ENVIRONMENTAL RESEARCH
LETTERS

TOPICAL REVIEW • OPEN ACCESS

Characterizing and communicating uncertainty: lessons from NASA's Carbon Monitoring System

To cite this article: Robert E Kennedy *et al* 2024 *Environ. Res. Lett.* **19** 123003

ENVIRONMENTAL RESEARCH
LETTERS

CrossMark

OPEN ACCESS

RECEIVED
31 August 2023

REVISED
7 October 2024

ACCEPTED FOR PUBLICATION
28 October 2024

PUBLISHED
26 November 2024

TOPICAL REVIEW

Characterizing and communicating uncertainty: lessons from NASA's Carbon Monitoring System

Robert E Kennedy[1,*], Shawn P Serbin[2], Michael C Dietze[3], Hans-Erik Andersen[4], Chad Babcock[5], David F Baker[6], Molly E Brown[7], Kenneth J Davis[8], Laura Duncanson[7], Sha Feng[9], Andrew T Hudak[10], Junjie Liu[11], Paul L Patterson[12], Brett Raczka[13], Mark A Cochrane[14], Edil A Sepúlveda Carlo[15] and Rodrigo Vargas[16]

[1] College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR, 97331, United States of America
[2] Biospheric Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, 20771, United States of America
[3] Department of Earth & Environment, Boston University, Boston, MA, 02215, United States of America
[4] Pacific Northwest Research Station, USDA Forest Service, Seattle, WA, 98040, United States of America
[5] Forest Resources Department, University of Minnesota, St. Paul, MN, 55108, United States of America
[6] Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO, 80524, United States of America
[7] Geographical Sciences Department, University of Maryland, College Park, MD, 20742, United States of America
[8] Department of Meteorology and Atmospheric Science and Earth and Environmental Systems Institute, The Pennsylvania State University, University Park, PA, 16801, United States of America
[9] Atmospheric, Climate, and Earth Sciences Division, Pacific Northwest National Laboratory, Richland, WA, 99354, United States of America
[10] Rocky Mountain Research Station, USDA Forest Service, Moscow, ID, 83843, United States of America
[11] California Institute of Technology, Jet Propulsion Laboratory, Pasadena, CA, 91109, United States of America
[12] Forest Inventory and Analysis Program, Rocky Mountain Research Station, Retired, USDA Forest Service, Ogden, UT, United States of America
[13] Computational & Information Systems Lab, Data Assimilation Research Section, National Center for Atmospheric Research, Boulder, CO, 80305, United States of America
[14] Appalachian Laboratory, University of Maryland Center for Environmental Science, Frostburg, MD, 21532-2307, United States of America
[15] NASA Goddard Space Flight Center, Greenbelt, MD, 20771, United States of America
[16] Department of Plant and Soil Science, University of Delaware, Newark, DE, 19716, United States of America
* Author to whom any correspondence should be addressed.

E-mail: robert.kennedy@oregonstate.edu

Keywords: uncertainty, Carbon Monitoring System, carbon monitoring

Supplementary material for this article is available online

Abstract

Navigating uncertainty is a critical challenge in all fields of science, especially when translating knowledge into real-world policies or management decisions. However, the wide variance in concepts and definitions of uncertainty across scientific fields hinders effective communication. As a microcosm of diverse fields within Earth Science, NASA's Carbon Monitoring System (CMS) provides a useful crucible in which to identify cross-cutting concepts of uncertainty. The CMS convened the Uncertainty Working Group (UWG), a group of specialists across disciplines, to evaluate and synthesize efforts to characterize uncertainty in CMS projects. This paper represents efforts by the UWG to build a heuristic framework designed to evaluate data products and communicate uncertainty to both scientific and non-scientific end users. We consider four pillars of uncertainty: origins, severity, stochasticity versus incomplete knowledge, and spatial and temporal autocorrelation. Using a common vocabulary and a generalized workflow, the framework introduces a graphical heuristic accompanied by a narrative, exemplified through contrasting case studies. Envisioned as a versatile tool, this framework provides clarity in reporting uncertainty, guiding users and tempering expectations. Beyond CMS, it stands as a simple yet powerful means to communicate uncertainty across diverse scientific communities.

1. Introduction

As society gains greater appreciation of the impacts of human activities on ecosystems and climate (IPCC 2023), it is imperative that the scientific community characterize the knowns and unknowns in our understanding of Earth's systems. To build that understanding, scientists utilize a wide-range of observational systems (e.g. satellites, surface measurements, experimental manipulations) and models both to monitor human activities and to gain insight into the ecological, physical and biogeochemical processes they affect.

Despite remarkable scientific progress in Earth system science, imperfections always exist in both the observations and models upon which our understanding is based. No observation (measurement) is made without error (ISO/IEC 2008), and models are simplified abstractions of complex physical and biological systems, making their outputs inherently uncertain (Walker *et al* 2003, Harmon *et al* 2015, Fisher and Koven 2020, Blyth *et al* 2021). As a result, scientists studying Earth systems must account for modeling and observational uncertainties to appropriately attribute a level of confidence in their results. These scientific uncertainties must be then framed within broader forms of uncertainty to form the basis for decisions with real-world consequences (Kwakkel *et al* 2010, Tak *et al* 2015, Gaudard and Romerio 2020). Finally, these considerations are not unique to Earth system science, but are shared challenges across a wide range of disciplines whenever uncertainties in data, models, or data-model fusion products are used, and are especially relevant to any physical, biological, or social-science discipline wrestling with spatial and/or temporal variability (Cassenti and Kaplan 2021).

As a microcosm of the diversity of research in Earth system science and beyond, the NASA Carbon Monitoring System (CMS) program (Hurtt *et al* 2022) regularly grapples with the practical challenges of characterizing uncertainty. CMS projects are selected competitively from PI-driven proposals (rather than dictated by NASA), and although anchored in the study of carbon, this grassroots element of the sponsored CMS research results in a diversity of projects across different Earth System Science subdomains (e.g. land, atmosphere, ocean). As a NASA-sponsored activity, many CMS projects involve data from Earth-observing satellites, but most also incorporate independent observations or models. Most projects involve characterization of carbon-cycle-relevant processes or states in specific places and times, often represented as maps, but do so at spatial scales ranging from the site to the globe and at temporal scales spanning minutes to centuries (Hurtt *et al* 2022). Despite this variety, all projects must include explicit plans to ***quantify uncertainty in any data, model, fused model-data products, or findings***. Because of this explicit requirement, CMS has helped speed the maturation of tools and approaches to quantify uncertainty, particularly for spatially-explicit carbon-related products (Dokoohaki *et al* 2021, Hurtt *et al* 2022).

Recognizing the potentially broad relevance of these advances, the CMS Uncertainty Working Group (UWG) was formed to review and synthesize approaches for handling uncertainty. The group was composed of representatives from the broad spectrum of sub-fields within CMS, including statisticians, remote sensing scientists, ecosystem modelers, atmospheric scientists, data scientists, and others. Through regular meetings, the group evaluated and assessed representations of uncertainty across CMS projects associated with group members. Articulating uncertainty concepts across disciplinary boundaries proved challenging, however. Although all CMS data products must represent some form of uncertainty, rarely is that estimate of uncertainty complete: different subfields have different expectations about what forms of uncertainty are tractable to represent, and even within a subfield, different projects include or omit possible sources of uncertainty. Evaluating these across projects was made even more challenging by the variability in terminology used across and within fields, as has been found in other contexts (Fischhoff and Davis 2014, Bevan 2022). It became clear that any attempt to review uncertainty across the diversity of CMS projects would first require development of a common lens: a conceptual framework through which discipline-specific approaches could be evaluated.

Development of such a framework would also address two other components specific to CMS, but also more broadly relevant in the discussion of uncertainty. First, with substantial collaboration across projects within CMS, outputs from one project are frequently used as inputs by other projects. Because these projects frequently span disciplines, differences in representation of uncertainty can lead to challenges in appropriate use. This is similar to the challenges faced by any user intending to make use of uncertainty information from an external source. Second, CMS requires that projects work with relevant stakeholders as they develop products (Hurtt *et al* 2022). Communication of uncertainty is critical for decision-making (Brown *et al* 2020, Meddens *et al* 2022), and any insights derived from CMS projects could be broadly relevant to other uses of uncertainty products to guide decisions.

Communicating and interpreting uncertainty are recognized as central goals in science, but concepts of uncertainty vary widely (Kwakkel *et al* 2010). In a general sense, uncertainty can be viewed as 'any deviation from the unachievable ideal of completely deterministic knowledge of the relevant system' (Walker et al 2003). The term *uncertainty* is itself polysemous, however, with interpretations and definitions evolving over the course of its emergence in the scientific literature (Bevan 2022). Many frameworks for characterizing uncertainty exist, largely emerging from policy analysis, decision support, and integrated modeling (*Ibid.*). Three broad themes emerge: uncertainty can arise at multiple points in the production of knowledge; it has different levels or severities; and it includes both uncertainty inherent in the stochasticity of a system (aleatoric uncertainty) as well as uncertainty that arises from incomplete knowledge or representation of the system (epistemic uncertainty). Despite robust scholarship on the topic, distinctions between and among various flavors of uncertainty are challenging to characterize (Dankers and Kundzewicz 2020), and adoption of consistent approaches has remained elusive (Kwakkel *et al* 2010, Bevan 2022).

