selected non-SFA staff are invited to the annual meeting to present innovative ideas for new exploratory activities. The PI makes final decisions on SFA scientific evolution and team members.

C.4.5 Team Science and Synergistic Activities

Team science is an essential element of the RC SFA, is rooted in inclusive and equitable work (**C.6**), and is foundational to everything we do. We facilitate a team science approach across the project with multidisciplinary ST teams that integrate modeling, experiments, and observations to design, perform, and publish research. Alignment of goals, milestones, and deliverables with key outcomes are regularly assessed and discussed in weekly all staff meetings so individuals understand how their contributions map into the SFA's research program. Teaming requires a project-wide commitment to open communication, including feedback and accountability (**C.6**), to create a safe and trusting environment that facilitates team growth and multidisciplinary design, integration, interpretation, and publication. Teaming and collaboration also extend beyond the immediate project team through various synergistic activities and long-term external relationship building (**C.6**). These include (1) coordination of WHONDRS, (2) development and execution of complementary university-led or SBIR research projects (**H.1**), (3) communication with regional agencies and interested parties in the Yakima River Basin, and (4) presentation of research at forums, conferences, and universities. Our long-term vision provides the motivation for teaming and working in unison, both in our PNNL team and with external participants, to accomplish a very challenging endeavor in a supportive, inclusive, and safe environment.

C.4.6 Links to Other BER-funded Activities

We actively engage with ESS program activities by partnering with the IDEAS-Watersheds and ExaSheds projects, participating in ESS Cyberinfrastructure Working Groups, and collaborating with other ESS investments via WHONDRS. We will continue active collaborations across BER-supported facilities and resources, including EMSL, JGI, KBase, NMDC, and ESS-DIVE. The RC SFA leadership has worked closely with the leads of other ESS watershed science SFAs to identify touchpoints and opportunities for coordination and mutual benefit, including the 'Watershed Function' SFA led by LBNL and SLAC and the 'Watershed Evolution and Dynamics' SFA led by ORNL. Both projects provide opportunities for synergistic handoffs of concepts, data, and approaches.

Our research is also well-aligned with the objectives of the larger EESSD program, which emphasizes the interplay among hydrology, biogeochemistry, and the function of human and Earth systems. The Earth and Environmental Systems Modeling program is developing innovative modeling capabilities for the integrated Earth system, such as the Energy Exascale Earth System Model (E3SM). Our research uses several codes related to E3SM, and we will fill a key gap in the land component of that model, namely the representation of watershed and stream network hydro-biogeochemical processes and how they are impacted by several globally relevant perturbations. Our SFA science advisory board (**C.7**) includes E3SM scientists to enable connections and synergy.

C.4.7 Community Connections

We will continue to actively empower others to use our research outcomes (e.g., partnering with institutions to train students and researchers to use RC SFA open-source models and data, facilitating open collaborative manuscripts co-written with the research community). This includes intentionally deepening our existing community connections rooted in mutual benefit in alignment with ICON principles, such as actively finding conceptual points of intersection with the Yakama Nation where they gain value from our data and models. We anticipate that the capabilities we develop, such as a dynamic remote sensing analysis pipeline and versions of ATS-PFLOTRAN with more complete river corridor processes, will likely be of significant interest to the watershed science community. We will continue collaborations to support synthesis and model-ready data needed to build transferable understanding (e.g., continuing to contribute microbial data to the Genome Resolved Open Watersheds database to enable community-led syntheses). A variety of additional community connections will be enabled via our PIER plan (**C.6**) and data management efforts (Appendix 5).

C.5 Data Management and Software Productivity and Sustainability Plans

Our approach to data management will build off our existing commitment to publishing open FAIR data on ESS-DIVE as early as possible in the research lifecycle, often prior to manuscript writing, to facilitate broad community use (**Appendix 5–Data Management Plan**). The PI has ultimate oversight and decision-making over project data management. The Data Management Cross-Cut Lead (Goldman) and Deputy (Forbes) (**C.4**) will be the primary points of contact and oversight for all data generation and data publishing across the project. Our extensive use of the ESS-DIVE Community Reporting Formats supports interoperability and discovery of our data using ESS-DIVE's fusion database. Our commitment to software productivity and sustainability is critical for our multidisciplinary team to accelerate scientific discovery while ensuring the longevity of proposed model developments (**Appendix 6– Software Productivity and Sustainability Plan**). The Software Sustainability Cross-Cut Lead (Hammond) will be the primary point of contact for software longevity.

C.6 Promoting Inclusive and Equitable Research (PIER) Plan

Our PIER Plan is integral to our scientific and technical advancements and supported by a PIER Steward. This section describes the SFA's plans to (1) create and sustain a culture of belonging and an inclusive, safe, and professional research environment through infrastructure, training, mentoring, and professional development; and (2) engage and recruit early career, diverse, and historically underserved and minoritized researchers through project staffing and long-term relationship building. All members of our project contribute to the culture of our team. The PI will be supported by a PIER Steward (Goldman) that will operate both at a project level and through the individual STs to oversee implementation of PIER activities and the milestones/metrics below. We will expand upon the project's current strengths and advance new project infrastructure to fill identified gaps. We will leverage PNNL's and our collaborators' institutional resources in training, mentoring, professional development, and diversity, equity, inclusion, and accessibility (DEIA) support. Additional detail about resources and commitments from PNNL and our partner institutions are included as **Appendix 8**.

C.6.1 Creating and Sustaining an Inclusive, Safe, and Professional Environment

This section describes creating and sustaining a culture of belonging through (1) infrastructure, (2) training, (3) mentoring, and (4) professional development. These interconnected roots are the foundation for recruiting (see Recruitment Strategy section) and are critical to our team generating impactful science.

Infrastructure for support and belonging. The PIER Steward will guide the project team's co-creation of a project-level governance documents, including a code of conduct, that will define expectations for behavior of project team members, including project leadership and funded external collaborators. The code of conduct will use existing external resources as a base, and it will describe structures for constructive feedback, reporting, conflict resolution, and accountability. The governance documents will also include expectations around accessibility of meetings and products created (i.e., list of accessibility recommendations) and guidance for creating authorship agreements. Annually, team members will have the opportunity to suggest revisions to the governance documents, and after any revisions are disseminated, will sign to indicate they have read and agree to follow the code of conduct. As a component of operationalizing the governance documents and project norms, we will continue holding weekly virtual project-wide meetings. Each weekly meeting will begin with three points of discussion in support of our team culture: (1) governance and code of conduct reminders, (2) celebrating wins across the project (i.e., publications, presentations, data packages, project milestones, professional achievements), and (3) new hire introductions. Meetings may be hybrid if appropriate, with attention to providing an environment in which everyone feels they can engage. Additional infrastructure for sustaining a culture of belonging will include increasing use of PNNL's rewards and recognition mechanisms (i.e., Outstanding Performance Awards); holding an annual all hands in-person meeting to support connection across our highly spatially distributed team; and continuing to distribute an annual psychological safety survey to all team members. **<u>Milestones</u>**: (1) Governance documents, (2) Signed code of conduct, (3) Introductory meeting expectations slide, (4) Annual team psychological safety survey and discussion, (5) Annual team knowledge check on accessing governance documents. **<u>Metrics</u>**: (1) List of celebrated "wins," (2) Annual assessment of psychological safety survey results.

Team training. The project leadership and mentors will train the project team on the governance infrastructure described above, including emphasizing enthusiastic support for revising the content in response to team feedback. The PIER Steward, with engagement from other project leaders and also early career researchers (ECRs), will create a set of RC SFA onboarding materials that will act as an orientation to the project science and governance documents. Mentor meetings (described below) will also include check-ins on this material. As part of the rotating weekly all-project meetings (described above), we will add a new PIER and Team Science meeting to the rotation that will occur approximately every 6 weeks. This recurring meeting will serve as an opportunity to learn and discuss topics as a team. Topics may include specific trainings offered by PNNL or external groups (e.g., anti-bias training; active bystander training; active listening training; inclusive hiring training; social styles); broader DEIA resources (e.g., Employee Resource Groups; LinkedIn Learning); and team science-specific readings and discussions. Training associated with physical and psychological field and laboratory safety will be informed by PNNL's required off-site risk management plans, Lab Assist, and the SFA's continued biweekly field team meetings with the PNNL Project Management Office Director. These biweekly meetings are used to raise project-specific field concerns up to PNNL management and across the team and hear new guidance from PNNL management. Field safety will also be supported by a supplemental fieldwork guidance document that points to the code of conduct, including reporting and accountability information, and includes specific elements for minoritized team member safety and travel. SFA field safety supporting guidance in the governance documents will include information for always having at least two people in the field and at least three people if there is an overnight involved. During overnight fieldwork, each team member will continue to have their own hotel rooms. These norms around field staffing numbers will continue to be discussed regularly individually and in groups to confirm field staff's comfort and sense of safety.

Milestones: (1) Onboarding materials provided to Project Coordinator and team for dissemination to new hires, (2) Off-Site Risk Management Plan, (3) Fieldwork guidance document. **Metrics**: List of DEIA and Team Science Meeting topics and dates.

Mentoring project staff. We will support specific mentoring approaches to reach both new hires and existing team members, with five distinct strategies for different sub-groups (i.e., new intern and research associate (RA) hires, new virtual hires, all new hires, modelers/experimentalists) and all team members. The project team will continue to identify formal mentors for all interns and RAs, in alignment with PNNL's systems, as a component of the hiring process. The mentor and new hire will work together to personalize a mentoring agreement provided to them alongside project onboarding information. The mentoring agreement will be a living document that includes information about expectations, goals, and minimum meeting frequency, and is updated as needed. If a new hire is fully virtual, the project team will also help them identify an additional virtual project mentor to aid in feeling connected to the team and navigating the virtual workplace. Virtual project mentors may be short-term or long-term, depending on the new hire's interest. For all new hires, their mentor (if applicable) and project leadership will help the new hire identify one or more peer mentors. In support of deeper ModEx understanding across the team, a cohort of ECRs will co-lead formation of mentoring peer groups between modelers and experimentalists (i.e., ModEx peer groups). These small (2-6 people) peer groups will aim to bridge the gap in high level technical understanding of what team members do, build connections among team members that may not otherwise interact frequently, and reduce barriers to asking questions across the team, and focus on specific science outcomes. All team members will also have recurring one-on-one meetings between project staff and SFA leadership. Depending on staff needs, these may develop into meetings with multiple SFA leaders and the staff member in a single meeting. The one-on-one leadership mentoring meetings will discuss professional development, feedback, and DEIA efforts. All team members will continue to be encouraged to create additional mentoring relationships within and outside of the project to facilitate growth. **Milestone**: Mentoring agreement template. **Metric**: Mentor/mentee matches matrix.

Facilitating professional development. We will continue to encourage all staff, including ECRs, to pursue all aspects of the research lifecycle that align with their goals and interests, including leading tasks, contributing to manuscripts, and presenting at and organizing sessions in conferences. We will continue to facilitate skill building by connecting team members with external collaborators, which also supports interoperability of methods across teams, in alignment with the Coordinated principle in ICON. The project structure has been created intentionally to include multiple types of leadership roles (e.g., ST leads, cross-cut leads and deputies, ModEx peer group organizers) to create more opportunities for development. We will continue to plan for a minimum of 8 ECRs to present at conferences each year. Project leadership and mentors will continue to encourage all staff to participate in trainings/workshops to develop, grow, and teach technical skills. Project leadership and the broader team will continue to widely share opportunities for professional development (e.g., workshops/trainings, Funding Opportunity Announcements (FOAs), proposal reviews, groups to engage) across communication mechanisms (i.e., a project-wide mailing list, meeting announcements, project-wide chat announcements, small group discussions, and direct emails). **Metric**: Tracking name of staff, job category, and conference, training, or skill building workshop attended.

C.6.2 Project Recruitment Strategy

This section describes two key components of project recruitment: (1) immediate staffing needs and (2) relationship building for long-term growth and engagement.

Project staffing. Our team will continue to include ECRs through multiple internship and RA mechanisms. Over the next quadrennial period, we plan to recruit and hire 3-8 postdoctoral, post-masters, and/or post-bachelors RAs and host 3-5 Science Undergraduate Laboratory Internship (SULI), Community College Internship (CCI), Minority-Serving Institution Partnership Program (MSIPP), and/or Office of Science Graduate Student Research (SCGSR) students. This will include the continuing support of one in progress SCGSR. To aid in future SCGSR connections, project leadership and other team members maintain profiles on the "View Potential Collaborating Scientists" DOE Workforce Development SCGSR webpage. The project's Science Advisory Board will include at least one faculty from a historically Black college or university (HBCU) or minority-serving institution (MSI) and one member associated with the Confederated Tribes and Bands of the Yakama Nation. The project will include sub-contracts with collaborators at Oregon State University (OSU) and University of Montana. The project will continue to support at least one Distinguished Graduate Research Program (DGRP) fellowships student from OSU. Project leadership will work together to coordinate hiring with an annual roadmap to facilitate alignment with PIER Plan goals, including job advertisement and recruiting approaches. Job advertisements will be distributed intentionally (i.e., job boards, mailing lists, announcements, and direct emails) to reach a diverse candidate pool that includes historically underserved and minoritized researchers. Job advertisements will describe the project's expectations of all staff supporting DEIA. Candidate interviews will include questions that highlight commitment to DEIA. Candidate references will be asked to reflect on the candidate's experiences in DEIA, safety, and creating a culture of belonging. **Milestones**: (1) Job advertisement and interviewing procedure. **Metrics**: (1) Number of SULI, CCI, MSIPP, and SCGSR students engaged with the project; (2) Number and prior institution of new hires.

Long-term relationship building. Long-term relationship building will include (1) local Yakima River Basin partners, (2) institutions and programs that support minoritized researchers, and (3) the broad WHONDRS participatory science collaborative community. Relationship building within the Yakima River Basin will continue to include collaborations with the Confederated Tribes and Bands of the Yakama Nation as a component of identifying mutual benefit between our immediate research team and those beyond it, in alignment with the Networked principle in ICON. Other engagement opportunities for mutual benefit with interested parties (i.e., stakeholder engagement) will be regularly assessed as research develops. Local community outreach (e.g., Salmon Summit) will support relationship building to younger age groups and will rely on collaboration with PNNL's STEM Ambassador program, for which six team members are trained. Project leadership and team members will identify select conferences for intentional attendance to support relationship building with minoritized researchers and to support professional development of our team members. These will include the Society for Advancing Chicanos/Hispanics and Native Americans in Science (SACNAS) National Diversity in STEM Conference and may include others (e.g., Out in STEM Conference; Society of Asian Scientists and Engineers Conference; National Association of Black Geoscientists Technical Conference). As appropriate, project team members will attend PNNL's monthly HBCU/MSI network seminar series and when possible, will present about project science. As FOAs are announced, project members will aim to partner with HBCUs/MSIs in RENEW, FAIR, and university proposals that align with project research. Broad relationship building as a component of WHONDRS participatory science will continue to include open discussions for study design feedback, collaborative open manuscript writing, and virtual classroom engagements and field activities that began as an outcome of identifying mutual benefit with WHONDRS collaborators. Classroom engagements will expand beyond WHONDRS-specific elements of the project and will expand to MSIs and other universities local to the Yakima River Basin and PNNL (e.g., Columbia Basin College (Hispanic Serving Institution; HSI), Heritage University (HSI/Native American-Serving Nontribal Institution; NASNTI), Northwest Indian College (TCU), Yakima Valley College (HSI)), partner institutions, and field sites. WHONDRS engagement opportunities will continue to be advertised widely, with an emphasis on reaching underserved and minoritized researchers and applicable disciplines. **Metric**: Type and outcome of mutually beneficial activities identified with external interested parties.

