Stability specifications for climate data records: their meaning and application in evaluating geophysical trend uncertainty

Article

Published Version

Merchant, C. J. ORCID: https://orcid.org/0000-0003-4687-9850, Woolliams, E. R., Dorigo, W., Bulgin, C. E. ORCID: https://orcid.org/0000-0003-4368-7386, Gobron, K., Hohensinn, R., Loizeau, X. and Tynan, C. P. J. (2026) Stability specifications for climate data records: their meaning and application in evaluating geophysical trend uncertainty. Surveys in Geophysics. ISSN 1573-0956 doi: 10.1007/s10712-026-09941-y

To link to this article DOI: http://dx.doi.org/10.1007/s10712-026-09941-y

Publisher: Springer

Surveys in Geophysics
https://doi.org/10.1007/s10712-026-09941-y

Stability Specifications for Climate Data Records: Their Meaning and Application in Evaluating Geophysical Trend Uncertainty

Christopher J. Merchant[1] · Emma R. Woolliams[2] · Wouter Dorigo[3] · Claire E. Bulgin[1] · Kevin Gobron[4,5] · Roland Hohensinn[6,7] · Xavier Loizeau[2] · Connor P. J. Tynan[2]

Received: 18 September 2025 / Accepted: 4 February 2026

Abstract

When quantifying changes over time in the natural environment, the stability of the observations used should be considered. Stability conceptually refers to how accurately true geophysical changes and trends are reflected in observational data. We argue the need for a better approach to defining and quantifying stability consistently across climate data records. We propose that the appropriate stability metric is the stability uncertainty for specified spatial and temporal scales. We formally define stability uncertainty by analogy with metrological measurement uncertainty. Informally, stability uncertainty informs data analysts about the plausible magnitude of a non-geophysical contribution to trend values arising solely from the observing system. Neglecting the stability uncertainty leads to overconfident assessment of the significance of geophysical trends inferred from observations. We recommend that adopting this metric would greatly improve the clarity and practical impact of the Global Climate Observing System (GCOS) statements of requirement for stability of essential climate variable (ECV) products. Moreover, GCOS stability requirements would then become a useful resource for users of ECV products when evaluating and interpreting trends in observations, helping them avoid unjustified claims for the significance of computed trends; a synthetic illustration of such usage is provided.

Keywords Essential climate variables · Climate data record · Stability · Trend analysis · Global climate observing system · Metrology

✉ Christopher J. Merchant
c.j.merchant@reading.ac.uk

1 National Centre for Earth Observation, University of Reading, Reading, UK

2 National Physical Laboratory, Teddington, UK

3 TU Wien, Vienna, Austria

4 Université Paris Cité, Institut de Physique du Globe de Paris, CNRS, 75005 Paris, IGN, France

5 Univ Gustave Eiffel, ENSG, 77455 Marne-La-Vallée, IGN, France

6 International Space Science Institute, Bern, Switzerland

7 GFZ German Research Centre for Geosciences, Helmholtz Centre Potsdam, Potsdam, Germany

Published online: 10 March 2026

1 Introduction: Observations to Quantify Climate Change

Earth's climate is changing due to greenhouse gases from fossil fuels and other human activities. Observations are vital (Hollmann et al. 2013) to quantify and understand these changes and to model flows of energy, water, and carbon. Observations are used to study climate changes, benchmark the current climate, inform government policy across sectors, and validate climate models. Observations are crucial for assessing temporal trends and underpinning projections of future climate risks, as well as for quantifying modes of geophysical variability.

The Global Climate Observing System (GCOS) defines 'essential climate variables' (ECVs) that require observation (GCOS 2022a, 2022b; GCOS 2016). Space agencies are among the organisations responding to GCOS because satellite-based Earth observations are key to globally quantifying over 20 ECVs (Hollmann et al. 2013; Schulz et al. 2009). This paper focusses on satellite-informed ECV products, here termed "climate data records" (CDRs; Mittaz et al. 2019). CDRs typically integrate observations from multiple satellite platforms, and often incorporate in situ elements of the observing system as well.

GCOS sets quantitative requirements for CDRs to be useful for climate applications (GCOS 2022b). These requirements cover spatial resolution, temporal frequency, measurement uncertainty, stability, and timeliness. This paper critically reviews the concept of stability, offering practical implications for CDR users. This paper is complementary to recent efforts to develop methods for setting CDR requirements objectively (Roebeling et al. 2025) and to aid users in assessing the suitability of CDRs for their needs (Dee et al. 2024).

A review of stability concepts is needed because the GCOS definition of the stability requirement, as detailed in Sect. 3 below, is imprecise, leading to ambiguity in whether a CDR is compliant. Consequently, stability assessment methods vary between different communities of practice and may be inconsistent. Furthermore, where a stability statement is provided for a CDR, its precise meaning and application may not be obvious to a user. This paper clarifies the meaning of stability as a concept and illustrates how users' analyses can benefit from quantitative stability information.

Previous work (Mittaz et al. 2019; Woolliams 2025; Hunt et al. 2020, 2018; Giering et al. 2019; Merchant et al. 2017, 2019; Woolliams et al. 2018; Woolliams et al. 2016) (much of it within the Horizon project FIDUCEO) exploited and extended concepts from metrology (the science of standards and measurement) to support the understanding, description, and structured quantification of uncertainty in CDRs. This paper extends this approach to stability, treating stability as a specific aspect of timeseries uncertainty.

The paper is structured as follows: Sect. 2 provides further context on CDRs and their climate change applications and discusses the nature of the observation errors in CDRs. Section 3 constructively critiques the GCOS approach to stability requirements, including its imprecise terminology. Section 4 proposes clearer stability vocabulary, informed by metrology. Section 5 uses a synthetic example to illustrate why users need stability information from data producers for quantifying changes and trends, and how GCOS requirements can still guide users when such information is absent. Section 6 concludes with a summary of implications and recommendations.

