Automated Integration of Continental-Scale Observations in Near-Real Time for Simulation and Analysis of Biosphere–Atmosphere Interactions

David J. Durden[1(✉)], Stefan Metzger[1,2], Housen Chu[3], Nathan Collier[4], Kenneth J. Davis[5], Ankur R. Desai[2], Jitendra Kumar[4,6], William R. Wieder[7,8], Min Xu[4], and Forrest M. Hoffman[4,9]

[1] National Ecological Observatory Network Program, Battelle, Boulder, CO 80301, USA
{ddurden,smetzger}@battelleecology.org

[2] Department of Atmospheric and Oceanic Sciences, University of Wisconsin, Madison, WI 53706, USA
desai@aos.wisc.edu

[3] Lawrence Berkeley National Laboratory, Berkeley, CA 94702, USA
hchu@lbl.gov

[4] Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
nathaniel.collier@gmail.com, {jkumar,forrest}@climatemodeling.org, xum1@ornl.gov

[5] Department of Meteorology and Atmospheric Science, and Earth and Environmental Systems Institute, The Pennsylvania State University, University Park, PA 16802, USA
kjd10@psu.edu

[6] Bredesen Center, University of Tennessee, Knoxville, TN 37996, USA

[7] Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, CO 80307, USA
wwieder@ucar.edu

[8] Institute of Arctic and Alpine Research, University of Colorado, Boulder, CO 80309, USA

[9] Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN 37996, USA

Abstract. The National Ecological Observatory Network (NEON) is a continental-scale observatory with sites across the US collecting standardized ecological observations that will operate for multiple decades. To maximize the utility of NEON data, we envision edge computing systems that gather, calibrate, aggregate, and ingest measurements in an integrated fashion. Edge systems will employ machine learning methods to cross-calibrate, gap-fill and provision data in near-real time to the NEON Data Portal and to High Performance Computing (HPC) systems, running ensembles of Earth system models (ESMs) that assimilate the data. For the first time gridded EC data products and response functions promise to offset pervasive observational biases through evaluating,

J. Nichols et al. (Eds.): SMC 2020, CCIS 1315, pp. 204–225, 2020.
https://doi.org/10.1007/978-3-030-63393-6_14

benchmarking, optimizing parameters, and training new machine learning parameterizations within ESMs all at the same model-grid scale. Leveraging open-source software for EC data analysis, we are already building software infrastructure for integration of near-real time data streams into the International Land Model Benchmarking (ILAMB) package for use by the wider research community. We will present a perspective on the design and integration of end-to-end infrastructure for data acquisition, edge computing, HPC simulation, analysis, and validation, where Artificial Intelligence (AI) approaches are used throughout the distributed workflow to improve accuracy and computational performance.

Keywords: Data-model integration · Eddy-covariance · Environmental observatory · National Ecological Observatory Network (NEON) · Edge computing systems · High performance computing · Earth system models · Land surface models · Model benchmarking · International Land Model Benchmarking (ILAMB)

1 Introduction

Advanced computational resources and new algorithmic developments have extended our environmental understanding over the past few decades. Now, an unprecedented volume of standardized observational data products (ODPs) are being realized through the National Ecological Observatory Network (NEON). NEON collects environmental and biological data with in situ sensors, observational sampling, and aerial overflights. Core components of NEON infrastructure are 47 tower sites, where eddy-covariance (EC) sensors are used to determine the surface–atmosphere exchange of momentum, heat, water, and carbon dioxide to assess interactions at the soil–vegetation–atmosphere interface. This continental-scale data set, having numerous contextual observations available in near-real time, affords new data-model integration opportunities to leverage such observations for new scientific understanding and to potentially enable viable ecological forecasting capabilities. This paper explores several ways that continued development of data-model integration, through new measurements, synthesized ODPs, and access to near-real-time data, contributes to improved scientific understanding of ecosystem processes and advances efforts to constrain uncertainty in Earth system models (ESMs) and subsequent benchmarking. First, we provide a background for the potential of data-model integration, the state of ESMs and benchmarking, and the growth of network-scale observations. Next, we discuss our vision for integrating network observations to improve model predictive capabilities, minimize prediction uncertainties, and advance forecast accuracy with scale-aware ODPs and near-real time data. Lastly, the roadmap to accomplishing our stated goals is outlined with considerations of emerging technologies that have the potential to broaden our goals.

1.1 Improving Scientific Understanding Through Data-Model Integration

Data-model integration is quickly becoming a fundamental component in efforts to evaluate and enhance our capabilities to simulate Earth system processes (Fer et al. 2018). Data-model integration improvements can be realized through improved parameterization of initial conditions, data assimilation techniques to inform model states or parameters during simulations, and comprehensive benchmarking of model structure and evaluation against observations (Dietze et al. 2014; Zobitz et al. 2011). Network-scale observations of ecosystem functions, such as surface-atmosphere exchange (SAE) of energy, water vapor, and trace gases, have historically (Stöckli et al. 2008) and continue to lead to novel advances in model performance (Fer et al. 2018).

Improved Model Optimization and Benchmarking

Additional Contextual Observations. Optimized model parameterization or constraints via data assimilation typically targets periods or conditions when model uncertainty is greatest. Enhanced access to numerous contextual observations can inform underlying model processes or elucidate missing information. Data assimilation constrains model predictions by comparing model output with ODPs, determining probabilistic differences, and advancing ensemble members with informed posteriors. The improved availability of repeated and interoperable in-situ, reanalysis, and remote sensing data with quantified uncertainty for weighting in assimilation and model benchmark scoring is expected to facilitate tuning process representations in ESMs and inform data providers of ODP requirements that are still unmet (Hoffman et al. 2017; Collier et al. 2018).

Resolving Scale Mismatch Between Simulations and Observations. Terrestrial ecosystem processes are widely recognized to be heterogeneous at spatial scales well below those resolved by most ESMs resulting in a spatial representativeness uncertainty when evaluating/informing models with single point observations (e.g., Riley and Shen 2014). Scaling has been shown to be non-linear with vegetation cover (e.g., Launiainen et al. 2016) and sensitive to resolution, scaling method, and the magnitude of heterogeneity (Wang et al. 2016; Liu et al. 2016). SAE observations based on the eddy-covariance (EC) flux technique (e.g., Aubinet et al. 2012) are one example of a process-scale benchmark for assessing the performance of ESMs (e.g., Fox et al. 2009; Williams et al. 2009; Schwalm et al. 2010; Schaefer et al. 2012) that suffers from such scale mismatch. Using site-based EC measurements for model benchmarking is thus complicated by biases arising from unmet assumptions on the observations. These include the limited and varying spatial representativeness of the observations at model grid scale (e.g., Chen et al. 2011; Griebel et al. 2020), and the observations violating the conservation of energy (e.g., Mauder et al. 2020). Both of these biases increase with spatial heterogeneity, which complicates regional-scale model benchmarking and improvement (e.g., Metzger 2018; Xu et al. 2020). Therefore, spatial scaling of site-based flux observations to ESM grid scales using multi-scale observations is needed to reduce uncertainties in flux estimates and constrain model benchmarking.

From Hindcasting to Forecasting. Ecosystem models are key to synthesizing process understanding, examining simulated ecosystem functioning against observations at local to regional scales, and can provide the scientific basis for field measurement campaigns (Dietze et al. 2014). The Predictive Ecosystem Analyzer (PEcAn) framework is a powerful ecoinformatics framework that utilizes Bayesian data assimilation techniques to inform models with ODPs. As such, PEcAn is a prime example of the synergistic improvements realized through data-model integration for both model parameterization and observational data requirements to reduce uncertainty (LeBauer et al. 2013; Dietze et al. 2013; Kattge et al. 2011). Access to low latency, repeated, and interoperable ODPs with quantified uncertainty is facilitating a movement to near-term ecological forecasting. These forecasts are envisioned to inform land-use decision makers with the most accurate predictions of ecosystem function via iterative model assessment and improvement through comparison with near-real-time data (Dietze et al. 2018). Similar model evaluation and benchmarking of ESMs can be realized; however, this approach likely involves a large number of perturbed parameter ensembles (PPE) of models or machine learning-based surrogate models running on high performance computing (HPC) systems.

