Understanding the Various Perspectives of Earth Science Observational Data Uncertainty

Authors: David F. Moroni[1], Hampapuram Ramapriyan[2], Ge Peng[3], Jonathan Hobbs[1], Justin C. Goldstein[4], Robert R. Downs[5], Robert Wolfe[6], Chung-Lin Shie[7], Christopher J. Merchant[8], Mark Bourassa[9], Jessica L. Matthews[3], Peter Cornillon[10], Lucy Bastin[11], Kenneth Kehoe[12], Benjamin Smith[1], Jeffrey L. Privette[13], Aneesh C. Subramanian[14], Otis Brown[3], Ivana Ivánová[15]

Affiliations:

1. Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA
2. Science Systems and Applications, Inc. and NASA Goddard Space Flight Center
3. North Carolina State University (NCSU), Cooperative Institute for Satellite Earth System Studies at NOAA's National Centers for Environmental Information (NCEI), Asheville, NC
4. Riverside Technology, Inc. supporting the NOAA NESDIS Technology, Planning, and Integration for Observation (TPIO) Division
5. Center for International Earth Science Information Network, Columbia University
6. NASA Goddard Space Flight Center
7. NASA Goddard Space Flight Center and University of Maryland, Baltimore County
8. University of Reading and National Centre for Earth Observation, Reading, UK
9. Center for Ocean-Atmospheric Prediction Studies, Florida State University
10. University of Rhode Island
11. Joint Research Centre, European Commission and Aston University, UK
12. University of Oklahoma - Cooperative Institute for Mesoscale Meteorological Studies
13. NOAA's National Centers for Environmental Information
14. University of Colorado Boulder
15. Curtin University, Australia

Recommended Citation:
Moroni, D. F., Ramapriyan, H., Peng, G., Hobbs, J., Goldstein, J. C., Downs, R. R., Wolfe, R., Shie, C.-L., Merchant, C. J., Bourassa, M., Matthews, J. L., Cornillon, P., Bastin, L., Kehoe, K., Smith, B., Privette, J. L., Subramanian, A. C., Brown, O., & Ivánová, I. (2019) Understanding the Various Perspectives of Earth Science Observational Data Uncertainty. *Figshare*. https://doi.org/10.6084/m9.figshare.10271450.

Abstract: Information about the uncertainty associated with Earth science observational data is fundamental to use, re-use, and overall evaluation of the data being used to produce science and support decision making. The associated uncertainty information leads to a quantifiable level of confidence in both the data and the science informing decisions produced using the data. The current breadth and cross-domain depth of understanding and application of uncertainty information, however, are still evolving as the practices associated with quantifying and characterizing uncertainty across various types of Earth observation data are diverse. Since its re-establishment in 2015, the Information Quality Cluster (IQC) of the Earth Science Information Partners (ESIP) has convened numerous sessions within the auspices of ESIP and the American Geophysical Union (AGU) to help collect expert-level information focusing on key aspects of uncertainty of Earth science data and addressed key concerns such as: 1) how uncertainty is quantified (UQ) and characterized (UC), 2) understanding the strengths and limitations of common techniques used in producing and evaluating uncertainty information, 3) implications using uncertainty information as a quality indicator 4) impacts of uncertainty on data fusion/assimilation, 5) various methods for documenting and conveying the uncertainty information to data users, and 6) understanding why certain user communities care about uncertainty and others do not. A key recommendation and action item from the ESIP Summer Meeting 2017 was for the IQC to develop a white paper to establish a clearer understanding of the concept of uncertainty and its communication to data users. The information gathered for this white paper has been provided by Earth science data and informatics experts spanning diverse disciplines and observation systems in the cross-domain Earth sciences. The intention of this white paper is to provide a diversely sampled exposition of both prolific and unique policies and practices, applicable in an international context of diverse policies and working groups, made toward quantifying, characterizing, communicating and making use of uncertainty information throughout the diverse, cross-disciplinary Earth science data landscape.

Keywords: Data Uncertainty, Uncertainty Quantification, Uncertainty Characterization, Earth Science, Earth Science Information Partners, Information Quality Cluster, Remote Sensing, Observation, In Situ, Calibration, Validation, Probability, Science Quality, Data Quality, Metrology

Table of Contents

1. Introduction

Earth Science Information Partners (ESIP) was established in 1998 as a non-profit, volunteer and community-driven organization. ESIP partners include federal data centers and university repositories, federal and academic research facilities, nonprofit and commercial enterprises. With over 140 partner organizations, ESIP aims to advance the use of Earth science data for societal applications like disaster response, climate, energy (i.e., commercial/industrial sources) and agriculture. An up-to-date list (note: this is periodically updated) of ESIP partner organizations is maintained here: https://www.esipfed.org/partners. Through its work, ESIP promotes open data and data sharing utilizing community best practices for data management, data stewardship, and information technology. Furthermore, ESIP provides "clusters" which function as open-membership groups to work together on more focused challenges pertaining to Earth science data informatics. The Information Quality Cluster (IQC) is one such cluster. IQC focuses on identifying solutions and best practices to address the overall quality of data and associated information throughout the data product lifecycle.

The creation and representation of Earth science data quality information is a significant and daunting challenge because "it is not uniquely defined, user dependent, difficult to be quantified, handled differently by different teams and perceived differently by data providers and data users" (Vicente, 2014). Moreover, different perspectives on "quality" are justifiable in different contexts, notably with regard to how the uncertainty of data is more strictly defined in a mathematical framework, yet disparately interpreted and applied in a user-centric, practical framework.

The IQC, re-established in 2015 by ESIP, published its seminal paper in July 2017 (Ramapriyan *et al*., 2017) that formally laid out the terminologies and paradigms of Information Quality. Ramapriyan *et al*. (2017) built on years of research, the efforts of the NASA Earth Science Data Systems Working Groups (ESDSWG) Data Quality Working Group (DQWG) from 2014 to 2016, and extensive discussions and collaborations within the ESIP IQC in subsequent years. Two primary components of information quality as described therein are: 1) scientific quality and 2) product quality. The scientific quality has been formally defined as "accuracy, precision, uncertainty, validity and suitability for use (fitness for purpose)" (Ramapriyan *et al*., 2017) for a given dataset or product of data. The product quality depends on the information provided from the scientific quality, but moreover it covers the following aspects of data and information quality: "the degree to which the scientific quality is assessed and documented; how accurate, complete and up-to-date the metadata and documentation are; the manner in which the data and metadata are formatted; the degree to which the associated information including provenance are published and traceable throughout the data lifecycle" (Ramapriyan *et al*., 2017).

In its attempts to place a greater emphasis on the scientific quality of Earth science data and information, the IQC has convened numerous sessions within the auspices of ESIP and the American Geophysical Union (AGU) to help collect expert-level information with regard to the following topics and challenges: 1) uncertainty quantification (UQ) and characterization (UC), 2) limitations of common techniques used in producing and evaluating uncertainty information, 3) using uncertainty information as a quality indicator, 4) various methods for documenting and conveying the uncertainty information to data users, and 5) understanding the overall lifecycle of uncertainty information, beginning with programmatic and inter-governmental initiatives and culminating in the dissemination and end user uptake of the uncertainty information.

The purpose and motivation for this white paper was cemented by a key recommendation and action item from the ESIP Summer Meeting 2017, which requested that the IQC develop and publish a paper to establish a clear understanding of the concept of uncertainty, and how it is generally evaluated via UQ/UC for Earth System Data Records (ESDRs) and leveraged by data producers and end users. The information gathered for this paper has been provided by science data and informatics experts spanning diverse disciplines and observation systems within the Earth science domain. This paper intends to provide an exposition of collective policies, practices and advances made toward quantifying, characterizing, communicating and making use of uncertainty information throughout the diverse, cross-disciplinary Earth science data landscape.

In the following sections, we present a variety of perspectives on uncertainty that are intended to paint a clear and comprehensive picture of the following: 1) foundational and advanced mathematical concepts and methods for UQ and UC (via the Mathematical section), 2) research and applications examples for how uncertainty information is used (via the User section), 3) institutional/agency/international drivers and policies for implementing UQ and UC for their data (via the Programmatic section), and 4) understanding the methods of UQ and UC employed across observational data covering diverse Earth science disciplines (via the Observational section). This paper will conclude with an assessment of both the differences and commonalities of these high-level perspectives. While use cases and discussion regarding uncertainty of computational model data were considered for both prognostic and diagnostic models, it was decided best to conserve the limited authorship resources committed to this paper and focus existing expertise on perspectives of observational data and digital products derived from them. While this paper primarily focuses on statistics-based techniques of evaluating uncertainty information, there are also knowledge-based techniques that are regularly used, but for the scope of this paper, such knowledge-based techniques were not earnestly explored.

2. Perspectives: Historical and Current

All measured values have a degree of uncertainty. Understanding the nature and degree of that uncertainty is important to users of that measurement, from scientists who need to understand how that uncertainty impacts their data products or models and predictions, to decision makers who need to understand the likelihood of various outcomes. The manner in which uncertainty information is derived, characterized, disseminated, and ultimately consumed and interpreted by the end user is quite diverse, yet there are shared commonalities. The international metrological community, which defines and promulgates measurement standards, has been a focus of philosophical, mathematical and practical developments in measurement science (JCGM, 2008) which increasingly informs observational practices in Earth sciences. There are clearly changes in practices that have evolved as a function of time due to technological advancements, more advanced computational methods and algorithms, discipline-specific observations, advancements in standardized specifications for representing science data information within formatted data files, progress within government and agency-specific priorities on Earth science programs, and variations in observing systems. To better assess both the differences and commonalities with regard to Earth science data uncertainty informatics, this paper contains an exposition of perspectives broken down as follows: a) Mathematical, b) User, c) Programmatic, and d) Observational. These four perspectives are discussed in more detail in the following subsections.