2 Context: Climate Data Records

The needs for climate data cover a wide range of time and space scales, from sub-hourly point samples to decadal global averages. CDRs are used to analyse temporal changes (including parameters describing trends) and spatial differences (gradients), as well as many other applications (Dee et al. 2024). Requirements may differ between applications: for example, stability requirements are usually most stringent for quantifying climatic trends on large scales.

Producers of CDRs aim to provide globally consistent, multidecadal data for ECVs (Popp and Mittaz 2022). However, CDRs typically must be made by combining the measured values of many sensors over more than a decade, which may introduce inconsistency through discrepancy of calibrations and because observing systems have evolved. Satellite constellations have generally improved in temporal density, spatial resolution, and/or spectral information content. Less obviously, changes in the mean local time-of-day of measurements are significant for CDRs of quantities with a diurnal cycle, as these measurement times have sometimes varied over decades (Embury et al. 2024).

Observation errors arise from various effects: errors of calibration or noise in sensor-measured values; retrieval errors; biased sampling within the time-location domain attributed to an observation; and misattribution of time or location to an observation. Analyses and interpretations of CDRs should therefore account properly for observation uncertainty that arises from such effects. The term 'observation uncertainty' requires clarification. CDRs are typically highly processed products derived from numerous (sometimes trillions of) measured values. In the case of satellite-based datasets, the sensor-measured values (e.g., radiance values) are used to infer ('retrieve') values of the ECV target quantity. Here, we use 'observations' to refer to retrieved values, space-and-time averages of retrieved values, and gap-filled estimates derived from retrieved values: these are the types of data commonly found in a CDR. Observation error is the difference of an observation from the true value (or true average value) of the quantity for the time and location (or time-location domain) attributed to the observation. The true value is generally unknown, except for rare instances where an independent, traceable reference measured value is available – which at best is for a limited number of locations and times. The observation standard uncertainty (hereafter simply 'uncertainty') is the standard deviation of the observation error distribution. Uncertainty informs the user about the dispersion of values it is reasonable to attribute to the ECV quantity given the observation. In other words, given an observed value, an associated uncertainty informs us how close to that observed value the true value likely was. This clarification of 'observation uncertainty' is intended to apply equally for individual observations and when averaging or gap-filling has been used to create a CDR, as is often done to make the dataset convenient to use.

Even in the absence of a reference, techniques are available to estimate uncertainty in CDR observations (Mittaz et al. 2019; Gruber et al. 2016). Inclusion of an uncertainty estimate per observation is good practice if the uncertainty varies significantly within a dataset (Merchant et al. 2017).

While some effects are independent between observations, many are not, so correlation of the errors between observations is usual in a CDR. Typically, correlated errors contribute a greater proportion of the uncertainty in the observations of CDRs whose production employed more averaging or gap-filling (Mittaz et al. 2019). Those components of observation error that correlate strongly on long timescales and/or large space scales might be referred to as 'bias'. There is no clean separation between 'noise' and 'bias' in a CDR;

rather, there are many effects spanning a spectrum of spatiotemporal scales of error correlation. Nevertheless, use of 'bias' to denote 'the mean error across large scales' is useful, even though the meaning of 'large' may be imprecise and application specific.

Users of CDRs are often interested in using a CDR to characterise long-term changes with time—i.e., trends. If bias (in the sense above, and after all applied corrections for bias) varies with time, this introduces an error into the evaluation of the trend. GCOS intend their stability requirements to constrain the magnitude of errors in trends evaluated from CDRs. The next section assesses the fitness for purpose of the formulation of GCOS stability requirements.

3 Critique of 'Stability' Requirements.

3.1 'Stability' in GCOS Documents

In GCOS (2022a), GCOS define the concept of stability as the property of 'minimal instrumental error and drift' in the observing systems that underpin CDRs. Stability is needed because climate monitoring requires the measurement of changes that are 'gradual and usually smaller than annual variability'(ibid.). 'Drift' (which is left undefined) typically refers to slow changes in sensor or observing system calibration from effects such as detector ageing.

In GCOS (2022b), GCOS define the quantity of stability as 'the change in bias over time … quoted per decade'. While not defined in the documentation, this 'change in bias' is to be interpreted as 'the maximum permissible cumulative effect of systematic changes' [Eggleston, S. (GCOS Secretariat), 2020, https://ceos.org/wp-content/uploads/2020/10/4WGClim12_Eggleston.pdf]. Implicitly, the 'systematic changes' and 'change in bias' originate in the observing system and affect the values in the CDR but do not reflect true change.

These definitions indicate GCOS's intent, but their imprecision hinders the development of consistent methodologies for quantifying CDR stability and consistently assessing compliance; the rest of this section explains this view. A degree of criticism is necessary to expose and resolve the current documentation's ambiguities and contradictions. Criticisms are offered constructively, setting the stage for proposed clarifications in the Sect. 4. Five criticisms are presented in the following subsections, mainly relating to the quantitative definition in GCOS (2022b).

3.2 Criticism 1: Literal Meaning Differs from Intent

The first criticism is that GCOS's literal wording for 'stability' ('change in bias over time') contradicts its intended meaning. GCOS (GCOS 2022b) specifies that a CDR is required to have a stability of *X* units per decade. The principle of substitutability (ISO 2022) is that a term can be substituted with its definition without a resulting change of meaning. Substituting this GCOS definition into a stability requirement implies that a CDR 'is required to have a change in bias over time of *X* units per decade'. Clearly, this not the intent since the ideal change in bias over time is zero.

The literal GCOS wording confuses a deterministic 'change in bias over time' with the probabilistic 'uncertainty in change in bias over time'. This uncertainty arises because CDR producers can correct for known observing system drifts only imperfectly, leaving residual biases that vary over time. The intent of the GCOS requirement is to constrain the likely magnitude of unknown changes in residual bias over time.

