

## **Moving from an experiment-dominated to an observation-dominated era in global change impact research on vegetation**

Sebastian Leuzinger, Belinda E. Medlyn

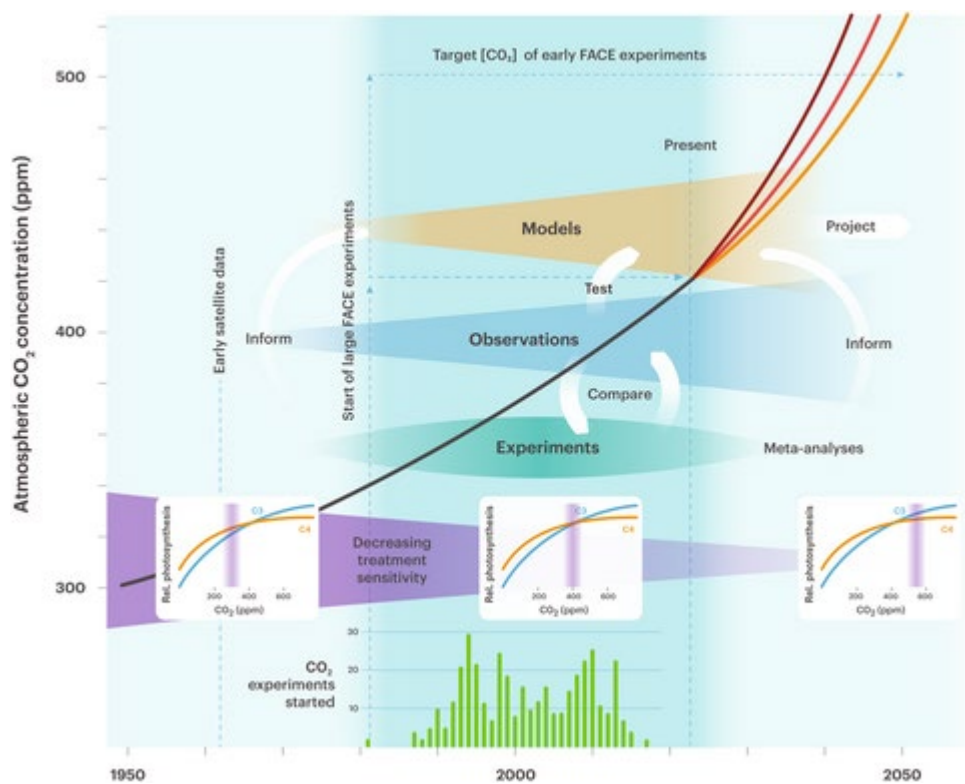
### **Summary**

In 1992, at the start of the senior author's career as a plant ecologist, climate change was perceived as something that would occur in future, and efforts were focussed on predicting what the impacts of those future changes on ecosystems would be, using the tools of manipulative experiments and models. This was still largely the case a decade later, at the start of the more junior author's career in the same field, although observations of impacts resulting from climate change that had already occurred had become more common. More than two decades on, we are deep into the trajectory of climate change, with atmospheric CO<sub>2</sub> having reached 50% above the pre-industrial average, and other drivers significantly beyond their 20th-century baselines. We therefore argue that we must rethink the way we conduct experiments. We will likely move from an experiment- to an observation-dominated era of climate change impact research. This transition could be more explicitly considered and may require adjustments in research policy and funding. We present a practical guide to plan experiments in this new era of researching global change impacts on plants.

### **The baseline is shifting**

The increasingly rapid shift in environmental conditions impacts the way we conduct and interpret research on how natural ecosystems respond to global change. This has been acknowledged in a marine context (Muldrow *et al.*, **2020**; Smith *et al.*, **2025**) and in biodiversity research (Gardner *et al.*, **2009**; Langley *et al.*, **2018**), but not explicitly in the context of global change experiments on plants and terrestrial ecosystems. Here, we specifically ask how the context of global change impact research over the last 40 yr has changed in terms of what we see as 'ambient' and 'future' conditions (Norby & Zak, **2011**). For example, open-top chambers and free-air CO<sub>2</sub> enrichment (FACE) experiments have contrasted ambient with elevated CO<sub>2</sub> concentrations (Norby & Zak, **2011**; de Alencar *et al.*, **2024**); air and soil warming experiments have contrasted ambient with warmed conditions (Melillo *et al.*, **2017**); and nitrogen (N) deposition experiments have contrasted ambient with N addition treatments (Lu *et al.*, **2011**). Such experiments have delivered the critical foundations for our understanding of how global change affects ecosystems, while meta-analysis of the results has supported generalisation across ecosystems and the development of process-based models (Curtis & Wang, **1998**; Song *et al.*, **2019**; Van Sundert *et al.*, **2023**). However, the future is now upon us. For many global change factors, the 'ambient' conditions are rapidly moving further and further from the pre-industrial baseline, and as a result, we need to rethink how we conduct and interpret such experiments in natural systems.