To develop a conceptual framework of uncertainty within CMS, we can find inspiration in these existing frameworks, but the specifics of CMS projects provide additional challenges. A primary distinction emerges from the core requirement in the CMS program that uncertainty be quantified for every product. In essence, CMS projects must make claims about the severity of uncertainty, one of the three core pillars of uncertainty, but little guidance is provided about the other two: how to characterize where in the production of knowledge that uncertainty emerges, and whether that claim of uncertainty includes or omits important knowledge of the system. Additionally, when products developed within one CMS project serve as inputs to other CMS projects, the downstream projects must make informed decisions about how to incorporate the uncertainty into their own calculation of uncertainty. A key component of this process is determining how to appropriately aggregate or disaggregate uncertainty in space or time. Finally, any effort to capture and communicate uncertainty must be generalized enough to allow each subfield to map its own concepts of uncertainty to it.

This paper represents a multiyear effort of the CMS UWG to build a conceptual framework to characterize and communicate uncertainty in quantified data products. Due to the central role of stakeholder interactions in CMS, we initiate an exploration into the relevance of uncertainty representations and characterizations to stakeholders and end users. We aim to discern whether and how these uncertainties should be communicated. Building on existing scholarship, we then articulate a conceptual framework and a descriptive heuristic, intending to enhance the communication of uncertainty both within and beyond the carbon science community. Finally, we illustrate the practical application of our framework by providing two in-depth examples and several brief ones, offering insights into reviewing uncertainty representation in projects affiliated with CMS and beyond.

2. Stakeholder feedback

Within CMS, the roles of data users and stakeholders are essential for improving product relevance and applicability. To comprehensively grasp the perception and utilization of uncertainty information among data users, we surveyed both external and internal users about uncertainty. Forty individuals responded, with nine self-identifying as end users of the information, 25 as researchers or scientists who used CMS data in their work, and six as either policymakers, private sector users, or consultants. Most respondents self-identified as being in either the government or academic sector, with 15% identifying as non-government or private sector. Here, we summarize the high-level themes.

Examining the overarching themes, a significant majority of surveyed users consider uncertainty to be an important component of the data they use (figure 1(a)). Roughly one-third use a summary of reported uncertainty in their own reporting, more than one-third use uncertainty directly in their analysis, and about a fifth formally use uncertainty in error propagation. Second, it was clear that communication of uncertainty could be better: one-fifth of respondents said that it was unclear how uncertainty information was generated, that it was difficult to interpret or use, or that they could not see the value added in the uncertainty information. Finally, of the respondents who used uncertainty (figure 1(b)), the majority require the information at pixel (i.e., grid cell) scale, but many also require the information in aggregated form (e.g., state or county). Thus, it is important that uncertainty be represented in a way that allows both forms of representation for a diverse group of users.

In a parallel CMS-sponsored study, Meddens *et al* (2022) surveyed 69 attendees of an Operational Lidar Inventory (OLI) meeting held in Olympia, WA in March 2020. Unlike the respondents to the CMS stakeholder survey, more OLI meeting attendees were from the forest industry or consulting sectors (41%) than from academia (25%), state/tribal agencies (20%) or federal agencies (14%). These natural resource professionals were queried on their preferences in geospatial data products used to monitor and manage forests across large landscapes. Concerning uncertainty, the majority of participants wanted estimates to be 'very precise' (RMSE <25%), while a minority would settle for 'precise'.

Although these were informal qualitative surveys, both studies illustrate that consumers of carbon-related products consider uncertainty important and

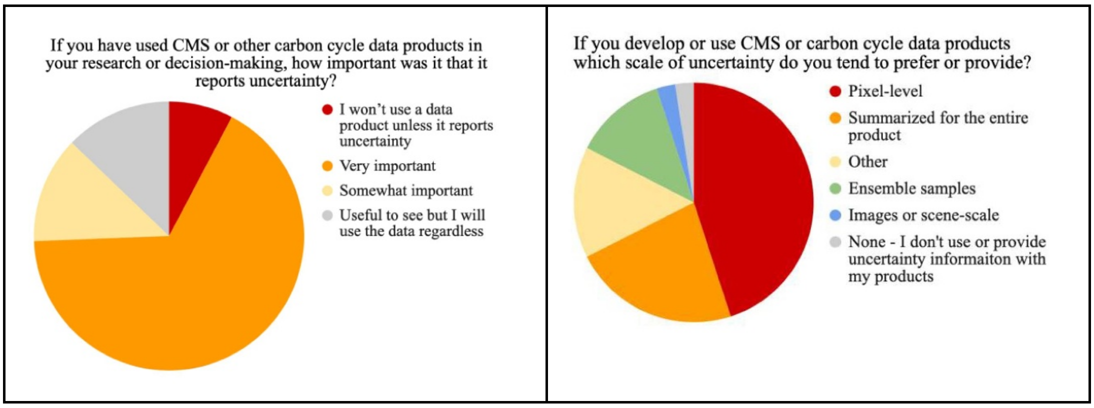

Figure 1. High-level summary of respondents to a NASA CMS stakeholder survey on uncertainty. (a) Respondent answers when asked how important uncertainty information was to them when they evaluate whether to use a data product. In a subsequent question, 66% of respondents said that uncertainty information was very important in their decision to use the data, whereas just10% responded that it had only minimal influence. (b) Respondent answers when asked which scale of uncertainty was preferred. Other answers included combining these scales—both pixel level and summary information or extracting the pixel level data into a polygon or other application-specific area. Respondents also emphasized that uncertainty reporting depends on the application and user needs.

useful. Recognizing the wide variations in the definitions and notions of uncertainty across disciplines (see also Bevan 2022), effective communication necessitates identifying core principles that underlie these concepts. Understanding the imperative nature of uncertainty information in designing data products becomes crucial for influencing decision-making processes.

3. Terminology and conceptual basis

As with any effort focused on uncertainty, the initial challenge lies in establishing a common conceptual basis and reconciling terminology (Bevan 2022). This is particularly challenging within the broad scope of CMS research projects that range from the statistical estimation of total biomass in a single forested ecosystem at a single point in time to physical modeling of global patterns of atmospheric CO_2 flux at sub-daily time-steps. To encourage interdisciplinary thinking, we have endeavored to use concepts and terms in their broad rather than specific sense, and provide a means to link these back to discipline-specific terms (table 1). We urge readers to relax strict sub-disciplinary definitions in favor of finding the conceptual similarities across disciplines.

We begin with a research goal common to nearly all CMS projects: characterizing some aspect of carbon in a specific place and time, often in the form of a geographic map. This structure is driven by both the applied nature of the program and the required involvement of stakeholders to use CMS products. It sets our goal apart from scientific efforts to discover physical laws or biological processes through experimentation or lab manipulation. In essence, our focus is on making claims about carbon-related quantities—for example, carbon density or carbon flux—in real-world situations, for which the true value is likely unsampled and therefore unknown for a given location or point in time.

We refer to these quantitative claims in specific places and times as 'predictions.' Although the term prediction has specific connotations in statistics and prognostic modeling, here we emphasize the generic notion of *making* ***a quantitative claim*** *in any situation where the true value is not known*. For example, a claim about carbon in an unmeasured spatial location is conceptually analogous to making a claim about carbon in an as-yet-unmeasurable future time. Similarly, a statistical estimate derived from samples can be broadly considered a prediction of what the true population parameter might be, and is relevant for a specific, bounded place and time. Broadly conceived, the core of most of our research projects can be conceptualized as a 'prediction process' whose goal is to produce claims about carbon-related quantities within defined spatial or temporal extent.

As a common conceptual basis, we posit a flexible and generic workflow to capture the prediction process (figure 2). While many projects feature more complex and interconnected workflows, our goal is to interpret the components of our simplified workflow broadly enough to encompass a wide range of projects within the generalized structure. Three core components of the workflow serve as pillars: models, predictors, and observations. We first describe and define these components (see table 1), and then describe where uncertainty enters into the 'prediction process'.

At the heart of the process is a model that translates predictor variables and/or observations into a prediction at a specific geographic location and time. The term 'model' implies specific forms in certain sub-disciplines, but here we adopt its broadest definition as a representation of how the world works

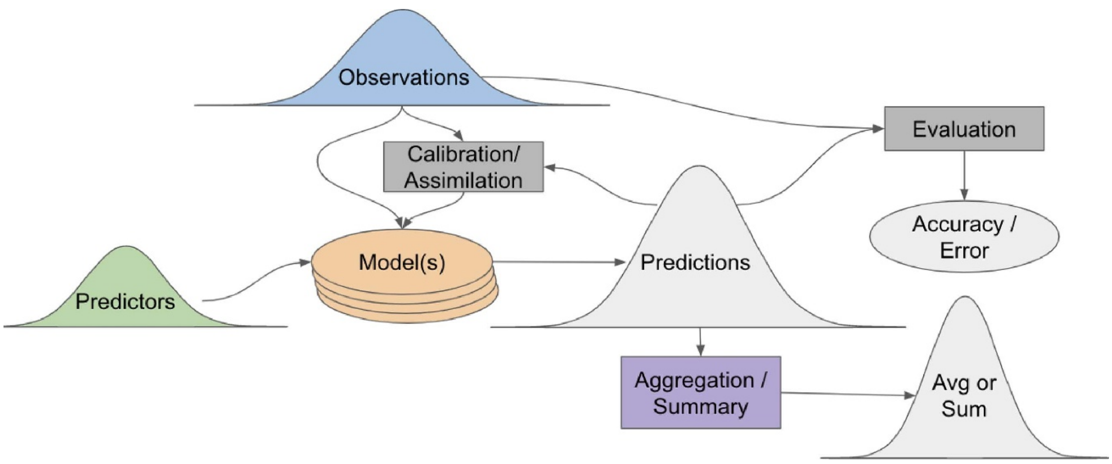

Figure 2. A generalized workflow for carbon prediction. At the core of the workflow is a model—either conceptual or actual—that translates predictors and observations (which could have a spread of possible values, as suggested by the normal curve symbols) into predictions for a given carbon-related attribute. These predictions can be evaluated directly for error, can be aggregated into larger spatial or temporal units and summarized, or fed back through a data assimilation framework to adjust the model. A specific project may only require a subset of the flows shown here, and some projects may require layered or sequential sets of these flows. The definitions of each category are framed from the perspective of the study or project of interest, recognizing that predictors and observations are often themselves the result of a separate workflow, which itself may depend on other prior workflows, etc.