3.3 Criticism 2: 'Stability' is the Wrong Name for the Numerical Requirement

Stability is qualitatively understood as a CDR's property of reliably representing true changes over time, which is the usage in GCOS (2022a). However, using "stability" also as the term for the quantitative property, as in GCOS (2022b), is problematic, for greater stability (concept) corresponds to a smaller quantified 'stability'. Consider the sort of statement this situation leads to: 'The new version of the CDR has greater stability (2 units per decade) than the previous version (3 units per decade)'. Clearly, the numbers here quantify something that is not exactly equivalent to the concept being discussed.

The solution is to separate the qualitative concept of 'stability' from the quantitative definition of 'stability uncertainty' as a specific component of the observation uncertainty. There is a parallel with the metrological distinction between the qualitative term 'accuracy' and its quantification, 'uncertainty' (Woolliams 2025; JCGM, 2008), as greater accuracy means smaller uncertainty. The proposals in Sect. 4 analogously resolve this issue for stability.

3.4 Criticism 3: Interpretation of the Numerical Requirement: Coverage Factor

The GCOS definition of the numerical requirement for uncertainty states clearly that the requirement refers to '2 standard deviations' of the estimated error distribution (a coverage factor of two). A similar level of clarity is needed to describe the coverage factor, for any numerical requirements targeting the uncertainty in change in bias over time.

Interpreting the 'stability' as 'the maximum permissible cumulative effect of systematic changes' [ibid. Eggleston, 2020] seems intuitive but does not account for the probabilistic nature of uncertainty. The change in bias over time after known effects have been corrected is unknown, and therefore, it cannot be claimed with certainty to be smaller than a maximum permissible magnitude.

Instead, CDR producers may be able to estimate the plausible distribution of residual changes in bias over time, or to assess the probability that the 'stability' is within the requirement. For a Gaussian distribution, a coverage factor of two implies acceptance of a ~ 5% chance that the magnitude of residual change of bias over time exceeds the numerical requirement. This low probability seems quite compatible with the intent of the 'maximum permissible' formulation quoted earlier. It is therefore plausible to interpret GCOS numerical requirements on stability as quantifying, with a coverage factor of two (or coverage probability of 95%), the uncertainty in change in bias over time. This needs definitive formal clarification.

3.5 Criticism 4: Interpretation of the Numerical Requirement: Applicable Spatial Scale

CDRs are complex datasets, unlikely to be spatially uniform in either bias or change of bias over time. This means that an evaluation of 'stability' will vary with the domain of evaluation. Five ECVs (sea surface temperature, sea surface salinity, regional mean sea level, ground water storage change, and relative humidity in the boundary layer) have a quantitative statement of the spatial scale of relevance for assessing the GCOS stability requirement (GCOS 2022b). This is good practice and should be built into the formulation of the requirement by default.

For example, consider a CDR for which a key application is the evaluation of trends across geographical areas typically ~ 1000 km in extent. The global ‘stability’ of such a CDR may well be better than the ‘stability’ for localised areas, so it would be relevant to state that the numerical requirement for ‘stability’ is applicable to ~ 1000 km scales.

Sometimes, a user may need to understand stability uncertainty on a range of scales. This requires error covariance information across observations separated in time and space: a more detailed set of information than can be represented in a statement of stability uncertainty. Practical methods for parameterising such error covariance information have been proposed (Mittaz et al. 2019; Merchant et al. 2019). Work to make these ideas operational for CDRs still needs to be done, but further discussion goes beyond the scope of this paper.

3.6 Criticism 5: Interpretation of the Numerical Requirement: Applicable Temporal Scale

The unit for numerical requirements on ‘stability’ is ‘units per decade’ in GCOS (2022b). What is the intent here? Some infer that the requirement applies to every 10-year interval within a record, others that the requirement should be assessed over a record’s full duration but expressed per decade, and others read this merely as a standardised presentation offering no methodological implications. GCOS’ intent here needs to be clarified, as evaluations against the requirements will differ based on the temporal scale used for assessment.

To illustrate this, Fig. 1 shows the stability uncertainty of a hypothetical CDR plotted against the interval over which the stability uncertainty is evaluated. For illustration, it is assumed that the observation errors are driven by an error process that operates constantly throughout the record (whereas reality will be more complicated). Four different error processes are depicted (see caption), which by design yield the same stability uncertainty when evaluated over 10 years. Over other evaluation intervals, the stability uncertainty differs from the 10-year “stability”, to a degree that depends on the correlation properties of the observation errors under different processes.

In summary, this section has identified some inconsistency and imprecision in the terminology employed by GCOS, and additional information needed to define more fully the meaning and evaluation of numerical requirements. Section 4 proposes a set of terminology intended to bring the clarity needed to make statements about stability practically useful to users of CDRs seeking to evaluate climatic trends.

4 A Systematic Terminology for Stability Requirements

In Table 1, we propose a systematic terminology for discussing stability requirements, by analogy with the metrological terminology for expressing uncertainty in measurement (Woolliams 2025; JCGM 2008). In Table 1, we use ‘trend’ to indicate a quantity characterising change of an ECV quantity with time.

