Surveys in Geophysics
https://doi.org/10.1007/s10712-025-09889-5

Making Sense of Uncertainties: Ask the Right Question

Alexander Gruber[1] · Claire E. Bulgin[2,3] · Wouter Dorigo[1] · Owen Embury[2,3] · Maud Formanek[1] · Christopher Merchant[2,3] · Jonathan Mittaz[2] · Joaquín Muñoz-Sabater[4] · Florian Pöppl[1] · Adam Povey[5,6] · Wolfgang Wagner[1]

Received: 12 December 2024 / Accepted: 30 April 2025

Abstract
Earth observation data should inform decision making, but good decisions can only be made if the uncertainties in the data are taken into account. Making sense of uncertainty information can be difficult, because uncertainties represent the statistical spread in the observations (e.g., expressed as $x \pm y$), which does not relate directly to one specific use case of the data. Here, we propose a Bayesian framework to transform Earth observation product uncertainties into actionable information, i.e., estimates of how confident one can be in the occurrence of specific events of interest given the data and their uncertainty. We demonstrate this framework using two case examples: (i) monitoring drought severity based on soil moisture and (ii) estimating coral bleaching risk based on sea surface temperature. In both cases, we show that ignoring uncertainties can easily lead to misinterpretation of the data, making any decisions based on these data unlikely to be the best course of action. The proposed framework is general and can, in principle, be applied to a wide range of applications. Doing so requires a careful dialogue between data users, to formulate meaningful use cases and decision criteria, and data producers, to provide a rigorous description of their data and its uncertainties. The next step would then be to confront the uncertainty-informed estimates of event probabilities (created by the framework proposed here) with the costs and benefits of possible courses of action in order to make the best possible decisions that maximize socioeconomic merit.

Keywords Uncertainties · Earth observation · Decision making

✉ Alexander Gruber
alexander.gruber@geo.tuwien.ac.at

[1] Department of Geodesy and Geoinformation, Technische Universität Wien (TU Wien), 1040 Vienna, Austria

[2] Department of Meteorology, University of Reading, Reading RG6 6ET, UK

[3] National Centre for Earth Observation (NCEO), University of Reading, Reading RG6 6ET, UK

[4] European Centre for Medium-Range Weather Forecasts (ECMWF), Reading RG2 9AX, UK

[5] School of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK

[6] NCEO, University of Leicester, Leicester LE4 5SP, UK

Article Highlights

- We present a framework to make sense of Earth observation product uncertainties by transforming them into event probabilities
- We apply this framework to assess the confidence in statements about drought and coral bleaching based on Earth observation data
- We demonstrate that transforming uncertainties into normalized levels of confidence can help make better decisions

1 Introduction

Earth observation (EO) data products are derived from active and passive satellite instruments observing the land, ocean, ice, and atmosphere across the electromagnetic spectrum (Tatem et al. 2008). Inevitably, though, EO products have an associated uncertainty arising from a number of error sources in the measurement process, including limitations of the measurement and calibration process, uncertainties in auxiliary data, approximations in the data processing, and inhomogeneities in the observed field. These errors can make it difficult to interpret EO products without careful consideration, which can limit trust in environmental and socioeconomic decisions based upon such data (Anderson et al. 2017; Bulgin et al. 2024).

EO data providers strive to provide reliable uncertainty estimates alongside their products (Ablain et al. 2019; Bulgin et al. 2016a, b; Loew et al. 2017; H-SAF 2018; Ghent et al. 2019; H-SAF 2022; Dorigo et al. 2023). Different communities have established best practices for this purpose depending on product error characteristics and reference data availability (Strahler et al. 2006; Guillevic et al. 2018; Gruber et al. 2020; Montzka et al. 2020; Duncanson et al. 2021; Bulgin et al. 2016a; Ablain et al. 2019). Moreover, there has been growing effort to reconcile these practices with guidelines developed by the metrological community (Merchant et al. 2017; Mittaz et al. 2019; Strobl et al. 2024).

Metrological guidelines are described in the Guide to the Expression of Uncertainty in Measurement (GUM; JCGM 2008), which is maintained by the Joint Committee for Guides in Metrology of the International Bureau of Weights and Measures. These guidelines dictate that all measurements be accompanied by a description of their "uncertainty", which refers to the probability distribution of the measurement errors (JCGM 2012). For example, Gaussian-distributed errors may be described by their standard deviation (i.e., random uncertainty) and by their mean (i.e., systematic uncertainty). Importantly, measurements ought to be SI-traceable. That is, linked to an SI reference standard through a documented, unbroken chain of data transformations, with a quantification of the uncertainties that are introduced at every step of that chain.

While this gold standard approach for characterizing uncertainties can—in theory—provide a reliable, quantitative description of the stochastic properties of deterministic estimates, the question remains: What shall users of these estimates make of this information? How can the information that an estimate of a variable of interest has a value of $x \pm y$ instead of just x influence any use or decision based on this estimate?

We conjecture that uncertainty estimates are rarely used *quantitatively* to inform specific decisions, in part, because $x \pm y$ does not answer the more relevant question: How much can I trust x, or any use or decision based upon x? For example, to most people it is much less relevant that the level of a river is $x \pm y$ than whether the river is likely to

flood their house. To decision makers, more generally, the importance of data uncertainty is how it affects their certainty that decisions based on these data really are the best course of action. We therefore propose that, in order to optimally utilize uncertainty estimates for decision making, they should first be transformed into a normalized measure of confidence that relates to the intended use of the data. This can be done using Bayes' rule (Efron 2013). Bayes' rule allows us to infer how confident we can be that, say, an event actually happened, given our data alongside knowledge about the statistical properties of these data and the events in question. While this confidence is, by and large, determined by the uncertainty in the data, it provides a more accessible and comprehensible presentation of the information contained in the uncertainty estimates.

In this paper, we first revisit Bayes' rule in the context of Earth system science. We then propose a general framework for using Bayes' rule to transform EO product uncertainties into estimates of confidence that relate to an intended use of EO data. Finally, we demonstrate the merit of this workflow using two case examples: (i) the monitoring of drought; and (ii) the estimation of coral bleaching risk. Throughout this paper, we will use uppercase letters to refer to random variables and lowercase letters to refer to instances of these variables.

2 Bayes' Rule

Herein, Bayes' rule is applied to a class of events, E, and measurements of these events, M. E may refer to an Earth system state variable, a categorical class derived from such a variable (e.g., drought severity), a flux, or anything else observable directly or indirectly. M could refer to direct measurements but also to estimates derived from related measurements or predictions from a numerical model. Both E and M may be continuous, discrete, or categorical. For simplicity and without loss of generality, we will, hereinafter, simply speak of 'events' and 'measurements'.

