frontiers | Frontiers in Climate

TYPE Original Research
PUBLISHED 08 January 2024
DOI 10.3389/fclim.2023.1283333

Check for updates

OPEN ACCESS

EDITED BY
Anders Hansson,
Linköping University, Sweden

REVIEWED BY
Daniel Andersson,
Linköping University, Sweden
Ryan O'Loughlin,
Queens College (CUNY), United States

*CORRESPONDENCE
Joseph Wilson
✉ joseph.wilson-1@colorado.edu

RECEIVED 25 August 2023
ACCEPTED 15 December 2023
PUBLISHED 08 January 2024

CITATION
Wilson J (2024) Modeling the measurement of carbon dioxide removal: perspectives from the philosophy of measurement.
Front. Clim. 5:1283333.
doi: 10.3389/fclim.2023.1283333

COPYRIGHT
© 2024 Wilson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Modeling the measurement of carbon dioxide removal: perspectives from the philosophy of measurement

Joseph Wilson*

Department of Philosophy, University of Colorado Boulder, Boulder, CO, United States

This paper explores how recent developments in the philosophy of measurement can frame and guide the way we measure successful carbon sequestration in carbon dioxide removal (CDR) projects. Recent efforts to mitigate carbon emissions, e.g., the forest offset program implemented in California, have been revealed to systematically over-credit projects relative to the benefits they produce for the climate. In this paper I utilize concepts from the philosophy of measurement, primarily those surrounding models of the measurement process, to diagnose this problem of over-crediting in the broader context of concerns about uncertainty and impermanence in CDR. In light of these measurement models, I argue for absolute measurement targets in favor of the standard comparative targets, the latter of which are significantly dependent on tenuous baseline projections. I go on to consider which contemporary approaches to CDR are successful in light of lingering uncertainty about the future, which puts particular emphasis on the permanence of carbon sequestration. Independent of the specific argument developed here, the paper also serves to introduce concepts from the philosophy of science and measurement to a broader audience, in the hopes they will benefit other areas of research.

KEYWORDS

carbon dioxide removal, carbon offsets, philosophy, measurement, climate, models

1 Introduction

Climate mitigation efforts have increasingly come to focus on carbon dioxide removal (CDR). The promise of CDR is that through the development of technologies and corresponding social incentives, national economies can smoothly transition toward an overall decrease in the quantity of carbon being released into the climate system. Considered technologies include afforestation, improved forest management, carbon dioxide mineralization, direct air capture, and many more (see Hovorka et al., 2021 and Pilorgé et al., 2021 for a general discussion of these technologies).

However, there are outstanding issues in the measurement and modeling of real-world outcomes of CDR technologies (Chay et al., 2022) arising from uncertainty in the application of these technologies and their underlying theory. The former, *execution uncertainty*, emerges when the operation of a project deviates from expectations, or when errors occur in the calculation or reporting of outcomes. As such, execution uncertainty is primarily error in implementation, "mitigated through careful deployment of existing tools and practice," (Chay et al., 2022). The latter, *scientific uncertainty*, emerges from an inadequate understanding of the relevant natural systems or processes. How would Atlantic Ocean circulation, for example, change in response to an influx of 250 billion tons of meltwater per year? Scientific uncertainty of this sort calls for research efforts directed toward novel analytic methods and improved understanding of the relevant systems.

There is also a third kind of uncertainty for CDR technologies (Chay et al., 2022), *counterfactual uncertainty*. We can only understand the benefits of CDR in the context of counterfactual scenarios where the technology was not implemented: *what would have happened to the carbon in a particular instance of CDR if it had not been removed?* Such counterfactuals play an indispensable role in establishing claims about baselines and additionality: what is the counterfactual *baseline* quantity of carbon sequestered without CDR and what *additional* carbon is sequestered as a result of the implementing the technology? However, there are deep challenges in the determination of baseline and additionality, challenges that require more than the kind of engineering and empirical solutions needed to address execution and scientific uncertainty. Indeed, some have argued that methods for estimating baseline and additionality values are inherently subjective (e.g., Gifford, 2020 in the context of forest carbon offsets). On top of this, even ignoring the potential subjectivity of these counterfactuals, errors due to over-crediting arise when credits do not correspond to real additionality (Badgley et al., 2022a).

Assessing the relative merit of a particular CDR method will require consideration of all three forms of uncertainty: execution, scientific, and counterfactual. Each form of uncertainty introduces distinct challenges for predicting how a project will perform in the future. While the ideal method would exhibit low degrees of uncertainty across the board, it will more often be the case that there are trade-offs between different kinds of uncertainty. As such, funding and policy decisions will ultimately need to be made on the basis of which uncertainties are more tolerable than others.

In this paper I approach the problem of uncertainty in CDR technologies from the perspective of the philosophy of measurement, with particular attention to issues of counterfactual uncertainty. My aim is to apply theoretical insights from the philosophy of measurement (e.g., Mari et al., 2017; Tal, 2017; Wilson and Boudinot, 2022) to address fundamental questions about the application and incentivizing of CDR technologies, mitigating some of the aforementioned uncertainties in carbon-crediting while suggesting a new way forward for conceptualizing CDR technologies. I argue that specific targets for CDR are more sensitive to counterfactual uncertainty than others. Indeed, the aforementioned subjectivity of additionality and baseline is partly a product of the specified target for carbon offsets being appropriately comparative, i.e., determined by comparison with a counterfactual baseline indicating what would have happened in the absence of the project. Alternative targets for CDR may thus be capable of mitigating some of this subjectivity. Ultimately, the paper uses ideas in the philosophy of science to frame and guide improvements for the quantification of carbon in environmental policy with the further aim of introducing those ideas that are useful to the broader group of experts interested in the climate. As such, I intend this paper to serve as an open invitation for further discussion and communication about the philosophy of carbon measurement, rather than as a definitive or conclusive proposal. I hope the arguments in this paper can get the ball rolling.

In Section 2 of the paper, I investigate the problem of over-crediting in California's forest carbon offset programs (Badgley et al., 2022a), in which standards failed to promote real climate benefits. A program only produces real climate benefits when its implementation results in a genuine reduction or removal of carbon in the atmosphere. I use California's forest offset program as a case study for raising a number of empirical and philosophical problems with the comparative approach to carbon measurement. In Section 3, I discuss some ideas from the philosophy of science, focusing on the development of models of the measurement process. These models facilitate measurement by appropriately representing the features of the world that make a difference to the measurement target and the apparatus used to get that measurement. Many shortcomings in carbon offset programs can be understood as a failure to capture the right difference makers. In Section 4, I develop an alternative framework for the valuation of carbon inspired by measurement models discussed in the philosophy of measurement. I argue that real climate benefits, the target of financial incentives, are better understood in terms of absolute carbon over comparative carbon. In Section 5 I evaluate CDR technologies in terms of their permanence, highlighting which methods are less vulnerable to remaining uncertainty about the future.

2 Diagnosing over-crediting in carbon offsets

Sections 2–4 of this paper are dedicated to identifying an appropriate measurement target for assessing CDR technologies. In short, what is it about the world that we are intending to learn when we conduct our carbon measurement, and does it correspond to our intended goals? In this Section I will consider California's forest carbon offset program as a case study for contemporary carbon markets. The program comprises the largest compliance market in operation, thus making it a useful case for drawing out some of the problems that arise when taking a *comparative approach*. Among other things, programs utilize a comparative approach when awarding credits or financial benefits on the basis of comparison with a counterfactual baseline or projection that indicates what would have happened in the absence of the program. First, I will discuss how California's offset program is vulnerable to a problem with over-crediting as a result of targeting inappropriate metrics for carbon sequestration (Section 2.1). That is, relying on standards derived from cross-species and cross-regional averages has led to systematic over-crediting in California's forest carbon offset program (Badgley et al., 2022a). This empirical problem arises because the utilized standards fail to capture the appropriate measurement target: the carbon target must result in real benefits to the climate. On top of this, I suggest several additional problems that limit the ability of California's offset program to produce real climate benefits, specifically problems resulting from methodological subjectivity, upfront crediting, and uncertainty about the future (Section 2.2). While I do focus on California's forest carbon offset program, the discussed problems for the comparative approach should generalize to programs that share in these features. I argue that some are endemic to the comparative approach. In a later section (Section 4) I develop an alternative approach to mitigate these problems, an approach partly inspired by work being done in the philosophy of measurement (Section 3).

