MS. EMILY KYKER-SNOWMAN (Orcid ID : 0000-0003-1782-1916)
DR. DANICA LOMBARDOZZI (Orcid ID : 0000-0003-3557-7929)
DR. ELIN M JACOBS (Orcid ID : 0000-0001-5697-7628)
DR. NICHOLAS GREGORY SMITH (Orcid ID : 0000-0001-7048-4387)
DR. WILLIAM R WIEDER (Orcid ID : 0000-0001-7116-1985)

Article type : Report

Increasing the spatial and temporal impact of ecological research: A roadmap for integrating a novel terrestrial process into an Earth system model

Running title: Bringing ecology into Earth system models

Emily Kyker-Snowman, Department of Natural Resources and the Environment, University of New Hampshire, Durham NH 03824, ek2002@wildcats.unh.edu
Danica L. Lombardozzi, Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder CO 80307, dll@ucar.edu
Gordon B. Bonan, Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder CO 80307, bonan@ucar.edu
Susan J. Cheng, Department of Ecology and Evolutionary Biology and Center for Research on Learning and Teaching, University of Michigan, Ann Arbor MI 48104, chengs@umich.edu
Jeffrey S. Dukes, Department of Forestry and Natural Resources, Purdue University, West Lafayette IN 47907; Department of Biological Sciences, Purdue University, West Lafayette IN 47907, jsdukes@purdue.edu
Serita D. Frey, Department of Natural Resources and the Environment, University of New Hampshire, Durham NH 03824, serita.frey@unh.edu

Elin M. Jacobs, Department of Forestry and Natural Resources, Purdue University, West Lafayette IN 47907, ekarlsso@purdue.edu

Risa McNellis, Department of Biological Sciences, Texas Tech University, Lubbock TX 79409, risa.mcnellis@ttu.edu

Joshua M. Rady, Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg VA 24061, jmrady@vt.edu

Nicholas G. Smith, Department of Biological Sciences, Texas Tech University, Lubbock TX 79409, nick.smith@ttu.edu

R. Quinn Thomas, Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg VA 24061, rqthomas@vt.edu

William W. Wieder, Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder CO 80307, Institute of Arctic and Alpine Research, University of Colorado, Boulder, CO 80309, wwieder@ucar.edu

A. Stuart Grandy, Department of Natural Resources and the Environment, University of New Hampshire, Durham NH 03824, stuart.grandy@unh.edu

Correspondence

Emily Kyker-Snowman, Department of Natural Resources and the Environment, University of New Hampshire, Durham NH 03824. Email: ek2002@wildcats.unh.edu

Abstract

Terrestrial ecosystems regulate Earth's climate through water, energy, and biogeochemical transformations. Despite a key role in regulating the Earth system, terrestrial ecology has historically been underrepresented in the Earth system models (ESMs) that are used to understand and project global environmental change. Ecology and Earth system modeling must be integrated for scientists to fully comprehend the role of ecological systems in driving and responding to global change. Ecological insights can improve ESM realism and reduce process uncertainty, while ESMs offer ecologists an opportunity to broadly test ecological theory and increase the impact of their work by scaling concepts through time and space. Despite this mutualism, meaningfully integrating the two remains a persistent challenge, in part because of logistical obstacles in translating processes into mathematical formulas and identifying ways to integrate new theories and code into large, complex model structures. To help overcome this interdisciplinary challenge, we present a framework consisting of a series of interconnected stages for integrating a new ecological process or insight into an ESM. First, we highlight the multiple ways that ecological observations and modeling iteratively strengthen one another, dispelling the illusion that the ecologist's role ends with initial provision of data. Second, we show that many valuable insights, products, and theoretical developments are produced through sustained interdisciplinary collaborations between empiricists and modelers, regardless of eventual inclusion of a process in an ESM. Finally, we provide concrete actions and resources to facilitate learning and collaboration at every stage of data-model integration. This framework will create synergies that will transform our understanding of ecology within the Earth system, ultimately improving our understanding of global environmental change and broadening the impact of ecological research.

Keywords: global ecology, Earth system models, data-model integration, collaborative bridging, modeling across scales, history of models, interdisciplinary workflow

I. The need to integrate ecology and Earth system models

Terrestrial ecosystems are an integral component of the Earth system. They govern the exchange of energy, water, and greenhouse gases between Earth's land surface and atmosphere and provide numerous services for society, including climate regulation and mitigation. For example, terrestrial ecosystems absorb approximately a third of anthropogenic carbon emissions (Friedlingstein et al., 2019), mitigating the impact of these emissions on climate change. They also play an essential role in regulating global water fluxes, from moderating soil water availability to influencing precipitation patterns and evaporative cooling. The physical properties of terrestrial ecosystems, including their surface reflectivity (i.e., albedo) and surface roughness, also help control the amount of energy absorbed and released by the land surface (Bonan, 2008, 2016). Human management of terrestrial ecosystems can change these biosphere-atmosphere interactions, for example by reducing carbon storage through deforestation and increasing greenhouse gas emissions through agricultural fertilization (Lade et al., 2019; Law et al., 2018). Given the importance of terrestrial ecosystems within the Earth system, modern ecological research papers frequently recommend updating existing ESMs to reflect new evidence or ideas about ecology that may have large-scale impacts on climate. This integration, however, has been slow (Fisher & Koven, 2020).

Historically, integration of ecological insights into ESMs has been hampered because of a disconnect between the scientists conducting empirical research and those engaging in modeling work (Fig. 1), a lack of cross-disciplinary training in modeling and empirical skills, and undervaluing of insights derived from modeling and data exercises completed along the way to incorporating an ecological process into an ESM. Although many scientists engage in both empirical and modeling work, the prevailing paradigm for integrating ecology into models tends to separate the tasks involved into the subdisciplines of empirical data collection and model development (Figs. 1, 2). Even when ecologists engage with model development, the models used in ecology often fall short of the global scale of ESMs. While these models generate valuable insights regardless of their ultimate contribution to ESMs, large-scale integrative understanding of global change impacts requires the use of ESMs because of the many interactions within and among the components of the Earth system. For clarity in terminology, we define “Earth system models” as models which represent the interactions among land, atmosphere, ocean, and cryosphere processes and follow the principles of energy and matter conservation. While we focus specifically on including ecology in the terrestrial component of ESMs, our recommendations can apply to similar challenges in other disciplines (e.g., marine ecology and modeling ocean-atmosphere interactions). The land component of ESMs can and should continue to incorporate ecological processes to improve model realism and to better understand the role of ecological processes within the larger Earth system.

Scientists in both empirical and modeling communities are aware of the need for and benefits of collaborating around ESMs. ESM developers understand that ecology plays an important role in controlling terrestrial ecosystems and that ecological insights can generate models that more faithfully represent real systems, both conceptually and in terms of model uncertainty. Ecological processes, for example, can generate amplifying or stabilizing feedbacks that can fundamentally alter climate and when incorporated will change model performance (e.g. nitrogen constraints on CO_2 fertilization of plant NPP changed the magnitude of model-projected future shifts in ecosystem carbon storage (Thornton et al., 2007)). Empiricists, on the other hand, understand the potential large-scale impact of their work and that ESMs can help to realize this impact (Fig. 3). For example, ESMs are useful for expanding the temporal and spatial scale of ecological research beyond the constraints of a particular set of sites or experiments. Additionally, models can be used to explore interactions and feedbacks between ecological and climate factors that might be prohibitively complex to measure directly. Models are an important means for ecologists to explore new concepts and generate insights about complex systems that can lead to testable hypotheses. Finally, models are a means to understand the impact of specific management and policy decisions and help stakeholders to make science-informed decisions.

Despite the mutual benefits that empirical and modeling communities receive from collaborating, obstacles remain to better integrating these communities (Leuzinger & Thomas, 2011; Reed et al., 2015). While most empiricists are adept at developing ecological theory for their specific species or system, translating that theory into a generalized mathematical formula can be challenging without decades of research gathering long-term data over broad scales. Next, empiricists face the formidable task of integrating this mathematical formulation into an ESM. ESMs can exceed millions of lines of code (Danabasoglu et al., 2020), and hunting for the right place to insert new code without breaking the rest of the model can be daunting. Working within the particular computing language or framework of an ESM can also be intimidating without extensive training in computational science and applied mathematics, which university ecology programs typically do not offer. Additionally, the overwhelming complexity and ambiguity of large models can make it difficult, without training, to assess the reliability of model results. Given these obstacles, an empirically-focused ecologist might question whether it is a good use of their time to put in the training and work involved with modeling ecological processes in the Earth system.

Modelers working to integrate ecological processes into ESMs, many of whom have formal ecological training, also face challenges in this partnership. Modelers must strive for parsimony in model development (i.e. avoiding unnecessary model complexity; see Table 1), and balancing this against the push to continuously incorporate more and more ecological detail can be difficult. Incorporating new processes can sometimes increase rather than decrease model uncertainty. Ecological and biological processes are inherently more complex and challenging to quantitatively define than the physical and/or chemical processes that drive most atmospheric or ocean models. As an example, the physiology of stomata does not conform to the principles of fluid dynamics that underpin the atmospheric and ocean components of ESMs. Quantitative ecology is a robust field, but the math of ecology is often defined in units of genes or whole organisms using statistical relationships, rather than the units of matter and energy and process representations that ESMs use, and translating between the two is persistently difficult.

Even when ecology can be quantified in a way that can be incorporated into an ESM, ecological data can be time- and resource-intensive to gather, and model development can be limited by the availability of all the necessary data to drive, tune, or test a new process. Including all ecological processes that impact water, energy, or biogeochemical cycles can lead to models that are overly complex and lack adequate foundations in measured data. Modelers are sometimes reluctant to add a new process without convincing evidence that its impact outweighs the uncertainty it adds to the model. Most ESMs strive to balance ecological realism with excessive complexity, which can lead empiricists to be frustrated with the disconnect between model parameters, processes, and reality. Meanwhile, modelers may grow frustrated and overwhelmed by the abundance of ecological data that "should" but cannot easily be incorporated into models. Resolving the realism-complexity dilemma requires modelers to understand the principles and constraints of researching ecological processes, while empiricists should be more involved in model development and aware of the unique data needed to translate ecological concepts for ESMs.

We address these challenges by providing a clearly defined map of the stages involved in the incorporation of a new ecological idea into an ESM. We seek to pull back the curtain on the complex, multi-scale workflow of coupled model-data-theory development (Fig. 1, 2, 3) and lower the barriers to interdisciplinary collaboration by articulating various phases and considerations along the way (Fig. 4). Below, we discuss the history of incorporating ecology into ESMs to provide context for the characteristics of modern ESMs. We then present our suggested workflow for integrating ecological processes into ESMs (Fig. 4). In this workflow, we describe the iterative procedure of data collection and model development for understanding ecological processes and models at different scales (Fig. 3). We highlight three stages through this workflow and the valuable outcomes at each stage, regardless of whether the endpoint of incorporating an ecological process into an ESM is reached. Finally, we include a list of resources to guide scientists through all the stages of this workflow. These guidelines and the suggested workflow will facilitate stronger connections between empirical and modeling communities, improving ESMs through realistic process representation and increasing the impact of ecological research.

II. History and context for current decision-making in ESM development

For many ecologists, Earth system modeling may seem a distant discipline, but in fact, ecology is already an important part of ESMs. The origin of ESMs is nearly 100 years old. In the early 20th century, an early model of weather forecasting (Richardson, 1922) required knowledge of land surface temperature, surface-absorbed radiation, and exchanges of heat, moisture, and momentum with the atmosphere. As a result, the model acknowledged the role of energy and moisture fluxes from plant canopies, and included rough representations of stomatal conductance and leaf fluxes in its calculations. In the 1960s, modelers expanded their work to the global scale with different labs and centers developing atmospheric general circulation models, which would form the foundation of some of our present-day ESMs (Edwards, 2011). As model development continued, terrestrial vegetation and human modification of the land became recognized as necessary aspects of climate science (Schneider & Dickinson, 1974), and prominent studies identified surface albedo, evapotranspiration, and deforestation as important climate regulators (Charney et al., 1975; Robert E. Dickinson & Henderson-Sellers, 1988; Sagan et al., 1979; Shukla & Mintz, 1982).

In the 1980s, attention turned to representing more than the atmosphere in global models. Models of the land surface, such as the Biosphere-Atmosphere Transfer Scheme (BATS; (R. E. Dickinson, 1986)) and Simple Biosphere model (SiB; Sellers et al., 1986), were developed for coupling with atmosphere models. These models initially focused on the biogeophysical processes of energy, moisture, and momentum fluxes and the associated hydrologic cycle. These models represented vegetation in more detail, including traits such as stomatal conductance, canopy height, leaf area index, and rooting depth. Photosynthesis was also recognized as an essential process to model, initially as a diagnostic (Robert E. Dickinson et al., 1981) and later as a predictor (Sellers et al., 1996) of carbon and water fluxes (Bonan, 1995; Denning et al., 1996). Building upon a history of ecosystem biogeochemical models first conceived during the International Biological Program (IBP) in the 1960s and 1970s, the carbon cycle was subsequently added to ESMs so that atmospheric CO_2 concentration automatically changed over time rather than being manually specified (Cox et al., 2000; Fung et al., 2005). Bioclimatic rules and simplified equations for competition for space were also added to allow vegetation composition and biogeography to change in relation to the simulated climate (Bonan et al., 2003; Foley et al., 1996; Sitch et al., 2003).

The current generation of ESMs now also includes models with nitrogen and phosphorus cycles, wildfires, biogenic volatile organic compound emissions, mineral dust emissions, methane, wetlands, agricultural management, and land use/land cover change (Bonan, 2016). That many ecological and biogeochemical processes are now included in ESMs is a defining feature in the evolution of climate models, which initially focused on the physical system, to today's more comprehensive ESMs that emphasize the interdisciplinary aspects of climate science (Bonan & Doney, 2018). For example, representations of the nitrogen and phosphorus cycles were added to some ESMs because of their role in regulating the carbon cycle (P. E. Thornton et al., 2009; Y. P. Wang et al., 2010; Yang et al., 2014; Zaehle & Friend, 2010). Similarly, more soil biogeochemical models are including direct representations of microbial populations because of their controls on nutrient and carbon cycling (Huang et al., 2021; Kyker-Snowman et al., 2020; K. Wang et al., 2017; Wieder et al., 2018; Wieder, Grandy, et al., 2015). However, many important processes are still absent from ESMs; for example, herbivores are recognized in ecology as important ecosystem drivers, but are not widely included in ESMs.

Conversations about including ecology in models have become increasingly common in the modeling community, particularly as modelers seek to better match model projections with observations. ESMs continue to be modified to include ecology that impacts model calculations of surface fluxes of energy, moisture, carbon, and momentum. What conditions need to be met for a process to be considered for integration into an ESM? The ecological properties and processes that have made their way into ESMs reflect choices by the modeling community about where to focus its efforts, as well as the practical limitations of the modeling work itself. In general, new ecological processes enter an ESM if:

- **The process can (or is hypothesized to) influence climate on large spatiotemporal scales.** Given the effort needed to code and test the addition of an ecological process into an ESM, the impact of this addition needs to be visible on large spatial scales or on long time frames. For example, explicit representations of vegetation were added to ESMs because they had a clear impact on and improved the performance of climate models through regulating water fluxes on long (e.g., decadal) timescales (Robert E. Dickinson, 1984; Robert E. Dickinson & Henderson-Sellers, 1988; Sato et al., 1989; Sellers et al., 1986).