C.7 Personnel

Each team member has been chosen for their excellence in the research areas that are needed to successfully fulfill project objectives. Our key personnel include senior-, mid-, and early-career scientists spanning a range of scientific capabilities in molecular-to-watershed scale modeling, ecohydrology, biogeochemistry, and remote sensing. We have a balance of empirical and computational scientists, as well as expertise in both field- and laboratory-focused research. Leadership team members have management experience and leadership skills or interest in developing these skills. Additional staff, including students and post-graduate researchers, complete the research team. Table C.2 lists the senior leadership team (PI, LS, co-PI, and operations steward) and key personnel. Curricula vitae for all key personnel are provided in Appendix 2. The full list of external collaborators and their institutions is provided in Appendix 3.

Table CC.2. Leadership and Key Personnel. Science Theme (ST) designations are 1 (Landscape controls on HBGC), 2 (Drivers of ERsed), and 3 (Integration and Scaling).

Name	ST	Primary Contributions	FTE/year	Inst.	Expertise
Senior Leadership					
James Stegen	2,3	PI. Oversees the project, budget, and performance; Science Theme 2 and 3 Co-Lead.	0.7	PNNL	*Has been task lead, campaign lead, or co-PI of RC SFA for the past 11 years. Co-founded and co-leads WHONDRS and the ICON Science Cooperative. Early Career Award project funded in 2019 focused on variably inundated sediment systems. Extensive experience in field, lab, computation, and analysis. Formal training in ecology and allometry (PhD on metabolic theory scaling). Member of the advisory boards for ESS-DIVE and MONet. Member of EBSD Diversity and Inclusion Council for 3 years.
Xingyuan Chen	1,3	Co-PI. Science Theme 1 and 3 Co-Lead.	0.6	PNNL	*Has been task lead, campaign lead, or co-PI of RC SFA for the past 14 years. Holds leadership roles in IDEAS-Watersheds, ExaSheds, AI4ESP, and COMPASS-FME in model development and coupling. Extensive experience in integrated watershed modeling, including overseeing river corridor model development and ATS-PFLOTRAN modeling. Formal training in ecohydrology and statistics.
Allison Myers-Pigg	1,2	Co-PI. Science Theme 1 and 2 Co-Lead.	0.6	PNNL	*Campaign lead of RC SFA for past 4 years. Holds leadership roles in COMPASS-FME EXCHANGE and community wildfire workshops. Extensive experience in wildfire science in aquatic systems, field, and lab. Formal training in aquatic BGC and analytical techniques, including wildfire biomarker analysis and methods development.
Amy Goldman	1,2,3	PIER and ICON Operations Steward. Data Management Cross-Cut Lead. WHONDRS co-lead.	0.7	PNNL	*Has been WHONDRS data management lead since 2018 and RC SFA data management co-lead for 4 years. Co-founded and co-leads WHONDRS and the ICON Science Cooperative. Co-founded and co-led PNNL's Prism Employee Resource Group for 5 years and was a member of PNNL's Diversity, Equity, and Inclusion Council for 3 years.
Tim Scheibe		Science Advisory Board Chair, Laboratory Research Manager		PNNL	*Outgoing PI of current SFA. PDO director. Decades of experience in hydrology, BGC, integration of microbes, and recognized community leadership.
Key Personnel					
Kevin Bladon	1,2	Wildfire impacts on stream nutrients and hydrologic connectivity expertise	(a)	OSU	Recognized expert in wildfire science… engaged in sfa already (Working with student)

Name	ST	Primary Contributions	FTE/year	Inst.	Expertise
Etienne Fluet-Chouinard	1,2,3	Remote Sensing Cross-Cut Lead		PNNL	Extensive experience with remote sensing of wetlands and surface water as well as geospatial data and tools. Developed several tools to map aquatic ecosystems. Currently contributing to remote sensing detection of surface hydrology in COMPASS-FME and for flood disaster in RADR-Flood.
Brieanne Forbes	1,2,3	Data Management Cross-Cut Deputy		PNNL	*Has supported RC SFA data management for 2 years, including development of automated workflows and extensive engagement with ESS-DIVE.
Vanessa Garayburu-Caruso	1,2,3	Laboratory Cross-Cut Lead. Will continue leading weekly lab coordination meetings.		PNNL	*Has been WHONDRS Chemistry lead since 2018 and managing RC SFA lab/field coordination for 5 years. Extensive experience with OM chemistry and thermodynamics, analytical instruments, lab experiments, data analysis, and data management.
Robert Hall	2,3	Stream metabolism expertise	(a)	UMT	World expert in reach-scale stream metabolism, stream BGC, and associated technology. Co-developer of streamMetabolizer. Extensive experience in biophysically diverse watersheds.
Glenn Hammond	1,2,3	Software Sustainability Cross-Cut Lead		PNNL	Lead PFLOTRAN Developer. Extensive experience in computational geosciences, high performance computing, and software development best practices and tools.
Peishi Jiang	1,2,3	Watershed-Scale ModEx Cross-Cut Lead		PNNL	*Extensive experience in data-driven analysis and modeling, integrated watershed modeling, and applied ModEx. Developed tools for causal inference, sensitivity analysis, and differentiable land surface modeling.
Heping Liu	1	Flux tower expertise	(a)	WSU	Recognized expert in flux towers and ecosystem exchange analysis… engaged in sfa already (Working with student)
Peter Regier	1,2,3	Field Cross-Cut Lead. Will lead the bi-weekly field coordination meetings.		PNNL	*Extensive experience with field sensors and field study design across CONUS. Leadership roles in COMPASS-FME and NEON aquatic BGC technical working group. Experience analyzing sensor data products, including machine learning applications. Familiar with RC SFA field activities, operates under COMPASS-FME field safety plan.
Jianqiu Zheng	1,2,3	Molecular-Scale ModEx Cross-Cut Lead		PNNL	Recognized expert in molecular-scale mechanistic modeling of BGC in environmental systems. Developed AquaMend. Developed advanced theory of the role of OM thermodynamics in redox-based BGC. Extensive experience

Name	ST	Primary Contributions	FTE/year	Inst.	Expertise
					in model development, lab experiments, analytical measurements, and applied ModEx.

Notes — (a) External collaborator
*Began at PNNL as a post-masters or post-doctoral research associate

C.8 Facilities and Resources

As the lead institution for the RC SFA project, PNNL has specialized and general laboratory capabilities that span a broad range of research facilities. Primary capabilities are associated with biogeochemistry, molecular separations, and computational support for river corridor science, including the Environmental Molecular Sciences Laboratory (EMSL) user facility. Our partners bring additional critical capabilities. Together we possess instrumentation and other resource capabilities and facility access capacity sufficient and critical to the success of our proposed research. Further details of these capabilities and facilities are included in **Appendix 4**. Our data management plan is included in **Appendix 5**, and our plan for software productivity, sustainability, and improvement is included in **Appendix 6**.

D. APPENDICES

Contents

Error! Bookmark not defined.

Appendix 1 - Bibliography

1. U.S. DOE. *Open Watershed Science by Design: Leveraging Distributed Research Networks to Understand Watershed Systems: Workshop Report*. doesbr.org/ openwatersheds/ (2019).
2. Butman, D. *et al. Chapter 14: Inland Waters. Second State of the Carbon Cycle Report*. https://carbon2018.globalchange.gov/chapter/14/ (2018) doi:10.7930/SOCCR2.2018.Ch14.
3. Drake, T. W., Raymond, P. A. & Spencer, R. G. Terrestrial carbon inputs to inland waters: A current synthesis of estimates and uncertainty. *Limnology and Oceanography Letters* **3**, 132–142 (2018).
4. Strahler, A. N. Quantitative analysis of watershed geomorphology. *Eos, Transactions American Geophysical Union* **38**, 913–920 (1957).
5. Cole, J. J. *et al.* Plumbing the global carbon cycle: integrating inland waters into the terrestrial carbon budget. *Ecosystems* **10**, 172–185 (2007).
6. Ward, N. D. *et al.* Where Carbon Goes When Water Flows: Carbon Cycling across the Aquatic Continuum. *Frontiers in Marine Science* **4**, (2017).
7. Battin, T. J. *et al.* River ecosystem metabolism and carbon biogeochemistry in a changing world. *Nature* **613**, 449–459 (2023).
8. McDowell, N. G. *et al.* Ecohydrological decoupling under changing disturbances and climate. *One Earth* **6**, 251–266 (2023).
9. Covino, T. Hydrologic connectivity as a framework for understanding biogeochemical flux through watersheds and along fluvial networks. *Geomorphology* **277**, 133–144 (2017).
10. Allen, D. C. *et al.* River ecosystem conceptual models and non-perennial rivers: A critical review. *Wiley Interdisciplinary Reviews: Water* **7**, e1473 (2020).
11. Abatzoglou, J. T., Williams, A. P., Boschetti, L., Zubkova, M. & Kolden, C. A. Global patterns of interannual climate–fire relationships. *Global Change Biology* **24**, 5164–5175 (2018).

12. Duane, A., Castellnou, M. & Brotons, L. Towards a comprehensive look at global drivers of novel extreme wildfire events. *Climatic Change* **165**, 43 (2021).

13. Zipper, S. C. *et al.* Pervasive changes in stream intermittency across the United States. *Environ. Res. Lett.* **16**, 084033 (2021).

14. Datry, T. *et al.* Non-perennial segments in river networks. *Nat Rev Earth Environ* **4**, 815–830 (2023).

15. McClain, M. E. *et al.* Biogeochemical Hot Spots and Hot Moments at the Interface of Terrestrial and Aquatic Ecosystems. *Ecosystems* **6**, 301–312 (2003).

16. Mirus, B. B., Ebel, B. A., Mohr, C. H. & Zegre, N. Disturbance Hydrology: Preparing for an Increasingly Disturbed Future. *Water Resources Research* **53**, 10007–10016 (2017).

17. Wymore, A. S., Ward, A. S., Wohl, E. & Harvey, J. W. Viewing river corridors through the lens of critical zone science. *Frontiers in Water* **5**, 1147561 (2023).

18. Vannote, R. L., Minshall, G. W., Cummins, K. W., Sedell, J. R. & Cushing, C. E. The River Continuum Concept. *Can. J. Fish. Aquat. Sci.* **37**, 130–137 (1980).

19. Raymond, P. A., Saiers, J. E. & Sobczak, W. V. Hydrological and biogeochemical controls on watershed dissolved organic matter transport: pulse-shunt concept. *Ecology* **97**, 5–16 (2016).

20. Bernhardt, E. S. *et al.* Control Points in Ecosystems: Moving Beyond the Hot Spot Hot Moment Concept. *Ecosystems* **20**, 665–682 (2017).

21. Battin, T. J., Besemer, K., Bengtsson, M. M., Romani, A. M. & Packmann, A. I. The ecology and biogeochemistry of stream biofilms. *Nat Rev Microbiol* **14**, 251–263 (2016).

22. Boano, F. *et al.* Hyporheic flow and transport processes: Mechanisms, models, and biogeochemical implications. *Reviews of Geophysics* **52**, 603–679 (2014).

23. Harvey, J. W. Chapter 1 - Hydrologic Exchange Flows and Their Ecological Consequences in River Corridors. in *Stream Ecosystems in a Changing Environment* (eds. Jones, J. B. & Stanley, E. H.) 1–83 (Academic Press, Boston, 2016). doi:10.1016/B978-0-12-405890-3.00001-4.

24. Son, K., Fang, Y., Gomez-Velez, J. D. & Chen, X. Spatial Microbial Respiration Variations in the Hyporheic Zones Within the Columbia River Basin. *JGR Biogeosciences* **127**, e2021JG006654 (2022).

25. Kaufman, M. H. *et al.* Sediment-associated processes are primary drivers of spatial variation in ecosystem respiration in the Yakima River Basin. (2024).

Appendix 2 – Project Publications

A2.1 Overview

In this appendix we provide a summary and analysis of SFA publications, which are numbered for ease of reference. The PNNL RCSFA's scientific impact in the past three years has increased as evidenced by total peer-reviewed publications and *h*-index. During the current reporting period, SFA researchers have published 125 articles and book chapters that were fully and partially supported by the project (CY2020–January 2024). For comparison, the SFA published 89 papers in the previous three-year reporting period (CY2017–February 2020). As shown in Figure A2.1, the trends in publication and citation rates in the past three years are very consistent with performance in the previous triennial periods, with the exception that overall publication numbers are up by approximately ten reflecting a large number of papers published from the group that were submitted at the time of the last research plan submittal (2020). A complete list of publication citations is provided in Section A2.2. An additional 26 articles have been submitted by the SFA and are in review as of February 1, 2020, for a total of 113 articles. For those manuscripts that are submitted (in review or available on preprint servers), we have included the article tracking number from the journal in the citation, and a listing with abstracts is provided in Section A2.3.

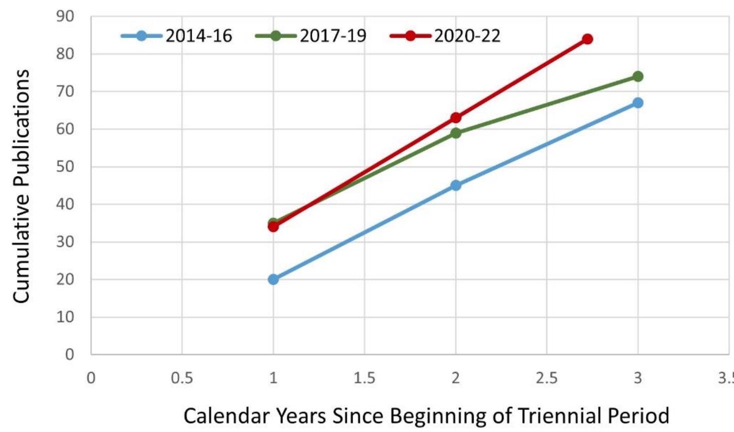


Figure A2.1. Trends in the number of peer-reviewed publications and citations are consistent between the current and previous reporting periods, but overall publication numbers have increased by more than ten articles.

Of the 125 total articles published during this reporting period, xx have current citation data indexed in the Web of Science (recently published or accepted articles do not yet appear in WoS). These indexed articles have been cited over xxx times to date. The total number of citations includes an aggregate set of citation statistics—including self-citations—for the publications that appear in the Web of Science Core Collection between dates stated. The PNNL SBR SFA *h*-index of ## during this time period means there were ## papers from January 1, 2020, through January 31, 2024, that were cited at least ## times.

A2.2 SFA Publications (2020 to February 2024)

CY2024 – 3 Publications Published or Accepted

1. *Hou, Z. J., Ward, N. D., Myers-Pigg, A. N., Lin, X., Waichler, S. R., Wiese Moore, C., Norwood, M. J., Regier, P., Yabusaki, S. B. 2024. Quantifying drivers of methane hydrobiogeochemistry in a tidal river floodplain system. *Water*, **16**, (1), doi: 10.3390/w16010171.
2. *Barnes, M. E., Johnson, D. W., Hart, S. C. 2024. The median isn't the message: Soil nutrient hot spots have a disproportionate influence on biogeochemical structure across years, seasons, and depths. *Biogeochemistry*, Accepted; doi: 10.1007/s10533-023-01107-x.
3. *Roebuck, J., Myers-Pigg, A., Garayburu-Caruso, V., Stegen, J. C. 2024. Investigating the impacts of solid phase extraction on dissolved organic matter optical signatures and the pairing with high-resolution mass spectrometry data across a freshwater stream network. *Limnology and Oceanography Methods*, Accepted; LOM-23-03-0024.R2.

CY2024 – 26 Manuscripts Submitted or under Review, including Pre-prints

See Section A2.3 for a list of abstracts of submitted publications (currently in review) and journal tracking numbers.