2.1 California's forest carbon offsets and real climate benefits

Carbon offset programs are intended to distribute credits to projects that reduce the amount of carbon in the atmosphere. They achieve this by either reducing emissions or removing carbon that is already there. Importantly, these carbon credits can regularly be used by polluters to emit more carbon than would otherwise be legally permitted. The owner of some wetlands may agree to preserve the land over the next century in favor of developing it, generating carbon credit. A coal refinery may then purchase that carbon credit so that it can expand emissions beyond legal limits in accordance with what is offset by the credit. If carbon offset programs allow polluters to generate more carbon emissions, then it is imperative that carbon credits correspond to genuine differences in carbon emissions. Credits ought only to be issued when a project produces real benefits to the climate, located with some real carbon in the world. The *real climate benefits* of a program are thus the quantity of net total carbon that is removed or reduced as a result of the program's implementation. We will see that there are a number of ways to make the notion of real climate benefits more precise. These real climate benefits are typically contrasted with intentional manipulation or statistical artifacts that generate a discrepancy between "real" benefits and the purported benefits indicated by a particular metric.

Indeed, there is evidence that the largest compliance market in active operation, California's forest offset program, systematically over-credits the carbon reduced by improved forest management projects (Badgley et al., 2022a). Credits are awarded to projects whose projected carbon stock doesn't fall below *common practice*. Common practice is a regionally specific baseline developed using the US Forest Service Forest Inventory and Analysis (FIA) database. The higher this projected carbon average (over common practice), and the lower the project's initial carbon stock, the more credits are earned. Common practice estimates are determined by applying the FIA data to specific geographic regions (*supersections*), which are subdivided into smaller regions represented by the dominant tree species (*assessment areas*). Within the Northern California Coast supersection, for example, all parcels are assigned to either the Oak Woodland assessment area or the Redwood/Douglas Fir Mixed Conifer assessment area based on which tree species are prominent. Estimated carbon stocks are determined by what the FIA data suggests about the carbon properties of these species.

However, in fixing each parcel to only one specific assessment area (*either* Redwood *or* Oak), species heterogeneity within a region is ignored. Badgley et al. (2022a) use an alternative assessment of common practice based on project specific reporting of local species to provide a more accurate representation of species diversity. With this alternative assessment, the authors discover higher baseline carbon estimates for the majority of the IFM projects involved. Single species carbon estimates were systematically lower than what we would expect in the diverse forests found in the real-world. In total, Badgley et al. (2022a), suggest that over thirty percent of upfront credits awarded by California's forest offset program are not grounded in real climate benefits, reflecting a statistical artifact of the chosen methodology for measuring carbon quantities. A significant portion of the credits do not correspond with a real quantity of sequestered carbon in the world. Local conditions differ from regional averages, and so accurately calculating how much carbon a project sequesters above baseline requires (among other things) "a more granular analysis of average carbon stocks across species and geographies" (Badgley et al., 2022a, p. 1443).

This analysis of systematic error in California's forest offset program highlights some basic ideas surrounding the more general project of carbon dioxide removal (CDR). CDR projects aim to produce real climate benefits, articulated in terms of reducing the amount of carbon that ends up in the atmosphere. One way in which these CDR projects can be prone to error is when their measures for success come apart from real climate benefits. In the case of forest offsets, the carbon estimates for a region's trees can come apart from the real carbon in those trees. Species designation can misrepresent the real species distribution of the region. In the following section (Section 3) I will consider more specifically how we should understand this "coming apart" from the perspective of the philosophy of measurement. For the remainder of this section, I investigate how California's forest offset program represents real climate benefits, and some further problems for the program.

2.2 The comparative approach and further challenges

California's forest carbon offset program provides a *comparative* quantification of real climate benefits. Credits are awarded to projects insofar as the sequestered carbon is greater than some designated common practice baseline. In ideal cases this baseline provides an empirically supported approximation for how much carbon would be sequestered in a counterfactual scenario in which there was no significant intervention on the land for the designated period of time. A real benefit to the climate is quantified as the (positive) difference between this status quo "do-nothing" baseline and the project.

This comparative approach falls in line with comparative accounts of harm in the philosophical literature. Comparative accounts of harm (Feinberg, 1984; Parfit, 1984) claim that an event harms someone if and only if the event makes her worse off than she otherwise would have been. I am harmed by a poisoned apple because eating the apple makes me worse off than I would have been had I not eaten the apple. Conversely, comparative accounts of benefit claim that an event benefits someone if and only if the event makes her better off than she otherwise would have been. I benefit by eating a non-poisoned apple because it makes me better off than I would have been had I not eaten the apple (assuming that had I not eaten the apple, I would have eaten nothing instead). One way of cashing out climate benefits, then, is the comparative approach suggested above: a project benefits our climate if the climate is better than it would have been without the project. Put another way, if a project puts us in a better position than the baseline condition, then it produces real climate benefits.

I suspect that something like the comparative approach is true for characterizing whether a project is genuinely beneficial or harmful. However, it is not immediately clear which of our chosen potential projects will achieve this comparative benefit. Furthermore, there are significant challenges in determining the baseline condition. As such, there are concerns with the comparative approach if it is to provide prescriptive guidance for policy and action. We've already seen with California's offset program how empirically imprecise methodologies can lead to systematic error in calculating baseline scenarios. Similarly, in contexts where multiple accounting protocols are permitted developers can earn unwarranted credits by selecting the most financially favorable method (Gifford, 2020). Project developers are in many cases told to "choose an accounting protocol that addresses a desired outcome" (Gifford, 2020, p. 296), introducing a kind of subjectivity into the measurement task that encourages systematic error. Developers are free to pursue specific metrics merely on the basis that they output the highest quantities of carbon. While concerns regarding subjectivity are not particular to the comparative approach, insofar as such subjectivity is permitted in the calculation of baseline carbon, we should expect baseline determinations to diverge from real climate benefits.

Nearby philosophical concerns also arise when we understand the baseline to represent the "status quo," as *what would have happened otherwise*. In particular, it is unclear whether we are positioned to reliably identify baseline conditions for what would have happened otherwise in the near future. Furthermore, it is unclear whether there is a singular baseline condition with which to compare future project estimates. Since the aim of California's forest carbon offset program is to reduce greenhouse gas emissions on timescales of decades and centuries, an estimate of what would have happened in the absence of the program (of how some particular forests would fare) will require some assumptions about how the earth's climate will progress over the next several decades. However, there are a number of different scenarios that the Intergovernmental Panel on Climate Change (IPCC) consider to be possible given our current circumstances. The IPCC utilizes several emissions scenarios in their projections of the future representative concentration pathways (RCPs) ranging from optimistic (limiting warming to 1.5C) to disastrous (exceeding 4C) warming by 2,100 (IPCC, 2023). It is unclear which specific RCP scenario we should understand ourselves to be currently tracking, and so it is unclear under which scenario we should interpret and project baseline conditions.

Even worse, once we decide which scenario best represents our current trajectory, there is still the issue of robust model disagreement for features of the climate that will influence carbon sequestration (e.g., mid-latitude precipitation change discussed in Zappa et al., 2021). If our goal in determining the baseline is to identify status quo carbon projections, we are not epistemically situated to determine a reliable baseline. Moreso, if we take the IPCC RPC pathways to illustrate real possibilities, trajectories that are still possible for us to achieve if we take the corresponding actions, then there is in fact no singular baseline.