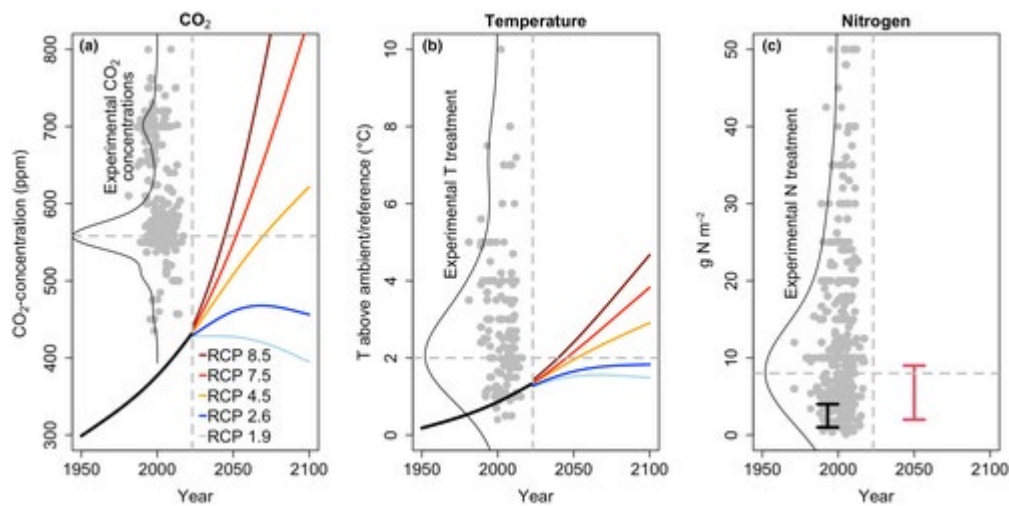
When the first manipulative experiments to study vegetation responses to global change were planned, the atmospheric CO<sub>2</sub> concentration was c. 330 parts per million (ppm) (Mortensen & Moe, **1983**; Kimball *et al.*, **1993**) (Fig. **1**). These early experiments often used double ambient CO<sub>2</sub> concentration as a target for their ‘elevated’ treatment. Across all elevated CO<sub>2</sub> experiments to date, a median treatment concentration of c. 560 ppm has been used (Fig. **2a**). As of 2024, we have passed the threshold of 420 ppm – more than a third of the way into the transition from 330 to 560 ppm (Keeling & Graven, **2021**) (shaded area in Fig. **1**). With high probability (representative concentration pathways (RCP) 7.5 and 8.5; IPCC, **2021**), the planet will cross the threshold of 560 ppm within the next 25 yr – about the time frame of a large FACE experiment from early planning stages to publication of long-term responses (Bader *et al.*, **2013**; Norby *et al.*, **2022**).



**Fig. 1**

Settings for manipulative global change experiments have changed substantially from their onset around the early 1980s to the present. The x-axis shows the timeline from the 1950s to the middle of our century, while the y-axis shows the historic (in black) and projected (RCP 2.6, 4.5, and 7.5 in orange, red, and brown) CO<sub>2</sub> concentrations (IPCC, **2021**). During the early stages of CO<sub>2</sub> experiments, global change was perceived as a future phenomenon, whereas now we are well into its trajectory. As ambient conditions start to approach those that were used in manipulative experiments, we enter a new era, in which observations gain importance. The background increase in atmospheric CO<sub>2</sub> also means that we are shifting to the right in the CO<sub>2</sub>-photosynthesis

response curve (inset graphs, relative (Rel.) net photosynthesis rates of typical C3 and C4 plants plotted against CO<sub>2</sub> concentration, data from Goodman, **2004**), and ecosystems will be increasingly less sensitive to rising CO<sub>2</sub>. Meta-analyses will be critical to comprehensively exploit the value of experiments, and models will increasingly be parameterised and tested against observations, rather than just manipulative experiments. In recent years, fewer CO<sub>2</sub> experiments have been started per calendar year (green bars at the bottom, data from Van Sundert *et al.*, **2023**).