The proposed definitions for stability, stability error, and stability (standard) uncertainty mirror the established metrological concepts of accuracy, measurement error, and

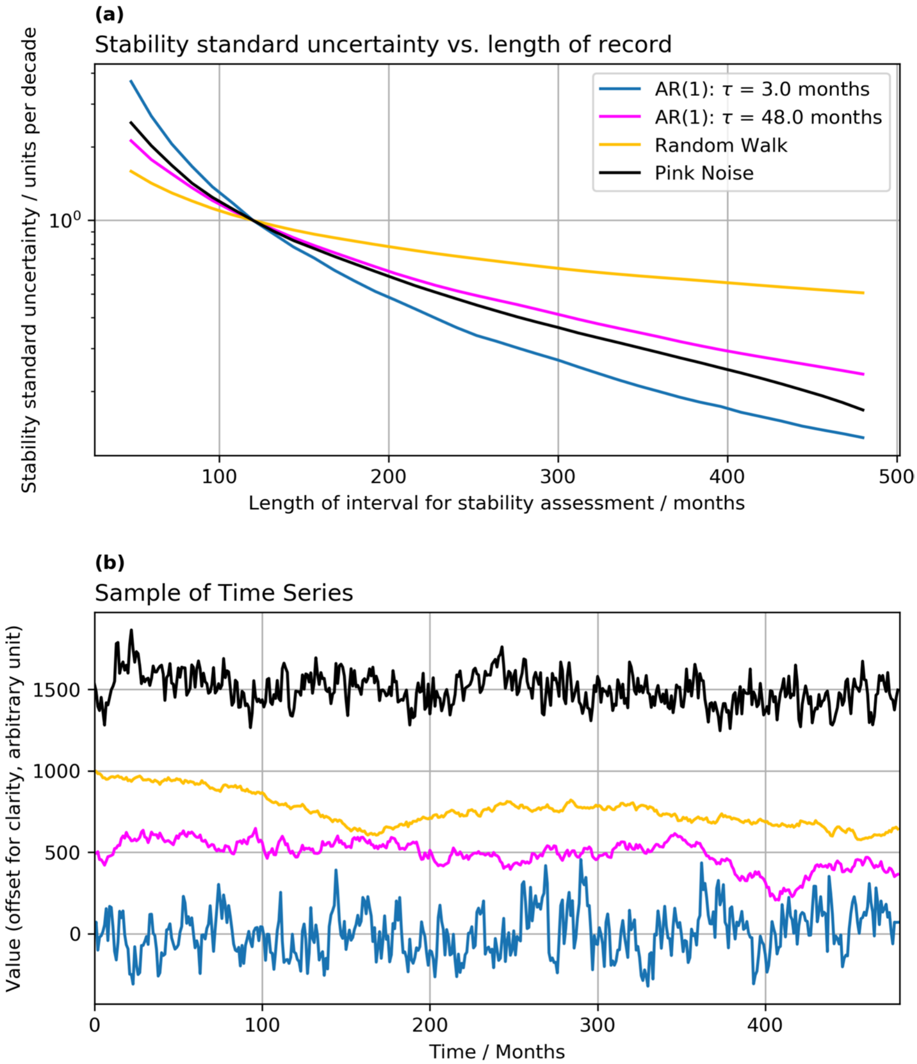


Fig. 1 **a** The stability uncertainty from observation errors depends on the time range over which the stability uncertainty is quantified. The uncertainty in computed trends arising from observation errors tends to decrease when the trend is computed over a longer interval, irrespective of the nature of the processes causing the errors. This is illustrated for four stationary error processes: an auto-regressive process (corresponding to errors that correlate over time with an exponential decay of time constant $\tau = 3$ months); an auto-regressive process with longer memory ($\tau = 48$ months); a random walk process; and pink noise. All processes have been normalised in magnitude to give a stability standard uncertainty ('k = 1') of 1 unit per decade when the stability is evaluated over 10 years. Real CDRs will have observation errors that may be characterised by a mixture of these and other types of error process that are not necessarily statistically stationary over time. **b** Randomly selected example timeseries corresponding to each of the four error processes. Each series is normalised to cause the same stability standard uncertainty over 10 years, as in panel (a) with the same colour coding. Processes with less memory tend to be 'noisier' between timesteps for a given level of decadal stability standard uncertainty. The series are shown offset in the vertical for legibility by setting the first value to 0 (AR τ=3), 500 (AR τ=48), 1000 (Random Walk) and 1500 (Pink Noise)

Table 1 Terminology for stability requirements, alongside the analogous terminology for accuracy of measurement

Terminology for accuracy (JCGM 2008)		Terminology for stability	
measurement accuracy[1]	closeness of agreement between a measured quantity value and a true quantity value of a measurand	(observational) stability[2]	closeness of agreement between an observed[3] value of change with time[4] and a true value of change with time[5]
error	difference of a measured value from a true (or reference) value of a measurand	stability error	difference of an observed trend value[6] from a true (or reference) trend value of an essential climate variable quantity
measurement uncertainty	parameter characterising the dispersion of the values being attributed to a measurand, based on the information used	stability uncertainty	parameter characterising the dispersion-arising-from-observation-errors[7] of the trend values being attributed to an essential climate variable quantity
measurement standard uncertainty	measurement uncertainty expressed as a standard deviation[8]	stability standard uncertainty	stability uncertainty expressed as a standard deviation

1. The concept 'measurement accuracy' is not a quantity and so is not given a numerical value. A measurement is said to be more accurate when it offers a smaller magnitude measurement error

2. The concept 'stability' is not a quantity and is not given a numerical quantity value. A CDR is said to be more stable when it offers a smaller magnitude error in the evaluation of a change with time. A CDR may be described as having good stability for a specific application when its stability uncertainty is sufficient to meet the requirements of that application

3. We refer to the quantity values in a CDR as observed values rather than measured values. 'Observed values' is intended to encompass all types of data in CDRs from measured values recorded by a satellite sensor to highly processed data. Observation errors arise both from effects in the observing system and in processing of data to the CDR

4. By referring to a 'change with time', the definition here does not specify a timescale, in order to be generally applicable. However, in a particular context, an explicit timescale of interest should be specified. Stability in the context of GCOS and CDRs might refer to 'decadal stability', for example. This applies to all the definitions

5. The true value of a climate trend is rarely known. A reference observing system can provide a practical true value if errors in reference observed values are negligibly small in magnitude

6. A trend value is the value of a parameter characterising the change over time of a quantity. Most commonly in reference to CDRs the parameter is an average rate of long-term change, such as the slope of a fitted linear representation of the change over a decade or more. However, other measures of change over time are not excluded from this terminology

7. The stability uncertainty is identified here as the component of uncertainty in an evaluated trend that arises from effects in the observing system and CDR production. Unfitted geophysical variability additionally contributes to uncertainty in trend values. The stability uncertainty is therefore not the only contributor to uncertainty in trend values (Gobron et al. 2026)

8. Expanded uncertainty values are expressed as a multiple of the standard uncertainty, where the multiple is described by the coverage factor, k. The expanded uncertainty is associated with an expanded 'coverage probability'. For a Gaussian distribution, a coverage factor of $k = 2$ represents a coverage probability of approximately 95%. For non-Gaussian distributions, other coverage factors are required for the same coverage probability, and, where distributions are asymmetric, a confidence interval with upper and lower values can better describe the expanded uncertainty

measurement (standard) uncertainty, allowing for precise distinctions between these related concepts.