Table 1. Generalized terminology used in this paper and its relationship to related terms found in common parlance or in discipline-specific cases.

Term used in this framework	Definition	Alternative or discipline-specific related terms
Carbon-related attribute	True carbon state or flux characteristics of a real-world system	Atmospheric carbon mixing ratio, terrestrial forest carbon, marine dissolved carbon, etc.
Observation	A measurement of the carbon-related attribute of interest for a specific place and time	Measurement, data
Prediction	A quantitative claim about a carbon-related attribute for a specific place and time	Estimate, output, forecast, hindcast
Predictor	A characteristic of the real-world system that is thought to help predict the carbon-related attribute of interest	Ancillary variable, explanatory variable, driver variable, feature
Model	A conceptual, statistical, or mechanistic tool to translate predictors and/or observations into predictions	Linear regression, terrestrial ecosystem models, atmospheric transport models, machine learning
Error	The difference between the true value of a carbon-related attribute and either the prediction or observation	Standard deviation, variability, bias, root mean square error
Uncertainty	A characterization of the spread or distribution of potential values for the carbon-related attribute	Standard error, variance. Inverse or opposite of: accuracy, confidence, precision.

(table 1). The specific manifestation of a model varies widely by sub-discipline, and may be conceptual, empirical, or mechanistic. For example, in survey sampling of forest carbon, mathematical models are often held as distinct from the sample-based approach used to describe characteristics of forest stands. However, we argue that even that situation contains an implicit model: an abstraction that guides the choice of statistical tool to represent the central tendency of a group of measurements. A model may also be an empirical regression or machine learning model calibrated by linking satellite imagery to ground-based measurements, or it may be a physical or mechanistic model of an ecosystem or of the atmosphere whose structure has been inferred from the literature or from physical laws. In all cases, however, we conceptualize a model as *a tool that translates knowledge* about certain characteristics of a specific place and time into a prediction for that place and time of a desired quantity, which in our case is carbon.

Knowledge about the characteristics of a specific place and time is captured in two ways: through observations and through predictors. Again, we adopt a generalized definition of 'observation': a measurement of our desired quantity (e.g. carbon pool, flux, etc), bounded in space and time. In essence, an observation is often a snapshot of the real world that provides information about the true state of the system. Conversely, 'predictors' serve as descriptors of the real world (e.g. temperature, tree height), different from observations in that they are *not the quantity of interest,* but rather are *related quantities* that we believe will enhance our ability to make predictions of that quantity of interest. When the goal of a carbon prediction process is to create a map, these predictors are often spatially explicit and spatially exhaustive: satellite imagery, maps of soil properties or topography, weather fields, etc.

Once predictions are formulated, they typically follow one of three trajectories. First, predictions can be compared with independent observations or with other model predictions to characterize the overall prediction accuracy (a.k.a. model verification or benchmarking). Second, predictions may be summarized or aggregated for end users, such as policymakers, or used as inputs into other modeling frameworks. Finally, in some modeling structures, the predictions from one cycle of modeling can be considered inputs in data assimilation (DA) approaches that harmonize observations and predictions.

Uncertainty arises because the true nature of all of these components is not known. Perhaps most fundamentally, the true value of the desired carbon-related quantity cannot be known exactly. Observations attempt to capture the true value, but as with any measurement, observations include error. Here, we consider 'error' to be the difference between the reported observation and the unknowable true value. Because of random and systematic effects, the observed value could be drawn from any number of reasonable—but slightly erroneous—values that are near the true value. Uncertainty characterization is a best effort to put constraints on the potential 'wrongness' of the reported observation. Thus, observation uncertainty can be conceptualized as a descriptor of the range or probability distribution of possible error for an observation. We indicate this graphically as a distribution of possible values (figure 2). *If all random and systematic effects are well understood and characterized, then the true, unknowable value will fall within the range of uncertainty.*

This concept extends beyond observations to other components. Predictors have uncertainties (for example, temperature observations or satellite imagery) or may themselves be the result of a prediction process from a different discipline (interpolated weather or climate data). Similarly, models may be inappropriately or incompletely specified for the system of interest. Taken together, the uncertainties in predictors, models, and observations result in uncertainty in predictions: prediction uncertainty. Prediction uncertainty is rarely the same for every prediction across a spatial and temporal space, and thus the prediction uncertainty is itself a spatially and temporally-varying field. While researchers strive to capture as many potential sources of uncertainty as possible, in practice doing so may not be tractable. For example, in the process of constructing a model to simulate the real world, it may be necessary to omit certain components to make the model computationally or conceptually manageable. In other cases, there are competing theories on how to best model a system, or which models apply best under different circumstances. Sometimes the model is thought to be well understood and fully described, but necessary predictors are known to be imprecise. In other cases, the observations are known to be imprecise or incomplete. Typically, the researcher engaged in the prediction process knows about these sources of uncertainty, and must make choices about which sources of uncertainty to include in their prediction uncertainty, but many problems also face unquantifiable 'unknown unknowns'.

Therefore, when faced with a prediction uncertainty product, users must discern which types of model, observation, and predictor uncertainty have been considered—and which types have not. At a minimum, this helps manage expectations. It is even more important when multiple products exist for ostensibly the same quantity, each with their own uncertainty fields—especially if the range of uncertainty among different products does not overlap. If the range of uncertainty is meant to indicate where the true value may lie, then non-convergence may suggest to a user that the entire enterprise is wrong. However, if attention is called to the fact that prediction uncertainty can incorporate different sources of uncertainty, then such disagreement can be understood as a case where the two different 'flavors' of uncertainty are incomplete or possibly wrong, and where further work is needed to incorporate more sources of uncertainty. Finally, the description of sources of uncertainty is critical when the prediction from one research endeavor becomes an input (either predictor or observation) in a subsequent research domain. For the downstream user to appropriately weigh or incorporate the output from the upstream project, there must be a frank assessment of which sources of uncertainty have been considered and which have not. Efficient and clear communication of uncertainty from producer to user is essential in all of these scenarios.

4. Heuristic framework

With these components in mind, we propose a generalized heuristic framework to accompany any product's quantified uncertainty (figure 3). This framework aims to facilitate producers in describing not only the quantification but also the entry points of

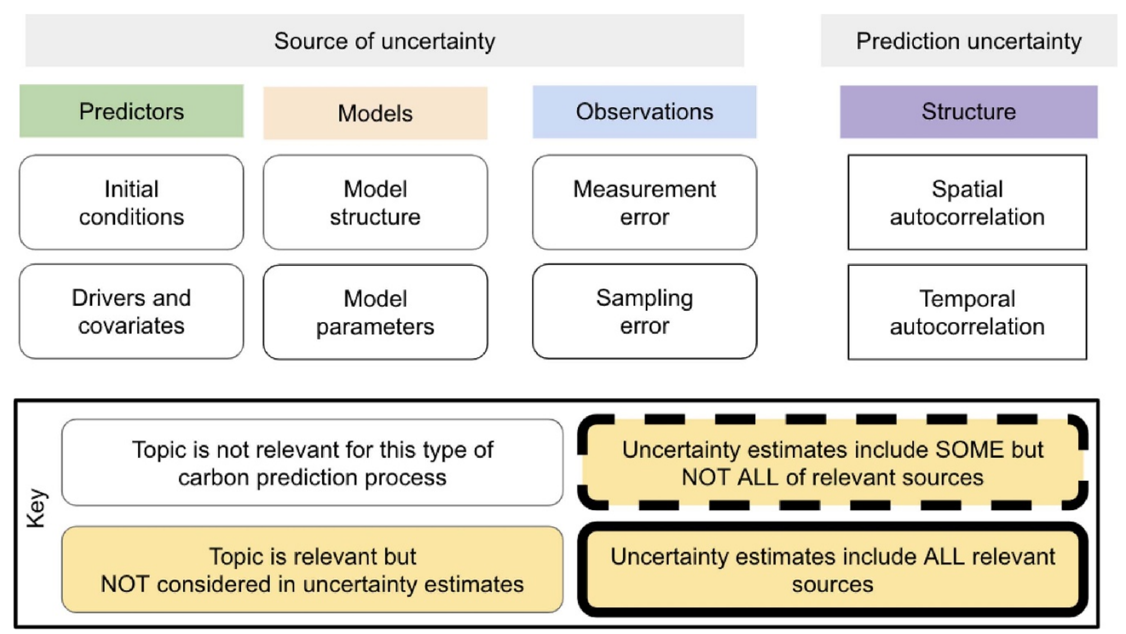

Figure 3. A heuristic framework to communicate uncertainty. When a research team develops a set of predictions along with their associated prediction uncertainties, they would also use symbolization to quickly impart which types of uncertainty are not relevant for their domain, which ones matter but cannot be accounted for in their product, and which ones have been given some level of treatment. Details of the categories are given in the main text.

