General guidance on a metrological approach to fundamental data records (FDR)

National Physical Laboratory

NPL
National Physical Laboratory

General guidance on a metrological approach to fundamental data records (FDR)

Emma Woolliams
Paul Green
Sam Hunt
Tom Gardiner
Peter Woolliams
David Medland

National Physical Laboratory

Hampton Road, Teddington, Middlesex, TW11 0LW

Version control

Issue	Date	Authors	Reviewed by	Notes
0.1	20/12/19	Emma Woolliams, Paul Green, Peter Woolliams, Tom Gardiner		Initial draft for internal comment
0.2	09/04/20	Emma Woolliams, Paul Green, Sam Hunt, Peter Woolliams, Tom Gardiner, David Medland	Paul Green	First full draft to consortium for comment
0.3	22/09/20	Paul Green	Paul Green	Incorporated ESA comments. Add project links to §1, Added footnote explanation to §3.2 eqn 3-4, Reorder §4, Consistent eqn numbering style, clarification in §7, correct a few typos
1.0	31/03/24	Paul Green	Günter Lichtenberg Gabriele Brizzi	Full release

Table of contents

Executive summary

This document provides guidance for the project consortium for the fdr4xxx series of projects on a metrological approach to the development of fundamental data records. The main document provides an overview to the approach that is sensor & project-agnostic.

Appendices provides information specific to the techniques and sensors targeted by a specific project

1 Introduction and motivation

In October 1959, the first Earth radiation budget sensor was launched on the Explorer 7 satellite. This initiated Earth Observation (EO) satellite programmes, which today are operated by a wide variety of space agencies, meteorological agencies and commercial operators and provide observations through an ever-growing variety of techniques for a wide range of social, scientific, environmental and commercial applications. Sustained EO programmes are now operated to support weather forecasting and, particularly through the Sentinel satellites that form the space arm of the Copernicus Services a holistic European programme to unify historical individual programs to provide routine, long-term observations of many aspects of the environment are now taken for the benefit of society.

Historical and current EO data provide information about environmental and climate change that is of great value to today's scientists and to decision makers in companies, in governments and in non-governmental organisations (NGOs). These data are also a legacy of immense value to future generations.

However, for this immediate and legacy value to be realised, it is important that EO data sets are interoperable and temporally stable. We need to be able to combine data from different sensors, to form multidecadal records from series of similar sensors and to understand the quality and uncertainty associated with data sets to assess their fitness for purpose.

Meeting these requirements is challenging satellite sensors, the design of the sensor, its operation and calibration modes (pre-flight and inflight) have evolved and become more sophisticated. New approaches and technology have been developed (and presumably will continue to be developed) meaning that each new generation of sensor have subtly different sensitivities & behaviours. Particularly, the earliest generations of historic sensors now used for long-term records, were never envisaged to be used in such a way and have not always been calibrated pre-flight to meet the challenging demands of long-term records. Sensors also all change (drift and degrade) in the harsh environment of space. In many cases, again particularly for the older sensors, there is also a lack of available information about the exact design and operational conditions for the sensors. Nevertheless, there is such scientific value in these long-term records that it is well worth the effort to develop so-called "fundamental data records" that provide detailed, uncertainty-quantified and traceable information on the origin and quality of long-term records.

Metrology, the science of measurement, is the discipline responsible for maintaining the International System of Units (SI) and the associated system of measurement. It is core to the SI, that measurements are stable over very long time periods, that measurement standards are equivalent worldwide and that measurements are coherent – that is different types of measurement can be combined because, for example, an electrical watt is equivalent to an optical watt is equivalent to a mechanical watt.

These properties of metrology: long term stability, international consistency and measurement coherence are principles that are desired for EO data records. It is for this reason that over the last two decades there has been increasing research in the collaborative field of EO Metrology. Clearly, uncertainty analysis is not new to the EO community, with many of the underlying principles of NWP and satellite data inter-comparisons based on sound metrological practise. Recent efforts in EO metrology brings together metrologists working in national metrology institutes (NMIs), satellite sensor experts and Earth observation practitioners developing long term data records, to develop methods to formalise to the core metrological principles of metrological traceability, comparison and uncertainty analysis to EO data sets.

Early collaborations between these communities led to the agreement, in 2010, by CEOS in the frame of the Global Earth Observation System of Systems (GEOSS) of the Quality Assurance Framework for Earth Observation (QA4EO). QA4EO has a core principle that EO data should be accompanied by a fully traceable indicator of their quality, where the quality indicator is sufficient to allow all users to readily assess the fitness for purpose for their applications and traceability requires this quality indicator to be based on "a documented and quantifiable assessment of evidence demonstrating the level of traceability to internationally agreed (where possible SI) reference standards." QA4EO stops short of requiring robust metrological traceability, but the accompanying guidelines are based on metrological principles adapted from the international metrology community.

Since 2010, collaborative EO-metrology projects have been developing robust methods to facilitate broader use of metrological principles in EO applications. Some of the core projects are:

- QA4ECV[1], an EU-funded FP-7 project, established methodologies for documenting and auditing climate data records of essential climate variables (ECVs). This involved developing 'traceability chains' that diagrammatically represent the data sources and processing algorithms used to obtain an ECV-record, stipulating a uniform set of underlying traceability metrics and reporting the quality assurance of the data sets. The QA4ECV methodologies were further developed in the EQC project funded by the Copernicus Climate Change Service and provide a documentary framework for ECV records provided to that data store.
- GAIA-CLIM[2], an EU-funded H2020 project, adapted methods from QA4ECV to establish standardised methods for documenting, combining and presenting quantitative uncertainty metrics to allow comparative uncertainty analysis of terrestrial observations from a range of measurements techniques used to validate satellite-derived products.
- Traceability and uncertainty assessments of terrestrial reference measurement networks have also formed part of the work being carried out in the development of the Climate Data Store within the Copernicus Climate Change Service. [3]
- The MetEOC[4] series of projects, funded by the EMRP and EMPIR programmes (joint programmes between EU H2020 and the participating states and led by EURAMET, the European Association of NMIs), performed research into approaches to apply metrology into a number of specific EO disciplines.
- FIDUCEO[5], an EU-funded H2020 project, developed a framework for long term historical records from passive radiometric sensors and established a way of evaluating error correlation structures in such records as well as developing methods for harmonising long-term data series
- The ESA "fiducial reference measurement" projects[6] and initiatives such as CEOS RadCalNet[7] have also applied such methodology to validation field site in-situ observations.

In this document we build on this legacy of activity. In particular, we expand the concepts, nascent in the FIDUCEO project [ref Mittaz et al.], generalising them beyond passive radiometric band sensors,

[1] http://www.qa4ecv.eu/
[2] http://www.gaia-clim.eu/ specifically D2.6 http://www.gaia-clim.eu/sites/www.gaia-clim.eu/files/document/d2_6_final.pdf
[3] https://cds.climate.copernicus.eu/#!/home
[4] http://www.meteoc.org/
[5] http://www.fiduceo.eu/
[6] https://earth.esa.int/web/sppa/activities/frm
[7] https://www.radcalnet.org/#!/

to establish "fundamental data records" (FDRs) and "thematic data products" (TDPs). This document is being written to support two ESA-funded projects: the fundamental data records for altimetry (FDR4ALT) and the fundamental data records for atmosphere (FDR4ATMOS) projects. This "core guidance document" is provided, identically in content, to both projects. The projects will then prepare "specific guidance documents" that tailor these ideas to their applications.

2 Fundamental Data Records and Thematic Data Products

2.1 Fundamental Data Records (FDRs)

A fundamental data record (FDR) is a generalised[8] concept of a fundamental climate data record (FCDR) that has been proposed to be defined[9] as:

> *An FCDR consists of a long, stabilised record of uncertainty-quantified sensor observations that are calibrated to physical units and located in time and space, together with all ancillary and lower-level instrument data used to calibrate and locate the observations and to estimate uncertainty.*

FDRs and FCDRs are produced as an initial step in the data processing chain – at the lowest pragmatic sensor level, be that defined as a 'pixel', 'datum', 'data unit' or 'element'. They are converted into Thematic Data Products (TDPs) or Climate Data Records (CDRs) of higher-level products, a process that can combine different FDR observations and additional information, e.g. from models.

An FDR should be stabilised, which means it should be harmonised or otherwise improved to maximise observational stability. Harmonization is the recalibration of the sensor observations based on a physical understanding of the instrument, using a stated reference and/or overlaps with other sensors in a series. The purpose of the recalibration is to bring consistency between sensors given the known (best estimate) differences in instrument characteristics (e.g., spectral response functions) between sensors. A harmonized (or otherwise stabilized) FDR should enable more stable TDPs or CDRs to be derived.

An FDR should be uncertainty-quantified. This means that alongside the FDR quantity values (the observations), the data sets should include all relevant uncertainty information. Such detailed uncertainty information should be provided at a per-pixel/datum/element/product[10] level (e.g. for an imaging sensor, per pixel of the image) and should also include the necessary information on the error correlation structures to enable users & TDP or CDR producers to understand the uncertainty on multiple spatial and temporal scales and combine multiple FDR data in a robust manner. It should also define the source, and ideally traceability, of each element of uncertainty information.

Finally, the FDR should be provided along with the information needed to evaluate it from raw measurements (e.g. a comprehensive ATBD). This data provision ensures the FDR data set meets long-term data preservation requirements and that the analysis and determination of the FDR is fully transparent and traceable.

---

[8] Generalised for applications beyond climate change monitoring that also require long-term stable records with robust uncertainty estimates

[9] This definition was suggested to the CEOS WG Climate by Chris Merchant, the principal investigator of the FIDUCEO project (www.fiduceo.eu). The definition builds on earlier definitions and was refined in a workshop on FCDRs hosted by the FIDUCEO project.

[10] The nomenclature for the basic data quality varies between communities and disciplines, although it is usual that any given sensor provider can easily identify and define this base data unit.

These are challenging requirements which may currently not be possible to meet for all observational records that are considered FDRs. This definition was developed around radiometric passive sensors operating in the solar reflective and thermal infrared/microwave spectral regions but should translate to more complex active or multi-dimensional sensors if applied to the lowest practical unit of data.

2.2 Climate Data Records and Thematic Data Products

The FIDUCEO project team developed the definition of a climate data record as:

> *A **Climate Data Record (CDR)** consists of a long, stabilised record of uncertainty-quantified retrieved values of a geophysical variable relevant to Earth's climate, together with all ancillary data used in retrieval and uncertainty estimation. The CDR is linked to (an) underlying fundamental climate data record(s).*

This definition links the CDR to an underlying FCDR. Again, it requires the CDR, like the FCDR, to be long-term, stabilised and provided with detailed uncertainty information. This definition can be broadened to a "thematic data product" as a long, stabilised records of uncertainty-quantified retrieved values of a geophysical variable linked to a fundamental data record.

Again, these are challenging requirements which may currently not be possible to meet for all observational records. In particular, the requirement for a CDR to be linked to an underlying FCDR may not be possible where no FCDR exists.

2.3 Developing FDRs and TDPs

In this report we consider both an ideal process for developing FDRs and the TDPs derived from them, and pragmatic options for where the ideal process cannot be followed. The principles here were originally developed for passive radiometric satellite sensors[11]. We have attempted to generalise these concepts for a wider variety of sensors and for FDRs as well as FCDRs; but recognise that concepts may be refined further during the FDR4ALT and FDR4ATMOS projects.

2.3.1 Ideal process for FDR and TDP/CDR generation

Ideally, an FDR would be derived first, and this would be used to generate a TDP or CDR. The FDR would be developed in the following steps. (Each described in subsequent sections of the report in more detail.)

1) A measurement function (or series of functions within a process) is defined that converts raw signal to the base sensor output, or FDR product. For example: for passive radiometric sensors this is almost always a function that converts counts to level 1 radiance at a per-pixel level; for spectrometers count arrays are converted to L1 radiance/irradiance spectra; and for active sensors, there are likely to be at least two steps – the conversion of raw signal to a processed signal (often a waveform) (level 1) and the conversion of the processed signal to derived observed quantity (level 2).
2) A diagram is used to document the traceability and sources of uncertainty for the terms in the measurement function. This will include an "uncertainty tree diagram" and /or "processing chain" or an embedded combination of these, as applicable

[11] FIDUCEO developed FCDRs for AVHRR (thermal infrared scanning imager in low earth orbit), for HIRS (thermal infrared scanning imager in low earth orbit), for the MHS and [check] microwave pushbroom sensors (low earth orbit) and for MVIRI, a broadband visible radiometer in geostationary orbit.