Crediting projects upfront for projected baseline quantities generates a more pragmatic concern regarding uncertainty in the permanence of that carbon. While uncertainties about the future can induce counterfactual uncertainty of the sort just mentioned, they also induce a factual uncertainty about the future of the project that is closely tied to execution uncertainty. Even if no fault lies with project managers, there are systematic factors beyond the manager's control that threaten the successful performance of CDR projects. Baseline comparisons are made on the presumption that carbon will be successfully sequestered for the duration of the program. It is presumed that projects will store carbon for the entire century. However, in the case of forest carbon sequestration this ignores the relevant possibility that carbon is released as a result of forest destruction via wildfire, pests, and drought. In order to accommodate this expected loss of carbon, California's forest carbon offset program creates a buffer pool of additional carbon. So long as the carbon lost does not exceed the carbon in the buffer pool, the program will result in net positive carbon storage (some amount of real climate benefit). This has the effect of diluting how much actual carbon there is per credit, but the more concerning problem is that estimated losses are soon expected to deplete the buffer entirely. Estimated wildfire losses in the next decade would consume ninety-five percent of what has been set aside for wildfires throughout the next century (Badgley et al., 2022b). It is thus incredibly likely that significant quantities of credited carbon will make its way back into the atmosphere.

It is common (mandatory in compliance markets) for managers of forest carbon projects to purchase insurance for the loss of carbon that occurs during such events. However, this insurance only serves to remediate financial loss, doing nothing to resolve the disparity between carbon credits and real climate benefits. That is, insurance permits a project to continue claiming carbon credits, even while the designated carbon roams freely in the atmosphere (Macintosh, 2013; D'Alisa and Kallis, 2016; Gifford, 2020). Insofar as projects are credited upfront in accordance with baseline estimates and are permitted to keep those credits in the case that the sequestered carbon is lost (in conjunction with some minor cost), carbon estimates and their associated credits come apart from real climate benefits.

In short, empirical imprecision, subjective methodology, and factual uncertainty about the future are all ways that a CDR project can fail to generate real climate benefits. On top of this, any baseline-driven comparative approach will run into challenges determining a reliable counterfactual baseline. While I have focused on a specific implementation of forest carbon sequestration, we should expect the lessons to generalize for any carbon removal techniques (e.g., enhanced weathering, direct air capture, ocean alkalinity enhancement) that exhibit these features. Problems determining counterfactual baselines are necessarily bound up with uncertainties about the future climate, making it difficult or impossible to answer what we should expect to happen to a natural system in the absence of any project. Given the pervasiveness of these uncertainties, I argue that it is more beneficial to consider which CDR methods are capable of sequestering carbon across a wide range of environmental circumstances. Some CDR methods are more insulated from the influences of the surrounding environmental changes, providing a more permanent method of carbon sequestration. I will return later (Section 5) to discuss respective permanence.

In light of these problems, it is incumbent on the environmentalist to seek solutions and alternatives. In the following two sections I will look to the philosophy of measurement for a theoretical framework that characterizes the over-crediting problem in California's forest carbon offset program (Section 3) and guides the development of an alternative approach to real climate benefits (Section 4).

3 Models of the measurement process

In this section I detail some of the work being done in the philosophy of science, with particular attention to the broader role of measurement models (Section 3.1). Importantly, models of the measurement process require a target phenomenon and an understanding of what systemic features influence variation in that target, i.e., *difference makers*. If we hope to reliably measure quantities of carbon, for example, then we will need to understand which natural processes and properties correspond to differences in carbon quantity. Models of the measurement process thus provide a useful theoretical framework for articulating the epistemic ideals of measurement, how workers work toward those ideals, and how to identify and resolve measurement problems. Within this framework I characterize the current problem of over-crediting in carbon offset programs (Section 3.2). While the empirical problem has been conceived in terms of being a problem of averages, I argue that it is more accurately a problem about standardizing the *wrong* average, averages that fail to capture the appropriate difference makers. However, since measurement models are necessarily grounded on theoretically sound dependencies, capturing the influence of difference makers, it is true that certain counterfactual claims are indispensable for any form of measurement. I will argue that these counterfactuals are not subject to the aforementioned *subjectivity* (Gifford, 2020) or *counterfactual uncertainty* (Chay et al., 2022).

3.1 The philosophy of science and measurement

Philosophers of science have shown a renewed interest in the investigation of measurement throughout the last few decades.[1] Expanding on earlier advances in the philosophy of scientific models (e.g., Cartwright, 1983; Giere, 1988; Giere et al., 2006; Godfrey-Smith, 2006), recent philosophical work has focused more specifically on the theoretical machinery required to ground scientific measurement. This work has led to a more careful understanding of what constitutes measurement and how it can be reliably achieved. What is it about a mercury thermometer, for example, that enables a user to reliably measure the local temperature? Different philosophers disagree (of course) on precisely how to understand the measurement process, though there seems to be increased attention to the need for *models of the measurement process.*

Models of the measurement process provide a rich description of the system in which our desired measurement target is present. As such, the models capture the components of the system and their interactions that are relevant to the measurement task. If our aim is to measure the ambient temperature of a room with our classic mercury thermometer, for example, then a model of the measurement process will represent the relevant features of the column of mercury, its material container, the air in the room, and their important dynamic relationships. By capturing these features and their interactions, the model will generate specific values for the ambient temperature of the room, given a specific height of the mercury column (as well as specific values for the height of the mercury column, given a particular ambient temperature). In this way, the model of the measurement process produces a framework for understanding how targeted interventions on specific variables would influence the system (a la Woodward, 2003). A model for a specific thermometer could imply, for example, that if the ambient temperature rises 3 degrees Celsius, then the height of the mercury column should grow 3 millimeters.

It is the robustness of this model of the measurement process that allows us to reliably use a thermometer. A thermometer works for us because we understand, across a variety of environmental conditions, the regular and robust dependency between the height of mercury and ambient temperature. It is in this context that Tal (2017) argues that the target outcome of a measurement will be the best predictor of the instrument indication given the model of the measurement process. As such, Tal grounds the reliability and objectivity of measurement in robust prediction. The mercury in a thermometer can be understood to measure ambient temperature because, given the model of the measurement process, ambient temperature is the best predictor for changes in the height of mercury. Among other things, the model of the measurement process captures how ambient temperature makes a difference to the height of mercury, and vice versa.

Mari et al. (2017) emphasize the role of background theory in the proper construction of a model of the measurement process. The first step, once the measurement task is identified, is to produce a general model constructed using the general laws that pertain to the general properties of the target system. For our thermometer, this means that thermodynamic laws pertaining to temperature, molecular motion, conductivity, and thermal expansion will be incorporated into a general model. It is thus important for Mari et al., that measurement be grounded, first, in established scientific laws. From here, we go on to specify the general model for the kind of object to be measured. This is where any necessary idealizations and approximations are introduced. A specific model for measuring the temperature of my living room, for example, will likely need to presume a homogenous temperature throughout the room even if there are in fact slight temperature variations throughout. Next, a model of the measuring system is constructed to include the instruments and techniques needed to identify the target property. The mercury thermometer and my living room are modeled (in accordance with prior general and specified models) to permit the calculation of my living room's temperature from the height of mercury in the thermometer.

For Mari et al. (2017), arriving at a model of the measurement process thus requires that workers first implement their

---

[1] Much of the credit for this resurgence in the philosophy of measurement belongs to Chang (2004) for his excellent and thoughtful investigation into the history of measuring temperature.

general physical understanding of the target, then introduce approximations to accommodate the specific features of the target, and finally integrate models of the measurement apparatus with models of the target. Reliable and objective measurement is the product of a model-building process that is grounded on general scientific laws and requisite idealizations. We should trust the dynamic relationship in the model, which represents how features of the system make a difference to other features of the system, because the relationship is derived from independently supported background theory.

For an instrument to provide a reliable measure, however, the influence of confounding factors must also be included in the measurement model. For the mercury in a thermometer to indicate temperature, and only temperature, the mercury must be held at a constant pressure. Variation in temperature *and* pressure both influence the volume of a fluid. Boudinot and Wilson (2020) and Wilson and Boudinot (2022) argue that standard measurements like the thermometer achieve this through *physical* control, while proxy measurements like tree rings or oxygen isotopes require *post-hoc* analysis, or *vicarious* control, to account for the influence of confounds. Whether standard or proxy, a reliable and robust model of the measurement process must represent those features of the real-world system that make a difference to the output, especially those that are distinct from our measurement target.