1.2 Earth System Models and Benchmarking

Earth system models (ESMs) are designed to simulate the coupled multiscale, multiphysics processes associated with interactive dynamics, physics, chemistry, and biology across the land, ocean, sea ice, land ice, and atmosphere that drive the Earth's climate system (Randall et al. 2018). Originally conceived as models of physics and dynamics, focused primarily on atmosphere and ocean processes, early global climate models evolved into ESMs with the inclusion of terrestrial and marine ecosystem processes, atmospheric chemistry, and human system interactions (Bonan and Doney 2018; Flato 2011). Research with these coupled ESMs has demonstrated that the carbon cycle responds to climate but also that large nonlinear climate feedbacks are produced by the biosphere (Friedlingstein et al. 2001, 2006; Arora et al. 2013). Terrestrial ecosystems in ESMs are represented by a variety of vegetation types, an amount of leaf area, functioning of stomata in leaves, and carbon and nutrient pools that interact with energy and water cycles (Bonan 2016). Relatively simplistic representations of vegetation and soil processes in land surface models (LSMs), typically contained within coupled ESMs, capture the mean state behavior of plants and soils over large spatial scales on annual time scales. However, process understanding limits the ability to reduce errors and biases when compared with observational data at local scales (Schimel et al. 1997).

Forecasting ecosystem responses to environmental forcing is important for resource management and understanding impacts of rapid climate change or land use change (Clark et al. 2001; Foley et al. 2005; Luo et al. 2011). While long-term EC flux measurements help to constrain energy, water, and carbon cycles for individual biomes (Baldocchi et al. 2001), more rapid integration of these data with models—employing data assimilation and benchmarking tools for uncertainty quantification, parameter optimization, and structural optimization—will improve understanding of these processes and lead to more mechanistic representations in models and more accurate ecosystem forecasts (Williams et al. 2009; Raupach et al. 2005).

LSMs rely on a collection of process representations, called parameterizations, embodied in numerical algorithms that employ many often-uncertain parameters to approximate the evolution of carbon, water, and energy in the natural world (Bonan 2019). Data assimilation methods are commonly used to calibrate and evaluate model accuracy and parameter uncertainty (Luo et al. 2011). Raupach et al. (2005) presented methods for assimilating diverse data and separating observational from model errors to produce more accurate forecasts of the global carbon cycle. These methods have been applied across scales, from global inversions (e.g., Ricciuto et al. 2008) to individual tree stands (e.g., Moore et al. 2008; Ricciuto et al. 2011), with a variety of approaches, including Kalman filters or ensemble Kalman filters (e.g., Quaife et al. 2008), other maximum likelihood techniques, and least squares optimization methods (e.g., Prihodko et al. 2008). Sophisticated data assimilation packages that ingest EC flux measurements are now being coupled directly to complex forward land surface models for use on HPC systems (Fox et al. 2018; Bastrikov et al. 2018). Perturbed physics ensembles (also called perturbed parameter ensembles) or PPEs employ thousands of ensemble simulations to develop an understanding of the sensitivity or importance of individual parameters or to quantify the impacts of their uncertainties on feedbacks, extremes, or model skill (Fischer et al. 2011; Sanderson et al. 2010). Conducting large numbers of ensemble simulations to search for optimal parameter combinations for complex ESMs has become so computationally intensive that in some cases surrogate models are being developed and used in place of running LSMs directly (Li et al. 2018; Lu et al. 2018). For example, Ricciuto et al. (2018) analyzed the sensitivity of five key carbon variables to 68 model parameters in the US Department of Energy's (DOE's) Energy Exascale Earth System Model (E3SM) land model using a global sensitivity analysis on 96 FLUXNET sites. Lu et al. (2018) further optimized 8 of 68 parameters of the E3SM land model using surrogate-based global optimization. Executing these direct or surrogate simulations is one part of the challenge; evaluating model results in a systematic fashion is another.

Systematic evaluation of model results, through comparison with observational data, is important for quantifying model fidelity (Randerson et al. 2009). As ESMs become more complex, routine assessment of model performance must be performed for verification of new parameterizations, evaluation of impacts on other model components, and validation of simulations under changing environmental conditions. The land modeling community has developed a variety of evaluation approaches for terrestrial carbon cycle models (Cadule et al. 2010;

Blyth et al. 2011; Abramowitz 2012; Anav et al. 2013; Piao et al. 2013). Some benchmarking approaches are based on an expected, pre-defined level of performance (Abramowitz 2005; Best et al. 2015), but most systematic benchmarking strategies produce a skill score based on a direct model-data comparison. Lack of standardized evaluation metrics and methods have limited adoption of model benchmarking and use of a wide diversity of observational data sets.

The International Land Model Benchmarking (ILAMB) project was organized to engage the research community in the development of standardized and internationally accepted benchmarks for land model performance. The ILAMB community aims to strengthen linkages among experimental, remote sensing, and climate modeling communities in the design of new model tests and new measurement programs, and supports the design and development of open source benchmarking tools through international workshops and working group activities (Hoffman et al. 2017). With support primarily from the US Department of Energy, community ILAMB activities have resulted in creation of an ILAMB benchmarking software package for evaluation of LSMs that incorporate biogeochemical cycles (Collier et al. 2018; Hoffman et al. 2017). The ILAMB package produces graphical and tabular diagnostics across a range of biogeochemistry, hydrology, radiation and energy, and forcing variables. It scores multi-model performance for period mean, bias, root-mean-square error (RMSE), spatial distribution, interannual coefficient of variation, seasonal cycle, and long-term trend. The design philosophy and details of its implementation and methodology are described by Collier et al. (2018). Efforts are underway to directly link ILAMB to PEcAn for more rapid assessment of site-level simulations over diurnal time scales. Being an open source and extensible package with a scalable design, so that it can run on the largest HPC systems, makes it a good choice for evaluating the results of ensemble simulations aimed at parameter optimization and uncertainty assessment.

1.3 Network-Scale Observations

Network-scale flux tower observations—such as those available from FLUXNET (Baldocchi et al. 2001), AmeriFlux (Novick et al. 2018), ICOS, TERN, or NEON (Metzger et al. 2019a)—are revolutionizing ecosystem science by providing observations that cover large spatial areas across a broad variety of eco-climatic zones. The proliferation of standardized and interoperable flux network ODPs through cross-network collaboration and integration strengthens the ability of observations to explain measured environmental variability. For instance, NEON provides data to AmeriFlux, which along with ICOS and TERN, feed into FLUXNET. However, limitations exist on standardized measurements across networks, and substantial latency can be incurred for fully quality controlled data sets with quantified uncertainties.

NEON is a continental-scale observatory with sites across the US that will operate for multiple decades. NEON produces data products, software, and services to facilitate research on the impacts of climate change, land-use change, and invasive species. NEON collects environmental and biological data with in-situ sensors, biometric observations, and aerial overflights. One of NEON's core components is its 47 tower sites, where EC sensors are used to determine the SAE of momentum, heat, water, and carbon dioxide to assess interactions at the soil–vegetation–atmosphere interface. These data are streamed from tower sites to a central NEON headquarters facility. There, calibration coefficients are applied, quality assurance and quality control are performed, and additional processing algorithms are applied to derive higher level data products. The resulting ODPs are served on the NEON data portal, currently with about a one month latency. The latency of biometric and airborne remote sensing data varies by ODP. One unique aspect of NEON ODPs is the standardization of sensor infrastructure, biometric protocols and algorithms for processing. This standardization and ubiquitous availability of "contextual" observations with respect to SAE processes, position NEON ODPs as a perfect test suite for ESM hypothesis testing and benchmarking.

2 Visions to Improve Model Performance with Network-Scale Observations

2.1 Scale-Aware Observational Data Products for ESM Evaluation

Improved understanding of model-data interfaces enables maximizing the usefulness of ODPs for ESM improvement. For data-model integration, we commonly rely on half-hourly intervals as the lowest common timestep denominator. That is, we expect both ODPs and models to capture in half-hourly slices the dynamics emerging from environmental processes at a much broader range of scales. From the observational perspective, inconsistencies arise when we interpret continuous, nonlinear environmental processes and non-symmetrical observation techniques through discrete data processing and analytics that assume linearity and Gaussianity. Resultant half-hourly ODPs may be biased on the order of several 10% due to space/time ambiguity associated with scaling (Xu et al. 2017), violation of energy conservation (Mauder et al. 2020), etc.: our models might perform better or worse than we think because we already know that our current ODP reference is off. Here we explore how we could rectify the situation by creating half-hourly ODPs that capture environmental processes at scales consistent with expectations for data-model integration.

