Earth Syst. Sci. Data, 9, 511–527, 2017
https://doi.org/10.5194/essd-9-511-2017

Open Access Earth System Science Data

Uncertainty information in climate data records from Earth observation

Christopher J. Merchant[1,2], Frank Paul[3], Thomas Popp[4], Michael Ablain[5], Sophie Bontemps[6], Pierre Defourny[6], Rainer Hollmann[7], Thomas Lavergne[8], Alexandra Laeng[9], Gerrit de Leeuw[10], Jonathan Mittaz[1,11], Caroline Poulsen[12], Adam C. Povey[13], Max Reuter[14], Shubha Sathyendranath[15], Stein Sandven[16], Viktoria F. Sofieva[10], and Wolfgang Wagner[17]

[1]Department of Meteorology, University of Reading, Reading RG6 6AL, UK
[2]National Centre for Earth Observation, University of Reading, Reading RG6 6AL, UK
[3]Department of Geography, University of Zurich, Winterthurerstr. 190, 8057 Zurich, Switzerland
[4]Deutsches Zentrum für Luft-und Raumfahrt e. V., Deutsches Fernerkundungsdatenzentrum, 82234 Oberpfaffenhofen, Germany
[5]Collecte Localisation Satellites, 11 Rue Hermès, 31520 Ramonville-Saint-Agne, France
[6]Earth and Life Institute, Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium
[7]Deutscher Wetterdienst, Frankfurterstr. 135, 63500 Offenbach, Germany
[8]Norwegian Meteorological Institute, 0313 Oslo, Norway
[9]Karlsrhue Institute for Technology, Institut für Meteorologie und Klimaforschung, 76021 Karlsruhe, Germany
[10]Finnish Meteorological Institute, 00101 Helsinki, Finland
[11]National Physical Laboratory, Teddington TW11 0LW, UK
[12]Science and Technology Facilities Council, Rutherford Appleton Laboratory, Didcot OX11 0QX, UK
[13]National Centre for Earth Observation, University of Oxford, Oxford OX1 3PU, UK
[14]Institute of Environmental Physics, University of Bremen, 28359 Bremen, Germany
[15]Plymouth Marine Laboratory, Prospect Place, Plymouth PL1 3DH, UK
[16]Nansen Environmental and Remote Sensing Center, Thormohlensgate 47, 5006 Bergen, Norway
[17]Department of Geodesy and Geoinformation, Vienna University of Technology, 1040 Wien, Austria

Correspondence to: Christopher J. Merchant (c.j.merchant@reading.ac.uk)

Received: 21 February 2017 – Discussion started: 28 February 2017
Revised: 19 June 2017 – Accepted: 19 June 2017 – Published: 25 July 2017

Abstract. The question of how to derive and present uncertainty information in climate data records (CDRs) has received sustained attention within the European Space Agency Climate Change Initiative (CCI), a programme to generate CDRs addressing a range of essential climate variables (ECVs) from satellite data. Here, we review the nature, mathematics, practicalities, and communication of uncertainty information in CDRs from Earth observations. This review paper argues that CDRs derived from satellite-based Earth observation (EO) should include rigorous uncertainty information to support the application of the data in contexts such as policy, climate modelling, and numerical weather prediction reanalysis. Uncertainty, error, and quality are distinct concepts, and the case is made that CDR products should follow international metrological norms for presenting quantified uncertainty. As a baseline for good practice, total standard uncertainty should be quantified per datum in a CDR, meaning that uncertainty estimates should clearly discriminate more and less certain data. In this case, flags for data quality should not duplicate uncertainty information, but instead describe complementary information (such as the confidence in the uncertainty estimate provided or indicators of conditions violating the retrieval assumptions). The paper discusses the many sources of error in CDRs, noting that different errors may be correlated across a wide range of timescales and space scales. Error effects that contribute negligibly to the total uncertainty in a single-satellite measurement can be the dominant sources of uncertainty in a CDR on the large space scales and long timescales that are highly relevant for some climate applications. For this reason, identifying and

characterizing the relevant sources of uncertainty for CDRs is particularly challenging. The characterization of uncertainty caused by a given error effect involves assessing the magnitude of the effect, the shape of the error distribution, and the propagation of the uncertainty to the geophysical variable in the CDR accounting for its error correlation properties. Uncertainty estimates can and should be validated as part of CDR validation when possible. These principles are quite general, but the approach to providing uncertainty information appropriate to different ECVs is varied, as confirmed by a brief review across different ECVs in the CCI. User requirements for uncertainty information can conflict with each other, and a variety of solutions and compromises are possible. The concept of an ensemble CDR as a simple means of communicating rigorous uncertainty information to users is discussed. Our review concludes by providing eight concrete recommendations for good practice in providing and communicating uncertainty in EO-based climate data records.

1 Introduction

Few scientists would dispute the principle that an estimate of uncertainty should be given with every measured value. However, meaningful adherence to this simple principle can be challenging, and in practice researchers commonly encounter datasets for which uncertainty information is generic, misleading, or absent. Climate data records (CDRs) are not immune to this problem, despite the fact that climatic signals are usually subtle (e.g., Kennedy, 2014; Mahlstein et al., 2012; Flannaghan et al., 2014; Barnett et al., 2005), which adds to the importance of rigorous uncertainty characterization in CDRs (e.g., Immler et al., 2010).

The question of how to derive and present uncertainty information in CDRs has received sustained attention within the European Space Agency (ESA) Climate Change Initiative (CCI; Hollman et al., 2013). Like the National Oceanic and Atmospheric Administration CDR programme (Bates et al., 2016), the CCI programme generates CDRs addressing a range of essential climate variables (ECVs; Global Climate Observing System, 2010; Bojinski et al., 2014). Here, we review the nature, mathematics, practicalities, and communication of uncertainty information in CDRs from Earth observations. We highlight some of the challenges that developing good uncertainty information presents and give examples of recent progress drawn from the experience of several CCI projects.

2 The requirement for uncertainty information

The environment and climate of Earth are changing (e.g., IPCC, 2013), and these changes reflect both profound human influences on the Earth system and natural variability. Scientific progress in understanding contemporary changes has great importance in constraining future changes that may have far-reaching consequences for society. For public understanding, policy development, and climate assessments, climatic changes and trends in recent decades need to be calculated. In this context, quantified observational uncertainties are required that reflect the degree to which the observing system is stable. The "system" here includes all components that can affect the values in the CDR, from the platform and sensor to software parameters and (where relevant) human judgements. Stability is the time rate at which systematic errors in the CDR may accumulate and needs to be understood so that artefacts arising from the limitations of observing systems are not misinterpreted as real changes or trends.

There is major international scientific effort in modelling the climate and its many component systems, and this is a major application that requires CDRs with quantified uncertainties. CDRs underpin climate model evaluation and improvement by providing references that can be used to identify model deficiencies. Model–data comparisons require appropriate skepticism about both the model and the data, since errors in both can be misleading (e.g., Notz, 2015; Massonnet et al., 2016). Modellers need confidence in discriminating model–data discrepancies that unambiguously indicate model deficiencies from those where observational errors are significant. Feedback gathered by CDR producers (e.g., Rayner et al., 2015) shows that modellers find it too time consuming to develop a level of appreciation of observational datasets that allows them to make confident judgements about such matters. For this reason, CDRs need to include validated uncertainty information that modellers trust for contextualizing model–data discrepancies. Until this is achieved, modellers will continue to rely on heuristics as being representative of observational uncertainty, a strategy that may or may not be valid depending on the case in point.

Uncertainty in CDRs also matters for data assimilation. Reanalysis runs of atmospheric forecasting models (e.g., Dee et al., 2011; Kobayashi et al., 2015) provide useful, dynamically consistent information about the climate system over recent decades. The analyses include inferred fields of variables that are practically unobservable and/or were not historically observed on a global scale. Reanalyses are among the most widely used datasets in geosciences because of their information content and spatio-temporal completeness. Reanalyses are created by data assimilation, which brings observations and models together by using the observations to constrain the evolution of the model towards reality. The combination involves weighting the impact of different observations together and weighting the influence of observa-