- **The process can be reasonably incorporated into existing model infrastructure.** New ESM developments build on earlier ones, which means there needs to be a clear plan for how to insert the code for the new process into the existing model code. In addition, this linking should be able to occur without major restructuring to the model's existing structure. For example, in order to integrate nitrogen cycling into an ESM, code needed to be developed to link nitrogen fluxes to the physics of the land surface and calculations of carbon fluxes (Bonan & Levis, 2010; Peter E. Thornton et al., 2007).
- **Process understanding and data are available to model the process globally.** The equations representing the process need to be solvable on a three-dimensional global grid (latitude, longitude, height) as well as on short time scales representing the model's timestep for calculations (e.g., 30 minutes). Ideally, any input data required by the new ecological process should be available globally as a gridded product or be calculable using existing variables simulated by the ESM. For example, the TRY database provides data that has been used to create global maps of plant traits that are used as the foundation for plant functional types (Kattge et al., 2011).
- **The mathematics of the process are tractable within the limits of current computing resources.** Computing resources have significantly expanded, allowing more ecological processes to enter models. However, there are still limits to numerical processing power. Processes must be reducible to a mathematical form that does not dramatically increase computing costs of the entire ESM, given that existing ESMs already push the capacity of the world's most powerful supercomputers (Washington et al., 2009). For example, representing biodiversity by modeling a large number of individual plant species or soil microbial taxa would greatly increase computing costs, so simplified representations of plant functional types and soil decomposition are typically used.
- **There is a community of researchers dedicated to developing, testing, and maintaining the process in the model.** Writing the code for a new ecological process is only one part of the process for integrating a new component into an ESM. Once code is written, it needs to be tested with different components of the ESM and under different simulation conditions before the process can be considered as an official addition to the ESM. In addition, the continued longevity of the process in the model requires there to be

one or more researchers continuing to maintain and update the modeled process as new data about the process and new changes to the ESM are made. As such, a community of researchers with the resources to both advocate for the inclusion of the process and support its inclusion in the model long-term is needed.

With the origin of ESMs in the atmospheric and physics communities, it is perhaps not surprising that the incorporation of ecology into ESMs started in these communities. The modeling community has initiated several grassroots efforts to bring more ecologists into ESM work. These efforts range from creating conference workshops and research coordination networks (e.g., (Cheng, 2018; Leuzinger & Thomas, 2011; Rogers et al., 2014) to leading tutorials and short courses to provide training for empiricists and modelers to bridge these subdisciplines (e.g., the CTSM tutorial at NCAR; FluxCourse; Bracco et al., 2015). However, these efforts are limited in the number of people they can reach. Larger, systematic changes in education and training, funding structures, and engagement across communities are critical to shifting the current siloed paradigm. We propose a new practical roadmap for empiricist-modeler collaboration that breaks down traditional disciplinary boundaries and fosters iterative, shared conceptual development.

III. Introducing the practical roadmap for integrating ecology and ESMs

New efforts to close the gap between ecological empiricists and Earth system modelers are growing, but the two communities could still be better integrated. To do so, each community needs to understand the approaches used by the other and work together both to develop the technical advancements needed to expedite data-model integration (e.g., Fer et al., 2021) and to address the social dimensions of collaboration. Focusing only on technical or mathematical aspects of data-model integration can perpetuate barriers through the use of discipline-specific language and dismissal of non-technical obstacles to participation (Bernard & Cooperdock, 2018; Duffy et al., 2021; Morales et al., 2020), which can lead to members feeling excluded and keep disciplines siloed (Marín-Spiotta et al., 2020; Mattheis et al., 2019). In general, effective cross-disciplinary collaboration depends on several key principles that facilitate team dynamics (O′Rourke et al., 2013) and need to be built into the start of a collaboration; namely: respect and trust among all team members, clear communication, common goals, and effective project leadership (Nancarrow et al., 2013). Research shows that clear team communication is essential for optimizing project outcomes (Anderson-Cook et al., 2019; Kuziemsky et al., 2009), as it is the foundation for identifying shared objectives and building interpersonal relationships that are necessary for teams to remain cohesive during times of conflict (Cooley, 1994). Breaking down barriers to interdisciplinary collaboration requires researchers to adopt practices that not only improve their collaboration, but also dismantle the inequitable and exclusionary dimensions of their disciplines (Chaudhary & Berhe, 2020; Duffy et al., 2021; Emery et al., 2021). Additionally, computing tools and frameworks evolve rapidly, and solutions that focus on facilitating collaboration will outlast any particular technological tool. To achieve better integration and collaboration among empirical and modeling communities, we outline a few necessary foundational principles of collaboration and educational change (Fig. 2). We also propose a workflow that highlights one possible pathway to improve collaboration between fields to improve the work of each (Fig. 4).

In addition to strengthening empiricist-modeler team dynamics, we emphasize the need to rethink ecological education to incorporate process modeling concepts and normalize regular collaboration between empirical and modeling subdisciplines. At many institutions, the ecology curriculum emphasizes field techniques and statistical analysis, but fewer options may exist for courses on ecological process-based modeling. While a given department may offer one or a few courses, often these are not required in ecological education, and programming skills development is limited to high-level statistics programs and languages like R and python that do not entirely prepare students for the computer science that powers modern ESMs. Conversely, educational requirements in other disciplines, such as atmospheric sciences, frequently include both field and modeling techniques and in-depth quantitative and programming skills in which computational science and applied mathematics are essential tools of the science. Ecologists wanting to learn modeling techniques often find themselves taking classes outside their discipline, attempting to separate content from technique and applying techniques to a different field, which is a challenging task. This can pose a large enough burden on the student that many do not follow through, finding it easier to continue with familiar skills. A detailed plan for modifying the way ecology programs teach quantitative skills is beyond the scope of this paper, but others have begun the difficult work of rethinking educational paradigms to address this problem (Hampton et al., 2017).

ESM communities also need to identify opportunities for redesigning their training so they can learn more about ecological concepts and data collection frameworks. Ecological data is complex and filled with caveats, and modelers often encounter data after it has been processed and organized and thus may be unfamiliar with the nuances of data collection and analysis. Modeler training in ecological concepts could take place at the student level, with classwork focused on the impacts of living organisms on biogeochemical, water and energy cycles, or at later career stages via field site visits, shared seminars, interdisciplinary conference sessions, etc. One powerful approach is for a modeler to take a day trip with an ecologist to engage in fieldwork. While we recognize that the outdoors are not a comfortable space for many people and this can be a barrier to participation (Anadu et al., 2020; Giles et al., 2020; Morales et al., 2020), direct experience with how an ecologist gathers data can be an invaluable insight into the the limitations and interpretation of data in a modeled context. Virtual site visits using recorded video are another alternative for those unable to visit in person.

Beyond these foundational shifts, we propose a new workflow for modeler-empiricist collaboration with three specific stages (Fig. 4). This workflow is meant as one (but not the only) route for any empiricist or modeler to understand the stages involved in integrating a new process or idea into an ESM. We strive to break down traditional disciplinary barriers between modelers and empiricists and highlight the iterative collaboration and shared skill sets that are necessary at each stage. The first stage in this workflow ("Assess process & potential impact") includes a list of questions that anyone (regardless of programming ability) can ask to assess the readiness of a process for incorporation into an ESM. The second stage ("Test process alone") involves the quantification and scaling of the new ecological concept using simple models and large-scale parameter determination. Finally, the last stage of the flowchart ("Test process with ESM") discusses the multiple steps involved in making modifications to an ESM, evaluating the impact of the new process on model-wide behavior, and projecting the large-scale impact of the new process within the Earth system. Importantly, each stage of this workflow generates valuable scientific products (e.g. hypotheses, new or improved theory, regional or ecosystem-scale models), regardless of whether the endpoint of "inclusion in an ESM" is reached. We recognize that tackling any part of this workflow is challenging for aspiring and seasoned modelers alike, and we encourage researchers to see it through. We include specific illustrative examples for each stage of the workflow (Boxes 1-3) and one that illustrates stepping through the entire workflow (Box 4), as well as resources for accomplishing each step (Table 1).

Workflow part 1: Identifying and understanding a new process

The first stage of the proposed workflow assesses the readiness of a new process for inclusion in an ESM based on how well the process can be quantified and understood in an ecosystem context. Many empiricists recognize the importance of their work for understanding global change and highlight the need to incorporate new processes into models. However, highlighting this need has minimal impact on ESMs unless coupled to an understanding of the stages of model development and the unique types of data necessary to progress through those stages. As such, the first part of the workflow provides three guiding questions empiricists should ask to assess whether a new process is ready for inclusion in an ESM, each of which will be discussed in more detail in the following paragraphs (Fig. 4, "Assess process & potential impact"). These questions can help identify data gaps and point to valuable targets for future experiments to facilitate downstream ESM integration. Importantly, these questions can be addressed by any empiricist without requiring formal modeling skills. While connecting with modelers is not required at this point, it can be helpful in co-designing future experiments to make process integration more streamlined (Fig. 2).

The first guiding question aims to evaluate the level of theoretical/empirical understanding of the targeted process: *Do you expect your process to respond consistently to environmental drivers, enabling scaling across space and time?* Consistent, quantified patterns are the heart of process modeling. Detailed understanding of a process or mechanism at a single location can help to identify whether the process is likely to scale. In order to develop a broad theoretical representation of a process, it can help to determine whether data are available across multiple sites and ecosystem types and at various timescales. For example, if a specific tropical soil owes its high carbon storage capacity to a unique volcanic mineral (Torn et al., 1997), it would be wise to evaluate the carbon storage capacity of soils without this mineral before generalizing observed patterns to a global scale. While it is not necessary at this stage to gather enough data to create a fully quantified global representation of a process, information gained in this step may help identify data gaps and guide the design of additional empirical experiments needed for large-scale modeling, such as repeating experiments across underexplored regions or a wider range of environmental conditions. This step also helps to identify conceptual areas where a large amount of data may be available but consistent relationships with environmental factors and process rates have not yet been identified. For instance, soil microbial biodiversity is being rapidly catalogued through metagenomics, but these data do not yet provide critical information for representing process rates at large scales (Fierer et al., 2021).

The second question in this stage of the workflow requires ecologists to get familiarized with ESMs and the way processes are represented: *Is your process already in or related to an existing process in an ESM?* Investigating this question will help identify existing model frameworks that can be used as scaffolding for building simple models and ultimately incorporating the process into an ESM. ESMs represent similar environmental processes using a variety of different approaches and equations, so it might help to start by identifying one or more ESMs that you may be interested in and reading model documentation to determine how related processes are represented and whether the model will fit your needs. For example, if you want to improve the representation of foliar nitrogen acquisition, it is vital that the model you choose already has a terrestrial nitrogen cycle. This is also an ideal time to discuss collaborations with ESM developers. We encourage ESM developers at this stage to welcome ecologists interested in working with ESMs by taking the time to explain modeling concepts in jargon-free language and providing resources to work through technical challenges.

If the selected ESM already contains a model of the process, the empiricist can consider how it can be improved or revised using new data or theoretical understanding. Many times a process is represented implicitly (e.g. soil microbial activity is often represented using a cascading decomposition scheme (Wieder, Allison, et al., 2015; Wieder et al., 2018)). Illustrating that explicit representation of the process will fundamentally change model behavior will help to determine whether an explicit representation is needed. In addition, if the current representation of the process connects multiple cycles (e.g. carbon and nitrogen, water and energy), exploring existing model structures can help empiricists understand all the connections between their process and various cycles that must be elucidated and quantified when updating the ESM. Like hooking up speakers to a television or finding the right dongle to plug in your phone, the new process will only work within the ESM if all the appropriate ins and outs are connected. If the process is not currently in a model, it is worth investigating why not (perhaps connecting with an ESM modeler) and whether it might be implicitly included through other model process representations. For example, plant hydraulic stress is not always explicitly included in ESMs (Kennedy et al., 2019), but may be implicitly included by existing connections between soil moisture and stomatal conductance.

The third and final question helps to identify ecological concepts that may be more appropriate to a different type of modeling because they are unlikely to alter climate simulations within an ESM: *Is the process likely to influence climate on scales of time and space consistent with other ESM processes?* Put another way, is the process likely to change the results of global climate simulations using ESMs? Generally, ecology in ESMs impacts climate prediction in two major ways: through biogeochemical (carbon and nutrient cycling) and biogeophysical (evapotranspiration and energy fluxes) processes. Coupling these processes provides a means for assessing feedbacks between ecosystems and climate that distinguish ESMs from stand-alone ecosystem models.

Simple estimates can be made to assess whether a process, when applied to large regions or the entire globe, has the potential to meaningfully influence climate. For example, the general process of insect herbivory, which responds to temperature (e.g., Deutsch et al., 2018; Edburg et al., 2011) and could meaningfully affect carbon fluxes through changing plant biomass, might influence climate (Box 1). On the other hand, temperature affects the distribution and abundance of mosquito species (Hunt et al., 2017), but if mosquitoes are not known to have a meaningful impact on climate, inclusion of mosquito species distributions would not change the outcome of ESM simulations, and may be better suited to a different type of model. In addition, new, climate-influencing processes must occur or change at a rate that is meaningful at ESM timescales. For example, changes in environmental conditions may alter the rates of soil microbial metabolic processes over the course of minutes or even seconds, but these rapid fluctuations are too fast to capture in the timestep of a typical ESM. On the other end of the spectrum, bedrock weathering is a process that releases nutrients for plants and may impact plant biomass (Morford et al., 2011), but it happens so slowly that it is unlikely to shift simulated plant productivity in an ESM over decade to century timescales.

Apart from facilitating ESM incorporation, these questions produce valuable intellectual products on their own: greater understanding of how a process fits into the terrestrial system, identification of knowledge gaps and a clear path towards future empirical work, and determining whether an ESM is the appropriate modeling tool for the process of interest.

Reflecting on these questions can help ecologists define "future directions" for their work with greater specificity than "inclusion in a model," and also generate valuable insights into the scale of an ecological process and its connections to water, energy, or biogeochemical cycles. In a classroom setting, these questions can be an effective way to practice "thinking like a modeler" without requiring any involvement with programming. Regardless of whether the answer to all of these questions for a given ecological concept is "yes", they are beneficial for ecologists to ask.