Bayes' rule allows us to consider the measurements M as evidence to improve our prior knowledge about the events E. More specifically, it allows us to infer the conditional probability that E actually happened, given that we measured M, from knowledge of how likely it is that E produces a measurement of M and knowledge of the probabilities of E and M independent of one another. Mathematically, this is expressed as:

$$p(E|M) = p(M|E)\frac{p(E)}{p(M)} \tag{1}$$

where $p(E|M)$ is the conditional probability of E given M. It is often called the 'posterior', telling us how likely it is that an event *actually occurred* after we take our measurements into consideration. $p(M|E)$ is the conditional probability of M given E. It is often called the 'likelihood', describing how likely it is that we obtain a particular measurement, given that a certain event *has occurred*. $p(E)$ is the marginal probability of E, often called the 'prior', referring to our knowledge about the frequency with which an event occurs before we measure anything. $p(M)$ is the marginal probability of M, often called 'evidence', which describes how often we obtain a certain measurement whatever the event.

In the following section, we propose a generic workflow for applying Bayes' rule to Earth system science problems in order to convert uncertainty estimates into normalized measures of confidence in statements derived from Earth system science data (i.e., 'we are between 0 and 100% certain that...').

3 Proposed Workflow for Bayesian Uncertainty Interpretation

Our proposed workflow comprises 4 steps:

1. **Asking the right question**, i.e., seeking confidence *in what exactly*?
2. **Defining the prior**, i.e., the marginal distribution of the events in question, $p(E)$.
3. **Selecting the error model**, i.e., the functional relationship between measurements, their errors, and the 'truth', which is required to identify the likelihood, $p(M|E)$, and the evidence, $p(M)$.
4. **Computing the posterior**, $p(E|M)$, i.e., our final confidence estimates.

The following subsections describe each of these steps in more detail. Afterward, we illustrate this workflow with a few examples.

3.1 Asking the Right Question

Questions in Earth system science usually focus on states or changes of biogeophysical or biogeochemical variables. We could simply ask how likely it is that a variable of interest has *exactly* the value we measure, but the answer will often be "very unlikely" due to noise and bias in the data. More importantly, the exact values of a variable are typically less important than how that value relates to ranges and thresholds that would trigger certain actions or changes in our beliefs. Uncertainties, then, should inform us how justified such an action or change in beliefs actually is.

For example, if we want to know whether an observed state variable exceeds a critical threshold, a better answer than "soil moisture is 0.07 ± 0.03 m^3m^{-3}" would be "we are 70% confident that soil moisture has dropped below a level at which plants start to suffer from water stress". This tells us directly how urgently we need to irrigate. Other examples of more meaningful questions could be: "How certain are we that this was the warmest year on record?", "How many years do we have left—with 95% confidence—before New York City will fall below sea level?", or simply "How likely is it that I will need an umbrella today?".

3.2 Defining the Prior

Bayesian inference starts from prior knowledge about the frequency with which the events occur, i.e., the marginal distribution $p(E)$. Suppose E follow a Gaussian distribution $\mathcal{N}$ with mean μ and variance σ^2 (this could be, for example, sea surface temperature anomalies). The prior can then be written as:

$$p(E) \sim \mathcal{N}(\mu, \sigma^2) \tag{2}$$

Priors of other variables may require more complex representations. Soil moisture, for example, usually follows a distribution that is both bounded and skewed or even bimodal, and rainfall is often described by a gamma distribution. Note also that good priors can be difficult to obtain. In such a case, priors might be conditioned on related proxy variables (e.g., variables that are easier to observe or for which more observations are readily

available) and the Bayesian analysis be conducted sequentially. For example, a prior for soil moisture could be conditioned on antecedent rainfall and temperature.

3.3 Selecting the Error Model

To update the prior using information about the measurements of the event and their uncertainty, which is represented by the likelihood $p(M|E)$ and the evidence $p(M)$, we must first identify an appropriate error model. The error model describes what kinds of error affect these measurements and how they do so.

Errors typically comprise both a systematic and a random component. Moreover, measurements can be subject to drift (e.g., due to instrument aging), which, too, can have a systematic and a random component. For example, the error model could assume additive, zero-mean Gaussian-distributed random errors with zero- and first-order systematic errors. Their impact on a particular measurement $M = m$ of a particular event $E = e$ can be written as:

$$m = \alpha + \beta e + \varepsilon \tag{3}$$

where α denotes a possibly space and/or time-dependent additive systematic error, or bias; β denotes a possibly space and/or time-dependent multiplicative systematic error, or scaling; and ε denotes zero-mean Gaussian-distributed random error $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ with possibly space and/or time-dependent random error variance (i.e., uncertainty) σ_ε^2. The time dependency allows one to account for drift. Since the events E were defined to have mean μ and variance σ^2 (see Eq. (2)), the evidence and the likelihood then follow from this error model as:

$$p(M) \sim \mathcal{N}(\beta\mu + \alpha,\ \beta^2\sigma^2 + \sigma_\varepsilon^2) \tag{4}$$

and

$$p(M|E = e) \sim \mathcal{N}(\beta e + \alpha,\ \sigma_\varepsilon^2) \tag{5}$$

, respectively. Note that this is just one possible error model (which is commonly assumed for soil moisture datasets, for example). Depending on the data, the appropriate error model may be more complex and include also covariance terms or higher-order biases, account for different error correlation time and length scales, etc. However, while it is easily possible to incorporate any form of systematic errors (e.g., additive or multiplicative biases) into the error model and thus in the computation of the posterior, it is usually more meaningful to correct for these systematic errors—if known—a priori because this increases the confidence one can have in statements derived from the data.

3.4 Computing the Posterior

How exactly we compute the posterior depends on whether we are interested in (i) the confidence of exact events given exact measurements ($E = e$ and $M = m$, respectively); (ii) a range of events given a range of measurements ($e_1 \leq E \leq e_2$ and $m_1 \leq M \leq m_2$, respectively); or (iii) any combination of the two. As mentioned in Sect. 3.1, posterior probabilities that exactly e happens if exactly m is measured, i.e., $p(E = e|M = m)$, can become very small. In most

cases, we will therefore be more interested in the posterior probability given some ranges of E and/or M and thus have to integrate Bayes rule over these ranges as:

$$P(e_1 \leq E \leq e_2 \,|\, m_1 \leq M \leq m_2) = \frac{\int_{e_1}^{e_2} \int_{m_1}^{m_2} p(M|E)\, p(E)\, \mathrm{d}M \mathrm{d}E}{\int_{m_1}^{m_2} p(M)\, \mathrm{d}M} \tag{6}$$

Should one be interested in exact values of E ($e_1 = e_2 = e$) and/or M ($m_1 = m_2 = m$), Eq. (6) is simplified by removing the appropriate integrals.

3.5 Practical Considerations

Modeling probability distributions and computing Bayesian posteriors using Eq. (6) can become complex and computationally demanding. In practice, the general Bayesian estimation problem is thus often reduced to a simpler problem by placing assumptions on the prior distribution, likelihood, and error model. One such special case, which is widely used in various fields (Bishop and Nasrabadi 2006; Rencher and Schaalje 2008; Barfoot 2024), is that of a Gaussian prior and likelihood with a linear error model (Bishop and Nasrabadi 2006). In that case, the Bayesian posterior is also Gaussian, and its mean and covariance equivalent to a least-squares estimator, which can be calculated more easily, and whose results are straightforward to interpret.