2.1 Mathematical

In this section, we will review the academic and computational foundations for the manner in which uncertainty information is estimated and interpreted from a statistical framework, noting that statistical evaluation is not the only means by which uncertainty may be quantified. Uncertainty characterization, hereafter referred to as UC, is the process of describing the functional dependencies, correlations, and distributions of errors for one or more quantities of interest (QOIs). UC is often aided by statistical, scientific, and/or engineering based analysis of potential sources of random and systematic errors in estimating a QOI, particularly when the QOI is estimated through a complex procedure and its behavior changes under different geophysical conditions. Uncertainty quantification, hereafter referred to as UQ, refers to the numerical description of the uncertainty or remaining lack of knowledge of the true value of a QOI. This description is often a summary of a probability distribution. Probability is foundational to UQ and offers a concise and coherent mathematical framework for UC. In this context, it is important to distinguish between a QOI (which in metrological parlance is a “measurand”) and an observation (i.e., a datum from which is obtained a “measured value” attributed to a measurand). In what follows we denote a QOI as X and an associated observation (or numerous observations) as Y. An Earth System Data Record (ESDR) will usually include routine processing of the observation into an estimate of the QOI. We denote this estimate $\hat{X} = R(Y)$, where R is a function. For some *in situ* climate data records, it is reasonable to assume that the best estimate of the QOI is the instrument's observation itself, that is, $\hat{X} \cong Y$. In most cases, even for *in situ* datasets, there is some type of $R(Y)$ function that translates from the observations to the estimate. For data records derived from remote sensing instruments (space-based, sub-orbital, or ground based optical/lidar/radar/sonar) the observations relate indirectly to the QOI; the function $R(Y)$ can then be complex and the task of estimating X can be appropriately cast as an *inverse problem.*

In any of these contexts, formal UQ can be viewed from a probabilistic perspective. Formally, the observation and often the QOI are viewed as random variables and are characterized with probability distributions. From here, the uncertainty can be viewed through the probability distribution as a way to provide additional knowledge about the range of values that can be reasonably attributed to the QOI given the observation (JCGM, 2008). This can be summarized through a conditional probability distribution, $p(X|Y)$, where p is a generic representation of a probability density function. Alternatively, in some settings, it can be convenient to instead summarize the estimation of the marginal error distribution, $p(\hat{X} - X)$. Connections between these distributions are described below. When appropriate, p is characterized as a parametric distribution, $p(\theta)$. That is, the full distribution is represented through a known mathematical form that may be a function of a small set of free parameters θ.

There are two key types of uncertainty that can be represented probabilistically. Aleatoric (or aleatory) uncertainty represents uncertainty associated with chance, and epistemic uncertainty represents uncertainty arising from lack of knowledge (Smith, 2014). Aleatoric uncertainty has been classically treated probabilistically with terms like "random error". Sources of uncertainty such as instrument measurement noise and computational artifacts in estimation algorithms fall into this category. Replication, or multiple observations of the same QOI, can reduce the impact of these sources of uncertainty on the evaluation of the QOI. However, in Earth science, the continuous change with time of the system being observed commonly excludes true replicate measurements. Nonetheless, techniques such as taking differences between two or more coincident observations of the same variable are used to approximate replicate measurements, with the need to account for imperfect colocation and changing variability in space and/or time. Further, the often mismatched spatial and temporal scales between the QOI and observations considered for validation can also contribute to the uncertainty. In contrast, epistemic uncertainty is realized through lack, or imprecision, of knowledge about the processes that relate the QOI and observation; the QOI and the observation itself may be directly affected by epistemic uncertainty. This can include incorrect specification of the true mathematical relationship, or *forward model*, that governs the relationship, as well as imperfect knowledge of validation data and ancillary parameters, such as calibration constants, that impact this relationship (Rodgers, 2000). These epistemic sources can contribute to either random or systematic errors in estimating the QOI.

Including uncertainty metric(s) in ESDRs has been an emerging goal for scientists, data archive managers, and agency administrators, and reducing uncertainty in key Earth system processes motivates long-term planning for future observing systems. The Joint Committee for Guides in Metrology (JCGM, 2008) has provided guidance on reporting uncertainty in a succinct manner for direct measurements of QOIs, and the principles outlined there can be extended to inferences involving inverse methods often encountered in ESDRs. In mathematical terms, UQ can be achieved by extracting appropriate summaries of the conditional error distribution $p(X|Y)$ or the marginal error distribution $p(\hat{X} - X)$. We will focus here on the latter. A minimal description of the distribution includes the mean μ and standard deviation σ. The mean is the expected value of the error distribution,

$$\mu = E(\hat{X} - X).$$

The mean of the error distribution is often called the *bias*. The standard deviation of the error distribution is

$$\sigma = \sqrt{E[(\hat{X} - E(X))^2]}$$

In situations where a single parameter is associated with a QOI's uncertainty, the standard deviation is very often selected to characterize the distribution of possible values of the QOI X given the observation $Y$and is formally the "standard uncertainty". This succinct summary of the distribution is reasonable when the QOI is univariate and the error distribution can be reasonably approximated by a Gaussian (or, "normal") distribution. Such a summary of the error distribution becomes increasingly problematic as the error distributions become more skewed.

Increasingly, ESDRs, particularly those produced from satellite remote sensing instruments, involve estimates of geophysical QOIs X that combine indirect observations Y with physical and/or statistical models. For example, Level 2 satellite data products result from processing Level 1 radiance spectra through operational retrieval algorithms, such as expressed by the Aquarius Salinity Retrieval ATBD (Meissner *et al*., 2017) and the MODIS Terra Land Surface Temperature ATBD (Hulley *et al*., 2016). These are examples of inverse problems which present several challenges for uncertainty characterization, including a tendency to be ill-posed and highly sensitive, particularly when the relationship between the QOI and the observation is nonlinear on scales finer than the uncertainty (NRC, 2012).

Bayesian inference represents an appealing framework for uncertainty quantification for inverse problems. In this context, the focus is on the conditional distribution of the QOI given the observation, which is constructed through Bayes' theorem,

$$p(X|Y) = \frac{p(Y|X)\,p(X)}{p(Y)}.$$

The conditional distribution $p(X|Y)$, often called the *posterior distribution*, is proportional to the product of a *prior distribution* for the QOI, $p(X)$, and a *likelihood*, $p(Y|X)$, that characterizes the conditional distribution of the observation given the QOI. Figure 1 illustrates this composition of information from the data and the prior distribution into a probabilistic description of the QOI given the observation.

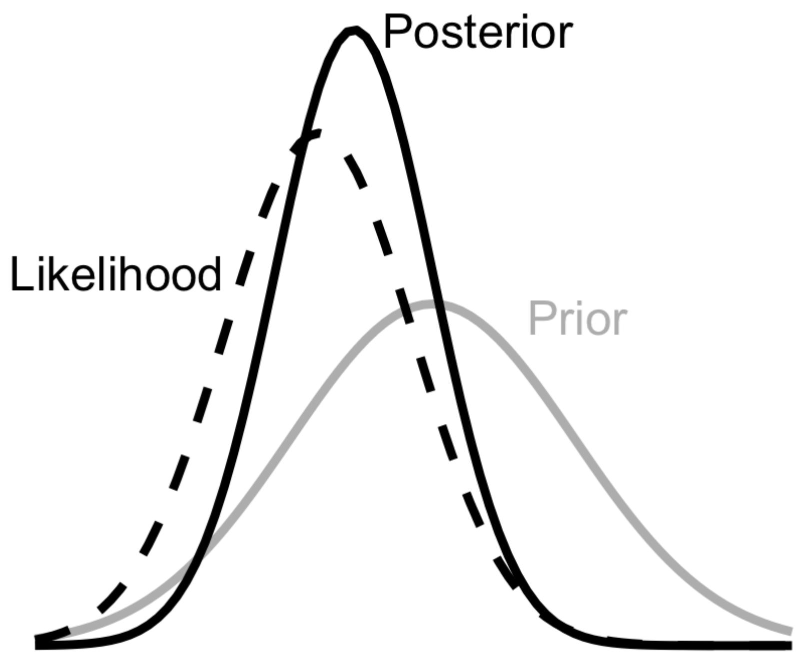

Figure 1: Schematic implementation of Bayes' theorem for a univariate QOI. The prior distribution is combined with information from an observation (via the likelihood) to produce a posterior distribution.

Some atmospheric remote sensing retrievals and their resulting data products adopt this Bayesian approach (Rodgers, 2000), although usually with the assumptions of normality and at most moderate nonlinearity. Products derived from data assimilation systems adopt the Bayesian paradigm, often with similar assumptions. While the construction is conceptually simple, challenges outlined above, particularly in formally representing aleatoric and epistemic uncertainty through probability distributions, remain in the Bayesian approach to inverse problems. In particular, the prior and likelihood may be characterized by additional statistical parameters that are unknown and need to be estimated. Further, the posterior distribution cannot be interrogated directly where normal and nearly linear assumptions do not hold; in such cases it is sampled with Monte Carlo methods (Gelman *et al*., 2014).

In either the case of direct, but uncertain, measurement of a QOI or in an inverse problem, the resulting distribution for the QOI may be complex and deviate from normality. This is a particular issue when the QOI is discrete or otherwise physically constrained or when the relationship between the QOI and observation is nonlinear. These situations are fairly common for geophysical variables and summarizing the QOI distribution may require additional information. In this general situation, the mean and standard deviation are two of many possible *moments* of the underlying distribution. Additional moments, particularly those describing, skewness and kurtosis, can be reported. Alternatively, the error or posterior distribution can be summarized by providing a set of *quantiles* or a collection of random draws from the distribution (Gelman *et al*., 2014). These generalizations are also applicable if the QOI is multivariate.

Many complex Earth system datasets, such as those from remote sensing instruments, are the result of multiple levels of data processing algorithms and quality control. Further, additional processing to combine, improve the observational stability of and standardize the format of multiple observation records of the same QOI can provide more usable data records. When possible, the data processing pipeline should be represented probabilistically to properly characterize the cascade of uncertainty (Mittaz *et al*., 2019). For example, Cressie (2018) discusses these issues in the context of a processing pipeline for greenhouse gas estimates from satellite data. In addition, a sensible probabilistic framework for a QOI that is informed by multiple observation records can exploit opportunities for shared and complementary information to reduce uncertainty on the QOI. Examples include fusion of data from multiple satellites to estimate atmospheric properties (Nguyen *et al*., 2012).