3) For each source of uncertainty identified in the diagrams, an 'effects table' will be filled in, which describes the magnitude of the uncertainty, its sensitivity, pdf and the error correlation structures in the relevant dimensions of the data product.
4) The Full FDR may be calculated, providing all the information in the effects table and all the data required to calculate the main FDR product, as well as to evaluate uncertainties. The FDR specifics may be driven by the proposed TDP/CDR with some aspects of the FCR more significant for this creation that others. The FDR should be completed in full, able to be used by the originally intended CDR or alternate analysis in future projects.
5) An Easy FDR is produced that summarises the Full FDR with the level 1 data product, uncertainties split into common, structured and independent components, and information to evaluate error correlation scales for structured effects and, where appropriate between spectral channels. This may be derived from the Full FDR or created directly.

Note that the process of analysing the uncertainties thoroughly, often leads instrument experts to propose an improved correction or otherwise adapt the measurement function. Furthermore, it may be appropriate to harmonise sensors to a common reference by recalibrating the calibration coefficients. Adding additional corrections or harmonising sensors, will lead to a new measurement function and may require iterating these steps.

Once an FDR has been derived, it will likely be used to develop a TDP/CDR. This will generally involve processing of the data, averaging over temporal and spatial scales, including auxiliary information, often from models. There is often also pre-processing steps such as scene determination and cloud screening, and post-processing steps such as gap filling. However, at the core of the process, there is a set of calculations, normally described in the ATBD. Whether this calculation can be written analytically or is determined only through a sequence of manipulation via a software algorithm, it can be described as the level 2/3 "measurement function". The steps to a TDP/CDR derived from an FDR mirror the FDR creation, requiring exactly the information, so are:

1) A measurement function (or series of functions within a process) is defined that converts FDR to the TDP/CDR
2) The processing chain is described diagrammatically as a series of steps, with indications of the origin of all information involved in each step. For the core calculation, where data are combined algebraically or algorithmically, a measurement-function centred 'uncertainty tree diagram' can be established.
3) For each source of uncertainty identified in the diagrams, an effects table is produced that describes the magnitude of the uncertainty and the error correlation structures in the relevant dimensions of the data product. Note that at this stage uncertainties are propagated from the FDR to the TDP/CDR and new uncertainties are considered.
4) The TDP/CDR is calculated and provided to the user with full or summary uncertainty information.

2.3.2 Pragmatic approach to FDR & TDP/CDR uncertainty analysis

Ideally all uncertainties are well known in the creation of the FDR and propagated from the FDR to the TDP/CDR, with additional uncertainty contributions included at the TDP/CDR processing stage due to auxiliary information included in that processing. However, there are situations where the relevant information is not available about the FDR uncertainties, or about some of the auxiliary information.

In these cases, it is necessary to understand the sensitivity of the overall uncertainty budget to these unquantified factors and assess the relative contribution. There are several approaches that can be used, with a 'worst case scenario' analysis typically the first assessment. The effects table includes a descriptor on the confidence in the uncertainty estimations, and the source of the uncertainty information. Missing information or un-quantified uncertainties is by no means uncommon. The underlying principle is to use the best data available and develop an uncertainty budget from all available information. This may include combining "bottom-up" uncertainty analysis (propagating uncertainties that have been evaluated at previous steps) and "top-down" uncertainty analysis (performing comparisons of the TDP/CDR against references that have better uncertainty analysis to estimate a total uncertainty).

Such "top-down" analyses are commonly used in Earth Observation at present, sometimes in part due to a lack of information about Level 1 uncertainties. Within the FDR projects we want, conceptually, to move away from such an approach and to reserve the use of "top-down" analyses to verify uncertainties (rather than to evaluate them). Therefore, such methods should be used only to fill in gaps that cannot be filled in from a bottom-up approach.

2.4 Metrological concepts

An FDR, and a TDP/CDR derived from an FDR, both require a metrological approach. Metrology has three basic principles:

> **Traceability**: Metrological traceability is a property of a measurement that relates the measured value to a stated metrological reference through an unbroken chain of calibrations or comparisons. It requires, for each step in the traceability chain, that uncertainties are evaluated and that methodologies are documented.
>
> > For Earth Observation traceability includes the formal calibration traceability of the instrument to appropriate references in pre-flight, in-orbit and vicarious calibrations, including the metrological traceability of any references (e.g. in situ observations).
>
> **Uncertainty Analysis**: Uncertainty analysis is the review of all sources of uncertainty and the propagation of that uncertainty through the traceability chain. It is based on the Guide to the Expression of Uncertainty in Measurement (the GUM).
>
> > For Earth Observation, uncertainty analysis needs to consider not only the calibration of satellite data (e.g. for L1 processing) but also the propagation of those uncertainties through retrieval algorithms for L2 and above processing.
>
> **Comparison:** Metrologists validate uncertainty analysis and confirm traceability through comparisons. The Mutual Recognition Arrangement (the inter-comparison of NMIs and the standards they disseminate) requires regular, formal international comparisons that are conducted under strict rules.
>
> > Earth Observation comparisons are carried out between sensors and between sensors and in situ observations using simultaneous observations, transfer standards (e.g. ground observations or pseudo-invariant sites/scenes), or large scale averages. Traditionally Earth Observation comparisons have been performed to estimate inter-sensor differences, we are only beginning to use them to validate uncertainties.

A discussion of these topics is provided in Appendix **Error! Reference source not found.**, which also p rovides links to reference and training material relating to these topics. Note that NPL is preparing an e-Learning course for FCDR production, with initial modules expected to be available in Spring/Summer 2020.

3 Establishing the measurement function

3.1 The measurement function of the FDR

The first step to establishing an FDR is to identify the measurement function. This is the function used to calculate the FDR from the raw, Level 0 product.

As an example, for a passive radiometric sensor, the measurement function will be a function that calculates the Earth radiance L_E for pixel p in channel c from the Earth counts C_E for pixel p in channel c . It will take a form similar to

$$L_{\mathrm{E},pc} = a_0 + a_1 C_{\mathrm{E},pc} + a_2 C_{\mathrm{E},pc}^2 + 0. \qquad \textbf{Eq 3-1}$$

Here a_0, a_1 and a_2 are calibration coefficients that may be determined from pre-launch characterisation, from post-launch vicarious calibration, from onboard calibration processes or through harmonisation (recalibration against a reference sensor or other sensors in a series). Note that the a_i terms may themselves be calculated from a more complex expression, for example the a_1 term is the radiometric gain and this may be measured from a view of an internal calibration target (and therefore depend on other observed signals), or it may be a function of detector pixel, or viewing angle. The "plus zero" term represents the extent to which this measurement function is an approximation, for example, due to assumptions that spectral band integration can be performed numerically and that the resultant Earth radiance can be assumed to be "monochromatic", or the assumption that a quadratic nonlinearity adequately represents the nonlinear behaviour of the sensor. In the uncertainty analysis we will consider the uncertainty associated with this plus zero assumption.

For atmospheric spectrometers the full measurement function may be complex, multi-stage and not suited to a single expression. However, the measurement function can be broken down into to the process steps and expressed as a model, that encapsulate the structure and interdependencies within the full measurement function, this process is described in §4.

For an active sensor, it is unlikely that the measurement function can be written explicitly. Furthermore, the Level 1 product is a waveform (and therefore has both time and power axes) rather than a single measurand. The processing from Level 0 to Level 1 for altimetry involves corrections of both the time and power axes and while most individual components can be expressed in measurement functions, there is no simple equation for the "Level 1 measurand". Despite this, we can think in terms of some kind of model:

$$P(t) = f(t, \boldsymbol{c}_t; P, \boldsymbol{c}_P) + 0 \qquad \textbf{Eq 3-2}$$

where $\boldsymbol{c}_t$ is a vector quantity representing the different time corrections (which could be individually provided) and $\boldsymbol{c}_P$ is similarly the different power corrections.

3.2 The measurement function of the TDP/CDR

For radiometric sensors the Level 2 product is some retrieved geophysical quantity, often corrected for atmospheric transmittance. Generally, the retrieval of L2 variables, $\boldsymbol{z}$, can often be written in a form

$$\boldsymbol{z} = g(\boldsymbol{y}, \boldsymbol{S}_\varepsilon, \boldsymbol{z}_a, \boldsymbol{S}_a, \boldsymbol{b}) + 0 \qquad \textbf{Eq 3-3}$$

Here, $\boldsymbol{z}$ is the 'retrieved state' for a given pixel at line-element (*l*, *e*); this 'state vector' may contain a single variable (at the surface or a particular atmospheric level), a vertical profile of a single variable, or a set of various variables. g is the measurement function for the L2 product. This may be a simple, analytic equation, or considerably more complex. The n_c values of radiance used for the retrieval are in the 'observation vector' $\boldsymbol{y} = \left[L_{1,l,e}, \dots, L_{n_c,l,e}\right]^{\mathrm{T}}$. $\boldsymbol{S_\varepsilon}$ is an evaluation of the error covariance matrix of the measured radiances. $\boldsymbol{z}_a$ is a prior estimate of the state with estimated error covariance $\boldsymbol{S}_a$. $\boldsymbol{S_\varepsilon}$, $\boldsymbol{z}_a$ and $\boldsymbol{S}_a$ are explicitly present in some retrieval algorithms such as optimal estimation (Rodgers, 2000), in which case they can be explicitly used to evaluate the error covariances of the retrieved state. In other cases, they are implicit and unevaluated. All retrieval algorithms have additional parameters used in retrieval, ***b***. This can be as simple as a set of weights for combining radiances, or could be tens of thousands of spectroscopic parameters embedded within a radiative transfer model (or 'forward model') used in the inversion. The '+ 0' indicates that not all aspects of the inverse problem are necessarily captured by terms in the measurement function; some uncertainty is associated with those aspects of the inverse problem, even though their net effect is assumed to be zero mean.

Note that prior to this retrieval step there may be some pre-processing steps that cannot be expressed analytically, for example cloud screening and pixel classification.

For active sensors, the Level 2 product is determined from the Level 1 waveform. There are usually two steps involved. For example, for altimetry in the first step the range is determined from the waveform either numerically using heuristic approaches (e.g based on the half peak time) or analytically by fitting a model to the waveform with the range one of the fitting parameters.

The second step involves correcting for geophysical and atmospheric properties. While the analysis of these corrections is complex (and involves modelled data), the final equation is often extremely simple as corrections are added or subtracted from the measured range, for instance here for an altimetry altitude correction,

$$z = h_{\mathrm{orbit}} - (R + C_{\mathrm{DTC}} + C_{\mathrm{WTC}} + C_{\mathrm{Iono}}) - z'_{\mathrm{geo}} + 0.\ \text{[12]} \qquad \textbf{Eq 3-4}$$

3.3 Uncertainty information essentials

The work done in the EO metrological projects centre around three key principles. There are a variety of ways of representing the information content required by these principles, graphically or in tabular form, but these are only variations on the underlying principles, which should not be lost sight of in the following sections. The framework in this document endeavours to standardise the approach and reporting of this essential information, in a way that can be translated between sensors, geophysical products and product levels adding value to the user understanding of the dataset.

Principle #1 – All contributing processes are considered. The graphical representation of the measurement function or processing steps as a measurement equation, uncertainty tree or process chain are fundamentally ways to represent the input data, transformations, algorithms, decisions & assumptions made. By methodically capturing the process, detailing their dependencies (and the dependencies of dependencies) then the full extent of the process is more robustly captured. See §4

Principle #2 – The essential uncertainty parameters associated with each process step are reported. Each input dataset, transformations, algorithm application, decisions & assumptions will (in the vast

[12] Eq 3-4 describes an altitude correction algorithm from altimetry, where h_orbit is the height above geoid, R is the range to the surface, plus corrections for DTC (dry tropospheric correction), WTC (wet tropospheric correction) and Ionospheric correction (Iono) with z_geo the geoid.

majority of cases) contribute to the uncertainty associated with the resultant data. The uncertainty parameters necessary to robustly describe the uncertainty propagation is the following:

1. The uncertainty value associated with the process step (§5.3)
2. The (SI) units of the process uncertainty contribution (§5.3)
3. The sensitivity of the underlying function to the process (§5.3)
4. The probability distribution function associated to the process (§5.3)
5. The correlation(s) of this process to other processes (§6)

Additionally, some evidence for the source of the above information is required.