In addition to modeling and controlling confounds, we may also consider our measurements to be reliable when multiple independent methods and techniques converge on similar results. Woodward (2003) calls this convergence *measurement robustness*. Insofar as different measurement devices are constructed employing different theoretical principles and different methodological assumptions, it is unlikely that the devices will fall victim to the same kinds of error. Because the errors are expected to be independent, there is unlikely to be something fundamentally wrong with the measurement results when agreement is achieved across distinct devices. Instrumental agreement in such cases would require an implausible convergence of independent errors. We should thus have increased confidence in our temperature measurement if our mercury thermometer agrees with a thermistor, constructed in accordance with electrical principles, since their agreement would require an unexpected agreement of independent error across the devices. In this way multiple models of the measurement process, models of distinct apparatus, can work together to improve the overall reliability of our measurements.

Taken together, these ideas from the philosophy of measurement help provide a framework for understanding reliable scientific measurement. Models of the measurement process provide a sufficiently detailed description of the target system, facilitating a prediction of the desired target using some indicator (e.g., temperature from a mercury column). The objectivity of these models should be constructed in accordance with empirically supported background theory, capturing the features of the system that make a difference on the target. As such, controls should be implemented to account for the influence of known confounding causes. Meanwhile, confidence in our measurements can be bolstered with the use of independent measurement techniques, so long as the results are robust.

3.2 Measurement problems as inadequate models of the measurement process

We can frame what has gone wrong with California's forest carbon offset program using this understanding of models of the measurement process. Remember that Badgley et al. (2022a) claim that over thirty percent of upfront credits awarded by California's forest offset program are the result of actual local conditions varying from regional averages. Common practice estimates were determined by "averaging dissimilar tree species across arbitrarily defined geographic regions" (Badgley et al., 2022a, p. 1442). Some have been quick to point to the problem as relating to statistical artifacts in the generation of averages (Badgley et al., 2021), though we should be careful not to think the problem is inherent to the methodology of averages.[2] Of course any application of FIA data will appeal to carbon averages of some sort, whether they be tree-species averages, averages for a tree-species within a specific region, or something more fine-grained. It is more precise to understand the fault here to be a reliance on the *wrong* average. What makes something the *wrong* average, I will show, is a failure to model the relevant difference makers in the measurement of real-world climate benefits.

California's forest carbon offset program affords credits on the basis of how well a project's carbon stock exceeds its projected baseline. This baseline is partly determined by the regional average of dissimilar trees. However, insofar as a region contains a variety of landscapes with a variety of diverse tree species, the regional average will wash away the influence of relevant difference makers. This is why, in addition to the over-crediting, some projects assessed by Badgley et al. were assessed to be a victim of under-crediting. If a region is comprised of diverse landscapes, with diverse species distributions over its numerous parcels, then a regional average will fail to capture the factors that influence the quantity of carbon stock in a given parcel of land. The regional average fails to capture features of the system that background scientific theory implies are important for determining forest carbon stock. The distribution of tree species is a crucial determinant of the amount of carbon, and the regional average (of necessity) ignores the real-world deviation from the mean. As a result of this, regional averages fail to provide a robust prediction of the forest carbon present at smaller scales within the region. The regional average is the *wrong* average to use in assessing baseline of a local project because it is a poor measure for forest carbon stock. Species variation within the region makes a difference to the carbon stock, and the measurement model ignores species variation.

Rather, an appropriate average must consider the sorts of real-world processes that make a difference to the target phenomenon. Hypothetically, if the primary difference maker for the carbon stock of a parcel is the presence of freshwater lakes, then a project's

---

[2] I suspect Badgley et al. (2021, 2022a) would agree with this point, though the framing of the issue on the CarbonPlan site might be misread as suggesting that averaging over diverse landscapes is sufficient for deviation from real climate benefit. Even if I am reading too much into the stated "problem with averages", we can understand my philosophical contribution to be that of making explicit what makes a *good* average.

baseline should be determined partly in accordance with whether the land does or does not contain freshwater lakes. There should at least be a carbon stock average for lake-containing land and one for land without lakes. Similarly, if the distribution of tree species is the primary difference maker for carbon stock, then a project's baseline will need to account for the distribution of trees. One could use the average carbon stock from land with 25% Douglas-Fir and 75% Oak to determine the baseline for land that is *evaluated as containing about 25% Douglas-Fir and 75% Oak*. Certainly, the species composition of a forest is one difference maker for the forests carbon stock, and for this reason it is a scandal that policy makers ignore local species composition when fixing the baseline.

In other words, we can say that the measurements of a site's carbon stock failed to be appropriately correlated to the site's *actual* carbon stock. The FIA informed regional averages failed to provide robust estimates of the quantity of carbon, and their application ignored species-based difference makers in carbon suggested by background theory. There is no robust model of the measurement process forthcoming that correlates a site's actual carbon with the current FIA informed regional averages.

The other problems with California's forest carbon offset program (Section 2.2) can also be viewed through the lens of measurement models. The difference between a project's carbon stock average over the next century and the counterfactual "do-nothing" average over the next century is difficult to estimate when: (1) a project's carbon stock is subject to hazardous uncertainties, threatening the permanence of the carbon stock, and (2) there is no single counterfactual scenario to consider. Uncertainties regarding the future generate uncertainties in the quantities that constitute the target of measurement, and so we do not know enough about the world to construct a reliable model of the measurement process. The way our climate system evolves in the near future will make a significant difference to the performance of particular CDR projects, and our best climate science suggests a significant range of viable possibilities. It isn't clear what epistemic reason we have for discriminating among the different possibilities, whether such possibilities are articulated in terms of emissions scenarios or individual model performance (given model disagreement). As such, it isn't clear what the model of the measurement process should look like, which features of the world need to be included and how to understand their dependency. It is like committing to the temperature indicated by a mercury thermometer fifty years in the future, even though there is a reasonable chance that the glass of the thermometer breaks and some of the mercury is lost.

Further concerns arise when we start awarding money and permitting pollution on the basis of such measurements, since it is not that unlikely for the quantity of sequestered carbon corresponding to the award or permitting the excess pollution to also end up in the atmosphere. Suddenly it turns out that the financial resources intended to mitigate the harms of climate change are achieving nothing, or, worse, have resulted in more carbon in the atmosphere than there would have been if we had done nothing at all. In short, the financial awards intended to drive mitigation efforts may no longer be making a difference in the right direction when a project is awarded upfront, failing to appropriately respect uncertainty about the future in the measurement model. Ultimately, it is the imperative to model all significant difference makers that will guide the development of an alternative *absolute* approach I advance in the next section (Section 4).

When Gifford (2020) raises concerns of *subjectivity* in carbon accounting, she cites several distinct ways in which carbon measurements might come apart from real benefits. We've already discussed the variability across interpretations of baseline and additionality, which Gifford flags as being "deeply subjective." Part of what Gifford is highlighting here is the distinct problem of *counterfactual uncertainty* as it relates to determinations of baseline and additionality. CarbonPlan describes these counterfactual uncertainties as those arising from "assumptions about what would have happened in the absence of a project" (Chay et al., 2022). A robust model of the measurement process ought to mitigate the impact of counterfactual uncertainty, but as mentioned above (Section 2.2) there are a number of deep counterfactual uncertainties that cannot be theoretically resolved or sufficiently constrained at the moment.