In the context of this paper, this is particularly relevant for time series analyses where we are interested not in how a single measurement changes, but rather in how the evolution of measurements can inform us about an event or its trajectory. For such problems, modeling all involved marginal and joint probabilities of correlated measurements and events—as would be required to solve Eq. (6)—is difficult for arbitrary distributions. One common solution is to model time series as stochastic processes with a linear error model and Gaussian noise instead (i.e., as Gauss-Markov processes). Latent parameters and their distributions can then be efficiently estimated using filtering (or smoothing) methods (Särkkä et al. 2013) or through a least-squares solution (Barfoot 2024, Ch. 3) as follows.

Suppose we have a series of measurements $\vec{m} = (m_1, \ldots, m_j)^\intercal$ and associated events $\vec{e} = (e_1, \ldots, e_i)^\intercal$ with Gaussian likelihood $p(m_1, \ldots, m_j, \,|\, e_1, \ldots, e_i)$ and Gaussian prior $p(e_1, \ldots e_i)$. Further suppose that the relationship between measurements and events is linear, i.e., described by the design matrix A in the linear system of equations $A\vec{e} = \vec{m}$, where $\vec{m}$ includes both measurements and prior information. Under this linear-Gaussian assumption, the Bayesian posterior $p(e_1, \ldots, e_i, \,|\, m_1, \ldots, m_j)$ is also Gaussian and its mean ($\vec{\mu}$) and covariance (Σ) are given by the maximum a-posteriori estimate, i.e., the least-squares solution:

$$\vec{\mu} = (A^T Q^{-1} A)^{-1} A^T Q^{-1} \vec{m} \tag{7}$$

and

$$\Sigma = (A^T Q^{-1} A)^{-1}, \tag{8}$$

respectively, where Q is the error covariance matrix (i.e., representing uncertainties and error correlation structures). Together, $\vec{\mu}$ and Σ fully describe the posterior probability of the events given the measurements as:

$$p(e_1, \ldots, e_i, \,|\, m_1, \ldots, m_j) \sim \mathcal{N}(\vec{\mu}, \Sigma) \tag{9}$$

Equation (9) provides us with an estimate not only of how confident we can be that an individual event happened given an individual measurement—as did Eq. (6)—but of how confident we can be that a series of events happened given a series of measurements. This could be, for example, our confidence that the duration of heatwaves has increased over the last decade due to climate change, i.e., that temperatures exceed a certain threshold for increasing amounts of time.

In the following sections, we demonstrate the merit of the proposed Bayesian uncertainty interpretation by applying the proposed workflow described above to two case examples: (i) assessing the occurrence and magnitude of drought using satellite soil moisture retrievals; and (iii) assessing the risk of coral bleaching using sea surface temperature estimates.

4 Case Example: Drought

In this example, we demonstrate how to estimate the confidence of actually experiencing a drought when a given soil moisture dataset suggests so, accounting for the dataset uncertainty. To this end, we consider droughts in soil moisture (van Hateren et al. 2021), also referred to as agricultural drought, using a definition based on soil moisture anomaly percentiles similar to that used by the Copernicus European Drought Observatory (https://edo.jrc.ec.europa.eu/; last accessed: 11 December 2024; Svoboda et al. 2016) or the United States Drought Monitor (USDM; https://droughtmonitor.unl.edu/About/AbouttheData/DroughtClassification.aspx; last accessed: 11 December 2024). More specifically, we define drought severity based on how much soil moisture drops below climatologically normal conditions, which we estimate from satellite soil moisture products. The specific anomaly percentiles we use as drought severity thresholds are shown in Table 1. We will not discuss different drought types (e.g., meteorological drought, agricultural drought, etc.) or drought indices and their respective advantages or disadvantages any further, because this has been addressed exhaustively (e.g., Zargar et al. 2011; Mukherjee et al. 2018) and is irrelevant for the purpose of this demonstration.

4.1 The Question

The two questions that we aim to answer in this example are:

- How does the confidence in drought severity estimates derived from satellite soil moisture retrievals change as a function of the uncertainty in these retrievals?

Table 1 Drought severity classification based on soil moisture anomaly percentiles. Class thresholds are slightly modified from those used by the United States Drought Monitor (USDM) to obtain classes with equal cumulative probability

Percentile	Drought severity
20–30%	Slight
12–20%	Moderate
7–12%	Severe
4–7%	Extreme
2–4%	Exceptional

- How much confidence can we put into drought severity estimates of two commonly-used satellite products, from SMOS and ASCAT, in different regions when considering their specific uncertainties?

4.2 The Prior

As mentioned, we define drought by means of soil moisture anomaly percentiles. Anomalies, here, refer to deviations from the mean seasonal cycle, which is here assumed to follow a zero-mean Gaussian distribution. The prior can thus be written as:

$$p(\Theta) \sim \mathcal{N}(0,\, \sigma_\Theta) \tag{10}$$

where Θ denotes (true) soil moisture anomalies, and σ_Θ their assumed standard deviation.

4.3 The Error Model

Satellite soil moisture retrievals as well as retrievals of soil moisture anomalies are usually assumed to follow a linear error model with additive and independent zero-mean Gaussian random errors (Gruber et al. 2020), like the one we showed in Eq. (3):

$$X = \alpha + \beta\,\Theta + \varepsilon \tag{11}$$

where X denotes the satellite soil moisture anomaly retrievals. The evidence and the likelihood thus follow as:

$$p(X) \sim \mathcal{N}(0,\, \beta^2\sigma_\Theta^2 + \sigma_\varepsilon^2) \tag{12}$$

and

$$p(X|\Theta = \theta) \sim \mathcal{N}(\alpha + \beta\theta,\, \sigma_\varepsilon^2) \tag{13}$$

respectively.

4.4 The Posterior

Since we are interested in how likely an observed drought *class* represents a real drought, our posterior needs to calculate the probability that soil moisture is within a certain range, given our measurements are in a given range, i.e., $p(\theta_{\min} \le \Theta \le \theta_{\max} | X_{\min} \le X \le X_{\max})$. However, the probability that a drought is exactly as severe as the satellite suggests is very small, so we will, instead, estimate how likely it is that there is any drought at all (i.e., at least a slight drought or worse) depending on the observed drought magnitude and the uncertainty of the measurement:

$$P\left(\Theta \le \theta_{\max}^{\text{slight}} | \theta_{\min}^{d} \le X \le \theta_{\max}^{d}\right) = \frac{\int_{-\infty}^{ppf(\Theta=\theta_{\max}^{\text{slight}})} \int_{ppf(X=\theta_{\min}^{d})}^{ppf(X=\theta_{\max}^{d})} p(X|\Theta)\, p(\Theta)\, \mathrm{d}X \mathrm{d}\Theta}{\int_{ppf(X=\theta_{\min}^{d})}^{ppf(X=\theta_{\max}^{d})} p(X)\, \mathrm{d}X} \tag{14}$$

where $\theta_{\min}^{d}$ and $\theta_{\max}^{d}$ represent the lower and upper soil moisture boundaries of drought class d, respectively, which are given by the class probability thresholds defined in Table 1.

$ppf(\cdot)$ is the percentile point function, which converts these soil moisture boundaries into the correct integration bounds for the associated probability distribution.