2.2 User

The user perspective is afforded by effective interpretation and communication of uncertainty information to facilitate understanding and applications of Earth observations and other Earth science data. Sufficient understanding of the uncertainty associated with Earth science data can enable applications of the data within different contexts. Communicating uncertainty information effectively can improve understanding and enable applications of Earth science data for both experts and non-experts, especially when offered with appropriate graphical representations, definitions, and guidance documentation. Considerations of human factors and differences among users and their applications also can contribute to effective communication of uncertainty information. Furthermore, the development of standards, tools, and services also offers opportunities to improve understanding of uncertainty information for using Earth science data.

Each application has its own context. Earth observations have time and space characteristics associated with them. It can be challenging to map observational uncertainty into the time and space context of user's applications when various spatio-temporal error characteristics are present.

Without sufficient information about the reasons why data were interpreted in a particular manner, users will not necessarily know how the findings were interpreted or be able to understand the practical implications of data, models, graphs, and other information that are presented in scientific products and services. Information about the evaluation of uncertainty can help to facilitate such understanding if communicated effectively. Graphical representations, such as figures and charts, could be used to communicate quantitative uncertainty to users in a manner that can be understood qualitatively (Morgan & Henrion, 1993). Similar approaches can be used to show the user particular characteristics of the data (e.g., spatial and temporal resolution) that help a user reduce errors related to assumptions, and to aggregate that data in a manner tuned to the user's application.

Communicating uncertainty measures can be valuable, even for non-expert populations that use such information. In a survey study of perceptions of the general population regarding precipitation probability as reported in weather forecasts, Murphy *et al*. (1980, p. 697) found that most participants recognized that numerical probabilities are reported to describe "uncertainty regarding the occurrence of precipitation". These authors also concluded that "participants not only understood but strongly preferred forecasts of precipitation occurrence in which uncertainty is expressed in probabilistic terms" (Murphy *et al*., 1980, p.699).

When communicating uncertainty measures, it would be valuable to use and refer to precise definitions of such measures, particularly those formally developed within the metrological community, and to provide guidance and examples of use for the uncertainty measures. This can facilitate correct interpretation of the data through proper use of the uncertainty information.

Earth science data obtained through crowdsourcing or citizen science (CS) initiatives comes under particularly harsh scrutiny from users who are suspicious of its quality. Numerous attempts have been made to evaluate and characterize the errors in volunteered data (e.g., Gardiner *et al*., 2012) with a number of recent studies concluding that sampling protocol can be more influential than expertise in delivering consistency and quality (Theobald *et al*., 2015; Van de Velde *et al*., 2017) and a consequent attention to standardized protocols and training (e.g. Tweddle *et al.*, 2012).

Kinkeldey *et al*. (2014) reviewed user studies that evaluated methods and tools for communicating and visually analyzing (Kinkeldey, 2014) uncertainty of geospatial information. They concluded that such studies should employ methods that foster systematic evaluation and should pay attention to the types of tasks being performed by the users of uncertainty visualizations. Subsequently, reviewing user studies of uncertainty visualization to support decision-making, Kinkeldey *et al*. (2017, p. 8) noted that, when decision makers use uncertainty information, "they have to accept the additional effort of incorporating it into decision-making". In an extensive web-based user survey, Senaratne *et al*. (2012) tested understanding of uncertainty visualization with users predominantly from five different domains: GIS, map visualization, statistics, decision-support and urban planning. Based on the results of their survey, the authors proposed a model of uncertainty visualization selector, a tool which aids producers to select the appropriate visualization technique, depending on the nature of QOI. In addition, the authors share their usability study design as a generic approach to help assessing usability of uncertainty visualization methods in the future (Senaratne *et al*., 2012). Sacha *et al*. (2016) analyzed the uncertainty propagation through visual analytical systems, its perception and understanding by the users, and implication of uncertainty understanding on users' trust in the results of visual analytics. The authors provide several guiding principles to avoid misconfiguration of uncertainty communication systems, which include: analysis of human factors involved in uncertainty communication, quantification of uncertainty at each stage of data processing (source, model building and usage, data processing and visualization of results), visualization of propagated uncertainty and enable users' interaction with uncertainty (Sacha *et al*., 2016).

There is no comprehensive standard for documenting and communicating uncertainty of ESDRs except for several well-known statistical measures included in the ISO 19157 data quality standard (ISO, 2013). UncertML is a markup language for documenting uncertainty information and its "descriptive capabilities range from summaries, such as simple statistics (e.g. the mean and variance of an observation), to more complex representations such as parametric distributions at each point of a regular grid, or even jointly over the entire grid" (Williams *et al*., 2009, p. 8). The Open Geospatial Consortium (OGC) proposes standards ("UncertML") for parametric distributions that may be used to characterize uncertainty (Williams *et al*., 2009), but does not refer to metrological standards for communicating uncertainty in measurement. UncertML has been published as an OGC discussion paper (*ibid.*), and its effective use has been demonstrated in various implementation efforts in communicating uncertainty of an ESDR. One successful example is the Image Quality and Accuracy Engineering Report, published as part of OGC's Testbed-12 innovation program (Masó and Zabala, 2017). In this report, the authors describe how to encode image quality (often represented by uncertainty measures) to aid decision on an ESDR's fitness for use. The proposed quality information communication model is based on ISO 19157, UncertML and QualityML, with the latter being an extended XML profile of ISO19157 (Masó and Zabala, 2017). The implication of such studies is that there is a trade-off that can be beneficial for decision-making if uncertainty is communicated effectively to facilitate understanding and interpretation by the intended users.

Tools and services are also another avenue for increasing the user uptake of uncertainty information, while also minimizing the effort required in applying pre-analyzed uncertainty estimates to multivariate geophysical data analysis. For example, the International Atomic Energy Agency (IAEA) Ocean Acidification International Coordination Centre (OA-ICC; https://www.iaea.org/services/oa-icc) has recently produced a consistent set of software tools released as public software packages to assist in the utilization of propagated uncertainty estimates in marine CO2 system variables (Orr *et al*., 2018). Orr *et al*. (2018) point out that these packages use a unique "error-space diagram" to assess the propagated uncertainty variations as a function of changes in the input uncertainty estimates. These tools are referred to as "uncertainty propagation add-ons" made available in the following four packages:

1. CO2SYS-Excel (https://github.com/jamesorr/CO2SYS-Excel);
2. CO2SYS-MATLAB (https://github.com/jamesorr/CO2SYS-MATLAB);
3. seacarb (https://cran.r-project.org/web/packages/seacarb/index.html);
4. mocsy (https://github.com/jamesorr/mocsy).

2.3 Programmatic

Programs at global, national and agency levels (see Figure 2) have emphasized the need for UQ and UC, as well as provision of such information to users of data and derived information. Examples of global entities stressing the importance of UQ and UC are the Group on Earth Observations (GEO, 2015), the Committee on Earth Observing Satellites (CEOS) Working Group on Calibration and Validation (WGCV), the International Panel on Climate Change (IPCC) (Mastrandrea *et al.,* 2010), the World Meteorological Organization (WMO), and the Working Group on Climate (WGClimate), a joint working group of CEOS and Coordination Group for Meteorological Satellites (CGMS). The GEO discusses the need for UC in the context of the Data Management Principles (DMP) on data traceability (DMP-5) and data quality control (DMP-6). The CEOS WGCV functions via smaller sub-working groups that focus on standardizing the UQ and UC for satellite observations within specific disciplinary frameworks. The IPCC provides detailed guidance to authors of the fifth Assessment Report (AR5) to communicate the degree of certainty in their key findings. Uncertainty requirements within WMO are mainly found in the context of requirements with respect to systematic observations (UNFCCC, 2008; GCOS, 2016). They call for the inclusion of measures of uncertainty for Essential Climate Variables (ECVs). This is not consistently defined for other types of data products. WGClimate has recently developed an inventory of ECVs (WGClimate, 2017) from various agencies around the world. The degree of maturity of uncertainty information is considered as a contributing criterion for overall maturity of the data production system (Schultz *et al*., 2016)~~.~~

Figure 2: Examples of global, national and agency level entities with documented emphasis on UC/UQ.

The participation in the aforementioned global organizations is much broader than it would appear on the surface. For example, a broad set of organizations contribute to the activities of the IPCC with over 800 contributing authors from many countries around the world. The synthesis report (IPCC, 2014) lists 51 individuals from 24 different countries as the "Core Writing Team". Likewise, over 20 organizations around the world have contributed 496 entries to the ECV inventory mentioned above.

Many agency programs in the U.S. owe their existence to legislative and executive mandates (e.g., 91st U.S. Congress, 1970; 94th U.S. Congress, 1976) as well as internal frameworks which reflect various science and application needs. An example of a Congressional mandate to improve accuracy in predictions, and reduce uncertainty by implication, is the Weather Research and Forecasting Innovation Act (115th U.S. Congress, 2017) that initiated an advanced weather forecasting program in NOAA. This mandate calls for increasing the lead-time for tornado warnings among others. Inherently, the forecasting of weather involves quite a bit of uncertainty due to spatial and temporal variations in model scale, atmospheric physics and parameterization. The resolution of such questions is in part responsible for the race to improve weather models.

Federal agencies and standards bodies have also recognized the importance of uncertainty quantification. The Guide for Uncertainty in Metrology (JGCM, 2008) establishes guidelines for evaluating and expressing measurement uncertainty. The National Research Council recently published a report on Verification, Validation and Uncertainty Quantification (VVUQ) of complex models (NRC 2012) that identifies the importance and best practices for VVUQ across a range of disciplines, including remote sensing and Earth Sciences.