A Complementary Benchmarking Framework. To resolve the scale mismatch between simulations and observations, participants of the DOE-funded 2019 RUBISCO-AmeriFlux Working Group Meeting (Hawkins et al. 2020) conceived a scale-aware benchmarking framework that complements top-down ODP constraints with bottom-up ODP process information across DOC, DOE, NASA and NSF projects (Fig. 1). The proposed approach will enable consistent regional-scale evaluations of carbon, water, and energy cycles in ESMs. At the center of the framework is the ILAMB package, which facilitates benchmarking ESMs in a modular fashion. The NCAR-NEON Community Land Model (CLM5) implementation is one example of an enhanced ESM module for use with ILAMB. Participants of the NSF-funded 2019 NCAR-NEON Workshop conceived an implementation of CLM5 that leverages an unprecedented range of contextual observations to constrain model uncertainty. In the past e.g. plot-based biometric observations, high-resolution airborne remote sensing, gas phase and water phase isotopes, replicate soil properties, as well as aquatic properties in adjacent lakes and streams have not been uniformly available at the flux tower network scale. With the advent of NEON these contextual observations are routinely available alongside traditional flux tower data from all 47 NEON terrestrial field sites, in standardized format via the NEON Data Portal and Application Programming Interface (API; Metzger et al. 2019a). A particular science focus of the NCAR-NEON CLM5 implementation is error characterization, including model structure, parameters, initial conditions, meteorological forcing, and observational error.

Large-scale observations of the atmospheric composition and its variation across time and space provide a first principal constraint on the benchmarking framework (e.g., Tans et al. 1990; Gurney et al. 2003; Battle et al. 2000; Pacala et al. 2001). The strength of this top-down ODP constraint is that it provides a direct measure of atmospheric stocks, though attribution to surface processes remains challenging (e.g., Houweling et al. 2017). These top-down constraints are available from tall towers (e.g., Miles et al. 2012; Andrews et al. 2014), airborne (e.g., Sweeney et al. 2015; Miller et al. 2016; Barkley et al. 2019) and spaceborne observations (e.g., Chen et al. 2020). One example is NASA's Atmospheric Carbon and Transport (ACT) - America campaign, which measured atmospheric carbon concentrations, trace gases and meteorological conditions via aircraft in five campaigns spanning all four seasons from 2016–2019 (Davis et al. 2019). ACT-America's airborne measurements are temporally sparse, but spatially extensive, covering four seasons and major ecoregions of the central and eastern United States. These flights are designed to provide regional-scale, seasonal constraints on carbon exchange rates by mapping out carbon and related trace gases (Baier et al. 2020) within synoptic weather systems (Pal et al. 2020), complementing the temporally-rich but relatively spatially sparse tower observations and spatially comprehensive column averaged space-borne observations.

Network-scale flux tower observations such as available from FLUXNET (Baldocchi et al. 2001), AmeriFlux (Novick et al. 2018) or NEON (Metzger et al. 2019a) provide the second principal constraint on the benchmarking framework. The strength of this bottom-up constraint is that SAE observations provide a direct and independent benchmark for assessing the process-scale performance of ESMs, though scale mismatch and surface energy imbalance remain challenging. Here, we seek to improve model benchmarking with flux tower data through two synergistic bottom-up approaches, an "extensive" and an "intensive" approach. The extensive bottom-up approach annotates AmeriFlux data with spatial attributes (e.g., land cover, vegetation indices, etc.; Chu et al. 2020). Thanks to comparatively weak data requirements this approach is readily applied to 200+ AmeriFlux sites. Site spatial representativeness can now be assessed by comparing spatial attributes in the flux surface source area vs. the target domain, such as a model grid cell. This approach facilitates shortlisting spatially representative sites (e.g., sites with similar plant functional type and vegetation characteristics between the flux source area and target domain) for initial model benchmarking, and improved model representation of compound ecosystems. The extensive approach also serves as a prior to identify and prioritize the sites where the intensive approach is deemed necessary, which we explore in more detail in the following section.

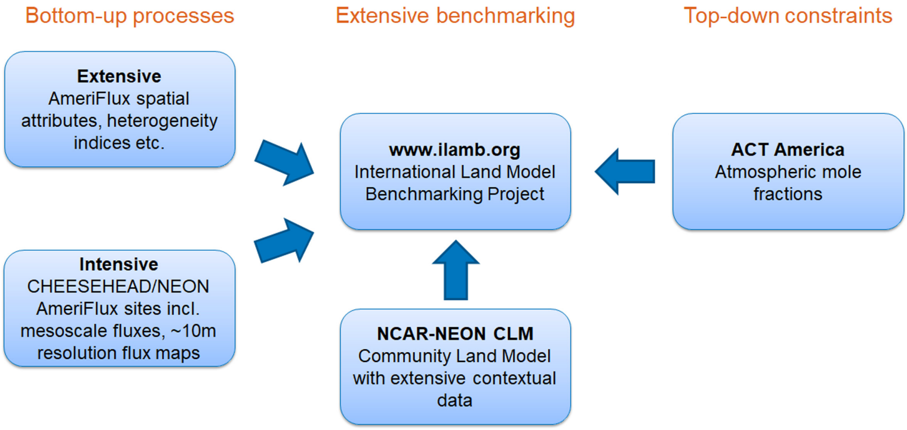

Fig. 1. Scale-aware benchmarking framework that complements bottom-up process information with top-down constraints across DOC, DOE, NASA and NSF projects. Presented during the AGU 2020 Fall Meeting NCAR-NEON Town Hall (Metzger et al. 2019b)

Scale-Equivalent Observational Benchmarks. In contrast to the shortlisting employed in the extensive bottom-up approach, the intensive bottom-up approach aims to fully utilize the variability inherent to changing flux tower sample characteristics. The aim here is to develop scale-aware ODPs from point and line observations for improved model benchmarking at equivalent space and time resolutions. This is achieved by fully incorporating the source area dynamics in source area-to-target-area upscaling (Fu et al. 2014; Metzger et al. 2013a; Ran et al. 2016; Xu et al. 2017). These approaches show great merits in providing space-time explicit flux ODPs that model predictions could be readily benchmarked against at designated grid cells. Furthermore, the Environmental Response Function (ERF) Virtual Control Volume (VCV) spatio-temporal data assimilation system shows promise to also close the surface energy imbalance frequently observed at flux towers (Metzger 2018; Xu et al. 2020), which to date hamstrings data synthesis and model-data fusion with a pervasive bias (e.g., Cui and Chui 2019; Mauder et al. 2020; Stoy et al. 2013).

While ERF promises complete data utilization it has comparatively strong data requirements. This includes EC high-frequency data, which are currently limited to AmeriFlux Core ($N = 14$) and NEON ($N = 47$) sites, and AmeriFlux Tech Team site visits ($N = 40$–50). Specifically, surface and meteorological controls on the fluxes change at minute timescales through transience of source areas, the passing of clouds, etc. Thus, performing ERF analyses at minute- and decameter-resolution allows separating meteorological and surface controls on the fluxes in unprecedented clarity: spectral averaging and source attribution of high-frequency data combined with machine learning connect fluxes to meteorological and surface properties, and ultimately transfer the joint information to the model grid scale. The utilization of high-frequency wavelet flux calculations produces response variable observation with large sample sizes and high signal-to-noise ratio. Thus, providing ample data for the boosted regression trees technique to extract the key driver-response relationships (Metzger et al. 2013a). Results include half-hourly flux maps and propagated uncertainties, alongside estimates of the spatial mean and land-cover specific fluxes and their variation across space (Fig. 2). Figure 2 illustrates the mapped projection of turbulent sensible heat flux, the transfer of heat inducing a change in temperature, throughout the day across a $30\ \text{km} \times 30\ \text{km}$ grid centered on the AmeriFlux Park Falls tall tower site. The derived spatially attributed fluxes from ERF are observed to transition from negative to positive as the surface warms during the day, with clear hot- and cold-spots observable due to the landscapes heterogenuous ecosystem. By including mesoscale motions in a continuous, fixed-frame representation of all hot- and cold-spots within a model grid cell ERF-VCV reduces advective errors by at least one order of magnitude, which effectively closes the surface energy balance (Xu et al. 2020). Where ESMs do not explicitly represent site heterogeneity, we integrate flux maps to probability density functions and from there to statistical measures of location and dispersion (Metzger 2018). We will add these to the ILAMB database of regional simulations to design new, probability-based model benchmarking metrics/scores, and inform the weighting of observations in the data assimilation, uncertainty quantification, and site-level validation processes.