**Fig. 2**

Projected trajectories for the main global change drivers that impact ecosystems. All five representative concentration pathways (RCPs) are plotted (IPCC, **2021**). (a) Historical and projected atmospheric CO<sub>2</sub>-concentrations from 1950 to 2100. The target concentrations and distribution of 414 elevated CO<sub>2</sub> experiment studies are plotted vertically, with a median c. 560 ppm. (b) The equivalent scenarios for global mean surface temperature and target temperatures from 224 warming studies. T, temperature. (c) Nitrogen (N) addition amounts for 731 N addition studies. Depending on the RCP, the mean experimental target concentrations/temperatures will be reached in < 20 yr. Data for experimental CO<sub>2</sub> levels and target temperatures extracted from the MESI data base (Van Sundert *et al.*, **2023**), approximate ranges of N deposition for 1993 (black) and 2050 (red) from Galloway *et al.* (**2004**). For a breakdown of target concentrations/temperatures according to latitude, see fig. 2 in Song *et al.* (**2019**).

Increases in surface temperature are much more heterogeneous spatially and temporally than the rise in atmospheric CO<sub>2</sub>, but the global annual average temperature was estimated to have increased by 1 degree C above pre-industrial levels by 2017, with larger warming over the land surface (IPCC, **2021**). Many ecosystem warming experiments have applied treatments of +2°C, a degree of warming that we could see on average in as little as 20–25 yr (Song *et al.*, **2019**) (Fig. **2b**; RCP 8.5, 7.5, and 4.5), but such conditions may already be experienced in some locations owing to spatial heterogeneity. For example, Drake *et al.* (**2018**) designed a ‘heatwave of the future’

experiment in 2016 around conditions expected for 2100, only to see such a heatwave occurs in the same region in 2019. Similarly, for N addition experiments, many locations are already experiencing deposition levels very close to those simulated in experiments (Fig. 2c), although there is even greater spatial heterogeneity in this driver.

The changing baseline is not a new issue (Gardner *et al.*, 2009; Langley *et al.*, 2018; Smith *et al.*, 2025), but it is becoming more acute as the rate of global change accelerates. It poses three distinct challenges for the interpretation of manipulative global change experiments. First, we may be measuring along a nonlinear response curve. For example, increasing CO<sub>2</sub> has a saturating effect on leaf photosynthesis: Adding 150 ppm to the ambient condition will have a considerably smaller impact on the photosynthetic response of a C3 or C4 plant in a large FACE experiment that commences now, at > 420 ppm, than it had in a similar experiment in 1996, at 360 ppm (Insets at the bottom of Fig. 1). The flow-on effects of photosynthesis to ecosystem function are also likely to saturate (although we note that many feedbacks modify the response at the ecosystem level, including changes in nutrient cycling (Reich *et al.*, 2018) or changes in species composition (Langley & Megonigal, 2010)). A clear understanding of where we sit on the response curve of the global change driver is also essential in the case of warming, precipitation changes, and N inputs. If the response curve is peaked, as is the case for most temperature responses, the response may be either positive or negative depending on whether species are currently below or above optimum conditions. If we are experimenting close to the point of saturation, as was the case in some N addition experiments in which the background levels of N deposition were already very high, we may not see a large ecosystem response (Emmett *et al.*, 1998). This nonlinearity of the response can be addressed by adjusting our hypotheses for the experiment, and by using quantitative, rather than qualitative, hypotheses. Such an approach would involve characterising the expected response curve for that global change driver based on the previous literature or targeted controlled-environment experiments, and then developing a testable estimate for effect size by identifying the distance of both the control and manipulated treatments from a quasistable baseline for that vegetation type. For example, in a warming experiment, we could use glasshouse studies to quantify temperature response curves to identify optimum temperature regimes for a particular species or community. We could then hypothesise positive or negative responses in the field, based on how both ambient and warmed treatments compare to the optimum. Generally, it has been shown that gradient studies are more powerful to capture nonlinear effects in ecology, particularly in the context of global change (Kreyling *et al.*, 2018).