A series of European Union (EU) projects has considered quality assurance, including uncertainty aspects, for Earth observation geophysical products and essential climate variables (e.g., http://www.qa4ecv.eu). The EU H2020 project, FIDUCEO (www.fiduceo.eu) demonstrated a metrological framework for including uncertainty in level 1 radiance products as a basis for propagating uncertainty across satellite data processing levels (Merchant *et al*., 2019; Mittaz *et al*., 2019). From 2006-2009, the FP6 INTAMAP project (Pebesma *et al*., 2011) developed automated real-time interpolation services for critical environmental variables such as radiation, with particular attention to calculation and communication of uncertainties associated with the interpolated maps. Uncertainty measures were standardized and exchanged in a machine-readable fashion using UncertML (Williams *et al*., 2009). In the 7th Framework Programme (FP7) UncertWeb project (Bastin *et al*., 2013) these standardized approaches to exchange of uncertainty information between models were extended to support a framework in which uncertainty-aware web-services underpinned workflows for integrated environmental modeling. Another FP7 project, GeoViQua (http://www.geoviqua.org/Overview.htm) developed rigorous data quality and provenance representations specifically for the GEOSS Common Infrastructure, including a GEOLabel which has been adopted to give visibility to dataset compliance with the GEO Data Management Principles (http://geolabel.info/facets.htm). Within the European Space Agency (ESA), uncertainty characterization has received close attention within the Climate Change Initiative (cci.esa.int). Eight general principles for inclusion of uncertainty information in climate data records were distilled from experience across a diverse set of ECVs, as summarized in Merchant *et al*. (2017). ESA is supporting the establishment of *in* *situ* Fiducial Reference Measurements across (presently) 8 key measurement types to help improve the SI traceability, validation and long-term stabilization of Earth Observation (EO) based geophysical records (https://earth.esa.int/web/sppa/activities/frm). National metrology services in Europe are collaborating within a framework called MetEOC to support the metrologically sound use of EO data for climate applications (http://www.meteoc.org).

Federal agencies in the U.S. are governed by information quality guidelines issued by the Office of Management and Budget (OMB) in response to the U.S. Information Quality Act (US Public Law 106-554 2001, Section 515), and are required to issue their own, agency-specific, implementing guidelines for ensuring the quality of disseminated information. A prime example of this implementation is for instances in which scientific conclusions are considered highly influential (i.e., having high societal and/or economic impact), thereby requiring federal agencies to be extra-vigilant in ensuring the quality of information that has been publicly disseminated. For instance, the NASA guidelines (NASA, 2002) state: "NASA requires a higher standard of quality for information that is considered influential. Influential scientific, financial, or statistical information is defined as NASA information that, when disseminated, will have or does have a clear and substantial impact on important public policies or important private sector decisions." Ensuring a high standard of information quality implies communication of information about errors and uncertainties to users. The National Climate Assessment (NCA) Reports are considered Highly Influential Scientific Assessments (HISA) and follow a rigorous set of rules about information quality. Expression of uncertainties with each of the findings in the reports is a key requirement. As indicated in the Climate Science Special Report, which is the Volume 1 of the Fourth NCA (NCA4), the uncertainties are expressed using two metrics (Wuebbles *et al*., 2017):

- "**Confidence** in the validity of a finding based on the type, amount, quality, strength, and consistency of evidence (such as mechanistic understanding, theory, data, models, and expert judgment); the skill, range, and consistency of model projections; and the degree of agreement within the body of literature.

- **Likelihood,** or probability of an effect or impact occurring, is based on measures of uncertainty expressed probabilistically (based on the degree of understanding or knowledge, e.g., resulting from evaluating statistical analyses of observations or model results or on expert judgment)."

This is consistent with the IPCC guidance referred to above (Mastrandrea *et al.,* 2010).

Regarding HISA, the NASA guidelines state: "NASA requires the highest standard of quality for publication of information that is considered highly influential scientific assessments, which are a subset of influential scientific information delineated in the OMB Bulletin. A scientific assessment is an evaluation of a body of scientific or technical knowledge that typically synthesizes multiple factual inputs, data, models, assumptions, and/or applies best professional judgment to bridge uncertainties in the available information." Other U.S. agencies have developed guidelines covering information quality as well, calling out uncertainty explicitly (EPA 2002; NOAA 2014).

Agency programs generally reflect the needs of the scientific and applications user communities, and a fundamental requirement is to understand the fitness of the data for their specific user needs. Inclusion of uncertainty information along with the measurements or derived parameters is essential for determining fitness for use. As such, many measurement programs require specification of error bounds for instrument measurements and require that quality information including uncertainty be included along with derived data products. As a prime example, targeted error bounds to meet scientific performance requirements of NASA's Earth observing missions are required to be specified within the Level 1 Requirements; a fairly recent example of this can be found for the ICESat-2 Mission (NASA, 2013). Other situations may call for competitive science proposals to target UQ and UC research for specified historical data records. For example, NASA had a call for proposals (NASA, 2010) for analysis of uncertainty in Earth System Data Records (ESDRs). Other NASA programs such as Making Earth System Data Records for Use in Research Environments (MEaSUREs) include a requirement to "Characterize uncertainties and quantify errors associated with the proposed ESDRs" (NASA, 2012). Similarly, NOAA requires producers of Climate Data Records (CDRs) to provide both the analytic details of their uncertainty estimation as well as affirmation of their application for each extension of the time series records.

While the importance of UQ and UC is recognized by agency programs, a consideration is how to spend the limited budgets – whether to estimate uncertainty, improve resolution, latency, update frequency, validation, access interface, quality assurance/quality control (QA/QC), metadata and documentation, etc. or add new features, extensions, or whole new products. Clearly, the answers to these concerns vary depending on the user community interests, existing user uptake of UQ and UC information, as well as the priority for and impact of the data products under consideration. Some of the other programmatic considerations are that often uncertainty information is ignored or misunderstood by users, and that it can be difficult to convince agencies to invest funds into UQ/UC when competed against other, and often more urgent, user needs. From a space-based remote sensing perspective, the majority of budgeted expenditures to support such missions has traditionally been focused on ensuring successful design, deployment, and operational function of the platform and sensor(s) for a particular mission. With improved and lower-cost launch vehicles and space-platform/sensor technologies (e.g., smallsat technology), it is a reasonable expectation that increasing opportunities will arise to support more comprehensive and standardized UQ/UC activities for Earth observing missions.

2.4 Observational

This section intends to explore and discuss the common, foundational approaches of UQ, UC, and the overall utilization of uncertainty information throughout the primary approaches to Earth observation. Throughout this section, the term "uncertainty" may be used in a broad interpretive context, from which we rely on the more specific definitions as established in the preceding Mathematical section; all other terminologies and concepts introduced in this section are generally unique to the observational perspective, and as such, those will be defined and/or explained in contextual detail. Many approaches exist for observing the Earth system's numerous components. These fall into one of two categories: point-based studies--invariant in space but not in time (e.g., Eulerian Specifications) – and those that conduct observations varying in *both* space and time (e.g., Lagrangian Specifications). Observations may either take place from within a system (i.e., *in situ*) or from a remote observation platform (i.e., remote sensing). UC is often tailored and dependent upon the artifacts particular to the specific approach of the observing technique used.

Satellite observations of geophysical quantities are, of course, remotely sensed estimates that are obtained through retrieval algorithms that invert the information available to a sensor (usually radiance across a number of wavelengths) to infer the quantities of interest[1]. Algorithms may use a combination of physical simulations, parameterizations and/or empirical relationships. They may be trained on *in situ* measurements referred to as ground "truth". The term "truth" in this expression is often a first-order approximation of the valid state of a physical system and the physical quantities that comprise that system, is commonly used with an unprescribed degree of both practical and interpretive latitude, and is typically characterized with its own level of uncertainty that is directly associated with the observations (or lack thereof) used to provide the requisite approximation which facilitates the calibration and validation (Cal/Val) of retrieval algorithms. It is a common practice for uncertainty estimation in algorithm results to be co-designed with initial algorithm determination. Any algorithm development on the basis of adequately comprehensive simulations can establish an associated uncertainty estimate (e.g., Merchant & Embury, 2014), and the uncertainty model will likely be able to discriminate between either more or less uncertain retrievals when this approach is taken. An often-used alternative has involved estimating generic uncertainties for retrieval algorithms from validation activities. This alternative approach, although more simplistic and less costly, has been shown to provide a more limited understanding of the variability of uncertainty in the retrieval process; more adequate quantification of retrieval uncertainty requires detailed knowledge of the uncertainty in the ground "truth" and of "point-to-pixel" differences (Immler *et al*., 2010). For bounded observations (e.g., wind speed must be greater than zero) knowledge of uncertainty is needed to explain the artificial appearance of large biases near the boundary (Freilich, 1997). Some algorithm retrieval methods have well established uncertainty estimates associated with their formulation – for example, optimal estimation and similar techniques (Rodgers, 2000), although these tend not to include uncertainties arising from the first validity/classification step and require good knowledge of uncertainty in their inputs for the output uncertainties to be realistic. Such methods also enhance understanding of the propagation of uncertainty throughout the retrieval process and understanding the constituents of the retrieval system and algorithms that provide varying degrees of uncertainty.

[1] A typical high-level structure would involve a first step towards establishing the suitability of the collected satellite data (at Level 1) for a retrieval algorithm (for example, cloud screening), followed by applying the inversion algorithm to the suitable cases (retrieving the Level 2). Additional processing such as gridding may subsequently occur, creating Level 3 data.

The uncertainty in a retrieval arises from the effects of uncertainty in the input quantities to the retrieval, which are the satellite measured values (such as radiances) and the retrieval parameters (all retrievals have some, even if implicit). These input-quantity uncertainties propagate to the result through a model imparting additional uncertainties, and this can be estimated by well-established conventional error analysis or Monte Carlo simulation methods (JGCM, 2008). The theoretical underpinning of uncertainty analysis is long established, but proper application in Earth observation has been limited. The development of uncertainty estimates for many algorithms has been hindered by the typical lack of uncertainty information in the input products, including the Level 1 data curated by space agencies, where Level 1 data is characterized as time-ordered, geospatially referenced and calibrated engineering data that is generated prior to being transformed into geophysical quantities, the latter of which is referred to as Level 2 data. Level 1 data should routinely be distributed with enough uncertainty information at a granule or pixel level to support useful propagation of uncertainty through the algorithms which use them to derive geophysical quantities as expressed by Level 2 data; at present this is rarely done (an example is Gorroño *et al*., 2017). This would enable algorithm developers to more efficiently and robustly develop and provide useful uncertainty estimates, and would be a significant step towards a more metrologically sound practice of Earth Observation (Mittaz *et al*., 2019).

Regarding errors in input data, one potential source of 'uncertain ground-truth' is crowdsourcing of field observations, photos or online image interpretation (Fonte *et al*. 2015; e.g., https://www.cocorahs.org/ or https://www.geo-wiki.org/). These approaches can boost the density and representativeness of validation points, but variation in individual contributors' skill can lead to different interpretations at the same location. Such inconsistencies can be tackled by requiring multiple user classifications at each sample location, or by eliciting an estimate of confidence from the volunteer along with their label (See *et al*., 2013). In the case of a rapidly changing variable (e.g., daily accumulated precipitation), the uncertainty related to volunteer observations can be far less than sampling errors associated with low Earth orbit satellites.