When a trend is derived from a CDR, the resulting trend value is our best estimate of the true trend. Because no CDR is perfect, there is a spread of plausible trend values around that best estimate associated with the unknown observation errors in the CDR. This dispersion equals the stability uncertainty of the CDR. It is typically quantified as a standard deviation (a coverage factor of $k = 1$), though other coverage factors could be chosen.

It is important to distinguish between stability uncertainty and statistical fitting uncertainty (Gobron et al. 2026). The former relates solely to the potential impact of observation errors on the estimated trend, while the latter arises from the limited length of a time-series and the presence of unfitted geophysical variability. ‘Unfitted geophysical variability’ here means ‘geophysical changes with time that are not represented by a functional time-dependence in the statistical model’. Gobron et al. (2026) formulate mathematically the propagation of the observational error covariance and unfitted geophysical variance to the uncertainty in a trend value.

The term ‘trend’ is used broadly in these definitions to refer to any representation of long-term change in an ECV, including straight-line or higher order polynomial fits. While the concept of stability uncertainty applies to any such trend parameter, it is most commonly associated with the linear trend component. GCOS requirements—given their units—are intended specifically to constrain the stability uncertainty of the linear trend component. This targeted approach is both practical and effective as it avoids undue complexity.

Based on this set of definitions, the GCOS stability requirements might be interpreted as specifying a quantitative limit on ‘decadal-scale stability uncertainty with coverage factor $k = 2$’, implicitly corresponding to a coverage probability of approximately 95%. The recommendations in Sect. 6 note that a definitive statement of the intended meaning of stability requirements is needed.

5 Practical implications for Analysing Trends in CDRs

An analyst using a CDR to evaluate a geophysical trend must consider the stability uncertainty of the observations to accurately interpret parameter uncertainties from their fit. If the CDR producers do not provide a clear statement on dataset stability (or, even better, an error covariance matrix), the analyst should use the GCOS stability requirement for that ECV as a guide. If the trend uncertainty from the fit is smaller than or comparable to this guide stability uncertainty, the true uncertainty is likely much larger than the fit suggests, as the effect of observational uncertainty cannot be ignored. This should, in turn, inform the analyst’s statistical and geophysical interpretations of the trend value.

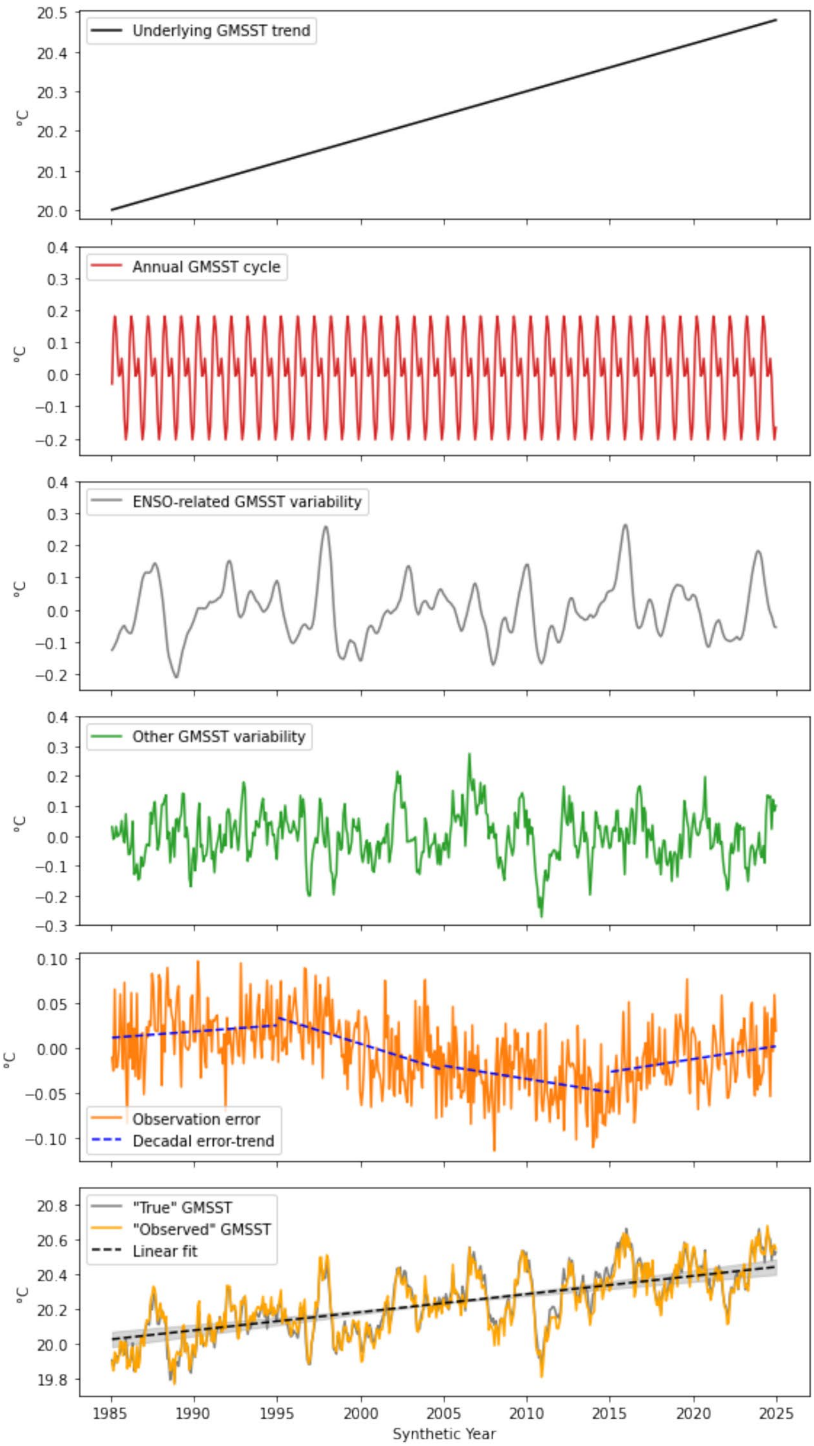
These practical implications are illustrated below with a synthetic example, based on characteristics of quasi-global (60°S to 60°N) monthly mean sea surface temperature (GMSST). GMSST over the past 40 years has followed a quadratic multidecadal trend, with the El Niño Southern Oscillation (ENSO) driving a notable share of the monthly to multi-annual variability (Merchant et al. 2025), along with volcanic activity and the solar irradiance cycle. Here, a GMSST-like timeseries is made as the sum of a linear trend, an annual cycle, a scaled ENSO index, and autocorrelated random variability with a seasonal (3-month) timescale. Using synthetic data allows us to know the true underlying trend and create multiple realisations of seasonal and observational variability to demonstrate their