This information needs to be reported for each process. Only with this information can the uncertainty associated with the output dataset be robustly determined. Where the information above is not known, either a best estimate should be provided, or the lack of information should be recorded.

Principle #3 – The uncertainty is reported clearly and transparently, in a relevant & accessible format for the user communities, allowing a robust uncertainty to be calculated on the temporal, spatial and spectral scale required by the user. As well as for long term preservation applications. See §7.

Uncertainty analysis assumes that all known corrections are applied, and the uncertainty analysis is applied to the residual effects. The process of a rigorous uncertainty analysis does tend to highlight additional corrections and improve the uncertainty as a result. It is understood that a full uncertainty assessment can be very challenging, but the principle remains, even if not fully met in any specific activity or project. The outcomes do however provide a list of recommendations for future work to advance the sensor knowledge towards a fuller understanding.

4 Diagrammatic representation of the traceability and sources of uncertainty

For simple radiometric sensors, the Level 1 processing to an FDR is usually a single equation (although input quantities in that equation may be themselves determined from prior analysis). The Level 2 processing often involves a series of pre-processing steps (e.g. cloud screening) that are performed prior to the main retrieval algorithm[13]. Two visualisation approaches to document the algorithms and sources of uncertainty have been developed, see §4.2 & §4.3. But initially, some general guidance will be presented.

4.1 Diagrams – general guidance

The Uncertainty Tree Diagram and Processing Chain Diagram are both methods for visualising where the measurand (the FDR or CDR/TDP quantity) comes from. We have found that this has two benefits:

- For the producer of an FDR or CDR/TDP, the diagrams help organise thinking, clarify interdependencies and often identify assumptions and sources of uncertainty that were previously missed.
- For the user of an FDR or CDR/TDP they provide information on where the product comes from and help users assess fitness for purpose.

Processing flow charts are relatively common in existing algorithm theoretical basis documents (ATBDs), but a metrological review will likely add to such diagrams, with further information on the

[13] Data filtering, although may not change the data value or its numerical uncertainty, may change the pdf of the distribution, sometimes truncating a normal distribution, for instance.

origin of the data that goes into such processing. The uncertainty tree diagram is not widely used at present but have been shown to be instructive where it has been applied.

4.2 The uncertainty tree diagram

The "Uncertainty Tree Diagram" takes the form shown below. The uncertainty tree diagram captures the measurement function and the structure of the dependencies, together with expressions for the sensitivities and short uncertainty contribution descriptors. The central box contains the measurement function, either written out in full, or written conceptually as a function of input parameters. This should include the "plus zero" term. Some terms in the measurement function are directly provided and have a single source of uncertainty (e.g x_3 in the diagram below). These are shown with the sensitivity coefficient between the term and the uncertainty (descriptor).

Others, such as x_1 in the diagram are directly measured but may be influenced by more than one "effect", each a separate source of uncertainty. Still others, e.g. x_2, are themselves calculated from other input quantities, which have their own sources of uncertainty. We should also document the uncertainties associated with the "plus zero" – these are the uncertainties associated with the assumptions implicit in the form of the function.

For some sensors such diagrams become extremely complex. In this case, it may not be possible to provide all information on a single figure. This has been resolved by nesting uncertainty tree diagrams (in some cases interactive), where sub-chains are represented separately on separate figures.

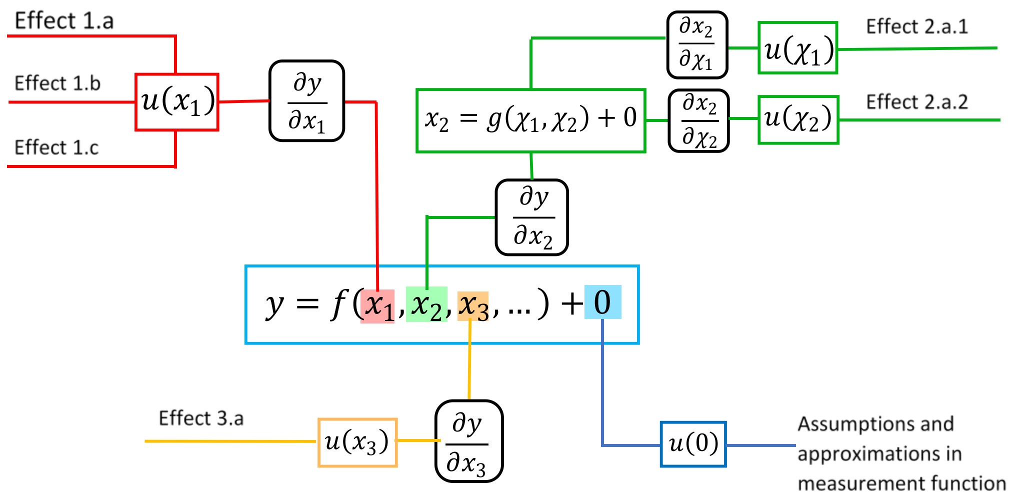


Figure 1 Conceptual Uncertainty Tree Diagram

4.3 Processing chain diagram

The uncertainty tree diagram is a useful approach to documenting a process that involves a single equation. Where corrections must be performed in a specified sequence, then a "processing chain diagram" is more appropriate. The concept of the processing chain diagram was developed in the QA4ECV project and improved in the GAIA-CLIM project, where it was applied to a broad range of non-satellite observation systems.

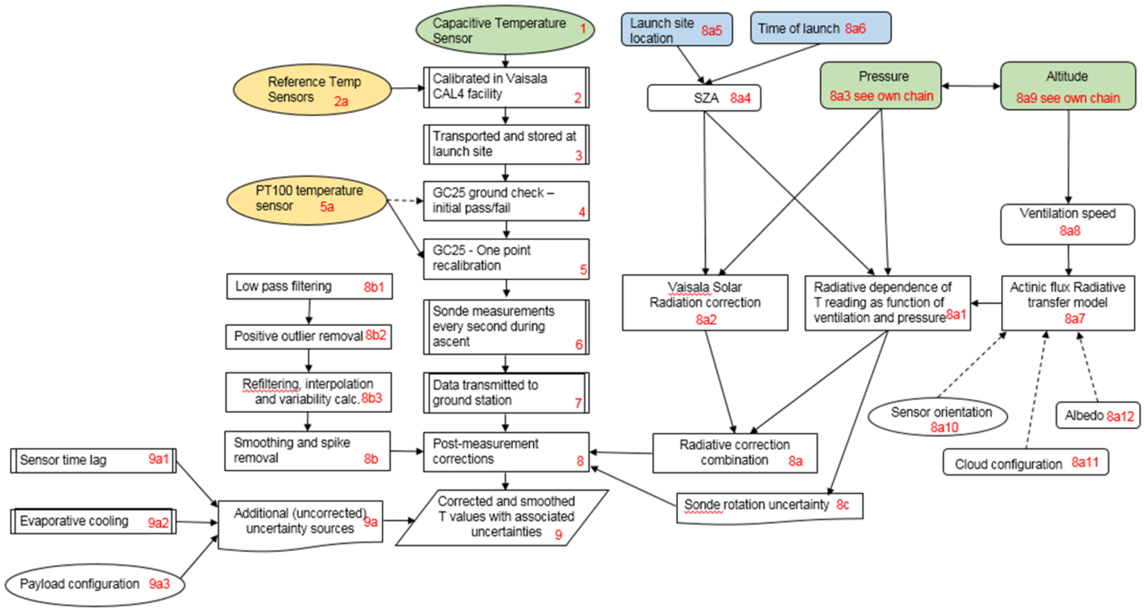


Figure 2. Example process chain for a RS92 radiosonde temperature measurement.

The chains should be drawn, graphically, as a series of boxes connected to one another via uni- or bi-directional arrows, as seen in Figure 2 (An example diagram for a RS92 radiosonde temperature product[14]). Guidance on the types of boxes for each type of chain element is given at Table 1. However, it is noted that the underlying process flow information is the important content, so excessive effort should not be spent in formatting the diagrams.

Table 1. Traceability Chain Shapes and Definitions

Input / Output dataset	Parallelogram	A dataset visible to the user, be that initial input, final output product or any intermediate product that is available to the user.
Process / processing step	Rectangle	A process within the chain, used to describe a transformation in the dataset that may or may not have an associated uncertainty. The default box shape. The dataflow within the process is typically invisible to the user.
Process	Rectangle with side-bars	Essentially identical to the process rectangle. However, sometimes used to represent a sub-chain or major processing block where more granular information is available.
Instrument / Physical item	Ellipse	Raw data from a measurement device central to the product value or its traceability. This can also include the data propagated from a previous Level.

14 GAIA CLIM D2.6 http://www.gaia-clim.eu/page/deliverables

Physical quantity	Rounded rectangle	An ancillary physical quantity dataset or product necessary in the processing chain or to give context to the product.
Isolated Uncertainty	Rectangle with wavy bottom	An uncertainty quantity not associated with (isolated from) an element in the traceability chain. Typically used to represent assumptions and known effects that are not directly corrected for (i.e. effects that become part of the +0 term).
Decision	Rhombus	A decision step that may affect whether specific data appears in the output product. Such decisions may impact the probability distribution function of the uncertainty.

If there is a complex sub process, this can be separated out, with an example shown in Figure 3

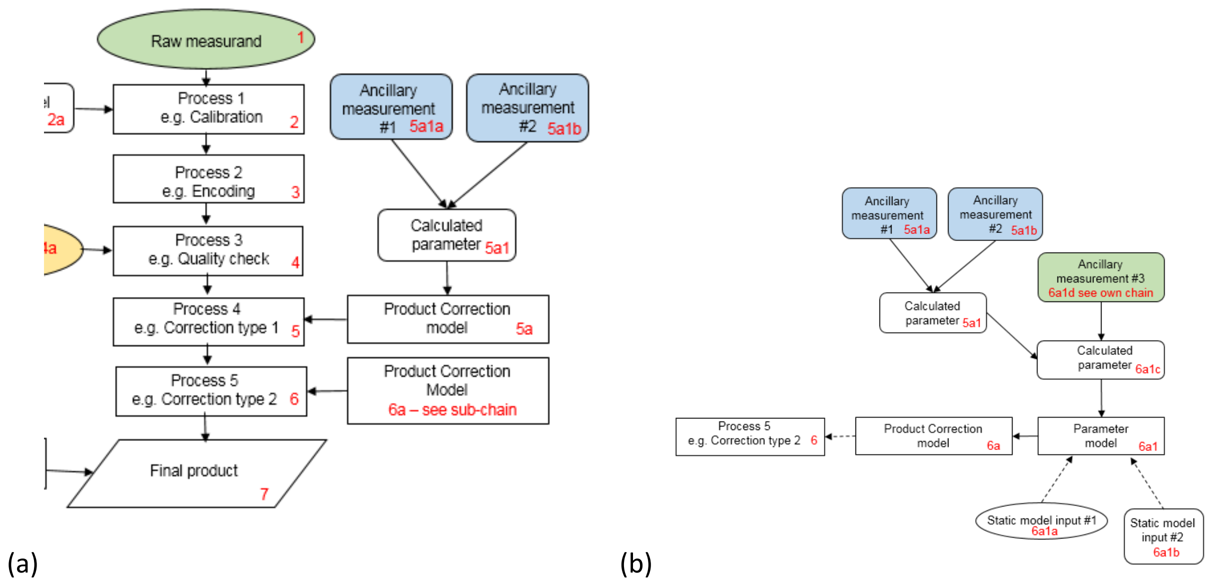


Figure 3 (a) extract from a processing chain which includes a reference to a sub chain, and (b) sub chain for that process.

4.4 Diagrams – details

4.4.1 Hybrid diagrams

It is likely that for all processes, both diagrams are needed. In many cases a basic chain is the starting point in visualising the process, with tree diagrams developed once the underlying structure is crystallised. The processing chain diagram is useful where processing steps must be performed in a specified order. Where corrections and calculations can be applied in any order, then the uncertainty tree diagram is more applicable. In many cases a hybrid scheme with tree diagrams embedded into a chain is most suitable. An initial assessment of the underlying structure for any given application will likely determine the most suitable construct.

For example, for many passive radiometric sensors, the uncertainty tree diagram is more suitable for the FDR and the processing chain more suitable for the TDP/CDR, with embedded uncertainty trees as specific stages, for example for retrievals.