CY2023 – 23 Publications Published or Accepted

1. *Ahamed, F., You, Y., Burgin, A., Stegen, J. C., Scheibe, T. D., Song, H.-S. 2023. Exploring the determinants of organic matter bioavailability through substrate-explicit thermodynamic modeling. *Frontiers in Water*, **5**, 1169701; doi: 10.3389/frwa.2023.1169701.
2. *Butler, Z., Good, S., Haagsma, M., Segura, C., Hu, H. 2023. Relationship between isotope ratios in precipitation and stream water across watersheds of the National Ecological Observation Network. *Hydrological Processes*, **37**, (11), doi: 10.1002/hyp.15018.
3. *Chen, K., Yang, S., Roden, E. E., Chen, X., Chang, K. Y., Guo, Z., Liang, X., Ma, E., Fan, L., Zheng, C. 2023. Influence of vertical hydrologic exchange flow, channel flow, and biogeochemical kinetics on CH_4 emissions from rivers. *Water Resources Research*, **59**, (12), doi: 10.1029/2023wr035341.
4. *Chen K., Chen X., Stegen J. C., Villa J. A., Bohrer G., Song X., Chang K. Y., Kaufman M., Liang X., Guo Z., Roden E. E. and Zheng C. (2023). Vertical hydrologic exchange flows control methane emissions from riverbed sediments. *Environmental Science & Technology* **57**,(9), 4014-4026; doi: 10.1021/acs.est.2c07676.
5. *Chen, X. Y., Dwivedi, D., Michael, H. A. 2023. Editorial: Hydrology, ecology, and nutrient biogeochemistry at the terrestrial-aquatic interface. *Frontiers in Environmental Science*, **11**, doi: 10.3389/fenvs.2023.1322708.
6. *Cui, Y., Wen, S., Stegen, J. C., Hu, A., Wang, J. 2023. Chemodiversity of riverine dissolved organic matter: Effects of local environments and watershed characteristics. *Water Research*, 121054; doi: 10.1016/j.watres.2023.121054.
7. *Danczak, R. E., Garayburu-Caruso, V. A., Renteria, L., McKever, S. A., Otenburg, O. C., Grieger, S. R., Son, K., Kaufman, M. H., Fulton, S. G., Roebuck, J. A., Myers-Pigg, A. N., Stegen, J. C. 2023. Riverine organic matter functional diversity increases with catchment size. *Frontiers in Water*, **5**, 1087108; doi: 10.3389/frwa.2023.1087108.
8. *Fulton, S. G., Stegen, J. C., Kaufman, M. H., Dowd, J., Thompson, A. 2023. Laboratory evaluation of open source and commercial electrical conductivity sensor precision and accuracy: How do they compare? *PLoS One*, **18**, (5), e0285092; doi: 10.1371/journal.pone.0285092.
9. *Graham, E. B., Song, H.-S., Grieger, S., Garayburu-Caruso, V. A., Stegen, J. C., Bladon, K. D., Myers-Pigg, A. N. 2023. Potential bioavailability of representative pyrogenic organic matter compounds in comparison to natural dissolved organic matter pools. *Biogeosciences*, **20**, (16), 3449-3457; doi: 10.5194/bg-20-3449-2023.
10. *Karra, S. K., Apte, S. V., He, X., Scheibe, T. D. 2023. Pore-resolved investigation of turbulent open channel flow over a randomly packed permeable sediment bed. *Journal of Fluid Mechanics*, **971**, doi: 10.1017/jfm.2023.636.
11. Kaufman, M. H., Torgeson, J., Stegen, J. C. 2023. Metabolic multireactor: Practical considerations for using simple oxygen sensing optodes for high-throughput batch reactor metabolism experiments. *PLoS One*, **18**, (7), e0284256; doi: 10.1371/journal.pone.0284256.
12. *McDowell N. G., Anderson-Teixeira K., Biederman J. A., Breshears D. D., Fang Y., Fernández-de-Uña L., Graham E. B., Mackay D. S., McDonnell J. J., Moore G. W., Nehemy M. F., Stevens Rumann C. S., Stegen J., Tague N., Turner M. G. and Chen X. (2023). Ecohydrological decoupling under changing disturbances and climate. *One Earth* **6**,(3), 251-266; doi: 10.1016/j.oneear.2023.02.007.

13. *Moreland K. C., Barnes M. E., O'Geen A., Dove N., Hart S. C. and Berhe A. A. (2023). Climatic controls on soil and saprock nitrogen distribution and persistence in the Sierra Nevada. *Journal of Plant Nutrition and Soil Science* **186**,(1), 116-129; doi: 10.1002/jpln.202200218.
14. *Patel, K. F., Myers-Pigg, A. N., Bond-Lamberty, B., Barnes, M. E. 2023. nmrrr: A reproducible workflow for binning and visualizing NMR spectra from environmental samples. *Journal of Geophysical Research: Biogeosciences*, **128**, (12), doi: 10.1029/2023jg007768.
15. *Pavlopoulos, G. A., Baltoumas, F. A., Liu, S., Selvitopi, O., Camargo, A. P., Nayfach, S., Azad, A., Roux, S., Call, L., Ivanova, N. N., Chen, I. M., Paez-Espino, D., Karatzas, E., Novel Metagenome Protein Families, C., Iliopoulos, I., Konstantinidis, K., Tiedje, J. M., Pett-Ridge, J., Baker, D., Visel, A., Ouzounis, C. A., Ovchinnikov, S., Buluc, A., Kyrpides, N. C. 2023. Unraveling the functional dark matter through global metagenomics. *Nature*, **622**, (7983), 594-602; doi: 10.1038/s41586-023-06583-7.
16. *Perez, G., Gomez-Velez, J. D., Chen, X., Scheibe, T. 2023. The directional unit hydrograph model: Connecting streamflow response to storm dynamics. *Journal of Hydrology*, **627**, (Part A), 130422; doi: 10.1016/j.jhydrol.2023.130422.
17. *Rodríguez-Ramos, J., Oliverio, A., Borton, M. A., Danczak, R., Mueller, B. M., Schulz, H., Ellenbogen, J., Flynn, R. M., Daly, R. A., Schopflin, L., Shaffer, M., Goldman, A., Lewandowski, J., Stegen, J. C., Wrighton, K. C. 2023. Spatial and temporal metagenomics of river compartments reveals viral community dynamics in an urban impacted stream. *Frontiers in Microbiomes*, **2**, doi: 10.3389/frmbi.2023.1199766.
18. *Roley, S., Hall, B., Perkins, W., Garayburu-Caruso, V., Stegen, J. C. 2023. Coupled primary production and respiration in a large river contrasts with smaller rivers and streams. *Limnology and Oceanography*, 9999, 1-15; doi: 10.1002/lno.12435.
19. *Shuai, P., Jiang, P., Coon, E. T., Chen, X. 2023. The importance of explicitly representing the streambed in watershed models. *Hydrological Processes*, **37**, (12), doi: 10.1002/hyp.15043.
20. *Stadler, M., Barnard, M. A., Bice, K., de Melo, M. L., Dwivedi, D., Freeman, E. C., Garayburu-Caruso, V. A., Linkhorst, A., Mateus-Barros, E., Shi, C., Tanentzap, A. J., Meile, C. 2023. Applying the core-satellite species concept: Characteristics of rare and common riverine dissolved organic matter. *Frontiers in Water*, **5**, 1156042; doi: 10.3389/frwa.2023.1156042.
21. *Stegen, J. C., Garayburu-Caruso, V. A., Danczak, R. E., Goldman, A. E., Renteria, L., Torgeson, J. M., Hager, J. 2023. Maximum respiration rates in hyporheic zone sediments are primarily constrained by organic carbon concentration and secondarily by organic matter chemistry. *Biogeosciences*, **20**, (14), 2857-2867; doi: 10.5194/bg-20-2857-2023.
22. *Turetcaia, A. B., Garayburu-Caruso, V. A., Kaufman, M. H., Danczak, R. E., Stegen, J. C., Chu, R. K., Toyoda, J. G., Cardenas, M. B., Graham, E. B. 2023. Rethinking aerobic respiration in the hyporheic zone under variation in carbon and nitrogen stoichiometry. *Environmental Science & Technology*, **57**, (41), 15499-15510; doi: 10.1021/acs.est.3c04765.
23. *Xiao Y., Sloan J., Hepworth C., Fradera-Soler M., Mathers A., Thorley R., Baillie A., Jones H., Chang T., Chen X., Yaapar N., Osborne C. P., Sturrock C., Mooney S. J., Fleming A. J. and Zhu X. G. (2023). Defining the scope for altering rice leaf anatomy to improve photosynthesis: a modelling approach. *New Phytol* **237**,(2), 441-453; doi: 10.1111/nph.18564.

CY2022 – 34 Publications Published or Accepted

24. *Acharya B. S., Ahmmed B., Chen Y., Davison J. H., Haygood L., Hensley R. T., Kumar R., Lerback J., Liu H., Mehan S., Mehana M., Patil S. D., Persaud B. D., Sullivan P. L. and URycki D. (2022). Hydrological Perspectives on Integrated, Coordinated, Open, Networked (ICON) Science. *Earth and Space Science* 9,(4), e2022EA002320; doi: https://doi.org/10.1029/2022EA002320.

25. *Apte S. V., Oujia T., Matsuda K., Kadoch B., He X. and Schneider K. (2022) Clustering of inertial particles in turbulent flow through a porous unit cell. *Journal of Fluid Mechanics*, **937**, A9; doi: 10.1017/jfm.2022.100.
26. Arora B., Briggs M., Zarnetske J., Stegen J. C., Gomez-Velez J., Dwivedi D. and Steefel C. (2022). Hot Spots and Hot Moments in the Critical Zone: Identification of and Incorporation into Reactive Transport Models. Biogeochemistry of the Critical Zone. Advances in Critical Zone Science. A. S. Wymore, W. H. Yang, W. L. Silver, W. H. McDowell and J. Chorover, Springer-Nature. doi: 10.1007/978-3-030-95921-0_2.
27. Bao J., Chen Y., Fang Y., Song X., Perkins W., Duan Z., Shuai P., Ren H., Hou Z., Richmond M., He X. and Scheibe T. (2022) Modeling framework for evaluating the impacts of hydrodynamic pressure on hydrologic exchange fluxes and residence time for a large-scale river section over a long-term period. *Environmental Modelling & Software*, **148**, 105277; doi: 10.1016/j.envsoft.2021.105277.
28. *Borton M. A., Collins S. M., Graham E. B., Garayburu-Caruso V. A., Goldman A. E., de Melo M., Renteria L., Stegen J. C. and W. C. C. (2022) It takes a village: Using a crowdsourced approach to investigate organic matter composition in global rivers through the lens of ecological theory. *Frontiers in Water*, **4**, 870453; doi: 10.3389/frwa.2022.870453.
29. *Chen Y., Bao J., Fang Y., Perkins W. A., Ren H., Song X., Duan Z., Hou Z., He X. and Scheibe T. D. (2022) Modeling of streamflow in a 30km long reach spanning 5 years using OpenFOAM 5.x. *Geoscientific Model Development*, **15**, 2917-2947; doi: 10.5194/gmd-15-2917-2022.
30. *Chen K. W., Chen X. Y., Song X. H., Briggs M. A., Jiang P. S., Shuai P., Hammond G., Zhan H. B. and Zachara J. M. (2022) Using ensemble data assimilation to estimate transient hydrologic exchange flow under highly dynamic flow conditions. *Water Resources Research*, **58**, 5, e2021WR030735, doi: 10.1029/2021WR030735.
31. *Dwivedi D., Steefel C. I., Arora B., Banfield J., Bargar J., Boyanov M. I., Brooks S. C., Chen X., Hubbard S. S., Kaplan D., Kemner K. M., Nico P. S., O'Loughlin E. J., Pierce E. M., Painter S. L., Scheibe T. D., Wainwright H. M., Williams K. H. and Zavarin M. (2022) From legacy contamination to watershed systems science: A review of scientific insights and technologies developed through DOE-supported research in water and energy security. Environmental Research Letters, 17, 043004; doi: 10.1088/1748-9326/ac59a9.
32. *Fremin B. J.Bhatt A. S.Kyrpides N. C.Sengupta A.Sczyrba A.Maria da Silva A.Buchan A.Gaudin A.Brune A.Hirsch A. M.Neumann A.Shade A.Visel A.Campbell B.Baker B.Hedlund B. P.Crump B. C.Currie C.Kelly C.Craft C.Hazard C.Francis C.Schadt C. W.Averill C.Mobilian C.Buckley D.Hunt D.Noguera D.Beck D.Valentine D. L.Walsh D.Sumner D.Lymperopoulou D.Bhaya D.Bryant D. A.Morrison E.Brodie E.Young E.Lilleskov E.Högfors-Rönnholm E.Chen F.Stewart F.Nicol G. W.Teeling H.Beller H. R.Dionisi H.Liao H.-L.Beman J. M.Stegen J.Tiedje J.Jansson J.VanderGheynst J.Norton J.Dangl J.Blanchard J.Bowen J.Macalady J.Pett-Ridge J.Rich J.Payet J. P.Gladden J. D.Raff J. D.Klassen J. L.Tarn J.Neufeld J.Gravuer K.Hofmockel K.Chen K.-H.Konstantinidis K.DeAngelis K. M.Partida-Martinez L. P.Meredith L.Chistoserdova L.Moran M. A.Scarborough M.Schrenk M.Sullivan M.David M.O'Malley M. A.Medina M.Habteselassie M.Ward N. D.Pietrasiak N.Mason O. U.Sorensen P. O.Estrada de los Santos P.Baldrian P.McKay R. M.Simister R.Stepanauskas R.Neumann R.Malmstrom R.Cavicchioli R.Kelly R.Hatzenpichler R.Stocker R.Cattolico R. A.Ziels R.Vilgalys R.Blumer-Schuette S., et al. (2022) Thousands of small, novel genes predicted in global phage genomes. Cell Reports, **39**, 110984; doi: https://doi.org/10.1016/j.celrep.2022.110984.
33. *Hammond G. E. (2022) The PFLOTRAN Reaction Sandbox. *Geoscientific Model Development*, 2022, 1-26; doi: 10.5194/gmd-15-1659-2022.
34. *Hills D. J., Damerow J. E., Ahmmed B., Catolico N., Chakraborty S., Coward C. M., Crystal-Ornelas R., Duncan W. D., Goparaju L. N., Lin C., Liu Z., Mudunuru M. K., Rao Y., Rovetto R. J., Sun Z., Whitehead B. P., Wyborn L. and Yao T. (2022) Earth and Space Science Informatics

Perspectives on Integrated, Coordinated, Open, Networked (ICON) Science. *Earth and Space Science*, **9**, e2021EA002108; doi: 10.1029/2021ea002108.