4.5 Results and Discussion

The following results are based on uncertainty estimates for monthly surface soil moisture anomalies from SMOS (Kerr et al. 2010) and ASCAT (Wagner et al. 1999; Naeimi et al. 2009) over Europe, obtained from triple collocation analysis (TCA; Stoffelen 1998) using the Quality Assurance for Soil Moisture online validation platform (QA4SM; https://qa4sm.eu/; last accessed: 11 December 2024). The SMOS and ASCAT products we use are the H-SAF H119 climate data record (H-SAF 2018) and the SMOS Level 2 v700 product (ESA 2021). Surface soil moisture simulations from the ERA5 reanalysis (Hersbach et al. 2020) are used to complete the triplet. TCA results and a description of processing parameters (spatial and temporal matching parameters, masking criteria, etc.) can be found at https://doi.org/10.5281/zenodo.11282947 (last accessed: 11 December 2024). Regions with insufficient data (e.g., due to dense vegetation or prolonged frozen periods) are masked out. Note that TCA provides estimates for both the uncertainty σ_ε^2 and the signal variance $\beta^2\sigma_\Theta^2$, and thus all information necessary to compute the posterior.

4.5.1 How Does Soil Moisture Retrieval Uncertainty Impact the Confidence in Drought Estimates?

Figure 1 aims to answer our first question: "How does the confidence in drought severity estimates derived from satellite soil moisture retrievals change as a function of the uncertainty in these retrievals?". It shows the probability of detecting at least a slight drought (Eq. (14)) as a function of the measurement uncertainty—expressed in terms of the signal-to-noise ratio ($SNR = \beta^2\sigma_\Theta^2/\sigma_\varepsilon^2$)—for various observed drought magnitudes. Note that the SNR scales linearly with β. Therefore, β and SNR affect the confidence in drought estimates in the same way (Fig. 1; x-axis).

As expected, the greater the observed drought severity or the lower the uncertainty, the more confident we can be that there really is *at least* a slight drought. The vertical lines in the figure indicate the points where this confidence exceeds 95%. That is, they show the minimum SNR that a product needs to have so that a given drought severity provides

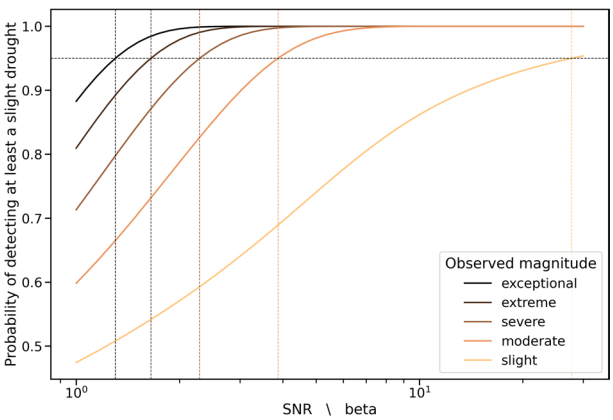

Fig. 1 Probability of detecting at least a slight drought as a function of the SNR for different observed drought magnitudes

us with 95% confidence that there is at least a slight drought. This helps us to answer our second question.

4.5.2 How Confident Can We be in Drought Estimates from ASCAT and SMOS?

To answer our second question—How much confidence can we put in the drought severity estimates of SMOS and ASCAT considering their specific uncertainties?—Fig. 2 shows, at each location of our study domain, how severe of a drought SMOS and ASCAT need to observe so that our confidence that there is, in fact, at least a mild drought (as calculated before) exceeds 95%.

This shows the regions in which we can be confident that a given sensor detects even the slightest droughts, and in which areas drought estimates are not to be trusted even when the soil moisture product suggests them to be extreme or even exceptional. This is arguably more tangible and readily interpreted than whether soil moisture product uncertainty is, e.g., 0.04 m^3m^{-3} or 0.06 m^3m^{-3}. For example, it shows that Spanish authorities could reliably monitor drought using these products, especially SMOS which exhibits lower uncertainties than ASCAT in this region. Germany, in contrast, would benefit from considering other (or at least additional) monitoring strategies because both ASCAT and SMOS soil moisture retrievals are too uncertain for this purpose.

5 Case Example: Coral Bleaching

In this example, we address an issue that has been exacerbated over recent years due to rising sea surface temperatures (SST), which is coral bleaching (Henley et al. 2024). While coral bleaching is a result of various complex, interdependent processes such as the amount of light, coral species and genetics, bathymetry etc., studies have found that one of the most important and robust predictors for coral bleaching is the existence of prolonged marine heat waves (McClanahan et al. 2019). Specifically, there seems to be a significant non-linear increase in coral bleaching events once extreme SST persists for long periods of time.

The Coral Reef Watch of the National Oceanic and Atmospheric Administration (NOAA) Satellite and Information Service (https://coralreefwatch.noaa.gov/; last accessed: 11 December 2024), for instance, uses satellite-derived SST to identify Coral

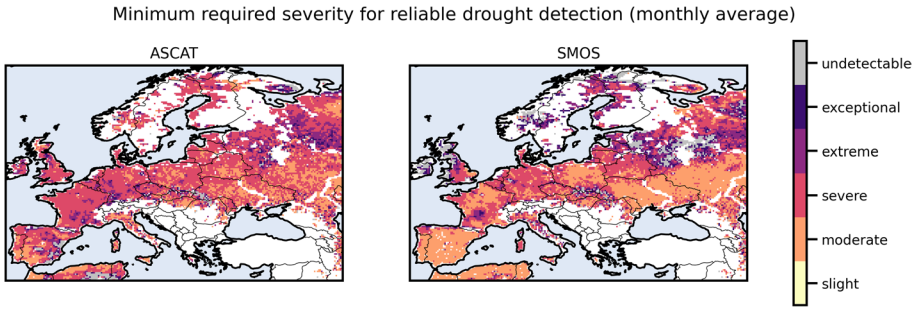

Fig. 2 Minimal required drought severity per location as reported by ASCAT (left) and SMOS (right) to reach 95% confidence in detection of at least a slight drought. Areas with insufficient input data or unreliable uncertainty estimates are masked out

Bleaching HotSpots. HotSpots are regions where SST is warmer than the highest climatological monthly mean (referred to as the Maximum Monthly Mean, MMM). A value of MMM + 1°C is used as a threshold to indicate heat stress which, if persisting over prolonged periods of time, has been associated with coral bleaching. However, such an approach does not take uncertainties in SST observations into account but instead merely tests whether or not deterministic SST estimates exceeded a certain threshold for a certain amount of time.

In this example, we demonstrate how the uncertainties in SST data can be interpreted meaningfully if one were to assess the occurrence of coral bleaching from SST observations based on the above-mentioned relation. This can transform statements such as "there have been conditions that are likely to promote coral bleaching" into statements such as "there is an *x*% probability that there have been conditions that are likely to promote coral bleaching". To that end, we use the ESA CCI SST L4 product (Embury et al. 2024) and focus on the Great Barrier Reef (GBR).

5.1 The Question

The two specific questions we aim to answer in this example are:

- How has the probability of SST conditions fostering coral bleaching changed over recent decades?
- How likely is it that SST conditions associated with coral bleaching events occur in certain years over specific locations within the GBR?