The flux maps are accompanied by a set of non-linear response functions, jointly extracted from ground, airborne, and spaceborne data (Fig. 3). These will serve as benchmarks for diagnosing calibrated models and attributing remote sensing data to surface processes. Ultimately, they allow designing new benchmarking metrics/scores based on ERF-observed vs. ESM-modeled driver-response relationships/surfaces (e.g., Koven et al. 2017).

The promise of scale-aware model benchmarking is that we can better ascribe differences between models and observations to process, parameter, driver, and random error (Dietze 2017), to which we might otherwise falsely attribute scale-related differences. In short: to what extent can we better evaluate or benchmark models with flux data when we consider a flux product that fully matches the scale of the model output and considers the mixing of spatial and temporal variability that occurs at many flux tower sites? The approach outlined here

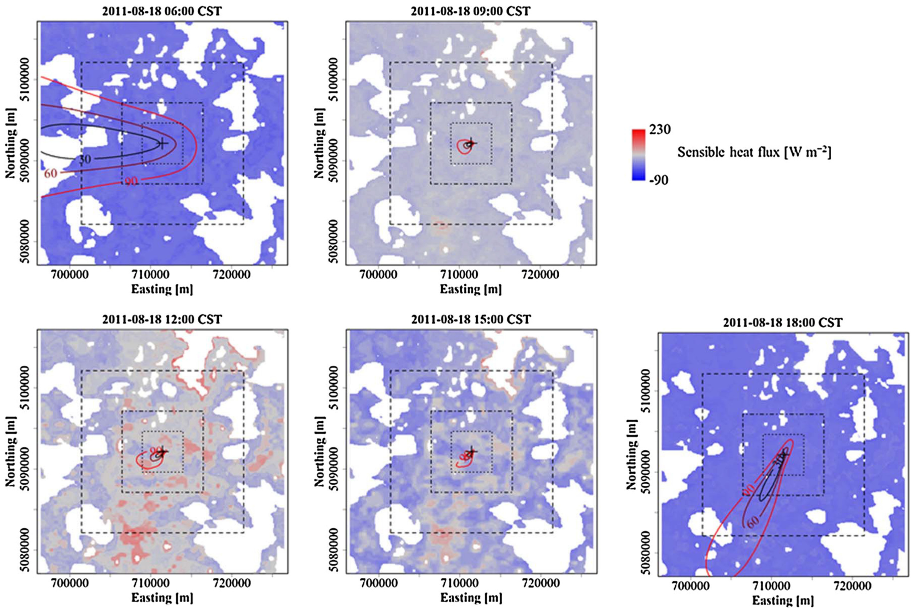

Fig. 2. Flux source area variations over time at the AmeriFlux Park Falls tall tower 122 m measurement height, modified after (Metzger et al. 2013b). The transient source areas are superimposed over the fixed-frame ERF-derived grids of turbulent sensible heat flux. Reprinted from *Agricultural and Forest Meteorology*, Volume 255, Stefan Metzger, Surface-atmosphere exchange in a box: Making the control volume a suitable representation for in-situ observations, Pages 68–80, Copyright (2018), with permission from Elsevier.

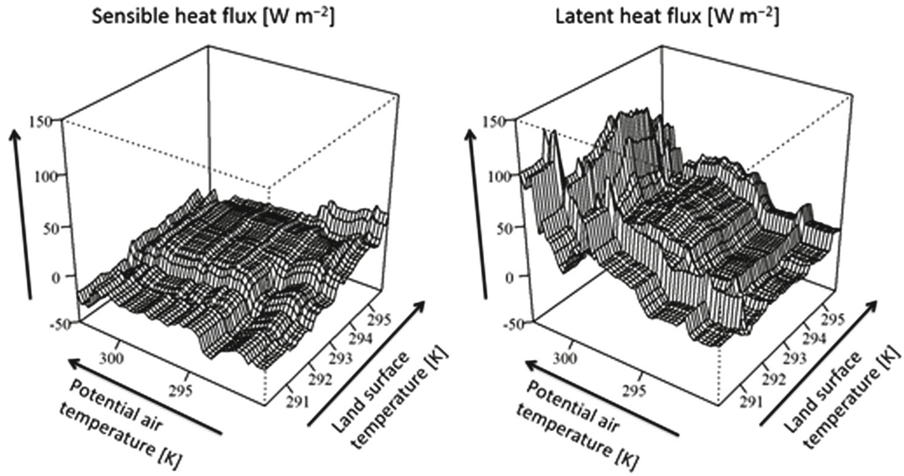

Fig. 3. Multi-dimensional flux response functions at the AmeriFlux Park Falls tall tower 122 m measurement height, modified after (Metzger et al. 2013b).

provides a framework to partition observational uncertainty into scale-related and instrument-related components. Benchmarking or data assimilation is not possible without proper characterization of uncertainty in both observation and model. A systematic approach is essential to make forward progress. A systematic application of a scale-aware benchmark also allows for identification of "ideal" sites or a complementary suite of measurements necessary for an observational site to be considered a high-quality benchmark.

To this last point, recent field experiments have exploited the "super-site" concept to better evaluate the mix of measurement types, extent, and frequency to develop a robust scale-aware benchmark. For example, the Chequamegon Heterogeneous Ecosystem Energy-balance Study Enabled by a High-density Extensive Array of Detectors 2019 (CHEESEHEAD19) field project deployed a quasi-random extensive set of EC flux towers within a "model grid", coupled with a range of airborne and ground based sampling of surface and atmospheric properties and expansive collection of satellite remote sensing imagery (Butterworth et al. 2020). Campaigns like this or the proposed NCAR-NEON super-site project provide a window into the capability of scale-aware benchmarks. They provide a framework for future experimental design of long-term super-sites or identification of core observables necessary to develop scale-aware benchmarks at other sites.

Similarly, nesting sub-grid models within global gridded ESMs provides another opportunity to incorporate scale dependencies within the model. The NOAA Climate Process Team (CPT) Coupling of Land and Atmospheric Subgrid Parameterizations (CLASP) is evaluating how large eddy simulations (LES) and parameterizations can be used to enhance representation of subgrid processes in a model. Such approaches further enhance the value of a scale-aware benchmark.

These experiments and developments thus provide a testbed for evolving the scale-aware benchmark approach. With these, we can start to ask: how much can we relax the high frequency and high resolution data requirements of the ERF approach and still reliably estimate grid-resolved fluxes and uncertainty? How does varying combinations of EC, concentration gradient, tower-mounted imaging, and new sensing techniques expand the reach of the methods into different trace gas fluxes or with higher accuracy? Can ERF also be used to map and predict state variables like biomass, leaf area, canopy chemistry, near-surface temperatures, and other sources of subgrid variability that facilitate space-time consistent ESM inputs and outputs? What are new ways to benchmark models once a space and time resolved benchmark or subgrid model is available? Is the information value of the benchmark limited to the single "grid-cell" of the land-surface model or is the spatial/temporal correlation structure useful for propagating the benchmark to other locations? A number of open research questions and exciting directions are currently foreseen, such as space/time gap-filling and partitioning to resolve issues inherent to current approaches, including confounding space/time transience with biophysical processes.

To summarize, ERF-derived ODPs fully match the scale of ESM inputs and outputs, and comply with previously unmet observational assumptions. The results are half-hourly flux maps of a model subgrid domain that facilitate consistent integration among multi-scale observations and models at flux tower sites. Individual flux pixels even provide a direct link to plot-scale surface observations, such as soil plots and biometric observations. Furthermore, in-situ response function benchmarks improve model diagnosis and remote sensing data interpretation. These scale-aware properties promise unequalled realism for integrating observations and models through overcoming long-standing differences in perception across disciplines.