Box 1:

Herbivores like insects and grazers have large impacts on plant biomass and productivity, yet they are still absent from ESMs. How do the conceptual questions in Part 1 of the workflow guide next steps in deciding whether to incorporate herbivores in ESMs? Although herbivores are broadly not yet included in ESMs (Question 2) and are known to have important impacts on plant biomass with feedbacks to climate (Question 3), ESMs also require that any new process behave consistently across space and time (Question 1) in a way that can be captured quantitatively. To move forward with incorporating herbivores into ESMs, the known impact of herbivores on plant biomass must be reduced down to quantifiable patterns that are consistent across space and time. For example, do herbivores reduce plant biomass by a fixed proportion, or by a proportion that depends on climate factors already present in ESMs like temperature and precipitation? Does the impact of herbivores vary in a predictable way across continents and ecoregions? If the answer is yes, then perhaps a simple model can be developed (Workflow part 2) or existing simple models can be considered for ESM incorporation (Workflow part 3).

Workflow part 2: Beginning to work with simple models

After assessing the theoretical understanding of a process and its likely importance for terrestrial ecosystems and climate, the next workflow steps involve the iterative development, implementation, and evaluation of simple models outside of the ESM, in addition to the collection and/or assembly of data necessary to apply the simple model at large scales (Fig. 4, "Test process alone"). The aim of these activities is to generate knowledge, highlight uncertainties, and refine understanding of the process(es) in question. At its core, this stage involves identifying formulas to represent our theoretical understanding of ecological systems.

This stage is a key precursor to working with ESMs because once a process is integrated into an ESM, it becomes harder to discern the cause of disagreement with observations, and uncertainty increases. For example, photosynthesis can be evaluated with leaf gas exchange data in highly controlled chambers. Gross primary productivity, on the other hand, is evaluated using eddy covariance flux towers. Errors can arise in the model's scaling from leaf to canopy, soil moisture, nitrogen availability, leaf area index, and aspects of the model other than the photosynthesis parameterization (Rogers et al., 2017). The "test process alone" stage is essential to identify the adequacy of a process model before compensating errors occur within the ESM. Although not a strict requirement, this phase of the workflow is best accomplished with equal, collaborative contributions from both empiricists and modelers (Fig. 2) including someone familiar with ESMs who can craft a bridge for future process incorporation.

Simple models are created at this stage by translating knowledge from conceptual models of organisms and ecosystems to mathematical representations of matter and energy. The development of simple models can start by creating a simple statistical model or using a pre-existing model. For example, R has a photosynthesis package (Duursma, 2015) that can be used as a starting point for modifications to photosynthesis like temperature acclimation (e.g., (Smith et al., 2017)) or ozone damage (e.g., Lombardozzi et al., 2012). Simple models can also be developed using any coding language (both R and Python are free and open source), or even start by using a spreadsheet program like Excel, and can range in complexity from a single equation to a complex web of variables and parameters. Unlike the first phase of the workflow, testing theory with data at this phase requires some comfort with programming and data management (for resources, see Table 1). These activities can be easily integrated into ecological coursework, and a variety of resources have been developed to facilitate this (e.g., (Carey et al., 2020)). Additionally, cross-disciplinary collaboration is beneficial at this stage, as it helps to formalize conceptual models, clarify assumptions, evaluate ideas within the scientific community about a process, connect various components of ecosystems and the Earth system, and test the broader applicability of theories over space and time.

In addition to simple model development, this phase of the workflow involves assembling the data necessary to estimate parameters and drive simple models at large scales. (Note: In a model, a "parameter" is the value of a variable in an equation. The word "parameterization" may seem like a derivative of "parameter", but is in fact a separate concept referring to representing a complex microscale process as an approximate bulk process. For example, model representations of photosynthesis are a parameterization of subcellular-level processes, and may use parameter values within the calculation (Bonan, 2019)). Necessary data fall into several distinct categories: data for parameter estimation during model development, driver data to feed into the model (e.g., climate or soil characteristics), and data for benchmarking the model following simulations (i.e., observational data to compare against model output).

At this stage, it is worth making a "shopping list" of the data necessary for a given modeling exercise and evaluating the availability of values at the relevant scale (Fig. 3). These data may come initially from a single site or lab experiments, but to eventually scale model results globally, data gathered across multiple regions and experiments become useful. ESMs use a variety of large-scale datasets for parameter estimation and evaluation, and it can be helpful to seek out datasets already in use before attempting to assemble a new dataset from scratch. Large-scale data can come from meta-analytical techniques and syntheses (e.g., Field & Gillett, 2010; Ainsworth & Long, 2005; Lombardozzi et al., 2013), pre-existing large synthesized datasets (e.g., SoDaH (Wieder et al., 2020), TRY (Kattge et al., 2011)), satellite data (e.g., Li & Xiao, 2019), or model-derived products (e.g., Fluxnet-MTE (Jung et al., 2020)). Direct measurements are generally preferable for parameter estimation and model evaluation but are not always feasible to collect. As a result, parameter estimation and model evaluation often use data products (i.e., data that have been modified by models) to achieve the spatial and temporal scales required by the ESM. Data products can be closely connected to the original data (i.e., data averages) or less closely connected (i.e., output of another mechanistic model that uses data as an input). Understanding the uncertainty of a data product is critical for determining the value of its use in parameter estimation and model evaluation (Dagon et al., 2020; Dietze, 2017). Simple models often get stuck here on the way to ESM incorporation because of gaps in data requirements to run models at global scales (e.g., lack of maps of soil edaphic properties or other input data that may be critical for further model development).

The creation and improvement of simplified mathematical models and large-scale synthesized datasets makes several valuable contributions to understanding and refining ecological theories, regardless of the eventual implementation in ESMs. Simple models help formalize, and make explicit, the underlying assumptions in the theories they represent and can illustrate weaknesses in existing theory. As such, they can be used to generate testable hypotheses that can be interrogated with existing data or new experiments. Estimating parameters for simple models with available observations helps identify data and knowledge gaps that can be addressed with further study. Compared to larger ESMs, simple models have greater traceability, allowing scientists to explore and understand model complexity, their associated uncertainties, and emergent properties that can be evaluated with independent observations. These simpler models also have the advantage of being easier to use, with greater flexibility and lower computation costs than running a full ESM, and can potentially be implemented in ESMs in a modularized manner that allows for testing multiple ecological theories (e.g., Fisher & Koven, 2020). Finally, these models help to clarify theory and develop concepts through independent community efforts to use them and improve their process representation.

Box 2:

After establishing that a new process is appropriate to consider including in an ESM (Part 1), what comes next? Current models of soil microbial activity highlight Part 2 of the workflow: simple quantified models evaluated at a variety of scales but not yet incorporated into ESMs. As an example, the MIcrobial-MIneral Carbon Stabilization (MIMICS) model was motivated by theories highlighting interactions among soil microbes and minerals that are responsible for soil organic matter decomposition and persistence. A simple process model was initially developed in R using measurements from laboratory experiments and rates of leaf litter mass loss. This model was tested first at a single site (Wieder et al., 2014), and subsequent evaluation across continental and global scale gradients illustrated reasonable agreement with litter decay rates and soil carbon stocks (Wieder, Grandy, et al., 2015) and a higher vulnerability of Arctic soil C stocks, compared to models that implicitly represent microbial activity (Wieder et al., 2019). MIMICS continues to undergo further development (e.g. to include coupled C-N biogeochemistry (Kyker-Snowman et al., 2020) and vertical resolution (Y. Wang et al., 2021)), refinement (Zhang et al., 2020), and evaluation (Basile et al., 2020; Koven et al., 2017; Shi et al., 2018; Sulman et al., 2018). All of these activities rely on conducting simulations across multiple study sites and at global scales, which is a valuable precursor to considering incorporating MIMICS into an ESM.

Workflow part 3: Integrating processes into ESMs

Developing and evaluating a simple model ultimately paves the way for integrating a process into an ESM, as illustrated in the final stage of the workflow (Fig. 4, “Test process with ESM”). The first step is deciding which ESM to use. Many ESMs exist and vary substantially in their ecological process representations (Fisher & Koven, 2020), and adding a new process requires an understanding of how processes of interest are currently represented in a given ESM (as in Stage 1) and a simple model that can be integrated within the framework of that ESM (developed in Stage 2). Additionally, some ESMs have proprietary or restricted access (e.g., GFDL-ESM, IPSL-CM5 (Dufresne et al., 2013; Dunne et al., 2020)) and require collaboration and/or approval by model developers, while others are open-source and community driven (e.g., CESM, E3SM (Danabasoglu et al., 2020; Golaz et al., 2019)). While not always required, incorporating new processes will be most efficient when building relationships with model developers who can help with technical aspects of code development. For example, developers with experience in running and testing the model can provide code structure guidance and highlight possible interactions or feedbacks among processes that might not be obvious to a novice model developer. ESM communities can be insular and siloed at times, and ESM developers at this stage can help build more integrated empirical-modeling collaborations by seeking out and remaining open to working with ecologists (see Table 1 for several opportunities).

Once access to model code is available, integrating the new process representation can begin. The first step is finding the location to integrate the new process. While this will vary depending on the ESM, code modules will often have descriptive names and the location of variables within the code can be searched using linux- and editor-based search tools (e.g., grep). It is also helpful to find a similar variable or process in the code (with similar inputs and outputs) that can be used as an example for how to structure the new process code. Having an example to mirror can be particularly useful in identifying other modules where the variables may be required (e.g., sometimes setting the initial value for variables happens in a different module). Additionally, it can be helpful to outline or diagram a work plan in advance, noting the modules and variables that will need to be added, modified, and connected.

Modifications should build on each other, starting with a simple change: for example, add a single variable, and then test that the code will compile and run for a short period of time. Sequentially add more complexity, connecting the new variable or process to existing model structure. Using this layered approach will help to identify any structural bugs early in the development process. Although the ultimate goal is to have a sophisticated representation that includes spatially-varying processes, simpler versions of the model can -- and should -- be tested to determine the sensitivity of the system to the new process. These simpler model iterations are excellent training tools for graduate students and postdoctoral trainees as they become more familiar with the model. Once the basic framework for the new process is in place, it can be tested to identify the magnitude of change in relevant processes, as well as any interactions with other ecosystem processes. Often, these proof-of-concept simulations can turn into publications that highlight the potential importance of the process at site or global scales and identify gaps in data that can help to improve the process representation.

Throughout the development, testing, and evaluation process, the simplest relevant version or component of the ESM available should be used. For example, if the new process does not rely on carbon cycling, it may be possible to leave out this portion of the model in your testing, allowing the model to run faster and reducing the complexity of model interactions. Often with ecological processes, the development process uses only the terrestrial component of an ESM driven by a gridded atmospheric data product (e.g., reanalysis), since fully coupled ESM runs are far more computationally expensive than smaller terrestrial-only runs. Additionally, running in the coarsest available resolution and for the smallest spatial domain possible (e.g., a single site) will expedite model testing. Once code is tested, running it globally (and eventually coupled to an atmospheric model) is necessary to ensure the simulation operates appropriately over the global domain.

An approach called “modular development” can also be useful for testing and evaluating different ecological theories, and can be employed when implementing new processes in ESMs (Fisher & Koven, 2020; see also Clark et al., 2015). This involves adding an alternate representation of a process that is already simulated in a model (not removing the process) and letting the user specify which theory the model will use in a given simulation. For example, testing multiple representations of stomatal conductance (Franks et al., 2018), soil carbon and nitrogen cycling (Wieder, Cleveland, et al., 2015; Wieder et al., 2018), and hydrology (Clark et al., 2008, 2011) have been helpful in testing different theories and highlighting when and where certain process representations perform best. This allows for refinement of existing theory and process representation, advancing the state of current knowledge.

Once the new process is incorporated, the model must be tested and evaluated. A first step is to determine whether the new process fundamentally changes model behavior relative to a simulation without this process. Does it affect other simulated processes, and by how much? Many processes do not exist in isolation within a model and thus cannot be modified for only one purpose. Better models of photosynthesis, for example, may be desired to improve the carbon cycle, but also impact energy and water fluxes to the atmosphere through stomatal conductance (Bonan et al., 2011). A second step is to evaluate model behavior against observations. Model evaluation is most effective if multiple processes are assessed, and is most useful when compared to evaluation of a baseline model simulation where the new process is not simulated. This step is similar to simple model evaluation in the second stage of this workflow, but this evaluation process should be repeated once the simple model is embedded within an ESM. One simple form of evaluation is to run a simulation at a single location where relevant observational or experimental manipulation data have been collected, such as a field site or a flux tower (Cheng et al., 2019; Medlyn et al., 2015). These data can be used to assess whether the new model behavior fundamentally changes model performance (De Kauwe et al., 2013, 2014; Smith et al., 2015; Thomas et al., 2013; Zaehle et al., 2014). It is also important to evaluate global responses. While global data can be more challenging to access, several resources are currently available. Perhaps the most useful is the International Land Model Benchmarking (ILAMB; Collier et al., 2018) project, which has developed internationally accepted benchmarking standards for ESM performance. This project has compiled global datasets for a range of variables and can help to identify where model performance is enhanced or degraded. Remotely sensed data products can also help with model evaluation at regional to global scales.

One of the greatest challenges in ESM development is ensuring parsimony while capturing the full range of biological complexity. This is particularly challenging for community models with contributors from multiple fields and institutions, which commonly suffer from "feature fatigue". Human instinct is to continue to add features to a solution, even when removing features may be more beneficial or efficient (Adams et al., 2021). While adding processes can improve model realism, care must be taken to avoid sacrificing model reliability, which can be degraded with the addition of uncertain parameters (Prentice et al., 2015). Eco-evolutionary optimality theory is one recent tool that can be used to improve model realism while limiting the number of new parameters (Box 3; Scott & Smith, 2021; H. Wang et al., 2017). Unlike statistical approaches where environmental responses are hard-coded with parameters, a theoretical approach allows process responses to emerge with fewer parameters (Prentice et al., 2015). These responses can then be tested with data that might, in a more statistical approach, be needed to estimate parameters.

The workflow so far has presented guidelines for incorporating a new process into an ESM, which requires substantial work in developing and incorporating new code into a model and then evaluating the responses of terrestrial processes. Often, the ecological workflow ends here with the assessment of the global-scale impact of a process and how it may change ecological functioning through time. Beyond this, an exciting next step is to understand whether this new process has climate feedbacks by comparing land-only and coupled model simulations. Land models can be coupled to other ESM components (atmosphere, ocean, ice, etc.) to investigate global feedbacks in water, energy or biogeochemical cycles. Connecting land and atmosphere components allows investigation of unexpected feedbacks with the atmosphere that may be different from land-only simulations.