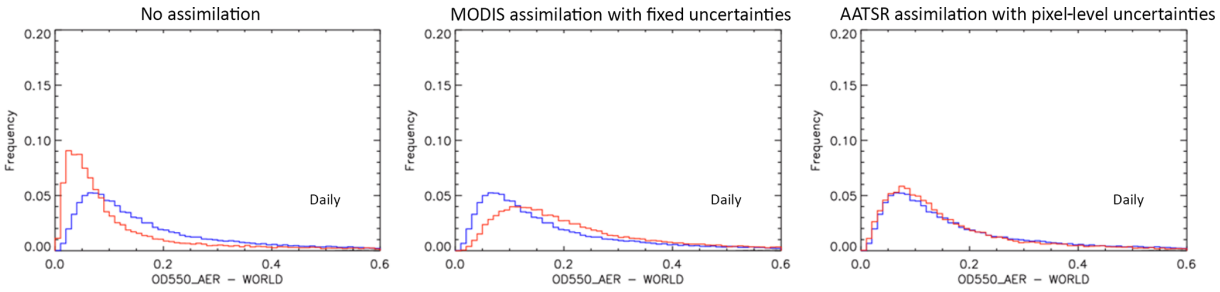

Figure 1. The benefit of pixel-level uncertainties in assimilating aerosol optical depth (AOD) estimated at 550 nm into the Monitoring Atmospheric Composition and Climate (MACC) model. Each panel shows a distribution of AOD in the MACC model (in red) matched to 29 528 AERONET ground-based AOD values (in blue): (left) no data assimilation; (centre) assimilation of MODIS retrievals; (right) assimilation of AATSR retrievals. The AERONET-measured values have negligible uncertainty compared to satellite data. The MODIS data were the dark target AOD dataset (collection 5.1), which was operational in MACC using fixed (generic) uncertainty estimates of 0.1 over land and 0.05 over ocean. These values were chosen after bias correction and thorough testing of alternative uncertainty assumptions (Benedetti et al., 2008). The AATSR dataset was from Aerosol CCI, and its pixel-level uncertainty estimates were used (with no bias correction). The improved agreement in aerosol distribution suggests that the use of pixel-level uncertainties is beneficial.

tions relative to the internal evolution of the model. Ideally, uncertainty estimates should be available for each observation so that more certain observations have more influence on the analysis. Densely sampled, numerous data (such as from satellites) can inappropriately overwhelm other observations if these data are subject to errors that correlate across space and time and therefore do not "average out". Ideally, spatio-temporal correlation should be understood and represented in the observational covariance matrices that weight satellite observations to avoid undue influence on the analysis. The requirement for uncertainty information goes beyond generic estimates at the dataset level: information is needed on which data are more or less certain and how their errors are structured in space and time. Where information provided in CDRs about observational uncertainties is limited, generic assumptions are generally made, leading to suboptimal outcomes; an example is shown in Fig. 1.

3 Terminology: error, uncertainty, and quality

The terms "error" and "uncertainty" are often unhelpfully conflated. Usage should follow international standards from metrology (the science of measurement), which bring clarity to thinking about and communicating uncertainty information. Formal definitions are found in the International Vocabulary of Metrology (Joint Committee for Guides in Metrology, 2008a). Adopting the "error approach" therein to describe the process of measurement, we have the following:

- the *measurand*: a quantity to be measured;
- *measurement*: the process of experimentally obtaining one or more measured values that can reasonably be attributed to a quantity;
- the *measured value*: the result of a measurement obtained to quantify the measurand:
- the *error*: the measured value minus the true value of the measurand. In practice the error is unknowable, except when the measured value can be compared with a reference value of negligible uncertainty;
- and the *uncertainty*: a non-negative parameter characterizing the dispersion (spread) of the quantity values attributed to a measurand, given the measured value and an understanding of the measurement.

Thus, a measured value results from the measurement of a target quantity, called the measurand. It is only an estimate of the measurand because various effects introduce errors into the process of measurement. These errors are unknown. Uncertainty information characterizes the distribution of values that it is reasonable to attribute to the measurand, given both the measured value and our characterization of effects causing error. Error is thus the "wrongness" of the measured value (and is unknown). Uncertainty describes the "doubt" we have about the measurand value, given the result of a measurement and our estimate of the error distribution. A classic question at a scientific meeting is the following: "What is the error in your measurement?" This is perhaps asked after a plot has been presented without "error" bars. The questioner is asking for information about uncertainty, but the technically correct answer to this question would be "I don't know the error, and if I did, I would correct for it".

Note that these technical definitions correspond well to the plain meaning of the words "error" (mistake) and "uncertainty" (doubt) as used by non-scientists. In addition to improving communication between scientists, careful usage will help scientists communicate beyond their community.

It is common for satellite datasets to include quality flags as a simple means to guide users in the usability and validity of data. This raises questions about the relationship between quality and uncertainty.

When a quantitative uncertainty estimate is provided for each pixel or datum, as advocated here, quality and uncertainty can be cleanly decoupled, giving different information to the user. The quality indicator should indicate whether both the measured value and its uncertainty estimate have been obtained under conditions such that they are expected to be quantitatively valid. With this approach, a highly uncertain measured value is not of lower quality provided that the high uncertainty is validly estimated. Data are flagged as lower quality in circumstances that violate the assumptions behind the measured value or its uncertainty estimate.

For example, consider a case in which the uncertainty estimates are known to be unrealistically small under certain conditions of illumination by the Sun. There may be contamination in the signal caused by stray radiance, for example, and no means to quantify the contamination. For these situations, a quality indicator can be used to indicate that an assumption or condition underlying the retrieval or the uncertainty estimate provided is not valid, i.e., that stray radiance may have biased the measured values by a non-negligible amount not accounted for in the uncertainty estimate.

4 Traceability of uncertainty

In addition to precise language for describing measurement uncertainty, metrology has developed a rigorous understanding of the issues surrounding measurement uncertainty in the context of developing and promulgating international measurement standards, particularly the Système International d'Unités (SI; Bureau International des Poids et Mesures, 2006). A key metrological concept is traceability through the chain of processes from the primary standard to an end-point measurement.

More generally, any measurement can be thought of as a series of transformations from the event observed to some final value. These include physical processes (such as the emission of light by a gas), measurement techniques (such as the observation of light by a detector), classifications (e.g., cloudy or clear sky), and mathematical analyses (e.g., inversion algorithms). Each transformation may be influenced by multiple effects that accumulate and propagate error. To develop a full uncertainty budget, every effect that may introduce error at any point in the chain needs to be considered, quantified (by one of various defined approaches), documented, and (if not negligible) appropriately propagated through the remainder of the chain.

Developing a more rigorous metrology of Earth observation (EO; Mittaz et al., 2017) is particularly important for CDR generation compared to EO applications in general. The applications of CDRs involve data analysis on a wide range of space scales and timescales, from process studies that are highly resolved in time and space, to decadal- and/or continental-scale assessments of subtle climate changes. To provide valid quantitative uncertainty information across this range of scales, all sources of error need to be assessed, and uncertainty propagation across scales needs to be rigorous. At larger scales of analysis, systematic effects that are small contributors to uncertainty in individual measured values may become the dominant sources of uncertainty (see Fig. 2).

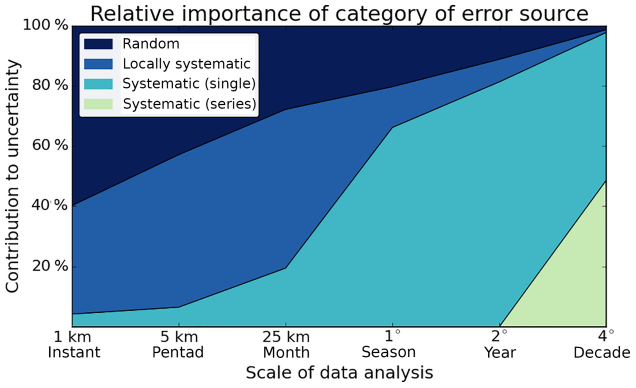

Figure 2. Contribution to the overall uncertainty from different error sources for different spatio-temporal scales of analysis of a climate data record (CDR). Conceptually, this figure is generally applicable to many climate CDRs. The particular case here is a sea surface temperature (SST) CDR derived from a series of typical meteorological sensors. The effects causing errors are characterized by their correlation properties: noise causes random errors in SST that average out rapidly when analysing change on larger or longer scales; retrieval errors for SST have a locally systematic aspect and average out more gradually with scale; systematic errors, particularly in calibration, for a single sensor become more significant over time as the sensor ages and the calibration tends to drift; and a long CDR is comprised of data from a series of sensors which are, inevitably, imperfectly harmonized so that systematic series effects become important for the longest timescales of analysis. Reproduced with permission from https://doi.org/10.6084/m9.figshare.1483408, where full details of the scenario underlying the figure are available.

Classic metrological concerns are firstly to assess and quantify all known sources of error, and secondly to propagate uncertainty rigorously through all steps to the end result. The analogy between the problems in EO-based climatology and metrology has prompted a developing dialogue and joint projects between these communities in recent years (e.g., World Meteorological Organization and Bureau Internationale de Poids et Mesures, 2010; Woolliams et al., 2016).

5 Origin and characterization of errors

5.1 A sequence of transformations

A datum in a CDR is the end result of a sequence of transformations. Consider a simplified scenario for the transformations involved in passive remote sensing using an infrared radiometer to create a multi-mission CDR.

1. Infrared radiation emitted from a particular field of view (originating from the Earth's surface and the atmosphere path above it) is collected by the aperture of a sensor and filtered during its passage through sensor optics.

2. The filtered radiance falls on a solid-state detector, causing a voltage signal.

3. The voltage is amplified electronically.

4. The amplified signal is quantized to "counts" and recorded.

5. The scene counts are compared with counts obtained when viewing two reference targets whose temperatures are measured; via this onboard calibration process, channel-integrated brightness temperature is determined using various parameters and assumptions.