Gifford's concern over subjectivity isn't just that there are different viable interpretations in the calculation of baseline and additionality, however, but that standards are sufficiently permissive as to encourage the systematic influence of self-interest. The Greenhouse Gas Management Institute (GHGMI) is one of the major organizations offering training and certification for the measurement and accounting of greenhouse gas emissions. However, accounting standards are flexible enough that participants in accounting courses offered by the GHGMI are instructed to select (or *create*) a method and criteria for quantifying carbon that "addresses a desired outcome" (Chay et al., 2022, p. 296). A robust model of the measurement process could not permit such a strong influence of "subjectivity" on the measurement of carbon, insofar as the subjective processes confound the relationship between measurement outcomes and actual carbon stock. A better approach should thus control the influence of these "subjective" processes (a la Wilson and Boudinot), better constraining the relationship between financial incentives and real-world climate benefits. To be clear, these concerns regarding subjective confounds are not specific to the comparative approach, but rather highlight the general need for measurement standards to better accord with our nuanced scientific understanding of the system.

It is true, however, that certain counterfactuals must be included in the measurement model if it is to capture the dynamic relationship between the indicator and the measurement target. We shouldn't understand the problem of counterfactual uncertainty as being a problem with counterfactual reasoning in general. For example, a good measurement model will help workers predict the quantity of carbon stock for any potential distribution of known tree species. But this is just a form of counterfactual reasoning. *If* the land contained a 50/50 split of Oak and Douglas Fir, it *would* contain such-and-such amounts of carbon. There is nothing problematic about such counterfactuals, since their truth can be empirically and theoretically supported, which is indeed what the construction of the measurement model is all about. Some conditional counterfactuals about the future can even be mitigated through the use of empirically and theoretically grounded simulation models, or appropriately targeted paleostudies (Wilson, 2023). As such, counterfactuals themselves are not a problem for the reliability of assessing CDR technologies, rather it is significant unresolved uncertainties that are the problem.

In this section I have drawn some important ideas from the philosophy of measurement, primarily those pertaining to the model of the measurement process and its implications for the relationship between a measurement target and the measurement outcome. Ultimately, I take this framework to highlight how an approach to carbon measurement that is less vulnerable to the financial speculation or manipulation engendered by counterfactual uncertainty, and more beholden to theoretically supported empirical methods for carbon measurement, would be preferable to the kind of comparative approach we find in the California forest carbon offset program. In the next section I explore what one such approach might look like.

4 An alternative approach to real climate benefits

As Cooper (2015) highlights, *"The work of metrology is fundamental to defining the 'thing' to be exchanged in a market through the assignment and verification of particular characteristics."* However, if the *thing* currently being exchanged does not adequately achieve our aims, then we can consider an alternative thing, with alternative characteristics, to be the target for environmentally oriented financial policies. More straightforwardly, if we are unhappy with how carbon measures currently credit carbon, then we can consider an alternative approach. In this section I will discuss one such alternative inspired by our consideration of models of the measurement process, an alternative way to understand the measurement of carbon and real climate benefits. This absolute (vs. comparative) approach quantifies real benefits in terms of *actual* carbon sequestered. I will argue that this absolute approach is less vulnerable to challenging counterfactual uncertainties, resulting in a model of the measurement process that more strongly links financial incentives to sequestered carbon. Furthermore, I will argue that taking this absolute perspective on carbon serves to better promote the ideal of *permanent* carbon storage.

An alternative approach to real climate benefits for carbon dioxide removal projects does not attempt to determine a status quo baseline for the next century, but rather interprets climate benefits exclusively in terms of the quantity of carbon that is presently sequestered. While comparative approaches understand benefit relative to some projection for the future, the proposed alternative quantifies climate benefits in terms of actual carbon. In short, the more carbon is sequestered the better it is for the climate. Instead of tying financial incentives to how well a project is expected to exceed average common practice over the next century, financial incentives would be tied to how much carbon is presently observed to be sequestered. Credits would thus be doled out on a regular (e.g., annual) basis in proportion to extant carbon, such that only actual carbon reserves would be paid.

Consider this hypothetical sketch of such a program. You own 20 acres of forestland, each acre containing 20 metric tons of carbon in aboveground biomass. You intend to manage the land, increasing the carbon that it will hold, and so you sign up for the carbon sequestration reward program. Suppose the program requires you commit to a century of management. Your land is estimated to contain 400 metric tons of carbon on the first annual assessment of carbon stock, and so you are awarded 400 (tons of carbon) divided by 100 (total year commitment) units of credit for the 1st year of your project.[3] If nothing changes, then the full 400 tons of carbon are rewarded by the time the century-long commitment is up. If improved management increases the amount of aboveground carbon every subsequent year, then more credits are earned at each annual assessment in proportion to the increase. If hazards strike, decreasing the carbon stock of the land, then fewer credits are earned at each annual assessment in proportion to the loss.

While many of the details of this hypothetical program are free to vary, it will continue to utilize the absolute approach to climate benefits insofar as it credits *actual* quantities carbon based on the *amount sequestered*. Neither counterfactual uncertainty about the future nor inadequate baseline determination will drive a wedge between real climate benefits and financial awards. Switching the target of measurement to *actual* carbon stock enables the construction of a more reliable and robust measurement model so far as empirical techniques are capable of deriving reliable carbon estimates from the observable properties of the land. Projections about the uncertain future are not necessary for generating a measurement model for financial awards. Instead, a much greater emphasis is placed on theoretically constrained empirical estimates for how a project was executed (execution uncertainty), given existing theoretical uncertainties (scientific uncertainty). Whereas many important counterfactual uncertainties are intractable, execution and scientific uncertainty can be tackled with careful application of tools and practices alongside targeted research efforts (as discussed by Chay et al., 2022).

A further upshot to this alternative approach to climate benefit is how it incentivizes a careful and persistent consideration of the carbon stock, over more ambitious (yet risky and empirically tenuous) projects. One major problem with providing upfront credits conjoined with insurance policies that do not remediate carbon loss is the failure to incorporate the *permanence* of carbon sequestration into the valuation (Macintosh, 2013; D'Alisa and Kallis, 2016; Gifford, 2020). A project is always overcredited if the carbon stock goes up in flames. I suspect that the financial structure that emerges from the alternative absolute approach affords more value to the permanence of carbon sequestration. With financial reward being tied to total carbon stock over longer intervals of time, managers are motivated to protect the carbon already being sequestered and implement reliable conservationist techniques for increasing carbon stock. The financial value of the asset (in the carbon-credit sense) is more directly tied to

---

[3] Let me note that the payout structure of this example will almost certainly need to be complicated to account for discounting and other economic realities (perhaps providing a greater payout for carbon near the end of the program). Furthermore, as I highlight later in the section, I also expect that such a program would need to be conjoined with a carbon tax to avoid the emergence of certain perverse incentives (to protect carbon stock after the policy duration has elapsed). However, the simplified example will suffice for instructive purposes.

the quantity of sequestered carbon, and the financial reward is disbursed throughout the desired period of sequestration.

Many financial aspects remain to be determined, like the credit value to be ascribed to a unit of carbon and the temporal structure of the award. I leave most of this task to the economists, though the pricing must ultimately suffice to incentivize sequestration over alternatives. The ultimate aim of establishing carbon markets is to guide agents, acting in their own self-interest, to act in ways that sequester carbon. Whatever else is true of the price of carbon, it must be such that land managers see a commitment to grow and preserve carbon stocks as a worthwhile or generally preferable financial option. The alternative is to admit that carbon markets are incapable of serving conservationist aims and ought to be scrapped in favor of more firm-handed environmentalist policies. Optimistic that there is some suitable approach to a regulated carbon market, I leave ironing out the details to the economists.

There are also a number of costs for the absolute approach that come along with a more focused attempt to link sequestered carbon to credits. First, I suggested above that the absolute approach puts a greater emphasis on resolving execution and scientific uncertainty, since it turns on more precise and developed models of the target. This means that successful implementation will place greater demands on background theory and engineering practices, encouraging potentially costly research efforts when error arises. While current approaches are also in need of targeted research efforts, a greater degree of precision and understanding may often be required for the more precise models of the measurement process encouraged by the absolute approach. This means that the absolute approach may expect more out of our scientists and engineers than competing approaches, making it more likely that workers will in practice bump up against the limits of our scientific understanding and engineering prowess.