Second, the 'ambient' (control) plots are likely to also be undergoing significant change. For example, in field-based warming experiments, vegetation in control experiments may already be physiologically acclimating to a warmer climate, or experiencing heat stress. The amount of global change experienced in the ambient treatment provides

important context for interpreting experimental outcomes. In an ideal world, we might envisage experiments that include ‘past’ or ‘baseline’ treatments as well as ‘ambient’ and ‘future’ treatments in order to quantify rapid changes in the ambient vegetation, but in many cases, it is difficult to devise such treatments: It is harder to remove CO<sub>2</sub> or N than to add it, and harder to cool ecosystems than to warm them. Instead, we recommend reporting the degree of global change experienced by the ambient treatments in experimental methods, particularly for spatially heterogeneous drivers, and considering the implications for the experimental results. Changes in the ambient treatment may also occur over the course of long-running experiments. For example, over the first 10 yr of fumigation at the Eucalyptus FACE (EucFACE) experiment, ambient CO<sub>2</sub> concentrations increased by 7% – a large enough increase that models applied to the experiment show CO<sub>2</sub>-driven vegetation dynamics in the ambient treatment (Jiang *et al.*, **2024**). In the 24-yr grassland diversity experiment BioCON, the ambient CO<sub>2</sub> controls have experienced an increase of c. +50 ppm CO<sub>2</sub> – nearly a third of the treatment of +180 ppm (Reich *et al.*, **2024**). In such long-running experiments, we may need to consider temporal CO<sub>2</sub> effects on the control as well as the manipulative treatment. Overall, the shifting baseline forces us to rethink experimental data carefully – their interpretation must acknowledge both the change that has already occurred and the change that is occurring during the course of the experiment, in our ambient treatment.

Third, the rapidly changing baseline also has implications for how we carry out meta-analyses of experimental data. While it is relatively straightforward for meta-analysis to be applied to experiments with comparable ambient and treatment conditions, more care and creativity will be required for syntheses of experiments in which the baseline conditions are no longer comparable. Treatment amplitudes have been considered in previous meta-analyses, with effect sizes estimated, for example per ppm of CO<sub>2</sub> increase or per degree of warming (Lin *et al.*, **2010**; Song *et al.*, **2019**). However, we argue that we need to go further than this given the shifting baseline: A given treatment magnitude applied in an experiment in 2020 can lead to a very different response compared to the same experiment conducted in 1980, because we have moved along the (almost always) nonlinear response curve. This is illustrated in the three insets in Fig. 1, using the example of rising atmospheric CO<sub>2</sub> as a global change driver. While in medical meta-analyses, efforts have been made to account for such varying baselines between studies (Abrams *et al.*, **2005**), we are not aware of similar developments in global change impact research. Meta-analyses could potentially account for shifting baselines in several ways. They could look at experiments conducted in different decades separately, or use decade as a covariate. Alternatively, they could include change in the driver variable since a baseline period as a covariate: for example, by using change in mean annual temperature (MAT) since a baseline period in addition to MAT in the baseline period as covariates for a meta-analysis of warming experiments.

Or, in drought experiments, by using both mean annual precipitation (MAP) *and* the difference between MAP and the ambient treatment as a covariate. Meta-analyses could also be based around response curves for the specific drivers, where available – for example, by normalising the response ratio in CO<sub>2</sub> experiments by the expected logarithmic response (Walker *et al.*, **2021**).

### **Earth is turning into a large-scale experiment**

As we move further into the trajectory of global change, the methods used to study its effects on ecosystems must necessarily adapt. Increasingly, the emphasis is shifting from experiments that mimic future change, to observational studies of the change that is already occurring, made possible by long-term records that document change. For example, we now have over 40 yr of high-resolution satellite data available (Piao *et al.*, **2019**), spanning an atmospheric CO<sub>2</sub> increase of c. 80 ppm and a global temperature increase of > 0.5°C, while the longest standing eddy covariance flux tower observations cover over 30 yr (Bautista *et al.*, **2021**). With the increasing availability of these data, there has been a notable increase in the number of observational studies of the effects of increasing atmospheric CO<sub>2</sub> (Walker *et al.*, **2021**) and a simultaneous decrease in the number of CO<sub>2</sub> experiments started in recent years, with the last spike c. 2010 (Fig. **1**, green bars). Such observational studies offer important opportunities to test our understanding of global change impacts at much larger scales than would ever be possible with experiments, and can identify critical processes that have been overlooked. A good illustration is the emerging global observational evidence demonstrating increasing tree mortality in droughts coupled with heatwaves (Allen *et al.*, **2015**). This phenomenon had been largely underestimated in models previously because it is relatively intractable to study in manipulative experiments; the observational data have clearly highlighted this underestimation and spurred extensive new scientific investigation to rectify it (McDowell *et al.*, **2022**). Kröel-Dulay *et al.* (**2022**) also demonstrate that observations are more robust than manipulative experiments, likely because of the ‘island effect’ that experiments create (Leuzinger *et al.*, **2015**).