uncertainty into the process and the sources that are considered or omitted in that quantification. Building on our structure above and borrowing conceptually from prior frameworks (Alcamo and Bartnicki 1987, Walker *et al* 2003, Harmon *et al* 2015, Bevan 2022, Blackhurst and Matthews 2022) we consider three broad categories where uncertainty can enter: observations, predictors, and models, each with subcategories. Acknowledging that most CMS products include space and time components, we highlight an important fourth component of uncertainty: the spatial and temporal structure of the prediction uncertainty fields that result from those source uncertainties. For each source, producers use graphical symbolization to impart which sources of uncertainty are relevant to the problem and the degree to which they are included in the quantification. We elaborate on these below.

Predictors (from figure 2) are split into two subcomponents: 'drivers and covariates' and 'initial conditions.' Covariates are a broad category of proxy variables that will be translated into predictions through the model based on statistical correlations or mechanistic relationships. Common examples are satellite imagery, soil properties, land cover, and political/-management units. Drivers are dynamic covariates, such as meteorology, that propel the model through time. We distinguish all of these from initial conditions, which are limited to the initial estimate of a quantity of interest that is later modeled dynamically to change through time. A common example would be initial estimates of carbon pools at the Earth's surface used in land models. Predictors may derive uncertainty both from measurement error, as with observations, and from the fact that they themselves are often the result of a separate prediction exercise. For example, a model may require gridded temperature as one of its drivers, and that gridded temperature surface is itself derived by a separate research team through a prediction process applied to weather station measurements. From the perspective of, for example, an atmospheric inversion exercise (see section 5.2), the prediction uncertainty of the weather variables (an output of a weather prediction exercise) becomes predictor uncertainty (an input to the atmospheric inversion exercise).

The translation of predictors into predicted variables is controlled by the structure of the model and by the parameters that control the model, both of which have uncertainties (Alcamo and Bartnicki 1987). Model structure uncertainty appears in both the empirical and the mechanistic realms. In the empirical modeling arena, the choice of parametric vs non-parametric statistics is critical, as well as the specific structure of the model within either domain. In the mechanistic modeling realm, the structure of the model can be driven by physics and/or biology, but often choices remain about which mechanisms are incorporated (or not), or about the scales at which those mechanisms are represented. In all of these cases, it is often impossible to know *a priori* which model structure is appropriate, or whether the appropriateness varies over space, time, or situation, and thus how to represent uncertainties in model choice.

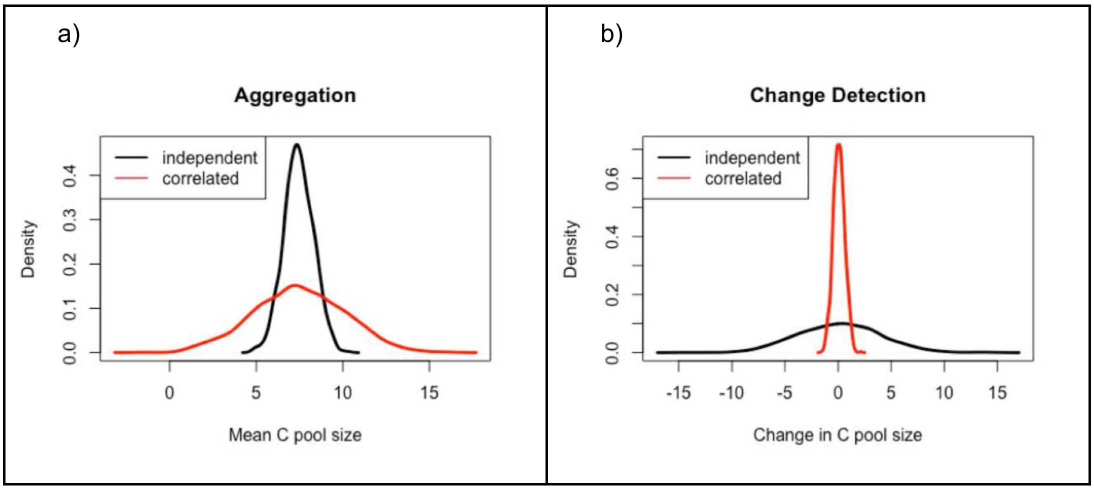

Figure 4. Simulated examples of the impact of ignoring autocorrelation in the spatial and temporal domain of a grid-cell product over time. (a) In a simple simulation of mean carbon pool size in a gridded landscape, if uncertainties at grid cells are assumed to be spatially-independent, then aggregation to a larger cell will result in an overly optimistic (narrow) estimate of the uncertainty at the coarser resolution compared to the case where spatial correlation is considered appropriately. (b) If the goal is to estimate uncertainties in the difference between two estimates over time, then assuming those estimates are temporally independent will result in unrealistically broad (pessimistic) uncertainty distributions.

Regardless of model structure, the model parameters that control the behavior of the model have uncertainty. When derived through calibration, they are affected by the choice of calibration method. When derived from prior studies or literature or even some form of direct measurement, it is uncertain whether a given set of parameters is applicable across the full domain of the model regime. From the perspective of incorporating uncertainty into model predictions, a study that perturbs parameters in a single model to vary predictions would only be considering parameter uncertainty, while an ensemble modeling approach that uses distinct mechanistic models, each calibrated with a single set of parameters, would only be considering model structure uncertainty. Note that stand-alone model sensitivity analyses are not strictly relevant to this heuristic, as their goal is not to develop predictions but rather to explore model structure. That said, model sensitivities do serve as inputs into certain uncertainty propagation methods (LeBauer *et al* 2013, Dietze 2017).

The assimilation of observations provides a means of iteratively updating key components of a model (e.g. parameters, model state) in an effort to reduce model prediction uncertainty. The extent to which the prediction uncertainty is reduced depends upon the certainty of the observations, and is ultimately limited by the predictor and model uncertainty. An example of DA application adapted to the heuristic is provided in section 5.2.

When considering observations, we distinguish between uncertainty arising from the measurements and that arising from the placement of observations (the sample design). Measurement error occurs for all measurements, of course, and can be exacerbated when attempting to measure carbon-related quantities in the real world, outside of a controlled lab environment. Moreover, uncertainty in the observed value can arise because the spatial and temporal bounds of a given measurement are sometimes not knowable with precision. Sampling error refers to the uncertainty that arises when a subset of a population (a sample) is meant to represent the entire population. The number of observations used, the means of distributing them (random or not), and their spatial and temporal autocorrelation can all be factors that ultimately expand or contract prediction uncertainty.

Finally, because we focus on predictions in a place and time, we must add a fourth pillar of uncertainty: the spatial and temporal structure of the uncertainty field itself. Assuming predictions are arrayed in space and time, the uncertainties in those predictions become additional modeled fields of their own (an 'uncertainty surface'), with concomitant spatial and temporal characteristics. If a given prediction product is spatially or temporally up- or downscaled by the next user, the uncertainty surface must also be scaled. However, calculating the distribution of uncertainty at a coarser scale requires careful consideration of spatial and temporal autocorrelation. If uncertainties are assumed to have no spatial autocorrelation, then offsetting random error across adjacent cells will appear to cancel out, resulting in an overly optimistic (narrow) estimate of uncertainty at the coarser spatial scale (figure 4(a)). This is particularly problematic in DA systems, where such narrow uncertainty bounds would be given proportionally too much weight in a solution, potentially leading to erroneous conclusions. Incorporating an assumption of spatially-correlated uncertainty broadens the estimate to a distribution that is more representative of the actual uncertainty at the coarse resolution. If the interest is estimating change over time, however, the penalties reverse. If two successive estimates (in time or space) of a quantity of interest have positively-correlated uncertainties, then much of the shared uncertainty between two estimates will cancel out when calculating ***the difference*** from one time step to another (figure 4(b)). If these correlations are ignored, then the corresponding uncertainty distributions will be too wide. Thus, in both the spatial and the temporal cases, it is necessary for a producer to specify the correlation structure of the unexplained variation in their predicted spatio-temporal fields.

Our heuristic framework is meant to benefit both users and producers of predictions and data products. We envision it as a core component of metadata accompanying an uncertainty product. For users it serves as a simple visual mechanism to gauge how to use any product, and it emphasizes that 'uncertainty' means different things for different products. Moreover, it can help clarify why predictions from different products may disagree: by explicitly stating which sources of uncertainty are considered and which are not, the heuristic helps temper expectations about the level of agreement among products and reduce their misuse. For producers of predictions and data products, the heuristic forces an honest reckoning of strengths and weaknesses in products. Perhaps most importantly, it can serve as a structural tool to focus efforts at improvement in products by highlighting gaps in the representation of uncertainty. Beyond graphical heuristics, the framework presented here can also inform the development of community conventions for sharing predictions and both data and metadata about their uncertainties (Dietze *et al* 2023).