Second, the absolute approach is likely to be less attractive to investors, making it a less marketable approach all other things being equal. Disbursing credits at more regular intervals on the basis of actual achievements, instead of disbursing them upfront on the basis of projections, places a respective limit on how much immediate financial gain is possible. Furthermore, tying carbon quantities more directly to credits also means that a loss in carbon should induce a corresponding financial loss. This exposes project managers to significant financial risks that were previously forgiven by insurance (remember that project managers could retain their credits even in the case of total carbon loss).

Third, the absolute approach would plausibly require additional policies to function as intended. A carbon tax would be needed to prevent perverse incentive structures from arising, e.g., incentivizing landowners to preserve carbon stock even after their policy lapses and there are no more credits to be gained. In this sense participation in the carbon market would be compulsory to a certain extent: while incentivized to participate in projects that award benefits for carbon storage, landowners would be legally required to pay for carbon losses. Furthermore, implementing the requisite regulatory infrastructure would be a fairly massive undertaking, requiring the collaborative efforts of stakeholders like policymakers, governmental agencies, environmental organizations, environmental lawyers, and so on. As such, a significant amount of additional work would be required to, both, iron out the details for the necessary policies (local, state, national, and international) and get the policies adopted within their respective locale. That is, while all carbon mitigation efforts require some intervention on policy, the absolute approach should require greater effort than approaches that rely primarily on features of the existing political-economic landscape.

While I am confident that each these costs help enable the absolute approach to better constrain the relationship between carbon sequestration and its financial incentives, I suspect that it is for some combination of these reasons that comparative approaches are more commonly discussed.

In this section I have outlined an alternative approach for understanding real benefits to the climate system, one that focuses on the absolute quantity of carbon over time instead of comparison to a baseline average. The approach places an emphasis on the regular and accurate measurement of actual carbon over projections, overgeneralized estimates, and regional averages. However, since the absolute approach imposes a greater financial risk to project managers, imposing greater costs in the case a projects carbon is lost, it is worth considering the degree to which specific CDR methods are vulnerable to future climate uncertainty. In the following section I consider the relative permanence for different CDR methods.

5 Permanence in carbon dioxide removal

Insofar as actual carbon stock is the target of measurement, and its sustained maintenance the aim of environmental financial policies, we can consider which are the most promising of the major approaches to carbon dioxide removal (CDR). Even if financial awards have been disentangled from counterfactual uncertainty, uncertainties about the future (including scientific and execution uncertainties) still impose a risk to CDR projects. Our goal is to sequester carbon and do it, all else being equal, for as long as possible, and so we might consider which methods of CDR best approach the ideal of permanence. The natural systems in which carbon is sequestered exhibit different degrees of sensitivity to surrounding environmental conditions, and so are more or less vulnerable to environmental uncertainties about the future. As such, models of the measurement process will exhibit differing degrees of robustness or permanence in the face of such uncertainties. I argue that we can divide methods for CDR into three broad categories with regard to their permanence: those that are *permanent* with respect to typical century and millennia timescales, those that are *risky*, and those that are *transient*.

Among the *permanent* methods are those that promote the geological storage of carbon via terrestrial mineralization or weathering, and direct air capture. What makes these methods permanent is ultimately their utilization of geological storage. Storage in causally isolated and inert, underground geological formations protects the sequestered carbon from the influence of destructive natural processes. This will typically result in carbon dioxide that is trapped in the pores of the rock (Krevor et al., 2015), dissolved in brine residing in those pores (Emami-Meybodi et al., 2015), or mineralized with rock and pore fluid (Matter et al., 2016; Zhang and DePaolo, 2017; Kelemen et al., 2019). This allows more reliable projection of the carbon stock going into the future, producing a simpler model of the measurement process: the measurement model must account for the dynamic processes influencing the measured quantity of carbon stock, and carbon sequestered in the right geological formations will be subject to fewer dynamic processes. If carbon is to be traded in an offset program, permitting the exchange of excess emissions for increases in carbon sequestration elsewhere, then these permanent CDR methods should be preferred in virtue of their minimizing the potential for unforeseen destructive processes. This carbon more precisely corresponds to real world climate benefits (i.e., less carbon in the atmosphere).

Terrestrial mineralization occurs when natural silicates or alkaline industrial mining waste mineralizes carbon. This can occur in *ex-situ, in-situ* (e.g., underground minerals), or surface contexts. *Ex-situ* use involves the extraction of the alkaline material for use "off-site" in locations like high pressure and high temperature reactors that permit enhanced reactivity (e.g., Pan et al., 2020). *In-situ* use keeps the alkaline material "on-site," producing subsurface mineralization by way of circulating carbon rich fluids through the alkaline rock (e.g., Wilcox et al., 2017). Surficial use emphasizes ambient weathering of alkaline mining waste (e.g., mafic and ultramafic mine tailings) via surface atmospheric and hydrological processes like precipitation (e.g., Mervine et al., 2018). All mineralization efforts result in the production of carbonate rock. This carbonate can then be stored in the appropriate geological contexts, removed from destructive natural processes, sequestering the carbon for as long as the geologic formation remains impermeable and isolated. The mineralized carbon can also be sold as building materials, or used to fertilize soil, though both uses significantly reduce the permanence of the carbon sequestration. Fertilizer qualifies as what I will be calling *transient* carbon storage.

Direct air capture (DAC) utilizes a variety of alternative chemical approaches to the capture of ambient carbon dioxide (Kumar et al., 2015; Keith et al., 2018). DAC devices are constructed so that fans circulate air to put it in contact with water-based solvents or synthetic sorbents. Carbon dioxide in the air ultimately binds with the reactive agent to form carbamate or carbonate bonds, while the remaining components (primarily nitrogen and oxygen) of the air circulate through the device unchanged. The chemical bonds are later broken to extract the collected carbon dioxide, before being compressed and transported. Insofar as this carbon is transported into the appropriate geological formations, the carbon is stored permanently.

Methods that are *risky* with regard to permanence include ocean alkalinity enhancement, and both terrestrial and coastal biomass sinking. What makes the methods *risky* is that the sequestered carbon stock remains well integrated in the uncertain and destructive natural processes occurring near the surface of the earth. As such, models of the measurement process need to incorporate the influence that these confounding causal processes will have on carbon stock in order to generate reliable projections of the future. We will see that while carbon stock is stable under a certain set of model assumptions, for each risky method there is at least one potential threat to permanence that our best scientific understanding of the climate suggests is reasonable to worry about.

Among terrestrial sinking projects is where we find the improved forest management methods credited in California's carbon offset program, as well as afforestation and reforestation efforts. The method should be fairly clear by now: forests are sites where a significant quantity of biotic carbon is stored. Thus, the generation of new forests, the regeneration of old forests, and the improved management of existing forests serves to increase the stock of terrestrial carbon. I suggested earlier that financial programs ought to incentivize project managers to protect and promote the development of such carbon stocks. However, there are relevant uncertainties in the preservation of forests: the possibility of wildfire or pests generates an existential threat to the carbon stock of a flourishing forest. Uncertainty in future precipitation patterns exacerbate those wildfire worries while generating additional concerns (Zappa et al., 2021): how permanent will a forest be if it no longer receives adequate rainfall? As such, there are a number of uncertain processes that threaten the reliable projection of terrestrial carbon stock.