Some opportunities for observational studies may arise serendipitously: For example, large data bases that have been created without focussing on global change impacts can now be mined for exactly that. The large plant trait data base TRY (Kattge *et al.*, **2011**) or the ‘Teacomposition’ initiative for studying litter decomposition using simple tea bags (Keuskamp *et al.*, **2013**) can for instance be examined for evidence of changing plant or ecosystem traits due to global environmental change. Global meta-analysis initiatives on multifactor experiments will become informative even by just looking at responses of control groups over time (Van Sundert *et al.*, **2023**), and new kinds of meta-analyses that combine experiments with observational studies become possible.

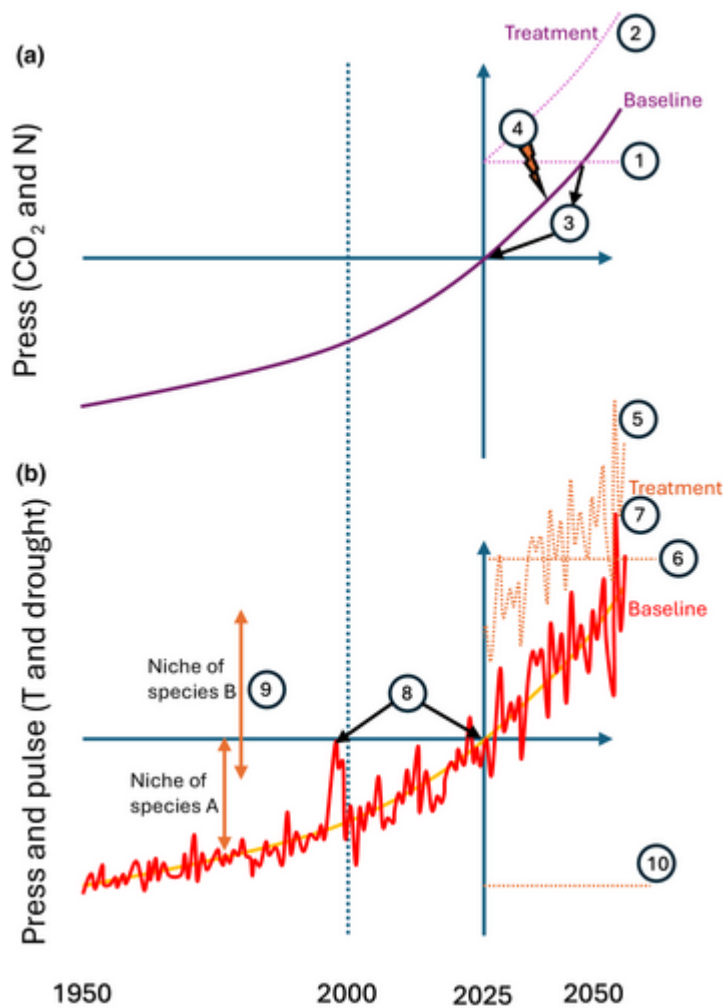
However, a shift to studies focussed on longitudinal trends poses its own challenges. First, we note the need for extreme care in interpreting changes in data that have been collected nonsystematically over time; changes in measurement methods or sampling coverage can obscure underlying trends (Bond-Lamberty & Thomson, **2010**). There is an urgent need for ongoing monitoring that is specifically designed to detect and examine long-term ecosystem changes in response to ongoing global change. Ideally, such monitoring would build on existing datasets, taking advantage of data collected before significant climate change had occurred. For example, long-term forest monitoring plots and long-term ecological research sites should be maintained and remeasured as a high priority, as they can now effectively be used as global change experiments (Anderson-Teixeira *et al.*, **2015**).