For some active sensors, this may be the other way around: the processing to obtain an FDR is a series of transformations that must be performed in a specific order. The final TDP/CDR however, is a simple additive expression for geophysical corrections, that can be described in an uncertainty tree diagram.

4.4.2 Diagram numbering schemes & diagram depth

Numbering the process steps allow easier reference to the element tables. The 'main chain' from raw measurand to final product forms the axis of the diagram, with top level identifiers (i.e. 1, 2, 3 etc.). GAIA-CLIM used the numbering convention as shown in **Error! Reference source not found.**, where s ide branch processes add sub-levels components to the top level identifier by adding alternate letters & numbers. The framework is that:

- a sub-level is added at dependency branching to help orientate the contribution within the overall scheme.
- The reported uncertainty at a level includes the combination of uncertainties for the sub-levels.

We should also think pragmatically about "how deep the diagram should go". For example, on a thermal infrared or microwave sensor, the onboard gain calibration is based on a blackbody, which is ultimately traceable on the SI ITS-90 standard and its derivatives[15].

Some pragmatic decision needs to be made about how far back to go and this will depend on the error correlation structures. In this example, in almost all cases, it is enough to have the temperature of the blackbody as the final parameter, with, as twigs, the uncertainty components: "noise on PRTs", "average PRT calibration bias", "PRT consistency" and "PRT representativeness". The noise on PRTs will be the uncertainty associated with the mean of the PRT readings due to noise, and the PRT representativeness will be the extent to which the mean of those PRTs represents the temperature seen by the radiometer.

4.4.3 Sources of common uncertainty

Where two quantities in the measurement function rely on exactly the same instance of the same information, we have a common uncertainty. This may come in artificially through the introduction of simplification quantities. For example, in the HIRS measurement function[16] we wrote

$$L_{\mathrm{E}} = \alpha\left(C_{\mathrm{E}} - \bar{C}_{\mathrm{S}}\right) + a_2\left(C_{\mathrm{E}}^2 - \bar{C}_{\mathrm{S}}^2\right) - \left(L_{\mathrm{self,E}}\left(T_{\mathrm{inst}}\right) - L_{\mathrm{self,S}}\left(T_{\mathrm{inst}}\right)\right) + a_4 + 0 \qquad \textbf{Eq 4-1}$$

[15] The blackbody radiance is determined by averaging the signal on platinum resistance thermometers (PRTs), whose temperature is determined by a fit of a measured count value to a polynomial, the coefficients of which are determined by pre-flight calibration by the manufacturer against a fixed-point cell that is traceable to the international temperature scale of 1990, which is based on thermodynamic measurements of fixed-points at national metrology institutes which were established using a transfer radiometer and so on. There is no need to take the diagram back to the origin of the international temperature scale!

[16] http://www.fiduceo.eu/publications

where α is defined by

$$\alpha = -\frac{\tilde{L}_{\mathrm{IWCT}} + L_{\mathrm{self,IWCT}}\left(T_{\mathrm{inst}}\right) - L_{\mathrm{self,S}}\left(T_{\mathrm{inst}}\right) - a_2\left(\bar{C}_{\mathrm{IWCT}}^2 - \bar{C}_{\mathrm{S}}^2\right)}{\bar{C}_{\mathrm{IWCT}} - \bar{C}_{\mathrm{S}}}. \qquad \textbf{Eq 4-2}$$

Clearly there are common quantities between α and other terms such as $\bar{C}_{\mathrm{S}}$ and $L_{\mathrm{self,S}}\left(T_{\mathrm{inst}}\right)$, but we also note that this is simply a space-saving way of writing the full measurement function and such correlations are easily handled by calculating sensitivity coefficients, such as $\frac{\partial L_{\mathrm{E}}}{\partial \bar{C}_{\mathrm{IWCT}}}$ in full.

There are some cases where such a process is more subtle. For example, it is typical to have a solar zenith angle term, θ , which feeds into multiple processes. This in turn is calculated from the solar declination, the latitude and the local hour angle. Both the solar declination and the local hour angle are determined from the acquisition time and therefore any error in that acquisition time is common to both the solar declination and local hour angle. Therefore, these two parameters are correlated. To avoid handling this correlation, however, we can consider the errors in the fundamental parameters – acquisition time, longitude and latitude – as the effects, and determine the sensitivity coefficient for the solar zenith angle to those three more fundamental parameters, even if computationally, and for understanding the process, it is easier to use the intermediate parameters.

This sort of decision must be made pragmatically and remembering the aim of this process is to get the right balance between making the diagram intuitively easy to understand and making the diagram show the full rigorous traceability of the measurement. It is also important to understand the contribution an individual uncertainty source has on the overall combined uncertainty as this can guide the appropriate level of detail required for a contribution.

4.4.4 Colour schemes and presentation details

The colour scheme is not defined but should be chosen by the producer to best illustrate the commonality in the specific traceability chains. For example, to indicate the raw data sources, the source of traceability, ancillary products, to group a set of boxes which contribute to a single process or, for interactive chains, that further information associated with the box is available, see §**Error! Reference source not found.**.

5 The effects tables

Once an uncertainty tree diagram and/or a processing chain has been produced, each source of uncertainty (the twigs on the tree, the input elements of the processing chain and each process step towards the output product) should be considered carefully and the results of that analysis should be recorded in an effects table. The effects table is a way to describe everything that needs to be known about the uncertainty component for the uncertainties to be propagated properly, as described in §3.3, but with some additional fidelity to capture the temporal and spatial scales over which these apply. The effects tables need to be reported within a (human-readable) document accessible to the data users and stored digitally (in a computer-readable form).

5.1 Effects table headers

The example effects table takes the following form. The first two rows give a name to the effect (source of uncertainty) and state which term in the measurement function is affected by that effect. Other rows are discussed in their sections below.

Where multiple instruments are being considered, the tables should be separate for the individual instruments where there is any difference. A common table is only acceptable if no content is different. In all cases, identification of the relevant instrument should be clear.

Table 2

Table descriptor		Value/Parameter	Notes
Name of effect		A unique name to describe the effect	
Effect identifier		A unique number used to identify the position in the uncertainty tree or process chain	
Affected term in measurement function		Name and standard symbol of affected term	Usually an effect will only affect a single term, though there may be exceptions. The next higher-level identifier should be reported.
Maturity of analysis	Maturity of uncertainty estimate	0 – Effect identified, no quantification performed (no further information in cells below) 1 – Rough estimates only 2 – Some analysis performed to estimate values 3 – Rigorous analysis performed	This allows for the fact in the FDR/CDR we haven't thought everything through in detail and makes that very clear to users. If the maturity is low, we may still be able to estimate if it is negligible or minor, or if it's possibly significant (and therefore needs more work soon) Reference to the evidence for the maturity assessment, e.g. publication, report, weblink etc.
	Maturity of correlation scale estimate	0 – Not done 1 – Estimated 2 – Based on analysis, unsure about correlation shape 3 – Strong evidence	
	If maturity of estimate is 0 or 1, how significant do you expect this effect to be?	Negligible, Minor or Significant? For pixel level results and for long-term / large scale results	
Correlation type and form	From level xx	Select one of the types defined in §6.3 and Table 4	See §6.3 Define the level of analysis from, then to, e.g. level 0 to level1, and the relevant scales, e.g. per scan, orbit, calibration cycle etc. If there is a correlation with another effect, state its identifier.
	temporal scale type & form [time]		
	spatial scale type & form [geospatial coordinates]		
	Spectral type & form		
Correlation scale	From level xx	What is the correlation scale	See §6.4
	temporal scale [time]		
	spatial scale [geospatial coordinates]		
	Spectral scale		
Uncertainty	PDF shape	Functional form of estimated error distribution for the term, see Table 3	
	units	Units in which PDF shape is expressed (units of term, or can be as percentage etc)	See comment in §5.3.1 where uncertainty and sensitivity cannot be separated
	magnitude	Value(s) or parameterisation estimating width of PDF	
Sensitivity coefficient		Value, equation or parameterisation of sensitivity of measurand to term Can also flag "included in uncertainty" (by making this equal 1)	Where the uncertainty and sensitivity coefficient cannot be separated the sensitivity coefficient should be one and the uncertainty is in units of the measurand.
Validation		A description of any validation of the uncertainty at effect level.	The source of the uncertainty information and validation should also be identified.

5.2 Effects table maturity

A robust metrological review of every source of uncertainty may not always be possible, either because information is not available or because project timescales require prioritisations to be made. We have therefore introduced the concept of the maturity of analysis. This was a four-point scale, based on your expert judgement, on the maturity of the evaluation of the uncertainty magnitude, of the error-correlation scale and form, and with an impact statement.

Maturity of analysis	Maturity of uncertainty evaluation	0 – Effect identified, no quantification performed 1 – Estimates only 2 – Some analysis performed to evaluate 3 – Rigorous analysis performed
	Maturity of correlation scale evaluation	0 – Not done 1 – Estimated 2 – Scale based on analysis, unsure about correlation shape 3 – Strong evidence for correlation scale and shape
	If maturity of estimate is 0 or 1, how significant do you expect this effect to be on pixel-level estimates?	Negligible, Minor or Significant? *(Preferably with explanation or evidence)*
	If maturity of estimate is 0 or 1, how significant do you expect this effect to be on large spatial scales / long time series?	Negligible, Minor or Significant? *(Preferably with explanation or evidence)*

This could be added to the effects table or could be described in a set of bullet points under a heading for that uncertainty component.

5.3 Uncertainty evaluation and sensitivity coefficient

5.3.1 Documenting the uncertainty

Uncertainty	PDF shape	Functional form of estimated error distribution for the term	
	units	Units in which PDF shape is expressed (units of term, or can be as percentage etc)	See comment below where uncertainty and sensitivity cannot be separated
	magnitude	Value(s) or parameterisation estimating width of PDF	
Sensitivity coefficient		Value, equation or parameterisation of sensitivity of measurand to term	Where the uncertainty and sensitivity coefficient cannot be separated the sensitivity coefficient should be one and the uncertainty is in units of the CDR measurand.

The uncertainty rows describe the shape, units and magnitude of the uncertainty and an expression for calculating the sensitivity coefficient. In the supporting documentation, some evidence is required to explain the origin of the values given here.

The uncertainty (“magnitude” row) is the parameter that characterises the dispersion (standard deviation) of values that could be attributed to the measurand based on the measurement. It is always a standard uncertainty (one standard deviation, and never an expanded uncertainty, e.g. for k = 2). The uncertainty will usually have the units of the effect, although in some cases uncertainties may be expressed in percentage.

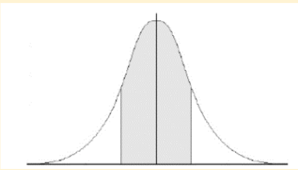
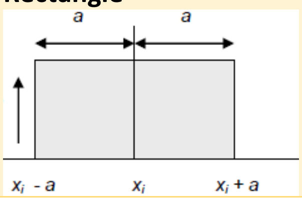
The sensitivity coefficient translates the uncertainty associated with the effect, in the units given in the “uncertainty units” row, into an uncertainty associated with the measurand in the units of the measurand. Where the measurand is calculated from an analytical expression, the sensitivity coefficient is the partial derivative of the measurement function with respect to the term that this uncertainty applies to, $\partial f / \partial x_i$. For uncertainty effects that are shown on the uncertainty tree diagram as a chained series of calculation, the sensitivity coefficient is calculated from the chain rule.

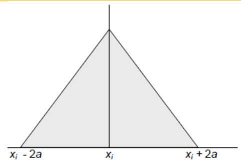
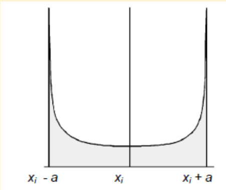
Where the sensitivity coefficient cannot be evaluated analytically, because, for example, the analysis is performed through an iterative software process rather than an analytical expression, it may be evaluated numerically, for example via Monte Carlo Methods. In this case, the uncertainty may be expressed in units of the measurand (as the effect it has on the measurand) and the sensitivity coefficient is 1.

The PDF shape will be one of a defined list of shapes given in Table 3. The actual PDF may not fit perfectly to one of these shapes, but they are likely to be sufficiently close to most actual PDFs, otherwise use the ‘Other’ option.

Table 3 describes common PDF shapes and what the standard uncertainty (the value in “magnitude” under uncertainty in the Effects tables) refers to.

