



Ideas and perspectives: Beyond model evaluation - combining experiments and models to advance terrestrial ecosystem science

Silvia Caldararu^{1,2}, Victor Rolo³, Benjamin D. Stocker^{4,5}, Teresa E. Gimeno⁶, Richard Nair^{1,2}

¹ Discipline of Botany, School of Natural Sciences, Trinity College Dublin, Ireland

5 ² Max Planck Institute for Biogeochemistry, Jena, Germany

³ University of Extremadura, Forest Research Group, INDEHESA, Plasencia, Cáceres, Spain

⁴ Institute of Geography, University of Bern, Bern, Switzerland

⁵ Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland

⁶ CREAM, 08193 Bellaterra (Cerdanyola del Vallès), Catalonia, Spain

10

Correspondence to: Silvia Caldararu (caldaras@tcd.ie) and Victor Rolo (rolo@unex.es)

Abstract. Ecosystem manipulative experiments are one of the most powerful tools to understand terrestrial ecosystem responses to global change because they measure real responses in real ecosystems. However, their scope is limited in space and time due to cost and labour intensity. This makes generalising results from such experiments difficult, which creates a conceptual gap between local scale process understanding and global scale future predictions. Recent efforts have seen results from such experiments used in combination with dynamic global vegetation models, most commonly to evaluate model predictions under global change drivers. However, there is much more potential in combining models and experiments. Here, we discuss the value and potential of a workflow for using ecosystem experiments together with process-based models to enhance the potential of both. We suggest that models can be used prior to the start of an experiment to generate hypotheses, identify data needs and in general guide experimental design. Models, when adequately constrained with observations, can also predict variables which are difficult to measure frequently or at all, and together with the data can provide a more complete picture of ecosystem states. Finally, models can be used to help generalise the experimental results in space and time, by providing a framework in which process understanding derived from site-level experiments can be incorporated. We also discuss the potential for using manipulative experiments together with models in formalised model-data integration frameworks for parameter estimation and model selection, a path made possible by the increasing number of ecosystem experiments and diverse observation streams. The ideas presented here can provide a roadmap to future experiment - model studies.

15
20
25



30 **1 Introduction**

In the face of a changing climate, understanding how ecosystems will respond to conditions unprecedented in the observational record is one of the biggest challenges in earth system science. To meet this challenge, we must employ all tools available to us - observations, experiments and meta-analyses, and process- and data-based models. All these different tools have strengths and weaknesses in terms of spatial and temporal scales, generality and predictive capacity. Therefore, not only using these
35 tools, but using them in combination, is essential for bridging the gap between process understanding and scalability.

Over the last few decades, there have been great advancements in the type and quantity of available ecosystem data, including large databases of field observations (Kattge et al., 2020); continuous ecosystem monitoring (Baldocchi, 2020) remotely sensed Earth observations (Xiao et al., 2019); spatial datasets with global coverage for climate (Fick & Hijmans, 2017; Harris et al.,
40 2020; Karger et al., 2020), soil (Food and Agriculture Organization & Food and Agriculture Organization of the United Nations, 2008; Hengl et al., 2017), and other spatially upscaled global ecosystem data products (Jung et al., 2020; Poggio et al., 2021). In and of themselves, such observational datasets provide a comprehensive picture of ecosystem properties and structure, carbon and water fluxes. However, the mounting data richness in ecosystem science has not led to a reduction in projection uncertainties of global biogeochemical cycles in a future climate (Arora et al., 2013; Friedlingstein et al., 2014).
45 Ecological and physiological process understanding - key for predicting outside the domain of observed conditions - often remains weakly constrained in spite of the large volume of available ecosystem data collected over recent decades. In this context, ecosystem manipulation experiments (EMEs) have a unique and particularly valuable role to play (Stocker et al., 2016; Wieder et al., 2019).

EMEs are outdoor experimental setups in which one or multiple driving factors (e.g., water availability, temperature, nutrient inputs, atmospheric CO₂ concentration, perturbations, species composition), or a combination of factors, are controlled to study their separate or interactive effects on ecosystem processes as well as subunits within ecosystems such as plants or microbes. They are a useful tool because they contain ecosystem process responses defined by treatments and hypotheses, while minimising confounding factors, and scaling biases and experimental artefacts from smaller and more limited lab or
55 mesocosm-based experiments (Dalling et al., 2013). EMEs provide a unique window into the responses of ecosystems to a potential future environment and can reveal complex responses at the level of plants, communities, and ecosystems, shaped by feedbacks between soil and vegetation. While extremely valuable for understanding ecosystem responses to future conditions, due to their realistic scale and multidimensionality of observations, EMEs generate small data - in contrast to, e.g., Earth observation. Therefore, it remains challenging to generalise from a relatively small set of local observations to future global
60 predictions. EMEs are usually operated at spatial scales on the order of 1-100 m, and are limited in time to typically 1-10 years due to high operating costs and laborious measurements. Thus, it can be difficult to scale the conclusions from experiments with confidence, both in time and in space. Here, we propose and discuss approaches for aiding generalisations from EME observations in combination with process-based models as a potential solution to overcome this scalability challenge.



65 Process-based ecosystem models are mathematical representations of how plant traits, and soil characteristics determine water, energy, and biogeochemical fluxes, and the accumulation of organic matter in a cascade of ecosystem pools, given a set of environmental boundary conditions. Depending on the scale and processes represented, such models are termed terrestrial biosphere models, dynamic vegetation models, mechanistic ecosystem models, or land surface models, and are implemented as coupled components in Earth System Models (Prentice et al., 2007). Irrespective of what we call them, all such models face
70 similar challenges, in that they need to balance realism, robustness, and reliability (Prentice et al., 2015) to achieve sound predictions for ecosystem fluxes and pools across the globe and across decades into the future. To achieve this such models need to be confronted with diverse data to leverage complementary constraints. Recent efforts have been made in using large observational datasets for validating and benchmarking these models (Collier et al., 2018; Seiler et al., 2022). However, as outlined above, model projections remain divergent ((Arora et al., 2013; Friedlingstein et al., 2014)), pointing to remaining
75 gaps in critical process understanding from such datasets and extrapolating observations from past climate conditions to the future often comes with increased uncertainty.

EMEs have been used to parameterise, validate and test models, ranging from drought (McDowell et al., 2013; Powell et al., 2013), warming (Ma et al., 2022; Parton et al., 2007; Zaehle et al., 2010), fertilisation (Meyerholt & Zaehle, 2015; Thomas et al., 2013), or decomposition (Bonan et al., 2013) experiments. A step change came about with the FACE Model Data Synthesis (MDS) project (De Kauwe et al., 2014; Medlyn et al., 2015; Walker et al., 2014; Zaehle et al., 2014) which synthesised data from two temperate forest Free Air Carbon Dioxide Enrichment EMEs. The project asked not only if the models match the data but why the models match the data, making full use of manipulative experiments and being able to plot a path to model improvement. This approach led to the identification of both key processes that need to be included in models to correctly
85 capture responses to elevated CO₂ (e.g., flexible but realistic tissue stoichiometry, flexible biomass allocation, flexible leaf mass per area, organic N uptake) alongside areas where more data is needed to identify realistic and reliable process representation (e.g. water limitation effects on photosynthesis and transpiration, wood turnover, detailed photosynthesis information, N losses). Other studies are now following up on these results to further improve models (Caldararu et al., 2020). Since then, there have been several other MDS studies following the same philosophy, including other FACE experiments
90 (Fleischer et al., 2019; Medlyn et al., 2016) and FACE plus warming experiments (De Kauwe et al., 2017). A few such modelling studies were performed prior to the start of the actual experiment (Fleischer et al., 2019; Medlyn et al., 2016; Parton et al., 2007) with the aim of informing experimental design and in the case of (Parton et al., 2007). Their model predictions were later tested against experimental results (Dijkstra et al., 2010), showcasing the power of such a study.

95 There has been an increase in the number of EMEs globally, as well as in the amount of open data and efforts to synthesise and bring together existing experiments (Cleland et al., 2019; Liang et al., 2020; Van Sundert et al., 2023). The present issue includes some examples of new experiments including a large-scale drought and fertilisation experiment (Vargas G. et al.,



2022) and EME-model integration, including studies looking at new nutrient models (Cornut, Delpierre, et al., 2022; Cornut, le Maire, et al., 2022; Eastman et al., 2023; Li et al., 2022), the effects of root distribution of water availability response (Kulmatiski et al., 2023), carbon storage in grasslands (Wilcox et al., 2022) and extreme events (Holm et al., 2022).