6. This brightness temperature is input to processing software that retrieves a geophysical variable to generate a CDR. This sixth step can itself be decomposed into many transformations and dependencies.

    a. Auxiliary information is also accessed by the processor, which may include a wide range of information. Some information is intrinsic to the observation and is highly certain (e.g., satellite view zenith angle, time). External geophysical datasets may be used, such as numerical weather prediction fields or surface classification, and these may or may not be provided with quantified uncertainties. All auxiliary information influences the CDR and gives rise to uncertainty.

    b. The processor typically involves a step to determine that the pixel properties are valid for the intended retrieval (screening cloudy pixels, for example). This influences the CDR through the sampling distribution of the observations.

    c. The set of observations is inverted to obtain an estimate of a geophysical quantity, such as an ECV. This inversion may be sensitive to the auxiliary information and may vary in its complexity and degree of non-linearity.

    d. A multi-mission CDR is created from datasets for several similar sensors by harmonizing discrepancies between sensors (using sensor overlap periods or other means), which modifies the datum to its final value.

    e. Many ECV estimates may be aggregated to a coarser space–time grid for the purpose of (for example) evaluating the results of a climate model run.

Every step in the above sequence is a transformation subject to effects that introduce errors. Characterizing these effects is the significant core work required to develop good uncertainty information in a CDR. The errors from each effect have certain properties which can be estimated to the degree that the effect is understood. There are several aspects to characterizing the errors from a given effect: the magnitude of uncertainty at the source, the shape of the error distribution, how the uncertainty propagates to the resulting data, and the correlation structure of the error from this source between observations.

5.2 Magnitude of uncertainty

The magnitude of uncertainty characterizes the dispersion (width) of the estimated distribution of errors. Standard uncertainty is the standard deviation of the distribution, although other coverage factors can also be used. The value of the standard uncertainty can be estimated from basic principles in some cases. An example is the uncertainty introduced by quantization of the signal, which in older sensors using relatively few bits could be a significant source of noise. In other cases, the uncertainty estimate may rely on empirical information. For example, the noise from an amplifier circuit may have been measured during pre-launch testing. Using pre-launch noise levels in an uncertainty estimate involves the assumption of stable behaviour of the amplifier during and after launch; that assumption itself can be tested for consistency with other instrument data or the noisiness apparent when observing relatively uniform targets.

In generating CDRs, we often have to deal with the multivariate case because several channels are combined to estimate a geophysical quantity. Errors in these channels are not necessarily independent, and in this case the generalization of the standard uncertainty is the error covariance matrix, which has as many rows and columns as there are channels (or other variates). The square root of an element on the diagonal of this matrix corresponds to the standard uncertainty for a particular variate.

With reference to the scenario described in Sect. 5.1, several sources of uncertainty can be identified with magnitudes that must be estimated. For example, at step 4, the combined effect of solid-state detector noise, amplifier noise, and digitization causes an uncertainty in counts. This uncertainty can be estimated by considering the dispersion of measured values when viewing a constant calibration reference. Another example is the retrieval uncertainty associated with the inverse solution that provides the geophysical retrieval from the satellite radiances (step 6c). Even with perfect data, the process of retrieval is usually ambiguous (more than one geophysical state can be associated with identical radiances). This component of uncertainty can be quantified by the simulation of retrieval outcomes compared to the simulation "truth" if a forward model for the satellite observations is available.

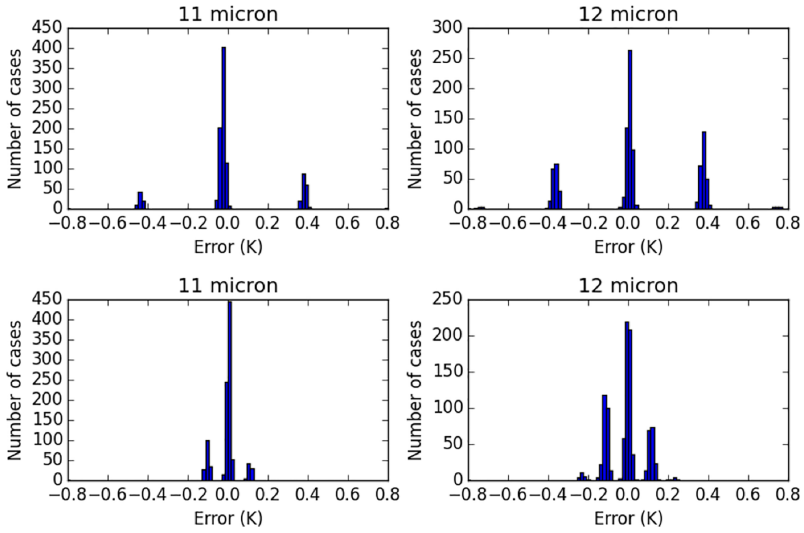

Figure 3. Distributions of single-pixel brightness temperature (BT) errors from a simulation for the detection and calibration system of an advanced very high resolution radiometer (AVHRR) for channels of different wavelength (columns) and two scene temperatures (upper row 200 K scene, lower row 300 K scene). The unit of frequency of occurrence is per thousand.

5.3 Shape of the error distribution

If the error distribution is zero-mean Gaussian, then the standard uncertainty fully describes the error distribution arising from the effect. Not all effects cause Gaussian-distributed errors. One example is the logarithmic distribution of radar backscatter errors associated with speckle. Another example is quantization (step 4 in the scenario in Sect. 5.1), as illustrated by Fig. 3, which shows a simulation of the distributions of brightness temperature for an Advanced Very High Resolution Radiometer (AVHRR) viewing a pixel with true scene temperatures of 230 and 300 K. This distribution was obtained by simulating detector noise, amplifier noise, quantization, and ideal (unbiased) onboard calibration. The separated peaks are the effect of the AVHRR 10-bit digitization of the detector and amplifier noise. Each separated spike has a nearly Gaussian distribution with a spread that arises from errors in the calibration process: the calibration applied for a given observation arises from a finite sample of the calibration target views (an internal black body and a space view), which therefore implies some statistical uncertainty. Cases such as this require a numerical representation of error distributions and a Monte Carlo simulation for the propagation of uncertainty (see the next subsection). When quantization is negligible, which is often the case for contemporary sensors, the Gaussian distribution realistically describes the signal noise and should be characterized by the standard deviation of the error distribution, which is the standard uncertainty.

5.4 Propagation of uncertainty

Uncertainty from effects associated with a particular transformation ultimately propagate to the contents of the CDR. Gaussian errors can be propagated through linear and nearly linear transformations by standard analytic means (Joint Committee for Guides in Metrology, 2008b). Let $\boldsymbol{Y} = f(\boldsymbol{X})$ represent any of the transformations between admitting Earth-leaving radiance into the aperture of a sensor and writing a datum in a climate data record. The function f describes how one or more inputs in vector $\boldsymbol{X}$ give rise to the output(s) of the transformation in vector $\boldsymbol{Y}$. The uncertainty in the output(s) is characterized by an error covariance matrix:

$$\mathbf{U}_y = \mathbf{C}_y \mathbf{U}_x \mathbf{C}_y^{\mathrm{T}}, \tag{1}$$

where $\mathbf{U}_x$ is the error covariance matrix of the inputs, and $\mathbf{C}_y$ is the matrix of sensitivity coefficients, in which $\frac{\partial f_i}{\partial x_j}$ quantifies the influence that the ith input in $\boldsymbol{X}$ has on the jth output in $\boldsymbol{Y}$. If there are several effects indexed by e, then

$$\mathbf{U}_x = \sum_e \mathbf{U}_{x,e}. \tag{2}$$

These analytic propagation equations are a first-order approximation and are strictly valid for Gaussian-distributed errors that are sufficiently small that f is linear over the range of likely errors.

For non-Gaussian distributions and/or transformations that are significantly non-linear, Monte Carlo approaches are necessary to propagate uncertainty. A common non-linear transformation in generating some CDRs is threshold-based categorization of a set of observations, either because the CDR is

comprised of a classification (such as land cover) or because the retrieval of the geophysical variable is valid only for certain classes (such as cloud-free scenes). When observations are near a threshold, errors can cause a change in classification. Simulating the retrieval process many times can characterize the propagation of uncertainty in observations into the classification results.

5.5 Correlation structure

It is important to understand the correlation of errors because failing to account for correlation generally leads to underestimation of uncertainty and unfounded confidence in the interpretation of the CDR.

A common example of error correlation arises when a geophysical variable is retrieved from satellite imagery (step 6c in Sect. 5.1). The estimation of geophysical quantities from radiance measurements is usually an inverse problem in which there is some ambiguity and dependence on auxiliary parameters (whether explicit or hidden). Both ambiguity and parameter dependence tend to cause retrieval errors that are shared to some degree between nearby image pixels; i.e., the errors are locally correlated. The correlation length scale for such retrieval errors depends on the effect.

For example, aerosol optical depth may be estimated across a particular scene in reflectance imagery assuming a size distribution and refractive index that systematically differ from reality; errors are therefore expected to be correlated between pixels on the scales of variation in true aerosol properties. More generally, retrieval errors are correlated on the space scales and timescales of atmospheric variability whenever retrieval ambiguity is related to atmospheric conditions (e.g., Merchant and Embury, 2014; Buchwitz et al., 2013). The errors may be de-correlated between different overpasses (because atmospheric conditions change; e.g., Reuter et al., 2014) but are strongly related for adjacent pixels from a single-orbit overpass. Figure 4 illustrates this for the case of a sea surface temperature retrieval (SST) with simulated retrieval errors that are correlated geographically and de-correlated in time.