35. *Isabwe A., Yang J. R., Wang Y., Wilkinson D. M., Graham E. B., Chen H. and Yang J. (2022). Riverine bacterioplankton and phytoplankton assembly along an environmental gradient induced by urbanization. *Limnology and Oceanography* **67**,(9), 1943-1958; doi: 10.1002/lno.12179.
36. *Jiang P. S., Son K., Mudunuru M. K. and Chen X. Y. (2022). Using Mutual Information for Global Sensitivity Analysis on Watershed Modeling. *Water Resources Research* **58**,(10), 2022WR032932; doi: 10.1029/2022WR032932.
37. *Jiang P. S., Chen X. Y., Missik J. E. C., Gao Z. M., Liu H. P. and Verbeke B. A. (2022). Encoding diel hysteresis and the Birch effect in dryland soil respiration models through knowledge-guided deep learning. Frontiers in Environmental Science 10, doi: 10.3389/fenvs.2022.1035540.
38. Johnson T., Thomle J., Stickland C., Goldman A. and Stegen J. (2022) Riverbed temperature and 4D ERT monitoring reveals heterogenous horizontal and vertical groundwater-surface water exchange flows under dynamic stage conditions. *Frontiers in Earth Science*, **10**; doi: 10.3389/feart.2022.910058.
39. Kaufman M. H., Ghosh R. N., Grate J., Shooltz D. D., Freeman M. J., Ball T. M., Loloee R., McIntire C. W., Wells J., Strickland C., Vermeul V., Rod K. A., Mackley R., Lin X., Ren H., Goldman A. and Stegen J. (2022) Dissolved oxygen sensor in an automated hyporheic sampling system reveals biogeochemical dynamics. *PLOS Water*, 1, e0000014; doi: 10.1371/journal.pwat.0000014.
40. *Leonard L. T., Vanzin G. F., Garayburu-Caruso V. A., Lau S. S., Beutler C. A., Newman A. W., Mitch W. A., Stegen J. C., Williams K. H. and Sharp J. O. (2022) Disinfection byproducts formed during drinking water treatment reveal an export control point for dissolved organic matter in a subalpine headwater stream. *Water Research X*, 15, 100144; doi: 10.1016/j.wroa.2022.100144.
41. *Little C. J., Rizzuto M., Luhring T. M., Monk J. D., Nowicki R. J., Paseka R. E., Stegen J. C., Symons C. C., Taub F. B. and Yen J. (2022) Movement with meaning: integrating information into meta-ecology. Oikos, 2022, e08892; doi: 10.1111/oik.08892.
42. *Maltz M. R., Carey C. J., Freund H. L., Botthoff J. K., Hart S. C., Stajich J. E., Aarons S. M., Aciego S. M., Blakowski M., Dove N. C., Barnes M. E., Pombubpa N. and Aronson E. L. (2022) Landscape topography and regional drought alters dust microbiomes in the Sierra Nevada of California. *Frontiers in Microbiology*, **13**, 856454; doi: 10.3389/fmicb.2022.856454.
43. *Mudunuru M. K., Son K., Jiang P., Hammond G. and Chen X. (2022). Scalable deep learning for watershed model calibration. F*rontiers in Earth Science* **10**, doi: 10.3389/feart.2022.1026479.
44. *Nelson A. R., Toyoda J., Chu R. K., Tolić N., Garayburu-Caruso V. A., Saup C. M., Renteria L., Wells J. R., Stegen J. C., Wilkins M. J. and Danczak R. E. (2022) Implications of sample treatment on characterization of riverine dissolved organic matter. *Environmental Science: Processes & Impacts*, **24**, 773-782; doi: 10.1039/D2EM00044J.
45. Ren H. Y., Cromwell E., Kravitz B. and Chen X. Y. (2022) Technical note: Using long short-term memory models to fill data gaps in hydrological monitoring networks. Hydrology and Earth System Sciences, 26, 1727-1743; doi: 10.5194/hess-26-1727-2022.
46. *Rodriguez-Ramos J. A., Borton M. A., McGivern B. B., Smith G. J., Solden L. M., Shaffer M., Daly R. A., Purvine S. O., Nicora C. D., Eder E. K., Lipton M., Hoyt D. W., Stegen J. C. and Wrighton K. C. (2022) Genome-resolved metaproteomics decodes the microbial and viral contributions to coupled carbon and nitrogen cycling in river sediments. *mSystems*, e0051622; doi: 10.1128/msystems.00516-22.
47. *Roebuck J. A., Bladon K. D., Donahue D., Graham E. B., Grieger S., Morgenstern K., Norwood M. J., Wampler K. A., Erkert L., Renteria L., Danczak R., Fricke S. and Myers-Pigg A. N. (2022) Spatiotemporal controls on the delivery of dissolved organic matter to streams following a wildfire. *Geophysical Research Letters*, **49**; doi: 10.1029/2022gl099535.

48. Rubinstein R. L., Borton M. A., Zhou H., Shaffer M., Hoyt D. W., Stegen J., Henry C. S., Wrighton K. C. and Versteeg R. (2022) ORT: A workflow linking genome-scale metabolic models with reactive transport codes. *Bioinformatics*, **38**, 778-784; doi: 10.1093/bioinformatics/btab753.
49. *Shaffer J. P., Nothias L. F., Thompson L. R., Sanders J. G., Salido R. A., Couvillion S. P., Brejnrod A. D., Lejzerowicz F., Haiminen N., Huang S., Lutz H. L., Zhu Q., Martino C., Morton J. T., Karthikeyan S., Nothias-Esposito M., Duhrkop K., Bocker S., Kim H. W., Aksenov A. A., Bittremieux W., Minich J. J., Marotz C., Bryant M. M., Sanders K., Schwartz T., Humphrey G., Vasquez-Baeza Y., Tripathi A., Parida L., Carrieri A. P., Beck K. L., Das P., Gonzalez A., McDonald D., Ladau J., Karst S. M., Albertsen M., Ackermann G., DeReus J., Thomas T., Petras D., Shade A., Stegen J., Song S. J., Metz T. O., Swafford A. D., Dorrestein P. C., Jansson J. K., Gilbert J. A., Knight R. and Earth Microbiome Project C. (2022). Standardized multi-omics of Earth's microbiomes reveals microbial and metabolite diversity. *Nature Microbiology* **7**,(12), 2128-2150; doi: 10.1038/s41564-022-01266-x.
50. *Son K., Fang Y., Gomez-Velez J. D. and Chen X. (2022). Spatial microbial respiration variations in the hyporheic zones within the Columbia River Basin. Journal of Geophysical Research: Biogeosciences 127,(11), e2021JG006654; doi: 10.1029/2021JG006654.
51. *Son K., Fang Y., Gomez-Velez J. D., Byun K. and Chen X. (2022). Combined effects of stream hydrology and land use on basin-scale hyporheic zone denitrification in the Columbia River Basin. *Water Resources Research* **58**,(12) doi: 10.1029/2021wr031131.
52. *Simmonds M. B., Riley W. J., Agarwal D. A., Chen X., Cholia S., Crystal-Ornelas R., Coon E. T., Dwivedi D., Hendrix V. C., Huang M., Jan A., Kakalia Z., Kumar J., Koven C. D., Li L., Melara M., Ramakrishnan L., Ricciuto D. M., Walker A. P., Zhi W., Zhu Q. and Varadharajan C. (2022). Guidelines for publicly archiving terrestrial model data to enhance usability, intercomparison, and synthesis. *Data Science Journal* **21**,(1) doi: 10.5334/dsj-2022-003.
53. *Ward A. S., Packman A., Bernal S., Brekenfeld N., Drummond J., Graham E., Hannah D. M., Klaar M., Krause S., Kurz M., Li A. G., Lupon A., Mao F., Roca M. E. M., Ouellet V., Royer T. V., Stegen J. C. and Zarnetske J. P. (2022) Advancing river corridor science beyond disciplinary boundaries with an inductive approach to catalyse hypothesis generation. *Hydrological Processes*, **36**, e14540; doi: 10.1002/hyp.14540.
54. *Yang J., Ye M., Chen X., Dai H. and Walker A. P. (2022). Process interactions can change process ranking in a coupled complex system under process model and parametric uncertainty. *Water Resources Research* **58**,(3) doi: 10.1029/2021wr029812.
55. *Yang Y., Berhe A. A., Hunsaker C. T., Johnson D. W., Safeeq M., Barnes M. E., McCorkle E. P., Stacy E. M., Bales R. C., Bart R. R., Goulden M. L. and Hart S. C. (2022) Impacts of climate and disturbance on nutrient fluxes and stoichiometry in mixed-conifer forests. *Biogeochemistry*, **158**, 1-20; doi: 10.1007/s10533-021-00882-9.
56. *Yang Y., Berhe A. A., Barnes M. E., Moreland K. C., Tian Z., Kelly A. E., Bales R. C., O'Geen A. T., Goulden M. L., Hartsough P. and Hart S. C. (2022). Climate warming alters nutrient storage in seasonally dry forests: insights from a 2,300 m elevation gradient. *Global Biogeochemical Cycles* **36**,(11) doi: 10.1029/2022gb007429.
57. *Yao J., Yuan W., Gao H., Liu H., Chen X., Ma Y., Arntzen E. and McFarland D. P. (2022). Impact of shifts in vegetation phenology on the carbon balance of a semiarid sagebrush ecosystem. *Remote Sensing* **14**,(23), 5924; doi: 10.3390/rs14235924.

CY2021 – 30 Publications Published or Accepted

1. Chen X., Lee R. M., Dwivedi D., Son K., Fang Y., Zhang X., Graham E., Stegen J., Fisher J. B., Moulton D. and Scheibe T. D. (2021) Integrating field observations and process-based modeling to

predict watershed water quality under environmental perturbations. *Journal of Hydrology*, **602**, 125762; doi: 10.1016/j.jhydrol.2020.125762.

2. *Chu H., Luo X., Ouyang Z., Chan W. S., Dengel S., Biraud S. C., Torn M. S., Metzger S., Kumar J., Arain M. A., Arkebauer T. J., Baldocchi D., Bernacchi C., Billesbach D., Black T. A., Blanken P. D., Bohrer G., Bracho R., Brown S., Brunsell N. A., Chen J., Chen X., Clark K., Desai A. R., Duman T., Durden D., Fares S., Forbrich I., Gamon J. A., Gough C. M., Griffis T., Helbig M., Hollinger D., Humphreys E., Ikawa H., Iwata H., Ju Y., Knowles J. F., Knox S. H., Kobayashi H., Kolb T., Law B., Lee X., Litvak M., Liu H., Munger J. W., Noormets A., Novick K., Oberbauer S. F., Oechel W., Oikawa P., Papuga S. A., Pendall E., Prajapati P., Prueger J., Quinton W. L., Richardson A. D., Russell E. S., Scott R. L., Starr G., Staebler R., Stoy P. C., Stuart-Haëntjens E., Sonnentag O., Sullivan R. C., Suyker A., Ueyama M., Vargas R., Wood J. D. and Zona D. (2021) Representativeness of Eddy-Covariance flux footprints for areas surrounding AmeriFlux sites. Agricultural and Forest Meteorology, 301-302, 108350; doi: https://doi.org/10.1016/j.agrformet.2021.108350.
3. *Conner A., Gooseff M. N., Chen X., Arntzen E. and Garayburu-Caruso V. (2021) Groundwater inflows to the Columbia River along the hanford reach and associated nitrate concentrations. *Frontiers in Water*, **3**; doi: 10.3389/frwa.2021.574684.
4. *Damerow J., Varadharajan C., Boye K., Brodie E. L., Burrus M., Chadwick D., Crystal-Ornelas R., Elbashandy H., Eloy Alves R., Ely K., Goldman A., Haberman T., Hendrix V., Kakalia Z., Kemner K. M., Kersting A., Merino N., O'Brien F., Perzan Z., Robles E., Sorensen P., Stegen J., Walls R., Weisenhorn P., Zavarin M. and Agarwal D. (2021) Sample identifiers and metadata to support data management and reuse in multidisciplinary ecosystem sciences. *Data Science Journal*, **20**, 1-19; doi: 10.5334/dsj-2021-011.
5. Danczak R. E., Goldman A. E., Chu R. K., Toyoda J. G., Garayburu-Caruso V. A., Tolic N., Graham E. B., Morad J. W., Renteria L., Wells J. R., Herzog S. P., Ward A. S. and Stegen J. C. (2021) Ecological theory applied to environmental metabolomes reveals compositional divergence despite conserved molecular properties. *Science of the Total Environment*, **788**, 147409; doi: 10.1016/j.scitotenv.2021.147409.
6. *Dove N. C., Veach A. M., Muchero W., Wahl T., Stegen J. C., Schadt C. W. and Cregger M. A. (2021) Assembly of the populus microbiome is temporally dynamic and determined by selective and stochastic factors. *Msphere*, **6**, e01316-01320; doi: 10.1128/mSphere.01316-20.
7. *Dwivedi D., Godsey S. E. and Scheibe T. D. (2021) Editorial: Linking hydrological and biogeochemical processes in riparian corridors. *Frontiers in Water*, **3**; doi: 10.3389/frwa.2021.693763.
8. *Fegel T. S., Boot C. M., Covino T. P., Elder K., Hall E. K., Starr B., Stegen J. and Rhoades C. C. (2021) Amount and reactivity of dissolved organic matter export are affected by land cover change from old-growth to second-growth forests in headwater ecosystems. *Hydrological Processes*, **35**, e14343; doi: 10.1002/hyp.14343.
9. *Fudyma J. D., Chu R., Grachet N. G., Stegen J. and Tfaily M. (2021) Coupled biotic-abiotic processes control biogeochemical cycling of dissolved organic matter in the Columbia River hyporheic zone. *Frontiers in Water*, **2**, 574692; doi: 10.3389/frwa.2020.574692.
10. *Graham E. B. and Smith A. P. (2021) Crowdsourcing global perspectives in ecology using social media. *Frontiers in Ecology and Evolution*, **9**; doi: 10.3389/fevo.2021.588894.
11. Graham E. B., Averill C., Bond-Lamberty B., Knelman J. E., Krause S., Peralta A. L., Shade A., Smith A. P., Cheng S. J., Fanin N., Freund C., Garcia P. E., Gibbons S. M., Van Goethem M. W., Guebila M. B., Kemppinen J., Nowicki R. J., Pausas J. G., Reed S. P., Rocca J., Sengupta A., Sihi D., Simonin M., Słowiński M., Spawn S. A., Sutherland I., Tonkin J. D., Wisnoski N. I., Zipper S. C., C. C., Staal A., Arora B., Oldfield C., Dwivedi D., Larson E., Santillan E., Aaron Hogan J., Atkins J.,

Zheng J., Lembrechts J., Patel K., Copes-Gerbitz K., Winker K., Mudge L., Wong M., Nuñez M., Luoto M. and Barnes R. (2021) Toward a generalizable framework of disturbance ecology through crowdsourced science. *Frontiers in Ecology and Evolution*, **9**, 588940; doi: 10.3389/fevo.2021.588940.

12. *He X., Apte S. V., Karra S. K. and Doğan Ö. N. (2021) An LES study of secondary motion and wall shear stresses in a pipe bend. *Physics of Fluids*, **33**, 115102; doi: 10.1063/5.0065338.
13. Hou Z., Ren H., Murray C. J., Song X., Fang Y., Arntzen E. V., Chen X., Stegen J. C., Huang M., Gomez-Velez J. D., Duan Z., Perkins W. A., Richmond M. C. and Scheibe T. D. (2021) A novel construct for scaling groundwater–river interactions based on machine-guided hydromorphic classification. *Environmental Research Letters*, **16**, 104016; doi: 10.1088/1748-9326/ac24ce.
14. Kaufman M. H., Warden J. G., Cardenas M. B., Stegen J. C., Graham E. B. and Brown J. (2021) Evaluating a laboratory flume microbiome as a window into natural riverbed biogeochemistry. *Frontiers in Water*, **3**, 596260; doi: 10.3389/frwa.2021.596260.
15. *Liu W., Graham E. B., Dong Y., Zhong L., Zhang J., Qiu C., Chen R., Lin X. and Feng Y. (2021) Balanced stochastic versus deterministic assembly processes benefit diverse yet uneven ecosystem functions in representative agroecosystems. *Environmental Microbiology*, **23**, 391-404; doi: https://doi.org/10.1111/1462-2920.15326.
16. Missik J. E. C., Liu H., Gao Z., Huang M., Chen X., Arntzen E., McFarland D. P. and Verbeke B. (2021) Groundwater regulates interannual variations in evapotranspiration in a riparian semiarid ecosystem. *Journal of Geophysical Research: Atmospheres*, **126**, e2020JD033078; doi: 10.1029/2020jd033078.
17. Moghaddam M. A., Ferré T. P. A., Chen X., Chen K., Song X. and Hammond G. (2021). Can simple machine learning tools extend and improve temperature-based methods to infer streambed flux? Water 13,(20), 2837; doi: 10.3390/w13202837.
18. *Mueller B. M., Schulz H., Danczak R. E., Putschew A. and Lewandowski J. (2021) Simultaneous attenuation of trace organics and change in organic matter composition in the hyporheic zone of urban streams. *Scientific Reports*, **11**, 4179; doi: 10.1038/s41598-021-83750-8.
19. Nayfach S., Roux S., Seshadri R., Udwary D., Varghese N., Schulz F., Wu D. Y., Paez-Espino D., Chen I. M., Huntemann M., Palaniappan K., Ladau J., Mukherjee S., Reddy T. B. K., Nielsen T., Kirton E., Faria J. P., Edirisinghe J. N., Henry C. S., Jungbluth S. P., Chivian D., Dehal P., Wood-Charlson E. M., Arkin A. P., Tringe S. G., Visel A., Woyke T., Mouncey N. J., Ivanova N. N., Kyrpides N. C., Eloe-Fadrosh E. A. and Consortium I. M. D. (2021) A genomic catalog of Earth's microbiomes (Nov, 10.1038/s41587-020-0718-6, 2020). *Nature Biotechnology*, **39**, 521-521; doi: 10.1038/s41587-020-0718-6
20. *Perez G., Gomez-Velez J. D., Chen X., Scheibe T., Chen Y. and Bao J. (2021) Identification of characteristic spatial scales to improve the performance of analytical spectral solutions to the groundwater flow equation. *Water Resources Research*, **57**, e2021WR031044; doi: https://doi.org/10.1029/2021WR031044.
21. *Perez G., Gomez-Velez J. D., Mantilla R., Wright D. B. and Li Z. (2021) The effect of storm direction on flood frequency analysis. *Geophysical Research Letters*, **48**; doi: 10.1029/2020gl091918.
22. *Ren H., Song X., Fang Y., Hou Z. J. and Scheibe T. D. (2021) Machine learning analysis of hydrologic exchange flows and transit time distributions in a large regulated river. *Frontiers in Artificial Intelligence*, **4**; doi: 10.3389/frai.2021.648071.
23. *Sengupta A., Volkmann T. H. M., Danczak R. E., Stegen J. C., Dontsova K., Abramson N., Bugaj A. S., Volk M. J., Matos K. A., Meira-Neto A. A., Barberán A., Neilson J. W., Maier R. M., Chorover J., Troch P. A. and Meredith L. K. (2021) Contrasting community assembly forces drive microbial structural and potential functional responses to precipitation in an incipient soil system. Frontiers in Microbiology, 12; doi: 10.3389/fmicb.2021.754698.