2.2 Near-Real Time Data Accessibility for ESM and Benchmarking

SAE ODPs for evaluating ESM are currently either available from individual sites in near-real-time, or from many networked sites with latencies on the order of 6 months to 1 year. Due to its central collection and processing structure NEON has the opportunity to push the boundaries of near-real-time data availability to facilitate ecological forecasting, data assimilation into ESMs, and ESM benchmarking. Currently, the vast majority of NEON's 53 terrestrial instrumented systems (TIS) data products are available with a 1-month latency via the NEON data portal (https://data.neonscience.org/) and API (https://data.neonscience.org/data-api/) due to a monthly publication cycle. However, NEON SAE processing pipeline improvements are in development to reduce data latency to 1–5 days. To our knowledge, this would be the largest EC tower data set provided in near-real-time globally.

A pilot project envisioned from the aforementioned NCAR-NEON workshop developed a workflow to grab NEON data from the API, perform some quality assurance and quality control, gap-fill data, partition fluxes, and package data in a netCDF data format that is ingestible by CLM5, ILAMB, and PEcAn. The workflow is being hosted on Github (https://github.com/NEONScience/NCAR-NEON), has been containerized (https://quay.io/repository/ddurden/ncar-neon-ddurden), and is deployable via command line for integration with job schedulers or workflow managers. The NEON data pipeline is transitioning to a microservices-based Pachyderm architecture (https://www.pachyderm.com/), a version control system for data that preserves data provenance. In the Pachyderm pipeline, any new commit to data, metadata, or processing code triggers the reprocessing of downstream derived products. Integration of ODP generation for model-data fusion into this architecture promises near-real time data access with full provenance. Work with the scientific community still remains to address where community modeling and benchmarking data sets should be hosted and determine the essential ODPs to be provided both for driving models and evaluating/benchmarking.

To support rapid and scalable assessment and benchmarking of LSM results, a Land Model Testbed (LMT) system is being developed through a pilot project at ORNL (Fig. 4). Aimed at delivering a workflow for very large ensemble simulations, the LMT provides software infrastructure for running multiple

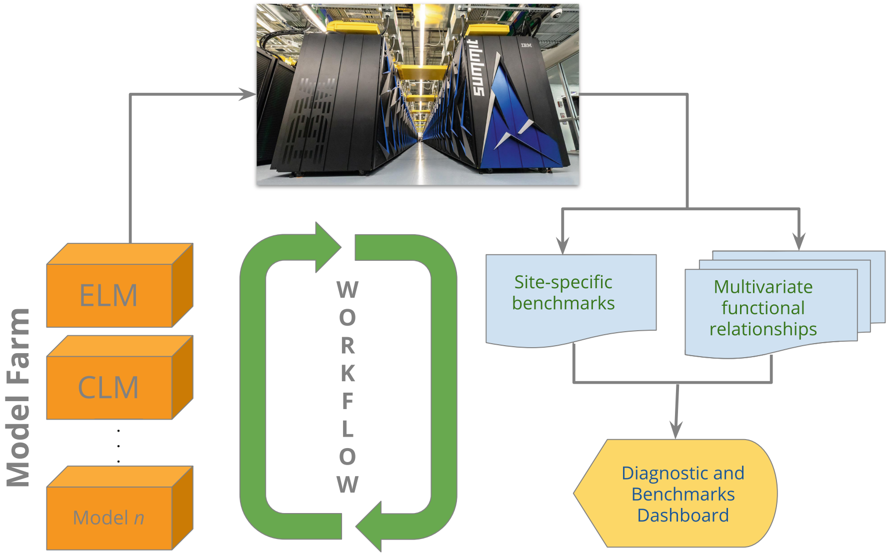

Fig. 4. A Land Model Testbed (LMT) workflow for running and evaluating large numbers of ensemble simulations for multiple LSMs on the Summit supercomputer system and dynamically provisioned cloud resources is being developed at ORNL. Site-specific benchmarks for EC super-sites and new functional relationship metrics are being incorporated into ILAMB, and a dynamic user interface is being developed to give users better control over how model-data comparison results are displayed through an interactive dashboard.

models on the Summit supercomputer system and dynamically provisioned cloud computing resources. New site-specific benchmarks for EC super-sites and new functional relationship metrics are being incorporated into ILAMB to support assessment of large ensembles and PPE simulations. An interactive dashboard is being designed to give users control over how benchmarking results and graphical diagnostics are displayed. Interfaces are also being developed around ILAMB for activation (executing an analysis) and linking to diagnostic results following the evolving Coordinated Model Evaluation Capabilities (CMEC) standards. CMEC interfaces will further enable connections to NOAA's Model Diagnostics Task Force that promotes development of process-oriented diagnostics for climate and weather forecasting models (Maloney et al. 2019). These improvements are key to informing parameterization improvements to address long-standing model biases and to delivering credible projection results for assessing climate change impacts and vulnerabilities for stakeholders and policy-makers (Eyring et al. 2019).

The LMT, combined with NEON's near-real time SAE ODPs, offers a truly scalable approach for rapidly conducting ecological forecasts on HPC systems and evaluating model performance as new measurements are made. We envision integrating the multi-scale observations from NEON's distributed edge computing systems with multiple LSMs running in the LMT framework on centralized HPC systems and distributed cloud computing resources. This data-model integration approach will advance ecological research and improve mechanistic understanding of Earth system processes important for environmental sustainability.

3 Roadmap to Scientific Understanding

The roadmap to extracting scientific understanding through data-model integration is contingent on multiple working groups working toward common underlying goals of maximizing our predictive capabilities, minimizing uncertainty associated with our predictions, and advancing our forecast accuracy with near-real-time data. Near-real-time data cyber-infrastructure is on the verge of being realized for multiple flux tower networks, and is opening new pathways to near-term ESM benchmarking, parameter optimization, and data-fusion techniques.

The 2019 RUBISCO-AmeriFlux Workshop (Hawkins et al. 2020) planned roadmap lays the foundation for the bottom-up scaling approaches to produce scale-aware ODPs and ingest them into the ILAMB benchmarking framework (Fig. 5). For the extensive bottom-up approach initial data processing is complete, and the manuscript by Chu et al. (2020) introduces the results and newly available spatial attributes to the community at large. Our planned goal for 2020 is to produce a shortlist of homogeneous sites for initial model benchmarking, with additional milestones through 2021 (Fig. 5). For the intensive bottom-up approach, the group is working on integrating the ERF-VCV data sets into ILAMB. At this time, the group has successfully ingested the NEON NetCDF file format into ILAMB, and is compiling the Metzger et al. (2019a) 30 min flux grids into these files. Planned goals for 2020 include regional ILAMB evaluations and site-level validations to design performance scores, with additional milestones through 2023 (Fig. 5). We further envision a hybrid "simplified high-res mapping" bottom-up approach to reduce ERF-VCV data requirements for use at all AmeriFlux sites, which is currently ahead of schedule.

The bottom-up approaches are complemented by the top-down syntheses of aircraft campaign data from ACT-America, an array of terrestrial ecosystem models, posterior flux estimates from atmospheric inverse flux estimates and AmeriFlux observations. The expected outcome is spatially and temporally comprehensive evaluation of the performance of these ecosystem models and inversion posteriors. This evaluation will provide insight into the process limitations of these models and the existing seasonal, regional biases in the inversion systems. The improved understanding will be used to improve the prior flux estimates used in atmospheric inversions, and to improve the process representation in regional to continental scale simulations of terrestrial carbon fluxes.