Systematic effects cause errors with structure across a whole dataset, or at least across large space scales and long timescales within a dataset. The term "systematic error" is sometimes loosely equated to "bias", but the concept of a systematic effect is in truth more subtle, since a systematic effect can produce zero-mean errors with no bias overall. Systematic effects can be defined as those that cause errors which one could in principle correct given the necessary quantitative understanding. For example, a CDR may be derived from a series of sensors with differing calibrations. Even if the series is adjusted to compensate for inconsistency between the calibration of different sensors (step 6d), there is uncertainty in doing this; errors in the adjustment parameters potentially affect the entire data record from a particular sensor. These systematic errors may correspond to an overall bias, but more commonly they have some geographical and/or temporal structure. However, in principle, given more complete information, corrections for these errors could be devised.

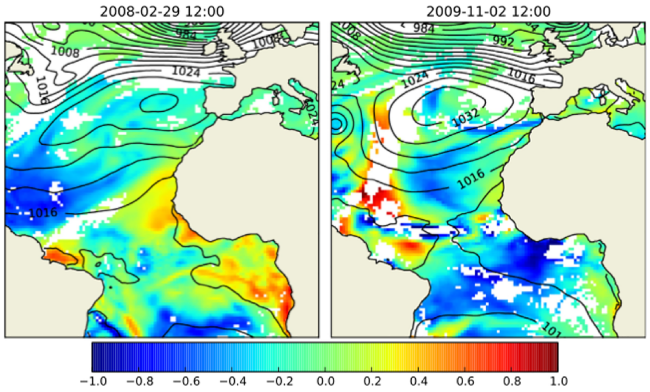

Figure 4. Simulation of locally correlated errors in the retrieval of sea surface temperature (SST) overlaid with surface pressure contours to indicate length scales of atmospheric variability. The simulated retrieval errors are for a noise-free sensor with a calibration that is perfectly known. The errors therefore arise solely from intrinsic ambiguity in inverting the observed radiances to SST. Note that there is no simple relationship between the SST errors and the atmospheric features associated with synoptic weather systems. The white areas indicate 100 % cloud cover. Reprinted from Merchant and Embury (2014) with permission from Elsevier© (2014).

Local correlations and the correlation of errors from systematic effects need to be properly accounted for when creating "level 3" versions of CDRs, i.e., gridded products involving the averaging of full-resolution data. If the correlated nature of the errors is neglected, the uncertainty estimate for the gridded data will be poor (usually an underestimate). In averaging data subject only to independent random errors, it is well known that the effect of the errors on the average decreases with the square root of the number of contributing data, but local correlation decreases the averaging-out of errors. In the extreme of pixel uncertainty dominated by an error source that is fully common across a grid cell, there is no reduction in uncertainty from averaging many pixels. The impacts of error correlation on the uncertainty of the grid-cell average can be evaluated using Eq. (1) with the required off-diagonal terms in U_x. When a grid cell is not completely sampled by the full-resolution data, there is an additional uncertainty not quantified by Eq. (1) associated with the unobserved part of the cell. See Reuter et al. (2010) and Bulgin et al. (2016a) for examples of parameterization development for subsampling uncertainty.

6 Which types of uncertainty information are used?

The previous section introduced four considerations that are useful in thinking about the uncertainty from a given effect:

the magnitude of uncertainty at the source, the shape of the error distribution, how the uncertainty propagates to the resulting data, and the correlation structure of the error from this source between observations. These considerations apply quite generally. However, the nature of the responses depends on the particularities of the CDR being considered. There is a range of forms which uncertainty information can take and a variety of empirical and theoretical methods used to estimate uncertainty.

Quantitative measures of uncertainty describe the doubt we have about the measurand, given the measured value, in numerical terms. Conceptually, the provided numbers quantify the dispersion (i.e., spread) of the estimated error probability distribution function (PDF). Options for characterization are varied, including percentiles, confidence intervals, maximum range of error, multiples of the standard deviation, covariance matrices, distribution histograms, and misclassification rates.

Standard uncertainty is a highly informative measure when the error distribution is close to Gaussian. For example, in the case of sea surface temperature (SST), errors are reasonably well described by a Gaussian distribution with a standard deviation that can be modelled by uncertainty propagation (Merchant and Le Borgne, 2004; Embury and Merchant, 2012). Even in this relatively simple case, there are subtleties. Sea water freezes at around −1.8 °C. Even though the measurement error distribution remains Gaussian when the retrieved temperature approaches the freezing point, the distribution of credible SSTs becomes asymmetric given the additional knowledge that SST below −1.8 °C is precluded.

The dispersion of errors is sometimes better described using fractional uncertainty. This approach is typically more appropriate for data such as ocean chlorophyll concentration or atmospheric aerosol optical depth (AOD). In both these cases there is a strict lower limit to valid data of zero, and both the measured values and the standard uncertainty can vary in value over orders of magnitude with larger uncertainty in absolute terms when the measured values are large. Quoting a fractional uncertainty is an appropriate approach and is equivalent to stating a standard uncertainty for logarithm-transformed data. However, for values near zero, standard uncertainty may be more representative. For example, effects associated with surface brightness introduce an uncertainty in AOD that is the dominant uncertainty for low-aerosol scenes. Thus, the Global Climate Observing System (2010) recommends uncertainty modelling using a combination of absolute and fractional uncertainty for CDRs of aerosol optical depth.

Some CDRs refer to categorical ECVs, such as the status of the land cover at a given place, whether the land at a given location has recently burned, or whether the land is covered by a glacier. Here an appropriate statement of uncertainty can be probabilistic: how probable is it that the status will be other than indicated? When the classification uses a Bayesian approach, like the maximum likelihood estimation, the probability to belong to the output class is naturally available. For non-probabilistic classifiers ("random forest" for instance), a proxy for class membership probability can be defined as the number of trees in the ensemble voting for the final class (Loosvelt et al., 2012). Similarly, the distance to the optimal separating hyperplane in the feature space can be used in support vector machine classifications (Giacco et al., 2010).

Table 1 shows the variety of ECVs and the corresponding uncertainty information in the CCI programme. The maturity of the uncertainty information presently provided varies, and for some cases uncertainty estimation is not yet achieved.

Most projects in the CCI programme adopted standard uncertainty as the provided uncertainty information (Table 1), which is a convergence that arose after sustained discussion across the programme and which is in line with metrological guidance. Exceptions include ECVs for which the geophysical data are categorical rather than numerical as discussed above. However, there is a wide range of methods employed to develop this uncertainty information documented in the varied contents of the uncertainty characterization reports prepared for each CDR. (For these reports and other documentation, refer to http://cci.esa.int.)

7 Validation of uncertainty

Quantified uncertainty information provided in CDRs needs to be validated, i.e., evaluated by independent means to establish quantitative realism and the credibility of the uncertainty estimates. Many validation studies in the literature consider the validation of measured values, but the validation of attached uncertainty information is less common. Indeed, where specific uncertainty estimates are not provided with measured values, measured-value validation is often seen as a method for deriving generic uncertainty information (based on the validation discrepancies).

The primary means of validating uncertainty estimates is to extend traditional measured-value validation in which satellite and in situ reference data are compared. Validating uncertainty information in a CDR is more challenging than validating a measured value because it requires the quantification of three contributions to the observed differences between the values measured from space and on the ground (e.g., Wimmer et al., 2012; Dils et al., 2014):

- the uncertainty for each CDR data value (the uncertainty estimate in the CDR product that is to be validated);
- the uncertainty for each reference measured value being used as a validation point; and
- the magnitude of real geophysical variability caused by the different nature of the satellite and validation measurements.

The third contribution can require significant effort. Real geophysical variability between measurements of nominally

Table 1. Essential climate variables addressed in the ESA Climate Change Initiative.