24. Song X., Fang Y., Bao J., Ren H., Duan Z., Perkins W., Zhou H., Hou Z., Chen Y. and Scheibe T. (2021) Scale-dependent spatial variabilities of hydrological exchange flows and transit time in a large regulated river. *Journal of Hydrology*, **598**, 126283; doi: 10.1016/j.jhydrol.2021.126283.
25. *Song H.-S., Stegen J. C., Graham E. B. and Scheibe T. D. (2021) Historical contingency in microbial resilience to hydrologic perturbations. *Frontiers in Water*, **3**; doi: 10.3389/frwa.2021.590378.
26. *Wu L., Gomez-Velez J. D., Krause S., Wörman A., Singh T., Nützmann G. and Lewandowski J. (2021) How daily groundwater table drawdown affects the diel rhythm of hyporheic exchange. *Hydrol. Earth Syst. Sci.*, 25, 1905-1921; doi: 10.5194/hess-25-1905-2021.
27. *Wu R., Chen X., Hammond G., Bisht G., Song X., Huang M., Niu G.-Y. and Ferre T. (2021) Coupling surface flow with high-performance subsurface reactive flow and transport code PFLOTRAN. *Environmental Modelling & Software*, **137**, 104959; doi: 10.1016/j.envsoft.2021.104959.
28. *Xiao Y., Sloan J., Hepworth C., Osborne C. P., Fleming A. J., Chen X. and Zhu X. G. (2021) Estimating uncertainty: A Bayesian approach to modelling photosynthesis in C3 leaves. *Plant Cell & Environment*, 44, 1436-1450; doi: 10.1111/pce.13995.
29. *Zhang Y., Zhou D., Wei W., Frame J. M., Sun H., Sun A. Y. and Chen X. (2021) Hierarchical Fractional Advection-Dispersion Equation (FADE) to quantify anomalous transport in river corridor over a broad spectrum of scales: Theory and applications. *Mathematics*, **9**, 790; doi: 10.3390/math9070790.
30. *Zhu B., Huang M., Cheng Y., Xie X., Liu Y., Bisht G. and Chen X. (2021) Impact of vegetation physiology and phenology on watershed hydrology in a semiarid watershed in the Pacific Northwest in a changing climate. *Water Resources Research*, **57**, e2020WR028394; doi: 10.1029/2020wr028394.

CY2020 – 35 Publications Published or Accepted

31. Ahmmed B., Mudunuru M. K., Karra S., James S. C., Viswanathan H. and Dunbar J. A. (2020) PFLOTRAN-SIP: A PFLOTRAN module for simulating spectral-induced polarization of electrical impedance data. *Energies*, **13**, 6552; doi: 10.3390/en13246552.
32. Bao Y., Feng Y., Stegen J. C., Wu M., Chen R., Liu W., Zhang J., Li Z. and Lin X. (2020) Straw chemistry links the assembly of bacterial communities to decomposition in paddy soils. *Soil Biology and Biochemistry*, **148**, 107866; doi: 10.1016/j.soilbio.2020.107866.
33. Chen X., Zachara J. M., Vermuel V. R., Hammond G., Freshley M. and Fang Y. (2020) Understanding contaminant migration within a dynamic river corridor through field experiments and reactive transport modeling. *Frontiers in Water*, **2**; doi: 10.3389/frwa.2020.533796.
34. Danczak R. E., Chu R. K., Fansler S. J., Goldman A. E., Graham E. B., Tfaily M. M., Toyoda J. and Stegen J. C. (2020) Using metacommunity ecology to understand environmental metabolomes. Nature Communications, 11, 6369; doi: 10.1038/s41467-020-19989-y.
35. Danczak R. E., Daly R. A., Borton M. A., Stegen J. C., Roux S., Wrighton K. C. and Wilkins M. J. (2020) Ecological assembly processes are coordinated between bacterial and viral communities in fractured shale ecosystems. *Msystems*, **5**, e00098-00020; doi: 10.1128/mSystems.00098-20.
36. Fang Y. L., Chen X. Y., Velez J. G., Zhang X. S., Duan Z. R., Hammond G. E., Goldman A. E., Garayburu-Caruso V. A. and Graham E. B. (2020) A multirate mass transfer model to represent the interaction of multicomponent biogeochemical processes between surface water and hyporheic zones (SWAT-MRMT-R 1.0). *Geoscientific Model Development*, **13**, 3553-3569; doi: 10.5194/gmd-13-3553-2020.
37. *Fang Y., Song X., Ren H., Perkins W. A., Shuai P., Richmond M. C., Hou Z., Bao J., Chen X. and Scheibe T. D. (2020) High-performance simulation of dynamic hydrologic exchange and implications

for surrogate flow and reactive transport modeling in a large river corridor. *Frontiers in Water*, **2**; doi: 10.3389/frwa.2020.564211.

38. Gao Z. M., Liu H. P., Arntzen E., Mcfarland D. P., Chen X. Y. and Huang M. Y. (2020) Uncertainties in turbulent statistics and fluxes of CO2 associated with density effect corrections. *Geophysical Research Letters*, **47**; doi: 10.1029/2020GL088859.
39. Gao Z., Liu H., Chen X., Huang M., Missik J. E. C., Yao J., Arntzen E. and McFarland D. P. (2020) Enlarged nonclosure of surface energy balance with increasing atmospheric instabilities linked to changes in coherent structures. *Journal of Geophysical Research: Atmospheres*, **125**, e2020JD032889; doi: 10.1029/2020jd032889.
40. *Garayburu-Caruso V. A., Danczak R. E., Stegen J. C., Renteria L., Mccall M., Goldman A. E., Chu R. K., Toyoda J., Resch C. T., Torgeson J. M., Wells J., Fansler S., Kumar S. and Graham E. B. (2020) Using community science to reveal the global chemogeography of river metabolomes. *Metabolites*, **10**, 518; doi: 10.3390/metabo10120518.
41. Garayburu-Caruso V. A., Stegen J. C., Song H.-S., Renteria L., Wells J., Garcia W., Resch C. T., Goldman A. E., Chu R. K., Toyoda J. and Graham E. B. (2020) Carbon limitation leads to thermodynamic regulation of aerobic metabolism. *Environmental Science & Technology Letters*, **7**, 517-524; doi: 10.1021/acs.estlett.0c00258.
42. Graham E. B. and Krause S. (2020) Social media sows consensus in disturbance ecology. *Nature,* **577**, 170; doi: 10.1038/d41586-020-00006-7.
43. Grant S. B., Gomez-Velez J. D., Ghisalberti M., Guymer I., Boano F., Roche K. and Harvey J. (2020) A one-dimensional model for turbulent mixing in the benthic biolayer of stream and coastal sediments. *Water Resources Research*, **56**, e2019WR026822; doi: 10.1029/2019wr026822.
44. Grant S. B., Monofy A., Boano F., Gomez-Velez J. D., Guymer I., Harvey J. and Ghisalberti M. (2020) Unifying advective and diffusive descriptions of bedform pumping in the benthic biolayer of streams. *Water Resources Research*, **56**, e2020WR027967; doi: https://doi.org/10.1029/2020WR027967.
45. Ji M., Kong W., Stegen J., Yue L., Wang F., Dong X., Cowan D. A. and Ferrari B. C. (2020) Distinct assembly mechanisms underlie similar biogeographical patterns of rare and abundant bacteria in Tibetan Plateau grassland soils. *Environmental Microbiology*, **22**, 2261-2272; doi: 10.1111/1462-2920.14993.
46. Kruegler J., Gomez-Velez J., Lautz L. K. and Endreny T. A. (2020) Dynamic evapotranspiration alters hyporheic flow and residence times in the intrameander zone. *Water*, 12, 424; doi: 10.3390/w12020424.
47. Li M. J., Li R., Gao Y. Q., Resch C. T., Qian W. J., Shi T. J., Shi L., Liu H. and Liu C. X. (2020) Nitrate bioreduction dynamics in hyporheic zone sediments under cyclic changes of chemical compositions. Journal of Hydrology, 585; doi: 10.1016/j.jhydrol.2020.124836.
48. Li B., Liu X., Kaufman M. H., Turetcaia A., Chen X. and Cardenas M. B. (2020) Flexible and modular simultaneous modeling of flow and reactive transport in rivers and hyporheic zones. *Water Resources Research*, 56; doi: 10.1029/2019wr026528.
49. Lin X., Ren H., Goldman A., Stegen J. C. and Scheibe T. D. (2020) WHONDRS-GUI: A web application for global survey of surface water metabolites. *PeerJ*, **8**, e9277; doi: 10.7717/peerj.9277.
50. Liu W., Graham E. B., Zhong L., Zhang J., Li W., Li Z., Lin X., Feng Y. and Wang J. (2020) Dynamic microbial assembly processes correspond to soil fertility in sustainable paddy agroecosystems. *Functional Ecology*, 34, 1244-1256; doi: 10.1111/1365-2435.13550.
51. Liu W., Graham E. B., Zhong L., Zhang J., Li S., Lin X. and Feng Y. (2020) Long-term stochasticity combines with short-term variability in assembly processes to underlie rice paddy sustainability. *Frontiers in Microbiology*, 11, 873; doi: 10.3389/fmicb.2020.00873.

52. Nelson W. C., Graham E. B., Crump A. R., Fansler S. J., Arntzen E. V., Kennedy D. W. and Stegen J. C. (2020) Distinct temporal diversity profiles for nitrogen cycling genes in a hyporheic microbiome. *PLOS ONE,* **15**, e0228165; doi: 10.1371/journal.pone.0228165.
53. *Ren H., Hou Z., Duan Z., Song X., Perkins W. A., Richmond M. C., Arntzen E. V. and Scheibe T. D. (2020) Spatial mapping of riverbed grain-size distribution using machine learning. *Frontiers in Water*, **2**; doi: 10.3389/frwa.2020.551627.
54. Rizzo C. B., Song X., de Barros F. P. J. and Chen X. (2020) Temporal flow variations interact with spatial physical heterogeneity to impact solute transport in managed river corridors. *Journal of Contaminant Hydrology*, 103713; doi: https://doi.org/10.1016/j.jconhyd.2020.103713.
55. Singh, T., Gomez-Velez, J. D., Wu, L. W., Worman, A., Hannah, D. M., Krause, S. 2020. Effects of successive peak flow events on hyporheic exchange and residence times. *Water Resources Research*, **56**, (8), doi: 10.1029/2020WR027113.
56. Song H.-S., Stegen J. C., Graham E. B., Lee J.-Y., Garayburu-Caruso V. A., Nelson W. C., Chen X., Moulton J. D. and Scheibe T. D. (2020) Representing organic matter thermodynamics in biogeochemical reactions via substrate-explicit modeling. *Frontiers in Microbiology*, **11**; doi: 10.3389/fmicb.2020.531756.
57. Song X. H., Chen X. Y., Zachara J. M., Gomez-Velez J. D., Shuai P., Ren H. Y. and Hammond G. E. (2020) River dynamics control transit time distributions and biogeochemical reactions in a dam-regulated river corridor. *Water Resources Research*, **56**; doi: 10.1029/2019WR026470.
58. Thomle J., Strickland C., Johnson T. C., Zhu Y. and Stegen J. (2020) A flux detection probe to quantify dynamic groundwater-surface water exchange in the hyporheic zone. *Ground Water*, **58**, 892-900; doi: 10.1111/gwat.13001.
59. Tso C. M., Johnson T. C., Song X., Chen X., Kuras O., Wilkinson P., Uhlemann S., Chambers J. and Binley A. (2020) Integrated hydrogeophysical modelling and data assimilation for geoelectrical leak detection. *J Contam Hydrol*, **234**, 103679; doi: 10.1016/j.jconhyd.2020.103679.
60. Villa J. A., Smith G. J., Ju Y., Renteria L., Angle J. C., Arntzen E., Harding S. F., Ren H., Chen X., Sawyer A. H., Graham E. B., Stegen J. C., Wrighton K. C. and Bohrer G. (2020) Methane and nitrous oxide porewater concentrations and surface fluxes of a regulated river. *Science of The Total Environment*, **715**, 136920; doi: 10.1016/j.scitotenv.2020.136920.
61. Wang J. J., Legendre P., Soininen J., Yeh C. F., Graham E., Stegen J. C., Casamayor E. O., Zhou J. Z., Shen J. and Pan F. Y. (2020) Temperature drives local contributions to beta diversity in mountain streams: Stochastic and deterministic processes. *Global Ecology and Biogeography*, 29, 420-432; doi: 10.1111/geb.13035.
62. Wu L., Gomez-Velez J. D., Krause S., Singh T., Wörman A. and Lewandowski J. (2020) Impact of flow alteration and temperature variability on hyporheic exchange. *Water Resources Research,* **56**; doi: 10.1029/2019wr026225.
63. Yao J., Liu H., Huang J., Gao Z., Wang G., Li D., Yu H. and Chen X. (2020) Accelerated dryland expansion regulates future variability in dryland gross primary production. Nature Communications, 11, 1665; doi: 10.1038/s41467-020-15515-2.
64. Zachara J. M., Chen X., Song X., Shuai P., Murray C. and Resch C. T. (2020) Kilometer-scale hydrologic exchange flows in a gravel bed river corridor and their implications to solute migration. *Water Resources Research,* **56**, e2019WR02525; doi: 10.1029/2019wr025258.
65. Zhu B., Huang M., Cheng Y., Xie X., Liu Y., Zhang X., Bisht G., Chen X., Missik J. and Liu H. (2020) Effects of irrigation on water, carbon, and nitrogen budgets in a semiarid watershed in the Pacific Northwest: A modeling study. *Journal of Advances in Modeling Earth Systems*, **12**, e2019MS001953; doi: 10.1029/2019ms001953.