As discussed in Sect. 3.5, these are time series problems that require an investigation of multiple, consecutive measurements, i.e., analyses of how likely it is that SST measurements continuously exceeded a certain threshold for a specific amount of time, taking into account the uncertainties in the measurements. We will thus model probabilities using a least-squares estimator.

5.2 The Prior

To obtain a prior for SST states that can account for their temporal evolution, we assume SST anomalies a_T (i.e., the deviations of actual temperature from the climatology) to characterize a random walk:

$$T_{i+1} = T_i + \delta C_i + a_T, \quad a_T \sim \mathcal{N}(0, \sigma_T^2) \tag{15}$$

so that

$$p(T_{i+1} | T_i) \sim \mathcal{N}(T_i + \delta C_i, \sigma_T^2) \tag{16}$$

where T_i represents SST states at time i, δC_i represents the climatological increment between time steps $i + 1$ and i, and σ_T^2 represents the random variability of SST anomalies between consecutive time steps.

5.3 The Error Model

We assume the measurements M_i of the SST states T_i to follow a linear error model with additive zero-mean Gaussian noise and negligible bias (Embury et al. 2024):

$$M_i = T_i + \varepsilon_M, \quad \varepsilon_M \sim \mathcal{N}(0,\, \sigma_M^2) \tag{17}$$

The likelihood of measurements M_i given temperatures T_i thus follows as:

$$p(M_i|T_i) \sim \mathcal{N}(T_i,\, \sigma_M^2) \tag{18}$$

where σ_M^2 is the standard uncertainty (i.e., random error variance) of the measurements. In the CCI SST L4 data product we use here, estimates for σ_M^2 are provided for each individual SST measurement.

5.4 The Posterior

As discussed in Sect. 3.5, since the prior and likelihood are both Gaussian and the error model is linear, the posterior $p(T_1, \ldots, T_n \,|\, M_1, \ldots, M_n)$ is also Gaussian and its mean and covariance can be obtained by solving the equivalent least-squares problem:

$$|WA\vec{t} - W\vec{m}|_2 \;\rightarrow\; \min! \tag{19}$$

Equation (19) is simply the matrix notation expressing our assumed relationship between SST states and measurements (see Eqs. (15) and (17)):

$$T_i - M_i = \varepsilon_M, \quad \varepsilon_M \sim \mathcal{N}(0, \sigma_M^2) \tag{20}$$

$$T_{i+1} - T_i - \delta C_i = a_T, \quad a_T \sim \mathcal{N}(0, \sigma_T^2) \tag{21}$$

A, $\vec{t}$, and $\vec{m}$ thus follow as:

$$A = \begin{pmatrix} 1 & 0 & 0 & \ldots \\ 0 & 1 & 0 & \ldots \\ 0 & 0 & 1 & \ldots \\ \vdots & \vdots & \vdots & \vdots \\ -1 & 1 & 0 & \ldots \\ 0 & -1 & 1 & \ldots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}, \; \vec{t} = \begin{pmatrix} T_1 \\ \vdots \\ T_n \end{pmatrix}, \quad \text{and } \vec{m} = \begin{pmatrix} M_1 \\ \vdots \\ M_n \\ \delta C_1 \\ \vdots \\ \delta C_{n-1} \end{pmatrix} \tag{22}$$

with the weight matrix W, defined as $W = Q^{-\frac{1}{2}}$ and $Q = \mathrm{diag}(\sigma_M^2, \ldots, \sigma_M^2, \sigma_T^2, \ldots, \sigma_T^2)$. The parameters of the posterior $p(T_1, \ldots, T_n \,|\, M_1, \ldots, M_n) \sim \mathcal{N}(\vec{t}, \Sigma)$ can then be estimated as:

$$\vec{t} = (A^T Q^{-1} A)^{-1} A^T Q^{-1} \vec{m} \tag{23}$$

and

$$\Sigma = (A^T Q^{-1} A)^{-1} \tag{24}$$

Finally, the probability that SST is above $T_{\max}$ degrees for k consecutive days starting at day i can be derived by integrating this posterior density over the hyperrectangle $[T_{\max}, \infty]^k$ (i.e., all temperatures at the k consecutive days between $T_{\max}$ and ∞):

$$P(T_{\max} < T_i, \ldots, T_{\max} < T_{i+k} \,|\, M_1, \ldots, M_n) = \int_{[T_{\max}, \infty]^k} p(T_i, \ldots, T_{i+k} \,|\, M_1, \ldots, M_n)\, \mathrm{d}T \tag{25}$$

5.5 Results and Discussion

The following results are based on the ESA CCI SST L4 product (Embury et al. 2024) version v3.0.1 (obtained from https://dx.doi.org/10.5285/4a9654136a7148e39b7feb56f8bb02d2; last accessed: 11 December 2024) (Good and Embury 2024), which provides daily gap-free SST data together with estimates of their standard uncertainty produced using a variational data assimilation system. We use these SST analyses and their uncertainty in Eq. (25) to calculate how likely it was when and where that the GBR experienced conditions that have been associated with an increased occurrence of coral bleaching events, that is, temperatures exceeding MMM + 1°C for at least 35 consecutive days. For simplicity, we will, hereinafter, refer to such conditions as "coral bleaching conditions" (this corresponds to alert level 1 of the Coral Reef Watch, which assigns alert levels from 1 to 5, the latter indicating a risk of near complete mortality). For comparison, we also calculate how often the deterministic SST observations exceeded the same threshold for the same period of time—as it is usually done—allowing SST to drop below the threshold for up to three consecutive days to mitigate the impact of noise. MMM was derived for the climatological period 1981–2010. The GBR region was extracted through a shape file obtained from https://www.marineregions.org/gazetteer.php?p=details&id=26847 (last accessed: 11 December 2024).

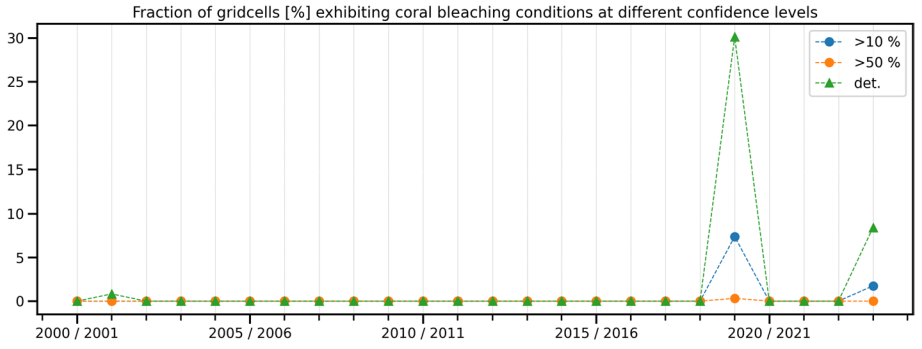

Fig. 3 Percentage [%] of grid cells that have more than a 10% (blue line) or 50% (orange line) probability (Eq. (25)) of having experienced coral bleaching conditions (i.e., temperatures above MMM + 1°C for at least 35 days) at least once in a given year. For comparison, the green line shows the fraction of grid cells where the deterministic SST estimates suggest such conditions (i.e., without considering uncertainties)

5.5.1 How Much Has the Problem of Coral Bleaching Exacerbated Over the Past Decades?

Figure 3 aims to answer our first question: "How has the probability of conditions that foster coral bleaching changed over the past decades?". It shows the percentage of grid cells that are likely (with different degrees of confidence) to have experienced coral bleaching conditions at least once in a given year. For comparison, the deterministic estimates that do not make use of uncertainty information (i.e., the percentage of grid cells where SST analyses exceeded said thresholds at least once in a given year) are also shown.