Essential climate variable	Comments on nature of variable	Product characteristics	Uncertainty information provided	Basis on which uncertainty is estimated
Aerosol optical depth (AOD)	AOD is a continuous, non-negative, log-normally distributed variable	Satellite swath ($10\times10\,km^2$ super pixels) and gridded (1° grid daily and monthly)	Standard uncertainty given for each pixel level in swath product; averaged uncertainty given for each cell in gridded products	Propagation of sensor noise through retrieval process; context-specific (surface, aerosol type) estimate of retrieval uncertainty
Cloud properties	Cloud properties are composed of several sub-variables (temperature, height, fraction) which are continuous non-negative variables	Satellite swath (≥ 5 km), gridded (0.05° grid) and averaged (0.5°, daily, monthly) estimates	Standard uncertainty given at pixel level in gridded swath product and averaged for each cell in gridded products	Propagation of sensor noise through retrieval process; (optimal estimation) context-specific (surface) estimate of retrieval uncertainty; propagation of uncertainty to grid boxes accounting for correlation
Glaciers	Glacier outlines derived from optical satellite data with manual intervention	Outlines are provided in a vector format, scene by scene; a geospatial database addresses 200 000 glaciers globally	Not regularly determined; some tests have been published	Various methods, including multiple digitizing by analysts; appropriate validation data are generally missing
Greenhouse gases (XCO_2, XCH_4)	XCO_2 and XCH_4 are defined as atmospheric, dry-air, column-averaged mole fractions of CO_2 and CH_4	One file per day, including XCO_2 and XCH_4, plus additional information for surface flux inversions; soundings have surface footprints of $\sim$ 10 to $\sim$ 60 km depending on sensor	Standard uncertainty of XCO_2 and XCH_4 (per sounding) plus averaging kernels (AK) and a priori concentration profiles	Propagation of sensor noise (and a priori uncertainty) through retrieval process and error scaling to match validation statistics
Land cover	A categorical variable describes the terrestrial surface annually in 22 discrete classes (from the UN Land Cover Classification System)	Annual land cover maps at 300 m depicting land cover change from 1992 to 2015	Class uncertainty is available at the map and class level; standard uncertainty of composite surface reflectance is provided at pixel level	Class uncertainty is computed from confusion matrix built on independent statistical validation process
Ocean colour	Variables of bio-optical relevance with high dynamic range (4 decades)	Chlorophyll *a* concentration, spectrally resolved inherent optical properties, diffuse attenuation coefficient at 490 nm, membership of optical classes	Standard uncertainty and bias estimates for all products except backscattering coefficient	Uncertainty assignment based on product comparison with match-up in situ data for each optical class, applied per pixel according to class membership
Ozone	Ozone total-column and vertical profiles	Ozone profiles from limb sounders with $\sim$ 3 km vertical and $\sim$ 300 km horizontal resolution; ozone profiles from nadir sounders with $\sim$ 4 km vertical resolution; analysed and gridded versions of profiles and total column	Standard uncertainty estimates are given for each ozone value in each record	Measurement noise propagated through the retrieval process and to the higher levels of data products, randomly varying parameter errors; sampling uncertainties

Table 1. Continued.

Essential climate variable	Comments on nature of variable	Product characteristics	Uncertainty information provided	Basis on which uncertainty is estimated
Sea ice	Sea ice concentration (SIC), thickness (SIT)	SIC: daily, gridded data at between ∼ 12 and ∼ 50 km grid spacing; SIT: presently Arctic winter only, monthly 100 km gridded freeboard and thickness	SIC: standard uncertainty estimated from retrieval and gridding; SIT: presently no uncertainty provided	SIC: the retrieval uncertainty is based on statistical spread of retrieval at tie points of known SIC; parameterization for gridding uncertainty
Sea level	Sea level is continuously variable in space and time; global variations should be consistent with the conservation of water mass in the climate system	Active remote sensing along ground tracks; analysed to monthly 0.25° grid	Standard uncertainty for each sea level determination along ground tracks; standard uncertainty in inter-annual global mean sea level (Ablain et al., 2015)	Uncertainty is inferred by generalized least squares, where the error covariance matrix is built from altimeter correction uncertainties
Sea surface temperature (SST)	Temperature is continuously variable in space and time with a lower bound at the freezing temperature of sea water	Satellite swath (≥ 1 km), gridded (0.05° grid) and gap-filled (0.05°, daily) SST estimates	Standard uncertainty given at pixel level in swath product and for each cell in gridded and gap-filled products; component uncertainty contributions also available	Propagation of sensor noise through retrieval process; context-specific estimate of retrieval uncertainty; sampling uncertainty estimate in cell means
Soil moisture	Microwave retrievals represent moisture content in a thin surface layer (1–5 cm); no data when soil is frozen, snow-covered, or overlain by very dense vegetation	Daily (00:00 UTC) gridded (0.25°) global data; three data records: (i) merged active, (ii) merged passive, (iii) merged active passive microwave data	Standard uncertainty given for each soil moisture value in each of the three data records; additionally, quality flags are provided	Propagation of sensor noise through retrieval process, including context-specific estimate of retrieval uncertainty; uncertainties introduced by sampling not yet characterized

the same measurand arises for many reasons, depending on the ECV considered. The spatial location of the measurements can differ (including the tolerance for spatial mismatch and the effect of point measurement vs. area average over a satellite pixel). The measurements are likely not perfectly synchronized, and the geophysical state may have evolved in the intervening time. Definitional differences are common between measurands even when nominally equivalent, such as a remotely sensed measurement being sensitive to a weighted average of some vertical profile of a variable, whereas the reference measurement is made at discrete heights or depths. In some cases, validation must be performed using reference data for a measurand that is closely related, but not exactly the same (a definitional discrepancy).

In the case of satellite CDR data, x_{sat}, containing standard uncertainty estimates, u_{sat}, the validation of the CDR uncertainty information can be based on the distribution of the ratio

$$\frac{x_{\mathrm{sat}} - x_{\mathrm{ref}}}{\sqrt{u_{\mathrm{sat}}^2 + u_{\mathrm{ref}}^2 + u_{\mathrm{mis}}^2}}, \quad (3)$$

where x_{ref} is the value of the reference (validation) data, u_{ref} is the uncertainty in the reference data, and u_{mis} is the geophysical variability arising from temporal, spatial, and definitional mismatch between the satellite and reference data. If the uncertainties and variability are correctly quantified, this ratio will be normally distributed with a standard deviation equal to unity. The better the quality of the reference data (the smaller u_{ref}) and the better the match of satellite to validation data (the smaller u_{mis}), the more sensitive the validation of u_{sat}.

An example validation of uncertainty based on this principle is shown in Fig. 5. In this case, the data are cloud-top height (CTH) from Cloud CCI retrievals driven by an interpretation of the cloud top temperature in thermal imagery and matched to independent CTH measurements made by CALIPSO using laser ranging. The CALIPSO validation data have, in this case, negligible uncertainty; mismatch uncertainty is also neglected. The plots therefore show the histogram of discrepancy in CTH between the two observations divided by the uncertainty estimated in the Cloud CCI retrieval process. The Gaussian that best fits the main peak is also shown with its calculated width. In the case of ice clouds, the product uncertainty is underestimated by around 10 %. For liquid clouds, the analysis reveals a systematic effect. For both ice and liquid clouds, there are tails to the

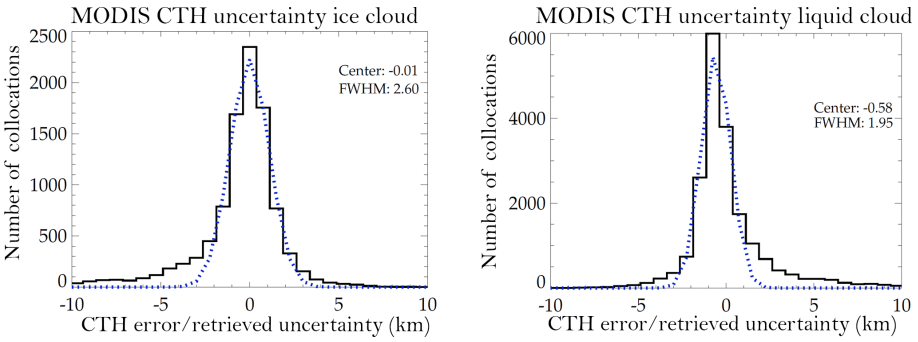

Figure 5. Example of uncertainty validation using the distribution of differences between matched cloud top heights measured by Cloud CCI (data) and CALIPSO (CALIPSO values minus those from MODIS AQUA Cloud CCI) for a single day, 20 June 2008 (solid black); (left) for ice clouds, (right) for liquid clouds. The plots show the histograms of the CTH error (the difference in retrieval compared to validation data that is assumed to have negligible uncertainty) divided by the stated retrieval uncertainty. For ideal uncertainty estimates the full width at half maximum (FWHM) of the fitted Gaussian distribution (dashed blue) would be 2.35.

distribution where the magnitude of disagreement exceeds 4 times the estimated uncertainty. This indicates that uncertainty is underestimated for these cases, since such outliers would be very rare if the estimated uncertainty were appropriate.

In addition to the extended validation described above, triple collocation techniques (McColl et al., 2014) have been used to assess uncertainty estimates in near-surface wind speed (Stoffelen, 1998), soil moisture (Gruber et al., 2016), and other remotely sensed variables. For valid quantitative estimation or validation of uncertainty, this technique requires three sources of collocatable data with errors that are independent and random (both between the data sources and within each data source) and assumes that sampling mismatches and differences in the definition of the measurands between the three types of data are negligible. Other uncertainty validation methods are briefly reviewed in Sofieva et al. (2014). The uncertainty arising specifically from instrument noise can be validated using an Earth target that is assumed not to vary, e.g., white sands in New Mexico for reflectance validation. In this case, validation is not against independent measurements, but it is performed by using repeated observations by the same instrument. Such analyses would be more robust if the geophysical standard could be traced to a more controlled reference, which would require more support for repeated accurate measurements of the Earth target from the ground (Loew et al., 2017). For categorical ECVs, such as land cover type, a degree of uncertainty validation can be obtained by verifying that the estimated misclassification rates in the product are stable with respect to reasonable ranges of classification parameters. For instance, if classification is based on training a classifier using a dataset split into calibration and validation ("train" and "test") subsets, the process can be repeated many times with a different random division into train and test subsets, which allows the dispersion in the misclassification rates to be characterized.