A2.3 CY2024 List of Submitted Articles Including Abstracts

Appendix 3 - Curricula Vitae

Appendix 4 - External Collaborations

Appendix 5 – Facilities and Resources

This appendix identifies the facilities and other resources to be used in the proposed work, including (when appropriate) their capacities, pertinent capabilities, relative proximity, and extent of availability to the project. The partner institutions included in this proposal have the expertise, instrumentation, and research space required to complete the proposed work. We rely on the capabilities and support infrastructure of all partners. These include unique facilities, laboratory spaces, and computational and other resources. Additional equipment descriptions may be found in Appendix 5. As the lead institution for the RC SFA project, Pacific Northwest National Laboratory (PNNL) has capabilities that span a broad range of research facilities. Primary capabilities are associated with biogeochemistry, molecular separations, and computational support for river corridor science, including the Environmental Molecular Sciences Laboratory (EMSL) user facility. Additional capabilities include the Marine and Coastal Research Laboratory (MCRL) and its central campus, which provide a wide range of watershed relevant measurements. Further details of these capabilities and facilities are included below.

A5.1 Pacific Northwest National Laboratory

PNNL, which is operated by Battelle for the Department of Energy (DOE), is a multi-program national laboratory that conducts a broad scope of research to address significant national problems and to facilitate scientific problem solving and collaboration. *Cross-disciplinary collaboration is a strong cultural feature at PNNL*. PNNL's distinctive focus on integrating computational and experimental capabilities as well as collaborating among disciplines yields a strong and synergistic scientific environment. Tools for environmental measurements and imaging are foundational to research across PNNL's research directorates and in the DOE user facilities on PNNL's campus—EMSL and the Atmospheric Radiation Measurement user facility. PNNL's strong collaborative research environment is facilitated by its deep staff experience across research divisions and its advanced capabilities.

Biogeochemistry and Microbiology Labs

The biogeochemistry laboratories in the 331 Building on PNNL's Richland, WA, campus are well equipped with pH meters, oxygen probes and meters, balances, ovens, shakers, centrifuges, wet sieving machines, fraction collectors, and controlled atmosphere chambers for studies with desired gas compositions. Capabilities include high-performance liquid chromatography (HPLC); gas chromatography (GC); fast protein liquid chromatography infrared gas analyzers; a Picarro G2301 CO_2/CH_4/H_2O Analyzer; capillary electrophoresis; inductively coupled plasma (ICP) emission spectrometers; ICP-MS (ICP linked to a mass spectrometer) for measuring major cations and trace metals; ion chromatographs (IC) and automatic titrators to measure major anions, alkalinity, and dissolved CO_2(g) in aqueous samples; multiple X-ray diffractometers for mineral analyses; a laser diffraction instrument for measuring sediment particle size distributions; elemental analyzers for dissolved particulate organic matter and nitrogen content; Hyprop and W4PC systems (Decagon) to determine water retention curves across a wide range of water tensions; a custom-built, porewater suction system with tunable suction pressures for the simultaneous sampling of 24 soil cores; and various microprobes and sensors for monitoring soil cores. We also have two environmental chambers for which temperature, humidity, and photons can be controlled in 90-minute steps (Conviron). One dedicated lab is outfitted with an MBraun anaerobic environment; an HPLC; the Elementar elemental analysis suite for the quantification of carbon, hydrogen, nitrogen, chloride, and sulfur; and a Vario Isotope cube connected to an IsoPrime isotope ratio mass spectrometer (IRMS) for stable isotope analysis of solid or liquid material.

The environmental microbiology laboratory occupies 1,163 ft^2 in PNNL's Biological Sciences Facility. It boasts a dedicated molecular biology laboratory that is outfitted for DNA isolation, polymerase chain reaction (PCR) amplification, DNA hybridization analysis, synthetic biology, protein purification, DNA sequencing, and related techniques. Equipment includes four PCR thermocyclers, an ABI StepOnePlus and a Bio-Rad 384CFX for real-time quantitative PCR, a Labcyte Echo 550 liquid handler, multiple epMotion liquid handling systems, a fast protein liquid chromatography system (Amersham Pharmacia Biotech) for protein purification, and an Illumina MiSeq desktop sequencer. Also available are a denaturing gradient gel electrophoresis system, controlled temperature water baths and incubators, several high-speed centrifuges, numerous microcentrifuges, an electroporator, and an ultraviolet (UV) crosslinker for DNA hybridization studies. Multiple Bio-Tek spectrometers and plate stackers capable of high-throughput fluorescence, optical density, and luminescent readings, as well as digital photograph systems, are available for image capture.

Stable Isotope Measurement Facilities

PNNL houses a well-equipped stable isotope facility using two Thermo Delta V plus IRMSs for measuring stable hydrogen, carbon, nitrogen, and oxygen isotopes. We maintain a large number of peripheral devices, including elemental analysis (for carbon and nitrogen analysis of solid or liquid samples), thermal conversion elemental analysis (for hydrogen and oxygen analysis of solid or liquid samples), gas bench (for carbon and oxygen analysis of dissolved organics, inorganics, or carbonate phases), GC (for compound-specific carbon, nitrogen, hydrogen, and oxygen analysis of volatile analytes), and dual-inlet sample introduction (for very-high-precision analysis of CO_2, N_2, and H_2 gas samples). The PNNL-developed laser ablation IRMS instrument is housed within the isotope facility. We also have access to cavity ring-down IR spectroscopy capabilities for measuring hydrogen and oxygen isotopes of water. We maintain a water extraction line for removing trace amounts of water from samples (cell pellets or other water-containing samples).

Separation Resources

Thirty-three capillary HPLC systems are available to meet the separation needs of any project based on sample type, quantity, and complexity. They include nine commercial nano-flow liquid chromatography (LC) systems for more routine workflows and 24 custom systems built around very-high-pressure ISCO pumps, Agilent nano-pumps, or a combination of both. These custom LC systems provide the following: unique and sophisticated advanced biological separation capabilities through custom software controls; operating pressures up to 20,000 psi; easily modified valve configurations; and custom-packed LC columns of any separation type (reversed-phase, ion exchange, size-exclusion, etc.) and dimension (down to 10 μm i.d.). The flexibility of the control software and valve configurations allows for rapid reconfiguration and automation to take immediate advantage of new separation techniques.

Metabolic Multi-reactor

The PNNL-developed Metabolic Multi-reactor provides high-throughput laboratory analysis of ex situ sediment samples for rates of O_2 consumption and gas production. Traditional bioreactor methods for these experiments require that either individual O_2 sensors are placed in a glass or plastic tube bioreactor or a technician manually moves part of a sensor system from reactor to reactor. This prevents highly time-resolved measurements of O_2 concentrations and restricts overall throughput. The Metabolic Multi-reactor consists of many individual reactors on a set of rollers, with optical O_2 sensors embedded in the individual tubes. All tubes on a roller rack can be interrogated at the same time by a single imaging system, providing high temporal resolution. Image processing software allows the system to automatically parse, process, and apply calibration factors to the individual tubes visible in each image, enabling high throughput. The system can be expanded by racking roller sets, allowing large numbers (100+) of tubes to be monitored concurrently. In addition to dissolved O_2, headspace gas, aqueous samples, and incubated sediments can be sampled and/or preserved without opening the Multi-reactor system. Development is underway to provide high-temporal-resolution optical pH and pCO_2 measurements as well.

Planar Optode Flow-Through Column Capability

The acrylic flow-through column allows for nondestructive dissolved oxygen measurement at high temporal and spatial resolutions using a novel application of planar optode sensors with [125]. The PNNL-developed design can be constructed to any length (within logistical constraints). The current column has an optical dissolved oxygen sensing dye system coated within one longitudinal half of the 7.6 cm diameter, 30 cm long column. The dye is imaged with a modified digital camera. The other longitudinal half of the column is equipped with sampling ports. The information provided by the spatially contiguous column optode, along with spatially distributed aqueous sample ports, allows for coupling respiration rates (changes in dissolved oxygen concentration) with changes in aqueous chemistry (including metabolomics) as solutes move through sediments packed in the column (i.e., sampling across different residence times and flow path lengths). This capability generates a temporal and spatial understanding of the mechanisms controlling biogeochemical rates. It is a tool to understand how ecosystems respond to disturbances by experimentally modifying injection flowrates and the concentration and composition of injected solutes (e.g., manipulation of DOM chemistry, dissolved oxygen, and NO_3^-). Flow-through in-line meters (e.g., conductivity, pH, and dissolved oxygen) can also be added to the outflow of the column to provide additional information about solute transport and responses to disturbance.

Marine and Coastal Research Laboratory

At approximately 100 acres, the MCRL contains indoor and outdoor facilities specifically designed for aquatic research. Aquatic laboratories (15,000 ft^2) are equipped with through-flowing seawater that can be heated, cooled, or diluted with freshwater to reflect water conditions from the tropics to the arctic. Experimental tank systems inside and outside the facility include flow- and wave-tanks and full-spectrum, computer-controlled lighting capabilities.

The organic chemistry labs at MCRL are well equipped for a variety of organic geochemical analyses. Equipment includes: a CHONS elemental analyzer; a full UV-Vis spectrum Aqualog capable of simultaneous absorbance/fluorescence detection; a programmable, ramping oxidation oven; a Dionex accelerated solvent extractor (ASE) for extraction of diverse matrices (soils, sediments, tissues, aerosol filters, solid-phase extraction disks, etc.); two Agilent GC-MS systems; one Agilent GC-ECD; two Agilent HPLC systems with diode-array, fluorescence, and refractive index detectors; an Agilent triple quadropole LC-MS with in-line diode-array and ELSD detectors; a Thermo-Scientific orbitrap LC-MS; and four high-sensitivity total organic carbon analyzers for dissolved TC/DOC/DIC/TDN analyses, among other equipment.

The inorganic chemistry labs at MCRL are located in Class-100 clean rooms with Class-100 laminar-flow benches for handling and analysis of ultratrace-level metals and metalloids. Equipment includes: a Dionex dual-channel ion chromatograph with two conductivity detectors and a multichannel variable wavelength detector for determination of anions and cations in both fresh and seawater; a Perkin-Elmer Optima 7300 inductively coupled plasma optical emission spectrometer (ICP-OES); a Thermo Elemental iCAP Q inductively coupled plasma mass spectrometer (ICP-MS) with SEAFAST sample introduction system for seawater samples; acid digestion capabilities for preparation of any solid or liquid matrices for major and trace elemental analyses; and a Mettler Toledo Excellence T7 automated titrator for high-throughput analyses of pH, alkalinity, and conductivity.

The gas exchange and in situ observations lab is equipped with: a Picarro G2508 CO_2, CH_4, N_2O, NH_3, H_2O Analyzer; a Los Gatos Methane Carbon Isotope Analyzer; a Los Gatos Carbon Dioxide Isotope Analyzer; a Los Gatos Water Isotope Analyzer; a portable Licor 7810 CO_2, CH_4, H_2O Analyzer with integrated Smart Chamber and soil moisture/temperature probe; two portable Licor 850 CO_2, H_2O Analyzers; three Teledyne Isco Autosamplers; two YSI Pro Plus meters with DO, pH, ORP, conductivity, and nitrate sensors; custom flux chambers for soil and tree stems; custom headspace equilibration chambers for analysis of dissolved gases; six Meter Pario soil particle size analyzers; three Meter Hyprop 2 precision tensiometers, among other equipment both in the lab and currently deployed in the field.


Figure A5.1. Cluster computing facility available at PNNL.

Computational Sciences Facility, Computer and Computational Resource

The Leadership in Energy and Environmental Design-certified CSF has a total of 12,500 ft^2 of raised floor space and can take advantage of geothermal heating and cooling technologies. In addition to recently updated HPC clusters, CSF will house the next-generation of computing resources in support of research in cybersecurity, information analytics, and bioinformatics, as well as HPC for extracting knowledge from massive amounts of heterogeneous data (**Fig. A5.1**).

PNNL has extensive computational capabilities and tools that are available to support all aspects of the proposed research including real-time processing, data-integration, and long-term storage. In addition to the general-purpose networking and desktop computing infrastructure at PNNL that are described later in this section, a range of advanced computer architectures are available. PNNL has multiple significant computer centers available to researchers on this proposed effort. These computer centers provide appropriate levels of network, power, uninterruptible power supply, fire detection and suppression, and physical security. Computer systems in these centers are monitored and managed by a team of system administrators who specialize in running computer systems that support scientific and computer science research.

Research Computing Capabilities. Research Computing is a laboratory-level investment in computational hardware, software, and consulting resources (**Fig. A5.2**). PNNL has been facilitating multidisciplinary collaborative scientific research via Research Computing since 2011. PNNL's supercomputer, Constance, is housed in CSF and is the successor to the original Research Computing base capability, Olympus. In addition, Research Computing supports long-term 5 petabytes (PB) of storage for projects. Projects also have the ability to use Research Computing resources through a consumption model in which the project is charged a competitive rate for cycles and storage as used. PNNL recently activated our next HPC platform, Deception. The initial configuration for Deception consists of 96 compute nodes boasting a total of 6,144 central processing unit (CPU) cores powered by AMD EPYC 7502 CPUs. The system has been designed to be easily scalable up to 330 compute nodes and 20,000 compute cores. Research Computing also has a dedicated graphical processing unit (GPU) computing cluster, Marianas, which is based on Intel Broadwell CPUs and NVIDIA P100 GPUs. With the GPU accelerators, Marianas is designed for deep learning/ML-type workloads. The cluster resources include 25 compute nodes with Dual NVIDIA P100 12GB PCI-e based GPUs, 6 nodes with Dual NVIDIA V100 16GB PCI-e based GPUs, 12 nodes with 8 Nvidia RTX 2080 Ti 11GB GPUs, 1 NVIDIA DGX-2 with 16x Nvidia V100, and 5 DGX-A100 systems with 8X Nvidia A100.

PNNL Scientific Data Management Capabilities. The PNNL DataHub is designed to help researchers address the full data lifecycle for their institutional projects and provides a path to creating findable, accessible, interoperable, and reusable (FAIR) data products. Open science data are a crucial focus of the core services of PNNL's DataHub, which offers state-of-the-art capabilities including (1) a performance

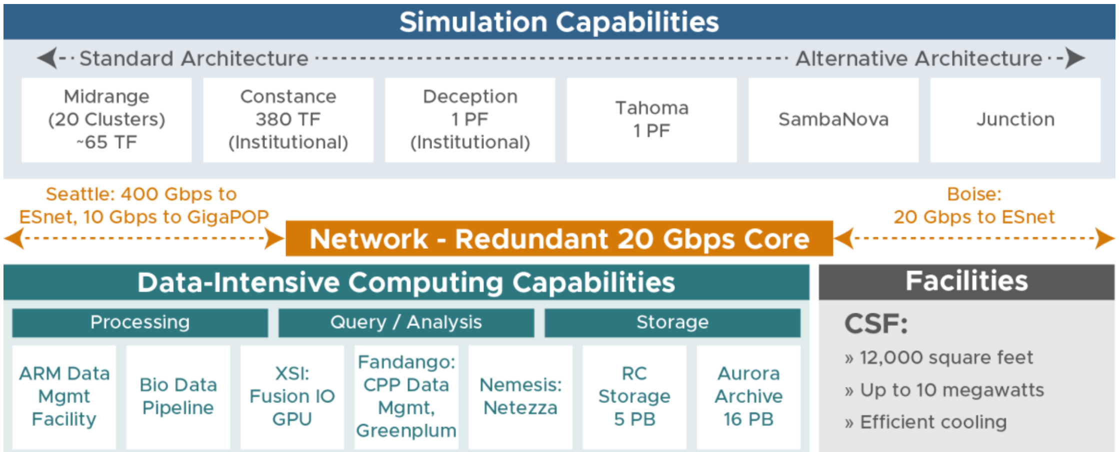


Figure A5.2. PNNL's Research Computing delivers computationally enabled science by providing researchers access to resources such as HPC, GPU-accelerated computation, software platforms, research data and services, and cloud computing. Notes—PB = petabytes; PF = petaFLOPS; RC = PNNL Research Computing; TF = teraFLOPS.

data catalog with metadata search and data (at rest) retrieval, (2) the ability to create or store custom metadata descriptions, (3) data access authorization for restricted or embargoed data, (4) data workflow and data repository integrations, (5) data standards consultations, (6) project data management and consultation across collaborators and sponsor, (7) globally unique persistent identifier generation (specifically digital object identifiers [DOIs]), (8) publication hosting, and (9) de-identified data workflows.