Box 3:

One example of how models have maintained parsimony (Part 3 of the workflow) is photosynthetic acclimation (Smith & Dukes, 2013). Initially, empirical models were developed to simulate temperature acclimation of photosynthetic biochemical capacity in ESMs based on observed responses (e.g., Kattge et al., 2009; Kattge & Knorr, 2007) and then incorporated in ESMs (Friend, 2010; Lombardozzi, Bonan, et al., 2015; Mercado et al., 2018; Smith et al., 2017; Smith & Dukes, 2013; Ziehn et al., 2011). However, more recently, eco-evolutionary optimality theory has been invoked to simulate photosynthetic biochemical capacity in a way that incorporates the processes without added parameters (configuration variables internal to a model that rely on observational data), thus increasing model realism without altering model reliability (Scott & Smith, 2021; Smith & Keenan, 2020; H. Wang et al., 2017). Eco-evolutionary optimality theory approaches rely on the assumption that natural selection will remove non-competitive traits from an environment, thus providing testable, theoretical trait responses to the *environment over short and long time scales, and offer potential promising avenues for adding biological processes to ESMs with little to no added parameters (Franklin et al., 2020). Eco-evolutionary optimality approaches are available to simulate processes at the leaf (Jiang et al., 2020; Prentice et al., 2014; Smith et al., 2019; Smith & Keenan, 2020; H. Wang et al., 2020; H. Wang et al., 2017), plant (Dybzinski et al., 2015; Farrior et al., 2013; Weng et al., 2015) and ecosystem (Baskaran et al., 2017; Franklin et al., 2020) scales.*

Box 4:

The following example illustrates the entire workflow, from initial conceptual development to simple modeling to working with ESMs. As part of her research, co-author Lombardozzi measured how leaf-level gas exchange changed in response to ground-level ozone. Upon analyzing her data, she found that leaf-level carbon (photosynthesis) and water (transpiration) fluxes decreased at different rates. Since these are both important greenhouse gases and affect fundamental plant processes (photosynthesis and stomatal conductance, which scale through time and space regardless of biome), she thought that ozone damage could have a global impact on climate feedbacks on model-relevant timescales and therefore should be included in large-scale models. Although Lombardozzi had no modeling or coding experience, she emailed several people working on the Community Land Model (CLM) to see if they might want to collaborate. She did some research about the photosynthesis and stomatal conductance models used in CLM and talked with modeling colleagues to decide how to best include this type of damage. After completing online Linux and Fortran tutorials, Lombardozzi started using a simple photosynthesis-stomatal conductance model provided by her colleagues. She applied linear regressions calculated from her experiment to the rates of maximum carboxylation (Vcmax) to simulate ozone damage to photosynthetic enzymes. She was able to show that including ozone damage improved simulated photosynthesis and stomatal conductance at the leaf scale (Lombardozzi et al., 2012).

Did these changes matter globally? Lombardozzi worked with model developers to find out, using the simple model to update code in the CLM to account for ozone damage. Using data from her experiment and a constant ozone concentration, she showed that ozone did have large consequences for carbon and water cycling globally (Lombardozzi et al., 2013). While this experiment highlighted the sensitivity of global processes to ozone damage, it did not provide a *realistic assessment of how ozone changes carbon and water cycling. Lombardozzi therefore synthesized existing published literature to determine how photosynthesis and stomatal conductance change in relation to ozone exposure, and identified a complete lack of data for tropical forests (Lombardozzi et al., 2013). Despite missing data for large biomes, these data were then used to update the CLM code to capture responses across different plant functional categories (e.g., broadleaf trees, needleleaf trees, herbaceous vegetation), and when combined with realistic ozone data, simulated that ozone decreases global photosynthesis by 10.8% and transpiration by 2.2%, with larger impacts in Eastern US, Europe, and Southeast Asia (Lombardozzi, Levis, et al., 2015).*

IV. Creating community change across scales

Empirical and modeling communities already work together and influence one another in many ways, yet integrating ecological processes into ESMs remains a persistently slow process with myriad challenges limiting efficient collaboration. Historically, ESMs have been developed by atmospheric and physical science communities while ecology has only been integrated relatively recently, and the disciplinary requirements in trainee education have not provided enough of a shared foundation to build strong conceptual bridges between ESMs and empirical ecologists. These communities must collectively address persistent obstacles including confusing technical language, lack of resources for skills development, and the need for better connections and integration across scientific communities. We provide resources to help expand terrestrial ecological process representation in ESMs (Table 1). With the advent of these and other tools, empiricists will be better poised to take advantage of technical workflows that can help streamline data-model integration (e.g., Fer et al., 2021).

The interdisciplinary work of developing an Earth system model is not only technical, but also social. As such, in addition to the workflow presented above, we offer specific suggestions for restructuring ecological education and interactions within collaborations (see Section III), both of which are key to ensuring that the workflow does not break down. For bridge-building between communities to be inclusive, the modeling and empirical communities need to examine their community practices, values, and norms. This work includes understanding the demographics of *who* is (and is not) represented in the research communities (Bernard & Cooperdock, 2018), *what* processes our communities are willing to consider (or dismiss) as valuable contributions to ESMs (e.g., microbes, moths, management), *where* data are collected and why some regions or ecosystems are over/under sampled (Martin et al., 2012; Metcalfe et al., 2018), *when* we overlook potential collaborators or fail to provide them with platforms for sharing their work, such as at conferences (Ford et al., 2019), and *why* we make the decisions that we do about where to focus efforts.

Improved collaboration between empirical and modeling communities will positively benefit each community. Adding modeling to empirical work can increase its impact while simultaneously advancing ecological theory, modeling capabilities, and model realism. To get started or go further with this work, we have assembled a list of resources for skills development at each stage of the workflow (Table 2). To maintain contemporary resources, please visit the regularly updated website (https://ecoesm.github.io/). Despite the many complex challenges involved in integrating terrestrial ecology and Earth system modeling, there has never been a better time to attempt such difficult work. Finding and communicating with scientists across the globe is getting easier every year, computing resources are rapidly evolving, and the internet provides an ever-growing assortment of free tools for developing new quantitative and programming skills. In addition, funding sources are increasingly recognizing the value of data-model integration (e.g. the NASA Modeling, Analysis, and Prediction program (https://map.nasa.gov/) or the USDA NIFA Data Science for Food and Agricultural Systems program (https://nifa.usda.gov/program/dsfas)) and grassroots efforts are creating a framework for these collaborations using workshops and tutorials. Our insights into the history of ecology in ESMs, workflow for developing and incorporating ecological processes into ESMs, and specific resource suggestions will advance this exciting progress and provide a scaffold for building fruitful bridges between empirical and modeling communities.

Acknowledgements

We thank graphic designer Elena Hartley for her excellent work designing the figures in this publication.

Funding Information

Funding for this project was provided by the USDA National Institute of Food and Agriculture (Project No. 2015-35615-22747).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author Contribution

E.K.S. and D.L. led the writing and editing of the manuscript. R.M. contributed to the artistic development of Figure 1. All authors contributed to the idea development, figure concepts, and writing and editing of the manuscript.

References

Adams, G. S., Converse, B. A., Hales, A. H., & Klotz, L. E. (2021). People systematically overlook subtractive changes. *Nature*, *592*(7853), 258–261.

Ainsworth, E. A., & Long, S. P. (2005). What have we learned from 15 years of free-air CO2 enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO2. *The New Phytologist*, *165*(2), 351–371.

Anadu, J., Ali, H., & Jackson, C. (2020). Ten steps to protect BIPOC scholars in the field. *Eos*, *101*. https://doi.org/10.1029/2020eo150525

Anderson-Cook, C. M., Lu, L., & Parker, P. A. (2019). Effective interdisciplinary collaboration between statisticians and other subject matter experts. *Quality Engineering*, *31*(1), 164–176.

Basile, S. J., Lin, X., Wieder, W. R., Hartman, M. D., & Keppel-Aleks, G. (2020). Leveraging the signature of heterotrophic respiration on atmospheric CO2 for model benchmarking. *Biogeosciences*, *17*(5), 1293–1308.

Baskaran, P., Hyvönen, R., Berglund, S. L., Clemmensen, K. E., Ågren, G. I., Lindahl, B. D., & Manzoni, S. (2017). Modelling the influence of ectomycorrhizal decomposition on plant nutrition and soil carbon sequestration in boreal forest ecosystems. *The New Phytologist*, *213*(3), 1452–1465.

Bernard, R. E., & Cooperdock, E. H. G. (2018). No progress on diversity in 40 years. *Nature*

Geoscience, *11*(5), 292–295.
Bonan, G. B. (1995). Land-atmosphere CO2 exchange simulated by a land surface process model coupled to an atmospheric general circulation model. *Journal of Geophysical Research*, *100*(D2), 2817.
Bonan, G. B. (2008). Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests. *Science*, *320*(5882), 1444–1449.
Bonan, G. B. (2016). Forests, Climate, and Public Policy: A 500-Year Interdisciplinary Odyssey. *Annual Review of Ecology, Evolution, and Systematics*, *47*(1), 97–121.
Bonan, G. B. (2019). *Climate Change and Terrestrial Ecosystem Modeling*. Cambridge University Press.
Bonan, G. B., & Doney, S. C. (2018). Climate, ecosystems, and planetary futures: The challenge to predict life in Earth system models. *Science*, *359*(6375). https://doi.org/10.1126/science.aam8328
Bonan, G. B., Lawrence, P. J., Oleson, K. W., Levis, S., Jung, M., Reichstein, M., Lawrence, D. M., & Swenson, S. C. (2011). Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data. *Journal of Geophysical Research*, *116*(G2). https://doi.org/10.1029/2010jg001593
Bonan, G. B., & Levis, S. (2010). Quantifying carbon-nitrogen feedbacks in the Community Land Model (CLM4). *Geophysical Research Letters*, *37*(7). https://doi.org/10.1029/2010gl042430
Bonan, G. B., Levis, S., Sitch, S., Vertenstein, M., & Oleson, K. W. (2003). A dynamic global vegetation model for use with climate models: concepts and description of simulated vegetation dynamics. *Global Change Biology*, *9*(11), 1543–1566.
Bracco, A., Long, M. C., Levine, N. M., Thomas, R. Q., Deutsch, C., & McKinley, G. A. (2015). NCAR's summer colloquium: Capacity building in cross-disciplinary research of earth system carbon–climate connections. *Bulletin of the American Meteorological Society*, *96*(8), 1381–1384.
Carey, C. C., Farrell, K. J., Hounshell, A. G., & O'Connell, K. (2020). Macrosystems EDDIE teaching modules significantly increase ecology students' proficiency and confidence working with ecosystem models and use of systems thinking. *Ecology and Evolution*, *10*(22), 12515–12527.

Charney, J., Stone, P. H., & Quirk, W. J. (1975). Drought in the sahara: a biogeophysical feedback mechanism. *Science*, *187*(4175), 434–435.

Chaudhary, V. B., & Berhe, A. A. (2020). Ten simple rules for building an antiracist lab. *PLoS Computational Biology*, *16*(10), e1008210.

Cheng, S. J. (2018, March 16). *Modeling global change ecology in a high–carbon dioxide world*. https://eos.org/meeting-reports/modeling-global-change-ecology-in-a-high-carbon-dioxide-world

Cheng, S. J., Hess, P. G., Wieder, W. R., Thomas, R. Q., Nadelhoffer, K. J., Vira, J., Lombardozzi, D. L., Gundersen, P., Fernandez, I. J., Schleppi, P., Gruselle, M.-C., Moldan, F., & Goodale, C. L. (2019). Decadal fates and impacts of nitrogen additions on temperate forest carbon storage: a data–model comparison. *Biogeosciences*, *16*(13), 2771–2793.

Clark, M. P., Kavetski, D., & Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling: Hypothesis testing in hydrology. *Water Resources Research*, *47*(9). https://doi.org/10.1029/2010wr009827

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Brekke, L. D., Arnold, J. R., Gochis, D. J., & Rasmussen, R. M. (2015). A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resources Research*, *51*(4), 2498–2514.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Gochis, D. J., Rasmussen, R. M., Tarboton, D. G., Mahat, V., Flerchinger, G. N., & Marks, D. G. (2015). A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies. *Water Resources Research*, *51*(4), 2515–2542.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., & Hay, L. E. (2008). Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models: Differences between hydrological models. *Water Resources Research*, *44*(12). https://doi.org/10.1029/2007wr006735

Collier, N., Hoffman, F. M., Lawrence, D. M., Keppel-Aleks, G., Koven, C. D., Riley, W. J., Mu, M., & Randerson, J. T. (2018). The international land model benchmarking (ILAMB) system: Design, theory, and implementation. *Journal of Advances in Modeling Earth*

Systems, *10*(11), 2731–2754.

Cooley, E. (1994). Training an Interdisciplinary Team in Communication and Decision-Making Skills. *Small Group Research*, *25*(1), 5–25.

Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A., & Totterdell, I. J. (2000). Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature*, *408*(6809), 184–187.

Dagon, K., Sanderson, B. M., Fisher, R. A., & Lawrence, D. M. (2020). A machine learning approach to emulation and biophysical parameter estimation with the Community Land Model, version 5. *Advances in Statistical Climatology Meteorology and Oceanography*, *6*(2), 223–244.

Danabasoglu, G., Lamarque, J. -F, Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T. M., Lindsay, K., Lipscomb, W. H., Mills, M. J., … Strand, W. G. (2020). The community earth system model version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, *12*(2). https://doi.org/10.1029/2019ms001916

De Kauwe, M. G., Medlyn, B. E., Zaehle, S., Walker, A. P., Dietze, M. C., Hickler, T., Jain, A. K., Luo, Y., Parton, W. J., Prentice, I. C., Smith, B., Thornton, P. E., Wang, S., Wang, Y.-P., Wårlind, D., Weng, E., Crous, K. Y., Ellsworth, D. S., Hanson, P. J., … Norby, R. J. (2013). Forest water use and water use efficiency at elevated CO2 : a model-data intercomparison at two contrasting temperate forest FACE sites. *Global Change Biology*, *19*(6), 1759–1779.

De Kauwe, M. G., Medlyn, B. E., Zaehle, S., Walker, A. P., Dietze, M. C., Wang, Y.-P., Luo, Y., Jain, A. K., El-Masri, B., Hickler, T., Wårlind, D., Weng, E., Parton, W. J., Thornton, P. E., Wang, S., Prentice, I. C., Asao, S., Smith, B., McCarthy, H. R., … Norby, R. J. (2014). Where does the carbon go? A model-data intercomparison of vegetation carbon allocation and turnover processes at two temperate forest free-air CO2 enrichment sites. *The New Phytologist*, *203*(3), 883–899.

Denning, A. S., Randall, D. A., Collatz, G. J., & Sellers, P. J. (1996). Simulations of terrestrial carbon metabolism and atmospheric CO2 in a general circulation model. Part 2: Simulated CO2 concentrations. *Tellus. Series B, Chemical and Physical Meteorology*, *48*(4), 543–567.

Deutsch, C. A., Tewksbury, J. J., Tigchelaar, M., Battisti, D. S., Merrill, S. C., Huey, R. B., &

Naylor, R. L. (2018). Increase in crop losses to insect pests in a warming climate. *Science*, *361*(6405), 916–919.

Dickinson, R. E. (1984). Modeling evapotranspiration for three-dimensional global climate models. *Climate Processes and Climate Sensitivity*, *29*, 58–72.