5. Application

The heuristic framework is envisioned as an organizing structure for two purposes: characterizing projects for evaluation and comparison, and communicating uncertainty to data users. In the former case, projects with similar workflows or product deliverables could be characterized using our framework, and differences in the heuristic could be easily identified to guide understanding of discrepancies among projects. In the latter case, a developer of an individual project would report both the graphical heuristic and an associated description to allow the data user to determine how to best interpret or use the uncertainty.

Both efforts require translation of discipline-specific concepts into the generalized notions of our framework. Using two examples, we illustrate a narrative approach to link project-specific approaches to our heuristic framework. The core goal of these examples is to map the terms in the figure to the specific disciplinary case, and then explain or defend the choices in the heuristic tool: How do the terms translate into the methods of the particular study? Why are some categories of uncertainty considered irrelevant (left blank)? When a category of uncertainty is relevant, but not considered in the study, what are the constraints that prevented it from being considered? If the producer states that a given category is fully or partially considered, what logic or evidence is available to back that claim?

5.1. Uncertainty arising from sample design

Our first example comes from 'design-based' statistical estimation theory as applied to estimation of forest carbon stocks. Under design-based estimation, a sample of observations is measured at locations under a known probability-based sample design and an estimator is used to estimate a population parameter. The sampling strategy (the combination of the sampling design and estimator) allows for calculation of errors in those estimates (Särndal *et al* 1992). Such approaches have long been used for assessing standing timber volume in traditional forestry applications, and this approach remains the standard on which forest carbon accounting must be based in the context of international carbon agreements (e.g. the TREES-2.0 protocol for projects under Reducing Emissions from Deforestation and Degradation: www.artredd.org/trees/).

Here, we apply the uncertainty heuristic framework (figure 5) to describe prediction uncertainty in Menlove and Healey (2020). In this study, the producers use the design-based approach to estimate forest biomass (and hence carbon) for each of more than 12 000 hexagons covering the lower 48 United States; the hexagons are approximately equal-area hexagons of 640 km^2 (Menlove and Healey 2020). For the purposes of this example, we translate the term 'estimation' used in the traditional parlance of design-based theory into the 'prediction' concept used in our framework: the goal of estimation is to *predict* the population-wide statistic of a particular property (for example, mean biomass density), which essentially involves making claims for the many locations for which there are no observations. The confidence intervals associated with design-based estimation can be considered the prediction uncertainty. Observations in this study come from the US Forest Inventory and Analysis (FIA) program. Because that program is built on a rigorous sampling design to enable design-based estimation (Bechtold and Patterson 2005), we can consider the uncertainty from the Sampling Design to be both a relevant source of uncertainty and completely treated ('Sampling error' symbol filled and outlined in figure 5).

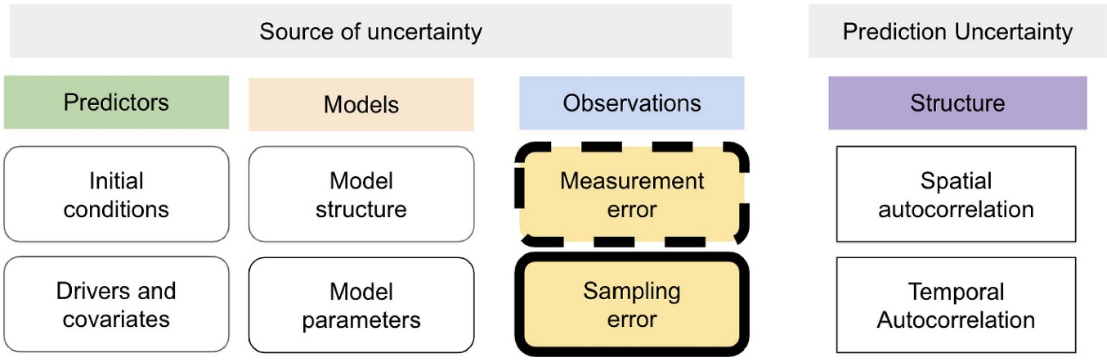

Figure 5. The heuristic framework applied to the study of Menlove and Healey (2020), which derived design-based estimation of forest carbon from Forest Inventory and Analysis (FIA) for large hexagons across the U.S. As with any design-based estimation, sampling error drives the uncertainty, but the design of the study raises several other decisions that could affect uncertainty (see text for details).

It is common in design-based studies to assume the observations are made without error, i.e. there is no measurement error. We are translating the definition of measurement error from observations such as diameter at breast height to observations which are modeled from measured observations; in the latter case measurement error could include the errors in the measured observations and the modeling process. In Menlove and Healey (2020), the observations of 'forest biomass' from the FIA plots are themselves derived from a modeling process that uses functions ('allometric equations') to translate field measurements of tree type and size characteristics into estimates of carbon or biomass. Because Menlove and Healey (2020) provide visual renditions of the impact of the choice of using these different measurement types, we can consider that measurement uncertainty is *partially treated*, but not fully incorporated ('measurement error' filled but outlined with a dashed line).

Spatial structure of the error deserves note. Using the independent samples for the hexagons, Menlove and Healey (2020) separately produce design-based estimates for each hexagon, and display these estimates in a map. Because the estimates are mapped, there may be a tendency to assign some spatial relationship between the estimates. However, in the design-based paradigm the samples are independent and the design-based covariance between the estimates for hexagons is zero, no matter the distance between hexagons. Thus, the 'spatial autocorrelation' symbol is left uncolored.

Many of the other sources of uncertainty are not relevant. There are no predictor variables, and because it is intended for a single point in time, temporal autocorrelation is not relevant. We further claim that model structure and parameters are not relevant, but the argument could be made that even a design-based study includes an implicit conceptual model that may or may not be relevant. This example illustrates the translation of terms in the framework into the specific disciplinary case, providing insights into choices made and considerations taken in addressing uncertainty in the study.

5.2. Uncertainty inferred from predictions and observations

Our second example comes from the DA domain, where disagreement between predictions and observations is used to infer and reduce uncertainties in the modeling process. In DA, a model is used to translate predictor variables and their uncertainties into predicted carbon-relevant values, often for full spatial surfaces. These predictions are then compared against observations, which may be spatially- and temporally-sparse, to refine the predictions at all sites, not solely those with available observations. Specifically, DA typically makes use of process-based models and relies on the covariance structure of the model itself to make inferences about unobserved location, times, and carbon cycle variables. Typically, DA approaches explicitly consider uncertainty in most components of the modeling process. Because DA strategies represent the state of the art in handling uncertainty, they provide an excellent counterpoint to the design-based estimation of error described in the prior example.

We apply the uncertainty framework (figure 6) to Raczka *et al* (2021), a CMS-supported study of terrestrial carbon pools and fluxes in the western United States. In Raczka *et al* (2021), the core model is the Community Land Model (CLM5.0; Lawrence *et al* 2019), which represents terrestrial ecosystems as horizontally and vertically distributed pools of vegetation, soil, and water, and uses physical or biological submodels to represent the transfer of carbon, nitrogen, water, and energy. The uncertainty in atmospheric conditions is used to generate 80 ensemble simulations that quantify the uncertainty in the modeled fluxes and pools. Observations of above-ground biomass and vegetation leaf-area index, derived from satellite remote sensing, are used to update important prognostic variables (i.e. carbon and nitrogen

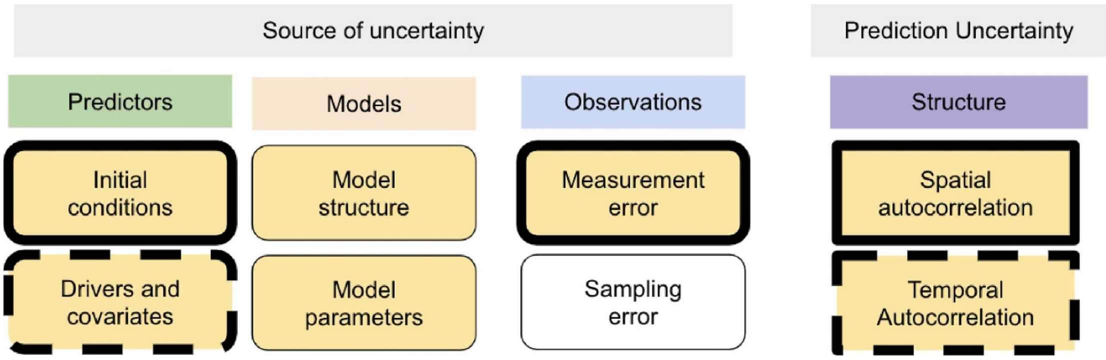

Figure 6. The heuristic framework applied to the terrestrial inversion study of Raczka *et al* (2021). Here, a single instance of a terrestrial biosphere model (TBM) is perturbed by meteorological data (driver) to predict an ensemble of terrestrial states. Then, through an Ensemble Kalman Filter (EnKF), these model states are adjusted by observations (with measurement error). This EnKF approach allows for the spatial and temporal correlation to be accounted for in the adjusted model state.

pools) based on an Ensemble Kalman Filter (EnKF) approach (Anderson *et al* 2009). Both observation uncertainty and initial condition uncertainty are explicitly considered through the EnKF assimilation framework and contribute to the updated uncertainty predictions. The uncertainty from atmospheric drivers influences the land-based uncertainty predictions, but uncertainty from other covariates (such as plant functional type and soil) was not explicitly considered. Both model structure (the core CLM model) and its controlling parameters were not varied nor contributed to the predicted uncertainty. In order to compensate for these missing sources of uncertainty, an adaptive inflation technique (Anderson 2007) is used to empirically enhance the model uncertainty. Spatial and temporal autocorrelation in the uncertainty of the output layers are considered through the EnKF adjustment. Because both data constraints were spatially exhaustive, sampling error was not considered relevant for this work. This example highlights the explicit consideration of various sources of uncertainty in a DA study and the methods employed to address them.