Global observations, ecosystem experiments, and process-based models all have their strengths and weaknesses in terms of scalability and process understanding (Figure 1). To advance ecosystem science an integration of complementary sources of observations and models is needed to achieve the full predictive potential of the tools available. The majority of EME-model integration studies have focused on constraining or evaluating models with experimental data, however, there is a large untapped potential to use models and EMEs bidirectionally, as they can bridge the process understanding - scalability gap. While models are capable of making global scale predictions on long timescales, they also include a detailed process representation which can be a reflection of the insight gained from EMEs. In this paper we set out a roadmap to using models together with EMEs not only as a one-way street for validating models but as an integrated workflow for improving experimental design and generalising experimental conclusions.

2 Not a one sided relationship: what can models do for experiments?

In past studies, the focus has been on integrating conclusions from EMEs into models and evaluating such models against experimental data to improve model process representation (Medlyn et al., 2015; Norby et al., 2016). However, this relationship can easily be turned around. Models can be used as digital twins or sandboxes to explore different possibilities for experimental design, set realistic and practical treatment levels and formulate quantitative, defensible hypotheses. Models can also be used to integrate observations gained from one or more experiments for generalising in time and space

2.1 Hypothesis generation and experimental design

Ecosystem manipulation experiments are designed around testable research questions and hypotheses, following established theory. Hypotheses are generally of a qualitative and first-order nature, i.e., ‘if higher A then B’, for example nutrient addition leads to increased growth and photosynthesis or drought reduces growth. However, ecosystem responses are affected by multiple feedbacks, operating at diverse time scales. Responses can be non-linear, even threshold-like, and interactive, and can be affected by plant adaptations, species-specific responses, and plant-soil interactions. Certain processes saturate at higher levels reached over time, such as nutrient addition or elevated CO₂, where continued application or the treatment does not lead to a further ecosystem response, as systems become limited by other factors. In the case of edaphic factors, such as nutrient availability, spatial factors, such as underlying fertility and history of deposition interact with future loading. Ecosystem responses have also been shown to vary with the intensity and duration of the treatment (Niu et al., 2022). This means that it is often difficult to (explicitly) generate hypotheses that reflect multiple known individual processes and their (unknown) interactive effects. Using a model, or multiple models, or multiple process representations within one modelling framework can aid the hypothesis generation.



130 As many responses to environmental drivers are non-linear, it can be uncertain when designing an experiment what and how many treatment levels are necessary, especially if choosing a response surface rather than a replicated approach (Kreyling et al., 2018) This is particularly important as ecosystem-level experiments are technically challenging and expensive, especially those on the scale of FACE experiments. There are financial, logistical, and practical constraints on how many treatment levels can be considered. Models can easily be run with multiple treatment levels, which would allow identification of key nonlinearities and a defined response surface which can be used as a basis for choosing experimental treatments.

135 Models can also potentially be used to define a sampling strategy. Initial model simulations can help identify which variables are likely to respond to the experimental treatment and at what time scales treatment responses are likely to be detected. If using multiple models or a model with multiple alternative process representations, this can also help identify variables that would help test the proposed hypotheses, by looking at model variables that respond differently for different model formulations. We propose a workflow (Figure 2) in which one or multiple models are run with local site conditions with multiple treatment levels. This then allows us to identify hypotheses based on predicted responses, ideal treatment levels and key variables to measure. This can be done multiple times to explore different experimental setups and ideas.

140 There is often a gap between the optimum sampling strategy and what is actually feasible in field experiments - variables often not be measured with a sufficient temporal or spatial frequency due to either costs or technical difficulties. While we as a community have come a very long way in automating certain measurements - carbon and water fluxes, spectral properties, soil moisture - many remain time and labour intensive, such as direct measurements of biomass above and below-ground or plant and soil chemical composition. Even more, other quantities which might be useful to know are not possible to measure, that is they are latent or conceptual variables such as plant nutrient demand. Such variables are often key in explaining the observed responses, but can usually only be guessed at.

145 In contrast, models produce all variables consistently at daily or sub daily intervals and predictions, including conceptual or latent quantities which can be expressed mathematically can therefore be used alongside experimental results to explain observed responses. For example, the issue of nutrient limitation under elevated CO₂ involves many plant and soil processes, many of which cannot be monitored easily but are included in nutrient-enabled models (Caldararu et al., 2022). One potential issue here is that models need to be able to accurately represent ecosystem processes at a given particular experimental site. Before being used in such a scenario, models should be evaluated or even calibrated for the site to make sure of the model performance.

2.2 Generalising in time and space

150 Most manipulative ecosystem experiments are commonly performed at the plot level and are only maintained for a short time period (1- 5 years), with a few exceptional experiments lasting in the range of ten years or more (Eastman et al., 2021; Magill et al., 2004). While a particular experiment can tell us a lot about the mechanistic process involved, it remains an open question how scalable they are in time and space. While there are increasing efforts to standardise experimental treatments through initiatives such as the Nutrient Network (Borer et al., 2014), only a small and inevitably imperfectly distributed area of the



Earth's vegetated surface can ever be covered by experimental treatments. However, models can be run over relatively long periods of time and globally, helping to test the generality of the process understanding informed by EMEs. .

The problem of spatial scaling is obvious. Can the particular ecosystem studied represent other ecosystems of the same type and can plant or soil processes inferred from observed response be generalised? Such upscaling has been previously performed by accounting for moderating factors of the response, such as climate or soil conditions (Terrer et al., 2021). However, such approaches lack the mechanistic and dynamic underpinning provided by a process-based model and could thus miss important feedbacks.

A particular challenge is posed by the step shape of imposed environmental change in EMEs, which contrasts with the gradual decadal-scale environmental change to which ecosystems are exposed. Several processes make the responses to a continuous change potentially different from those to a step change. Plants acclimate to the new conditions over time (Reich et al., 2018), a process that has been primarily studied for changes in temperature but is likely to occur for most environmental drivers (see Section 5). Species composition is likely to change under different environmental conditions, but it can likely only be observed during the duration of the experiment in systems with high species turnover such as grasslands (Avolio et al., 2014). Although ecosystem responses in observable quantities can often not be assumed to be time-invariant, particularly if slow plant-soil feedbacks are involved and even if conditions were held constant after a step-change, underlying processes often can. Mechanistic models that simulate responses to step changes, and that are constrained by data from step-change experiments are thus a device for translating empirical insights gained from step-change experiments into predictions that respect *a priori* knowledge of slow processes and their role in shaping decadal-scale ecosystem dynamics.

A challenge for model-enabled temporal upscaling are slow processes that are not represented in models or insufficiently constrained by relatively short-term observations (1-10 years in EMEs), but can modify ecosystem responses at the longer term. Acclimation of photosynthesis and respiration to changes temperature and CO₂ (Prentice et al., 2014; Way & Yamori, 2014) and plasticity in allocation in response to carbon-nutrient balance changes (Poorter et al., 2012; Zaehle et al., 2014) largely lack reliable representations in models, and may have a strong influence on decadal to centennial carbon, nitrogen, and phosphorous supply and demand, and ecosystem dynamics. On even longer timescales, genetic adaptation through selection under new environmental pressures and changes in species compositions may emerge as strong drivers of ecosystem responses. Some experiments on shorter-lived, herbaceous species have shown shifts in species composition in response to elevated CO₂ (Reich et al., 2018), similar changes in longer-lived organisms are harder to observe in EMEs. These types of responses and their lack of representation in ecosystem models imply limits for data-constrained and model-based temporal up-scaling. In terms of short-term plasticity or acclimation, recent advances in using eco-evolutionary optimality in models (Harrison et al., 2021) can represent plastic plant responses, and have successfully been used together with EMEs (Caldararu et al., 2020; Sabot et al., 2022). However, current ecosystem models are not well suited for dealing with changes in species composition, demography, and competition processes. However, trait-based, individual or cohort-based models (Fisher et al., 2015). Particularly for evaluating simulated acclimation and other phenotypic plasticity that operate at time scales of months to years, it will be important to make targeted use of insights gained from EMEs. Since this is a knowledge gap in both EMEs and



models, it is an opportunity for both communities to work side by side, rather than sequentially, to advance our knowledge of plant adaptation.