8 Presenting uncertainty information in climate datasets

When determining how uncertainty information is to be included in the CDR, various requirements can conflict (Table 2). The core conflict is between providing for applications that require only summary information that discriminates more and less uncertain data, and providing for applications that demand detail about uncertainties that is sufficient to calculate uncertainty in the quantities derived from the CDR (averages in space and time, temporal differences, integrals, trends, and fluxes). Data producers themselves are users of their low-level (e.g., full resolution, orbital) products when they create higher-level products (e.g., gridded datasets and gap-filled analyses). In order to provide realistic uncertainty information at the higher level, they may require fine-grained uncertainty information for the low-level CDR, such as separate quantification of uncertainty at the pixel level from effects with distinct spatio-temporal correlation properties. Such detailed information is complex for non-expert users and is an unnecessary data volume for those with applications requiring, for example, only the total uncertainty.

The increase in data volume involved in providing uncertainty information is far from a minor consideration. The volume of data required for a comprehensive description of uncertainty, including the degree of error correlation, can be many times the volume of the measured values. For example, a full error covariance matrix for N measured values is $N \times N$. Data volume and processing limits are thus significant obstacles to comprehensive brute-force calculations of uncertainty. Insight and imagination are required to develop treatments of uncertainty that meet the requirements for rigour in CDR applications and are computationally tractable. Data producers can develop different versions of products that are light and heavy with respect to uncertainty information. Data delivery systems can be developed that allow users to select on download consistent uncertainty information to the degree

Table 2. Generic requirements for uncertainty information in climate data records, illustrating potential contradictions between the requirements for different data applications.

Requirement	Implications	Conflicts & solutions
1. Minimize data volume for users to download.	Provide only key summary information on uncertainty, such as the total uncertainty (or the means to calculate uncertainty) for each measured value.	Conflicts with need for detailed uncertainty information for some purposes (cf. 3, 4, and 5). More complete uncertainty information can be made available separately to core data products.
2. Data should be easy to read and understand.	Use standard metrological vocabulary to express uncertainty. Uncertainty data should be easy to associate with measured values.	Some established community standards and conventions include uncertainty vocabulary that is inconsistent with best practice. Work with community standards to converge practices.
3. Provide sufficient uncertainty information to allow correct propagation of uncertainty to spatial and temporal averages of data.	Uncertainty components from errors with different spatial correlations need to be separately quantified with correlation information (e.g., length scales, covariance matrix).	Increases data volume (cf. 1). Increases complexity of dataset (cf. 2). Could provide two versions of data, one with summary and the other with comprehensive uncertainty information, with guidance as to which is needed for different purposes.
4. Provide information about temporal stability of observations and/or evolution of trend uncertainty over time (up to decades).	Information is provided on temporal correlation of errors, particularly arising from long-term systematic effects.	Full spatio-temporal covariance matrix for CDR is challenging to calculate or parameterize and is likely infeasible to distribute. More general estimates of overall stability can be made. Ensemble approaches have been proposed.

of detail they require. There is likely no single strategy that is optimal for every ECV.

A user consultation meeting on uncertainty information in SST CDRs (Rayner et al., 2015) explored these issues with a range of users, including "power users", in applications such as data assimilation for reanalyses and centennial-scale climate modelling. An interesting conclusion from the workshop is that many users are interested in ensemble versions of EO-based CDRs, despite the multiplied data volume this implies. The purpose of the ensemble CDR is to represent the effect of all error sources on all spatio-temporal scales. The motivation behind the ensemble approach is two-fold (e.g., Morice et al., 2012). First, the user does not need to engage deeply with the origins and correlation structure of errors in the CDR or their implications for the application, since these are captured in the differences between ensemble members. Second, for some applications it is simpler to rerun a process several times with different ensemble members than to propagate uncertainties through the process, particularly when error structures exist across a wide range of scales. These motivations do not apply to every application, and the ensemble approach is less attractive to users facing constraints on data volume or processing power. The ensemble approach raises issues and opportunities for the data provider. Uncertain auxiliary parameters for in the processing can be sampled across their plausible range rather than relying on a single best estimate. However, the strategy for creating an ensemble requires careful design, and there are subtleties to be addressed, such as whether a "best" member is supplied, how large an ensemble is appropriate, and what the ensemble spread represents. Within the CCI programme, the ensemble approach has been adopted only experimentally thus far (e.g., Reuter et al., 2013).

The producers of CDRs therefore have to reflect on the expected applications of their data and make a judgement about the balance to strike between conflicting requirements, such as ease of use versus the completeness of the uncertainty information. Nonetheless, the collective experience across the CCI ECVs represented in Table 1 shows that the provision of per-datum standard uncertainty has emerged as a rigorous but simple approach adopted for all ECVs (other than products comprised of classifications). The standard uncertainty provided is generally the total from all sources of error, although uncertainty components with different error correlation structures are additionally provided in one case. Although not sufficient for every possible application, quantifying the total standard uncertainty per datum in a CDR product emerges as a baseline standard for future good practice.

9 Good practice for uncertainty quantification

One perspective on what constitutes good practice in uncertainty quantification has been embedded in recently proposed metrics for CDR maturity. Building on the work of Bates and Privette (2012) for the NOAA Climate Data Records Program, Schulz et al. (2015) have proposed a system maturity matrix (SMM) for assessing CDR-generating capacity. The SMM includes criteria for assessing the maturity of uncertainty characterization, including linkage to standards,

Table 3. Criteria for scoring the maturity of aspects of a CDR generation system, including criteria for uncertainty characterization, taken from the system maturity matrix of Schulz et al. (2015).

Climate data record (CDR) maturity evaluation guidelines			
Maturity	Standards	Validation	Uncertainty quantification
1	None	None	None
2	Standard uncertainty nomenclature is identified or defined	Validation using external reference data for limited locations and times	Limited information on uncertainty arising from systematic and random effects in the measurement
3	Score 2 +; standard uncertainty nomenclature is applied	Validation using external reference data for global and temporal representative locations and times	Comprehensive information on uncertainty arising from systematic and random effects in the measurement
4	Score 3 +; procedures to establish SI traceability are defined	Score 3 +; (inter)comparison against corresponding CDRs (other methods or models)	Score 3 +; quantitative estimates of uncertainty provided within the product characterizing more or less uncertain data points
5	Score 4 +; SI traceability partly established	Score 4 +; data provider participated in one international data assessment	Score 4 +; temporal and spatial error covariance quantified
6	Score 5 +; SI traceability established	Score 4 +; data provider participated in multiple international data assessments and incorporates feedback into the product development cycle	Score 5 +; comprehensive validation of the quantitative uncertainty estimates and error covariance

degree of validation, the approach to uncertainty quantification, and the degree of automation of quality monitoring. The originators are clear that the purpose of assessing a CDR system against the SMM is to identify priorities for investment in developing a CDR in support of routine climate information and assessments. The overall maturity score is not an indicator of the scientific value of a dataset, which could be very high for a new variable obtained by a system with low maturity.

For multiple factors in CDR generation, the SMM maps the status of a CDR system on a scale from 1 (low maturity) to 6 (high maturity). The content of the SSM relevant to uncertainty, validation, and quality is reproduced as Table 3. A score of 2 on the uncertainty quantification criterion corresponds to the provision of limited information, such as estimates of uncertainty that are generic (i.e., that describe the typical uncertainty for the dataset as a whole). At the next maturity score, the provided information is still at the level of the dataset, but it is comprehensively described and quantified, which suggests that the nature of the effects causing error is determined. To move to a score of 4, this understanding is applied to develop uncertainty information in the product that is specific to each datum and capable of discriminating between more and less certain data. A score of 5 corresponds to providing a quantification of the correlation structures in errors via covariance information or other means. For practical purposes, since covariance matrices can be large, this provision is not necessarily required to be within the product per datum. However, feasible approaches may be found that satisfy this maturity criterion at a per-datum level, such as the decomposition of total uncertainty into dominant components arising from effects with distinct, quantified correlation structures (e.g., Bulgin et al., 2016b). The highest maturity score of 6 is obtained when the estimated uncertainty magnitudes and error correlation structures are thoroughly validated.

It is not the purpose of this paper to discuss the general merits of the maturity matrix approach to evaluating CDR systems. However, it is clear that if CDR producers address uncertainty using the perspectives in this paper, they will achieve a high maturity score in this aspect of the SMM.

This paper has demonstrated the complexity of developing good uncertainty information for climate dataset users. The aspiration to provide per-datum uncertainty estimates at all product levels and all product versions at all spatio-temporal scales is very challenging and not fully achieved. It is clear that developing and validating uncertainty estimates involves effort comparable to developing the retrieval itself. There is a lot of diversity in the nature of CDRs and the errors present in them. The details for good practice in describing the uncertainty in CDRs vary accordingly. Nonetheless, it is useful to state some general principles that emerge from the previous sections.