5.3. Other examples

The lens of the heuristic framework can be used to evaluate or compare other projects. We omit the graphical depiction and the full narrative in the interest of brevity and highlight key characteristics of uncertainty that would be illuminated by the heuristic. In the domain of forest carbon estimation, for example, Patterson *et al* (2019) apply a 'hybrid estimation' approach to forest biomass estimation that blends design and model-based considerations to improve consideration of uncertainty in calculating biomass from the GEDI (Dubayah *et al* 2020) lidar sensor. The sources of uncertainty considered are the sampling error of modeled predictions of biomass at the sampled locations and the modeling error of the biomass predictions, along with autocorrelation of the residual errors. In a CMS study in the western United States, Kennedy *et al* 2018 linked field measurements with remotely-sensed data to predict forest biomass as a continuous field for many years. Although several versions of driver and parameter values were used to estimate uncertainty, many other model structures or drivers could have been used; moreover, spatial and temporal autocorrelation were not considered at all, though these would affect the structure of the uncertainty fields. Also in the western U.S., the CMS study by Hudak *et al* (2020) used a two-stage modeling approach: the first to link field data to airborne-lidar estimates of biomass, then linking those with satellite-based estimates, and quantified uncertainty at both levels using random forest non-parametric models. The uncertainty layer associated with the lidar-based biomass map was not explicitly propagated to the satellite-based biomass map; rather, the uncertainties associated with predictions from both were reported as separate CMS products. In both of these studies, neither measurement nor allometric errors were explicitly included in estimates of uncertainty. Saatchi *et al* (2011) included measurement error in a suite of uncertainty sources propagated through a modeling process as they calibrated nonparametric maximum entropy models using satellite-derived optical and radar image data against biomass observations. As with the other studies above, model structure was not varied, and as with all prior examples, spatial autocorrelation in uncertainty was not considered. Spatial autocorrelation was explicitly modeled in Babcock *et al* (2016), where hierarchical Bayesian strategies were used to explicitly solve for spatial dependence of uncertainty terms, while propagating uncertainties from plots and models in a small project area involving field, lidar, and satellite-based measurements.

The uncertainty framework also illuminates how DA approaches in both atmospheric and terrestrial modeling continue to advance their representation of uncertainty. For example, in another CMS project, Dokoohaki *et al* (2021) use terrestrial DA strategies to develop a proof-of-concept for a national-scale carbon cycle reanalysis product, incorporating uncertainties in the parameters and initial conditions driving their terrestrial biomass model, and explicitly modeling the spatial and temporal structure of prediction uncertainty—effectively covering most aspects of the heuristic framework. In the realm of atmospheric modeling, the advent of satellite-based CO_2 sounding instruments has provided new options for improved sampling of the global atmosphere, and these can then be used in a DA framework to improve overall uncertainty (Crowell *et al* 2019, Peiro *et al* 2022, Byrne *et al* 2023), again leading to improved representation of spatial structure of uncertainty that would be captured in the heuristic. All DA methods depend highly on the uncertainty structures of the model and observations, and biases in these (particularly observations) are not easily handled (Cameron *et al* 2022). This potential source of bias in observations could be represented graphically as an incomplete consideration of observation uncertainty (i.e. a dashed line), accompanied by further description in the narrative. Therefore, in the case of atmospheric inversion, some studies have evaluated uncertainties by comparing optimized fields to independent observations not used in the original assimilation (Liu *et al* 2021). In this case, the graphical heuristic would be similar to other studies, but an accompanying narrative or metadata would illuminate the contrasts in approach.

6. Discussion and recommendations

Characterization of uncertainty is a central goal in any scientific endeavor, but the definitions of uncertainty differ across scientific disciplines (Bevan 2022). The NASA CMS program incorporates a broad diversity of scientific fields and requires all products to estimate uncertainty, providing a unique opportunity to evaluate how uncertainty is handled across fields. This paper represents an effort by the NASA CMS UWG to build a conceptual lens through which projects can be evaluated and uncertainty can be communicated across projects and to external stakeholders. Our goals were to find commonality of concepts across disciplines, to develop a tool for quickly summarizing and communicating information about the uncertainties in different data products, and to acknowledge the particular need to consider both spatial and temporal properties of uncertainty.

The resulting heuristic framework borrows from the substantial body of scholarship characterizing and communicating uncertainty (Walker *et al* 2003, Kwakkel *et al* 2010, Dankers and Kundzewicz 2020, Gaudard and Romerio 2020, Bevan 2022, Blackhurst and Matthews 2022), but adapts and augments those concepts to the conditions represented in CMS. Of the three common pillars frequently recognized in uncertainty (source of uncertainty, type [stochastic, epistemic], and severity), the CMS requirement to quantify uncertainty addressed only one: severity. But the CMS UWG recognized that quantification of uncertainty was insufficient by itself. Uncertainty mismatches between ostensibly similar products and difficulty propagating uncertainty among projects exposed fundamental gaps in our representation of the concept. This required a mechanism to communicate sources of uncertainty, and the degree to which some sources of uncertainty are not represented in a given product. Moreover, we recognized that predicting the when and where of uncertainty also required addition of a fourth component: the spatial and temporal characteristics of uncertainty. Without consideration of all of these topics, users both inside and outside the scientific community would have difficulty making sense of our products.

Our heuristic framework complements important benchmarking and intercomparison projects that have already been used in carbon cycle science to quantify uncertainty. For example, the ILAMB project (Collier *et al* 2018) assesses model performance of Earth system model land representations against a suite of observational data, providing insight into model performance. We believe a heuristic like the one developed here would be highly valuable for such benchmarking projects: when considering what data products to use as benchmarks, in communicating the uncertainties in those product uncertainties to users, and when selecting skill scores to use when evaluating models (e.g. the scoring of pixels should not be treated as independent if the uncertainties in the benchmark data are highly autocorrelated). Relatedly, process-based models are frequently used to predict earth system processes (including the carbon cycle) into the future, either under the status quo or alternative emissions scenarios. Model intercomparison projects (MIPs) are also often used to evaluate across modeled predictions, with the spread in model results providing a measure of uncertainty. For example, the Coupled Climate–Carbon Cycle Model Intercomparison Project (Jones *et al* 2016) has been used to assess uncertainty in the carbon cycle component of climate models, while the OCO-2 Flux Inversion MIP (Byrne *et al* 2023) has been used to evaluate carbon sources and sinks derived 'top-down' from atmospheric CO_2 measurements, including those from satellites. The heuristic developed here is equally valuable when discussing the uncertainties associated with the projections from individual models or MIPS. While others have recognized the challenges of interpreting the uncertainty of a multi-model ensemble, as such models are a non-random and non-independent sample of possible models (Tebaldi and Knutti 2007, Sanderson and Knutti 2012), less attention is often given to a full -accounting of the uncertainties within individual models, and the 'spin-up' protocols of many MIPs often result in complex within-model convolutions of uncertainties (e.g., parameters, initial conditions, drivers, and process error) (Raiho *et al* 2020). In all of these cases, our framework is not focused on attempting to score or rank the relative impact of different sources of uncertainty, but rather to provide a means of describing which broad categories are considered, with associated narrative to back up the choices made. In this way, our work complements within-discipline efforts to explore quantitatively the relative impacts of different sources of uncertainty (e.g. Bonan and Doney 2018, Duarte *et al* 2021).

More broadly, the explosion of data production in all domains of science and commerce necessitates a concomitant attempt to clearly and honestly reckon with the uncertainties in those data. By encompassing such a wide range of predictive frameworks, we believe the heuristic framework we have proposed can be used directly or adapted across a wide range of fields and paradigms.

An important feature of the heuristic is to highlight what types of uncertainty are *not* considered in a given uncertainty product. This distinguishes our framework from other efforts to organize sources of uncertainty in carbon-related projects (e.g. Blackhurst and Matthews 2022). It leans on the assumption that the scientists who develop predictions know best the degree of belief they have in their own understanding (Gaudard and Romerio 2020). This includes both the factors that are unknowable (true epistemic uncertainty) and the known stochastic factors that could introduce uncertainties into their products. Of these, they know which of those are not considered or not tractable when building predictions of uncertainty. The heuristic tool can provide three benefits in this domain. First, it can clarify why uncertainty estimates (or the predictions themselves) may vary from one data product to the next. As products with ostensibly similar goals proliferate, it becomes critical to provide users a means of differentiating them, and, importantly, to dampen expectations among users that all products must agree. Second, it provides a venue for producers to speculate about unknown unknowns—places where our knowledge of the system is truly incomplete. And finally, it can encourage producers of predictions to critically self-evaluate where and why different sources of uncertainty are not included, and indeed help point to the areas where improvements can be made.