Observational studies are also much less controlled than manipulative experiments. Ecosystems are being affected by many global change factors simultaneously, including changes in land use and land management, with the potential for very different interpretations of the processes involved (Williams *et al.*, **2025**). For example, debate has raged over the relative roles of climate warming, forest harvesting, and fire suppression, in driving observed increases in fire frequency (Bowman *et al.*, **2022**), while studies probing the drivers of the terrestrial C sink have disagreed significantly about the relative importance of increasing atmospheric CO<sub>2</sub> vs changes in land use (Zhu *et al.*, **2016**; Walker *et al.*, **2021**). Understanding how simultaneously changing drivers combine to produce the observed dynamics is a major scientific challenge and requires that we combine observations with manipulative experiments and models in novel ways, which we outline below.

### **New ways forward for combining experiments, observations, and models**

Entering this new era, there is less justification for manipulative experiments that simply aim to ‘simulate the future’ as has often been stated; instead, experiments must be designed around more targeted goals, including understanding and disentangling processes (Beier *et al.*, **2012**; De Boeck *et al.*, **2015**), interpreting and attributing observational data (Anderegg *et al.*, **2015**), and supporting forecasts of future ecosystem states (Medlyn *et al.*, **2015**). Classic global change experiments may be complemented by more controlled experiments in microcosms or ‘ecotrons’ (Roy *et al.*, **2021**), in which response surfaces to multiple global change drivers can be characterised in a more mechanistically focussed way, allowing us to situate the treatment levels of a large-scale experiment within the trajectory of global change (Orr *et al.*, **2020**). One way to gain such mechanistic insights is to use multiple treatment levels with (sometimes unrealistic) factor combinations that allow us to create response surfaces (Collins *et al.*, **2022**), or playing with treatment levels that are either fixed or keep pace with the baseline shift (see (1), (2), (5), and (6) in Fig. **3a** illustrating press, and 3B illustrating a press/pulse global change driver). For

example, the interpretation of field-based warming experiments is challenging when plants under ‘ambient’ conditions may already be experiencing heat stress during the experimental period (see (7) in Fig. 3b). Field experiments can be complemented by controlled glasshouse experiments to quantify plant responses to a range of temperatures, both above and below current ambient temperatures ((10) in Fig. 3b), including controlled interactions with water availability, to develop mechanistically based expectations of plant performance that can be tested in the field.



**Fig. 3**

Considerations during the planning phase of future large manipulation experiments. The context in which such experiments are planned has changed substantially since before the 2000s, when the baseline shift was much less apparent (horizontal dashed line). (a) shows global change drivers that exert a press change on ecosystems (e.g. CO<sub>2</sub> and N). Traditionally, a certain future scenario is chosen as the (constant) treatment level (1). With a rapidly changing baseline, this would create a diminishing treatment effect. In an extreme case, the baseline could intersect with the treatment level. An alternative to the constant treatment is a treatment level that keeps pace with the baseline shift (2) to the effect of a constant treatment – control difference. The choice between (1) and (2)

largely depends on the projected scenarios. During a long-term experiment, an opportunity arises to make comparisons within the control plots over time (3): the conditions towards the end of an experiment will have changed measurably and might be large enough that they can be detected and quantified. (4) Extreme events of other global change factors can occur over the course of an experiment and provide a unique opportunity to study interactions. (b) shows global change drivers that impose both a press and a pulse change (e.g. temperature (T) or drought). Usually, treatments are layered on top of an ambient scenario (5). If the treatment is constant (6), or unable to be stacked on top of the ambient conditions, the control scenario can overshoot the treatment (7, Drake *et al.*, **2018**), causing the experimental treatment to coincide with the control. With pulse drivers, a way forward could be to use historic extremes (e.g. via dendrochronology) that may match the current or future baseline shift (8) – an efficient way to characterise ecosystem responses without the use of experiments. Taking into account a species' niche (9) in terms of its tolerance to a global change driver is essential, as the baseline shift during an experiment may span its entire space. The application of treatment levels that represent historic levels (10) is confined to glasshouse experiments.

Manipulative experiments in a changing climate can also yield a unique experimental setting that allows interactions among factors to be examined, with the control (ambient) plots and treatment (target) plots both subject to other global-change-driven disturbances ((4) in Fig. **3a**). For example, at the EucFACE experiment, ambient and elevated CO<sub>2</sub> plots have both experienced extremes of heat and rainfall over the course of the experiment, enabling the interaction of these extremes with atmospheric CO<sub>2</sub> concentrations to be studied (Griebel *et al.*, **2023**). Another fascinating example is a saltmarsh experiment representing the longest-running CO<sub>2</sub> enrichment experiment to date (33 yr), which witnessed a background increase in atmospheric CO<sub>2</sub> from 348 to 410 ppm, and a sea-level rise of 23 cm (C. Zhu *et al.*, **2022**). In an extreme case, the ambient baseline over the course of an experiment might reach or get into the vicinity of the initially chosen treatment level, allowing for interesting within-experiment comparisons ((3) in Fig. **3a**).