In recent years, there has been a global effort to overcome the limitations of the different approaches, and try to solve the need of standardised controlled experiments on wide temporal and spatial scales. (Fraser et al., 2013) coined the term “coordinated distributed experiments” (CDEs) to describe a global network of standardised experiments distributed across a wide range of ecosystems and climate zones to account for issues of spatial and temporal scales. While initial CDEs were limited in their spatial coverage to Europe or North America, recent ones are striving to be truly global. Examples of global CDEs include the International Drought Experiment (Knapp et al., 2017) that studies the sensitivity of ecosystems to extreme drought events, the Nutrient Network (Borer et al., 2014) which addresses how grasslands are affected by eutrophication and grazing or the Tree Diversity Network (Paquette et al., 2018) which includes a global collection of tree biodiversity experiments. Coordinated EME networks with standardised protocols and homogenised compilations of published EME data (Van Sundert et al., 2023) can drastically facilitate model-data integration studies, performed for an extended set of sites and experiments. Multi-experiment modelling may be essential for powerful generalisability tests and uncertainty quantification of model predictions, e.g., by a leave-experiment-out cross-validation (see also Section 3). Thus, using experiment networks and compilations together with models would increase confidence in the generality of model predictions and partially deal with the upscaling issues discussed in Section 3.3. The use of CDE results could also help to improve the parametrization of process-based models in ecosystems that are underrepresented, but this would require better coordination between response variables measured in the field and those processes included in the models (N. G. Smith et al., 2014).

3 Model-data integration

Data assimilation (DA) and model data integration (MDI) are broad umbrella terms for a variety of statistical methods that fit process based models parameters to observations. The methods are well established and widely used with remote sensing (Exbrayat et al., 2019; W. K. Smith et al., 2020) and eddy covariance data (Fox et al., 2009; Kuppel et al., 2012), at both site and global scales. There are now a number of DA methods available to the community (Anderson et al., 2009; Fer et al., 2021; Huang et al., 2019), which significantly lowers the barrier to entry for those who want to use such tools.

One of the main issues with using DA methods is that the data used to constrain models are observations in present or past conditions, raising questions about the capacity of resulting models to predict ecosystem responses under future conditions. Therefore, using data from manipulative experiments can be extremely valuable in providing information of yet unobserved conditions. One other common issue with using remote sensing data to parameterise models, is that the observations used in DA need to be variables that are actually represented in models (MacBean et al., 2022), so that most remote sensing indices need to be processed further before they can be used. In contrast, experimental observations provide information that can easily be mapped to model variables - biomass, ecosystem fluxes, soil pools, etc.



MDI provides a formalised approach to make best use of naturally sparse EME observations and combine them with *a priori* understanding embodied in model structures. Integrating diverse ecosystem data can help estimating the system state, given physical constraints that are built into the model (e.g., mass conservation) (Jiang et al., 2020). MDI can also provide an approach to formalised model selection (Mark et al., 2018) and a systematic treatment for trading off model complexity and fit to the data.. EMEs, in contrast with observational data commonly used in MDI studies, often have a particularly strong leverage in discriminating between predictions of alternative models, which other data types often lack. Only if an ecosystem's slow biogeochemical cycling is "hit hard" in an experimental setup, underlying processes are revealed. Thus, EMEs provide key information that is required for robust model selection - the discrimination of alternatively formulated model structures that reflect alternative hypotheses of how ecosystem processes work.

However, typical models used for global biogeochemical cycle and climate change impact simulations are often complex and contain a large number of weakly constrained parameters. In view of the sparsity of EME data, this poses a risk of overfitting. Over-fitting may be mitigated using a "leave-experiment-out cross-validation" approach where one experiment is systematically left out of the model fitting procedure and used as an out-of-sample test. This may be a way to handle the overfitting risk, enable a more robust calibration of model parameters, and provide a more reliable estimate of the spatial generalisation error. However, the environmental space currently covered by EMEs and their available data is limited and gaps remain particularly in the tropics and for CO₂ experiments in all biomes except temperate forests and grasslands (Van Sundert et al., 2023).

4 New data for EME-model synthesis

Despite the potential for model-informed ecosystem experiments discussed above, data availability still defines model use because without data, models are impossible to constrain. Advances in measurement techniques and data processing offer an opportunity to increase the types and frequency of measurement that can be gathered at EMEs. EMEs are one of the best locations to develop new data streams because they are typically already well parameterised, offering established data in addition to that from novel sources and potentially benefiting from frequent site visits necessary for non-standard instrument development. Thus in addition to the direct feedback with individual EMEs, models can set the agenda for data development through transparent discussion of the uncertainty and parameter sensitivity.

Perhaps the most straightforward, EME experiments can be equipped with proximal sensing devices such as phenocams (Brown et al., 2016) or sun induced fluorescence (SIF) sensors (Yang et al., 2015). These provide continuous regular measurements of (generally) canopy properties. While these types of measurements are available from spaceborne sensors, local measurements provide more spatial detail and often the possibility of measuring each individual plot separately, thus identifying treatment effects in smaller scale EMEs. Unlike spaceborne instruments, proximal sensors usually require calibrating sensors between different treatments but on the other hand provide regular, location specific information. Such



260 continuous measurements are extremely useful for model evaluation and development as they are often done at a temporal
frequency similar to that of a model timestep and can provide information on short-term changes in vegetation responses.
Additionally, such data can also be used in DA contexts (Section 3) much more easily than irregular, spot measurements. One
application of note on the horizon is capturing patterns of functional diversity, which can be validated in the field (Pacheco-
Labrador et al., 2022) and through EME explore causative mechanisms behind changes in environment, changes in function
265 and changes in species or functional diversity. As models move more towards representing species diversity (Kauwe et al.,
2015) and new traits based demographic models become more widespread (Xu & Trugman, 2021), such measurements in the
context of EMEs can provide an invaluable source of process understanding.

Given widespread coverage, data captured from proximal sensing devices or field measurements may also potentially be
upscaled via wide networks (which do not necessarily rely on EMEs). Many modern approaches to this problem rely on
270 machine learning (ML) methods to reach from regional or global scales (Lapeyre et al., 2020; Poggio et al., 2021)). But in
EMEs and similar contexts, ML also offers another complementary tool: data streams which are either difficult to capture or
process (Nair et al., 2023) can be gap filled or interpreted with higher confidence and representativeness. Further development
of techniques currently only possible at laboratory or homogenous agricultural scale may allow dynamic sub annual time series
of difficult to measure parameters such as photosynthetic capacity (Heckmann et al., 2017), nutrient pools both in biomass and
275 available in soil (Tan et al., 2022), or phenological dynamics beyond leaves, especially those belowground (Wang et al., 2022).
This is particularly relevant belowground, where data are particularly sparse. However technical and logistical limitations
remain and on-site instrumentation is considerably more challenging than laboratory studies. A key challenge here is capturing
both spatial and temporal dynamics at the same time. In many cases many sensor nodes are preferable because ecosystem
properties in general and belowground properties in particular are very heterogeneous. However, when such technologies are
280 fully deployable in the field and an EME context, they can provide datastreams for models which were not previously available,
reducing uncertainty in difficult to measure processes.

Eddy covariance (EC) data are one of the most frequently used datastreams for model evaluation, parameterization and data
assimilation. The advantage of using EC data is its high temporal frequency and the large number of globally distributed sites
285 where such data is available. However, as detailed above (Fig. 1), they are purely observational measurements and lack the
specific advantages of an EME. A small number of EC sites do use a treatment, albeit on an unreplicated treatment scale (El-
Madany et al., 2021; Zhao et al., 2022). EC experiments are laborious to construct and expensive to operate in tall stature
ecosystems, and even in shorter ecosystems the footprint of an EC tower is a challenge for application of many global change
relevant treatments. Indeed, only a small number of manipulations can be conducted effectively on this scale and thus far are
290 limited to nutrient treatments. These designs also necessitate a well-supported baseline of major ecosystem parameters before
treatments to partially substitute for replication. This is a necessary compromise versus the scale and unprecedented realism
which can be achieved.