A key area where frank assessment is needed is in possible sources of bias. If an estimate of uncertainty is attempted by comparing predictions with observations, any bias in the acquisition of those observations could lead to misleading conclusions about the uncertainty in the system. For example, a calibration-based modeling strategy can be compared to observations to provide some insight into uncertainty in the model structure or parameterization, but if the observations are sampled in a biased manner, the representativeness of the uncertainty will be incorrect. Similarly, atmospheric inversion techniques (or estimation theory more broadly) generally assume unbiased errors in observations (Baker *et al* 2006), and if this is not the case, then the assimilation of uncertainty will be incorrect. Such observation or sampling bias can be highlighted graphically in the sampling and measurement portions of the graphical heuristic, and elaborated in the accompanying narrative. Identification of possible sources of bias can also provide a path forward to improvement in future efforts.

Similarly, the explicit calculation of uncertainty surfaces can provide insight into the spatial and temporal domains where more effort is needed, through improvement in either observations or models. For example, the spatial analysis of uncertainty inferred from atmospheric inversion of satellite-based soundings pointed to the need for improved *in situ* observations over tropical Africa (Peiro *et al* 2022). Similarly, Feng *et al* (2021) directly ingested an ensemble carbon flux product created by parameter perturbation into a high-resolution regional atmospheric transport model, and compared the modeled CO_2 with aircraft and tower measurements. The uncertainty analysis illustrated sink processes that were not well-accounted for in the terrestrial biome model. Alternatively, some studies have tested ensemble model inversions across a range of drivers, initial conditions, and model structure, and have used rank histograms (Hamill *et al* 2001) to evaluate the match of predicted to actual uncertainty, and thereby choose among ensemble members and best represent uncertainties (Diaz-Isaac *et al* 2018, Feng *et al* 2019). The shape of a rank-histogram (a.k.a. posterior predictive quantiles or Bayesian *p*-values) indicates whether the uncertainty estimates are over- or under-dispersive, which in turn can provide insight into whether key processes generating predictive uncertainty are appropriately represented.

Even if uncertainty is adequately calculated, passing on to users information about correlated errors (e.g. spatiotemporal autocorrelation) is a nontrivial problem. From a user's perspective, ensemble-based data products that represent Monte Carlo samples are by far the simplest way to ensure that uncertainties are propagated correctly. In this approach, users simply apply their calculation (e.g. upscaling or differencing) to every ensemble-member and the distribution across ensemble members represents the uncertainty. However, ensemble-based data products can significantly increase storage requirements and, at the moment, most users and producers lack familiarity with such products. Alternatively, reporting covariance information about data products can be much more compact, and in some cases easier to produce, but is significantly more challenging to understand how to use and apply correctly, especially when spatial or temporal covariance is nontrivial. It will also generally introduce larger approximation errors since assumptions about multivariate normality, stationarity, and anisotropy typically need to be invoked. Despite these challenges, we assert that consideration of spatiotemporal autocorrelation is critical when communicating prediction uncertainty.

Appropriate specification of spatiotemporal autocorrelation is also helpful when translating space- and time-specific predictions into the language of uncertainty used in broader carbon and climate risk assessments. Often, the language in those assessments is framed as uncertainty in the aggregate. For example, a community assessment of the global carbon budget (Friedlingstein *et al* 2019) reports on bulk estimates and uncertainties of global flux of carbon among the five major components of the carbon cycle: emissions from fossil carbon and land use change, and accumulation in sinks in the atmosphere, land and ocean. Each separate component is accumulated from space- and time-specific bookkeeping and modeling efforts, often using multiple sources, all of which include their own assessments of uncertainty. The uncertainty framework presented here could inform the aggregation process. First, it would help clarify which types of uncertainty are represented or not represented in the sub-component sources, aiding in the provision of a roadmap to the community for improvement of estimates. Second, it could draw attention to the impacts of spatial and temporal autocorrelation in estimates of uncertainty from component sources. In both cases, the effort focuses on providing a simple means of communicating uncertainty across groups in the scientific community. This is particularly important in global synthesis efforts, where a certain degree of expert opinion is needed to put sideboards on confidence of results. Clear tools to communicate uncertainty among experts will aid in this effort.

Clear communication of uncertainty is also essential in the emerging domain of carbon markets and in the monitoring, reporting, and verification of global climate agreements. In the agriculture, forestry, and land use sector, for example, remotely-sensed maps of deforestation, degradation, and carbon density can serve as central elements of county- or project-level carbon mitigation strategies (Bos *et al* 2019). However, both economic and compliance arenas require some measure of the confidence in the findings derived from these maps, and these maps are rarely accompanied with robust per-pixel estimates of uncertainty. Moreover, the proliferation of mapping products can lead to differing results across maps ostensibly claiming the same information (Neeti and Kennedy 2016), and without robust communication of uncertainty, policymakers and markets have no way to judge progress. Our framework could be applied to aid in this work.

More generally, our framework can help stakeholders navigate any situation where competing geospatial products disagree. Without a sense of source and completeness of uncertainty components, divergence among those products can lead to mistrust of all products. By highlighting divergence in components used to estimate uncertainty, the framework may dampen expectations of agreement among products, allowing users to focus on which types of uncertainty are more important in their use case, or to find means to blend together products from different sources to dampen risk. Moreover, a framework that articulates discussion among producers and users can lead to greater understanding and trust among users, and can clarify for producers which components of uncertainty need to be improved.

Ultimately, we envision our framing being relevant beyond carbon to any discipline dealing with the quantification uncertainties in data, models, fused model-data products, or findings, especially when space and time are important parts of such analyses. The emergence of spatially- and temporally-explicit estimates of uncertainty in any discipline will require similar considerations: source of uncertainty, frank appraisal of which uncertainties are not considered, and how to aggregate or disaggregate the uncertainty when ingesting or summarizing it. Although frameworks for uncertainty have a long pedigree, we hope that the lessons reflected in the NASA CMS Uncertainty Heuristic can provide a practical tool to capture these critical considerations.

Data availability statement

No new data were created or analysed in this study.

Acknowledgment

The work described here was conducted by the Uncertainty Working Group within NASA's Carbon Monitoring System (CMS) program. All co-authors received research support from the CMS program. Baker: NASA NNX14A077G, Cochrane: 80NSSC20K0408; Davis: NASA 80NSSC21K1060; Dietze: NASA 80NSSC21K0965, NASA 80NSSC17K0711, NSF 1638577 and NSF 1926388; Feng at PNNL NASA #80HQTR21T0069 (PNNL is operated by the US Department of Energy by the Battelle Memorial Institute under contract DE-A05-76RL0 1830); Hudak: 80HQTR20T0002; Kennedy: NASA NNX16AQ25G, Kennedy, Babcock, Hudak: 80NSSC21K1001; Liu: acknowledges the support of the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA (80NM0018D0004); Raczka (Lin): NASA 80NSSC20K0010; Serbin: NASA 80HQTR21T0066, NASA 80GSFC22TA016.

ORCID iDs

Robert E Kennedy https://orcid.org/0000-0002-5507-474X
Shawn P Serbin https://orcid.org/0000-0003-4136-8971
Hans-Erik Andersen https://orcid.org/0000-0002-5546-9522
Molly E Brown https://orcid.org/0000-0001-7384-3314
Kenneth J Davis https://orcid.org/0000-0002-1992-8381
Laura Duncanson https://orcid.org/0000-0003-4031-3493
Sha Feng https://orcid.org/0000-0002-2376-0868
Andrew T Hudak https://orcid.org/0000-0001-7480-1458
Junjie Liu https://orcid.org/0000-0002-7184-6594
Brett Raczka https://orcid.org/0000-0003-0582-1118
Rodrigo Vargas https://orcid.org/0000-0001-6829-5333

References

Alcamo J and Bartnicki J 1987 A framework for error analysis of a long-range transport model with emphasis on parameter uncertainty *Atmos. Environ.* **21** 2121–31

Anderson J L 2007 An adaptive covariance inflation error correction algorithm for ensemble filters *Tellus A* **59** 210–24

Anderson J, Hoar T, Raeder K, Liu H, Collins N, Torn R and Avellano A 2009 The data assimilation research testbed: a community facility *Bull. Am. Meteorol. Soc.* **90** 1283–96

Babcock C, Finley A O, Cook B D, Weiskittel A and Woodall C W 2016 Modeling forest biomass and growth: coupling long-term inventory and LiDAR data *Remote Sens. Environ.* **182** 1–12

Baker D F *et al* 2006 TransCom 3 inversion intercomparison: impact of transport model errors on the interannual variability of regional CO_2 fluxes, 1988–2003: TRANSCOM 3—INTERANNUAL VARIABILITY OF CO $_2$ SOURCES *Glob. Biogeochem. Cycles* **20** n/a

Bechtold W A and Patterson P L 2005 *The Enhanced Forest Inventory and Analysis Program—National Sampling Design and Estimation Procedures (No. SRS-GTR-80)* (U.S. Department of Agriculture, Forest Service, Southern Research Station) (https://doi.org/10.2737/SRS-GTR-80)