Through the convergence of high throughput computational frameworks processing EC data and applying machine learning algorithms to develop scale-aware ODPs with multiple instances of ESMs running on HPC, we can make substantial strides to our understanding of Earth systems processes across spatiotemporal scales that have previously restricted such studies. The advancement

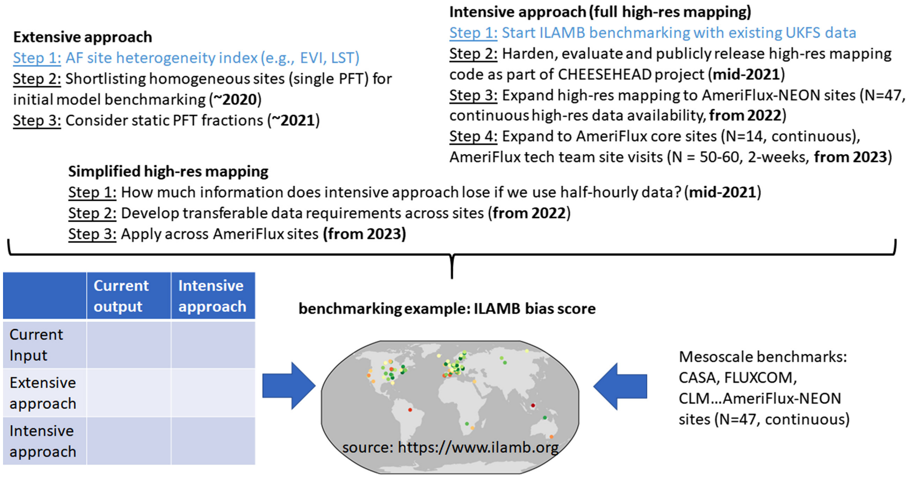

Fig. 5. Status and roadmap of the bottom-up scaling approaches. Blue font indicates areas of currently active work. (Color figure online)

of ecosystem understanding is not confined to the described work though. The development of the Waggle, an open sensor platform for edge computing, by the Array of Things (AoT) opens the door to enhanced distributed data collection, advanced reactive measurements, and manipulative studies (Beckman et al. 2016). NEON has the observational infrastructure, such as sufficient power and network connectivity at tower sites and advanced command and control capabilities, to utilize such compute infrastructure in the future.

Acknowledgements. The National Ecological Observatory Network is a project sponsored by the National Science Foundation and managed under cooperative agreement by Battelle. Some of this material is based upon work supported by the National Science Foundation (Grant DBI-0752017). We acknowledge contributions for some of this material from participants of NSF-sponsored joint NCAR and NEON workshop, Predicting life in the Earth system – linking the geosciences and ecology, and the continued efforts in post-workshop working groups. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Portions of this research were sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory and used resources of the Oak Ridge Leadership Computing Facility (OLCF) at Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract DE–AC05–00OR22725. Additional support was provided by the Reducing Uncertainties in Biogeochemical Interactions through Synthesis and Computation Science Focus Area (RUBISCO SFA), which is sponsored by the Regional and Global Model Analysis (RGMA) activity of the Earth & Environmental System Modeling (EESM) Program in the Earth and Environmental Systems Sciences Division (EESSD) of the Office of Biological and Environmental Research (BER) in the U.S. Department of Energy Office of Science. Some contributions to this research were supported by NASA's Earth

Science Division (Grant NNX15AG76G). We acknowledge contributions for some of this material from all participants in the RUBISCO-AmeriFlux Working Group, which is supported by the RUBISCO SFA.

References

Abramowitz, G.: Towards a benchmark for land surface models. Geophys. Res. Lett. **32**(22), L22702 (2005). https://doi.org/10.1029/2005GL024419

Abramowitz, G.: Towards a public, standardized, diagnostic benchmarking system for land surface models. Geosci. Model Dev. **5**(3), 819–827 (2012). https://doi.org/10.5194/gmd-5-819-2012

Anav, A., et al.: Evaluating the land and ocean components of the global carbon cycle in the CMIP5 earth system models. J. Clim. **26**(18), 6801–6843 (2013). https://doi.org/10.1175/JCLI-D-12-00417.1

Andrews, A., et al.: CO_2, CO, and CH_4 measurements from tall towers in the NOAA Earth System Research Laboratory's Global Greenhouse gas reference network: instrumentation, uncertainty analysis, and recommendations for future high-accuracy greenhouse gas monitoring efforts. Atmos Meas Tech **7**(2), 647 (2014). https://doi.org/10.5194/amt-7-647-2014

Arora, V.K., et al.: Carbon-concentration and carbon-climate feedbacks in CMIP5 earth system models. J. Clim. **26**(15), 5289–5314 (2013). https://doi.org/10.1175/JCLI-D-12-00494.1

Aubinet, M., Vesala, T., Papale, D.: Eddy Covariance: A Practical Guide to Measurement and Data Analysis. Springer, Dordrecht (2012). https://doi.org/10.1007/978-94-007-2351-1

Baier, B.C., et al.: Multispecies assessment of factors influencing regional and enhancements during the Winter 2017 ACT-America Campaign. J. Geophys. Res. Atmos. **125**(2), e2019JD031339 (2020). https://doi.org/10.1029/2019JD031339

Baldocchi, D., et al.: FLUXNET: a newtool to study the temporal and spatial variability of ecosystem-scale carbondioxide, water vapor, and energy flux densities. Bull. Am. Meteorol. Soc. **82**(11), 2415–2434 (2001). https://doi.org/10.1175/1520-0477(2001)082%3C2415:FANTTS%3E2.3.CO;2

Barkley, Z.R., et al.: Forward modeling and optimization of methane emissions in the south central United States using aircraft transects across frontal boundaries. Geophys. Res. Lett. **46**(22), 13564–13573 (2019). https://doi.org/10.1029/2019GL084495

Bastrikov, V., MacBean, N., Bacour, C., Santaren, D., Kuppel, S., Peylin, P.: Land surface model parameter optimisation using in situ flux data: comparison of gradient-based versus random search algorithms (a case study using ORCHIDEE v1.9.5.2). Geosci. Model. Dev. **11**(12), 4739–4754 (2018). https://doi.org/10.5194/gmd-11-4739-2018

Battle, M., et al.: Global carbon sinks and their variability inferred from atmospheric O_2 and δ13C. Science **287**(5462), 2467–2470 (2000). https://doi.org/10.1126/science.287.5462.2467

Beckman, P., Sankaran, R., Catlett, C., Ferrier, N., Jacob, R., Papka, M.: Waggle: an open sensor platform for edge computing. In: 2016 IEEE SENSORS, pp. 1–3. IEEE (2016). https://doi.org/10.1109/ICSENS.2016.7808975

Best, M.J., et al.: The plumbing of land surface models: benchmarking model performance. J. Hydrometeor. **16**(3), 1425–1442 (2015). https://doi.org/10.1175/JHM-D-14-0158.1

Blyth, E., et al.: A comprehensive set of benchmark tests for a land surface model of simultaneous fluxes of water and carbon at both the global and seasonal scale. Geosci. Model. Dev. **4**(2), 255–269 (2011). https://doi.org/10.5194/gmd-4-255-2011

Bonan, G.B.: Ecological Climatology: Concepts and Applications, 3rd edn. Cambridge University Press, New York (2016). https://doi.org/10.1017/CBO9781107339200

Bonan, G.B.: Climate Change and Terrestrial Ecosystem Modeling. Cambridge University Press, New York (2019). https://doi.org/10.1017/9781107339217

Bonan, G.B., Doney, S.C.: Climate, ecosystems, and planetary futures: The challenge to predict life in Earth system models. Science **359**(6375), eaam8328 (2018). https://doi.org/10.1126/science.aam8328

Butterworth, B.J., et al.: Connecting land-atmosphere interaction to surface heteorogeniety in CHEESEHEAD 2019 (2020, in preparation)

Cadule, P., et al.: Benchmarking coupled climate-carbon models against long-term atmospheric CO_2 measurements. Glob. Biogeochem. Cycles **24**(2), GB2016 (2010). https://doi.org/10.1029/2009GB003556

Chen, B., et al.: Assessing eddy-covariance flux tower location bias across the Fluxnet-Canada research network based on remote sensing and footprint modelling. Agric. Forest Meteorol. **151**(1), 87–100 (2011). https://doi.org/10.1016/j.agrformet.2010.09.005

Chen, Z., Liu, J., Henze, D.K., Huntzinger, D.N., Wells, K.C., Miller, S.M.: Linking global terrestrial CO_2 fluxes and environmental drivers using OCO-2 and a geostatistical inverse model. Atmos. Chem. Phys. Discuss **2020**, 1–24 (2020). https://doi.org/10.5194/acp-2020-285

Chu, H., et al.: Footprint representativeness of eddy-covariance flux measurements across AmeriFlux sites (2020, in preparation)