295 A key aspect of several modern designs is dispensing of replication. While this is controversial in the classic model of the
scientific methods, as we highlighted with the EC measurements example this offers advantages in this case for realism which
could not reasonably be captured in other fashions. Similarly, quantitative gradients allow parameterisation of response
surfaces (as opposed to step changes in standard EMEs) but limit possible replication (Kreyling et al., 2018; Piepho et al.,
2017). In a model-EME context, this allows the models to be evaluated at multiple treatment levels, circumventing the step
change issues. Large, full ecosystem level experiments with sometimes groundbreaking treatments such as the TEMPEST
flooding experiment (Hopple et al., 2023), are also not replicated but provide invaluable insights into ecosystem functioning.
300 An alternative example is the unique Biosphere2 setup, which allowed a fully traceable experiment on a ‘whole ecosystem’,
albeit without replication (Werner et al., 2021). This would allow intensive tracing of treatment effects in models through a
process-by-process validation, rather than the usual validation against end results such as biomass responses.

5 Conclusions and outlook

305 Ecosystem manipulative experiments and process-based models are both powerful tools of understanding how terrestrial
ecosystems are going to respond to future global change. However, on their own, they each suffer from their own particular
short-comings, including lack of generality in time and space in the case of experiments and lack of process realism in the case
of models. We propose that used in combination, these two tools can achieve their full potential. Going beyond a one-sided
process where data from experiments is used to inform models, we lay out a roadmap (Figure 2) for using experiments and
models alongside each other, and alongside new types of measurements and techniques, as complementary sources of
310 information. This approach will further our understanding of how terrestrial ecosystems work and increase our predictive
capability of the future of terrestrial ecosystems.

Acknowledgments

We would like to acknowledge the help of Dr. Sönke Zaehle and Dr. Karin Rebel for their comments on early drafts of this
opinion piece. B.D.S. was funded by the Swiss National Science Foundation grant PCEFP2_181115. This work is a
315 contribution to the LEMONTREE (Land Ecosystem Models based On New Theory, obseRvations and ExperimEnts) project,
funded through the generosity of Eric and Wendy Schmidt by recommendation of the Schmidt Futures program (B.D.S.).

Conflict of interest

The authors declare no conflict of interest.



References

- Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., & Avellano, A. (2009). The Data Assimilation Research Testbed: A Community Facility. *Bulletin of the American Meteorological Society*, 90(9), 1283–1296. <https://doi.org/10.1175/2009BAMS2618.1>
- Arora, V. K., Boer, G. J., Friedlingstein, P., Eby, M., Jones, C. D., Christian, J. R., Bonan, G., Bopp, L., Brovkin, V., Cadule, P., Hajima, T., Ilyina, T., Lindsay, K., Tjiputra, J. F., & Wu, T. (2013). Carbon-concentration and carbon-climate feedbacks in CMIP5 earth system models. *Journal of Climate*, 26(15), 5289–5314. <https://doi.org/10.1175/JCLI-D-12-00494.1>
- Avolio, M. L., Koerner, S. E., La Pierre, K. J., Wilcox, K. R., Wilson, G. W. T., Smith, M. D., & Collins, S. L. (2014). Changes in plant community composition, not diversity, during a decade of nitrogen and phosphorus additions drive above-ground productivity in a tallgrass prairie. *The Journal of Ecology*, 102(6), 1649–1660. <https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/1365-2745.12312>
- Baldocchi, D. D. (2020). How eddy covariance flux measurements have contributed to our understanding of Global Change Biology. *Global Change Biology*, 26(1), 242–260. <https://doi.org/10.1111/gcb.14807>
- Bonan, G. B., Hartman, M. D., Parton, W. J., & Wieder, W. R. (2013). Evaluating litter decomposition in earth system models with long-term litterbag experiments: an example using the Community Land Model version 4 (CLM4). *Global Change Biology*, 19(3), 957–974. <https://doi.org/10.1111/gcb.12031>
- Borer, E. T., Harpole, W. S., Adler, P. B., Lind, E. M., Orrock, J. L., Seabloom, E. W., & Smith, M. D. (2014). Finding generality in ecology: a model for globally distributed experiments. *Methods in Ecology and Evolution / British Ecological Society*, 5(1), 65–73. <https://doi.org/10.1111/2041-210x.12125>
- Brown, T. B., Hultine, K. R., Steltzer, H., Denny, E. G., Denslow, M. W., Granados, J., Henderson, S., Moore, D., Nagai, S., SanClements, M., Sánchez-Azofeifa, A., Sonnentag, O., Tazik, D., & Richardson, A. D. (2016). Using phenocams to monitor our changing Earth: toward a global phenocam network. *Frontiers in Ecology and the Environment*, 14(2), 84–93. <https://doi.org/10.1002/fee.1222>
- Caldararu, S., Thum, T., Yu, L., Kern, M., Nair, R., & Zaehle, S. (2022). Long-term ecosystem nitrogen limitation from foliar $\delta^{15}\text{N}$ data and a land surface model. *Global Change Biology*, 28(2), 493–508. <https://doi.org/10.1111/gcb.15933>
- Caldararu, S., Thum, T., Yu, L., & Zaehle, S. (2020). Whole-plant optimality predicts changes in leaf nitrogen under variable CO_2 and nutrient availability. *The New Phytologist*, 225(6), 2331–2346. <https://nph.onlinelibrary.wiley.com/doi/abs/10.1111/nph.16327>
- Cleland, E. E., Lind, E. M., DeCrappeo, N. M., DeLorenze, E., Wilkins, R. A., Adler, P. B., Bakker, J. D., Brown, C. S., Davies, K. F., Esch, E., Firn, J., Gressard, S., Gruner, D. S., Hagenah, N., Stanley Harpole, W., Hautier, Y., Hobbie, S. E., Hofmockel, K. S.,



- 350 Kirkman, K., ... Seabloom, E. W. (2019). Belowground Biomass Response to Nutrient Enrichment Depends on Light Limitation Across Globally Distributed Grasslands. In *Ecosystems* (Vol. 22, Issue 7, pp. 1466–1477). <https://doi.org/10.1007/s10021-019-00350-4>
- 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. <https://doi.org/10.1029/2018ms001354>
- 355 Cornut, I., Delpierre, N., Laclau, J.-P., Guillemot, J., Nouvellon, Y., Campoe, O., Stape, J. L., Fernanda Santos, V., & le Maire, G. (2022). Potassium-limitation of forest productivity, part 1: A mechanistic model simulating the effects of potassium availability on canopy carbon and water fluxes in tropical eucalypt stands. *EGUsphere*, 2022, 1–37. <https://doi.org/10.5194/egusphere-2022-883>
- Cornut, I., le Maire, G., Laclau, J.-P., Guillemot, J., Nouvellon, Y., & Delpierre, N. (2022). Potassium-limitation of forest productivity, part 2: CASTANEA-MAESPA-K shows a reduction in photosynthesis rather than a stoichiometric limitation of tissue formation. *EGUsphere*, 2022, 1–27. <https://doi.org/10.5194/egusphere-2022-884>
- 360 Dalling, J. W., Winter, K., Andersen, K. M., & Turner, B. L. (2013). Artefacts of the pot environment on soil nutrient availability: implications for the interpretation of ecological studies. *Plant Ecology*, 214(2), 329–338. <https://doi.org/10.1007/s11258-013-0172-3>
- De Kauwe, M. G., Medlyn, B. E., Walker, A. P., Zaehle, S., Asao, S., Guenet, B., Harper, A. B., Hickler, T., Jain, A. K., Luo, Y., Lu, X., Luus, K., Parton, W. J., Shu, S., Wang, Y.-P., Werner, C., Xia, J., Pendall, E., Morgan, J. A., ... Norby, R. J. (2017). Challenging terrestrial biosphere models with data from the long-term multifactor Prairie Heating and CO₂ Enrichment experiment. *Global Change Biology*, 23(9), 3623–3645. <https://doi.org/10.1111/gcb.13643>
- 365 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. <https://doi.org/10.1111/nph.12847>
- 370 Dijkstra, F. A., Blumenthal, D., Morgan, J. A., Pendall, E., Carrillo, Y., & Follett, R. F. (2010). Contrasting effects of elevated CO₂ and warming on nitrogen cycling in a semiarid grassland. *The New Phytologist*, 187(2), 426–437. <https://doi.org/10.1111/j.1469-8137.2010.03293.x>
- Eastman, B. A., Adams, M. B., Brzostek, E. R., Burnham, M. B., Carrara, J. E., Kelly, C., McNeil, B. E., Walter, C. A., & Peterjohn, W. T. (2021). Altered plant carbon partitioning enhanced forest ecosystem carbon storage after 25 years of nitrogen additions. *The New Phytologist*, 230(4), 1435–1448. <https://doi.org/10.1111/nph.17256>