Case study example: In 2010, the NASA/NOAA Sea Surface Temperature (SST) Science Team prepared an SST Error Budget white paper (Cornillon *et al*., 2010 and summarized in Wu *et al*., 2017) detailing the various contributions to the SST error budget from Level 1 through Level 4 products. Although this error budget was prepared for SST products, the resulting framework applies equally well to most other geophysical parameters derived from satellite-borne sensors. A particularly important observation which emerged from this effort is that errors introduced to the SST values are often characterized by two spatial scales - a pixel-to-pixel scale resulting from instrument errors and classification errors and a longer wavelength scale resulting from atmospheric contamination. This distinction becomes increasingly important as the resolution of the products increases - for products with a sampling interval greater than, say 25 km, the pixel-to-pixel scale and the scale associated with atmospheric contamination are similar. For higher resolution (i.e., smaller sampling interval) products this is not the case. The reason that the distinction between these scales is important is that gradients of the derived products are becoming of increasing interest to, at least in the ocean, sub-mesoscale studies and air-sea coupling (Shi and Bourassa, 2019). The long wavelength variability introduced by atmospheric contamination has little impact on the small-scale gradient in high resolution spatial fields, hence the standard error measures, which include contributions from both scales, tend to overestimate the error in the gradient field. One would be led to believe that SST fields obtained from most infrared sensors are inadequate to study the gradient field for gradients weaker than order 0.3 degrees K/km - the vast majority of SST gradients in the ocean. This is in fact not the case as estimates of pixel-to-pixel errors needed to determine this are generally not available; what is more prevalent in such cases are the bias and variance of the pixel values, generally when compared with *in situ* observations. These tend to range from 0.3 to 0.5 degrees K for most SST products obtained from infrared sensors, while the pixel-to-pixel uncertainty appears to range from 0.05 to 0.2 degrees K for the same suite of sensors (Wu *et al*., 2017). Although this discussion is presented in the context of SST observations, we believe that it applies equally well to other parameters obtained from high resolution (<10 km) satellite-borne sensors for which atmospheric corrections are critical.

2.4.1 Cal/Val

Where uncertainty estimates are provided with Earth science measurements including satellite retrievals, the process of validation (comparison of measurements with independent ground truth) informs both the estimates and their estimated uncertainties, provided some understanding is available about the ground-truth uncertainties and the bridging of scales between different measurements (Immler *et al*., 2010).

Estimates of instrument measurement uncertainty at the beginning of a new satellite mission are based on models, pre-flight instrument characterization and surrogate data (laboratory experiments, heritage instruments and *in situ* data, including tower, aircraft and balloon data). Alternatively, satellite to satellite comparisons can be extremely effective for wide swath instruments that have temporally and spatially coincident observations several times a day. Such a situation produces vastly more colocations than seen from point *in situ* observations (Wentz *et al*., 2017). On-orbit calibration uncertainty is dependent on the characteristics and knowledge of the accuracy of any internal calibration sources (e.g. blackbody, solar diffuser) as well as external calibration sources (e.g. moon and sun) and *in situ* data. Pre-launch error budgets based on all of these factors are developed based on expert knowledge and prior experience, including modeling, and updated during the design, pre-launch testing and on-orbit checkout phase. Many of the calibration uncertainty requirements are driven by the needed accuracy of the derived geophysical parameters, including requirements for spatial resolution, sampling and accuracy, as well as temporal resolution, sampling, and stability.

After the post-launch check-out phase, the mission calibration team monitors the instrument using on-orbit calibration sources and by comparing the Level 1 data to independent *in situ* measurements and data collected from other remotely-deployed instruments. Note that each of these independent calibration sources will have their own uncertainty in addition to those embedded in the inter-comparison process itself. For example, spatial variability causes

Root Means Square (RMS) differences (often used as a first-order quantification metric of uncertainty but assumes a Gaussian error distribution) to grow as spatial constraints in colocation are relaxed. Feedback from retrievals of Level 2 geophysical parameter retrievals are also key to understanding the Level 1 data uncertainty. In addition, modeling activities (through assimilation or direct comparison with model output) may be used to determine the Level 1 data uncertainty. In some cases, these uncertainty estimates will be estimated as a function of some instrument characteristic (e.g. scan angle) and/or as a function of time on-orbit. In addition, there may be periods of time when the instrument is in a non-nominal state (e.g. attitude maneuvers, black-body warm-up/cool-down) when the uncertainty could be larger. Furthermore, even when the platform/instrument is in a stable state, there may be properties outside of the control of the instrument(s) that could further impact uncertainty, such as fluctuations in terrestrial radiation within the wavelength of the instrument (i.e., noise floor variability), cloud/aerosol/rain contamination, and differences in atmospheric attenuation due to fluctuations in the vertical column of air mass (i.e., controlled by mostly temperature and water vapor). Due to the complexities that comprise uncertainty in remote sensing, the initial uncertainty estimates are generally not well characterized *a priori*. Through careful analysis of more continuous observations over time, the knowledge as well as confidence in the uncertainty estimates typically improves, and it is through this iterative process that instrument and algorithm calibration teams are expected to provide their best estimates in final products, beginning with improvements in the Level 1 calibration and associated uncertainties; when various reprocessing events occur over time, these estimates are typically updated.

The uncertainty requirements for Cal/Val are usually very different from those for wider applications. For Cal/Val, it is usually preferred to reject all data for which the observations are overly uncertain (i.e., outside of the threshold of science quality as typically defined by science teams). For example, many remotely sensed surface observations are adversely impacted by rain. Consequently, such types observations through rain are often removed from intercalibration datasets, except where rain is the observable and thus correctable in a manner that minimizes the impact of rain on the targeted observation (Hilburn *et al*., 2006; Stiles and Dunbar, 2010; Weissman *et al*., 2012). In many cases this quality control is done to remove concerns about systematic errors; however, it should be kept in mind that most uncertainties are estimated for typical conditions rather than adverse conditions. If biases related to adverse conditions can be accounted for, then additional uncertainty can be added for adverse conditions. Such efforts are complicated by the difficulty in determining such systematic errors, and in obtaining a sufficiently large sample of observations from adverse conditions.

Case study example: The Land Product Validation (LPV) sub-group of the CEOS WGCV was established in 2000 to provide an international forum for developing protocols for quantifying accuracies of Land bio- and geophysical products (Justice *et al*., 2000). The term 'validation' is now commonly used by the observational community to describe the process of quantifying the observational product accuracy, and inversely uncertainty. The first of several protocols published by this group describes the 'validation good practices' for the global Leaf Area Index (LAI) products (Fernandes *et al*., Version 2.0.1, Aug. 2014). This document first defines the measurement and other key terms related to validation so that it is clear what is being measured and how the measurements are made. It also describes how different comparison approaches and stratified spatial sampling can be used to quantify product uncertainty. It also discusses how the measurement frequency and continuity affects the knowledge of the measurement of LAI. This detailed approach to validation of Land products is an excellent example of how the international research community can develop a common understanding of uncertainty for a particular global measurement. This approach has since been replicated for other Land products.

2.4.2 Product Development

Earth science and satellite data products should be available to users already "uncertainty quantified", that is, with useful uncertainty information included in the products and their metadata. The preceding User Perspectives section goes into more detail on examining a variety of use cases and studies on the considerations behind what makes uncertainty information useful. This could be as simple as a single number in cases where uncertainty (or fractional uncertainty) is adequately invariant across the data set, but a common situation would be that uncertainty is variable, and needs to be provided per datum (i.e., for each unique observation) or otherwise parameterized. This is particularly important with large and heterogeneous databases, such as OpenStreetMap, Weather Underground or the Global Biodiversity Information Facility; these databases are aggregated from multiple sources and whose individual granules or observations are usually filtered, selected and assembled into many different 'datasets' depending on user requirements. In such cases, a single summary of uncertainty is of little value, and per-datum quality records offer more flexibility.

Uncertainty of data products is not always probabilistic or expressed in a quantified manner, but may be constrained to only include qualitative or categorical factors which can be tested against standard assertions of quality to assess the usefulness of each individual record. An example of such an approach is the list of assertion codes used by *Atlas of Living Australia* to label georeferenced data according to their geographic and attribute validity (Belbin *et al*., 2018). In another example, Bell *et al*. (2015) combine data from two volunteered weather station networks with authoritative measurements from the UK Met Office's official observing stations, identifying bias and sensor errors on the 'unofficial' data which permit it to be corrected to a sufficient standard to be used to improve interpolations and local-scale forecasts. This is a case where, for each volunteered dataset as a whole, there is no meaningful single summary of uncertainty because of the wide variation between instruments and the microclimates in which they are placed. However, characterizing the particular uncertainty at each amateur weather station permits a reduction in the variability across the network to a more comparable residual uncertainty at each sampling location. In addition to quantifying uncertainty, as defined more thoroughly in the preceding Mathematical section, there may be a need to provide information in products characterizing the correlations of errors (Bell *et al*., 2015). This could arise when a multivariate dataset is jointly estimated, such that errors in one variable are linked to errors in others. Where a product may be used on a number of spatio-temporal scales, users are likely to prefer to average or otherwise aggregate data. Propagation of uncertainty correctly to coarser scales requires information from the quantifying or parameterizing of the spatio-temporal error correlation in the product to be available. More complete information about uncertainty and error correlation is commonly expressed through the error covariance matrix of the data. However, providing this for a product of any significant size is infeasible. Therefore, fairly complex choices have to be made regarding what it is most important to provide and how to do so (e.g., Merchant *et al*., 2017).

3. Discussion

All of the above perspectives demonstrate a collective interest and need for UQ/UC information for ESDRs, spanning a variety of observational data and approaches to data uncertainty. Foundationally, the mathematical perspective establishes a common and robust framework for UQ and UC, providing clarity on often misconstrued and mis-applied metrics and definitions for all of the constituents that define and characterize the overall uncertainty for a QOI. The mathematical approach is generally applied to ESDRs through both data production and in Cal/Val phases. However, there are significant differences in the way geophysical algorithm developers (i.e., those developing transfer functions and geophysical model functions) and Cal/Val teams apply the mathematical approach. Examples of differences are in:

- Whether the PDF is considered;
- Assumptions of Gaussian vs. non-Gaussian;
- When to start UQ/UC in the data processing workflow (i.e., Level 1, Level 2, etc.);
- Whether or not to provide uncertainty estimates for every observational data point;
- Whether or not to include uncertainty information as part of the primary ESDR dataset; and
- Consistency in representation of uncertainty information in data, metadata and documentation.