1. Include quantitative uncertainty information within the dataset. (Don't expect users to find uncertainty information by reading related papers.)

2. Follow metrological practice for quantifying uncertainty. The baseline good practice is to provide the total standard uncertainty for numerical variables.

3. Uncertainty estimates (or the means to calculate them) should be provided per datum in CDRs for which uncertainty varies significantly so that the uncertainty information discriminates which data are more and less certain.

4. Assuming per-datum uncertainty information is provided, avoid redundancy of this information with quality flags. Do not flag high-uncertainty data as "bad" if a valid estimate of that high uncertainty is provided; instead, use quality flags to indicate the level of confidence in the validity of the provided uncertainty and retrieval assumptions.

5. Define what uncertainty information is given in the CDR in the product documentation.

6. Describe in the product documentation the main effects causing errors, how uncertainty varies within the dataset, how errors may be correlated in time and space, and under what circumstances estimated uncertainty may be invalid (and flagged as such).

7. Use validation to evaluate both retrieved quantities and associated uncertainty estimates.

8. Propagate uncertainty appropriately (accounting for error correlation) and consistently when creating aggregated products.

10 Data availability

Examples in this paper draw on datasets from the ESA CCI programme, which are available from http://cci.esa.int.

11 Conclusion

Quantifying and validating uncertainty information is challenging. The challenge is particularly great when using complex observational systems to meet the data requirements for climate applications. The form of uncertainty information may differ according to the nature of the target essential climate variable. In general, however, the aim is to provide a justified (validated) quantification of uncertainty that allows users to know which data are more or less certain within the product.

There are many sources of error (effects) that influence the values populating a climate data record. Uncertainty is not generally provided in fundamental climate data records (level 1 products) in a form sufficient to support per-datum propagation to estimate uncertainty in derived climate data records, so there are constraints on what is practical. The efforts of CDR producers must focus on identifying dominant sources of error, bearing in mind that effects of a relatively small magnitude in a single datum may be the dominant effect on a large space–time scale and therefore may be relevant for climate applications. There is unavoidably a need to develop a good understanding of many error sources and not just "instrument noise". At the same time, one cannot wait for the perfect uncertainty budget: producers must provide CDRs using the best available knowledge. When some error sources are as yet unquantifiable, users benefit from simple, accessible descriptions of the potential uncertainty not estimated in the product.

The means of quantifying uncertainty vary across ECVs, depending on factors such as the nature of the geophysical retrieval (ranging across physics-based inversion methods, to empirical relationships and manual interpretation) and the availability of validation data. Uncertainty contributions may be modelled using a detailed uncertainty budget or estimated from the spread of outcomes across Monte Carlo simulations. Again, pragmatism is often required to obtain a timely estimate.

The idea of validation should encompass the validation of the data and the uncertainty information associated with data. The validation of uncertainties (and the measured values themselves) can be limited by the availability of reference data.

Uncertainty concepts can be confusing, and user needs vary. CDR producers can help by providing versions of products with "simple" (but inevitably partial and approximate) uncertainty information. Documentation must make clear what the provided information is (and is not) telling users. We have noted that ensemble methods may be able to provide users with conceptual simplicity and quantitative rigour, although at the expense of practical issues in terms of data volume.

The use of well-defined, internationally agreed standards for naming and calculating uncertainty information in CDRs is highly desirable wherever possible and will clarify interaction with and feedback from user communities. These standards come from the field of metrology and cover most situations encountered in developing CDRs. Engagement between Earth observers and metrologists is increasing. These interactions will make progress on the aspects of EO that go beyond the definitions developed for laboratory-based metrology. In particular, quantifying uncertainty over large scales of space and time (the low temporal and spatial frequencies in CDRs) remains a major research challenge and involves an understanding of complex error correlation structures (effects that cause neither independent random nor fixed systematic errors). This area of research cannot be neglected because users apply climate data to the full range of space–time scales spanned by Earth observation. Significant progress needs to be made in order to provide the users of climate data records with the certainty they need regarding uncertainty.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The primary funding for the work represented in this paper was through the European Space Agency Climate Change Initiative Phases 1 and 2. Christopher J. Merchant was additionally supported by the National Centre for Earth Observation, UK Natural Environment Research Council. Further contributions were developed within the project "Fidelity and Uncertainty in Climate data records from Earth Observation", which received funding from the European Union Horizon 2020 Research and Innovation Programme under grant agreement 638822. We thank Angela Benedetti at the ECWMF for support in summarizing the MACC data assimilation experiments. Max Reuter was partly funded by the Bremen government and the University of Bremen.

Edited by: David Carlson
Reviewed by: two anonymous referees

References

Ablain, M., Cazenave, A., Larnicol, G., Balmaseda, M., Cipollini, P., Faugère, Y., Fernandes, M. J., Henry, O., Johannessen, J. A., Knudsen, P., Andersen, O., Legeais, J., Meyssignac, B., Picot, N., Roca, M., Rudenko, S., Scharffenberg, M. G., Stammer, D., Timms, G., and Benveniste, J.: Improved sea level record over the satellite altimetry era (1993–2010) from the Climate Change Initiative project, Ocean Sci., 11, 67–82, https://doi.org/10.5194/os-11-67-2015, 2015.

Barnett, T., Zwiers, F., Hegerl, G., Allen, M., Crowley, T., Gillett, N., Hasselmann, K., Jones, P., Santer, B., Schnur, R., Scott, P., Taylor, K., and Tett, S.: Detecting and Attributing External Influences on the Climate System: A Review of Recent Advances, J. Climate, 18, 1291–1314, https://doi.org/10.1175/JCLI3329.1, 2005.

Bates, J., Privette, J., Kearns, E., Glance, W., and Zhao, X.: Sustained Production of Multidecadal Climate Records: Lessons from the NOAA Climate Data Record Program, B. Am. Meteorol. Soc., 97, 1573–1581, https://doi.org/10.1175/BAMS-D-15-00015.1, 2016.

Bates, J. J. and Privette, J. L.: A maturity model for assessing the completeness of climate data records, Eos T. Am. Geophys. Un., 93, 441, https://doi.org/10.1029/2012EO440006, 2012.

Benedetti, A., Morcrette, J.-J., Boucher, O., Dethof, A., Engelen, R. J., Fisher, M., Flentjes, H., Huneeus, N., Jones, L., Kaiser, J. W., Kinne, S., Mangold A., Razinger, M., Simmons A. J., Suttie, M., and the GEMS-AER team: Aerosol analysis and forecast in the ECMWF Integrated Forecast System: Data assimilation, Technical Memoranda ECMWF 571., European Centre for Medium-range Weather Forecasting, Reading, UK, 2008.

Bojinski, S., Verstraete, M., Peterson, T., Richter, C., Simmons, A., and Zemp, M.: The Concept of Essential Climate Variables in Support of Climate Research, Applications, and Policy, B. Am. Meteorol. Soc., 95, 1431–1443, https://doi.org/10.1175/BAMS-D-13-00047.1, 2014.

Buchwitz, M., Reuter, M., Bovensmann, H., Pillai, D., Heymann, J., Schneising, O., Rozanov, V., Krings, T., Burrows, J. P., Boesch, H., Gerbig, C., Meijer, Y., and Löscher, A.: Carbon Monitoring Satellite (CarbonSat): assessment of atmospheric CO_2 and CH_4 retrieval errors by error parameterization, Atmos. Meas. Tech., 6, 3477–3500, https://doi.org/10.5194/amt-6-3477-2013, 2013.

Bulgin, C. E., Embury, O., and Merchant, C. J.: Sampling uncertainty in gridded sea surface temperature products and Advanced Very High Resolution Radiometer (AVHRR) Global Area Coverage (GAC) data, Remote Sens. Environ., 117, 287–294, https://doi.org/10.1016/j.rse.2016.02.021, 2016a.

Bulgin, C. E., Embury, O., Corlett, G., and Merchant, C. J.: Independent uncertainty estimates for coefficient based sea surface temperature retrieval from the Along-Trck Scanning Radiometer instruments, Remote Sens. Environ., 178, 213–222, https://doi.org/10.1016/j.rse.2016.02.022, 2016b.

Bureau International des Poids et Mesures: The International System of Units (SI), 8th Edn., available at: http://www.bipm.org/en/publications/si-brochure/ (last access: 21 February 2017), 2006.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteor. Soc., 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Dils, B., Buchwitz, M., Reuter, M., Schneising, O., Boesch, H., Parker, R., Guerlet, S., Aben, I., Blumenstock, T., Burrows, J. P., Butz, A., Deutscher, N. M., Frankenberg, C., Hase, F., Hasekamp, O. P., Heymann, J., De Mazière, M., Notholt, J., Sussmann, R., Warneke, T., Griffith, D., Sherlock, V., and Wunch, D.: The Greenhouse Gas Climate Change Initiative (GHG-CCI): comparative validation of GHG-CCI SCIAMACHY/ENVISAT and TANSO-FTS/GOSAT CO_2 and CH_4 retrieval algorithm products with measurements from the TCCON, Atmos. Meas. Tech., 7, 1723–1744, https://doi.org/10.5194/amt-7-1723-2014, 2014.