- Eastman, B. A., Wieder, W. R., Hartman, M. D., Brzostek, E. R., & Peterjohn, W. T. (2023). Can models adequately reflect how long-term nitrogen enrichment alters the forest soil carbon cycle? *Biogeosciences Discussions*, 2023, 1–33. <https://doi.org/10.5194/bg-2023-36>
- 380 El-Madany, T. S., Reichstein, M., Carrara, A., Martín, M. P., Moreno, G., Gonzalez-Cascon, R., Peñuelas, J., Ellsworth, D. S., Burchard-Levine, V., Hammer, T. W., Knauer, J., Kolle, O., Luo, Y., Pacheco-Labrador, J., Nelson, J. A., Perez-Priego, O., Rolo, V., Wutzler, T., & Migliavacca, M. (2021). How nitrogen and phosphorus availability change water use efficiency in a Mediterranean Savanna ecosystem. *Journal of Geophysical Research. Biogeosciences*, 126(5). <https://doi.org/10.1029/2020jg006005>
- Exbrayat, J.-F., Bloom, A. A., Carvalhais, N., Fischer, R., Huth, A., MacBean, N., & Williams, M. (2019). Understanding the Land Carbon Cycle with Space Data: Current Status and Prospects. *Surveys in Geophysics*, 40(4), 735–755. [https://doi.org/10.1007/s10712-](https://doi.org/10.1007/s10712-019-09506-2)
- 385 019-09506-2
- 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., Hoffman, F. M., Johnston, M. R., Kooper, R., LeBauer, D. S., Mantooth, J., Parton, W. J., Poulter, B., Quaife, T., Raiho, A., ... Dietze, M. C. (2021). Beyond ecosystem modeling: A roadmap to community cyberinfrastructure for ecological data-model integration. *Global Change Biology*, 27(1), 13–26. <https://doi.org/10.1111/gcb.15409>
- 390 Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. In *International Journal of Climatology* (Vol. 37, Issue 12, pp. 4302–4315). <https://doi.org/10.1002/joc.5086>
- Fisher, R. A., Muszala, S., Versteinstein, M., Lawrence, P., Xu, C., McDowell, N. G., Knox, R. G., Koven, C., Holm, J., Rogers, B. M., Lawrence, D., & Bonan, G. (2015). *Taking off the training wheels: the properties of a dynamic vegetation model without climate envelopes* [Data set]. <https://doi.org/10.5194/gmdd-8-3293-2015>
- 395 Fleischer, K., Rammig, A., De Kauwe, M. G., Walker, A. P., Domingues, T. F., Fuchslueger, L., Garcia, S., Goll, D. S., Grandis, A., Jiang, M., Haverd, V., Hofhansl, F., Holm, J. A., Kruijt, B., Leung, F., Medlyn, B. E., Mercado, L. M., Norby, R. J., Pak, B., ... Lapola, D. M. (2019). Amazon forest response to CO₂ fertilization dependent on plant phosphorus acquisition. *Nature Geoscience*, 12(9), 736–741. <https://doi.org/10.1038/s41561-019-0404-9>
- Food and Agriculture Organization, & Food and Agriculture Organization of the United Nations. (2008). *Harmonized World Soil Database: Fao/Iiasa/isric/isscas/jrc (Version 1.0)*. Food & Agriculture Org. https://books.google.com/books/about/Harmonized_World_Soil_Database.html?hl=&id=N0VoPgAACAAJ
- 400 Fox, A., Williams, M., Richardson, A. D., Cameron, D., Gove, J. H., Quaife, T., Ricciuto, D., Reichstein, M., Tomelleri, E., Trudinger, C. M., & Wijk, M. T. V. (2009). The REFLEX project: Comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data. *Agricultural and Forest Meteorology*, 149(10), 1597–1615.



- 405 <https://doi.org/10.1016/j.agrformet.2009.05.002>
- Fraser, L. H., Henry, H. A. L., Carlyle, C. N., White, S. R., Beierkuhnlein, C., Cahill, J. F., Jr, Casper, B. B., Cleland, E., Collins, S. L., Dukes, J. S., Knapp, A. K., Lind, E., Long, R., Luo, Y., Reich, P. B., Smith, M. D., Sternberg, M., & Turkington, R. (2013). Coordinated distributed experiments: an emerging tool for testing global hypotheses in ecology and environmental science. *Frontiers in Ecology and the Environment*, 11(3), 147–155. <https://doi.org/10.1890/110279>
- 410 Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., & Knutti, R. (2014). Uncertainties in CMIP5 Climate Projections due to Carbon Cycle Feedbacks. *Journal of Climate*, 27(2), 511–526. <https://doi.org/10.1175/JCLI-D-12-00579.1>
- Harris, I., Osborn, T. J., Jones, P., & Lister, D. (2020). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data*, 7(1), 109. <https://doi.org/10.1038/s41597-020-0453-3>
- Harrison, S. P., Cramer, W., Franklin, O., Prentice, I. C., Wang, H., Brännström, Å., de Boer, H., Dieckmann, U., Joshi, J., Keenan, T. F., 415 Lavergne, A., Manzoni, S., Mengoli, G., Morfopoulos, C., Peñuelas, J., Pietsch, S., Rebel, K. T., Ryu, Y., Smith, N. G., ... Wright, I. J. (2021). Eco-evolutionary optimality as a means to improve vegetation and land-surface models. *The New Phytologist*, 231(6), 2125–2141. <https://doi.org/10.1111/nph.17558>
- Heckmann, D., Schlüter, U., & Weber, A. P. M. (2017). Machine Learning Techniques for Predicting Crop Photosynthetic Capacity from Leaf Reflectance Spectra. *Molecular Plant*, 10(6), 878–890. <https://doi.org/10.1016/j.molp.2017.04.009>
- 420 Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangquan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PloS One*, 12(2), e0169748. <https://doi.org/10.1371/journal.pone.0169748>
- Holm, J. A., Medvigy, D. M., Smith, B., Dukes, J. S., Beier, C., Mishurov, M., Xu, X., Lichstein, J. W., Allen, C. D., Larsen, K. S., Luo, 425 Y., Ficken, C., Pockman, W. T., Anderegg, W. R. L., & Rammig, A. (2022). Exploring the impacts of unprecedented climate extremes on forest ecosystems: hypotheses to guide modeling and experimental studies. *Biogeosciences Discussions*, 2022, 1–45. <https://doi.org/10.5194/bg-2022-65>
- Hopple, A. M., Doro, K. O., Bailey, V. L., Bond-Lamberty, B., McDowell, N., Morris, K. A., Myers-Pigg, A., Pennington, S. C., Regier, P., Rich, R., Sengupta, A., Smith, R., Stegen, J., Ward, N. D., Woodard, S. C., & Megonigal, J. P. (2023). Attaining freshwater and 430 estuarine-water soil saturation in an ecosystem-scale coastal flooding experiment. *Environmental Monitoring and Assessment*, 195(3), 425. <https://doi.org/10.1007/s10661-022-10807-0>
- Huang, Y., Stacy, M., Jiang, J., Sundi, N., Ma, S., Saruta, V., Jung, C. G., Shi, Z., Xia, J., Hanson, P. J., & Others. (2019). Realized



ecological forecast through an interactive Ecological Platform for Assimilating Data (EcoPAD, v1. 0) into models. *Geoscientific Model Development*, 12(3), 1119–1137. <https://gmd.copernicus.org/articles/12/1119/2019/>

435 Jiang, M., Medlyn, B. E., Drake, J. E., Duursma, R. A., Anderson, I. C., Barton, C. V. M., Boer, M. M., Carrillo, Y., Castañeda-Gómez, L., Collins, L., Crous, K. Y., De Kauwe, M. G., Dos Santos, B. M., Emmerson, K. M., Facey, S. L., Gherlenda, A. N., Gimeno, T. E., Hasegawa, S., Johnson, S. N., ... Ellsworth, D. S. (2020). The fate of carbon in a mature forest under carbon dioxide enrichment. *Nature*, 580(7802), 227–231. <https://doi.org/10.1038/s41586-020-2128-9>

440 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., Lombardozi, 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. <https://doi.org/10.5194/bg-17-1343-2020>

Karger, D. N., Schmatz, D. R., Detting, G., & Zimmermann, N. E. (2020). High-resolution monthly precipitation and temperature time series from 2006 to 2100. In *Scientific Data* (Vol. 7, Issue 1). <https://doi.org/10.1038/s41597-020-00587-y>