UQ/UC implementation is generally applied to data once it has been transformed into geophysical quantities, or Level 2 data. Rarely is UQ/UC done for Level 1 data, but it is nonetheless considered to be a critical step in understanding the propagation of uncertainty from Level 1 to Level 2. Since many of the calibration uncertainty requirements are driven by the needed accuracy of the derived geophysical parameters (i.e., requirements for spatial resolution, sampling and accuracy, as well as temporal resolution, sampling, and stability), this can lead to differences in the approach for instrument design, which is much more complex for space-based remote sensing than for *in situ* observational platforms. *In situ* platforms are generally assumed to be ubiquitous enough for there to be a higher level of confidence and consistency based on the sheer numbers of instruments deployed throughout the globe; this assumption often leads to a very optimistic presumption that all *in situ* observation networks are consistently calibrated, continually validated and well-coordinated internationally, which is generally not the case across disparate observation networks. Even where *in situ* platforms are large in number, the sparsity of coverage coupled with lack of cross-network coordination leaves much to be desired with regard to a more coordinated, assessment sharing of uncertainty for more uniform UQ and UC assessment. Even though the number of space-based platforms are few and far between compared to *in situ*, spatial and temporal coverage for even a single space-based platform is enormous. However, in many cases the space-based deployments represent brand new technology in which the QOI estimated by the deployed instrument at pre-launch can only be simulated through models or estimated by proxy with similar instruments; while there may be more cross-network and international coordination in the space-based remote sensing domain, UQ/UC challenges still remain, namely for cutting edge technological design of the instrumentation intended to either improve upon existing remotely sensed QOI estimation or to estimate an entirely new QOI never before acquired from space. Aside from technological challenges, there are challenges associated with the prescribed deployment (timing and orbital placement) of select sensors/platforms that are often not part of a coordinated constellation to provide more adequate spatial and temporal coverage.

As a result of this higher level of complexity for space-based remote sensing compared to *in situ* observations, the UQ and UC during the pre-launch phase often takes a back seat to the overall health and quality of the instrumentation itself, and *a priori* uncertainty estimation is often the easiest solution that allows for pre-launch constraints to be met. In the post-launch phase, instrument health is still a consideration, but the Cal/Val period is typically when UQ/UC is assessed and leveraged to help ensure that the instrument coupled to the downstream data processing is meeting the mission requirements. Since there is no standardized implementation for UQ/UC across the disparate disciplines and domains of Earth observation, notably within the Cal/Val phase for space-based remote sensing, particularly at the Level 1 stage of data production, there is plenty of room for diverging approaches. What often results in such cases is a prescribed implementation for UQ/UC that is determined by a select group of experts who specialize in that particular domain unique to the instrument and platform; that guidance is often what is sought after by data producers and Cal/Val teams. Data producers tend to put a lot of thought into leveraging UQ/UC for their own iterative algorithm development (e.g., improving the performance accuracy and removing bias produced by the geophysical model function), but this does not always translate into making the UQ/UC information available for the data users. The traditional approach in such cases is publishing a Cal/Val paper to summarize the overall data uncertainty and performance of the algorithm(s). Only in recent years has there been a stronger push to make the UQ/UC information more accessible, such as including it in the data files.

Programmatically, there seems to be a growing trend in requesting or even requiring the presence of uncertainty information for specific types of ESDRs, particularly those that contain an ECV. The CEOS WGCV is arguably leading the way in developing these types of requirements in the international Earth observation community, at least from a space-based remote sensing perspective. Sub-orbital and *in situ* observations have their own smaller niche communities, and it is less clear how well organized these groups are with regard to developing international standards and requirements for UQ and UC in their data collections. The ESIP IQC appears to be the first instance of any concerted effort to bring together the space-based remote sensing, sub-orbital and *in situ* communities in one common venue to discuss these topics. While this is encouraging to see, it is also recognized that there remain differences in thought and approach as a result of the historical “stove-piping” of the information that is exchanged within these observational data communities. From the U.S. national perspective, there are numerous laws on the books at the federal level that have promulgated as a result of the need for more open data and open science. While this can be perceived as progress at least legislatively, there appears to be a lot of room for interpretation and application of these laws at the agency level (e.g., NOAA, NASA, DOE, etc.). The majority of the progress toward UQ/UC implementation and standardization at the U.S. agency level is currently driven by cross-agency and international organizations, such as CEOS, ESIP, OGC, and others.

In examining the user perspective, we find that communication and interpretation of UQ/UC information go hand-in-hand, and that differences in communication styles may likewise lead to differences in interpretation. For instance, many times uncertainty is communicated using error bars, which often do not include a probability-based estimate of the confidence that can be attributed to those error estimates, thus leading to what may be an overconfident interpretation of the uncertainty. This consistent disconnect may help explain why studies have demonstrated a considerable lack of understanding among users of UQ/UC information of how probability relates to uncertainty. Standards for representing UQ/UC information within data files and as metadata are quickly emerging and gaining steam in the Earth science data informatics and stewardship communities, particularly with regard to ISO 19157 and OGC's UncertML standards. The area for opportunistic community adoption in this regard is with the data producers themselves, as these informatics approaches tend to be focused more on communicating UQ/UC in a post-hoc containerized sense, rather than something that is directly integrated with the data production workflow. Just as important to these standards are the even more recent emergence of analytics and visualization tool kits and software packages that provide a relatively consistent user experience in both analyzing and visualizing UQ/UC information. The sentiment at this stage is that these toolkits and software packages are still relatively new to the data informatics and data science world, so it is too early to assess user uptake, let alone, whether or not they are operationally mature enough to be used directly by data producers and Cal/Val teams. Despite these early stages of maturity in common solutions, we expect significant value and impact to the users if these informatics approaches to communicating UQ/UC were to be more thoroughly evaluated for wider adoption by data producers to help assess the potential for direct integration with the data production and publication workflow.

4. Conclusions

The manner in which uncertainty information is derived, characterized, disseminated, and ultimately consumed and interpreted by the end user is quite diverse, yet there are shared interests and growing commonalities of approach. While this paper primarily focuses on statistics-based techniques of evaluating uncertainty information, there are also knowledge-based techniques that are regularly used. But, for the scope of this paper, such knowledge-based techniques were not earnestly explored. Regardless of the observation, whether remotely sensed or *in situ*, UQ/UC is ubiquitously recognized as a fundamental step toward providing vital information that helps characterize the scientific quality of the Earth science data being collected, evaluated, assimilated, and applied. Despite the differences in the approach of estimation, interpretation, dissemination, and utilization of uncertainty information, there are compelling opportunities within the cross-domain and cross-discipline Earth observation communities to work more closely together toward adopting a more uniform implementation of this valuable information. Effective utility of UQ/UC information depends primarily on effective communication and dissemination of that information. When it comes to expressing uncertainty in a probabilistic framework, very few users (perhaps depending on their level of expertise) have a proper understanding of that framework to correctly interpret the data; this user limitation may be overcome by improved visualization coupled with consistent packaging of the UQ/UC information, such as at the pixel level (i.e., 1:1 uncertainty:observation availability) and with the inclusion of ancillary information such as probability-based confidence levels.

Since the summer of 2017, the ESIP IQC has undergone an extensive collection effort of expert-level information focusing on key aspects of scientific quality of Earth science data and addressed key concerns regarding uncertainty. This collection of information has been summarized and discussed at a foundational, expository level that is intended to be ubiquitously applicable to producers, managers, and users of data across all domains of Earth science observation; such information is also meant to be helpful for fundamental researchers and decision/policy makers. This paper provides foundational information on the ways in which Earth science data uncertainty information is estimated, characterized, and ultimately communicated. A natural and expected next step beyond this white paper will be a continued analysis of historical and ongoing practices (as discovered through this paper) to provide a set of recommended best practices that we expect will serve the interests of all interested parties noted above. Through these efforts, it is the hope of the ESIP IQC that more cross-domain and cross-discipline coordination can be achieved for the continued advancement of science through Earth observation and the study of Earth's complex processes.

Acknowledgements

a. Primary, and co-authorship contributions:

   i. Introduction section: this was championed by David Moroni and was supported with secondary material contributions provided directly by Hampapuram Ramapriyan, Ge Peng, Otis Brown, and Chung-Lin Shie.
   ii. Mathematical section: this was championed by Jonathan Hobbs and was supported with secondary material contributions made by the following co-authors: Jessica Matthews, David F. Moroni, and Mark A. Bourassa.
   iii. User section: this was championed by Robert Downs and was supported with secondary material contributions made by the following co-authors: Lucy Bastin, and Ivana Ivánová.
   iv. Programmatic section: this was championed by Hampapuram Ramapriyan and was supported by contributions made by the following co-authors: Otis Brown, Justin C. Goldstein, Ivana Ivánová, Christopher Merchant, David Moroni, Ge Peng, Jeffrey L. Privette, and Benjamin Smith
   v. Observational section: this was championed by Justin Goldstein and was supported with secondary material contributions made by the following co-authors: Lucy Bastin, Kenneth Kehoe, Peter Cornillon, Christopher Merchant, Hampapuram Ramapriyan, Robert Wolfe, Mark A. Bourassa, Aneesh C. Subramanian, and Ivana Ivánová.
   vi. Discussion section: this was championed by David Moroni and was supported with secondary material contributions provided directly by Hampapuram Ramapriyan and Ge Peng.
   vii. Conclusions section: this was championed by David Moroni and was supported with secondary material contributions provided directly by Hampapuram Ramapriyan and Ge Peng.

b. Pre-white paper information exchange through discussions and presentations:

   i. Amy Braverman, Jet Propulsion Laboratory, California Institute of Technology.
   ii. Carol Anne Clayson, Woods Hole Oceanographic Institution.
   iii. Matthew Plumlee, Northwestern University
   iv. Isla Simpson, National Center for Atmospheric Research

c. Rose Borden is credited with the final copy editing and styling adjustments made just prior to publication.

d. All authors have reviewed and approved the final version.

e. Institutional statements of acknowledged support:

   i. This work was supported by NOAA grants/contracts to the:
      - North Carolina State University, Raleigh, NC
      - Riverside Technology, Inc.
   ii. This work was supported by NASA grants/contracts to:
      - The Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA

- Science Systems and Applications, Inc., Lanham, MD.
- The University of Maryland, Baltimore County
- The Center for International Earth Science Information Network, Columbia University

iii. U.S. Government sponsorship acknowledged for all institutional affiliations funded under various government grants/contracts.