With some rare exceptions in the summer of 2001/2002, SST data suggest the occurrence of critical conditions that could have led to coral bleaching only in the summers of 2019/2020 and 2023/2024. Importantly, looking at deterministic SST estimates alone can lead one to be overconfident about the occurrence of coral bleaching events. In the summer 2019/2020, for instance, deterministic estimates suggest that about 30% of the GBR have experienced coral bleaching conditions. When taking uncertainties in the SST data into consideration, however, we learn that for only about 7–8 % of the GBR we can be more than 10 % confident in such a statement, and virtually nowhere does the probability of such conditions exceed 50 %.

It is important to note that the shown values are the estimated probabilities that *actual* SST *did* exceed certain thresholds given the SST observations and their uncertainty. That is, low values do *not* indicate a low level of confidence whether the thresholds were exceeded, but a high level of confidence that they were not. In other words, when ignoring uncertainties, SST-based estimates of when or where coral bleaching conditions might have occurred are often worse than a coin flip would be.

5.5.2 How Much Does the Likelihood of Coral Bleaching Events Vary Across Years and Locations?

Here we aim to answer the second question: "How likely did coral bleaching conditions occur over individual locations of the GBR?". To that end, Fig. 4 shows the maximum observed probability of coral bleaching conditions for the two potential incidence summers of 2019/2020 and 2023/2024. For comparison, Fig. 4 also shows binary maps of where the deterministic SST estimates suggest such conditions in the same years.

As expected from Fig. 3, deterministic estimates suggest coral bleaching conditions in many regions where they are actually highly unlikely. To understand the significant overestimation of coral bleaching conditions when using deterministic SST estimates alone, Fig. 5 shows SST time series and their uncertainty together with the estimated probability for coral bleaching conditions at three randomly selected locations with different (maximum) probabilities of coral bleaching conditions. In all cases, SST time series do not exceed MMM + 1°C by a large margin, and their uncertainties seem quite comparable. Nevertheless, the estimated probabilities that these time series exceeded MMM + 1°C for 35 days or more are quite different (~0.1, ~0.4, and ~0.8). Strikingly, despite having a clear visualization of SST estimates and their uncertainty, it is difficult to guess visually even in which cases coral bleaching conditions are more likely than not (i.e., above or below 0.5).

This illustrates well how difficult it can be to correctly interpret—or even distinguish—uncertainty estimates directly. This is, in part, because the usual expression of a one (two) error standard deviation range around an estimate does *not* indicate the range within which

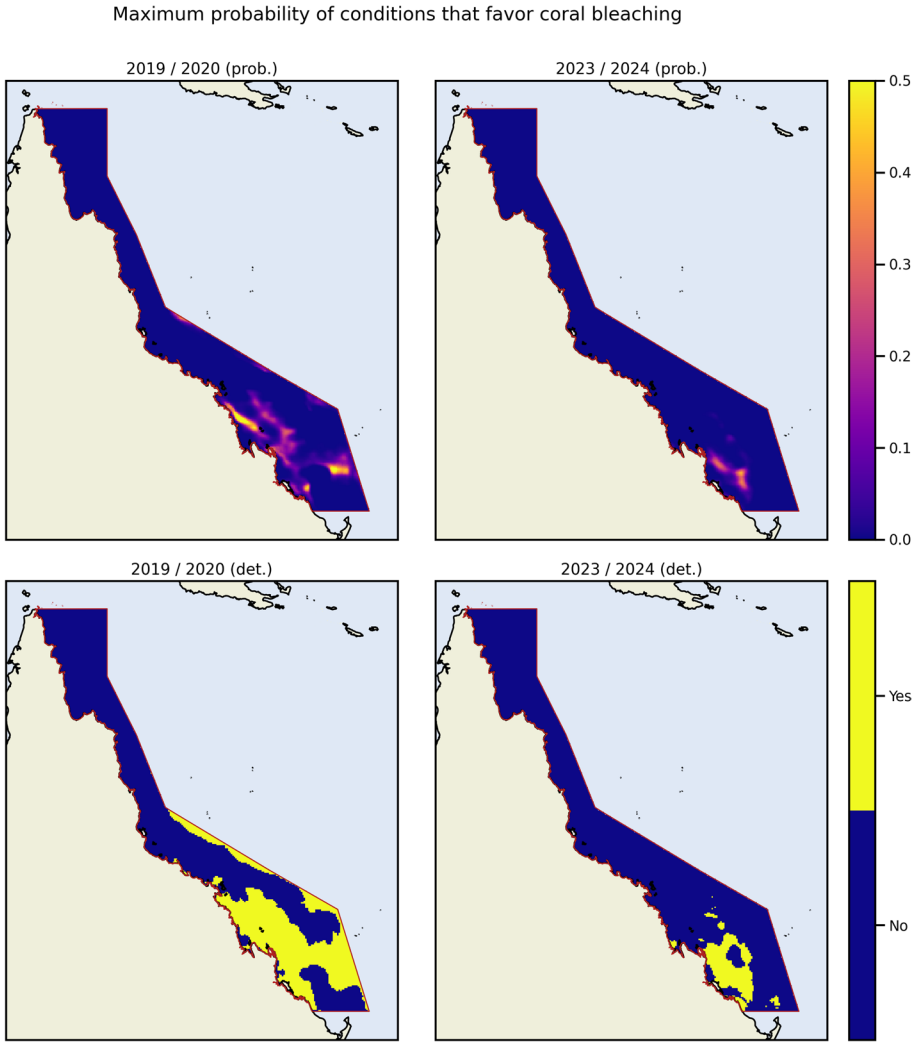

Fig. 4 Maximum probability that coral bleaching conditions (i.e., temperatures above MMM + 1°C for at least 35 days) occurred at least once in the summers of 2019/2020 (top left) and 2023/2024 (top right); and grid cells where deterministic SST observations exceed these thresholds at least once in the same years (bottom left and bottom right, respectively)

the true value lies with a 68% (95%) probability, as is a common misconception akin to the misconception of p values and confidence intervals (Greenland et al. 2016). Such a probability of where the true value is likely to be is arguably more meaningful, and precisely the one we estimate here using Bayes' theorem and integrate to get a holistic assessment of the uncertainties of a series of measurements.