445 Kattge, J., Bönisch, G., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Tautenhahn, S., Werner, G. D. A., Aakala, T., Abedi, M., Acosta, A. T. R., Adamidis, G. C., Adamson, K., Aiba, M., Albert, C. H., Alcántara, J. M., Alcázar C. C., Aleixo, I., Ali, H., ... Wirth, C. (2020). TRY plant trait database - enhanced coverage and open access. *Global Change Biology*, 26(1), 119–188. <https://doi.org/10.1111/gcb.14904>

450 Kauwe, M. G. D., De Kauwe, M. G., Zhou, S.-X., Medlyn, B. E., Pitman, A. J., Wang, Y.-P., Duursma, R. A., & Prentice, I. C. (2015). Do land surface models need to include differential plant species responses to drought? Examining model predictions across a mesic-xeric gradient in Europe. In *Biogeosciences* (Vol. 12, Issue 24, pp. 7503–7518). <https://doi.org/10.5194/bg-12-7503-2015>

Knapp, A. K., Avolio, M. L., Beier, C., Carroll, C. J. W., Collins, S. L., Dukes, J. S., Fraser, L. H., Griffin-Nolan, R. J., Hoover, D. L., Jentsch, A., Loik, M. E., Phillips, R. P., Post, A. K., Sala, O. E., Slette, I. J., Yahdjian, L., & Smith, M. D. (2017). Pushing precipitation to the extremes in distributed experiments: recommendations for simulating wet and dry years. *Global Change Biology*, 23(5), 1774–1782. <https://doi.org/10.1111/gcb.13504>

455 Kreyling, J., Schweiger, A. H., Bahn, M., Ineson, P., Migliavacca, M., Morel-Journel, T., Christiansen, J. R., Shtickzelle, N., & Larsen, K. S. (2018). To replicate, or not to replicate - that is the question: how to tackle nonlinear responses in ecological experiments. *Ecology Letters*, 21(11), 1629–1638. <https://doi.org/10.1111/ele.13134>

460 Kulmatiski, A., Holdrege, M. C., Chirvasa, C., & Beard, K. H. (2023). Root distributions predict shrub-steppe responses to precipitation intensity. *Biogeosciences Discussions*, 2023, 1–29. <https://doi.org/10.5194/bg-2023-13>



- Kuppel, S., Peylin, P., Chevallier, F., Bacour, C., Maignan, F., & Richardson, A. D. (2012). Constraining a global ecosystem model with multi-site eddy-covariance data. *Biogeosciences Discussions*, 9(3), 3317–3380. <https://doi.org/10.5194/bg-9-3317-2012>
- Lapeyre, C. J., Cazard, N., Roy, P. T., Ricci, S., & Zaoui, F. (2020). Reconstruction of Hydraulic Data by Machine Learning. *Advances in Hydroinformatics*, 701–715. https://doi.org/10.1007/978-981-15-5436-0_54
- 465 Liang, X., Zhang, T., Lu, X., Ellsworth, D. S., BassiriRad, H., You, C., Wang, D., He, P., Deng, Q., Liu, H., Mo, J., & Ye, Q. (2020). Global response patterns of plant photosynthesis to nitrogen addition: A meta-analysis. *Global Change Biology*, 26(6), 3585–3600. <https://doi.org/10.1111/gcb.15071>
- Li, S., Waring, B. G., Powers, J. S., & Medvigy, D. (2022). Tropical Dry Forest Response to Nutrient Fertilization: A Model Validation and Sensitivity Analysis. *Biogeosciences Discussions*, 2022, 1–31. <https://doi.org/10.5194/bg-2022-243>
- 470 MacBean, N., Liddy, H., Quaife, T., Kolassa, J., & Fox, A. (2022). Building a Land Data Assimilation Community to Tackle Technical Challenges in Quantifying and Reducing Uncertainty in Land Model Predictions. *Bulletin of the American Meteorological Society*, 103(3), E733–E740. <https://doi.org/10.1175/BAMS-D-21-0228.1>
- Magill, A. H., Aber, J. D., Currie, W. S., Nadelhoffer, K. J., Martin, M. E., McDowell, W. H., Melillo, J. M., & Steudler, P. (2004). Ecosystem response to 15 years of chronic nitrogen additions at the Harvard Forest LTER, Massachusetts, USA. *Forest Ecology and Management*, 196(1), 7–28. <https://doi.org/10.1016/j.foreco.2004.03.033>
- 475 Mark, C., Metzner, C., Lautscham, L., Strissel, P. L., Strick, R., & Fabry, B. (2018). Bayesian model selection for complex dynamic systems. In *Nature Communications* (Vol. 9, Issue 1). <https://doi.org/10.1038/s41467-018-04241-5>
- Ma, S., Jiang, L., Wilson, R. M., Chanton, J. P., Bridgham, S., Niu, S., Iversen, C. M., Malhotra, A., Jiang, J., Lu, X., Huang, Y., Keller, J., Xu, X., Ricciuto, D. M., Hanson, P. J., & Luo, Y. (2022). Evaluating alternative ebullition models for predicting peatland methane emission and its pathways via data–model fusion. In *Biogeosciences* (Vol. 19, Issue 8, pp. 2245–2262). <https://doi.org/10.5194/bg-19-2245-2022>
- 480 McDowell, N. G., Fisher, R. A., Xu, C., Domec, J. C., Hölttä, T., Mackay, D. S., Sperry, J. S., Boutz, A., Dickman, L., Gehres, N., Limousin, J. M., Macalady, A., Martínez-Vilalta, J., Mencuccini, M., Plaut, J. A., Ogée, J., Pangle, R. E., Rasse, D. P., Ryan, M. G., ... Pockman, W. T. (2013). Evaluating theories of drought-induced vegetation mortality using a multimodel-experiment framework. *The New Phytologist*, 200(2), 304–321. <https://doi.org/10.1111/nph.12465>
- 485 Medlyn, B. E., De Kauwe, M. G., Zaehle, S., Walker, A. P., Duursma, R. A., Luus, K., Mishurov, M., Pak, B., Smith, B., Wang, Y.-P., Yang, X., Crous, K. Y., Drake, J. E., Gimeno, T. E., Macdonald, C. A., Norby, R. J., Power, S. A., Tjoelker, M. G., & Ellsworth, D. S. (2016). Using models to guide field experiments: a priori predictions for the CO₂ response of a nutrient- and water-limited native



Eucalypt woodland. *Global Change Biology*, 22(8), 2834–2851. <https://doi.org/10.1111/gcb.13268>

490 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, 528.
<https://doi.org/10.1038/nclimate2621>

Meyerholt, J., & Zaehle, S. (2015). The role of stoichiometric flexibility in modelling forest ecosystem responses to nitrogen fertilization.
495 *The New Phytologist*, 208(4), 1042–1055. <https://doi.org/10.1111/nph.13547>

Nair, R., Strube, M., Hertel, M., Kolle, O., Rolo, V., & Migliavacca, M. (2023). High frequency root dynamics: sampling and
interpretation using replicated robotic minirhizotrons. *Journal of Experimental Botany*, 74(3), 769–786.
<https://doi.org/10.1093/jxb/erac427>

Niu, G., Wang, Y., Wang, R., Ning, Q., Guan, H., Yang, J., Lu, X., Han, X., & Huang, J. (2022). Intensity and duration of nitrogen
500 addition jointly alter soil nutrient availability in a temperate grassland. *Journal of Geophysical Research. Biogeosciences*, 127(3).
<https://doi.org/10.1029/2021jg006698>

Norby, R. J., De Kauwe, M. G., Domingues, T. F., Duursma, R. A., Ellsworth, D. S., Goll, D. S., Lapola, D. M., Luus, K. A., MacKenzie,
A. R., Medlyn, B. E., Pavlick, R., Rammig, A., Smith, B., Thomas, R., Thonicke, K., Walker, A. P., Yang, X., & Zaehle, S. (2016).
Model-data synthesis for the next generation of forest free-air CO₂ enrichment (FACE) experiments. *The New Phytologist*, 209(1),
505 17–28. <https://doi.org/10.1111/nph.13593>

Pacheco-Labrador, J., Migliavacca, M., Ma, X., Mahecha, M. D., Carvalhais, N., Weber, U., Benavides, R., Bouriaud, O., Barnoiaea, I.,
Coomes, D. A., Bohn, F. J., Kraemer, G., Heiden, U., Huth, A., & Wirth, C. (2022). Challenging the link between functional and
spectral diversity with radiative transfer modeling and data. *Remote Sensing of Environment*, 280, 113170.
<https://doi.org/10.1016/j.rse.2022.113170>