Marine and coastal biomass sinking projects suffer a structurally similar concern, though the processes threaten at decidedly slower rates. Plants and soils in coastal ecosystems provide another source for the sequestration of biotic carbon stock in the generation and management of seagrass meadows, mangrove forests, marshes, and other coastal wetlands (Pendleton et al., 2012; Kroeger et al., 2017). While these landscapes are not particularly vulnerable to wildfire, impending changes in sea level, temperature, salinity, nutrient availability, and even pollution do threaten the stability of coastal ecosystems. Our inability to better constrain the changing of our oceans into the next century thus generates a similar concern regarding our measurement models of coastal carbon. While we can be confident that things will change, the specific changes for many coastal regions cannot be sufficiently pinned down for biomass sinking projects to be permanent.[4]

Ocean alkalinity enhancement, our last of the *risky* methods, works to increase the uptake of carbon dioxide by the ocean itself, primarily by expanding and accelerating the dissolution of carbonate and silicate minerals into carbonate and bicarbonate ions (and their associated cations). Methods for increasing alkalinity include the deposition of alkaline minerals (Renforth et al., 2013) and the construction of seawater reactors to promote weathering (Rau, 2011). The result is carbon dioxide that is sequestered in the water itself. This method is *risky* with regard to permanence, however, because carbonate ions in the ocean are used by marine organisms in the construction of their calcite shells, a process that itself releases carbon dioxide [It is also risky in the more traditional sense with regard to potential impacts on ocean ecosystems (e.g., Bach et al., 2019)]. Thus, biological calcite formation serves as a negative feedback on the storage of carbonate ions in the ocean, resulting in an ever-present leak that is proportional to the concentration of carbonate ions (given the presence of shell-forming organisms).

The remaining *transient* CDR method is biomass energy with carbon capture and storage (BECCS). I refer to BECCS as *transient* because the carbon, as biomass *energy*, is sequestered with the intention of being released again into the atmosphere. Workers treat BECCS as a method for CDR when the carbon drawn from the atmosphere is more than is released in the production and utilization of biomass energy. While there are a number of methods

---

[4] Ultimately, there may be additional scientific reasons to be concerned with the viability of coastal carbon efforts (Williamson and Gattuso, 2022).

for producing biomass energy, one promising approach lies in capturing the carbon dioxide emitted in the production of ethanol via fermentation (e.g., Lynd et al., 2017). For example, yeast or bacteria can be used to ferment corn products into ethanol, and the carbon dioxide released in the process may be captured for storage. Thus, it may be more precise to understand BECCS as conjoining two carbon-relevant processes: (1) the generation of biomass energy and (2) the storage of carbon dioxide emitted during the process. It is biomass energy that qualifies as *transient*, shortly to be used as fuel and returned to the atmosphere. The captured emissions may qualify as permanent, being a form of DAC described above, so long as the carbon dioxide makes its way to one of our trusty geological formations.

So, there we have our three categories of permanence for CDR techniques. Insofar as our goal is to promote more permanent carbon sequestration over less permanent carbon storage, we should prefer CDR methods that are permanent to those that are risky, and those that are risky to those that are transient, all else being equal. But of course, not all else is equal, and there are finite spaces available for geological storage. As such, in providing this analysis I do not mean to suggest that CDR methods achieving relative permanence ought to, for that reason alone, be preferred to more risky or transient methods of CDR. Rather, I mean to highlight that the contribution of certain CDR methods to the goal of keeping carbon out of the atmosphere will be more readily quantifiable, such that a project's carbon stock is more reliably tied to real world climate benefits. The carbon removed in the generation of biomass energy should not be credited, for example, if it is soon to be released back into the atmosphere. Carbon should be evaluated differently depending on how securely it is sequestered.

6 Concluding remarks

In this paper I have investigated some problems with California's forest carbon offset program, including problems pertaining to uncertain future performance, subjective methodologies, and the establishment of wrong averages, as well as problems determining baseline arising from counterfactual uncertainty. These latter problems emerge for any comparative approach that quantifies baseline estimates in terms of what would happen over the course of the next century in the absence of a specific project. Confronted with these problems, I have drawn from some ideas from the philosophy of science, many from the philosophy of measurement, to provide a general theoretical framework for reliable measurement. This framework focuses on the development of a robust model of the measurement process, which represents the features of the world that make a difference to how the measurement target relates to our measurement technique. In this framework we can understand the sequestered carbon necessary for real world climate benefits to be our measurement target, while understanding the noted problems in California's forest offset program to introduce error and uncertainty into our model of the measurement process. One of the key insights afforded by this framework is that current over-crediting in the program is not the result of standardizing averages, but the result of standardizing the wrong kinds of averages. Standardized averages must be appropriately sensitive to the real-world processes that make a difference to the measurement target, unlike current forest carbon standards that coarsely represent the distribution of tree species in a region.

Problems quantifying counterfactual baselines are specifically intractable, and so in striving to reduce the error in our model of the measurement process, I sketch an alternative proposal to the comparative approach of carbon valuation. This absolute alternative does not rely on comparison with counterfactual baselines to determine financial awards, but rather looks to quantities of actual sequestered carbon in the present. While this alternative does have its constraints, it seeks to minimize error in our model of the measurement process predominantly by linking financial awards more directly to actual carbon quantities over time. However, while the absolute approach better links financial metrics to real climate benefits, uncertainties about the future still impose varying levels of risk for particular methods for carbon dioxide removal (CDR). In light of this, I conclude with a consideration of the relative permanence of the carbon sequestered by different CDR methodologies. Different physical systems will exhibit differing degrees of sensitivity to general sorts of changes we expect the earth's climate to experience in the next century, suggesting that a model of the measurement process for each method will incorporate more or fewer potential causal confounds (e.g., destructive wildfires or pests in the case of forest sequestration). In short, in the face of uncertainty about the future, some CDR methods will be more secure as a result of being causally insulated from certain environmental phenomena.

An incidental upshot of my argument, though perhaps its most important consequence, is that in engaging with CDR from a philosophical perspective (primarily from the philosophy of measurement), the kinds of ideas, insights, and frameworks that have been useful in more general philosophical theorizing might travel and find use across disciplinary boundaries. While admittedly abstract, I suspect that many workers will find it helpful to have a general schema for what goes into reliable measurement, and how to interpret and address measurement problems. The notion that measurements and quantitative estimates require a model of the measurement process with particular properties may provide a useful lens through which workers can view some of their work. Addressing climate change is a wildly transdisciplinary task, requiring the cooperation of experts across numerous domains, and so it is beneficial in the assessment and application of CDR methods that researchers have access not just to empirical results from other fields but also the ideas, insights, and frameworks that have been fruitful in other domains. I hope this paper helps further the present transdisciplinary discussion of climate change and environmental policy.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

JW: Conceptualization, Writing – original draft.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

A special thanks goes to F. Garrett Boudinot for introducing me to much of the important scientific work being done in Carbon Dioxide Removal and entertaining some of the thoughts and questions I had throughout the development of this paper. I would also like to thank two referees for their constructive comments.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bach, L. T., Gill, S. J., Rickaby, R. E., Gore, S., and Renforth, P. (2019). CO2 removal with enhanced weathering and ocean alkalinity enhancement: potential risks and co-benefits for marine pelagic ecosystems. *Front. Clim.* 1, 7. doi: 10.3389/fclim.2019.00007

Badgley, G., Chay, F., Chegwidden, O. S., Hamman, J. J., Freeman, J., and Cullenward, D. (2022b). California's forest carbon offsets buffer pool is severely undercapitalized. *Front. Forests Global Change* 5, 154. doi: 10.3389/ffgc.2022.930426

Badgley, G., Freeman, J., Hamman, J. J., Haya, B., Trugman, A. T., Anderegg, W. R., et al. (2021). "Systematic over-crediting of forest offsets" CarbonPlan. Available online at: https://carbonplan.org/research/forest-offsets-explainer (accessed August 25, 2023).

Badgley, G., Freeman, J., Hamman, J. J., Haya, B., Trugman, A. T., Anderegg, W. R., et al. (2022a). Systematic over-crediting in California's forest carbon offsets program. *Global Change Biol.* 28, 1433–1445. doi: 10.1111/gcb.15943

Boudinot, F. G., and Wilson, J. (2020). Does a proxy measure up? A framework to assess and convey proxy reliability. *Clim. Past* 16, 1807–1820.

Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford: Oxford University Press. doi: 10.1093/0198247044.001.0001

Chang, H. (2004). *Inventing Temperature: Measurement and Scientific Progress*. Oxford: Oxford University Press. doi: 10.1093/0195171276.001.0001

Chay, F., Klitzke, J., Hausfather, Z., Martin, K., Freeman, J., and Cullenward, D. (2022). *"Verification Confidence Levels for carbon dioxide removal" CarbonPlan.* Available online at: https://carbonplan.org/research/cdr-verification-explainer (accessed August 25, 2023).