Dickinson, R. E. (1986). Biosphere/Atmosphere Transfer Scheme (BATS) for the NCAR Community Climate Model. *Technical Report, NCAR*. https://ci.nii.ac.jp/naid/10009851528/

Dickinson, R. E., & Henderson-Sellers, A. (1988). Modelling tropical deforestation: A study of GCM land-surface parametrizations. *Quarterly Journal of the Royal Meteorological Society, 114*(480), 439–462. https://doi.org/10.1002/qj.49711448009

Dickinson, R. E., Jaeger, J., Washington, W. M., & Wolski, R. (1981). *Boundary subroutine for the NCAR global climate model*. National Center for Atmospheric Research.

Dietze, M. C. (2017). Prediction in ecology: a first-principles framework. *Ecological Applications: A Publication of the Ecological Society of America*, *27*(7), 2048–2060.

Duffy, M. A., García-Robledo, C., Gordon, S. P., Grant, N. A., Green, D. A., Kamath, A., Penczykowski, R. M., Rebolleda-Gómez, M., Wale, N., & Zaman, L. (2021). Model Systems in Ecology, Evolution, and Behavior: A Call for Diversity in Our Model Systems and Discipline. *The American Naturalist*.

Dufresne, J.-L., Foujols, M.-A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., & Others. (2013). Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. *Climate Dynamics*, *40*(9), 2123–2165.

Dunne, J. P., Horowitz, L. W., Adcroft, A. J., Ginoux, P., Held, I. M., John, J. G., Krasting, J. P., Malyshev, S., Naik, V., Paulot, F., Shevliakova, E., Stock, C. A., Zadeh, N., Balaji, V., Blanton, C., Dunne, K. A., Dupuis, C., Durachta, J., Dussin, R., … Zhao, M. (2020). The GFDL earth system model version 4.1 (GFDL-ESM 4.1): Overall coupled model description and simulation characteristics. *Journal of Advances in Modeling Earth Systems*, *12*(11). https://doi.org/10.1029/2019ms002015

Duursma, R. A. (2015). Plantecophys--An R Package for Analysing and Modelling Leaf Gas Exchange Data. *PloS One*, *10*(11), e0143346.

Dybzinski, R., Farrior, C. E., & Pacala, S. W. (2015). Increased forest carbon storage with increased atmospheric CO_2 despite nitrogen limitation: A game-theoretic allocation model for trees in competition for nitrogen and light. *Global Change Biology*, *21*(3), 1182–1196.

Edburg, S. L., Hicke, J. A., Lawrence, D. M., & Thornton, P. E. (2011). Simulating coupled carbon and nitrogen dynamics following mountain pine beetle outbreaks in the western United States. *Journal of Geophysical Research*, *116*(G4). https://doi.org/10.1029/2011jg001786

Edwards, P. N. (2011). History of climate modeling. *Wiley Interdisciplinary Reviews. Climate Change*, *2*(1), 128–139.

Emery, N. C., Bledsoe, E. K., Hasley, A. O., & Eaton, C. D. (2021). Cultivating inclusive instructional and research environments in ecology and evolutionary science. *Ecology and Evolution*, *11*(4), 1480–1491.

Farrior, C. E., Dybzinski, R., Levin, S. A., & Pacala, S. W. (2013). Competition for water and light in closed-canopy forests: a tractable model of carbon allocation with implications for carbon sinks. *The American Naturalist*, *181*(3), 314–330.

Fer, I., Gardella, A. K., Shiklomanov, A. N., Campbell, E. E., Cowdery, E. M., De Kauwe, M. G., Desai, A., Duveneck, M. J., Fisher, J. B., Haynes, K. D., & Others. (2021). Beyond ecosystem modeling: A roadmap to community cyberinfrastructure for ecological data-model integration. *Global Change Biology*, *27*(1), 13–26.

Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *The British Journal of Mathematical and Statistical Psychology*, *63*(Pt 3), 665–694.

Fierer, N., Wood, S. A., & Bueno de Mesquita, C. P. (2021). How microbes can, and cannot, be used to assess soil health. *Soil Biology & Biochemistry*, *153*, 108111.

Fisher, R. A., & Koven, C. D. (2020). Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems. *Journal of Advances in Modeling Earth Systems*, *12*(4). https://doi.org/10.1029/2018ms001453

Foley, J. A., Prentice, I. C., Ramankutty, N., Levis, S., Pollard, D., Sitch, S., & Haxeltine, A. (1996). An integrated biosphere model of land surface processes, terrestrial carbon balance, and vegetation dynamics. *Global Biogeochemical Cycles*, *10*(4), 603–628.

Ford, H. L., Brick, C., Azmitia, M., Blaufuss, K., & Dekens, P. (2019). Women from some under-represented minorities are given too few talks at world’s largest Earth-science conference. *Nature*, *576*(7785), 32–35.

Franklin, O., Harrison, S. P., Dewar, R., Farrior, C. E., Brännström, Å., Dieckmann, U., Pietsch, S., Falster, D., Cramer, W., Loreau, M., Wang, H., Mäkelä, A., Rebel, K. T., Meron, E., Schymanski, S. J., Rovenskaya, E., Stocker, B. D., Zaehle, S., Manzoni, S., … Prentice, I. C. (2020). Organizing principles for vegetation dynamics. *Nature Plants*, *6*(5), 444–453.

Franks, P. J., Bonan, G. B., Berry, J. A., Lombardozzi, D. L., Holbrook, N. M., Herold, N., & Oleson, K. W. (2018). Comparing optimal and empirical stomatal conductance models for application in Earth system models. *Global Change Biology*, *24*(12), 5708–5723.

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., Bakker, D. C. E., Canadell, J. G., Ciais, P., Jackson, R. B., Anthoni, P., Barbero, L., Bastos, A., Bastrikov, V., Becker, M., … Zaehle, S. (2019). Global carbon budget 2019. *Earth System Science Data*, *11*(4), 1783–1838.

Friend, A. D. (2010). Terrestrial plant production and climate change. *Journal of Experimental Botany*, *61*(5), 1293–1309.

Fung, I. Y., Doney, S. C., Lindsay, K., & John, J. (2005). Evolution of carbon sinks in a changing climate. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(32), 11201–11206.

Giles, S., Jackson, C., & Stephen, N. (2020). Barriers to fieldwork in undergraduate geoscience degrees. *Nature Reviews Earth & Environment*, *1*(2), 77–78.

Golaz, J., Caldwell, P. M., Van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., Abeshu, G., Anantharaj, V., Asay-Davis, X. S., Bader, D. C., Baldwin, S. A., Bisht, G., Bogenschutz, P. A., Branstetter, M., Brunke, M. A., Brus, S. R., Burrows, S. M., Cameron-Smith, P. J., Donahue, A. S., … Zhu, Q. (2019). The DOE E3SM coupled model version 1: Overview and evaluation at standard resolution. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2089–2129.

Hampton, S. E., Jones, M. B., Wasser, L. A., Schildhauer, M. P., Supp, S. R., Brun, J., Hernandez, R. R., Boettiger, C., Collins, S. L., Gross, L. J., Fernández, D. S., Budden, A., White, E. P., Teal, T. K., Labou, S. G., & Aukema, J. E. (2017). Skills and Knowledge for Data-Intensive Environmental Research. *Bioscience*, *67*(6), 546–557.

Huang, Y., Guenet, B., Wang, Y. L., & Ciais, P. (2021). Global simulation and evaluation of soil organic matter and microbial carbon and nitrogen stocks using the microbial decomposition model ORCHIMIC v2.0. *Global Biogeochemical Cycles*. https://doi.org/10.1029/2020gb006836

Hunt, S. K., Galatowitsch, M. L., & McIntosh, A. R. (2017). Interactive effects of land use, temperature, and predators determine native and invasive mosquito distributions. *Freshwater Biology*, *62*(9), 1564–1577.

Jiang, C., Ryu, Y., Wang, H., & Keenan, T. F. (2020). An optimality-based model explains seasonal variation in C3 plant photosynthetic capacity. *Global Change Biology*, *26*(11), 6493–6510.

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N., Chevallier, F., Gans, F., Goll, D. S., Haverd, V., Köhler, P., Ichii, K., Jain, A. K., Liu, J., Lombardozzi, D., … Reichstein, M. (2020). Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach. *Biogeosciences*, *17*(5), 1343–1365.

Kattge, J., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Bönisch, G., Garnier, E., Westoby, M., Reich, P. B., Wright, I. J., Cornelissen, J. H. C., Violle, C., Harrison, S. P., Van Bodegom, P. M., Reichstein, M., Enquist, B. J., Soudzilovskaia, N. A., Ackerly, D. D., Anand, M., … Wirth, C. (2011). TRY - a global database of plant traits. *Global Change Biology*, *17*(9), 2905–2935.

Kattge, J., & Knorr, W. (2007). Temperature acclimation in a biochemical model of photosynthesis: a reanalysis of data from 36 species. In *Plant, Cell & Environment*, *30*(9), 1176–1190. https://doi.org/10.1111/j.1365-3040.2007.01690.x

Kattge, J., Knorr, W., Raddatz, T., & Wirth, C. (2009). Quantifying photosynthetic capacity and its relationship to leaf nitrogen content for global-scale terrestrial biosphere models. *Global Change Biology*, *15*(4), 976–991.

Kennedy, D., Swenson, S., & Oleson, K. W. (2019). Implementing plant hydraulics in the community land model, version 5. *Journal of Advances*. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001500

Koven, C. D., Hugelius, G., Lawrence, D. M., & Wieder, W. R. (2017). Higher climatological temperature sensitivity of soil carbon in cold than warm climates. *Nature Climate Change*, *7*(11), 817–822.

Kuziemsky, C. E., Borycki, E. M., Purkis, M. E., Black, F., Boyle, M., Cloutier-Fisher, D., Fox, L. A., MacKenzie, P., Syme, A., Tschanz, C., Wainwright, W., Wong, H., & Interprofessional Practices Team. (2009). An interdisciplinary team communication framework and its application to healthcare “e-teams” systems design. *BMC Medical Informatics and Decision Making*, *9*(1), 43.

Kyker-Snowman, E., Wieder, W. R., Frey, S. D., & Grandy, A. S. (2020). Stoichiometrically coupled carbon and nitrogen cycling in the MIcrobial-MIneral Carbon Stabilization model version 1.0 (MIMICS-CN v1. 0). *Geoscientific Model Development*, *13*(9), 4413–4434.

Lade, S. J., Steffen, W., de Vries, W., Carpenter, S. R., Donges, J. F., Gerten, D., Hoff, H., Newbold, T., Richardson, K., & Rockström, J. (2019). Human impacts on planetary boundaries amplified by Earth system interactions. *Nature Sustainability*, *3*(2), 119–128.

Law, B. E., Hudiburg, T. W., Berner, L. T., Kent, J. J., Buotte, P. C., & Harmon, M. E. (2018). Land use strategies to mitigate climate change in carbon dense temperate forests. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(14), 3663–3668.

Leuzinger, S., & Thomas, R. Q. (2011). How do we improve Earth system models? Integrating Earth system models, ecosystem models, experiments and long-term data. *New Phytologist*, *191*(1), 15–18. https://doi.org/10.1111/j.1469-8137.2011.03778.x

Li, X., & Xiao, J. (2019). A Global, 0.05-Degree Product of Solar-Induced Chlorophyll Fluorescence Derived from OCO-2, MODIS, and Reanalysis Data. *Remote Sensing*, *11*(5), 517.

Lombardozzi, D. L., Bonan, G. B., Smith, N. G., Dukes, J. S., & Fisher, R. A. (2015). Temperature acclimation of photosynthesis and respiration: A key uncertainty in the carbon cycle-climate feedback. *Geophysical Research Letters*, *42*(20), 8624–8631.

Lombardozzi, D. L., Levis, S., Bonan, G. B., Hess, P. G., & Sparks, J. P. (2015). The Influence of Chronic Ozone Exposure on Global Carbon and Water Cycles. *Journal of Climate*, *28*(1), 292–305.

Lombardozzi, D. L., Levis, S., Bonan, G. B., & Sparks, J. P. (2012). Predicting photosynthesis and transpiration responses to ozone: decoupling modeled photosynthesis and stomatal conductance. *Biogeosciences*, *9*(8), 3113–3130.

Lombardozzi, D. L., Sparks, J. P., & Bonan, G. B. (2013). Integrating O3 influences on terrestrial processes: photosynthetic and stomatal response data available for regional and global modeling. *Biogeosciences*, *10*(11), 6815–6831.

Marín-Spiotta, E., Barnes, R. T., Berhe, A. A., Hastings, M. G., Mattheis, A., Schneider, B., &

Williams, B. M. (2020). Hostile climates are barriers to diversifying the geosciences. *Diversity and Equality in the Geosciences (EGU2019 EOS6.1 & US4, AGU2018 ED41B, JpGU2019 U-02)* , *53*, 117–127.

Martin, L. J., Blossey, B., & Ellis, E. (2012). Mapping where ecologists work: biases in the global distribution of terrestrial ecological observations. *Frontiers in Ecology and the Environment*, *10*(4), 195–201.

Mattheis, A., Murphy, M., & Marin-Spiotta, E. (2019). Examining intersectionality and inclusivity in geosciences education research: A synthesis of the literature 2008–2018. *Journal of Geoscience Education*, *67*(4), 505–517.

Medlyn, B. E., Zaehle, S., De Kauwe, M. G., Walker, A. P., Dietze, M. C., Hanson, P. J., Hickler, T., Jain, A. K., Luo, Y., Parton, W., Prentice, I. C., Thornton, P. E., Wang, S., Wang, Y.-P., Weng, E., Iversen, C. M., McCarthy, H. R., Warren, J. M., Oren, R., & Norby, R. J. (2015). Using ecosystem experiments to improve vegetation models. *Nature Climate Change*, *5*(6), 528–534.

Mercado, L. M., Medlyn, B. E., Huntingford, C., Oliver, R. J., Clark, D. B., Sitch, S., Zelazowski, P., Kattge, J., Harper, A. B., & Cox, P. M. (2018). Large sensitivity in land carbon storage due to geographical and temporal variation in the thermal response of photosynthetic capacity. *The New Phytologist*, *218*(4), 1462–1477.

Metcalfe, D. B., Hermans, T. D. G., Ahlstrand, J., Becker, M., Berggren, M., Björk, R. G., Björkman, M. P., Blok, D., Chaudhary, N., Chisholm, C., Classen, A. T., Hasselquist, N. J., Jonsson, M., Kristensen, J. A., Kumordzi, B. B., Lee, H., Mayor, J. R., Prevéy, J., Pantazatou, K., … Abdi, A. M. (2018). Patchy field sampling biases understanding of climate change impacts across the Arctic. *Nature Ecology & Evolution*, *2*(9), 1443–1448.

Morales, N., Bisbee O'Connell, K., McNulty, S., Berkowitz, A., Bowser, G., Giamellaro, M., & Miriti, M. N. (2020). Promoting inclusion in ecological field experiences: Examining and overcoming barriers to a professional rite of passage. *Bulletin of the Ecological Society of America*, *101*(4). https://doi.org/10.1002/bes2.1742

Morford, S. L., Houlton, B. Z., & Dahlgren, R. A. (2011). Increased forest ecosystem carbon and nitrogen storage from nitrogen rich bedrock. *Nature*, *477*(7362), 78–81.