Table 3 Parameters defined for different PDFs. For an explanation of these standard uncertainty values, see the GUM section 4.4.

PDF shape	What is the standard uncertainty	Description
Gaussian 	$u = \sigma$	Be careful when using published literature, or a calibration certificate, to provide u . If an expanded uncertainty is quoted, then it’s important to divide by k (often $k = 2$ in certificates).
Digitised_Gaussian	Unknown	The most appropriate standard uncertainty for a digitised Gaussian has not been fully evaluated. Please treat as a Gaussian, but keep this option open for the future
Rectangle 	$u = a/\sqrt{3}$ where a is the half width	Useful for when we know a quantity must be in a range $\pm a$, but it’s equally likely to be anywhere in that range, e.g. digitisation
Triangular	$u = a/\sqrt{6}$ where a is the half base	Useful for where we know there is a range a quantity is in but it’s more likely to be in the middle of

		that range (e.g. when a quantity is the difference between two digitised values)
U-distribution 	$u = a/\sqrt{2}$ where a is the half base	Useful for where we know there is a range a quantity is in but it's more likely to be at the edges of that range (e.g. where there is a feedback loop that switches on and off and encourages drift to the two ends of a temperature range)
Other		If the PDF is not one of these, but a standard uncertainty can be provided, then this is also acceptable, a note should be added in documentation.

5.3.2 Evaluating the uncertainty

There are many ways to do the uncertainty evaluation and the choice will depend on the nature of the uncertainty and the available information. In Appendix 10B.2 we consider some specific examples. These generally fall into one of the following methods:

- **Provided uncertainties** – if a calibration coefficient is determined through harmonisation or through pre-flight laboratory-based calibration, an uncertainty should be provided with the quantity. It is important to consider the provenance of this uncertainty statement. If it has been rigorously analysed with a "fiducial" QA4EO-compliant method, or is audited to ISO 17025, then it is likely to be directly useable. If it is based on a less rigorous analysis it may be appropriate to review the uncertainty calculation independently (where information is available).
- **Noise estimates** – one of the challenges in EO is that, because the scene is changing, the signal is varying all the time and therefore laboratory approaches of making repeat measurements of a stable source are not possible. However, most satellite sensors have some information about noise performance, from, e.g. for radiometric sensors, a stable scene, onboard calibrator, or deep space views. The Allan deviation can be useful here[17]. For active sensors, noise information is available from the repeatability of the individual waveforms that are averaged to give the final waveform (these are provided in some instances) and from onboard calibration modes.
- **Modelling processes** - sometimes it is possible to estimate the scale of a particular source of uncertainty by modelling the processes on board. In the FIDUCEO project this was done for example for the AVHRR onboard calibration target, where thermal gradients caused by direct solar heating were modelled based on a physical model of the instrument and the available information.[18]
- **Comparison to a reference** - there are occasions when an independent reference measurement is available [e.g. in-situ data], and comparisons to that reference can be used to evaluate the uncertainty. This has been a common method in Earth Observation to evaluate

---

[17] See Mittaz, J., 2016, Instrument noise characterization and the Allan/M-sample variance

[18] See Appendix A.3.2 in Taylor, M., Desmos, M. & Woolliams, E., 2017. D2.2 (AVHRR): Report on the AVHRR FCDR uncertainty

measurement uncertainties and is sometimes the only option. Care needs to be taken to consider the uncertainty associated with the reference, and it is better if this comparison is performed on specific input parameters and not on the resultant measurand. In addition, consideration of the collocation uncertainty – due to any spatial and temporal mismatch between the two measurements should form a part of any comparison exercise.

6 Error correlation scales

6.1 Different types of error correlation

Each effect will have an associated uncertainty and that means that it will lead to an (unknown) error. In order to use the FCDR we need to understand the correlation in three different senses:

- The error correlation between different quantities in the measurement function is needed for combining such quantities. Such error correlations arise from effects in common between quantities, and where quantities are determined from a common fit to data. Where possible, we avoid this type of error correlation, but sometimes it must be considered.
- The spatial-temporal error correlation between measured values taken in different locations and/or at different times is needed for higher level processing that combines data spatially or temporally (e.g. for spatial or temporal averages and trend analysis). For example, such correlation occurs when there is a common calibration approach, due to corrections that are applied which are common over multiple individual observations or due to methods such as rolling averaging or spline interpolation.
- For a multi-channel instrument with different spectral channels or operating modes, then the error correlation between measured values in different channels / modes must be considered for any application that combines such data. Such error correlation could arise if a common calibration source is used to calibrate different channels.

These are fundamentally different types of correlation and have to be considered separately.

Note that the measurand itself could be spatially or temporarily correlated due to geophysical processes. Unless such processes are used to determine corrections, this correlation is not considered in uncertainty analysis.

The error-correlation between different quantities in the measurement function needs to be considered separately from the effects tables as this is considered only when uncertainties are combined. Although these correlations should be identified and noted within the effects table – to make transparent to the user that they exist.

6.2 Dimensions for error correlation

The error correlation needs to be considered along the appropriate dimensions. For the low-earth-orbiting radiometric sensors considered in FIDUCEO, these were cross track (pixel-to-pixel), along track (scanline to scanline), orbit-to-orbit and, separately, between spectral bands. For the geostationary radiometric sensor, the dimensions were the same, but image-to-image replaced “orbit-to-orbit”. For other operating modes, different error correlation dimensions are needed. These may include spatial and temporal dimensions, and between detectors, channels or modes. They may also depend on surface type, so, for example in an altimetry sensor there could be components that are correlated over ocean, or over sea ice, for example. Each assessment should identify the relevant scales for the sensor under review and report accordingly.

For the CDR/TDP, there are also error correlation that depend not on the operating mode of the satellite, but the temporal and spatial scales of the auxiliary information, for example, information from a modelled atmospheric condition.

The same dimensions should be used for all effects tables for a particular measurement equation. For the FCDR of LEO radiometric sensors, the following table form has been used:

Correlation type and form	Pixel-to-pixel [pixels]
	from scanline to scanline [scanlines]
	between images [images]
	Between orbits [orbit]
	Over time [time]

With a second section for the spectral channel correlations:

Channels/ bands	List of channels / bands affected
	Error correlation coefficient matrix

For the CDR, the following as also included,

Correlation type and form	From level 1
	Larger scale temporal [time]
	Larger scale spatial [geospatial coordinates]

6.3 Error correlation structures and correlation scale

Within any one of the dimensions described above, the error correlation can take different forms. The list below (& Table 4) is not exhaustive, but is a menu of error correlation forms that are sufficiently close to those expected in practical cases. The error correlation form describes the correlation coefficient between any two measured values in the dimension for which it is defined.

The defined error correlation forms are:

- **Random:** In this there is no error correlation with any other measured value.
- **Rectangular absolute (contains systematic):** In this the error correlation is constant for a particular range of values defined absolutely, rather than relative to the measured value. This includes the following cases:
    - Where a single measured value is used over an explicit range, e.g. where a single calibration value is used for all measurements over several scan cycles, or in a particular year, and a different calibration value is used for all measurements outside that range.
    - For an effect that is fully systematic in that dimension (common to all measured values in that dimension). This is described with the dimensions $[-\infty,+\infty]$

- **Triangular relative:** In this the error correlation drops linearly (in the dimension of interest) relative to a particular measured value. This comes from a running average with constant weights.
- **Bell-shaped relative:** In this the error correlation drops faster than linearly (in the dimension of interest) relative to a particular measured value. This can come from:
    - A weighted running average (e.g. over neighbouring scan cycles), which weights the central reading more than the others involved in the average.
    - Any other form of weighted averaging (e.g. through a spline fit in geolocation)
    - Other cases where our expectation is that the correlation drops off over distance in some way.

  In none of these cases is the error correlation form exactly Gaussian, but a truncated Gaussian form is a practical approximation for the Bell-shaped form, and is typically used. What this correlation form represents is the situation where "nearby" errors are relatively highly correlated, but this correlation drops off over a distance. By defining the Gaussian width and the truncation range (beyond which there is no error correlation), it is possible to define a reasonable range of realistic correlation forms.

- **Repeating rectangles:** This comes from something for which the error correlation coefficient is constant within a small range (1 pixel or a range of pixels), then repeats on a regular cycle. It could come from a push-broom sensor where every *n*th scan cycles are from a common detector element, or from a seasonal effect that occurs annually.
- **Repeating bell-shapes:** This is another repeating effect, but one where locally there is a drop off of correlation (partially correlated with neighbouring pixels/scan cycles) which then has a repeating effect.
- **Stepped triangle absolute:** This accounts for the situation where there is a calibration cycle, so that the instrument measures the calibration target once every few scan cycles (in HIRS, this was once every 48 scanlines [ref]), and then there is a rolling average between scan cycles. The correlation to neighbouring scan cycles takes the form of a stepped triangle.
- **Exponential Decay:** This accounts for the situation where the error correlation drops off exponentially across the dimension. This was introduced for propagation of uncertainties from the FCDR to the CDR.
- **Provided per pixel:** This can be used where detailed correlation information can be provided, and introduced for the propagation of uncertainties from the FCDR to the CDR.
- **Other:** Although true correlation structures may be more complicated than the ones given above, the above are sufficiently representative for the correlation structures encountered thus far. However, there may be situations where an FCDR producer needs to define a new error correlation form.

Depending on the type of error correlation form, different information is required. This is listed in Table 4.

Table 4 Parameters defined for different correlation forms

Correlation form	Parameters	Description
random	none required	For fully random effects there is no correlation with any other pixel
rectangle_absolute	**[-a,+b]** (rectangle limits). Provide these per pixel/scan cycle/orbit as	An effect is systematic within a range and different outside that range. For

	required. Allow for a way of representing **[-∞,+∞]** **[rmax]** States correlation coefficient for all pixel / scan cycle / orbit pixels. Default is rmax = 1 (fully correlated)	each pixel / scan cycle / orbit in range say number of pixels / etc either side that it shares a correlation with. For fully systematic effects notation to say "systematic with all". If rmax is defined, then the correlation coefficient is one for the pixel with itself, and is rmax with all other pixels.
triangle_relative	**[n]** – number of pixels/scan cycles being averaged in simple rolling average (should be an odd number)	Suitable for rolling averages over a window from (–n-1)/2 to (+n-1)/2 (i.e. for n pixels/scan cycles being averaged) Assumes a simple mean, not a weighted mean. No rmax is needed, since it is always 1.
bell_shaped_relative	[n] – number of pixels being averaged in a weighted rolling average, from which truncation range and standard deviation for Gaussian representation follow (truncation beyond ±n pixels, $\sigma = \frac{n/2-1}{\sqrt{3}}$) (n should be odd) OR [n,sigma] n: truncation from –n to +n, sigma: width of Gaussian representation (n should be odd) Typically provided once per orbit file (some further consideration needed about first/last scan cycles in an orbit)	Suitable for rolling averages over a window from (–n-1)/2 to (+n-1)/2 (i.e. for n pixels/scan cycles being averaged). Assumes a weighted mean, for any weights (and thus also includes things like spline fitting). Also suitable for anything else where the assumption is that "closer pixels/scan cycles are more correlated than further pixels". This can use two terms – n gives the truncation range outside which the assumption is there is no (or negligible) correlation, and sigma gives how fast the correlation drops off.
repeating_rectangles	[-a,+b,rmax,L,h,imax] per pixel/scan cycle/orbit etc (rmax,L,h will be same for different pixels)	Correlation coefficient assumed to be rmax for pixels/scan cycles from –a to +b, and h for pixels/scan cycles from L-a to L+b and from 2L-a to 2L+b and so on (iL-a to iL+b) for all integers i up to imax.
repeating_bell-shapes	[n,sigma,L,h, imax]	Correlation coefficient assumed to drop off as a truncated Gaussian for local pixels/scan cycles etc in the range defined by n and a similar Gaussian with a peak of h and the same width for pixels/scan cycles iL pixels apart on either side, for all integers I up to imax.

Stepped_triangle_absolute	[-a,+b,n] per pixel/scan cycle/orbit etc (n will be same for different pixels)	The step is a rectangular absolute from –a to +b with a correlation coefficient of one, after which the correlation coefficients drops for another a+b+1 lines, and then again. n is the number of calibration windows averaged.
Exponential_decay	[el,unit]	el: Length scale of exponential decay. Unit: unit of that length scale (we need to think about how we make this machine-interpretable! – could be time or space depending on which row you're considering]
Provided_by_pixel	[vector of relative correlation]	
Other	A function describing the correlation	Not yet implemented.