Clark, J.S., et al.: Ecological forecasts: an emerging imperative. Science **293**(5530), 657–660 (2001). https://doi.org/10.1126/science.293.5530.657

Collier, N., et al.: The international land model benchmarking (ILAMB) system: design, theory, and implementation. J. Adv. Model. Earth Sy. **10**(11), 2731–2754 (2018). https://doi.org/10.1029/2018MS001354

Cui, W., Chui, T.F.M.: Temporal and spatial variations of energy balance closure across FLUXNET research sites. Agric. Forest Meteorol. **271**, 12–21 (2019). https://doi.org/10.1016/j.agrformet.2019.02.026

Davis, K.J., et al.: ACT-America: L3 merged in situ atmospheric trace gases and flask data. Eastern USA (2019). https://doi.org/10.3334/ORNLDAAC/1593

Dietze, M.C., LeBauer, D.S., Kooper, R.: On improving the communication between models and data. Plant Cell Environ. **36**(9), 1575–1585 (2013). https://doi.org/10.1111/pce.12043

Dietze, M.C., et al.: A quantitative assessment of a terrestrial biosphere model's data needs across North American biomes. J. Geophys. Res. Biogeosci. **119**(3), 286–300 (2014). https://doi.org/10.1002/2013JG002392

Dietze, M.C., et al.: Iterative near-term ecological forecasting: needs, opportunities, and challenges. Proc. Natl. Acad. Sci. **115**(7), 1424–1432 (2018). https://doi.org/10.1073/pnas.1710231115

Eyring, V., et al.: Taking climate model evaluation to the next level. Nat. Clim. Change **9**(2), 102–110 (2019). https://doi.org/10.1038/s41558-018-0355-y

Fer, I., Kelly, R., Moorcroft, P.R., Richardson, A.D., Cowdery, E.M., Dietze, M.C.: Linking big models to big data: efficient ecosystem model calibration through Bayesian model emulation. Biogeoscience **15**(19), 5801–5830 (2018). https://doi.org/10.5194/bg-15-5801-2018

Fischer, E.M., Lawrence, D.M., Sanderson, B.M.: Quantifying uncertainties in projections of extremes-a perturbed land surface parameter experiment. Clim. Dyn. **37**(7), 1381–1398 (2011). https://doi.org/10.1007/s00382-010-0915-y

Flato, G.M.: Earth system models: an overview. WIREs Clim. Change **2**(6), 783–800 (2011). https://doi.org/10.1002/wcc.148

Foley, J.A., et al.: Global consequences of land use. Science **309**(5734), 570–574 (2005). https://doi.org/10.1126/science.1111772

Fox, A., et al.: The REFLEX project: comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data. Agric. Forest Meteorol. **149**(10), 1597–1615 (2009). https://doi.org/10.1016/j.agrformet.2009.05.002

Fox, A.M., et al.: Evaluation of a data assimilation system for land surface models using CLM4.5. J. Adv. Model. Earth Syst. **10**(10), 2471–2494 (2018). https://doi.org/10.1029/2018MS001362

Friedlingstein, P., et al.: Positive feedback between future climate change and the carbon cycle. Geophys. Res. Lett. **28**(8), 1543–1546 (2001). https://doi.org/10.1029/2000GL012015

Friedlingstein, P., et al.: Climate-carbon cycle feedback analysis: Results from the C^4MIP model intercomparison. J. Clim. **19**(14), 3373–3383 (2006). https://doi.org/10.1175/JCLI3800.1

Fu, D., et al.: Estimating landscape net ecosystem exchange at high spatial-temporal resolution based on landsat data, an improved upscaling model framework, and eddy covariance flux measurements. Remote Sens. Environ. **141**, 90–104 (2014). https://doi.org/10.1016/j.rse.2013.10.029

Griebel, A., Metzen, D., Pendall, E., Burba, G., Metzger, S.: Generating spatially robust carbon budgets from flux tower observations. Geophys. Res. Lett. **47**(3), e2019GL085942 (2020). https://doi.org/10.1029/2019GL085942

Gurney, K.R., et al.: TransCom 3 CO_2 inversion intercomparison: 1. Annual mean control results and sensitivity to transport and prior flux information. Tellus B **55**(2), 555–579 (2003). https://doi.org/10.3402/tellusb.v55i2.16728

Hawkins, L.R., Kumar, J., Luo, X., Sihi, D., Zhou, S.: Measuring,monitoring, andmodeling ecosystem cycling. EOS Trans. AGU **101** (2020). https://doi.org/10.1029/2020EO147717

Hoffman, F.M., et al.: International land model benchmarking (ILAMB) 2016 workshop report. Technical report DOE/SC-0186, U.S. Department of Energy, Office of Science, Germantown, Maryland, USA (2017). https://doi.org/10.2172/1330803

Houweling, S., et al.: Global inverse modeling of CH_4 sources and sinks: an overview of methods. Atmos. Chem. Phys. **17**(1), 235–256 (2017). https://doi.org/10.5194/acp-17-235-2017

Kattge, J., et al.: TRY - a global database of plant traits. Glob. Change Biol. **17**(9), 2905–2935 (2011). https://doi.org/10.1111/j.1365-2486.2011.02451.x

Koven, C.D., Hugelius, G., Lawrence, D.M., Wieder, W.R.: Higher climatological temperature sensitivity of soil carbon in cold than warm climates. Nat. Clim. Change **7**(11), 817–822 (2017). https://doi.org/10.1038/nclimate3421

Launiainen, S., et al.: Do the energy fluxes and surface conductance of boreal coniferous forests in Europe scale with leaf area? Glob. Change Biol. **22**(12), 4096–4113 (2016). https://doi.org/10.1111/gcb.13497

LeBauer, D.S., Wang, D., Richter, K.T., Davidson, C.C., Dietze, M.C.: Facilitating feedbacks between field measurements and ecosystem models. Ecol. Monogr. **83**(2), 133–154 (2013). https://doi.org/10.1890/12-0137.1

Li, J., Duan, Q., Wang, Y.P., Gong, W., Gan, Y., Wang, C.: Parameter optimization for carbon and water fluxes in two global land surface models based on surrogate modelling. Int. J. Climatol. **38**(S1), e1016–e1031 (2018). https://doi.org/10.1002/joc.5428

Liu, S., et al.: Upscaling evapotranspiration measurements from multi-site to the satellite pixel scale over heterogeneous land surfaces. Agric. Forest Meteorol. **230**, 97–113 (2016). https://doi.org/10.1016/j.agrformet.2016.04.008

Lu, D., Ricciuto, D., Stoyanov, M., Gu, L.: Calibration of the E3SM land model using surrogate-based global optimization. J. Adv. Model. Earth Syst. **10**(6), 1337–1356 (2018). https://doi.org/10.1002/2017MS001134

Luo, Y., et al.: Ecological forecasting and data assimilation in a data-rich era. Ecol. Appl. **21**(5), 1429–1442 (2011). https://doi.org/10.1890/09-1275.1

Maloney, E.D., et al.: Process-oriented evaluation of climate and weather forecasting models. Bull. Am. Meteorol. Soc. **100**(9), 1665–1686 (2019). https://doi.org/10.1175/BAMS-D-18-0042.1

Mauder, M., Foken, T., Cuxart, J.: Surface-energy-balance closure over land: a review. Boundary-Layer Meteorol. (2020). https://doi.org/10.1007/s10546-020-00529-6

Metzger, S.: Surface-atmosphere exchange in a box: making the control volume a suitable representation for in-situ observations. Agric. Forest Meteorol. **255**, 68–80 (2018). https://doi.org/10.1016/j.agrformet.2017.08.037

Metzger, S., et al.: Spatially explicit regionalization of airborne flux measurements using environmental response functions. Biogeoscience **10**(4), 2193–2217 (2013a). https://doi.org/10.5194/bg-10-2193-2013

Metzger, S., et al.: From NEON field sites to data portal: a community resource for surface-atmosphere research comes online. Bull. Am. Meteorol. Soc. **100**(11), 2305–2325 (2019a). https://doi.org/10.1175/BAMS-D-17-0307.1