Finally, note that the probabilities shown here reflect only our confidence in certain SST states and not in the actual occurrence of coral bleaching, which is only to some degree predictable from SST data alone (McClanahan et al. 2019). Remember, however, that the

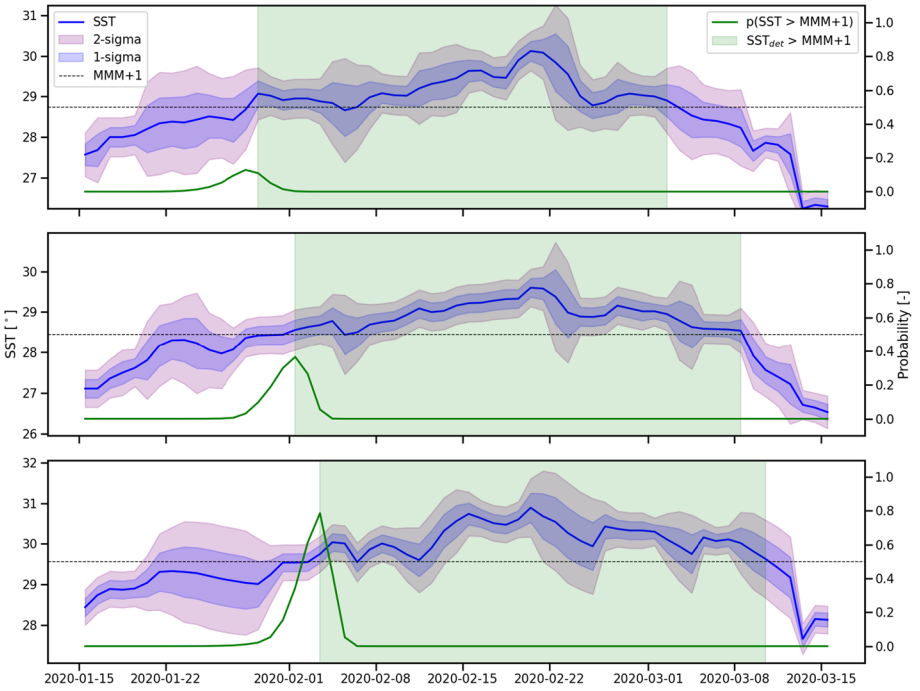

Fig. 5 SST time series (blue solid line) at three random grid cells of the GBR region shown in Fig. 4. Blue- and purple-shaded areas indicate SST ± one and two times their standard uncertainty, respectively. The green solid line shows the probability, for each day, that SST exceeded the MMM + 1°C threshold (dashed line) for that and the subsequent 35 days. The green-shaded area marks the total (≥ 35 day) time period during which deterministic SST estimates consistently exceeded MMM + 1°C (allowing for drops below that threshold for up to three days)

purpose of this example is not to obtain the most reliable estimates of when and where coral bleaching occurs, but to demonstrate how data uncertainties can be interpreted meaningfully by converting them into statements about the confidence in certain events given some measurements. As this example shows, this arguably leads to better insights than using deterministic estimates alone.

6 Conclusions

Earth observation data should help inform decision making. Good decisions can only be made if the uncertainties in the data are taken into account. Data producers thus strive to provide reliable uncertainty estimates alongside their products. Data users, however, often struggle to make sense of uncertainty information, because it represents the statistical spread in the error distribution of the observations (for example, the random error standard deviation), which does not relate to any specific use of the data. That is, expressing data and their uncertainty as something like "$x \pm y$" does not directly answer the really important question: How much can I trust x, or any use of or decision based upon x?

In this paper, we propose a framework to convert estimates of EO data uncertainty into more meaningful, actionable information. To that end, we use Bayes' rule to convert

uncertainty estimates into normalized statements about the probability of a given event of interest. This can turn statements of the kind: "The state of this variable is $x \pm y$" into more meaningful statements like, "We can be $z\%$ certain that this event actually happened.". We demonstrate this framework in two case study examples. In the first example, we use satellite soil moisture retrieval uncertainty to estimate how confident we can be that a certain region really is experiencing a drought when the soil moisture retrievals suggest so. In the second example, we use the uncertainties in sea surface temperature analyses to estimate how confident we can be that extreme temperature conditions prevailed sufficiently long that coral bleaching is possible. In both cases, we show that the deterministic (soil moisture or sea surface temperature) estimates alone are easily misinterpreted if their uncertainties are ignored, which would make any decisions based on these estimates unlikely to be the best course of action.

The proposed framework can, in principle, be applied to a wide variety of applications, from the design of new satellite missions to environmental action. Turning it into practice, however, will require close collaboration between data users, to formulate meaningful use cases and decision criteria, and data producers, to provide a rigorous description of the data (i.e., estimates about product uncertainties including possible correlation structures, reliable priors, an appropriate error model, etc.).

The next step toward meaningful decision making, then, must be to confront event probabilities with possible courses of action to estimate their socioeconomic value. For example, knowing the probability of crop yield loss due to drought allows one to weigh the expected monetary costs of this loss against the costs of irrigation or other mitigation measures, *including the costs of getting it wrong some of the time*. The best decision, then, is the one that requires minimum costs for maximum socioeconomic benefit.

Acknowledgements This paper is an outcome of the Workshop "Remote Sensing in Climatology: Essential Climate Variables and their Uncertainties" held at the International Space Science Institute (ISSI) in Bern, Switzerland (13–17 November 2023). The study has received funding from the European Space Agency's Fiducial Reference Measurements for Soil Moisture (FRM4SM) project (ESA Contract No: 4000135204/21//I-BG). Parts of this study were funded through the National Environmental Research Council's (NERC's) support of the National Centre for Earth Observation (NCEO) with Contract No. NE/R016518/1.

Funding Open access funding provided by TU Wien (TUW).

Declarations

Conflict of interest The authors have no conflict of interest to declare.

References

Ablain M, Meysignac B, Zawadzki L et al (2019) Uncertainty in satellite estimates of global mean sea-level changes, trend and acceleration. Earth Syst Sci Data 11:1189–1202. https://doi.org/10.5194/essd-11-1189-2019

Anderson K, Ryan B, Sonntag W et al (2017) Earth observation in service of the 2030 agenda for sustainable development. Geo-spat Inf Sci 20(2):77–96. https://doi.org/10.1080/10095020.2017.1333230

Barfoot TD (2024) State estimation for robotics. Cambridge University Press, Cambridge

Bishop CM, Nasrabadi NM (2006) Pattern recognition and machine learning, vol 4. Springer, Berlin

Bulgin CE, Embury O, Corlett G et al (2016a) Independent uncertainty estimates for coefficient based sea surface temperature retrieval from the Along-Track Scanning Radiometer. Remote Sens Environ 178:213–222. https://doi.org/10.1016/j.rse.2016.02.022

Bulgin CE, Embury O, Merchant CJ (2016b) Sampling uncertainty in gridded sea surface temperature products and Advanced Very High Resolution Radiometer (AVHRR) Global Area Coverage (GAC) data. Remote Sens Environ 177:287–294. https://doi.org/10.1016/j.rse.2016.02.021

Bulgin CE, Green P, Gruber A, et al (2024) The importance of scale in the definition of uncertainties: how do we best communicate this to data users? Surv Geophys (submitted)

Dorigo W, Preimesberger W, Stradiotti P, et al (2023) Algorithm Theoretical Baseline Document (ATBD) supporting product version 08.1. Technical report, ESA Climate Change Initiative Plus—Soil Moisture