Fig. 2 A synthetic global 'sea surface temperature' dataset with a known linear trend (top, 'true' slope = 0.12 K decade^{-1}), annual cycle (second panel), ENSO-related variability (third panel), and other geophysical variability (fourth panel). The synthetic 'true' series (bottom) is a combination of these, whereas the 'observed' series additionally includes observation errors (fifth panel); the deseasonalised timeseries are shown here for clarity. The annual cycle and ENSO index used are from SST CCI (Embury et al. 2024), and the GMSST response is one-tenth of the ENSO index (neglecting the real-world lag in response of around 3 months). The characteristics of the 'other geophysical variability' are based on the properties of the residuals to the multi-factor linear fit of GMSST in Merchant et al. (2025), and are synthesised here as an autocorrelated random (AR 1) process with time constant of 3 months and standard deviation of 0.08 K. The 'observation errors' are a combination of: an AR 1 process with exponential time constant of 48 months and standard deviation over the series of 0.215 K; and independent noise with standard deviation 0.0337 K. The combined series of observation errors gives a standard uncertainty of 0.04 K (root sum of squares of the standard deviations from the two error components) and stability standard uncertainty of 0.035 K/decade (evaluated over one decade). These values are purely illustrative but were selected to be compatible with GCOS requirements for SST. Both the other geophysical variability and the observation errors are generated with 1000 realisations, of which the first examples are shown here. The quoted stability standard uncertainty equals the standard deviation across the 1000 observational error series of the linear trend over the first decade. An impression of the decadal trend errors associated with this value of stability standard uncertainty is given by the decadal linear fits to the observation errors (dashed lines in the fifth panel). In the bottom panel, the dashed line is the fit (slope = 0.104 K decade^{-1}) to the 'observed' series shown. The shaded region indicates the true trend uncertainty (with $k = 2$) arising from the combination of stability uncertainty and unfitted geophysical variability. The shading spans the set of straight lines with slopes within ± 0.022 K decade^{-1} of the best-fit slope. The observation error series shown in the fifth panel five is a single illustrative realisation; the stability uncertainty shown in the bottom panel is derived from the ensemble spread of trend estimates across many such realisations (and cannot be inferred from the illustrated realisation alone). For the code used to generate this figure, see (Merchant 2026)

effect on fitting uncertainty. Observation errors are added, also randomly generated in multiple realisations to mimic noise and imperfectly corrected sensor biases. The observation error series is created with an autocorrelation timescale of 4 years, a typical interval between major changes in the satellite observing system for SST (Embury et al. 2024). The synthetic timeseries are illustrated in Fig. 2, with more details given in the figure caption.

Each realisation of the synthetic SST data is fitted, and the distribution of the fitted slopes across these realisations represents the true trend uncertainty. This can then be compared to the uncertainty estimates provided by three different fitting methods: ordinary least squares (OLS) fitting of an offset, a linear trend slope, and magnitude of the annual cycle, with the standard equations for estimation of parameter uncertainty assuming independent errors; OLS as before, but explicitly fitting the impact of the ENSO cycle using an ENSO index as an additional function in the linear fit; and OLS using the ENSO index and additionally using an improved method for parameter uncertainty. The improved method in the latter case is implementation of the heteroscedasticity and autocorrelation consistent (HAC) covariance estimator (Newey and West 1987), available in Python as an option to the function statsmodels.regression.linear_model. The uncertainty estimates of parameters in all cases are purely data driven and cannot account for any component of the observational errors series that happens not to be orthogonal to the fit functions (annual cycle, linear increase and ENSO variability); all results therefore underestimate the true trend uncertainty, as shown below.