Access to External High-Bandwidth Communication. High-bandwidth, fully redundant internet connectivity is provided by PNNL's existing two 400 Gbps connections to the DOE ESnet in Boise, Idaho, and Seattle, Washington, as well as to the Pacific Northwest GigaPOP core hub in Seattle. A data transfer node provides a high-speed file transfer capability to and from PNNL's 5 PB institutional computing storage capability. PNNL uses the Northwest Open Access Network to connect the Richland campus with PNNL facilities in Sequim, Washington, and Portland, Oregon.

Office space for the PI, co-PIs, collaborators, staff, postdoctoral fellows, and undergraduate interns is located at PNNL and we also support a hybrid work environment, allowing staff to work from their home. Administrative, document preparation, technical artists, and other support capabilities (e.g., graphics and photography) are available at PNNL and will be used on an as-needed basis to prepare progress reports, manuscripts, posters, and presentations.

PNNL provides modern laboratory and office spaces, including standardized conferencing capabilities for video- and teleconferencing (using Microsoft Teams and other desktop collaboration tools). Campus conference spaces feature high-definition web cameras and audio devices, large video displays, and built-in table electrical connections. Virtual collaboration is a distinguishing strength at PNNL, and we have elaborated a vision to build on that strength during the award period. In response to the pandemic and to ongoing trends in online collaboration, PNNL is continuously deploying and refining new collaboration tools to enable robust scientific communication.

Research staff members also use laptops, tablets, or desktop computers that can be connected to the PNNL network for remote or on-campus internet access, file sharing, data backup, email, and printing. PNNL staff network can connect to a laboratory-wide information network, databases, internal collaboration software (Confluence), and/or other institutional computer resources (as appropriate per Classified Matter Protection and Control guidelines). In addition, all-in-one voice systems provide options for phone calls, video chat, instant messaging, or desktop sharing.

A5.2 Access to DOE User Facilities

The Environmental Molecular Sciences Laboratory

EMSL is a 200,000 ft^2 multidisciplinary national scientific user facility funded by DOE's BER Program at located at PNNL-Richland. EMSL provides integrated experimental and computational resources for discovery and technological innovation in the environmental, molecular, and energy sciences to support the needs of DOE and the nation. EMSL is known for integrating computational and experimental capabilities across multiple scientific disciplines to create synergistic research environments. EMSL capabilities that will be accessed for this project include those associated with state-of-the-art MS and nuclear magnetic resonance spectroscopy instrumentation, flow cytometry, equipment for transcriptional profiling, characterization of sediment physical properties, experimental flow cells, and high-performance computing. All scientific instruments are accessed via user proposals submitted by project PIs on an as-needed basis, using EMSL's year-round submittal and approval process for user-executed projects.

Figure A5.3 EMSL boasts the largest MS and nuclear magnetic resonance capability for molecular measurements among the national laboratories and is a world-class facility. *This capability will be critical for obtaining the molecular-level* understanding of biological and chemical interactions within the coastal zone.

Mass Spectrometer Resources. EMSL offers a variety of MS-based approaches for characterizing soil, microbial, and water samples (**Fig. A5.3**), including the following:

- *Panomics* – Advanced global proteomics, metabolomics, glycomics, and activity-based omics research using cutting-edge tools, including customized hardware and sophisticated bioinformatics tools.
- *Natural Organic Matter* – Several workflows targeting different classes of organic compounds in soil and the environment.

Specific instrumentation includes:

- **One 12-Tesla high-performance FTICR-MS.** Bruker Solarix MS with ETD capabilities and sub-ppm mass accuracy, high-resolving power, wide bandwidth detection, high dynamic range, and tandem MS capabilities beneficial for characterizing complex mixtures such as natural organic matter. This instrument is coupled to an automated direct-infusion liquid chromatograph that allows a high throughput of samples (~50/day).
- **One 21-Tesla high-performance FTICR-MS.** This capability combines the highest magnetic field strength available for MS with advanced technologies such as UV photodissociation (UVPD). These advances were specifically designed to address analytical challenges in metabolomics, natural organic matter characterization, top-down proteomics, and other intact protein measurements. This system is equipped with MALDI-2 and LAESI sources.
- **Seven ThermoScientific Q Exactive hybrid quadrupole-Orbitrap MS.** These benchtop LC-MS/MS systems combine quadruple precursor ion selection with high-resolution, accurate-mass (HR/AM) Orbitrap detection to deliver exceptional performance and versatility for both global LC-MS/MS and

targeted quantification by applying parallel reaction monitoring. In addition, one of the Q Exactive instruments is implemented with a 193 nm UVPD Coherent Existar XS ArF excimer laser, providing an alternative fragmentation method as needed for the analysis of specific post-translational modifications.

- **Five ThermoScientific LTQ-Orbitrap Velos hybrid MS**. These instruments combine the proven mass accuracy and ultrahigh resolution of the Orbitrap mass analyzer with the increased sensitivity and improved cycle time of the LTQ Velos. Each instrument has a dual-pressure linear trap, HCD collision cell, and multiple fragmentation techniques, including ETD, for optimal structural characterization.
- **Two Agilent 5975C GC/MSD gas chromatography-mass selective detector systems.** These mass spectrometers operate in electron ionization mode and are coupled with Agilent high-performance 7890A GCs. These systems are routinely used for untargeted metabolomics measurements and can also be operated in SRM mode for targeted compound analyses. The National Institute of Standards and Technology MS library and Agilent Fiehn Metabolomics Retention Time Locked Library are used with these instruments to facilitate compound identification.
- **Two Thermo Delta V plus IRMS instruments.** These instruments measure stable carbon, nitrogen, hydrogen, and oxygen isotopes. We maintain a large number of peripheral devices, including elemental analysis (for carbon and nitrogen analysis of solid or liquid samples), thermal conversion elemental analysis (for hydrogen and oxygen analysis of solid or liquid samples), gas bench (for carbon and oxygen analysis of dissolved organics, inorganics, or carbonate phases), GC (for compound-specific carbon, nitrogen, hydrogen, and oxygen analysis of volatile analytes), and dual-inlet sample introduction (for very-high-precision analysis of CO_2, N_2, and H_2 gas samples).
- **Instrument support and maintenance**. All the above-listed spectrometers are on a common computer network with direct high-speed access to the EMSL data storage and archive facility, a multi-terabyte storage system that allows large numbers of data files to be archived indefinitely. Researchers have full access to both the resources and support personnel in the Instrument Design Laboratory to support instrument development and repair.

Nuclear Magnetic Resonance Resources (EMSL). The EMSL user facility houses 11 instruments capable of performing liquid- and solid-state nuclear magnetic resonance (NMR) experiments in support of environmental and biological research. Three instruments are exclusively dedicated to high-resolution liquids NMR metabolomics and are optimized for environmentally derived microbiome small-molecule characterization. These dedicated instruments are one Bruker 750 MHz NMR with an auto-sample changer and two Agilent/Varian 600 MHz NMRs. All three have cryogenically cooled 5 mm triple-resonance probes, allowing for the high sensitivity and resolution needed for microbiome extracts. A new Bruker 800 MHz system with an auto-sample changer and equipped with a state-of-the-art, cryogenically cooled 5 mm quadruple-resonance probe (HPCN) allows the highest-sensitivity 1H detection paired with P/C/N decoupling. This probe will enable enhanced detection of small organic molecules, including phosphorylated metabolites. The associated metabolomics data are typically screened with Chenomx software using established methods and an enhanced custom database that currently represents ~1,000 metabolites. NMR researchers have two 32-node Linux clusters available for NMR data processing and structure calculation. In addition to the open-source NMR software, specialized commercial software for data processing and analysis (Felix, MestReNova) is available. This equipment is expected to be used for screening aqueous-based environmental samples and LC fractions. It is fully accessible to EMSL researchers. Solid-state NMR capabilities include 850 MHz, 750 MHz, 600 MHz, 500 MHz, and 300 MHz instruments and a leading-edge, novel magic-angle-spinning rotor technology to allow in situ studies at precisely controlled sample environments, including at high temperatures and pressures. All of the equipment is on a common computer network with direct high-speed access to the PNNL data storage and archive facility, a multi-terabyte storage system that allows large quantities of data files to be archived indefinitely.

Mass Spectrometer Resources. The computing and data operations are located within EMSL and are supported by EMSL's project and operations funding from DOE-BER. EMSL's distinctive focus on integrating computational and experimental capabilities, as well as collaborating among disciplines, provides for a major synergistic scientific research environment. The computing and data operations provide an integrated production computing environment that supports a range of computational activities in environmental molecular research, archive storage, scientific expertise, and the Molecular Science Software Suite (including NWChem and the Global Arrays Toolkit).

Systems include Tahoma, an advanced scientific computing cluster with 184 Intel "Cascade Lake"™ nodes with a clock speed of 3.1 GHz, including 24 analytical nodes with GPUs.

The Tahoma system has an aggregate of 98 terabytes (TB) of memory, 12 PB of global storage in a BeeGFS file system, and an aggregate of 1.2 PB of local disk. Its peak performance is 1,050 teraflops. Computing and data operations also operate a 77 PB hierarchical archive storage system, Aurora.

Appendix 6 – Data Management Plan

This appendix describes strategies the RC SFA will use for management of observed and model-generated (meta)data, including data generation, standards, storage, and sharing. The PI has ultimate oversight and decision-making over project data management and is supported by the Data Management Cross-Cut Lead and Deputy (**C.4**). Our data management plan (DMP) is intended to coordinate storage, preservation, and (meta)data capture in a timely and user-friendly way to facilitate user engagement, and ultimately to enhance the reproducibility and reusability of our scientific results to maximize impact. By doing so, we will not only comply with DOE's Office of Science (DOE-SC's) requirements on digital data management, but also build a foundation to facilitate model–data integration needed to accelerate our predictive understanding. Our DMP draws upon lessons learned from our previous work as well as best practices established by the broader community. We will continue to implement the FAIR (findable, accessible, interoperable, reusable) data management and stewardship principles [112] in alignment with ICON (integrated, coordinated, open, networked) science.

A6.1 Data Collected, Generated, or Used

This project will generate two broad categories of data: (1) observational and experimental laboratory and field data; and (2) simulation data generated from our multiscale modeling activities. Field deployable sensors include but are not limited to distributed monitoring data of water level, specific conductance, temperature, dissolved oxygen, and soil moisture. Soil and river corridor water/sediment samples collected will generate multiple data types, including organic matter characterization and microbial data. If any observational social data focused on community and regional partner priorities are collected that result in any personally identifying information (PII) related to Human Subject Research (HSR), HSR guidelines will be followed as appropriate, including Institutional Review Board (IRB) approval. The project will leverage external observational data including state and federal water quality monitoring and water monitoring by USGS as well as those from other observational networks (e.g., LTER, CZN). We will leverage existing NASA, USGS, and ESA satellites to produce estimates of land surface properties. Our modeling activities heavily rely on observational and experimental data to parameterize, calibrate, and validate numerical models. Most of our model simulations are performed on DOE computing facilities and DOE/PNNL supercomputers. While the raw model outputs can be massive, they are typically post-processed to produce results needed to address science questions and generate plots to be used in journal publications. There are multiple value-adding steps that can reduce the dataset size. Raw data will undergo QA/QC appropriate for the data type prior to analysis and prior to publishing. QA/QC documentation is required as one type of metadata when datasets are disseminated and published. Published (meta)data and datasets will use persistent, unique identifiers for samples collected (i.e., international generic sample numbers (IGSNs)) in accordance with ESS-DIVE Community Reporting Formats and will use persistent identifiers for published datasets (i.e., digital object identifiers (DOIs)) generated by data repositories.

A6.2 Data Standards

Datasets here are defined holistically as raw (meta)data, derived (meta)data products, QA/QC documentation, and algorithms and settings, including model output. Datasets published on ESS-DIVE will use existing, open-source, standard data formats (e.g., csv, netcdf) and follow standards in metadata and QC procedures whenever possible, including ESS-DIVE Community Data Reporting Formats. (Meta)data formats will align with the formats anticipated for publishing as soon as possible after data generation to facilitate standardized data ingestion for analysis and modeling. Model output will be stored on the Earth System Grid Federation (ESGF) nodes at Lawrence Livermore National Laboratory (LLNL) as needed due to size considerations on ESS-DIVE, will adhere to guidance provided for the Coupled Model Intercomparison Project phase 6[4], and output files will be in netCDF4 format. We will work closely with other BER-funded data repositories/facilities to follow their existing formats and when needed identify other (meta)data standards or reporting formats. Where no standard or reporting format exists, we will review best practices to generate consistent outputs. We will use standard methods developed by AmeriFlux for flux towers, by WHONDRS for water and sediment sampling, and by NEON for soil sampling. We will use methods consistent with past RC SFA studies when possible to enable synthesis.

A6.3 Related Tools, Software and/or Code

We will use the ESGF and ESS-DIVE open source RESTful API to enhance model output access and dataset upload, respectively. We will continue to use tools developed in the prior quadrennial period of the SFA, including R code to automate use of ESS-DIVE Community Reporting Formats and extensive team guidance documents for navigating dataset generation. Our team will continue to make heavy use of Jupyter notebooks to document computational workflows including data sources, codes used, and all associated metadata. Git products' control system will be used to store and share code, scripts, algorithms, and software used to analyze experimental data and produce value-added datasets with the research team.

A6.4 Data Sharing

All project data except HSR-limited data will continue to be organized into datasets and published openly on ESS-DIVE under a Creative Commons Attribution 4.0 International License. Datasets may include references to external public data or previously published project datasets. For each project manuscript, we will continue to require that a dataset be published on ESS-DIVE prior to manuscript publication and contain all necessary (meta)data needed to reproduce SFA-associated manuscript results (i.e., "manuscript data packages"). We will also continue to publish raw (meta)data associated with discrete field and/or laboratory efforts to facilitate wider data reuse (i.e., "study data packages") as soon after data generation and QAQC as possible, striving for release within one year of generation, and often achieving release prior to use of the data for project analyses. Data generated in collaboration with DOE user facilities (e.g., JGI and EMSL) will be made available by those facilities in accordance with their data management and release policies. ESS-DIVE datasets will indicate connections to other user facilities. Data stored in repositories other than ESS-DIVE will have metadata mirrored on ESS-DIVE to support their goal of a unified, searchable ESS database. Data users are anticipated to continue to be SFA team members, known close collaborators, the broad open WHONDRS community, and users of opportunity. SFA team members will continue to be strongly encouraged to use published datasets, rather than using early or incomplete data products or transmitting local data copies.

A6.5 Data Preservation

(Meta)data generated during our research will be captured and stored in the appropriate public data repository (primarily ESS-DIVE; if needed ESGF) except HSR-limited data (**Section A6.4** Data Sharing). All raw and processed (meta)data will first be stored on PNNL institutional computing and, when needed due to size, National Energy Research Scientific Computing (NERSC) High Performance Storage System (HPSS), before being uploaded to ESS-DIVE. For laboratory and field observational data, including experimental and sensor-based data, we will preserve raw and processed (meta)data. We will store and share the algorithms and software used to analyze data and produce value-added datasets. For model-generated data, we will capture all necessary information to reproduce published results, which include but are not limited to all model input files, specific simulator change sets or versions, computing environment settings, and all algorithms needed to produce the results used to generate tables or figures from raw model outputs.