Nancarrow, S. A., Booth, A., Ariss, S., Smith, T., Enderby, P., & Roots, A. (2013). Ten principles of good interdisciplinary team work. *Human Resources for Health*, *11*, 19.

O′Rourke, M., Crowley, S., Eigenbrode, S. D., & Wulfhorst, J. D. (2013). *Enhancing Communication & Collaboration in Interdisciplinary Research*. SAGE Publications.

Prentice, I. C., Dong, N., Gleason, S. M., Maire, V., & Wright, I. J. (2014). Balancing the costs of carbon gain and water transport: testing a new theoretical framework for plant functional ecology. *Ecology Letters*, *17*(1), 82–91.

Prentice, I. C., Liang, X., Medlyn, B. E., & Wang, Y.-P. (2015). Reliable, robust and realistic: the three R's of next-generation land-surface modelling. *Atmospheric Chemistry and Physics*, *15*, 5987–6005.

Reed, S. C., Yang, X., & Thornton, P. E. (2015). Incorporating phosphorus cycling into global modeling efforts: a worthwhile, tractable endeavor. *The New Phytologist*. https://doi.org/10.1111/nph.13521

Richardson, L. F. (1922). *Weather Prediction by Numerical Process*. Cambridge University Press.

Rogers, A., Medlyn, B. E., & Dukes, J. S. (2014). Improving representation of photosynthesis in Earth System Models. *The New Phytologist*, *204*(1), 12–14.

Rogers, A., Medlyn, B. E., Dukes, J. S., Bonan, G., von Caemmerer, S., Dietze, M. C., Kattge, J., Leakey, A. D. B., Mercado, L. M., Niinemets, Ü., Prentice, I. C., Serbin, S. P., Sitch, S., Way, D. A., & Zaehle, S. (2017). A roadmap for improving the representation of photosynthesis in Earth system models. *The New Phytologist*, *213*(1), 22–42.

Sagan, C., Toon, O. B., & Pollack, J. B. (1979). Anthropogenic Albedo Changes and the Earth's Climate. *Science*, *206*(4425), 1363–1368.

Sato, N., Sellers, P. J., Randall, D. A., Schneider, E. K., Shukla, J., Kinter, J. L., Hou, Y.-T., & Albertazzi, E. (1989). Effects of Implementing the Simple Biosphere Model in a General Circulation Model. *Journal of the Atmospheric Sciences*, *46*(18), 2757–2782.

Schneider, S. H., & Dickinson, R. E. (1974). Climate modeling. *Reviews of Geophysics*, *12*(3), 447. https://doi.org/10.1029/rg012i003p00447

Scott, H. G., & Smith, N. G. (2021). A model of C4 photosynthetic acclimation based on least-cost optimality theory suitable for Earth System Model incorporation. *Earth and Space Science Open Archive ESSOAr; Washington*. https://doi.org/10.1002/essoar.10505842.1

Sellers, P. J., Mintz, Y., Sud, Y. C., & Dalcher, A. (1986). A Simple Biosphere Model (SIB) for Use within General Circulation Models. *Journal of the Atmospheric Sciences*, *43*(6), 505–

531.

Sellers, P. J., Randall, D. A., Collatz, G. J., Berry, J. A., Field, C. B., Dazlich, D. A., Zhang, C., Collelo, G. D., & Bounoua, L. (1996). A Revised Land Surface Parameterization (SiB2) for Atmospheric GCMS. Part I: Model Formulation. *Journal of Climate*, *9*(4), 676–705.

Shi, Z., Crowell, S., Luo, Y., & Moore, B., 3rd. (2018). Model structures amplify uncertainty in predicted soil carbon responses to climate change. *Nature Communications*, *9*(1), 2171.

Shukla, J., & Mintz, Y. (1982). Influence of Land-Surface Evapotranspiration on the Earth's Climate. *Science*, *215*(4539), 1498–1501.

Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J. O., Levis, S., Lucht, W., Sykes, M. T., Thonicke, K., & Venevsky, S. (2003). Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Global Change Biology*, *9*(2), 161–185.

Smith, N. G., & Dukes, J. S. (2013). Plant respiration and photosynthesis in global-scale models: incorporating acclimation to temperature and CO2. *Global Change Biology*, *19*(1), 45–63.

Smith, N. G., & Keenan, T. F. (2020). Mechanisms underlying leaf photosynthetic acclimation to warming and elevated CO2 as inferred from least-cost optimality theory. *Global Change Biology*, *26*(9), 5202–5216.

Smith, N. G., Keenan, T. F., Prentice, I. C., Wang, H., Wright, I. J., Niinemets, Ü., Crous, K. Y., Domingues, T. F., Guerrieri, R., Yoko Ishida, F., Kattge, J., Kruger, E. L., Maire, V., Rogers, A., Serbin, S. P., Tarvainen, L., Togashi, H. F., Townsend, P. A., Wang, M., … Zhou, S.-X. (2019). Global photosynthetic capacity is optimized to the environment. *Ecology Letters*, *22*(3), 506–517.

Smith, N. G., Lombardozzi, D. L., Tawfik, A., Bonan, G. B., & Dukes, J. S. (2017). Biophysical consequences of photosynthetic temperature acclimation for climate. *Journal of Advances in Modeling Earth Systems*, *9*(1), 536–547.

Smith, N. G., Malyshev, S. L., Shevliakova, E., Kattge, J., & Dukes, J. S. (2015). Foliar temperature acclimation reduces simulated carbon sensitivity to climate. *Nature Climate Change*, *6*(4), 407–411.

Sulman, B. N., Moore, J. A. M., Abramoff, R. Z., Averill, C., Kivlin, S., Georgiou, K., Sridhar, B., Hartman, M., Wang, G., Wieder, W. R., Bradford, M. A., Luo, Y., Mayes, M. A., Morrison, E., Riley, W. J., Salazar, A., Schimel, J. P., Tang, J., & Classen, A. T. (2018).

Multiple models and experiments underscore large uncertainty in soil carbon dynamics. *Ecology Letters*, *14*, 109–123.

Thomas, R. Q., Zaehle, S., Templer, P. H., & Goodale, C. L. (2013). Global patterns of nitrogen limitation: confronting two global biogeochemical models with observations. *Global Change Biology*, *19*(10), 2986–2998.

Thornton, P. E., Doney, S. C., Lindsay, K., Moore, J. K., Mahowald, N., Randerson, J. T., Fung, I., Lamarque, J.-F., Feddema, J. J., & Lee, Y.-H. (2009). Carbon-nitrogen interactions regulate climate-carbon cycle feedbacks: results from an atmosphere-ocean general circulation model. *Biogeosciences Discussions*, *6*, 2099–2120.

Thornton, P. E., Lamarque, J.-F., Rosenbloom, N. A., & Mahowald, N. M. (2007). Influence of carbon-nitrogen cycle coupling on land model response to CO2fertilization and climate variability. *Global Biogeochemical Cycles*, *21*(4). https://doi.org/10.1029/2006gb002868

Torn, M. S., Trumbore, S. E., Chadwick, O. A., Vitousek, P. M., & Hendricks, D. M. (1997). Mineral control of soil organic carbon storage and turnover. *Nature*, *389*(6647), 170–173.

Wang, H., Atkin, O. K., Keenan, T. F., Smith, N. G., Wright, I. J., Bloomfield, K. J., Kattge, J., Reich, P. B., & Prentice, I. C. (2020). Acclimation of leaf respiration consistent with optimal photosynthetic capacity. *Global Change Biology*. https://doi.org/10.1111/gcb.14980

Wang, H., Prentice, I. C., Davis, T. W., Keenan, T. F., Wright, I. J., & Peng, C. (2017). Photosynthetic responses to altitude: an explanation based on optimality principles. *The New Phytologist*, *213*(3), 976–982.

Wang, H., Prentice, I. C., Keenan, T. F., Davis, T. W., Wright, I. J., Cornwell, W. K., Evans, B. J., & Peng, C. (2017). Towards a universal model for carbon dioxide uptake by plants. *Nature Plants*, *3*(9), 734–741.

Wang, K., Peng, C., Zhu, Q., Zhou, X., Wang, M., Zhang, K., & Wang, G. (2017). Modeling global soil carbon and soil microbial carbon by integrating microbial processes into the ecosystem process model TRIPLEX-GHG. *Journal of Advances in Modeling Earth Systems*, *9*(6), 2368–2384.

Wang, Y. P., Law, R. M., & Pak, B. (2010). A global model of carbon, nitrogen and phosphorus cycles for the terrestrial biosphere. *Biogeosciences*, *7*(7), 2261–2282.

Wang, Y., Zhang, H., Ciais, P., Goll, D., Huang, Y., Wood, J. D., Ollinger, S. V., Tang, X., & Prescher, A. (2021). Microbial activity and root carbon inputs are more important than soil carbon diffusion in simulating soil carbon profiles. *Journal of Geophysical Research. Biogeosciences*, *126*(4). https://doi.org/10.1029/2020jg006205

Washington, W. M., Buja, L., & Craig, A. (2009). The computational future for climate and Earth system models: on the path to petaflop and beyond. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *367*(1890), 833–846.

Weng, E. S., Malyshev, S., Lichstein, J. W., Farrior, C. E., Dybzinski, R., Zhang, T., Shevliakova, E., & Pacala, S. W. (2015). Scaling from individual trees to forests in an Earth system modeling framework using a mathematically tractable model of height-structured competition. *Biogeosciences*, *12*(9), 2655–2694.

Wieder, W. R., Allison, S. D., Davidson, E. A., Georgiou, K., Hararuk, O., He, Y., Hopkins, F., Luo, Y., Smith, M. J., Sulman, B. N., Todd-Brown, K., Wang, Y.-P., Xia, J., & Xu, X. (2015). Explicitly representing soil microbial processes in Earth system models. *Global Biogeochemical Cycles*, *29*, 1782–1800.

Wieder, W. R., Cleveland, C. C., Lawrence, D. M., & Bonan, G. B. (2015). Effects of model structural uncertainty on carbon cycle projections: biological nitrogen fixation as a case study. *Environmental Research Letters*, *10*(4), 044016.

Wieder, W. R., Grandy, A. S., Kallenbach, C. M., & Bonan, G. B. (2014). Integrating microbial physiology and physio-chemical principles in soils with the MIcrobial-MIneral Carbon Stabilization (MIMICS) model. *Biogeosciences*, *11*(14), 3899–3917.

Wieder, W. R., Grandy, A. S., Kallenbach, C. M., Taylor, P. G., & Bonan, G. B. (2015). Representing life in the Earth system with soil microbial functional traits in the MIMICS model. *Geoscientific Model Development*, *8*(6), 1789–1808.

Wieder, W. R., Hartman, M. D., Sulman, B. N., Wang, Y. P., Koven, C. D., & Bonan, G. B. (2018). Carbon cycle confidence and uncertainty: Exploring variation among soil biogeochemical models. *Global Change Biology*, *24*(4), 1563–1579.

Wieder, W. R., Pierson, D., Earl, S., Lajtha, K., Baer, S., Ballantyne, F., Berhe, A. A., Billings, S., Brigham, L. M., Chacon, S. S., & Others. (2020). SoDaH: the SOils DAta Harmonization database, an open-source synthesis of soil data from research networks, version 1.0. *Earth System Science Data Discussions*, 1–19.

Wieder, W. R., Sulman, B. N., Hartman, M. D., Koven, C. D., & Bradford, M. A. (2019). Arctic soil governs whether climate change drives global losses or gains in soil carbon.

Geophysical Research Letters, *46*(24), 14486–14495.

Yang, X., Thornton, P. E., Ricciuto, D. M., & Post, W. M. (2014). The role of phosphorus dynamics in tropical forests - A modeling study using CLM-CNP. *Biogeosciences*, *11*(6), 1667–1681.

Zaehle, S., & Friend, A. D. (2010). Carbon and nitrogen cycle dynamics in the O-CN land surface model: 1. Model description, site-scale evaluation, and sensitivity to parameter estimates. *Global Biogeochemical Cycles*, *24*(1), 1–13.

Zaehle, S., Medlyn, B. E., De Kauwe, M. G., Walker, A. P., Dietze, M. C., Hickler, T., Luo, Y., Wang, Y.-P., El-Masri, B., Thornton, P., Jain, A., Wang, S., Warlind, D., Weng, E., Parton, W., Iversen, C. M., Gallet-Budynek, A., McCarthy, H., Finzi, A., … Norby, R. J. (2014). Evaluation of 11 terrestrial carbon-nitrogen cycle models against observations from two temperate Free-Air CO2 Enrichment studies. *The New Phytologist*, *202*(3), 803–822.

Zhang, H., Goll, D. S., Wang, Y.-P., Ciais, P., Wieder, W. R., Abramoff, R., Huang, Y., Guenet, B., Prescher, A.-K., Viscarra Rossel, R. A., Barré, P., Chenu, C., Zhou, G., & Tang, X. (2020). Microbial dynamics and soil physicochemical properties explain large-scale variations in soil organic carbon. *Global Change Biology*, *26*(4), 2668–2685.

Ziehn, T., Kattge, J., Knorr, W., & Scholze, M. (2011). Improving the predictability of global CO2 assimilation rates under climate change. *Geophysical Research Letters*, *38*(10). https://doi.org/10.1029/2011gl047182

Figures

Table 1. Glossary of commonly used words in Earth System Modeling.