Cooper, M. H. (2015). Measure for measure? Commensuration, commodification, and metrology in emissions markets and beyond. *Environ. Plan. A.* 47, 1787–1804. doi: 10.1068/a130275p

D'Alisa, G., and Kallis, G. (2016). A political ecology of maladaptation: insights from a Gramscian theory of the State. *Global Environ. Change* 38, 230–242. doi: 10.1016/j.gloenvcha.2016.03.006

Emami-Meybodi, H., Hassanzadeh, H., Green, C. P., and Ennis-King, J. (2015). Convective dissolution of CO2 in saline aquifers: Progress in modeling and experiments. *Int. J. Greenhouse Gas Control* 40, 238–266. doi: 10.1016/j.ijggc.2015.04.003

Feinberg, J. (1984). *Harm to Others*. Oxford University Press.

Giere, R. N. (1988). *Explaining Science: A Cognitive Approach*. Chicago, IL: University of Chicago Press. doi: 10.7208/chicago/9780226292038.001.0001

Giere, R. N., Bickle, J., and Mauldin, R. F. (2006). *Understanding Scientific Reasoning*. London: United Kingdom: Thomson/Wadsworth.

Gifford, L. (2020). "You can't value what you can't measure": a critical look at forest carbon accounting. *Clim. Change* 161, 291–306. doi: 10.1007/s10584-020-02653-1

Godfrey-Smith, P. (2006). The strategy of model-based science. *Biol. Philos.* 21, 725–740. doi: 10.1007/s10539-006-9054-6

Hovorka, S., Kelemen, P., Wilcox, J., Kolosz, B., and Freeman, J. (2021). *The Building Blocks of CDR Systems: Geological Sequestration*. Oxford: CDR Primer.

IPCC (2023). Summary for Policymakers. In: *Climate Change 2023: Synthesis Report. A Report of the Intergovernmental Panel on Climate Change. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change,* eds. Core Writing Team, H. Lee and J. Romero (Geneva, Switzerland: IPCC).

Keith, D. W., Holmes, G., Angelo, D. S., and Heidel, K. (2018). A process for capturing CO2 from the atmosphere. *Joule* 2, 1573–1594. doi: 10.1016/j.joule.2018.05.006

Kelemen, P., Benson, S. M., Pilorgé, H., Psarras, P., and Wilcox, J. (2019). An overview of the status and challenges of CO2 storage in minerals and geological formations. *Front. Clim.* 1, 9. doi: 10.3389/fclim.2019.00009

Krevor, S., Blunt, M. J., Benson, S. M., Pentland, C. H., Reynolds, C., Al-Menhali, A., et al. (2015). Capillary trapping for geologic carbon dioxide storage–From pore scale physics to field scale implications. *Int. J. Greenhouse Gas Control* 40, 221–237. doi: 10.1016/j.ijggc.2015.04.006

Kroeger, K. D., Crooks, S., Moseman-Valtierra, S., and Tang, J. (2017). Restoring tides to reduce methane emissions in impounded wetlands: A new and potent Blue Carbon climate change intervention. *Scient. Rep.* 7, 11914. doi: 10.1038/s41598-017-12138-4

Kumar, A., Madden, D. G., Lusi, M., Chen, K. J., Daniels, E. A., Curtin, T., et al. (2015). Direct air capture of CO2 by physisorbent materials. *Angew. Chem. Int. Ed.* 54, 14372–14377. doi: 10.1002/anie.201506952

Lynd, L. R., Liang, X., Biddy, M. J., Allee, A., Cai, H., Foust, T., et al. (2017). Cellulosic ethanol: status and innovation. *Curr. Opin. Biotechnol.* 45, 202–211. doi: 10.1016/j.copbio.2017.03.008

Macintosh, A. (2013). Coastal climate hazards and urban planning: how planning responses can lead to maladaptation. *Mitig. Adapt. Strat. Global Change* 18, 1035–1055. doi: 10.1007/s11027-012-9406-2

Mari, L., Carbone, P., Giordani, A., and Petri, D. (2017). A structural interpretation of measurement and some related epistemological issues. *Stud. History Philos. Sci. Part A* 65, 46–56. doi: 10.1016/j.shpsa.2017.08.001

Matter, J. M., Stute, M., Snæbjörnsdottir, S. Ó., Oelkers, E. H., Gislason, S. R., Aradottir, E. S., et al. (2016). Rapid carbon mineralization for permanent disposal of anthropogenic carbon dioxide emissions. *Science* 352, 1312–1314. doi: 10.1126/science.aad8132

Mervine, E. M., Wilson, S. A., Power, I. M., Dipple, G. M., Turvey, C. C., Hamilton, J. L., et al. (2018). Potential for offsetting diamond mine carbon emissions through mineral carbonation of processed kimberlite: an assessment of De Beers mine sites in South Africa and Canada. *Mineral. Petrol.* 112, 755–765. doi: 10.1007/s00710-018-0589-4

Pan, S. Y., Chen, Y. H., Fan, L. S., Kim, H., Gao, X., Ling, T. C., et al. (2020). CO2 mineralization and utilization by alkaline solid wastes for potential carbon reduction. *Nat. Sustain.* 3, 399–405. doi: 10.1038/s41893-020-0486-9

Parfit, D. (1984). *Reasons and Persons*. Oxford University Press.

Pendleton, L., Donato, D. C., Murray, B. C., Crooks, S., Jenkins, W. A., Sifleet, S., et al. (2012). Estimating global "blue carbon" emissions from conversion and degradation of vegetated coastal ecosystems. *PLoS ONE* 7, e43542. doi: 10.1371/journal.pone.0043542

Pilorgé, H., Kolosz, B., Wu, G. C., Freeman, J., and Wilcox, J. (2021). Global mapping of CDR opportunities. *Carbon Dioxide Removal Primer (cdrprimer. org).* Available online at: https://cdrprimer.org/read (accessed August 25, 2023).

Rau, G. H. (2011). CO2 mitigation via capture and chemical conversion in seawater. *Environ. Sci. Technol.* 45, 1088–1092. doi: 10.1021/es102671x

Renforth, P., Jenkins, B. G., and Kruger, T. (2013). Engineering challenges of ocean liming. *Energy* 60, 442–452. doi: 10.1016/j.energy.2013.08.006

Tal, E. (2017). Calibration: Modelling the measurement process. *Stud. History Philos. Sci. Part A* 65, 33–45. doi: 10.1016/j.shpsa.2017.09.001

Wilcox, J., Psarras, P. C., and Liguori, S. (2017). Assessment of reasonable opportunities for direct air capture. *Environ. Res. Lett.* 12, 065001. doi: 10.1088/1748-9326/aa6de5

Williamson, P., and Gattuso, J. P. (2022). Carbon removal using coastal blue carbon ecosystems is uncertain and unreliable, with questionable climatic cost-effectiveness. *Front. Clim.* 4, 853666. doi: 10.3389/fclim.2022.853666

Wilson, J. (2023). Paleoclimate analogues and the threshold problem. *Synthese* 202, 17. doi: 10.1007/s11229-023-04202-6

Wilson, J., and Boudinot, F. G. (2022). Proxy measurement in paleoclimatology. *Eur. J. Philos. Sci.* 12, 1–20 doi: 10.1007/s13194-021-00444-8

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation.* Oxford: Oxford University Press. doi: 10.1093/0195155270.001.0001

Zappa, G., Bevacqua, E., and Shepherd, T. G. (2021). Communicating potentially large but non-robust changes in multi-model projections of future climate. *Int. J. Climatol.* 41, 3657–3669. doi: 10.1002/joc.7041

Zhang, S., and DePaolo, D. J. (2017). Rates of CO2 mineralization in geological carbon storage. *Acc. Chem. Res.* 50, 2075–2084. doi: 10.1021/acs.accounts.7b00334