The observation errors were synthesised to have stability standard uncertainty of 0.035 K decade^{-1} evaluated over 10 years. Over the 40-year period, this translates to a smaller stability standard uncertainty of 0.007 K decade^{-1}. This factor of 5 decrease is similar to the decrease as shown in Fig. 1 for pink noise. The first thing to note from Table 2 is that the stability standard uncertainty at 0.007 K decade^{-1} contributes a major part of the true trend uncertainty of 0.011 K decade^{-1}. The difference between



Underlying GMSST trend
Annual GMSST cycle
ENSO-related GMSST variability
Other GMSST variability
Observation error
Decadal error-trend
"True" GMSST
"Observed" GMSST
Linear fit
°C
Synthetic Year
1985
1990
1995
2000
2005
2010
2015
2020
2025

Table 2 Examples of fitting uncertainty using different least-square fitting approaches

Assumptions of fit	Mean of errors in fitted trend (trend bias) / K decade^{-1}	Standard deviation of errors in fitted trend (true uncertainty) / K decade^{-1}	Parameter estimate of trend uncertainty from fit / K decade^{-1}
OLS, fitting an offset, linear trend and annual cycle	0.005	0.011	0.005
OLS as above, additionally fitting ENSO cycle	0.000	0.011	0.003
OLS fitting ENSO and using a HAC covariance estimator	0.000	0.011	0.007

0.007 K decade^{-1} arising from observation errors and 0.011 K decade^{-1} is the statistical effect of the unfitted 'other GMSST variability' on fitted trends.

When OLS is applied without accounting for the ENSO cycle, the fitted trend has a mean error (bias) of 0.005 K decade^{-1}. This bias occurs because the ENSO index is more positive in later decades (see Fig. 2, third panel). The uncertainty in the linear parameter from the standard OLS equations is 0.005 K decade^{-1}, only 45% of the true value. This happens because the fit cannot distinguish between the geophysical trend and the observational drift.

Adding the ENSO cycle to the OLS fit largely eliminates the bias in the trend estimate but has only a minor impact on the true trend uncertainty, as it is dominated by observation errors and other GMSST variability. However, this also causes the fitted parameter uncertainty to become even more underestimated (27% of the true value) because fitting the ENSO cycle reduces the residuals from which the uncertainty is estimated. Despite this, fitting the ENSO cycle is the better choice as it reduces the bias in the underlying trend estimate.

Plotting the residuals from the OLS fit reveals autocorrelation (not shown). This is due to both the added GMSST variability and the observation errors having some persistence. This means that the assumption of errors independently drawn from one distribution that underlies the standard equations for parameter uncertainty is violated. The implementation of OLS with HAC covariance allows for autocorrelation and heteroscedasticity. Using OLS HAC does not change the fits but does increase the estimate of the uncertainty in the linear parameter to 0.007 K decade^{-1}, which is 64% of its true value. While improved, this remains an underestimate because the trend in the observation errors cannot be isolated from the trend signal from investigation of the residuals alone.

Ideally, a timeseries analyst would have access to information about the stability standard uncertainty of the observations. Compared to a true trend uncertainty of 0.011 K decade^{-1} evaluated over 40 years, the trend uncertainty estimates from the fits are clear underestimates. The true trend standard uncertainty cannot be less than the stability standard uncertainty. Such information could also inform detailed statistical modelling for trend uncertainty assessment as presented in Gobron et al. (2026), although it is clear from Fig. 1 that statements of stability uncertainty alone do not fully specify the temporal error covariance needed.

If a statement of stability uncertainty is not available for a CDR, an analyst can check the plausibility of obtained trend uncertainty against the GCOS stability requirements. For SST, 'threshold' and 'goal' requirements for 'stability'" are stated to be 0.1 K decade^{-1} and 0.01 K decade^{-1}, respectively. The stability uncertainty attained by the best CDRs is likely to be somewhere between the threshold and goal values. Notwithstanding the ambiguities in the present GCOS definitions, reference to the requirements would alert an analyst that it would not be justified to quote 0.003 K decade^{-1} (as obtained from the second OLS fit) as the uncertainty in a linear trend, because the stability of the observations is not sufficient to warrant such precision.

6 Conclusion and Recommendations.

The concept of stability needs to be paired with a corresponding clearly defined quantity in order to set requirements intelligibly, to drive self-assessment and innovation on the part of observation scientists, and to inform users of climate data records about the observation-driven uncertainty in calculated trends.

Known effects (biases) are estimated and/or corrected by producers of climate data records, which can be done only imperfectly, that is, with some remaining uncertainty. Knowledge of the trend artefacts caused by the observational errors in the data is therefore probabilistic. The appropriate metric is therefore probabilistic, namely, stability uncertainty, which is a parameter characterising the uncertainty of the trend values attributed to a climate variable that arises from observation errors in a CDR. Stability standard uncertainty is this metric expressed as one standard deviation. To an analyst of data, the stability standard uncertainty provides a minimum value of standard uncertainty attributable to a linear trend parameter derived from the data. Data producers should therefore endeavour to provide evaluations of the stability standard uncertainty of their timeseries to inform users of their data.

GCOS set numerical requirements for ECV products to which many international initiatives respond. To maximise the value and impact of requirements setting, we recommend that the meaning of those numerical requirements is clarified. Only then can data producers coherently and quantifiably respond to them.

Specifically, we recommend the following:

(1) Define the numerical quantity that describes stability as a form of uncertainty – because that is what it is. For example, adopt the formal structure of the proposed definitions in Table 1, particularly ‘stability uncertainty’.
(2) Clarify the temporal scale over which the value of stability uncertainty is to be quantified, not merely the units of expression.
(3) Clarify for each requirement the relevant spatial scale(s) for quantification of the stability uncertainty.
(4) Clarify the coverage factor applying to stability uncertainty. For example, adopt stability standard uncertainty (coverage factor of $k = 1$), or state what expanded uncertainty (coverage factor or coverage probability) is intended.
(5) Foster efforts by data producers to evaluate the stability uncertainty and to communicate it with their products.
(6) Foster understanding by data users that the stability uncertainty can usefully inform their interpretation of trends in datasets, avoiding overconfident interpretations.