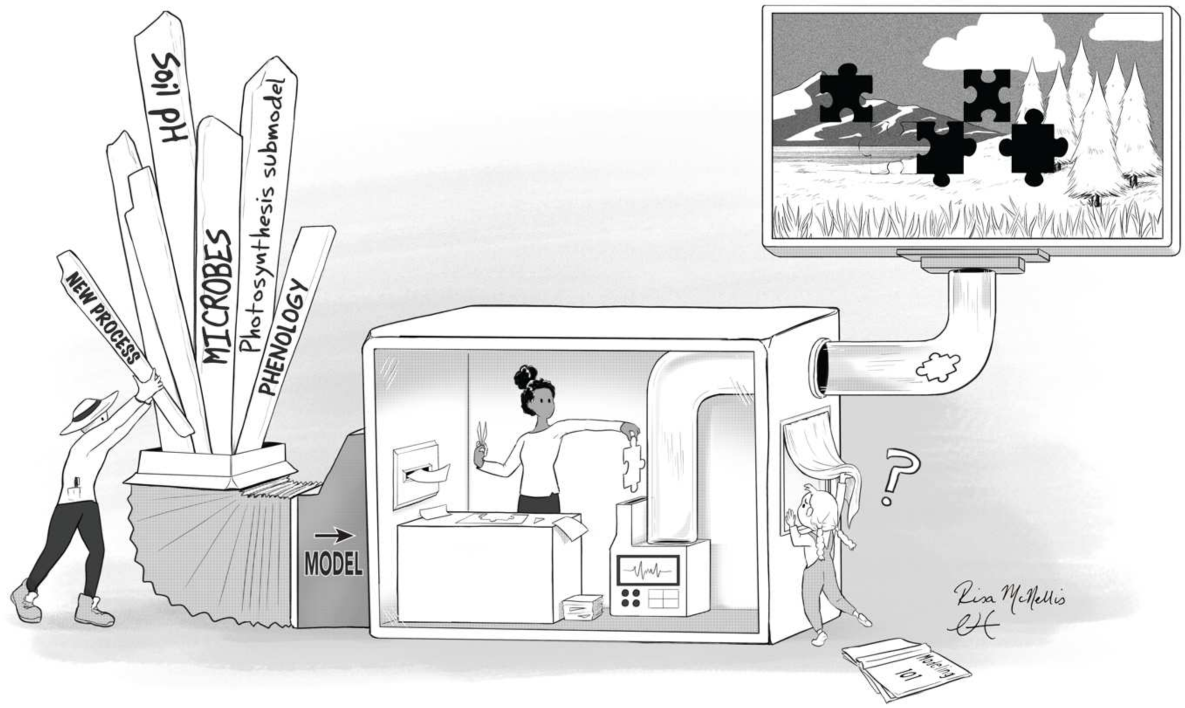
Term	Definition
Benchmarking	Comparing models against a consistent set of observational data to document the performance of multiple models or improvements with newer versions of a particular model.
Calibration	Setting or adjusting model parameters based on model performance against a training dataset. Separate from validation.
Data assimilation	Adjusting model states at regular time intervals based on observations.
Ensemble	Multiple model simulations from one or more models that follow a standard protocol, including "multi-model" ensembles of multiple models and "multi-member" ensembles that differ in initial conditions or parameter values. Ensembles are used to understand model variability and uncertainty.

Equifinality	The ability of multiple model configurations or parameter sets to explain the same set of observations.
Evaluation	Assessing model performance, often using a validation or benchmarking approach.
Feature fatigue	The continual addition of new model processes, often with diminishing returns on model performance.
Fluxes	Movement of matter or energy between the components of a model. Alternatively: flows.
Forcing	Driver inputs external to a model.
Forecasting	A type of prediction that generates model outputs of future conditions based on current knowledge and initial states.
Modularity	A property of models in which one representation of a process can be swapped out for another to allow comparison of model formulations.
Parameter	Constant within an equation in a model.
Parameterize	To represent a complex process as a simplified equation that relates parameters and variables to one another.
Parsimony	Avoiding unnecessary model complexity; only including those model components that contribute to the goals of model development.
Prediction	Model outputs beyond the scope of observed data.
Projection	Model outputs based on a certain scenario or set of conditions occurring as represented in the forcing data.
Realism	The adherence of model representations to the actual properties and behavior of ecosystems.
Sensitivity	How model output changes in response to shifts in inputs or individual model parameters.
States	The current values of components of a model system, which typically change through time. For example, soil moisture, soil temperature, biogeochemical pools.
Toy model	A simple model that allows for exploration of a subset of ecosystem processes.
Traceability	The ability to connect model sensitivity or uncertainty back to a particular model component.
Trait	Property of an ecosystem component that maps onto model parameters.
Validation	Evaluating model performance against an independent dataset without modifying parameters. Separate from calibration.

Table 2. Table of textbooks and free resources for developing cross-disciplinary skill sets in empirical and modeling work and learning to traverse the stages of integrating new processes into an Earth System model. For a regularly updated list of resources, visit https://ecoesm.github.io/.

Skill/ Category	Item	Description	Link
Programming	NCAR Python tutorials	Basic introduction to the Python language from the National Center for	https://ncar.github.io/python-tutorial/

		Atmospheric Research	
Programming	PEcAn project tutorials	Introduction to working with the Predictive Ecosystem Analyzer	https://pecanproject.github.io/tutorials.html
Programming	The Unix Shell	The basics of file systems and the shell	http://swcarpentry.github.io/shell-novice/
Programming	Udacity	Free courses on basic programming competency with github, linux, R, python, and many others	https://www.udacity.com/
Programming	Software Carpentry	Free courses on basic programming competency with github, linux, R, python, and many others	https://software-carpentry.org/lessons/index.html
Programming	R tutorial	Basic introduction to working with R	https://education.rstudio.com/learn/beginner/
Simple modeling	InsightMaker	Tools for developing quantitative stock-and-flow diagrams of processes	https://insightmaker.com/
Simple modeling	Teaching Resources	Lessons and other resources developed for teaching basic principles of ecological modeling	https://matthesecolab.com/teaching/ http://www.maryheskel.com/teaching.html https://onlinelibrary.wiley.com/doi/full/10.1002/ece3.6757
Simple modeling	Modeling the Environment	Textbook on environmental modeling by Andrew Ford	https://islandpress.org/books/modeling-environment-second-edition
Simple modeling	EDDIE	Modeling/forecasting teaching modules developed for NEON sites	https://serc.carleton.edu/eddie/macrosystems/index.html
Simple modeling	Excel modeling tutorial	Tutorial on building simple models in Excel	http://www.mbaexcel.com/excel/how-to-build-an-excel-model-step-by-step/
Earth system modeling	Climate Change and Terrestrial Ecosystem Modeling	Textbook on global-scale ecosystem modeling by Gordon Bonan	https://www.cgd.ucar.edu/staff/bonan/ecomod/index.html https://www.cgd.ucar.edu/staff/bonan/ecoclim/index.html
Earth system modeling	CESM tutorial	Workshop on working with the Community Earth System Model	https://www.cesm.ucar.edu/events/tutorials/
Earth system modeling	Earth System Modeling Framework	Introduction to working with Earth System Models	https://earthsystemmodeling.org/tutorials/
Earth system modeling	CESM-Lab	Cloud version of CLM	https://github.com/NCAR/CESM-Lab-Tutorial



Dorothy pulled back the curtain to find that the model wasn't magic after all...

Figure 1. Historically, the process of integrating ecology in Earth System models (ESMs) has often separated tasks along disciplinary lines, with empirical ecologists feeding data into a mysterious "modeling" process and modelers modifying and using data without a thorough understanding of data collection procedures and caveats. The newest generation of scientists has the opportunity to pull back the curtain by developing cross-disciplinary skill sets and building stronger, more collaborative bridges between empirical and modeling communities, with the goal of accelerating the integration of ecological concepts into ESMs.

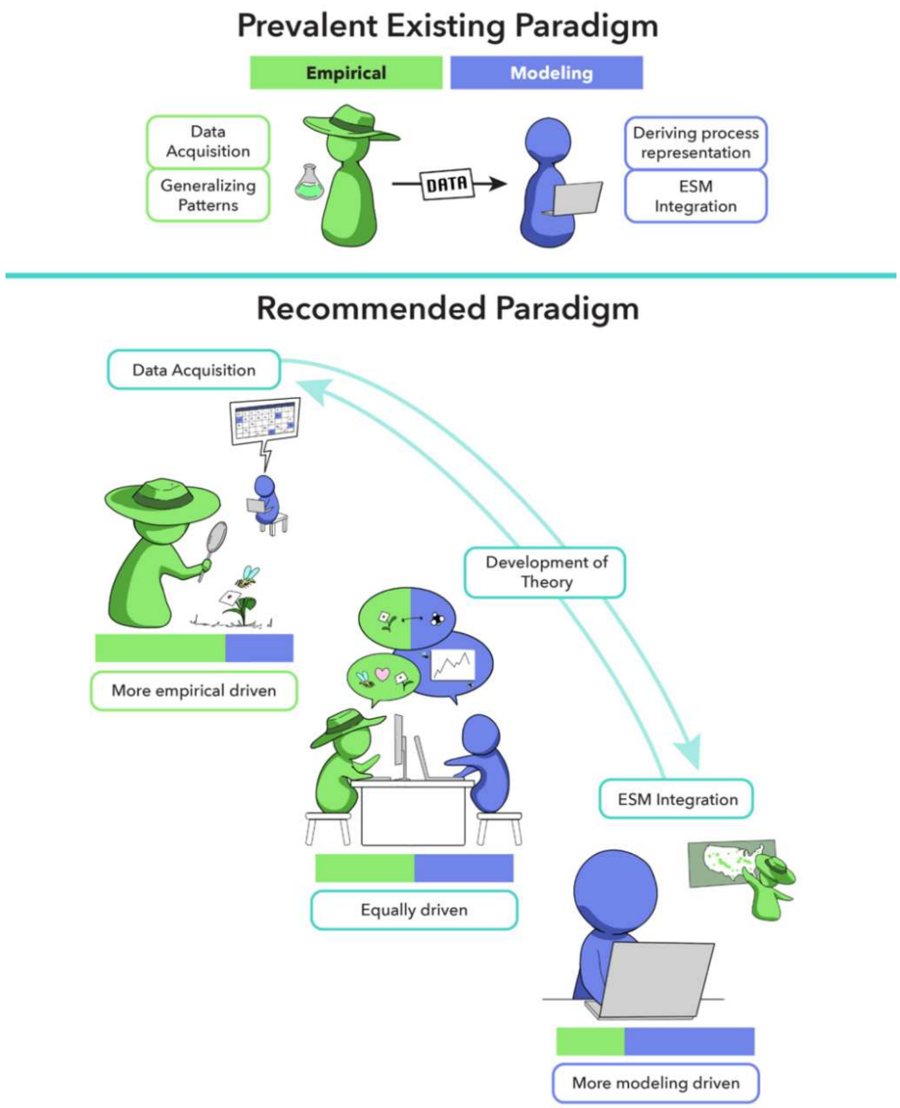


Figure 2. The prevalent existing paradigm in ecology-Earth System model (ESM) integration separates tasks along disciplinary lines, with empirical scientists giving data and generalized patterns to modelers who then develop quantitative models and work with ESMs. We recommend a shift away from this historical paradigm towards a more collaborative one in which empiricists and modelers are involved in co-producing knowledge (with differing degrees of contribution) at every stage of data collection, theory development, and model integration. We also emphasize the two-way exchange of ideas, insights, and data between empirical and modeling driven activities.

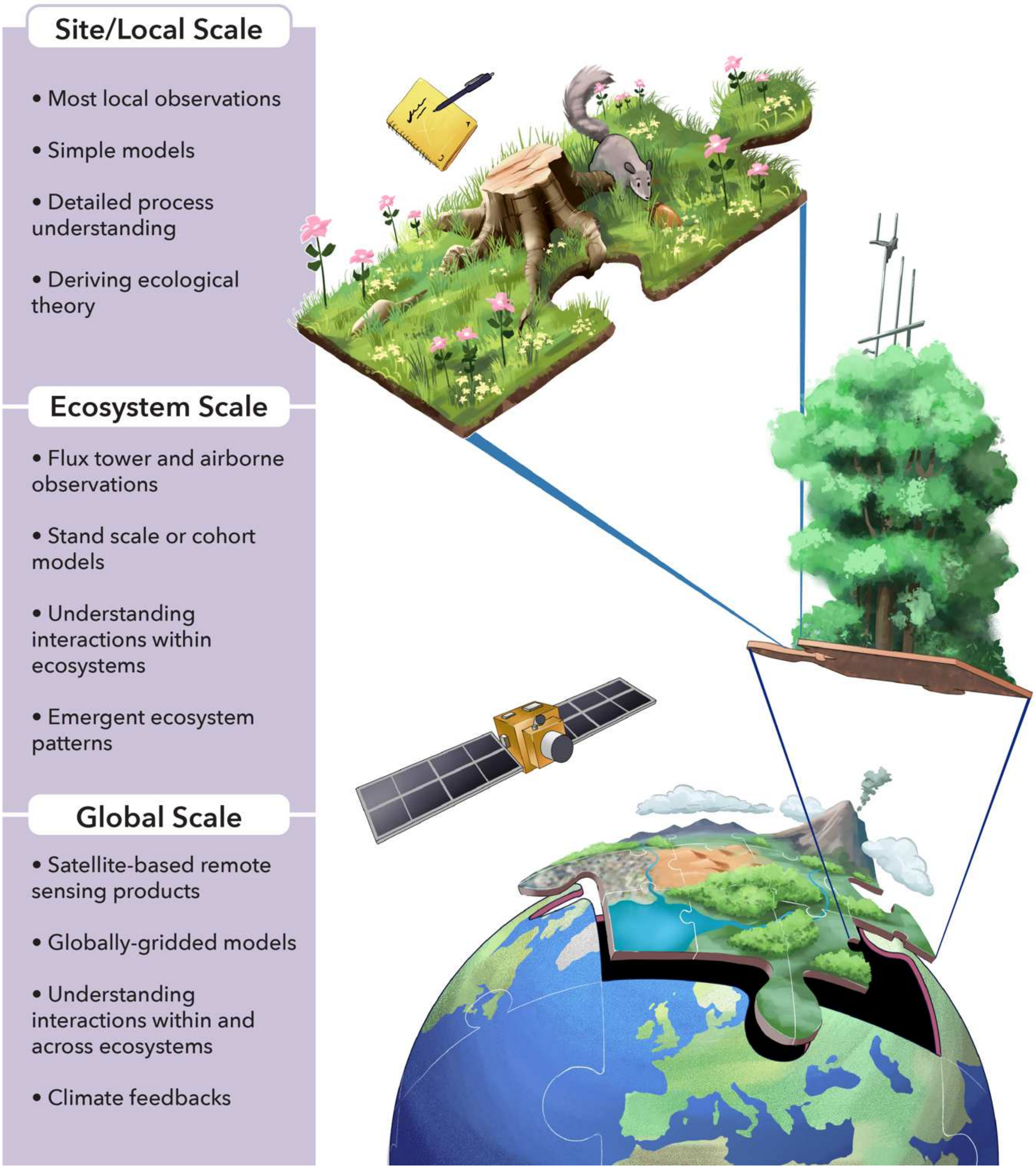


Figure 3. In the hierarchy of model development, simple models of individual processes, classes of organisms, and inorganic components (site/local scale) are often pieced together to form larger models of ecosystems and regions (ecosystem scale) and ultimately combined to form Earth system models (ESMs; global scale). Data gathered at each of these scales can be used to inform model development at the same scale.