6.4 Evaluating error correlation scales

Error correlation scales can be evaluated by one of the following methods:

1. In some cases, the error correlation scale is obvious from the mechanism of how the instrument takes a measurement and where the uncertainty comes from. For example, an onboard calibration that takes place every 50 scan cycles, will have an (unknown) error that is common for all measurements within those 50 scan cycles. If a fresh calibration is used for the next 50 scan cycles, then the error correlation form would be a rectangular absolute form, if a rolling average is used then it would be a stepped triangle. Considering the onboard processes can directly identify the error correlation scales in these cases.
2. There are certain situations, where the error correlation can be evaluated using statistical analysis. This method must be applied carefully to avoid confusing value correlation with error correlation and requires an independent estimate of the error. In the FIDUCEO project, this was done for HIRS, where the noise in different spectral channels was found to be correlated by comparing individual observations of the stable onboard calibration source with an average observation and statistically determining the error correlation structure[19]. While with MVIRI the error correlation structures in the longitude and latitude errors as a function of distance were established using ground reference points.
3. For other processes, particularly those that involve reanalyses, it is not so obvious what the error correlation scale and form is. Where a reanalysis model is used to determine a parameter that is used in a retrieval, for example, then it is necessary to consider whether the error correlation scale is that of the grid scale of the retrieval or of the underlying physical phenomenon. If, for example, a relevant meteorological condition has a very large typical scale (spatially or temporally) compared to the grid scale, then the error (difference between modelled condition and 'true' condition) is likely to be correlated within a grid cell. On the other hand, if the phenomenon changes within a grid cell, then the error-correlation length scale is likely to be correlated only over the real range of the phenomenon as different points within the cell will have the same model, but different "true values" and therefore different errors (differences between those). So, for the FIDUCEO CDR on aerosols, we considered the

[19] Holl, G.; Mittaz, J.P.D.; Merchant, C.J. Error Correlations in High-Resolution Infrared Radiation Sounder (HIRS) Radiances. *Remote Sens.* **2019**, *11*, 1337. **DOI: 10.3390/rs11111337**

aerosol type error to be correlated within a reanalysis grid (1° longitude/latitude), but for a temporal scale we considered the monthly-gridding to be longer than the typical duration of a steady aerosol type, so temporarily we considered the error-correlation scale to be one week.

4. Finally, there are some error correlation length scales that are extremely difficult to quantify. In these cases, it is necessary to rely on "expert estimation" and record this in the "maturity of correlation scale" cell in the Effects Table.

6.4.1 Evaluating spectral error correlation

This applies to radiometric sensors with multiple spectral channels or bands

Most commonly, the error correlation between different spectral bands due to a particular uncertainty effect can be evaluated from an understanding of the processes on board the sensor. For example, most commonly, the instrument noise is fully independent between spectral channels[20], while the properties of an onboard calibration source may be fully correlated between spectral channels.

For a thermal infrared or microwave radiometer that calibrates all spectral channels against an onboard blackbody, then the uncertainty in the temperature of that blackbody leads to an error that is common for all channels. The uncertainty is not, however, the same for all spectral channels as the effect of a particular temperature uncertainty differs with channel wavelength..

In FIDUCEO we accounted for this using an analysis method which was called the "CURUC" method. In the effects table we gave the inter-channel correlation matrix. This is a square matrix, denoted $\boldsymbol{R}$, with the number of rows/columns equal to the number of channels. The cells give the error correlation (the extent to which the original error is common) between channels. In almost all cases this would be a matrix of zeros and ones. If for example, the instrument had three channels, and all were calibrated against the same blackbody whose temperature was determined once per calibration, then the error correlation matrix would look like this:

$$\boldsymbol{R} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

The uncertainty matrix, is a square matrix with the off-diagonals equal to 0 and the on-diagonals equal to the uncertainty associated with this effect in units of the effect, so, e.g.

$$\boldsymbol{U} = \begin{bmatrix} u(T) & 0 & 0 \\ 0 & u(T) & 0 \\ 0 & 0 & u(T) \end{bmatrix}.$$

Finally, the sensitivity coefficient matrix is also a diagonal matrix, with the sensitivity coefficients along the diagonal, thus

$$\boldsymbol{C} = \begin{bmatrix} \frac{\partial y_1}{\partial T} & 0 & 0 \\ 0 & \frac{\partial y_2}{\partial T} & 0 \\ 0 & 0 & \frac{\partial y_3}{\partial T} \end{bmatrix}.$$

[20] An exception was for the HIRS instrument, as described above.

The error covariance matrix, due to this effect is then given as:

$$\boldsymbol{S}_T = \boldsymbol{CURU}^{\mathsf{T}}\boldsymbol{C}^{\mathsf{T}}.$$

For spectrometer sensors, the spectral correlations will be considerably more complex and assessed as part of the overall sensor performance.

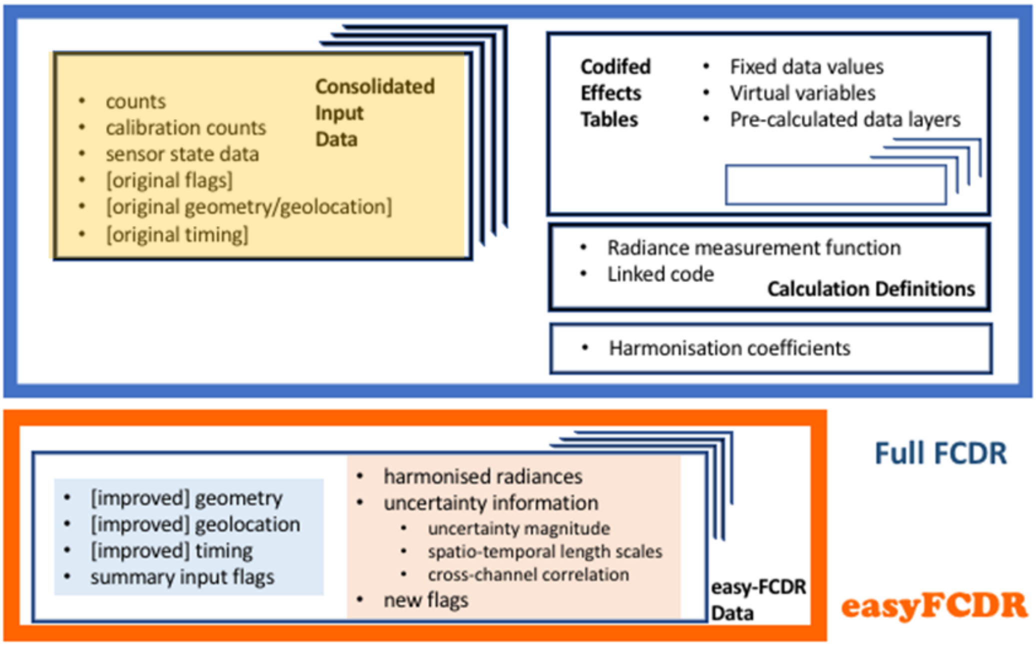
7 Full and Easy FDRs (and equivalent)

The effects tables, when completed should be used to propagate uncertainties between the FDR and the CDR/TDP and from the CDR/TDP to user applications, including stored information about the traceability and origins of the data record, which is important information for future researchers reviewing this work in later decades.

The exact format of the digitally stored will be project specific with the specific needs identified via the uncertainty tree and process chain analysis. Two levels of FDR are specified.

The 'Full FDR' provides a fully traceable data set, which includes all the information needed to perform a fully-robust uncertainty propagation from level 1 to higher level products, along with all necessary ancillary data to generate the FDR from raw data. The Full FCDR may also provide all the raw Level 0 data needed to generate the FCDR, it practical although different projects will start from different starting datasets. What is important, is that this is clearly stated and communicated in the FDR.

The 'Easy FCDR' is a compromise between simplicity and robustness which provides level 1 data users with sufficient radiance uncertainty information to propagate uncertainty to higher-order geophysical products with adequate rigour (or to use the radiance in data assimilation with knowledge of the radiance observation error covariances).



The Easy FDR provides the FDR data product together with uncertainty information at pixel-level that fits into three categories:

- Uncertainties associated with independent effects
- Uncertainties associated with common effects
- Uncertainties associated with structured effects

The selection of the categories and how they are reported should be driven by the user need and typical data use in terms of temporal, spatial and spectral averaging & processing. The data should be sufficiently detailed to allow a user to calculate an uncertainty relevant to the scales used.

Obtaining the Easy FCDR from the Full FCDR requires the classification of each source of uncertainty in the Full FCDR into one of the three Easy FCDR uncertainty classifications. Uncertainties associated with effects with similar error correlation structures are combined using the Law of Propagation of Uncertainties.

To obtain the error correlation structure for the structured effects and to determine the spectral channel error correlation structure, it is necessary to combine effects that may have different individual error correlation structures. This was done using an adaption of the CURUC technique[21]

8 Improving the measurement function

Uncertainty analysis presumes that the result of a measurement has been corrected for all recognized systematic effects, and that every effort has been made to identify such effects (GUM 2008 Section 3.2.4). This means that the measurand will be as accurate as possible given the current state of knowledge. It is a common experience in doing FDR and CDR/TDP analyses that as uncertainties are investigated, effects are identified that could be fully or partially corrected for. Where this is identified, the identified correction should be applied, leaving a smaller residual uncertainty after correction. Performing uncertainty analysis therefore often has the co-benefit of improving EO data and uncertainty analysis is likely to be an iterative process.

9 Harmonisation

9.1 Principles of Harmonisation

Harmonisation is the process of recalibrating a satellite sensor. This analysis is typically based on match-up information with another sensor, or series of sensors. Ideally, it takes in “match-ups” – moments where two instruments saw (almost) the same location at (almost) the same time and uses these to determine new harmonisation coefficients for the sensor. However, where direct match-ups are not possible, alternative methods should be employed. The use of the resultant dataset needs to be considered in designing the harmonisation process to ensure the drivers for the harmonisation are met.

These harmonisation coefficients generally refer to physical attributes of an instrument that can be better estimated from the harmonisation process than from available pre-flight or on-board calibration information and harmonisation can be considered a recalibration of the sensor.

The harmonisation coefficients can also enable re-calibration of sensors to account for underlying evolutions in sensor technology and changes in instrument design, operation and performance to allow a unified dataset to be produced.

Harmonisation attempts to remove biases between sensors in a series. It is, however, important to distinguish harmonisation from simple bias correction and from homogenisation. In bias correction such differences are removed through a simple scaling (addition of or multiplication by a parameter

[21] Merchant, C.J.; Holl, G.; Mittaz, J.P.D.; Woolliams, E.R. Radiance Uncertainty Characterisation to Facilitate Climate Data Record Creation. *Remote Sens.* **2019**, *11*, 474, DOI: 10.3390/rs11050474.

that brings data sets closer together), while harmonisation reconsiders the measurement function from first principles and recalibrates the sensor's calibration coefficients based on match up data. No two sensors have identical responses and we would not necessarily expect them to give the same value. For example, for radiometric sensors, two sensors may have different spectral response functions. In homogenisation, such differences are removed, and one sensor is made to look like the other sensor. In harmonisation, such differences are respected, and the unexplained biases are removed. See §10A.3.2

9.2 Practicalities of Harmonisation

Harmonisation takes the data from the periods during which sensors operated concurrently and finding instances, known as match-ups, where two sensors simultaneously observe the same location on the Earth with compatible viewing geometry (within a given tolerance). In addition, match-ups are also found to modern reference sensors which have calibrations considered to be high quality. The recalibration then involves calibrating the whole sensor series to the reference sensor. This recalibration is a large non-linear regression problem, solving for new calibration parameters in the measurement equation of each sensor based on the information provided by the match-ups.

Ordinary least squares (LSQ) is a commonly used approach for regression, however, it can only consider uncertainties in the derived quantity (simplistically – the y-axis) and further it treats all the observations as having independent errors. For the harmonisation approach here, we want to be able to respect the uncertainties associated with all measured values and the error correlation between match-ups. Aside from these philosophical requirements, in practice, LSQ solutions have been found to cause biases which will affect the long-term stability of the series, and therefore the ability to determine a climate trend, whereas a more robust error-in-variables regression models (EIV), which can consider uncertainties associated with all variables including the 'x-axis' perform much better.