Funding National Centre for Earth Observation, NE/RO16518/1, Christopher John Merchant

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Dee D, Obregon A, Buontempo C (2024) Are our climate data fit for your purpose? Bull Am Meteorol Soc 105:E1723–E1733. https://doi.org/10.1175/bams-d-23-0295.1

Embury O, Merchant CJ, Good SA, Rayner NA, Hoyer JL, Atkinson C, Block T, Alerskans E, Pearson KJ, Worsfold M et al (2024) Satellite-based time-series of sea-surface temperature since 1980 for climate applications. Sci Data 11:326. https://doi.org/10.1038/s41597-024-03147-w

Giering R, Quast R, Mittaz JPD, Hunt SE, Harris PM, Woolliams ER, Merchant CJ (2019) A novel framework to harmonise satellite data series for climate applications. Remote Sens. https://doi.org/10.3390/rs11091002

Global Climate Observing System (GCOS) (2016) The Global Observing System for Climate: Implementation Needs: GCOS-200; World Meteorological Organisation, Geneva, Switzerland

Global Climate Observing System (GCOS) (2022a) The 2022 GCOS Implementation Plan: GCOS-244; World Meteorological Organisation, Geneva, Switzerland

Global Climate Observing System (GCOS) (2022b) The 2022 GCOS ECVs Requirements: GCOS-245; World Meteorological Organisation, Geneva, Switzerland

Gobron K, Hohensinn R, Loizeau X, Bulgin CE, Merchant CJ, Woolliams ER, Cox MG, Dorigo W, Howard T, Langsdale M et al (2026) A unified framework for trend uncertainty assessment in climate data record: application to global mean sea level. Surv Geophys 5:52. https://doi.org/10.1007/s10712-025-09922-7

Gruber A, Su CH, Zwieback S, Crowd W, Dorigo W, Wagner W (2016) Recent advances in (soil moisture) triple collocation analysis. Int J Appl Earth Obs Geoinf 45:200–211. https://doi.org/10.1016/j.jag.2015.09.002

Hollmann R, Merchant CJ, Saunders R, Downy C, Buchwitz M, Cazenave A, Chuvieco E, Defourny P, de Leeuw G, Forsberg R et al (2013) The ESA climate change initiative: satellite data records for essential climate variables. Bull Am Meteorol Soc 94:1541–1552. https://doi.org/10.1175/BAMS-D-11-00254.1

Hunt SE, Mittaz JPD, Smith D, Polehampton E, Yemelyanova R, Woolliams ER, Donlon C (2020) Comparison of the Sentinel-3A and B SLSTR tandem phase data using metrological principles. Remote Sens. https://doi.org/10.3390/rs12182893

Hunt, S.E.; Quast, R.; Harris, P.M.; Mittaz, J.P.D.; Woolliams, E.R.; Giering, R.; Dilo, A.; Merchant, C.J.; Ieee. A Metrological Approach To Producing Harmonised Fundamental Climate Data Records From Long-Term Sensor Series DATA. In *Igarss 2018 - 2018 Ieee International Geoscience and Remote Sensing Symposium*, 2018; pp. 3397–3400.

ISO. Terminology work – Principles and methods. In *Geneva, Switzerland*, International Organization for Standardization: 2022.

Joint Committee for Guides in Metrology (JGCM) (2008) Guide to the expression of Uncertainty in Measurement

Merchant CJ (2026) Code Notebook for Paper. Stability specifications for climate data records. https://doi.org/10.6084/m9.figshare.31126723

Merchant CJ, Paul F, Popp T, Ablain M, Bontemps S, Defourny P, Hollmann R, Lavergne T, Laeng A, de Leeuw G et al (2017) Uncertainty information in climate data records from Earth observation. Earth Syst Sci Data 9:511–527. https://doi.org/10.5194/essd-9-511-2017

Merchant CJ, Holl G, Mittaz JPD, Woolliams ER (2019) Radiance uncertainty characterisation to facilitate climate data record creation. Remote Sens. https://doi.org/10.3390/rs11050474

Merchant CJ, Allan RP, Embury O (2025) Quantifying the acceleration of multidecadal global sea surface warming driven by Earth's energy imbalance. Environ Res Lett. https://doi.org/10.1088/1748-9326/adaa8a

Mittaz J, Merchant CJ, Woolliams ER (2019) Applying principles of metrology to historical Earth observations from satellites. Metrologia. https://doi.org/10.1088/1681-7575/ab1705

Newey WK, West KD (1987) A simple, postivie semidefinite, heteorskedasticity and autocorrelation consistent covariance-matrix. Econometrica 55:703–708. https://doi.org/10.2307/1913610

Popp T, Mittaz J (2022) Systematic propagation of AVHRR AOD uncertainties-a case study to demonstrate the FIDUCEO approach. Remote Sens. https://doi.org/10.3390/rs14040875

Roebeling RA, Bojinski S, Poli P, John VO, Schulz J (2025) On the determination of GCOS ECV product requirements for climate applications. Bull Am Meteorol Soc 106:E868–E893. https://doi.org/10.1175/bams-d-24-0123.1

Schulz J, Albert P, Behr HD, Caprion D, Deneke H, Dewitte S, Durr B, Fuchs P, Gratzki A, Hechler P et al (2009) Operational climate monitoring from space: the EUMETSAT satellite application facility on climate monitoring (CM-SAF). Atmos Chem Phys 9:1687–1709. https://doi.org/10.5194/acp-9-1687-2009

Woolliams ER (2025) A metrological framework for addressing uncertainty in satellite and in situ Earth environmental observations. Surv Geophys. https://doi.org/10.1007/s10712-025-09916-5

Woolliams ER, Mittaz J, Merchant CJ (2016) Harmonisation and Recalibration: a FIDUCEO perspective. GSICS Quart. https://doi.org/10.7289/V5GT5K7S

Woolliams, E.R.; Mittaz, J.P.D.; Merchant, C.J.; Hunt, S.E.; Harris, P.M. Applying Metrological Techniques to Satellite Fundamental Climate Data Records. In *13th International Conference on New Developments and Applications in Optical Radiometry*, Grobner, J., Ed. 2018; Vol. 972.