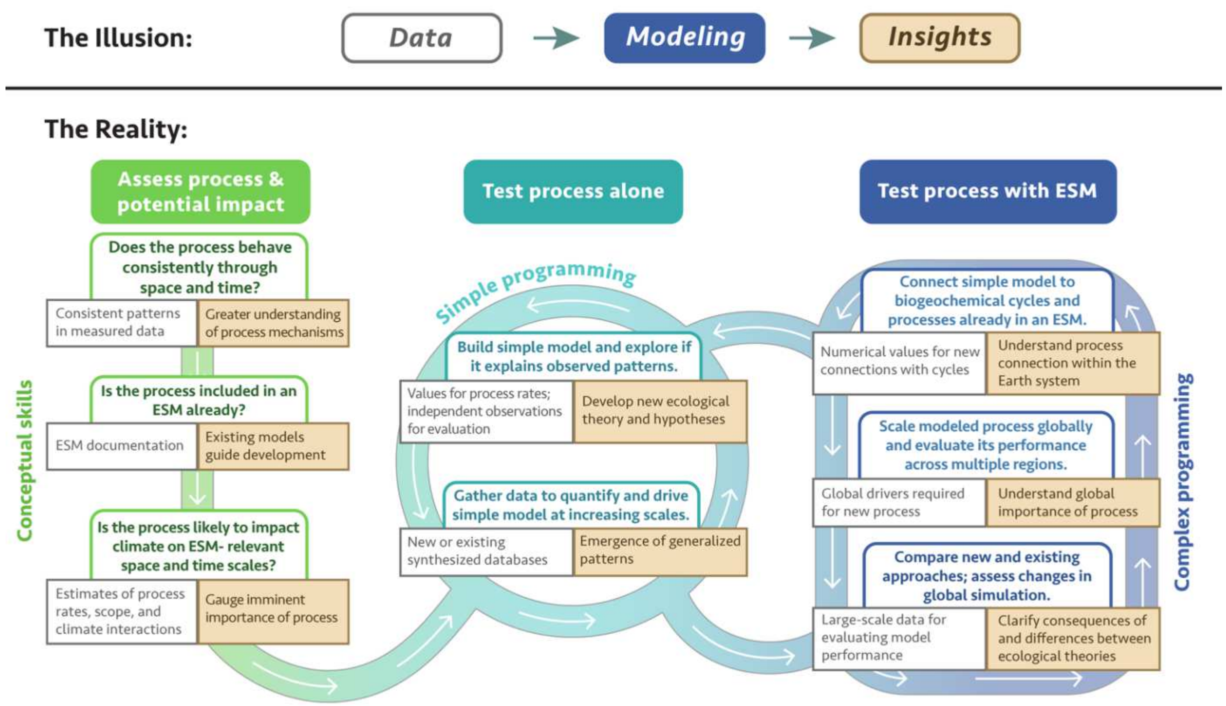
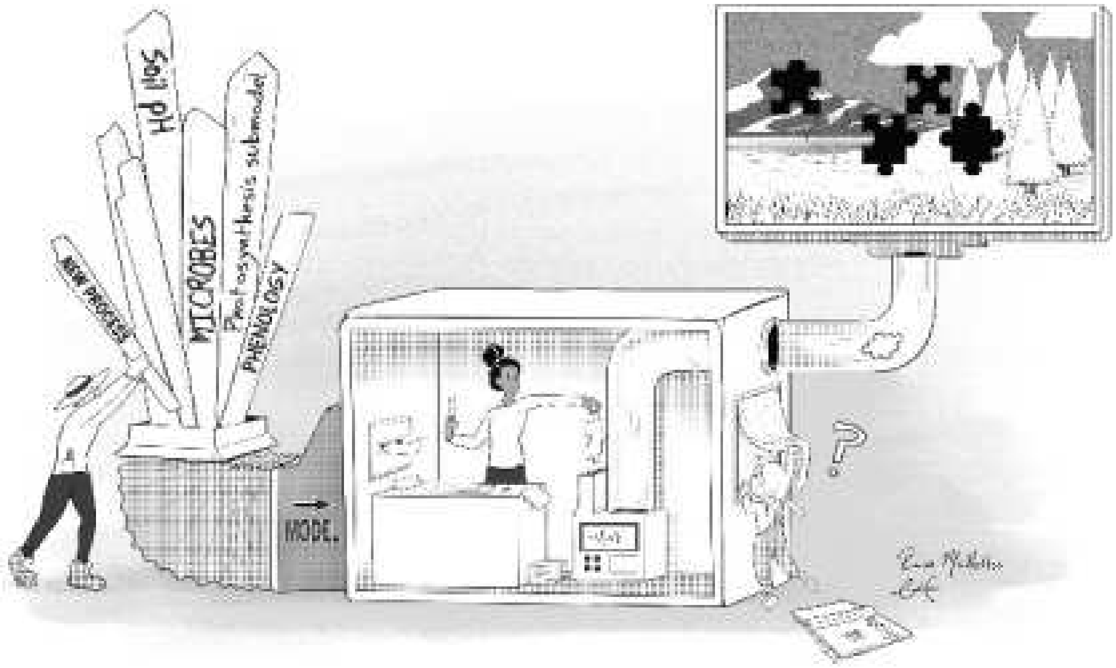
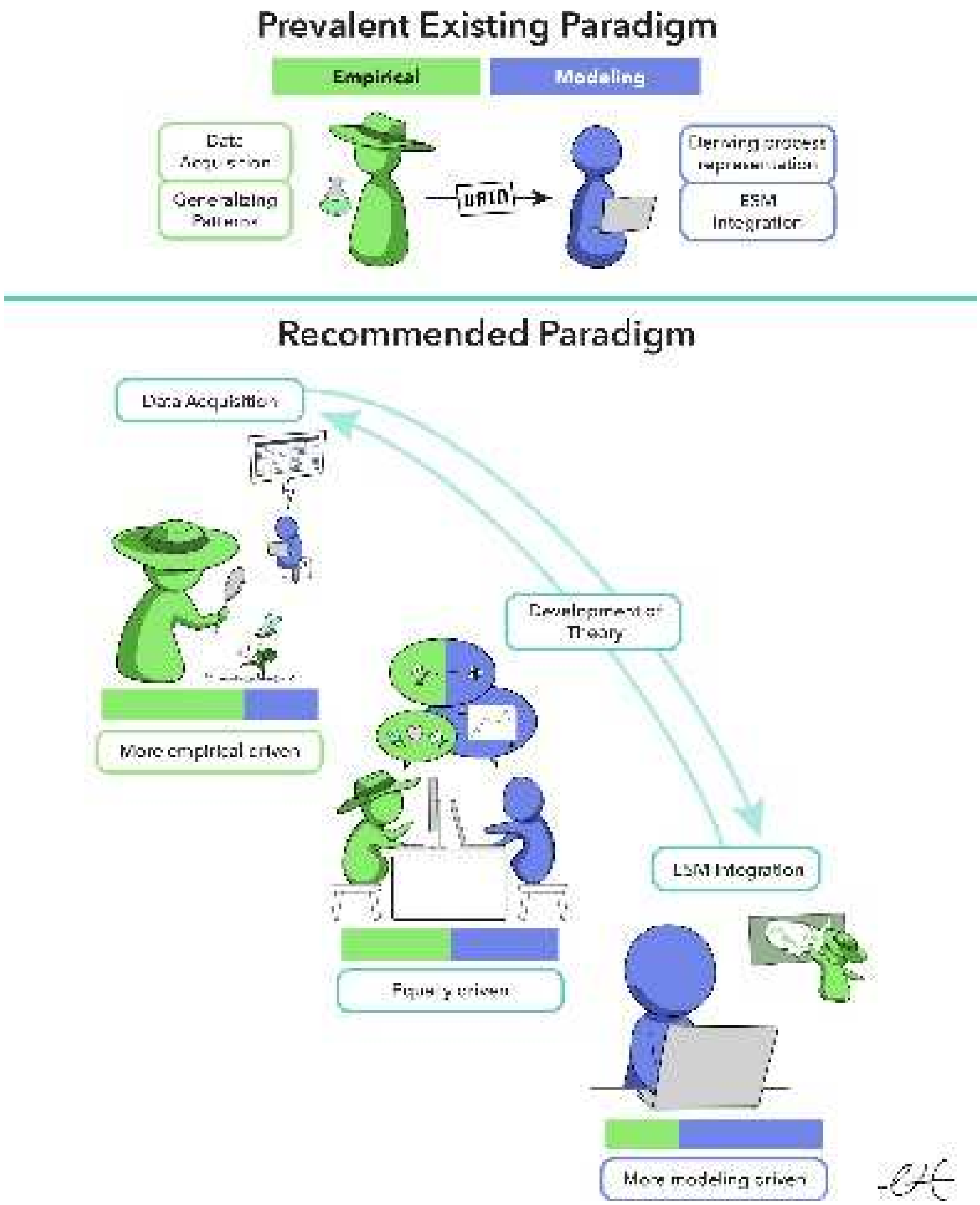


Figure 4. Although scientists sometimes think "The Illusion" (top panel) is the way that ecological concepts are integrated into Earth system models (ESMs), the reality is more like a complex metabolic cycle or eddy-filled stream, with different data inputs (gray boxes) and valuable insights (tan boxes) throughout the process. We identify three key phases in integrating a new process into an ESM, namely: "Assess process & potential impact", which emphasizes conceptual skills (green boxes), "Test process alone", which involves simple programming (teal), and "Test process with ESM", which involves more complex programing (blue). Within each phase, we offer specific questions to guide empiricists and modelers along the way.

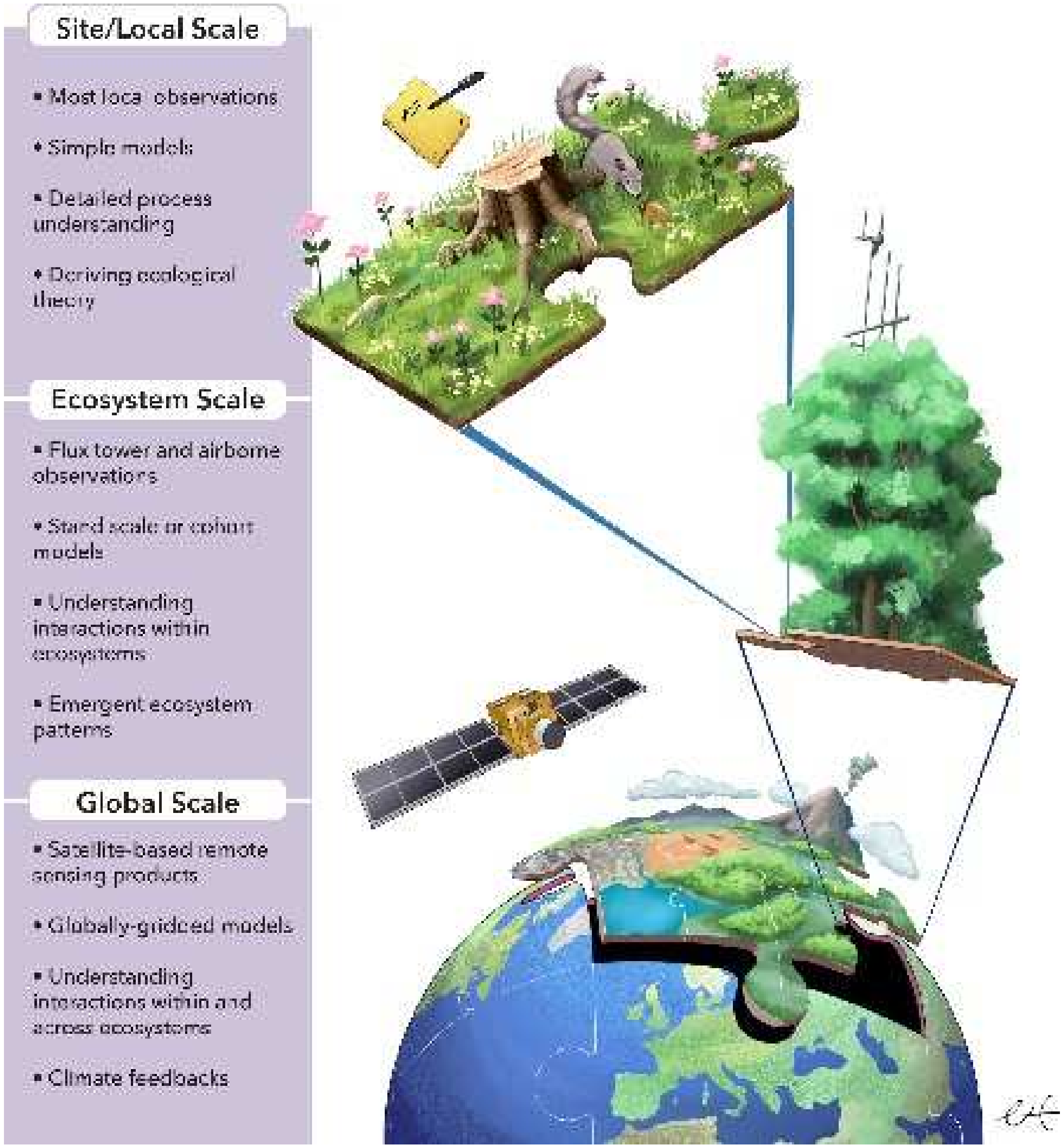


Dorothy pulled back the curtain to find that the model wasn't magic after all...

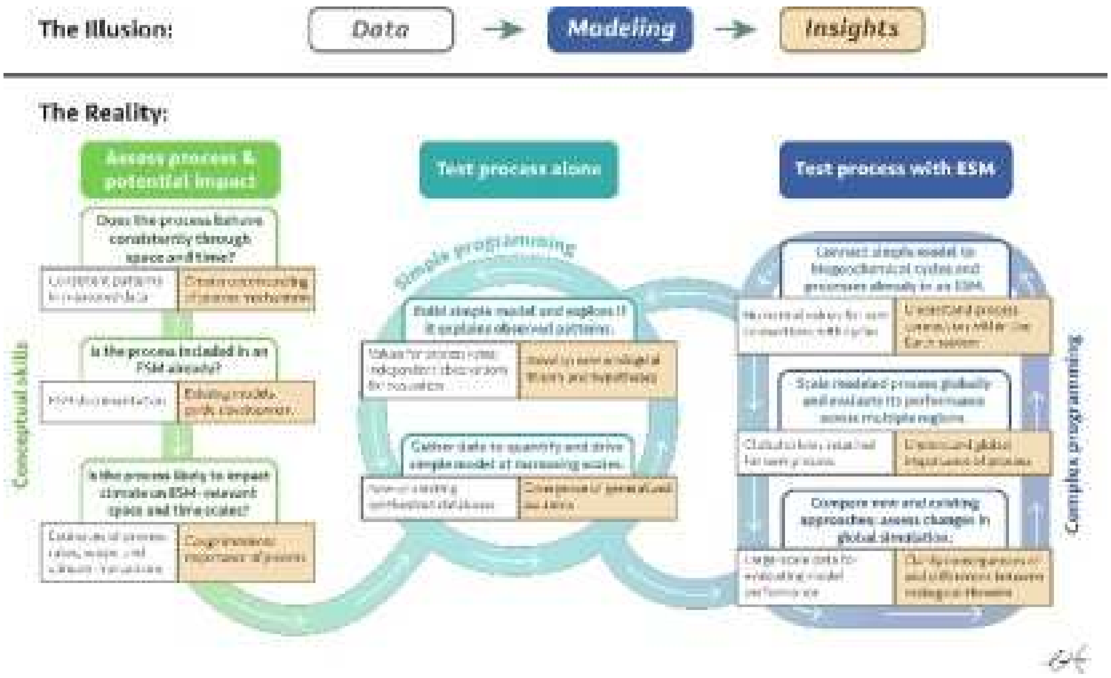
gcb_15894_f1.jpg



gcb_15894_f2.png



gcb_15894_f3.png



gcb_15894_f4.png