A6.6 Data Protection: Security and Integrity

We will apply the standard PNNL protocols for release of information; therefore, data that need to be protected for security, economic, or business reasons will have been protected (e.g., via patents or provisional patents) before they are disseminated. Our data management system is designed to protect confidentiality, personal privacy, PII, and information that is important to U.S. national, homeland, and economic security. Our hardware and software systems leverage the firewalls and authentication systems of PNNL so that only authorized personnel have access to the data. The data management system itself allows assignment of access privileges at multiple levels, including visitors, user-defined groups, project members, institutions, and individual people. Data stored locally are regularly backed up and restorable in case of a hardware failure. All public access is write-protected, preventing modification of published data.

A6.7 Oversight of Data Management

The Principal Investigator (PI) has ultimate oversight and decision-making over project data management. The Data Management Cross-Cut Lead and Deputy will be the primary points of contact and oversight for all data generation and data publishing across the project. The Data Management Cross-Cut Lead's shared role as ICON Steward will be a support to assure alignment with ICON principles C (coordinated) and O (open). Each team generating data will inform data management strategy with input from the Data Management Cross-Cut Lead and Deputy, and teams must align with project approach, including FAIR data stewardship principles and ICON principles. The PI, co-PIs, and Data Management Cross-Cut Lead and Deputy may update the DMP after discussion and consensus. Intellectual property will follow requirements from institutions and funding agencies.

A6.8 Rationale

Data reproducibility and reuse relies on well-structured and enforced data management strategies. Our approach prioritizes FAIR data in order to maximize the benefit of the data we generate to our team, our collaborators, other scientists, and interested parties. This aligns with ICON principles, extends the reach of the project, and maximizes the value of the resources used. Limitations of our DMP are informed by privacy considerations of HSR and PII.

Appendix 7 – Software Productivity, Sustainability, and Improvement Plan

A Software Productivity and Sustainability Improvement Plan (S-PSIP) is critical for our multidisciplinary, multi-institutional RC SFA team to accelerate scientific discovery while ensuring the longevity of the proposed code developments. We define the productivity (feasibility, sustainability, and verifiability) activities and identify some of the documentation that will be created and maintained during the entire software lifecycle. Our S-PSIP draws upon lessons learned from our previous and ongoing projects (e.g., E3SM, COMPASS, and IDEAS Watersheds) as well as the best practices established by the broader scientific community (e.g., Better Scientific Software [BSSs]) and open-source software development community. The application and development of software packages in this project can be classified into three categories:

1. Application and development of existing software such as ATS, PFLOTRAN, ELM, FATES-HYDRO, and SWAT.

2. Development of machine learning models and hybrid AI-physics models.

3. Development of new software modules including codes for workflows and data analysis.

We now describe the key elements of the S-PSIP of the project.

A7.1 Software Availability

All software development of existing and new codes will use an open-source, open-development strategy. We will ensure that the code repositories for all codes generated by our project will be made publicly available, making it easy to get the latest from the master branch, or a tagged release can be provided to comply with new software requirements imposed by many peer-reviewed journals. We will ensure that their licenses are compatible so that distribution, integration, and use of the code is not hindered. New software will be released under a compatible open-source license, such as the three-clause BSD open-source license. If there is more than one contributing laboratory or institution, they will jointly assert copyright and agree to use the same open-source license. We will ensure that new development of existing software is compatible with their existing software licenses. Our open-development strategy, which is extensively used by project team members in multiple ongoing projects, will make the latest version of the code available to the project team, our research collaborators, and the broader scientific community. We will make a versioned release of our code when significant new features have been added since the last release. Code examples will be included in software releases to provide a demonstration of new code features. Git tags will be used to archive the software version used in a publication and documentation on how to reproduce the results reported in the publication will be made available on an open access code repository (e.g., Bitbucket GitHub, GitLab, etc.). To make pre-built versions of the software ecosystem readily accessible, we will explore using containerized builds (e.g., Docker).

A7.2 Software Development Process and Tools

Project software will conform to established community standards and apply best practices when implementing new capabilities within existing codes or coupling different codes. Specifically, we will follow the software DevOps doctrines by encouraging collaboration and iteration among developers, testing team, and interested parties throughout the code development lifecycle. Our development team will identify must-have functionality with the project leadership team, POC, and related collaborators, and then iterate over short software development and test cycles until a baseline is achieved. The baseline will be established as open source on GitHub, GitLab or Bitbucket with good documentation, defined software contribution standards, and a modular framework. The software development process will include the following stages:

1. **Feature determination**: Cross-cut leads and POC in consultation with the project team will determine high-priority features for the next software release.

2. **Design document**: Software developers will prepare a design document that will include information such as governing equations of a model, dependency of the new feature on third-party software libraries, and development of an interface for coupling to other software.

3. **Code implementation**: The code will be developed adhering to the approved design document. New code will be developed in a feature branch that will be integrated into the main code branch.

4. **Testing**: A testing suite will be developed to ensure new code development does not break existing capabilities. Software developers will be able to run the test suite on their local machine and the project will also use automated continuous software integration services (e.g., Bitbucket Pipelines, GitHub Actions, GitLab CI/CD, Travis-CI, etc.). The development of each new software feature will use regression and unit testing.

5. **Code integration**: Before the integration of a feature branch in the main code branch, the developer will perform regression testing on at least one of the project-supported platforms to ensure unexpected changes will not be introduced in the main branch. A pull request will be issued by the developer to integrate new code and code reviewers will be assigned. After addressing all reviewers' comments, the new code will be integrated by an assigned code integrator.

6. **Deployment**: The developed software will be verified for the correctness of numerical methods and validated by comparing simulation results against existing benchmarks and observations. The software will be supported in multiple DOE high-performance computing centers that will include the National Energy Research Scientific Computing Center (NERSC) and the Environmental Molecular Sciences Laboratory (EMSL).

7. **Documentation**: A user guide and technical guide will be written for all software developed in this project. This documentation will also be version controlled using Git and the repositories hosting the documentation will use the open-source, open-development strategy. Additionally, the project will develop code style guides for developing new software. The development of existing software will adhere to the code style guide of the existing software. Data analysis scripts developed by the project team will be documented for reproducing the research findings.

8. **Issue tracker**: Requests for new features and reported code bugs will be tracked using the issue tracker feature of cloud code hosting services. Each issue will be labeled to at least one code category, and new development to close issues will be prioritized.

A7.3 Software Developer and User Training

Training of new software developers and users is the key to ensuring software productivity and sustainability. The proposed development of extensive documentation such as a user guide, a technical guide, and a coding style guide will help lower the barrier to using and developing software. We will set up an email list for software developers and users to seek help using the software packages. New software developers will be mentored by one of the senior software developers during their onboarding period. Additionally, the software development team members will attend webinars on software development, such as those organized by the IDEAS-productivity project, BSSw, and various DOE computational facilities.

Turnover of software developers on this five-year project is inevitable, and we will establish a plan to minimize overall disruption to project development. Firstly, detailed development documentation will help other developers navigate ongoing development. Secondly, we will build in redundancy for a few key software development responsibilities, such as having at least two code integrators at all times on the project. Thirdly, once a software developer has notified the PI about their departure from the project, we will schedule a shadowing session between the departing and new software developer to ensure a smooth transition of code development responsibilities.

A7.4 Software Improvement Strategies

The PI, co-PIs, and cross-cut leads will be responsible for evaluating the S-PSIP annually. Input from the entire project team will be gathered regarding the current S-PSIP and necessary improvements will be made accordingly. We will establish connections with the software development team of other DOE-funded projects (e.g., COMPASS and ExaSheds) to exchange and leverage each other's development workflows and tools. To continuously improve our software tools, at the end of each development and test cycle, a retrospective of that cycle will be performed. This will lead to updates in processes, checklists, and acceptance criteria going into the next planning cycle. Routine examination creates a continual improvement practice, enhancing software productivity and increasing the quality of the software produced. Best practices observed in other projects will be adopted. More effective tools will also be adopted as they become available. Developers who practice better productivity and sustainability will naturally be given greater weight in prioritizing future directions for the code. Their names will be publicly recognized through association and listed among developers of associated community codes if the contributions are significant.

Appendix 8 – Institutional Support for PIER Plan

The PIER Plan, as detailed in **Section C.6**, describes SFA project-specific PIER elements. This appendix provides additional detail about the institutional structures that intersect and support the project-specific elements.

A8.1 Procedures for Reporting Workplace Issues or Concerns

Here, reporting applies to workplace issues or concerns ranging from constructive feedback to reportable incidents of harassment or misconduct. RC SFA governance documents (**C.6**) will summarize reporting procedures for PNNL and non-PNNL team members, including providing relevant institutional contact information (i.e., names, emails, phone numbers). There are options for reporting both within and outside of the project team. Institutional resources outside the project team are described below.

PNNL Employees. If PNNL team members would like to report to someone outside of SFA project staff, they can report workplace issues or concerns to their line manager, their Human Resources Strategic Partner, the Employee Concerns Program manager, anonymously to EthicsPoints (a PNNL-contracted independent third-party provider), or anonymously on the DOE Employee Concerns Hotline.

Non-PNNL Employees. If team members from non-PNNL institutions would like to report something to PNNL outside of SFA project staff, they can report to the PI's line manager, the PI's Human Resources Strategic Partner, or the other resources listed above for PNNL employees. They can also report concerns to the institutional contact for the SFA collaboration (Bladon, Hall, or Liu). If they would like to report about an issue within their institution, some resources are described below, but they should follow individual institutional policy. **OSU**: The Office of Equal Opportunity and Access (EOA) documents reports. Individuals can also report incidents through an on-line reporting form. There is a separate reporting process for bias incidents. **UMT**: XXXXX. **WSU**: XXXXX.

A8.2 Institutional Support for Field Safety

Field work conducted at PNNL-permitted field sites, when PNNL staff are present, will follow PNNL institutional field safety and reporting procedures. See the PIER Plan (**C.6**) for information about the project-specific RC SFA supporting field safety guidance.

PNNL Employees. Prior to conducting field work, PNNL employees review and acknowledge the associated Offsite Risk Management Plan (ORMP) and Lab Assist offsite activities, which together provide detailed descriptions of the activity scope, identify mitigations to hazards, contain contact information in the case of emergencies, and initiate necessary trainings. Hard copies of these documents and permits are carried with the field crew every time field work is done. Any site-specific permits, such as those for conducting work on the Yakama Nation Reservation, are kept updated to reflect the current field crew.

Non-PNNL Employees. For field work conducted at PNNL-permitted sites but without PNNL staff present, the field crew will follow their own institutional safety policies. For the flux towers, if the work is done with the funding provided by the SFA but all work and staffing is done by WSU collaborators, then staff will follow safety and reporting procedures for WSU. PNNL field safety policies do not govern participatory science collaborators (i.e., WHONDRS). WHONDRS collaborators sign a consent agreement informed by PNNL's general counsel and follow provided WHONDRS field protocols.

A8.3 Institutional Support for Laboratory Safety

Laboratory work in PNNL labs will follow PNNL institutional lab safety and reporting procedures, including Lab Assist. Lab Assist includes hazards, mitigations, procedures, and identified trainings. Each individual worker is listed on individual lab activities that correspond to their specific work scope and the lab space where that work is permitted.

PNNL Employees. PNNL staff that travel to non-PNNL institutions to perform lab work will follow the non-PNNL institution's safety policies. They will also follow PNNL travel, ORMP, and off-site Lab Assist policies.
Non-PNNL Employees. For lab work conducted at non-PNNL institutions by non-PNNL staff, collaborators will follow their own institutional safety policies.

A8.4 Institutional Training Resources

All team members have institutional requirements to take certain trainings at their institutions, which will also benefit their project participation, including annual trainings on responsible conduct of research. Optional institution-specific trainings, including those described below, will also be encouraged. See the PIER Plan (**C.6**) for information about the project-specific RC SFA PIER and Team Science Meetings that may include trainings relevant across institutions.
PNNL Employees. All PNNL SFA staff will be informed about the self-enrollment option for PNNL online trainings: "Diversity, Inclusion, and Belonging Course" and "Managing Bias Course" and encouraged to register. If their line management will not cover labor costs associated with attending these courses, the project will cover the staff time to attend.
Non-PNNL Employees. Non-PNNL staff will also be encouraged to take optional trainings at their institutions. Some resources are described below. **OSU**: trainings from the social justice education initiative on implicit bias, microaggressions, cultural appropriation, and facilitating dialogue. **UMT**: XXXX. **WSU**: XXXX.

A8.5 Institutional Accessibility, Safety, and Inclusion Resources

RC SFA governance documents will include accessibility recommendations, which will identify both specific recommendations and also select institutional resources. Some examples of these institutional resources are linked below. Note that most PNNL links are internally facing.

Resource Topic	PNNL	OSU	UMT	WSU
General accessibility	Employee resource group for diverse abilities	Accessibility		
Affinity Groups	Employee Resource Groups			
Lactation support	Answers about parental leave	Lactation room support		
Supporting transgender staff	Guides for supporting transgender staff			
Bathrooms (all gender/single user)	All gender and single user facilities	Single user bathroom; gender inclusive restrooms		
Interviews and unconscious bias	Interview resources			
Inclusive language	Inclusive language and style guide			
DEIA Learning	Diversity and inclusion resources			
DEIA and EEO Offices	Diversity, equity, inclusion, and accessibility	Office of Institutional Diversity; Office of Equal Opportunity and Access		

A8.6 Office of STEM Education

As described in the PIER Plan (**C.6**), the SFA engages with students across DOE internship, graduate, and postgraduate opportunities (i.e., SULI, CCI, SCGSR, DGRP). The mechanism for this engagement is through the interface of the PNNL Office of STEM Education (OSE). OSE connects prospective student applications with researchers upon researcher request. OSE can also support projects that have prospective students identified in advance of the application process. OSE provides general onboarding support and development experiences for each student cohort. OSE is also responsible for the PNNL STEM Ambassador program. Select researchers interested in outreach, engagement, and communication are trained under the STEM Ambassador program and develop a display specific to their project research. As described in the PIER Plan, the SFA has six trained STEM Ambassadors and an interactive display that we will continue to use as an engagement tool.

A8.7 General institutional commitments to DEIA

PNNL. PNNL takes a proactive approach to promoting inclusive and equitable research by recruiting researchers from diverse and underrepresented backgrounds. PNNL's talent acquisition strategy prioritizes attracting undergraduate students, graduate students, and early-stage investigators from diverse backgrounds and experiences, which enables us to expand the diversity of our staff and bring a wider range of approaches, ideas, and insights to bear on our science mission, leading to more innovative and inclusive research. Our research leaders and project teams routinely leverage PNNL's DEIA support system and institutional strengths in recruiting, training, professional development, and mentoring for the development and implementation of our project PIER Plans described in **Section C.6.**

OSU. Inclusive excellence is the recognition that an organization's success is dependent on, and tied directly to, how well it values, engages and includes the rich diversity of its community members, including its students, faculty, staff, alumni, friends and affiliates. Over the last two years, Oregon State has taken significant steps to prioritize and more rapidly advance the pursuit of inclusive excellence in all that it does. The primary goals, include: (a) Integrate and advance inclusive excellence within all aspects of the university; (b) Improve recruitment of students and employees from underrepresented communities; (c) Create an inclusive university climate to support the retention and success of all students and employees; (d) Provide innovative and transformative learning experiences enabling all students and employees to advance inclusive excellence; (e) Communicate Oregon State's accomplishments, initiatives and innovations as the university advances inclusive excellence. The College of Forestry strategic plan has DEIJA initiatives integrated throughout.

UMT. xxx

WSU. xxx

Appendix 9 –

E. BUDGET

Please see PAMS PDF attachment for PNNL and subaward budget.

F. BUDGET JUSTIFICATION

Please see PAMS PDF attachment for PNNL and subaward budget justification.