References

91st U.S. Congress. (1970). National Environmental Policy act of 1969, Public Law No. 94-265 (01/01/1970). Retrieved from https://www.gpo.gov/fdsys/pkg/STATUTE-83/pdf/STATUTE-83-Pg852.pdf.

94th U.S. Congress. (1976). Fishery Conservation and Management Act of 1976, Public Law No. 94-265 (04/13/1976). Retrieved from https://www.congress.gov/bill/94th-congress/house-bill/200?r=1.

115th U.S. Congress. (2017). Weather Research and Forecasting Innovation Act of 2017, Public Law No: 115-25 (04/18/2017). Retrieved from https://www.congress.gov/bill/115th-congress/house-bill/353.

Bastin, L., Cornford, D., Jones, R., Heuvelink,G. B. M., Pebesma, E., Stasch, C., Nativi, S., Mazzetti, P. & Williams, M. (2013). Managing uncertainty in integrated environmental modelling: The UncertWeb framework. *Environmental Modelling & Software, 39*, pp. 116-134. https://doi.org/10.1016/j.envsoft.2012.02.008.

Belbin, L., Chapman, A., Wieczorek, J., Zermoglio, P., Thompson, A., & Morris, P. (2018). Data Quality Task Group 2: tests and assertions. *Biodiversity Information Science and Standards 2*: e25608. https://doi.org/10.3897/biss.2.25608.

Bell, S., Cornford, D., & Bastin, L. (2015). How good are citizen weather stations? addressing a biased opinion. *Weather*, *70*(3), pp. 75-84. https://doi.org/10.1002/wea.2316.

Committee on Earth Observing Satellites. Working Group on Calibration & Validation. http://ceos.org/ourwork/workinggroups/wgcv/.

Cornillion, P., Castro, S., Gentemann, C., Jessup, A., Kaplan, A., Lindstrom, E., & Maturi, E. (2010). SST Error Budget - White Paper. Retrieved from https://works.bepress.com/peter-cornillon/1/.

Cressie, N. (2018). Mission CO_2ntrol: A Statistical Scientist's Role in Remote Sensing of Atmospheric Carbon Dioxide. *Journal of the American Statistical Association*, *113*, 152-168. https://doi.org/10.1080/01621459.2017.1419136.

Environmental Protection Agency. (2002). Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility and Integrity of Information Disseminated by the Environmental Protection Agency. Retrieved from https://www.epa.gov/sites/production/files/2018-11/documents/epa-info-quality-guidelines_1.pdf

Fernandes, R., Plummer, S., Nightingale, J., Baret, F., Camacho, F., Fang, H., Garrigues, S., Gobron, N., Lang, M., Lacaze, R., LeBlanc, S., Meroni, M., Martinez, B., Nilson, T., Pinty, B., Pisek, J., Sonnentag, O., Verger, A., Welles, J., Weiss, M., & Widlowski, J.L. (2014). Global Leaf Area Index Product Validation Good Practices. Version 2.0. In G. Schaepman-Strub, M. Román, & J. Nickeson (Eds.), *Best Practice for Satellite-Derived Land Product Validation* (p. 76): Land Product Validation Subgroup (WGCV/CEOS). https://doi.org/10.5067/doc/ceoswgcv/lpv/lai.002.

Fonte, C., Bastin, L., See, L., Foody,G.. & Lupia, F. (2015). Usability of VGI for validation of land cover maps. *International Journal of Geographical Information Science*, *29*(7), pp. 1269-1291. https://doi.org/10.1080/13658816.2015.1018266.

Freilich, M. H. (1997). Validation of vector magnitude datasets: Effects of random component errors. *Journal of Atmospheric and Oceanic Technology*, *14*(3), pp. 695-703. https://doi.org/10.1175/1520-0426(1997)014%3C0695:VOVMDE%3E2.0.CO;2.

Gardiner, M. M., Allee, L. L., Brown, P. M., Losey, J. E., Roy, H. E., & Smyth, R. R.. (2012). Lessons from lady beetles: accuracy of monitoring data from US and UK citizen-science programs. *Frontiers in Ecology and the Environment, 10*(9), pp. 471-476. Retrieved from https://nora.nerc.ac.uk/20992/1/N020992JA.pdf.

Global Observing System for Climate (GCOS). (2016). The Global Observing System for Climate: Implementation Needs. GCOS-200. Geneva, Switzerland. pp. 315, Retrieved from https://library.wmo.int/opac/index.php?lvl=notice_display&id=19838.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian Data Analysis*. Boca Raton, FL: CRC Press. ISBN 9781439840955 - CAT# K11900. Retrieved from https://www.crcpress.com/Bayesian-Data-Analysis/Gelman-Carlin-Stern-Dunson-Vehtari-Rubin/p/book/9781439840955.

Gorroño, J., Fomferra, N., Peters, M., Fascon, F., Underwood, C. I., Fox, N. P., Kiches, G., & Brockmann, C. (2017). A Radiometric Uncertainty Tool for the Sentinel 2 Mission. *Remote Sensing*, *9(*2), 178. https://doi.org/10.3390/rs9020178.

Group on Earth Observations. (2015). Data Management Principles Implementation Guidelines. GEO-XII, Document 10. Retrieved from http://bit.ly/GeossDmpImpGuide.

Hilburn, K. A., Wentz, F. J., Smith, D. K., & Ashcroft, P. D. (2006). Correcting active scatterometer data for the effects of rain using passive microwave data. *Journal of Applied Meteorology and Climatology, 45*(3), 382–398. https://doi.org/10.1175/JAM2357.1.

Hulley, G., Malakar, N., Freepartner, R. (2016). MODIS Land Surface Temperature Emmissivity Product (MxD21) ATBD Collection-6. *JPL Publication 12-17.* Jet Propulsion Laboratory:

California Institute of Technology. Retrieved from https://lpdaac.usgs.gov/documents/107/MOD21_ATBD.pdf.

Immler, F. J., Dykema, J., Gardiner, T., Whiteman, D. N., Thorne, P. W., & Vomel, H. (2010). Reference quality upper air measurements: guidance for developing GRUAN data products. *Atmospheric Measurement Techniques*, *3*, pp. 1217-1231. Retrieved from https://doi.org/10.5194/amt-3-1217-2010.

International Organisation for Standardisation. (2013). ISO 19157:2013 Geographic information - Data Quality. Geneva: ISO. Retrieved from https://www.iso.org/standard/32575.html.

Intergovernmental Panel on Climate Change. 2014. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Geneva, Switzerland: IPCC. ISBN 978-92-9169-143-2. Retrieved from https://archive.ipcc.ch/pdf/assessment-report/ar5/syr/AR5_SYR_FINAL_Front_matters.pdf.

Joint Committee for Guides in Metrology (JGCM). (2008). Guide to the expression of Uncertainty in Measurement. Bureau International des Poids et Measures, 100:2008. Retrieved from https://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf.

Justice, C., Belward, A., Morisette, J., Lewis, P., Privette, J., & Baret, F. (2000). Developments in the 'validation' of satellite sensor products for the study of the land surface. *International Journal of Remote Sensing, 21*(17), pp. 3383-3390. https://doi.org/10.1080/014311600750020000.

Kinkeldey, C. (2014). A concept for uncertainty-aware analysis of land cover change using geovisual analytics. *ISPRS International Journal of Geographic Information, 3*(3), pp. 1122-1138. https://doi.org/10.3390/ijgi3031122

Kinkeldey, C., Maceachren, A. M., and Schiewe, J. 2014. How to assess visual communication of uncertainty? A systematic review of geospatial uncertainty visualization user studies. *The Cartographic Journal, 51*, pp. 372–386. https://doi.org/10.1179/1743277414Y.0000000099

Kinkeldey, C., MacEachren, A. M., Riveiro, M., and Schiewe, J. (2017) Evaluating the effect of visually represented geodata uncertainty on decision-making: systematic review, lessons learned, and recommendations. *Cartography and Geographic Information Science, 44*(1), pp. 1-21. https://doi.org/10.1080/15230406.2015.1089792.

Mastrandrea, M.D., Field, C.B., Stocker, T.F., Edenhofer, O., Ebi, K.L., Frame, D.J., Held, H., Kriegler, E., Mach, K.J., Matschoss, P.R., Plattner, G.-K., Yohe, G.W., & Zwiers, F.W. (2010). Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. Intergovernmental Panel on Climate Change (IPCC). Retrieved from http://www.ipcc.ch.

Meissner, T., Wentz, F., and Le Vine, D. (2017). Aquarius Salinity Retrieval Algorithm Theoretical Basis Document (ATBD), End of Mission Version. Retrieved from https://doi.org10.5067/DOCUM-AQR04.

Merchant, C. J. & Embury, O. (2014). Simulation and inversion of satellite thermal measurements. In: Zibordi, G., Donlon, C. J. and Parr, A. C. (eds.) *Optical radiometry for ocean climate measurements. Experimental methods in the physical sciences*, 47. Academic Press, pp. 489-526. https://doi.org/10.1016/B978-0-12-417011-7.00015-5.

Merchant, C. J., Holl, G., Mittaz, J. P. D. and Woolliams, E. R. (2019). Radiance uncertainty characterisation to facilitate climate data record creation. *Remote Sensing, 11*(5), p. 474. doi: https://doi.org/10.3390/rs11050474

Merchant, C. J., Paul, F., Popp, T., Ablain, M., Bontemps, S., Defourny, P., Hollmann, R., Lavergne, T., Laeng, A., de Leeuw, G., Mittaz, J., Poulsen, C., Povey, A. C., Reuter, M., Sathyendranath, S., Sandven, S., Sofeiva, V. F. & Wagner, W. (2017). Uncertainty information in climate data records from Earth observation. *Earth System Science Data*, *9*(2), pp. 511-527. https://doi.org/10.5194/essd-9-511-2017.