510 Paquette, A., Hector, A., Castagneyrol, B., Vanhellefont, M., Koricheva, J., Scherer-Lorenzen, M., Verheyen, K., & TreeDivNet. (2018).
A million and more trees for science. *Nature Ecology & Evolution*, 2(5), 763–766. <https://doi.org/10.1038/s41559-018-0544-0>

Parton, W. J., Morgan, J. A., Wang, G., & Del Grosso, S. (2007). Projected ecosystem impact of the Prairie Heating and CO₂ Enrichment
experiment. *The New Phytologist*, 174(4), 823–834. <https://doi.org/10.1111/j.1469-8137.2007.02052.x>

Piepho, H.-P., Herndl, M., Pötsch, E. M., & Bahn, M. (2017). Designing an experiment with quantitative treatment factors to study the
515 effects of climate change. *Journal of Agronomy and Crop Science*, 203(6), 584–592. <https://doi.org/10.1111/jac.12225>

Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., & Rossiter, D. (2021). SoilGrids 2.0: producing



soil information for the globe with quantified spatial uncertainty. *SOIL*, 7(1), 217–240. <https://doi.org/10.5194/soil-7-217-2021>

Poorter, H., Niklas, K. J., Reich, P. B., Oleksyn, J., Poot, P., & Mommer, L. (2012). Biomass allocation to leaves, stems and roots: meta-analyses of interspecific variation and environmental control. *The New Phytologist*, 193(1), 30–50. <https://doi.org/10.1111/j.1469-8137.2011.03952.x>

Powell, T. L., Galbraith, D. R., Christoffersen, B. O., Harper, A., Imbuzeiro, H. M. A., Rowland, L., Almeida, S., Brando, P. M., da Costa, A. C. L., Costa, M. H., Levine, N. M., Malhi, Y., Saleska, S. R., Sotta, E., Williams, M., Meir, P., & Moorcroft, P. R. (2013). Confronting model predictions of carbon fluxes with measurements of Amazon forests subjected to experimental drought. *The New Phytologist*. <https://doi.org/10.1111/nph.12390>

Prentice, I. C., Colin Prentice, I., Bondeau, A., Cramer, W., Harrison, S. P., Hickler, T., Lucht, W., Sitch, S., Smith, B., & Sykes, M. T. (2007). Dynamic Global Vegetation Modeling: Quantifying Terrestrial Ecosystem Responses to Large-Scale Environmental Change. In J. G. Canadell, D. E. Pataki, & L. F. Pitelka (Eds.), *Terrestrial Ecosystems in a Changing World* (pp. 175–192). https://doi.org/10.1007/978-3-540-32730-1_15

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. <https://doi.org/10.1111/ele.12211>

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(10), 5987–6005. <https://doi.org/10.5194/acp-15-5987-2015>

Reich, P. B., Hobbie, S. E., Lee, T. D., & Pastore, M. A. (2018). Unexpected reversal of C3 versus C4 grass response to elevated CO2 during a 20-year field experiment. *Science*, 360(6386), 317–320. <https://doi.org/10.1126/science.aas9313>

Sabot, M. E. B., De Kauwe, M. G., Pitman, A. J., Ellsworth, D. S., Medlyn, B. E., Caldararu, S., Zaehle, S., Crous, K. Y., Gimeno, T. E., Wujeska-Klaue, A., Mu, M., & Yang, J. (2022). Predicting resilience through the lens of competing adjustments to vegetation function. *Plant, Cell & Environment*, 45(9), 2744–2761. <https://doi.org/10.1111/pce.14376>

Seiler, C., Melton, J. R., Arora, V. K., Sitch, S., Friedlingstein, P., Anthoni, P., Goll, D., Jain, A. K., Joetzjer, E., Lienert, S., Lombardozzi, D., Luysaert, S., Nabel, J. E. M. S., Tian, H., Vuichard, N., Walker, A. P., Yuan, W., & Zaehle, S. (2022). Are Terrestrial Biosphere Models Fit for Simulating the Global Land Carbon Sink? *Journal of Advances in Modeling Earth Systems*, 14(5), e2021MS002946. <https://doi.org/10.1029/2021MS002946>

Smith, N. G., Rodgers, V. L., Brzostek, E. R., Kulmatiski, A., Avolio, M. L., Hoover, D. L., Koerner, S. E., Grant, K., Jentsch, A., Fatichi, S., & Niyogi, D. (2014). Toward a better integration of biological data from precipitation manipulation experiments into Earth system models. *Reviews of Geophysics*, 52(3), 412–434. <https://doi.org/10.1002/2014RG000458>



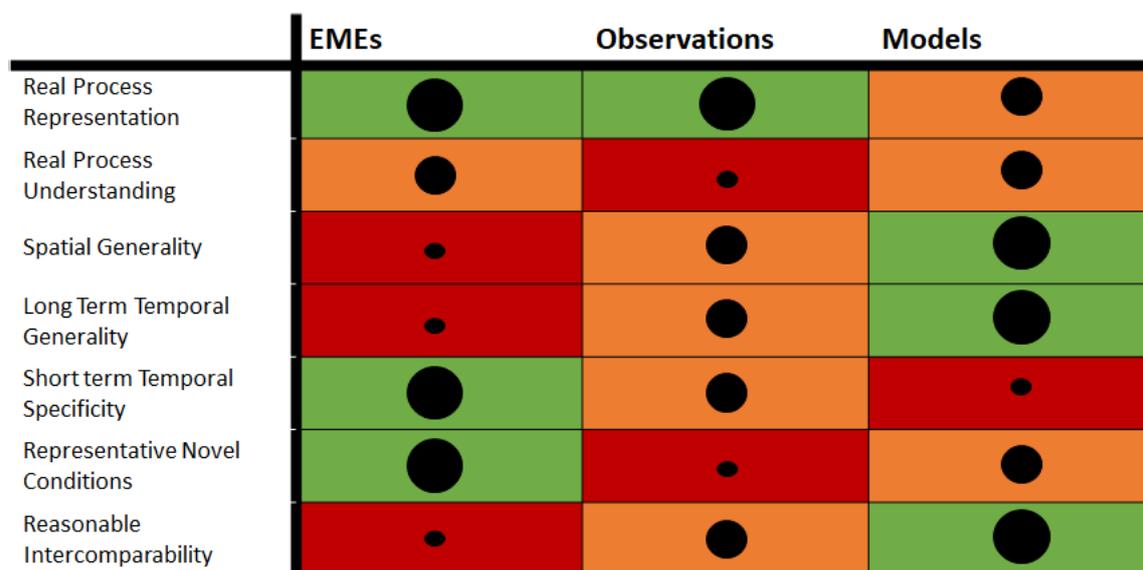
- 545 Smith, W. K., Fox, A. M., MacBean, N., Moore, D. J. P., & Parazoo, N. C. (2020). Constraining estimates of terrestrial carbon uptake: new opportunities using long-term satellite observations and data assimilation. *The New Phytologist*, 225(1), 105–112. <https://doi.org/10.1111/nph.16055>
- Stocker, B. D., Prentice, I. C., Cornell, S. E., Davies-Barnard, T., Finzi, A. C., Franklin, O., Janssens, I., Larmola, T., Manzoni, S., Näsholm, T., Raven, J. A., Rebel, K. T., Reed, S., Vicca, S., Wiltshire, A., & Zaehle, S. (2016). Terrestrial nitrogen cycling in Earth system models revisited. *The New Phytologist*, 210(4), 1165–1168. <https://doi.org/10.1111/nph.13997>
- 550 Tan, B., You, W., Tian, S., Xiao, T., Wang, M., Zheng, B., & Luo, L. (2022). Soil Nitrogen Content Detection Based on Near-Infrared Spectroscopy. *Sensors*, 22(20). <https://doi.org/10.3390/s22208013>
- Terrer, C., Phillips, R. P., Hungate, B. A., Rosende, J., Pett-Ridge, J., Craig, M. E., van Groenigen, K. J., Keenan, T. F., Sulman, B. N., Stocker, B. D., Reich, P. B., Pellegrini, A. F. A., Pendall, E., Zhang, H., Evans, R. D., Carrillo, Y., Fisher, J. B., Van Sundert, K., Vicca, S., & Jackson, R. B. (2021). A trade-off between plant and soil carbon storage under elevated CO₂. *Nature*, 591(7851), 599–603. <https://doi.org/10.1038/s41586-021-03306-8>
- 555 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. <https://doi.org/10.1111/gcb.12281>
- Van Sundert, K., Van Sundert, K., Leuzinger, S., -F. Bader, M. K., Chang, S. X., De Kauwe, M. G., Dukes, J. S., Adam Langley, J., Ma, Z., Mariën, B., Reynaert, S., Ru, J., Song, J., Stocker, B., Terrer, C., Thoresen, J., Vanuytrecht, E., Wan, S., Yue, K., & Vicca, S. (2023). When things get MESI: The Manipulation Experiments Synthesis Initiative—A coordinated effort to synthesize terrestrial global change experiments. In *Global Change Biology*. <https://doi.org/10.1111/gcb.16585>
- 560 Vargas G., G., Perez-Aviles, D., Raczka, N., Pereira-Arias, D., Tijerín-Triviño, J., Pereira-Arias, L. D., Medvigy, D., Waring, B. G., Morrisey, E., Brzostek, E., & Powers, J. S. (2022). Throughfall exclusion and fertilization effects on tropical dry forest tree plantations, a large-scale experiment. *Biogeosciences Discussions*, 2022, 1–32. <https://doi.org/10.5194/bg-2022-203>
- Walker, A. P., Hanson, P. J., De Kauwe, M. G., Medlyn, B. E., Zaehle, S., Asao, S., Dietze, M., Hickler, T., Huntingford, C., Iversen, C. M., Jain, A., Lomas, M., Luo, Y., McCarthy, H., Parton, W. J., Prentice, I. C., Thornton, P. E., Wang, S., Wang, Y.-P., ... Norby, R. J. (2014). Comprehensive ecosystem model-data synthesis using multiple data sets at two temperate forest free-air CO₂ enrichment experiments: Model performance at ambient CO₂ concentration : FACE MODEL-DATA SYNTHESIS. *Journal of Geophysical Research: Biogeosciences*, 119(5), 937–964. <https://doi.org/10.1002/2013JG002553>
- 570 Wang, B., Luke McCormack, M., Ricciuto, D. M., Yang, X., & Iversen, C. M. (2022). Embracing fine-root system complexity to improve the predictive understanding of ecosystem functioning. In *bioRxiv* (p. 2022.10.07.511037). <https://doi.org/10.1101/2022.10.07.511037>