Embury, O. and Merchant, C. J.: A reprocessing for climate of sea surface temperature from the Along-Track Scanning Radiometers: a new retrieval scheme, Remote Sens. Environ., 116, 47–61, https://doi.org/10.1016/j.rse.2010.11.020, 2012.

Flannaghan, T. J., Fueglistaler, S., Held, I. M., Po-Chedley, S., Wyman, B., and Zhao, M.: Tropical temperature trends in Atmospheric General Circulation Model simulations and the impact of uncertainties in observed SSTs, J. Geophys. Res.-Atmos., 119, 13327–13337, https://doi.org/10.1002/2014JD022365, 2014.

Giacco, F., Thiel, C., Pugliese, L., Scarpetta, S., and Marinaro, M.: Uncertainty analysis for the classification of multispectral satellite images using SVMs and SOMs, IEEE T. Geosci Remote Sens., 48, 3769–3779, 2010.

Global Climate Observing System: Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC (2010 Update), GCOS-138 WMO-TD/No. 1523, 2010.

Gruber, A., Su, C. H., Zwieback, S., Crowd, W., Dorigo, W., and Wagner, W.: Recent advances in (soil moisture) triple collocation analysis, Int. J. Appl. Earth Obs., 45, 200–211, 2016.

Hollmann, R., Merchant, C. J., Saunders, R., Downy, C., Buchwitz, M., Cazenave, A., Chuvieco, E., Defourny, P., de Leeuw, G., Forsberg, R., Holzer-Popp, T., Paul, F., Sandven, S., Sathyendranath, S., van Roozendael, M., and Wagner, W.: The ESA Climate Change Initiative: Satellite Data Records for Essential Climate Variables, B. Am. Meteorol. Soc., 94, 1541–1552, https://doi.org/10.1175/BAMS-D-11-00254.1, 2013.

Immler, F. J., Dykema, J., Gardiner, T., Whiteman, D. N., Thorne, P. W., and Vömel, H.: Reference Quality Upper-Air Measurements: guidance for developing GRUAN data products, Atmos. Meas. Tech., 3, 1217–1231, https://doi.org/10.5194/amt-3-1217-2010, 2010.

IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, UK and New York, NY, USA, 1535 pp., 2013.

Joint Committee for Guides in Metrology: International vocabulary of metrology – Basic and general concepts and associated terms (VIM), JCGM 200:2008, available at: http://www.bipm.org/en/publications/guides/gum.html (last access: 21 February 2017), 2008a.

Joint Committee for Guides in Metrology: Evaluation of measurement data – Guide to the expression of uncertainty in measurement, JCGM 100:2008, available at: http://www.bipm.org/en/publications/guides/gum.html (last access: 21 February 2017), 2008b.

Kennedy, J. J.: A review of uncertainty in in situ measurements and data sets of sea surface temperature, Rev. Geophys., 52, 1–32, https://doi.org/10.1002/2013RG000434, 2014.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K.: The JRA-55 Reanalysis: General Specifications and Basic Characteristics, J. Meteorol. Soc. Jpn, 93, 5–48, https://doi.org/10.2151/jmsj.2015-001, 2015.

Loew, A., Bell, W., Brocca, L., Bulgin, C. E., Burdanowitz, J., Calbet, X., Donner, R. V., Ghent, D., Gruber, A., Kaminski, T., Kinzel, J., Klepp, C., Lambert, J.-C., Schaepman-Strub, G., and Schröder, M.: Validation practices for satellite based earth observation data across communities, Rev. Geophys., https://doi.org/10.1002/2017RG000562, 2017.

Loosvelt, L., Peters, J., Skriver, H., De Baets, B., and Verhoest, N. E.: Impact of reducing polarimetric SAR input on the uncertainty of crop classifications based on the random forests algorithm, IEEE T. Geosci. Remote Sens., 50, 4185–4200, 2012.

Mahlstein, I., Hegerl, G., and Solomon, S.: Emerging local warming signals in observational data, Geophys. Res. Lett., 39, L21711, https://doi.org/10.1029/2012GL053952, 2012.

Massonnet, F., Bellprat, O., Guemas, V., and Doblas-Reyes, F. J.: Using climate models to estimate the quality of global observational data sets, Science, 354, 452–455, https://doi.org/10.1126/science.aaf6369, 2016.

McColl, K. A., Vogelzang, J., Konings, A. G., Entekhabi, D., Piles, M., and Stoffelen, A.: Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target, Geophys. Res. Lett., 41, 6229–6236, 2014.

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

Merchant, C. J. and Le Borgne, P.: Retrieval of sea surface temperature from space based on modeling of infrared radiative transfer: capabilities and limitations, J. Atmos. Ocean. Tech., 21, 1734–1746, https://doi.org/10.1175/JTECH1667.1, 2004.

Mittaz, J., Woolliams, E., and Merchant, C. J.: Applying Principles of Metrology to Historical Earth Observations from Satellites, Metrologia, in preparation, 2017.

Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D.: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set, J. Geophys. Res., 117, D08101, https://doi.org/10.1029/2011JD017187, 2012.

Notz, D.: How well must climate models agree with observations?, Philos. T. R. Soc. A, 373, 20140164, https://doi.org/10.1098/rsta.2014.0164, 2015.

Rayner, N. A., Merchant, C. J., and Corlett, G. K.: Communicating uncertainties in sea surface temperature, Eos, 96, https://doi.org/10.1029/2015EO030289, 2015.

Reuter, M., Thomas, W., Mieruch, S., and Hollmann, R.: A method for estimating the sampling error applied to CM-SAF monthly mean cloud fractional cover data retrieved from MSG SEVIRI, IEEE T. Geosci. Remote Sens., 48, 2469–2481, 2010.

Reuter, M., Bösch, H., Bovensmann, H., Bril, A., Buchwitz, M., Butz, A., Burrows, J. P., O'Dell, C. W., Guerlet, S., Hasekamp, O., Heymann, J., Kikuchi, N., Oshchepkov, S., Parker, R., Pfeifer, S., Schneising, O., Yokota, T., and Yoshida, Y.: A joint effort to deliver satellite retrieved atmospheric CO_2 concentrations for surface flux inversions: the ensemble median algorithm EMMA, Atmos. Chem. Phys., 13, 1771–1780, https://doi.org/10.5194/acp-13-1771-2013, 2013.

Reuter, M., Buchwitz, M., Hilker, M., Heymann, J., Schneising, O., Pillai, D., Bovensmann, H., Burrows, J. P., Bösch, H., Parker, R., Butz, A., Hasekamp, O., O'Dell, C. W., Yoshida, Y., Gerbig, C., Nehrkorn, T., Deutscher, N. M., Warneke, T., Notholt, J., Hase, F., Kivi, R., Sussmann, R., Machida, T., Matsueda, H., and Sawa, Y.: Satellite-inferred European carbon sink larger than expected, Atmos. Chem. Phys., 14, 13739–13753, https://doi.org/10.5194/acp-14-13739-2014, 2014.

Schulz, J., John, V., Kaiser-Weiss, A., Roebeling, R., Tan, D., and Swinnen, E.: Core-Climax Climate Data Record Capacity Assessment Report, CORE-CLIMAX Technical Report, CC/EUM/REP/15/001, 253 pp., available at: http://www.eumetsat.int/website/home/Data/ClimateService/index.html (last access: 21 February 2017), 2015.

Sofieva, V. F., Tamminen, J., Kyrölä, E., Laeng, A., von Clarmann, T., Dalaudier, F., Hauchecorne, A., Bertaux, J.-L., Barrot, G., Blanot, L., Fussen, D., and Vanhellemont, F.: Validation of GOMOS ozone precision estimates in the stratosphere, Atmos. Meas. Tech., 7, 2147–2158, https://doi.org/10.5194/amt-7-2147-2014, 2014.

Stoffelen, A.: Toward the true near-surface wind speed: Error modeling and calibration using triple collocation, J. Geophys. Res.-Oceans, 103, 7755–7766, 1998.

Wimmer, W., Robinson I. S., and Donlon, C. J.: Long-term validation of AATSR SST data products using shipborne radiometry in the Bay of Biscay and English Channel, Remote Sens. Environ., 116, 17–31, https://doi.org/10.1016/j.rse.2011.03.022, 2012.

Woolliams, E., Mittaz, J., Merchant, C. J., and Dilo, A.: Harmonization and Recalibration: A FIDUCEO perspective, Global Space-based Inter-calibration System Quarterly, 10, 1–2, https://doi.org/10.7289/V5GT5K7S, 2016.

World Meteorological Organisation and Bureau Internationale de Poids et Mesures, Measurement Challenges for Global Observation Systems for Climate Change Monitoring: Traceability, Stability and Uncertainty, WMO/TD-No. 1557, Rapport BIPM-2010/08, ISBN 13 978-92-822-2239-3, available at: http://www.bipm.org/en/conference-centre/bipm-workshops/wmo-bipm_workshop/ (last access: 21 February 2017), 2010.