Mittaz, J., Merchant, C. J., and Woolliams, E. (2019). Applying principles of metrology to historical Earth observations from satellites. *Metrologia, 56*(3). https://doi.org/10.1088/1681-7575/ab1705.

Morgan, M. G., and Henrion, M. (1993). *Uncertainty: A guide to Dealing with Uncertainty in Qualitative Risk and Policy Analysis*. New York, NY: Cambridge University Press. https://doi.org/10.1002/(SICI)1099-0771(199606)9:2<147::AID-BDM199>3.0.CO;2-8.

Murphy, A. H., Lichtenstein, S., Fischhoff, B., & Winkler, R. L. (1980). Misinterpretations of precipitation probability forecasts. *Bulletin of the American Meteorological Society*, *61*(7), pp. 695-701. https://doi.org/10.1175/1520-0477(1980)061<0695:MOPPF>2.0.CO;2.

National Aeronautics and Space Administration (NASA). (2002). National Aeronautics and Space Administration Guidelines for Ensuring the Quality of Information. Retrieved from http://www.nasa.gov/sites/default/files/517756main_FINAL_NASA_guidelines.pdf

National Aeronautics and Space Administration (NASA). (2010). EARTH SYSTEM DATA RECORDS UNCERTAINTY ANALYSIS, NASA Solicitation under Research Opportunities in Space and Earth Sciences (ROSES 2010), Appendix A.32. Retrieved from https://smd-prod.s3.amazonaws.com/science-pink/s3fs-public/atoms/files/ESDRERR_2010_Solicitation.pdf

National Aeronautics and Space Administration (NASA). (2012). MAKING EARTH SYSTEM DATA RECORDS FOR USE IN RESEARCH ENVIRONMENTS, Solicitation under Research Opportunities in Space and Earth Sciences (ROSES 2012), Appendix A.41. Retrieved from https://smd-prod.s3.amazonaws.com/science-pink/s3fs-public/atoms/files/MEaSUREs_2012_Solicitation.pdf

National Aeronautics and Space Administration (NASA). (2013). The ICESat-2 Mission: Level-1 Requirements and Mission Success Criteria. Retrieved from https://icesat.gsfc.nasa.gov/icesat2/project/ICESat2_program_level_requirements.pdf.

National Research Council (NRC) Committee on Mathematical Foundations of Verification, Validation, and Uncertainty Quantification. (2012). *Assessing the Reliability of Complex Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification.* National Academies Press. https://doi.org/10.17226/13395.

Nguyen, H., N. Cressie, and A. Braverman. (2012). Spatial statistical data fusion for remote sensing applications. *Journal of the American Statistical Association*, *107*, pp. 1004-1018. https://doi.org/10.1080/01621459.2012.694717.

National Oceanic and Atmospheric Administration (NOAA). (2014). National Oceanic and Atmospheric Administration Information Quality Guidelines. Original Publication 2002. Last Revised October 2014. Retrieved from https://www.cio.noaa.gov/services_programs/IQ_Guidelines_103014.html.

Masó, J and Zabala, A (eds.). (2017). Testbed-12 Imagery Quality and Accuracy Engineering Report. Wayland: OGC. Retrieved from https://docs.opengeospatial.org/per/16-050.html.

Orr, J. C., J.-M. Epitalon, A. G. Dickson, J.-P. Gattuso. (2018). Routine uncertainty propagation for the marine carbon dioxide system. *Chemistry, 207*, pp. 84-107. https:doi.org/10.1016/j.marchem.2018.10.006.

Pebesma, E., Cornford, D., Dubois, G., Heuvelink, G. B. M., Hristopulos, D., Pilz, J., Stöhlker, U., Morin, G. and Skøien, J. O. (2011). INTAMAP: The design and implementation of an interoperable automated interpolation web service. *Computers & Geosciences, 37*(3), pp. 343-352. https://doi.org/10.1016/j.cageo.2010.03.019.

Ramapriyan, H., G. Peng, D. Moroni, and C.-L. Shie. (2017). Ensuring and Improving Information Quality for Earth Science Data and Products. *D-Lib Magazine, 23*(7/8). https://doi.org/10.1045/july2017-ramapriyan.

Rodgers, C.D. (2000). *Inverse Methods for Atmospheric Sounding: Theory and Practice*. World Scientific Publishing Co. Pte.Ltd. https://doi.org/10.1142/3171.

Sacha, D., Senaratne, H., Kwon, B. C., Ellis, G. and Keim, D. A. (2016). The Role of Uncertainty, Awareness, and Trust in Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics, 22*(1), pp. 240-249. https://doi.org/10.1109/TVCG.2015.2467591.

Schultz, J., Nunes, A., & Pechorro, E. (2016). ECV Inventory Questionnaire Guide, ECO-WGCL-EOPS-TN-16-0022, Version 2.4. Retrieved from http://ceos.org/document_management/Working_Groups/WGClimate/Documents/WGClimate_ECV-Inventory-Questionnaire-Guide_v2-2_Feb2016.pdf.

See, L., Comber, A., Salk, C., Fritz, S., van der Velde, M., Perger, C., Schill, C., McCallum, I., Kraxner, F., & Obersteiner, M. (2013). Comparing the quality of crowdsourced data contributed by experts and non-experts. *PLOS One 8*(7), e69958. https://doi.org/10.1371/journal.pone.0069958

Senaratne, H., Gerharz, L., Pebesma, E., & Schwering, A. (2012). Usability of spatio-temporal uncertainty visualisation methods. In J. Gensel, D. Josselin, & D. Vandenbroucke (Eds.), *Bridging the geographic information sciences* (p. 3-23). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-29063-3_1.

Shi, Q. and Bourassa. M.A. (2019). Coupling Ocean Currents and Waves with Wind Stress over the Gulf Stream. *Remote Sens. 2019, 11*(12), p. 1476. https://doi.org/10.3390/rs11121476.

Smith, R. (2014). *Uncertainty Quantification: Theory, Implementation, and Applications*. Philadelphia: Society for Industrial and Applied Mathematics. ISBN 10: 161197321X. ISBN 13: 9781611973211. https://isbn.nu/9781611973211.

Stiles, B.W., and Dunbar, S. (2010). A Neural Network Technique for Improving the Accuracy of Scatterometer Winds in Rainy Conditions. *IEEE Transactions on Geoscience and Remote Sensing, 48*(8), pp 3114-3122. https://doi.org/10.1109/TGRS.2010.2049362.

Theobald, E. J., Ettinger, A. K., Burgess, H. K., DeBey, L. B., Schmidt, N. R., Froehlich, H. E., Wagner, C., HilleRisLambers, J., Tewksbury, J., Harsch, M. A., & Parrish, J. K. (2015). Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. *Biological Conservation, 181*, pp. 236-244.
Retrieved from https://www.sciencedirect.com/science/article/pii/S0006320714004029.

Tweddle, J. C., Robinson, L. D., Pocock, M. J. O., & Roy, H. E. (2012). Guide to Citizen Science: Developing, implementing and evaluating citizen science to study biodiversity and the environment in the UK. NERC/Centre for Ecology & Hydrology.
Retrieved from https://openupsci.wixsite.com/citizensciencecourse/evaluation.

United Nations Framework Convention on Climate Change (UNFCCC). 2008: Reporting on global observing systems for climate, Decision 11/CP.13. Retrieved from https://unfccc.int/sites/default/files/resource/docs/2007/cop13/eng/06a02.pdf.

Williams, M, Cornford, D., Bastin, L. and Pebesma, E. (eds.). (2009). Uncertainty Markup Language (UncertML). Wayland: OGC. Retrieved from https://portal.opengeospatial.org/files/?artifact_id=33234.

US Public Law 106-554 2001 Information Quality Act. *Public Law, pp. 106-554.* Retrieved from https://www.gpo.gov/fdsys/pkg/PLAW-106publ554/html/PLAW-106publ554.htm.

Van der Velde, T, Milton, D., Lawson, T., Wilcox, C. Lansdell, M., Davis, G., Perkins, G. and Hardesty, B. (2017). Comparison of marine debris data collected by researchers and citizen scientists: Is citizen science data worth the effort? *Biological Conservation, 208*, pp. 127-138. https://doi.org/10.1016/j.biocon.2016.05.025.

Vicente, G. (2014). Data Quality. Presented at ESIP 2014 Summer Meeting., Earth Science information Partners, Frisco, CO. Retrieved from https://commons.esipfed.org/node/2327.

Weissman, D. E., Stiles, B. W., Hristova-Veleva, S. M., Long, D. G., Smith, D. K., Hilburn, K. A., & Jones, W. L. (2012). Challenges to satellite sensors of ocean winds: Addressing precipitation effects. *Journal of Atmospheric and Oceanic Technology, 29*(3), pp. 356–374. https://doi.org/10.1175/JTECH-D-11-00054.1.

Wentz, F., Ricciardulli, L., Rodriguez, E., Stiles, B., Bourassa, M., Long, D., Hoffman, R., Stoffelen, A., Verhoef, A., O'Neill, L., Farrar, T., Vandermark, D., Fore, A., Hristove-Veleva, S., Turk, J., Gaston, R., & Tyler, D. (2017). Evaluating and extending the ocean winds data climate record. *Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 99*. https://doi.org/10.1109/JSTARS.2016.2643641.

WGClimate. (2017). Climate Monitoring from Space - ECV Inventory. Retrieved from http://climatemonitoring.info/ecvinventory/

Wu, F., Cornillon, P., Boussidi, B., & Guan L. (2017). Determining the Pixel-to-Pixel Uncertainty in Satellite-Derived SST Fields. *Journal of Remote Sensing, 9*(9), p. 877. https://doi.org/10.3390/rs9090877.

Wuebbles, D.J., D.W. Fahey, K.A. Hibbard, D.J. Dokken, B.C. Stewart, and T.K. Maycock (eds.). (2017). Climate Science Special Report: Fourth National Climate Assessment, Volume I. Page 6. Washington, DC, USA: U.S. Global Change Research Program. https://doi.org/10.7930/J0J964J6.