A simple set of simulations can illustrate the fundamental problem. We have taken a simple straight line equation ($Y = A + B \times X$) with fixed values $A = 0.0$, $B = 1.0$ in the range $0.0 < X < 1.0$ and have generated X,Y pairs where noise of 0.05 has been added to both X and Y values. We have then fitted a straight line to the data where

a. Only the uncertainty associated with Y has been included in the fitting process (LSQ)
b. Uncertainties associated with both X and Y have been included explicitly with Orthogonal Distance Regression (ODR)

The first two columns of Figure 1 show the distribution of the deviation of the fitted parameters (denoted as p[0] and p[1]) from the true values (A, B) as a function of the estimated uncertainty associated with p[0] and p[1] based on the solution's covariance matrix. Also shown in red are the predicted normal distributions for a statistically consistent set of values relative to the truth. Figure 1 makes it clear that LSQ is not capable of returning the correct value of A and B whereas the ODR solution is completely consistent within the estimated uncertainty. The right hand set of plots show the deviation of the estimated Y value from the fits from the true Y value for an X of 2.0. Again the LSQ fits are biased whereas the ODR values are not.

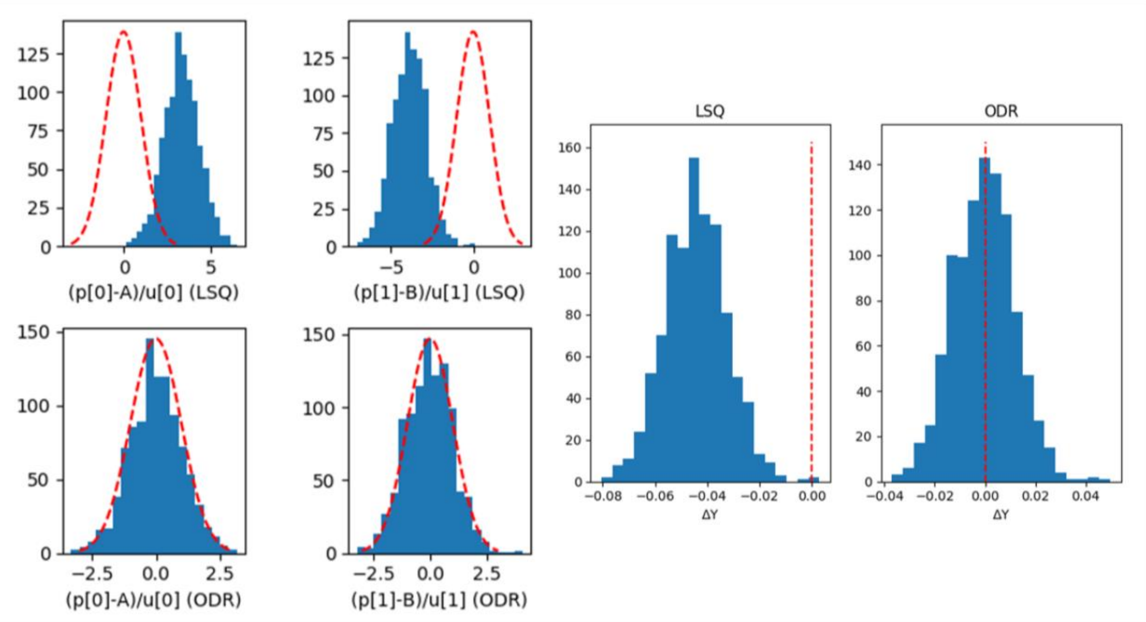


Figure 4 Left hand plots show the error in the fitted parameters for a straight-line model from their true values in terms of their deviation relative to their estimated uncertainty. The red curve shows a normal distribution centred at zero which is the expected distribution for a correct fit. The ODR example is completely statistically consistent with the true values, whereas the LSQ case is biased. The right hand pair of plots shows the error in Y from the fitting process and again shows the bias in the LSQ method. The diagram and example text are taken from http://www.fiduceo.eu/content/beyond-least-squares-analysis-regression-considering-correlation

This is not, however, the end of the story as existing EIV implementations, such as ODR, still do not capture the error correlation structure between the data for the match-ups and so cannot provide an optimal solution. In the FIDUCEO project novel methods for a rigorous, metrological solution to the EIV regression which fully respects the match-up error correlation structure were developed[22][23].

The implementation of this methodology is challenging due to the possibly complex error and geophysical correlation structures and high data volume.

9.2.1 Generating match up datasets

We define a matchup as a "point" measurement that is matched by another "point" measurement sufficiently close in space and time – excluding the trivial case of neighbouring measurements from the same acquisition, of course. In other words, we required matchup pixels that covered the same place on earth acquired at almost the same time. The smaller the time-delta aimed at becomes, the

[22] S. E. Hunt et al., "A Metrological Approach to Producing Harmonised Fundamental Climate Data Records from Long-Term Sensor Series Data," IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, 2018, pp. 3397-3400. Provided with this guidance document

[23] Giering, R.; Quast, R.; Mittaz, J.P.D.; Hunt, S.E.; Harris, P.M.; Woolliams, E.R.; Merchant, C.J. A Novel Framework to Harmonise Satellite Data Series for Climate Applications. *Remote Sens.* **2019**, *11*, 1002.

more challenging the detection algorithms, especially when getting below the acquisition duration of a satellite data product (which is in the range of 45 minutes for polar-orbiting instruments).

In most cases, we extended this definition to also include the neighbouring pixels of a sensor acquisition, covering a symmetrical window of n by m pixels around the matchup point. This allowed us to perform some calculations to get a feeling for the data (e.g. check for homogeneity, cloud shadow detection, etc.).

Understanding the conditions under which the match-ups take place is important, so occasions where ancillary complementary information is available (in-situ monitoring site data, for instance.)

In previous projects a variable software framework was developed to initial provide coarse match-up database using sensor meta-data refined with a fine-grained analysis to better determine the exact observational co-incidence (primarily in time). This can be made efficient by using the sensor orbital data, rather than interrogating the data product files themselves.

Once the matchup locations and times – and the associated data files are identified, the specific product data can be extracted to perform checks on additional constraints on the data, like cloud detection or checks for data homogeneity or ground surface constraints.

The algorithm can then only store the needed for later scientific analysis, packaged as temporal segments (typically, one file per month.)

9.2.2 File formats

It is important to ensure that the match up process recognises the error correlation between the radiance observations in different match ups. The file format definition provides a common input file format for all the project harmonisation processes and also shows how error correlation structures can be built into the input files.

The FIDUCEO file formats can be shared as a starting point example.

10 Summary and conclusions

In this report, we have described the approach to take towards the generation of an FDR or CDR/TDP. The methods described here have been generalised with respect to previous project, specifically FIDUCEO and its preceding, sister and follow on projects. The report has described the concepts behind diagrammatic representations of the traceability, such as the uncertainty tree diagram and processing chain.

Such diagrams are helpful to explore the origin of all data being used within an FDR and CDR/TDP, and also provide a means to identify the different effects – that is, the different sources of uncertainty – that influence the measurement result. The methodology presented here describes how an “effects table” can be used to document the magnitude of the uncertainty, its sensitivity coefficient (that is how an uncertainty in this effect affects the measurand), and to document error correlation structures due to this effect.

By its nature, this report has been very generic in style and the only provided examples are from related projects. During the FDR4ALT and FDR4ATMOS projects we will apply these methods to a greater variety of sensors working with new teams. It is expected that these guidelines will need updating using the developments made within the current projects.

Below follow appendices on some of the core concepts behind uncertainty analysis.

A Appendix on metrological concepts

A.1 Measurement Function

The measurement function is defined from the 'measurement model', which establishes the mathematical relations between the measurand (Y) and the input quantities (X_i). Here, we use the word 'function' in the most general sense. Often, we can explicitly write the measurement model in terms of an analytic expression of quantities:

$$Y = f(X_1, X_2 \dots; \boldsymbol{A}) + \Delta \qquad \textbf{Eq A1-1}$$

where:

- Y represents the output quantity (the measurand).
- X_1, X_2 ... are the input quantities, for example, the counts.
- $\boldsymbol{A}$ is the vector of calibration parameters (which are also input quantities but are usefully distinguished).
- Δ is an input quantity introduced to represent any inadequacy of the function f to represent all physical phenomena that affect the measurand.

Sometimes it is necessary to define the measurement function in a different way, for example as the iterative solution of a measurement model through code. In this case, we can still think of a measurement function as a function of certain input quantities, even though we cannot explicitly write the function out.

The equations used to populate e.g. L1 products, evaluate the measurand using estimates of the input quantities. In the GUM[24], the convention is for estimates to be represented with the lower-case characters corresponding to the quantities written in upper case. The measured output value is therefore determined from the expression:

$$y = f(x_1, x_2, \dots, \boldsymbol{a}) + \delta \qquad \textbf{Eq A1-2}$$

where the input estimates include the recorded sensor counts and values of calibration coefficients, etc. Since 0 is our best estimate of δ, which is the expectation of Δ (assuming we are using the best measurement model we can formulate), then we may practically write this as

$$y = f(x_1, x_2, \dots, \boldsymbol{a}) + 0. \qquad \textbf{Eq A1-3}$$

Here, this 'plus zero' term does not alter the value of the measurand, but will have an associated uncertainty in recognition of the fact that all measurement functions are approximations to the physical process they describe. In other words, this term considers the extent to which the equality of the measurement function may not hold. For example:

- If the measurement function is a linear equation, the 'plus zero' term considers the extent to which the instrument may be non-linear.
- If the measurement function is a spectral integral determined numerically using a trapezium or rectangular rule, the 'plus zero' term considers the extent to which this rule acts as an approximation of the integrated quantity.

[24] Guide to the Expression of Uncertainty in Measurement (see uncertainty section, below).

- If there is an assumption that quantities or effects cancel each other out, the 'plus zero' term considers the uncertainty in the extent to which they cancel

Once we have a clear picture of the extent to which the measurement function describes the true physical state of the measurement process and the effects that influence each input quantity, we can determine the uncertainty in the measurand through the process of uncertainty analysis. Further material can be found in tutorials[25] publications[26] or the measurement model in the GUM[27]

A.2 Uncertainty and errors

A.2.1 The Guide to the Expression of Uncertainty in Measurement (GUM)

The *Guide to the Expression of Uncertainty in Measurement*, known as 'the GUM', provides guidance on how to determine, combine and express uncertainty. It was developed by the JCGM (Joint Committee for Guides in Metrology), a joint committee of all the relevant standards organisations (e.g. ISO) and the BIPM (*Bureau International des Poids et Mesures*). This heritage gives the GUM authority and recognition. The JCGM continues to develop the GUM and has recently produced a number of supplements which cover topics such as Monte Carlo Methods for uncertainty analysis and an extension of the Law of Propagation of Uncertainties to multiple output quantities. All of these supplements, as well as the 'VIM' (International Vocabulary of Metrology) are freely downloadable from the BIPM website.

Further material:

- The GUM
- The VIM
- Introductory and advanced training courses on uncertainty analysis are available on the NPL website. Note that the "explained" series and the Earth Observation courses are free, while the standard uncertainty courses are charged.

A.2.2 Errors and Uncertainties

The terms 'error' and 'uncertainty' are not synonyms, although they are often confused. To understand the distinction, consider the result of a measurement – the measured value. The value will differ from the true value for several reasons, some of which we may know about. In these cases, we apply a **correction**. A correction is applied to a measured value to account for known differences, for example the measured value may be multiplied by a gain determined during the instrument's calibration, or a measured optical signal may have a dark reading subtracted. This correction will never be perfectly known and there will also be other effects that cannot be corrected, so after correction there will always be a residual, unknown **error** – an unknown difference between the measured value and the (unknown) true value.

The specific error in the result of a particular measurement cannot be known, but we describe it as a draw from a probability distribution function. The **uncertainty** associated with the measured value is a measure of that probability distribution function; in particular, the **standard uncertainty** is the standard deviation of the probability distribution.

[25] https://research.reading.ac.uk/fiduceo/archive/tutorials/
[26] https://doi.org/10.1088/1681-7575/ab1705
[27] https://www.bipm.org/en/publications/guides/gum.html

There are generally several 'sources of uncertainty' that jointly contribute to the uncertainty associated with the measured value. These will include uncertainties associated with the way the measurement is set up, the values indicated by instruments, and residual uncertainties associated with corrections applied. The final (unknown) error on the measured value can be considered to be drawn from the overall probability distribution described by the uncertainty associated with the measured value. This is built up from the probability distributions associated with all the different sources of uncertainty.