A key challenge of the observational era is the attribution of trends in observational studies to causal factors, particularly where there are competing explanations for trends. Attribution requires a quantitative model framework in which the expected effects of individual factors, and their interactions, can be compared against the magnitude of the observed changes (Kogan, **1997**; Grace *et al.*, **2002**; Ruehr *et al.*, **2023**; Williams *et al.*, **2025**). This overall approach requires a strong integration of classic global change field experiments with observations and models. Observations can be used to help design ecosystem manipulations that will provide insight into mechanisms and support model development. For example, observations of increasing tree mortality rates over time could be a result of several alternative mechanisms (Reyer *et al.*, **2013**;

Allen *et al.*, **2015**). Experiments may be designed to apply extreme treatments that attempt to generate mortality in order to understand the processes at play, such as testing different water stress thresholds (Beier *et al.*, **2012**; De Boeck *et al.*, **2015**). Mechanistic models parameterised with these data can be applied to long-term observational studies to attribute observed changes to underlying causes. Similarly, as longer time series become available, it is now possible to track the recent terrestrial carbon sink in great detail and over periods of simultaneous large-scale changes in land use and steeply increasing atmospheric CO<sub>2</sub> (Xu *et al.*, **2021**). Our understanding of the causal factors behind the carbon sink hinges on the application of mechanistic models incorporating these potential drivers to test which drivers can explain observed patterns (e.g. Sitch *et al.*, **2024**).

A new opportunity arises if we are able to use the climatic history of a site to compare past responses to extreme temperature events that now make up the mean ambient conditions, for example using dendrochronological methods ((8) in Fig. **3**). However, we have to be mindful of the fact that if we are moving along a nonlinear response curve, the comparison may be biased. If it is a natural ecosystem that we are warming, we need to appreciate that the plants may already be stressed with the rise in T that has already occurred. In that case, it is useful to take into consideration where the species are currently sitting with respect to their thermal niche ((9) Fig. **3**).

Finally, we get an opportunity to review and test modelling work from the late 20th century. Most of these model projections worked on a 100-yr horizon – typically until 2100. As we complete the first quarter of this time period, we have a full 25 yr to validate modelling results with observations (space between the solid and dashed line in Fig. **3**). For instance, Cramer *et al.* (**2001**) published a model comparison projecting key environmental variables (Net Ecosystem Productivity (NEP), total terrestrial biomass) until the end of the 21st century. These projections can now be compared to robust long-term observations, and model performances can be analysed and evaluated in an unprecedented way (Fig. **1**). Running historic and state-of-the-art models in retrospect to predict the past 25 yr for which we now have observational data allows us to pinpoint the key processes that improved our models. We argue that the potential of such model – observation evaluations – has not fully been exploited by the vegetation modelling community, particularly in the case of near-term forecasting (Dietze *et al.*, **2018**).

We conclude that the transition from a pre- to within-global change era challenges the way we conduct manipulative experiments. Observational and monitoring studies are becoming increasingly important and are often more realistic than experimentally imposed treatments, and their integration with experimental studies and models offers unprecedented explanatory power. While funding decisions ultimately need to be in alignment with the needs of local stakeholders and national priorities, we propose a focus on (1) the analysis of long observational records that make use of the planet's

trajectory from pre- to within-global change conditions, (2) experiments that are increasingly designed to specifically isolate interactive and mechanistic process understanding rather than ‘simulating’ global change, and (3) the use of observations to ground-truth and test both models and experiments that have been developed during the early global change era. This new dynamic between models, experiments, and observations will help us understand global change responses to tackle the global sustainability goals as we move towards the middle of the 21<sup>st</sup> century.

### **Acknowledgements**

The idea for this perspective was born during the 6th iLEAPS (integrated Land Ecosystem Atmosphere Processes Study, a Future Earth Global Research Network) conference in Auckland, New Zealand. Artwork partly by Phil Dunstan-Brown, Ghostdynamics Ltd. We thank three anonymous reviewers for their valuable comments.