Bevan L D 2022 The ambiguities of uncertainty: a review of uncertainty frameworks relevant to the assessment of environmental change *Futures* **137** 102919

Blackhurst M and Matthews H S 2022 Comparing sources of uncertainty in community greenhouse gas estimation techniques *Environ. Res. Lett.* **17** 053002

Blyth E M *et al* 2021 Advances in Land Surface Modelling *Curr. Clim. Change Rep.* **7** 45–71

Bonan G B and Doney S C 2018 Climate, ecosystems, and planetary futures: the challenge to predict life in Earth system models *Science* **359** eaam8328

Bos A B, De Sy V, Duchelle A E, Herold M, Martius C and Tsendbazar N-E 2019 Global data and tools for local forest cover loss and REDD+ performance assessment: accuracy, uncertainty, complementarity and impact *Int. J. Appl. Earth Obs. Geoinf.* **80** 295–311

Brown M E, Cooper M W and Griffith P C 2020 NASA's carbon monitoring system (CMS) and arctic-boreal vulnerability experiment (ABoVE) social network and community of practice Environ. Res. Lett. **15** 115014

Byrne B *et al* 2023 National CO_2 budgets (2015–2020) inferred from atmospheric CO_2 observations in support of the global stocktake *Earth Syst. Sci. Data* **15** 963–1004

Cameron D, Hartig F, Minnuno F, Oberpriller J, Reineking B, Van Oijen M and Dietze M 2022 Issues in calibrating models with multiple unbalanced constraints: the significance of systematic model and data errors *Methods Ecol. Evol.* **13** 2757–70

Cassenti D N and Kaplan L M 2021 Robust uncertainty representation in human-AI collaboration *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications III. Presented at the Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications III* (SPIE) pp 249–62

Collier N, Hoffman F M, Lawrence D M, Keppel-Aleks G, Koven C D, Riley W J, Mu M and Randerson J T 2018 The International Land Model Benchmarking (ILAMB) System: Design, Theory, and Implementation J Adv Model Earth Syst **10** 2731–2754

Crowell S *et al* 2019 The 2015–2016 carbon cycle as seen from OCO-2 and the global in situ network *Atmos. Chem. Phys.* **19** 9797–831

Dankers R and Kundzewicz Z W 2020 Grappling with uncertainties in physical climate impact projections of water resources *Clim. Change* **163** 1379–97

Diaz-Isaac L I, Lauvaux T and Davis K J 2018 Impact of physical parameterizations and initial conditions on simulated atmospheric transport and CO_2 mole fractions in the US Midwest *Atmos. Chem. Phys.* **18** 14813–35

Dietze M C 2017 Ecological Forecasting (Princeton University Press)

Dietze M C, Thomas R Q, Peters J, Boettiger C, Shiklomanov A N and Ashander J 2023 A community convention for ecological forecasting: output files and metadata 1.0 *Ecosphere* **14** e4686

Dokoohaki H, Morrison B D, Raiho A, Serbin S P and Dietze M, 2021. A novel model–data fusion approach to terrestrial carbon cycle reanalysis across the contiguous U.S using SIPNET and PEcAn state data assimilation system v. 1.7.2 *Copernicus GmbH* pp 1–28

Duarte H F, Raczka B M, Bowling D R, Wang A, Buotte P C and Lin J C 2021 How can biosphere models simulate enough vegetation biomass in the mountains of the western United States? Implications of meteorological forcing *Environ. Model. Softw* 148

Dubayah R *et al* 2020 The global ecosystem dynamics investigation: high-resolution laser ranging of the Earth's forests and topography *Sci. Remote Sens.* **1** 100002

Feng S, Lauvaux T, Keller K, Davis K J, Rayner P, Oda T and Gurney K R 2019 A road map for improving the treatment of uncertainties in high-resolution regional carbon flux inverse estimates *Geophys. Res. Lett.* **46** 13461–9

Feng S, Lauvaux T, Williams C A, Davis K J, Zhou Y, Baker I, Barkley Z R and Wesloh D 2021 Joint CO_2 mole fraction and flux analysis confirms missing processes in CASA terrestrial carbon uptake over North America *Glob. Biogeochem. Cycles* **35** e2020GB006914

Fischhoff B and Davis A L 2014 Communicating scientific uncertainty *Proc. Natl. Acad. Sci. U.S.A.* **111** 13664–71

Fisher R A and Koven C D 2020 Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems *J. Adv. Model. Earth Syst.* **12** e2018MS001453

Friedlingstein P *et al* 2019 Global carbon budget 2019 *Earth Syst. Sci. Data* **11** 1783–838
Gaudard L and Romerio F 2020 A conceptual framework to classify and manage risk, uncertainty and ambiguity: an application to energy policy *Energies* **13** 1422
Hamill T, Whitaker J and Snyder C 2001 Distance-dependent filtering of background error covariance estimates in an ensemble kalman filter *Mon. Weather Rev.* **129** 2776–90
Harmon M E, Fasth B, Halpern C B and Lutz J A 2015 Uncertainty analysis: an evaluation metric for synthesis science *Ecosphere* **6** 63
Hudak A T, Fekety P A, Kane V R, Kennedy R E, Filippelli S K, Falkowski M J, Tinkham W T, Smith A M, Crookston N L and Domke G M 2020 A carbon monitoring system for mapping regional, annual aboveground biomass across the northwestern USA *Environ. Res. Lett.* **15** 095003
Hurtt G C *et al* 2022 The NASA Carbon Monitoring System Phase 2 synthesis: scope, findings, gaps and recommended next steps *Environ. Res. Lett.* **17** 063010
IPCC 2023 *Climate Change 2023: Synthesis Report* (Intergovernmental Panel on Climate Change)
ISO/IEC 2008 Uncertainty of measurement—Part 3: Guide to the Expression of Uncertainty in Measurement
Jones C D *et al* 2016 C4MIP – The Coupled Climate–Carbon Cycle Model Intercomparison Project: experimental protocol for CMIP6 *Geosci. Model Dev.* **9** 2853–80
Kennedy R E *et al* 2018 An empirical, integrated forest biomass monitoring system Environ. Res. Lett. **13** 025004
Kwakkel J H, Walker W E and Marchau V A W J 2010 Classifying and communicating uncertainties in model-based policy analysis *Int. J. Technol. Policy Manage* **10** 299–315
Lawrence D M *et al* 2019 The Community Land Model Version 5: Description of New Features, Benchmarking, and Impact of Forcing Uncertainty *J. Adv. Model Earth Syst.* **11** 4245–4287
LeBauer D S, Wang D, Richter K T, Davidson C C and Dietze M C 2013 Facilitating feedbacks between field measurements and ecosystem models *Ecol. Monogr.* **83** 133–54
Liu J *et al* 2021 Carbon monitoring system flux net biosphere exchange 2020 (CMS-Flux NBE 2020) *Earth Syst. Sci. Data* **13** 299–330
Meddens A J, Steen-Adams M M, Hudak A T, Mauro F, Byassee P M and Strunk J 2022 Specifying geospatial data product characteristics for forest and fuel management applications Environ. Res. Lett. **17** 045025
Menlove J and Healey S P 2020 A comprehensive forest biomass dataset for the USA allows customized validation of remotely sensed biomass estimates *Remote Sens.* **12** 4141
Neeti N and Kennedy R 2016 Comparison of national level biomass maps for conterminous US: understanding pattern and causes of differences *Carbon Balance Manage.* **11** 19
Patterson P L *et al* 2019 Statistical properties of hybrid estimators proposed for GEDI—NASA's global ecosystem dynamics investigation *Environ. Res. Lett.* **14** 065007
Peiro H *et al* 2022 Four years of global carbon cycle observed from the Orbiting Carbon Observatory 2 (OCO-2) version 9 and in situ data and comparison to OCO-2 version 7 *Atmos. Chem. Phys.* **22** 1097–130
Raczka B, Hoar T J, Duarte H F, Fox A M, Anderson J L, Bowling D R and Lin J C 2021 Improving CLM5.0 biomass and carbon exchange across the western united states using a data assimilation system *J. Adv. Model. Earth Syst.* **13** e2020MS002421
Raiho Ann, Dietze Michael C, Dawson Andria, Rollinson Christine R, Tipton John and McLachlan Jason 2020 Towards understanding predictability in ecology: A forest gap model case study *biorxiv* **2020.05.05** 079871
Saatchi S S *et al* 2011 Benchmark map of forest carbon stocks in tropical regions across three continents *Proc. Natl Acad. Sci.* **108** 9899–904
Sanderson B M and Knutti R 2012 On the interpretation of constrained climate model ensembles *Geophys. Res. Lett.* **39**
Särndal C-E, Swensson B and Wretman J 1992 *Model Assisted Survey Sampling* (Springer)
Tak S, Toet A and van Erp J 2015 Public understanding of visual representations of uncertainty in temperature forecasts *J. Cogn. Eng. Decis. Mak.* **9** 241–62
Tebaldi C and Knutti R 2007 The use of the multi-model ensemble in probabilistic climate projections *Phil. Trans. R. Soc.* **365** 2053–75
Walker W E, Harremoës P, Rotmans J, van der Sluijs J P, van Asselt M B A, Janssen P and Krayer von Krauss M P 2003 Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support *Integr. Assess. J.* **4** 5–17