Metzger, S., et al.: Synthesized observations and processes for plot- to landscape-scale research. In: NCAR and NEON Town Hall TH13M, 2019 American Geophysical Union (AGU) Annual Fall Meeting, CA, USA, San Francisco (2019b)

Metzger, S.: Spatio-temporal rectification of tower-based eddy-covariance flux measurements for consistently informing process-based models. In: 2013 American Geophysical Union (AGU) Annual Fall Meeting, CA, USA, San Francisco (2013b)

Miles, N.L., et al.: Large amplitude spatial and temporal gradients in atmospheric boundary layer co2mole fractions detected with a tower-based network in the U.S. Upper Midwest. J. Geophys. Res. Biogeosci. **117**(G1) (2012). https://doi.org/10.1029/2011JG001781. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011JG001781. https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2011JG001781

Miller, S.M., et al.: A multiyear estimate of methane fluxes in Alaska from CARVE atmospheric observations. Glob. Biogeochem. Cycles **30**(10), 1441–1453 (2016). https://doi.org/10.1002/2016GB005419

Moore, D.J.P., Hu, J., Sacks, W.J., Schimel, D.S., Monson, R.K.: Estimating transpiration and the sensitivity of carbon uptake to water availability in a subalpine forest using a simple ecosystem process model informed by measured net CO_2 and H_2O fluxes. Agric. Forest Meteorol. **148**(10), 1467–1477 (2008). https://doi.org/10.1016/j.agrformet.2008.04.013

Novick, K.A., et al.: The AmeriFlux network: a coalition of the willing. Agric. Forest Meteorol. **249**, 444–456 (2018). https://doi.org/10.1016/j.agrformet.2017.10.009

Pacala, S.W., et al.: Consistent land-and atmosphere-based US carbon sink estimates. Science **292**(5525), 2316–2320 (2001). https://doi.org/10.1126/science.1057320

Pal, S., et al.: Observations of greenhouse gas changes across summer frontal boundaries in the Eastern United States. J. Geophys. Res. Atmos. **125**(5), e2019JD030526 (2020). https://doi.org/10.1029/2019JD030526

Piao, S., et al.: Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO_2 trends. Glob. Change Biol. **19**(7), 2117–2132 (2013). https://doi.org/10.1111/gcb.12187

Prihodko, L., Denning, A.S., Hanan, N.P., Baker, I., Davis, K.: Sensitivity, uncertainty and time dependence of parameters in a complex land surface model. Agric. Forest Meteorol. **148**(2), 268–287 (2008). https://doi.org/10.1016/j.agrformet.2007.08.006

Quaife, T., et al.: Assimilating canopy reflectance data into an ecosystem model with an Ensemble Kalman Filter. Remote Sens. Environ. **112**(4), 1347–1364 (2008). https://doi.org/10.1016/j.rse.2007.05.020

Ran, Y., et al.: Spatial representativeness and uncertainty of eddy covariance carbon flux measurements for upscaling net ecosystem productivity to the grid scale. Agric. Forest Meteorol. **230**, 114–127 (2016). https://doi.org/10.1016/j.agrformet.2016.05.008

Randall, D.A., et al.: 100 years of Earth system model development. Meteor. Monogr. **59**, 12.1–12.66 (2018). https://doi.org/10.1175/AMSMONOGRAPHS-D-18-0018.1

Randerson, J.T., et al.: Systematic assessment of terrestrial biogeochemistry in coupled climate-carbon models. Glob. Change Biol. **15**(9), 2462–2484 (2009). https://doi.org/10.1111/j.1365-2486.2009.01912.x

Raupach, M.R., et al.: Model-data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications. Glob. Change Biol. **11**(3), 378–397 (2005). https://doi.org/10.1111/j.1365-2486.2005.00917.x

Ricciuto, D., Sargsyan, K., Thornton, P.: The impact of parametric uncertainties on biogeochemistry in the E3SM land model. J. Adv. Model. Earth Syst. **10**(2), 297–319 (2018). https://doi.org/10.1002/2017MS000962. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017MS000962. https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2017MS000962

Ricciuto, D.M., Davis, K.J., Keller, K.: A Bayesian calibration of a simple carbon cycle model: the role of observations in estimating and reducing uncertainty. Glob. Biogeochem. Cycles **22**(2) (2008). https://doi.org/10.1029/2006GB002908

Ricciuto, D.M., King, A.W., Dragoni, D., Post, W.M.: Parameter and prediction uncertainty in an optimized terrestrial carbon cycle model: effects of constraining variables and data record length. J. Geophys. Res. Biogeosci. **116**(G1) (2011). https://doi.org/10.1029/2010JG001400

Riley, W.J., Shen, C.: Characterizing coarse-resolution watershed soil moisture heterogeneity using fine-scale simulations. Hydrol. Earth Syst. Sci. **18**(7), 2463–2483 (2014). https://doi.org/10.5194/hess-18-2463-2014

Sanderson, B.M., Shell, K.M., Ingram, W.: Climate feedbacks determined using radiative kernels in a multi-thousand member ensemble of AOGCMs. Clim. Dyn. **35**(7), 1219–1236 (2010). https://doi.org/10.1007/s00382-009-0661-1

Schaefer, K., et al.: A model-data comparison of gross primary productivity: results from the North American Carbon Program site synthesis. J. Geophys. Res. Biogeosci. **117**(G3) (2012). https://doi.org/10.1029/2012JG001960

Schimel, D.S., VEMAP Participants, Braswell, B.H.: Continental scale variability in ecosystem processes: models, data, and the role of disturbance. Ecol. Monogr. **67**(2), 251–271 (1997). https://doi.org/10.1890/0012-9615(1997)067[0251:CSVIEP]2.0.CO;2

Schwalm, C.R., et al.: A model-data intercomparison of CO_2 exchange across North America: results from the North American Carbon Program site synthesis. J. Geophys. Res. Biogeosci. **115**(G3) (2010). https://doi.org/10.1029/2009JG001229

Stöckli, R., et al.: Use of FLUXNET in the community land model development. J. Geophys. Res. Biogeosci. **113**(G1) (2008). https://doi.org/10.1029/2007JG000562

Stoy, P.C., et al.: A data-driven analysis of energy balance closure across FLUXNET research sites: the role of landscape scale heterogeneity. Agric. Forest Meteorol. **171**, 137–152 (2013). https://doi.org/10.1016/j.agrformet.2012.11.004

Sweeney, C., et al.: Seasonal climatology of co2 across North America from aircraft measurements in the NOAA/ESRL global greenhouse gas reference network. J. Geophys. Res. Atmos. **120**(10, 5155–5190 (2015). https://doi.org/10.1002/2014JD022591. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014JD022591. https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2014JD022591

Tans, P.P., Fung, I.Y., Takahashi, T.: Observational constraints on the global atmospheric CO_2 budget. Science **247**(4949), 1431–1438 (1990). https://doi.org/10.1126/science.247.4949.1431

Wang, Y.Q., Xiong, Y.J., Qiu, G.Y., Zhang, Q.T.: Is scale really a challenge in evapotranspiration estimation? A multi-scale study in the Heihe oasis using thermal remote sensing and the three-temperature model. Agric. Forest Meteorol. **230**, 128–141 (2016). https://doi.org/10.1016/j.agrformet.2016.03.012

Williams, M., et al.: Improving land surface models with FLUXNET data. Biogeoscience **6**(7), 1341–1359 (2009). https://doi.org/10.5194/bg-6-1341-2009

Xu, K., Metzger, S., Desai, A.R.: Upscaling tower-observed turbulent exchange at fine spatio-temporal resolution using environmental response functions. Agric. Forest Meteorol. **232**, 10–22 (2017). https://doi.org/10.1016/j.agrformet.2016.07.019

Xu, K., Sühring, M., Metzger, S., Durden, D., Desai, A.R.: Can data mining help eddy covariance see the landscape? A large-eddy simulation study. Boundary-Layer Meteorol. **176**(1), 85–103 (2020). https://doi.org/10.1007/s10546-020-00513-0

Zobitz, J.M., Desai, A.R., Moore, D.J.P., Chadwick, M.A.: A primer for data assimilation with ecological models using Markov Chain Monte Carlo (MCMC). Oecologia **167**(3), 599 (2011). https://doi.org/10.1007/s00442-011-2107-9