Duncanson L, Armston J, Disney M et al (2021) Aboveground woody biomass product validation good practices protocol. CEOS Land Prod Valid Subgroup. https://doi.org/10.5067/doc/ceoswgcv/lpv/agb.001

Efron B (2013) Bayes' theorem in the 21st century. Science 340(6137):1177–1178

Embury O, Merchant CJ, Good SA et al (2024) Satellite-based time-series of sea-surface temperature since 1980 for climate applications. Sci Data 11(1):326. https://doi.org/10.1038/s41597-024-03147-w

ESA (2021) SMOS L2 SM V700, version 700 [dataset]. European Space Agency. https://doi.org/10.57780/SM1-857c3d7

Ghent D, Veal K, Trent T et al (2019) A new approach to defining uncertainties for MODIS land surface temperature. Remote Sens. https://doi.org/10.3390/rs11091021

Good S, Embury O (2024) ESA sea surface temperature climate change initiative (SST_cci): level 4 analysis product, version 3.0. https://doi.org/10.5285/4a9654136a7148e39b7feb56f8bb02d2. 09 Apr 2024

Greenland S, Senn SJ, Rothman KJ et al (2016) Statistical tests, p values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol 31(4):337–350. https://doi.org/10.1007/s10654-016-0149-3

Gruber A, De Lannoy G, Albergel C et al (2020) Validation practices for satellite soil moisture retrievals: what are (the) errors? Remote Sens Environ 244:111806. https://doi.org/10.1016/j.rse.2020.111806

Guillevic P, Göttsche F, Nickeson J et al (2018) Land surface temperature product validation best practice protocol. version 1.1. CEOS Land Prod Valid Subgroup 60:10. https://doi.org/10.5067/doc/ceoswgcv/lpv/lst.001

H-SAF (2018) Algorithm Theoretical Baseline Document (ATBD) Metop ASCAT soil moisture CDR and offline products, v0.7. Technical report. https://hsaf.meteoam.it/Products/ProductsList?type=soil_moisture. Accessed 25 Nov 2024

H-SAF (2022) Product Validation Report (PVR) Metop ASCAT surface soil moisture climate data record v7 12.5 km sampling (H119) and extension (H120), v1.2. Technical report. https://hsaf.meteoam.it/Products/ProductsList?type=soil_moisture. Accessed 25 Nov 2024

Henley BJ, McGregor HV, King AD et al (2024) Highest ocean heat in four centuries places Great Barrier Reef in danger. Nature 632(8024):320–326. https://doi.org/10.1038/s41586-024-07672-x

Hersbach H, Bell B, Berrisford P et al (2020) The ERA5 global reanalysis. Q J R Meteorol Soc 146(730):1999–2049. https://doi.org/10.1002/qj.3803

JCGM (2008) Evaluation of measurement data–guide to the expression of uncertainty in measurement (GUM). Technical Report, JCGM 100:2008, Bureau International des Poids et Mesures (BIPM), Joint Committee for Guides in Metrology (JCGM). https://www.bipm.org/en/publications/guides/gum.html

JCGM (2012) International vocabulary of metrology–basic and general concepts and associated terms (VIM 3rd edition). Technical report. JCGM 200:2012, Bureau International des Poids et Mesures (BIPM), Joint Committee for Guides in Metrology (JCGM). https://www.bipm.org/en/publications/guides/vim.html

Kerr Y, Waldteufel P, Wigneron JP et al (2010) The SMOS mission: new tool for monitoring key elements of the global water cycle. Proc IEEE 98(5):666–687. https://doi.org/10.1109/JPROC.2010.2043032

Loew A, Bell W, Brocca L et al (2017) Validation practices for satellite-based earth observation data across communities. Rev Geophys 55(3):779–817. https://doi.org/10.1002/2017RG000562
McClanahan TR, Darling ES, Maina JM et al (2019) Temperature patterns and mechanisms influencing coral bleaching during the 2016 El Niño. Nat Clim Chang 9(11):845–851. https://doi.org/10.1038/s41558-019-0576-8
Merchant CJ, Paul F, Popp T et al (2017) Uncertainty information in climate data records from Earth observation. Earth Syst Sci Data 9(2):511–527. https://doi.org/10.5194/essd-9-511-2017
Mittaz J, Merchant CJ, Woolliams ER (2019) Applying principles of metrology to historical Earth observations from satellites. Metrologia 56(3):032002. https://doi.org/10.1088/1681-7575/ab1705
Montzka C, Cosh M, Bayat B, et al (2020) Soil moisture product validation good practices protocol version 1.0. In: Montzka C, Cosh M, Nickeson J, Camacho F (eds) Good practices for satellite derived land product validation. Land Product Validation Subgroup (WGCV/CEOS), p 123. https://doi.org/10.5067/doc/ceoswgcv/lpv/sm.001
Mukherjee S, Mishra A, Trenberth KE (2018) Climate change and drought: a perspective on drought indices. Curr Clim Change Rep 4:145–163. https://doi.org/10.1007/s40641-018-0098-x
Naeimi V, Scipal K, Bartalis Z et al (2009) An improved soil moisture retrieval algorithm for ERS and METOP scatterometer observations. IEEE Trans Geosci Remote Sens 47(7):1999–2013. https://doi.org/10.1109/TGRS.2008.2011617
Rencher AC, Schaalje GB (2008) Linear models in statistics. Wiley, New York
Särkkä S, Solin A, Hartikainen J (2013) Spatiotemporal learning via infinite-dimensional Bayesian filtering and smoothing: a look at Gaussian process regression through Kalman filtering. IEEE Signal Process Mag 30(4):51–61. https://doi.org/10.1109/MSP.2013.2246292
Stoffelen A (1998) Toward the true near-surface wind speed: error modeling and calibration using triple collocation. J Geophys Res 103(C4):7755–7766. https://doi.org/10.1029/97JC03180
Strahler AH, Boschetti L, Foody GM et al (2006) Global land cover validation: recommendations for evaluation and accuracy assessment of global land cover maps. Eur Communities Luxemb 51(4):1–60
Strobl PA, Woolliams ER, Molch K (2024) Lost in translation: the need for common vocabularies and an interoperable thesaurus in Earth observation sciences. Surv Geophys. https://doi.org/10.1007/s10712-024-09854-8
Svoboda M, Fuchs B et al (2016) Handbook of drought indicators and indices. Drought and water crises: integrating science, management, and policy, pp 155–208
Tatem AJ, Goetz SJ, Hay SI (2008) Fifty years of earth observation satellites: views from above have lead to countless advances on the ground in both scientific knowledge and daily life. Am Sci 96(5):390. https://doi.org/10.1511/2008.74.390
van Hateren TC, Chini M, Matgen P et al (2021) Ambiguous agricultural drought: characterising soil moisture and vegetation droughts in Europe from earth observation. Remote Sens 13(10):1990. https://doi.org/10.3390/rs13101990
Wagner W, Lemoine G, Rott H (1999) A method for estimating soil moisture from ERS scatterometer and soil data. Remote Sens Environ 70(2):191–207. https://doi.org/10.1016/S0034-4257(99)00036-X
Zargar A, Sadiq R, Naser B et al (2011) A review of drought indices. Environ Rev 19(NA):333–349. https://doi.org/10.1139/A11-013