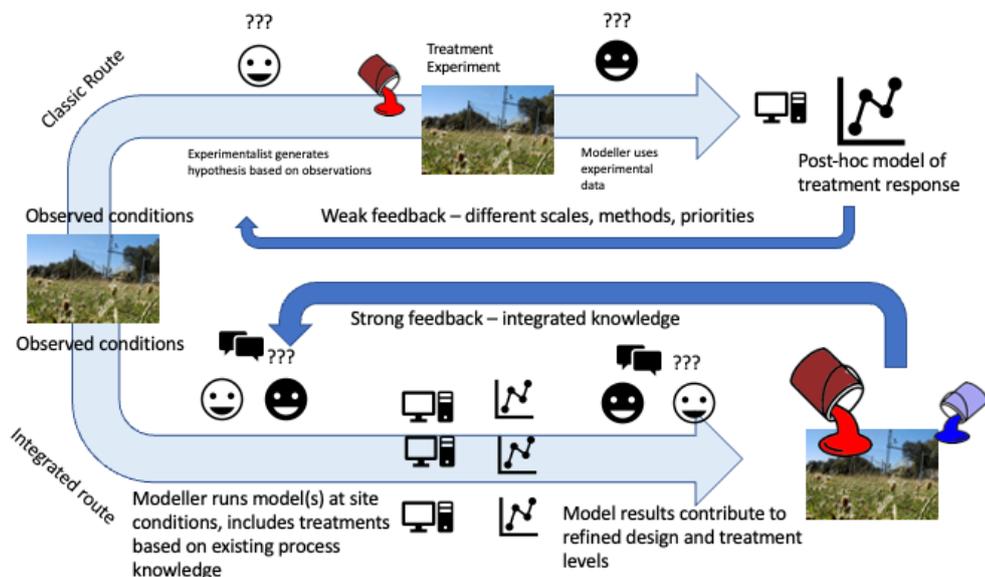
- Way, D. A., & Yamori, W. (2014). Thermal acclimation of photosynthesis: on the importance of adjusting our definitions and accounting for thermal acclimation of respiration. *Photosynthesis Research*, 119(1-2), 89–100. <https://doi.org/10.1007/s11120-013-9873-7>
- 575 Werner, C., Meredith, L. K., Ladd, S. N., Ingrisch, J., Kübert, A., van Haren, J., Bahn, M., Bailey, K., Bamberger, I., Beyer, M., Blomdahl, D., Byron, J., Daber, E., Deleeuw, J., Dippold, M. A., Fudyma, J., Gil-Loaiza, J., Honeker, L. K., Hu, J., ... Williams, J. (2021). Ecosystem fluxes during drought and recovery in an experimental forest. *Science*, 374(6574), 1514–1518. <https://doi.org/10.1126/science.abj6789>
- Wieder, W. R., Lawrence, D. M., Fisher, R. A., Bonan, G. B., Cheng, S. J., Goodale, C. L., Grandy, A. S., Koven, C. D., Lombardozzi, D.,
580 L., Oleson, K. W., & Thomas, R. Q. (2019). Beyond Static Benchmarking: Using Experimental Manipulations to Evaluate Land Model Assumptions. *Global Biogeochemical Cycles*, 33(10), 1289–1309. <https://doi.org/10.1029/2018GB006141>
- Wilcox, K. R., Collins, S. L., Knapp, A. K., Pockman, W., Shi, Z., Smith, M., & Luo, Y. (2022). Assessing carbon storage capacity and saturation across six central US grasslands using data-model integration. *Biogeosciences Discussions*, 2022, 1–52. <https://doi.org/10.5194/bg-2022-164>
- 585 Xiao, J., Chevallier, F., Gomez, C., Guanter, L., Hicke, J. A., Huete, A. R., Ichii, K., Ni, W., Pang, Y., Rahman, A. F., Sun, G., Yuan, W., Zhang, L., & Zhang, X. (2019). Remote sensing of the terrestrial carbon cycle: A review of advances over 50 years. *Remote Sensing of Environment*, 233, 111383. <https://doi.org/10.1016/j.rse.2019.111383>
- Xu, X., & Trugman, A. T. (2021). Trait-Based Modeling of Terrestrial Ecosystems: Advances and Challenges Under Global Change. *Current Climate Change Reports*, 7(1), 1–13. <https://doi.org/10.1007/s40641-020-00168-6>
- 590 Yang, X., Tang, J., Mustard, J. F., Lee, J.-E., Rossini, M., Joiner, J., Munger, J. W., Kornfeld, A., & Richardson, A. D. (2015). Solar-induced chlorophyll fluorescence that correlates with canopy photosynthesis on diurnal and seasonal scales in a temperate deciduous forest. *Geophysical Research Letters*, 42(8), 2977–2987. <https://doi.org/10.1002/2015gl063201>
- Zaehle, S., Friedlingstein, P., & Friend, A. D. (2010). Terrestrial nitrogen feedbacks may accelerate future climate change. *Geophysical Research Letters*, 37(1). <https://doi.org/10.1029/2009gl041345>
- 595 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. <https://doi.org/10.1111/nph.12697>
- Zhao, P., Chi, J., Nilsson, M. B., Löfvenius, M. O., Högberg, P., Jocher, G., Lim, H., Mäkelä, A., Marshall, J., Ratcliffe, J., Tian, X.,
600 Näsholm, T., Lundmark, T., Linder, S., & Peichl, M. (2022). Long-term nitrogen addition raises the annual carbon sink of a boreal



forest to a new steady-state. *Agricultural and Forest Meteorology*, 324, 109112. <https://doi.org/10.1016/j.agrformet.2022.109112>



605 Figure 1 - Strengths and weaknesses of ecosystem manipulation experiments, observations and process based models. Green areas with large circles represent areas that a particular tool is strong at, orange with medium circles areas of medium strength and red with small circles areas where each tool is weak.



610 Figure 2 - Classic workflow (A) contrasted with proposed workflow (B) for using models together with ecosystem manipulative experiments for experimental design and hypothesis generation. Through iterative synthesis of models and experiments in B) and tuning of experimental parameters based on modelled feedback, stronger predictions are generated.