(Strictly the GUM describes uncertainty as the dispersion of values that could reasonably be attributed to the measurement. In this way it is not strictly from the probability distribution of errors around a true value, but a distribution of possible values around the measured value. Thus, the explanation above is an oversimplification that is not strictly GUM compliant. It can, however, clarify the distinction between "error" and "uncertainty"). See §10B.2

A.2.3 Error effects

Each input quantity in the measurement function (whether written formally as in $Y = f(X_1, X_2 \dots; \boldsymbol{A}) + \Delta$, or as it is calculated, as in $y = f(x_1, x_2, \dots, \boldsymbol{a}) + 0$), has associated with it at least one source of error. This includes measured counts, calibration parameters (for Level 1) or retrieval parameters (for level 2) and the "plus zero" or Δ term.

It is important to identify and consider each of these error effects in turn. Every error effect is a source of uncertainty and we should quantify the size of that uncertainty, and understand the sensitivity coefficient (the 'translation' between an error in the input quantity and the error in the output quantity). For example, if an error effect affects the input quantity X_1 and has an uncertainty $u(X_1)$, then the uncertainty associated with Y due to this effect is given by $\frac{\partial f}{\partial X_1} u(X_1)$, where $\frac{\partial f}{\partial X_1}$is the sensitivity coefficient.

We should also consider the error correlation introduced by the effect by considering the influence of a single error in that input quantity on spatial and temporal scales.

A.2.4 Error correlation

A.2.4.1 Independent, structured and common errors, correlation

The measured value in each pixel of an EO image is the result of a sequence of steps and transformations. In transforming from raw data (L0) to calibrated radiances (L1), many measured values relevant to calibration measurements may be combined. In transforming L1 radiances to L2 geophysical products radiances from different spectral bands may be used. In L2 to L3+ processing, data across different pixels are then combined. (This description is biased towards radiometric sensors; similar principles apply in other sensor applications). Where correlation exists between errors in different measured values (different wavelengths and/or pixels), this error correlation needs to be considered in the uncertainty analysis.

The GUM defines systematic and random errors, concepts that are widely used in science.

Random errors are errors manifesting independence: the error in one instance is in no way predictable from knowledge of the error in another instance. A complication arises in EO imagery when one instance of a parameter in the radiance measurement function is used in the calculation of the Earth radiance across many pixels. That component of the error in the radiance image is then correlated across pixels, even though the originating effect is random. Put another way, the originating random error contributes errors with a particular structure to the image.

Systematic errors are those that could in principle be corrected for if we had sufficient information to do so: that is, they arise from unknowns that could in principle be estimated rather than from chance processes. All systematic errors in EO are structured in that there is a pattern of influence on multiple data. They include, but are not limited to, effects that are constant for a significant proportion of a satellite mission—i.e., biases, for which the structure is a simple error in common.

When considering EO imagery, it is can be useful to categorise effects primarily according to their cross-pixel error correlation properties, as independent, structured or common effects.

A.2.4.2 Independent errors

Independent errors arise from random effects causing errors that manifest independence between pixels, such that the error in $L(l', e')$ is in no way predictable from knowledge of the error in $L(l, e)$, were that knowledge available. Independent errors therefore arise from random effects operating on a pixel level, the classic example being detector noise.

A.2.4.3 Structured errors

Structured errors arise from effects that influence more than one measured value in the satellite image, but are not in common across the whole image. The originating effect may be random or systematic (and acting on a subset of pixels), but in either case the resulting errors are not independent, and may even be perfectly correlated across the affected pixels. Since the sensitivity of different pixels/channels to the originating effect may differ, even if there is perfect error correlation, the error (and associated uncertainty) in the measured radiance can differ in magnitude. Structured errors are therefore complex, and, at the same time, important to understand, because their error correlation properties affect how uncertainty propagates to higher-level data.

A.2.4.4 Common errors

Common errors are constant (or nearly so) across the satellite image, and may be shared across the measured radiances for a significant proportion of a satellite mission. Common errors might typically be referred to as biases in the measured radiances. Effects such as the progressive degradation of a sensor operating in space mean that such biases may slowly change.

A.3 Metrological comparisons

A.3.1 How metrologists use comparisons

National Metrology Institutes (NMIs) have always used comparisons for scientific purposes, to test their methods and especially their uncertainty analysis. In the early stages of research these scientific comparisons show up the unknown unknowns – the differences between participants that are not (yet) considered in the uncertainty analysis. At this point, comparisons tend to be informal and performed, for example, through participants visiting each other's facilities. As a field matures and the technical approaches move from research to operational services, comparisons show increasing agreement between participants. At this point the role of comparisons changes from research into auditing and peer review.

This second purpose was formalised in 1999 by the signing of the Mutual Recognition Arrangement[28] (MRA) by the world's NMIs. The MRA says that 'within an appropriate degree of equivalence' the results of one NMI can be considered equivalent to the results of another NMI. In practice this enables

[28] http://www.bipm.org/en/cipm-mra/

world trade and the use of artefacts and instruments calibrated in another country. Being a legal process, the MRA relies on NMIs regularly reviewing each other's calibration and measurement capabilities through a combination of formal peer review and auditing and through formal 'key comparisons' that compare the measurement capability of laboratories – both at the international level (by a handful of laboratories with, generally, the lowest uncertainties) and at the regional level (e.g. within Europe or within Asia-Pacific).

The formal key comparisons are run with strict guidelines and are always blind comparisons (only one 'pilot' laboratory has access to the results before they are published). There is ongoing discussion about the best ways of analysing such comparisons, and in particular about the choice of the Key Comparison Reference Value (KCRV) against which all participants are compared. In very mature fields, where the differences between the measured values of the different participants and the KCRV are consistent with uncertainties, the most common KCRV is the weighted mean of the results of the different NMIs. In fields where there is more spread, this may not be the appropriate choice and alternatives (including 'weighted mean with cut-off' which limits the weight assigned to the laboratories with the lowest uncertainties, or simply using a median value) are considered.

It is important to note that for metrologists the purpose of comparisons is to test and validate uncertainty claims. Comparisons are not performed to estimate uncertainties.

A.3.2 E_N ratio

Ideally a bilateral comparison is made between two independent observations, each with full uncertainty analysis, to calculate the equivalence ratio:

$$E_N = \frac{|\rho_1 - \rho_2|}{k\sqrt{u_1^2 + u_2^2 + u_{\mathrm{comp}}^2}} \qquad \textbf{Eq A3-2}$$

where ρ_1 and ρ_2 are the two independent measured values and u_1 and u_2 are the two standard uncertainties associated with those measured values. u_{comp} is the standard uncertainty associated with the comparison itself (e.g. from a known difference between the observation conditions – matchup uncertainties) and k is the coverage factor for the appropriate confidence interval (usually $k = 2$ to have a confidence interval of 95 %).

An equivalence ratio $E_N < 1$ suggests that the two measured values agree within their uncertainties, while a larger E_N ratio suggests that at least one uncertainty is underestimated.

For a multilateral comparison (with multiple independent observations, e.g. by several participants) then it is possible either to do every pair of bilateral comparison and to present data in a table showing whether the E_N ratio is greater or less than 1 for each pair of observations, or to determine a comparison reference and calculate the E_N ratio for each participant with respect to the reference, using:

$$E_{N,i} = \frac{\rho_i - \rho_{\mathrm{ref}}}{k\sqrt{u_i^2 + u_{\mathrm{ref}}^2 + u_{\mathrm{comp}}^2}} \qquad \textbf{Eq A3-2}$$

Note that while the reference can be arbitrarily chosen, some care must be taken in choosing it as it is easy to interpret an $E_N > 1$ as implying "bad" data, which may have political and potentially financial repercussions. Within the metrology community a weighted mean of all values may be chosen as a reference.

This principle has been developed in EO with the U_{comp} term descriptive of a representation & co-location error when comparing sensor datasets.

B Appendix on Nomenclature

B.1 Appendix on notation

As with all disciplines, vocabulary and notation have very specific meanings in metrology. When we work in multidisciplinary teams, issues of vocabulary and notation may seem, originally to be a barrier, but, with care, can open to a greater mutual understanding.

These rules are based on the ISO 80000 standard series.

- In equations, italics represent a variable/quantity. Upright text is used for labels, e.g. L_{Earth} for radiance of the Earth, or λ_i for the ith wavelength
- Write spaces before and between units, use a non-breaking space to stop the units being on a different line to the measured value. E.g. 300 K and 3 mW m^{-2} sr^{-1} nm^{-1}. Avoid using the / in units, unless it is a very simple expression, e.g. m / s.
- A space should be used between a number and the % sign, e.g. 2.5 %. A space is not used for the angular degrees, e.g. 45°, but is for temperatures, e.g. 30 °C.
- Units are always written with a small letter when written out in full (e.g. 300 kelvin) except for "degrees Celsius"
- Unit symbols are always in upright type.

Further material:

- ISO 80000 on units and mathematical notation
- NIST guide on notation rules
- FIDUCEO project notation guide

B.2 Terminology glossary

In the 'glossary' below, a few important words are explained, taken from [5]. Precise or rigorous definitions are not given here. They can be found elsewhere, for example in the *International Vocabulary of Basic and General Terms in Metrology*. A useful and correct set of definitions can also be found in UKAS publication M 3003 *The Expression of Uncertainty and Confidence in Measurement*

accuracy - closeness of the agreement between a measurement result and true value of that measurand. (Accuracy is a qualitative concept only and is not given a numerical quantity value. It is often misused as uncertainty or precision.)

bias (of a measurement) – estimate of a systematic measurement error

bias (of a measuring instrument) - systematic error of the indication of a measuring instrument

calibration - operation that, under specified conditions, in a first step, establishes a relation between the quantity values with measurement uncertainties provided by measurement standards and corresponding indications with associated measurement uncertainties and, in a second step, uses this information to establish a relation for obtaining a measurement result from an indication. In other words, the comparison of an instrument against a reference or standard, to find any errors in the values indicated by the instrument. In some cases, calibration assigns a relationship between the input and output of an instrument; for example, calibration of a resistance thermometer could relate its output (in ohms) to an input temperature (in degrees Celsius, or in kelvins).

confidence level - number (e.g. 95 %) expressing the degree of confidence in a result

correction (calibration correction) - compensation for an estimated systematic effect. A number added to an instrument reading to correct for an error, offset, or bias. (Similarly, a reading may be multiplied or divided by a ***correction factor*** to correct the value.)

correlation - interdependence, or relationship, between data or measured quantities

coverage factor - number larger than one by which a combined standard measurement uncertainty is multiplied to obtain an expanded measurement uncertainty, for a particular level of confidence

error - measured quantity value minus a reference quantity value. The offset or deviation (either positive or negative) from the correct value

estimated standard deviation - estimate of the standard deviation of the 'population' based on a limited sample

expanded uncertainty - product of a combined standard measurement uncertainty and a factor larger than the number one. Standard uncertainty (or combined standard uncertainty) multiplied by a coverage factor *k,* to give a particular level of confidence

Gaussian distribution - (See ***normal distribution***)

influence quantity - quantity that, in a direct measurement, does not affect the quantity that is actually measured, but affects the relation between the indication and the measurement result; e.g., cloudiness in the field-of-view of an instrument can influence the accuracy of its measurement

interval (confidence interval) - interval containing the set of true quantity values

of a measurand with a stated probability, based on the information available. The margin within which the 'true value' being measured can be said to lie, with a given level of confidence

level of confidence - number (e.g. 95 %) expressing the degree of confidence in the result

mean (arithmetic mean) - average of a set of numbers

measurand - quantity intended to be measured. The particular quantity subject to measurement

normal distribution - distribution of values in a characteristic pattern of spread (Gaussian curve) with values more likely to fall near the mean than away from it

operator error - a mistake

precision - closeness of agreement between indications or measured quantity values obtained by replicate measurements on the same or similar objects under specified conditions. A term meaning 'fineness of discrimination' but often misused to mean 'accuracy' or 'uncertainty'. Its use should be avoided if possible.

random error - component of measurement error that in replicate measurements varies in an unpredictable manner. An error whose effects are observed to vary randomly.

range - absolute value of the difference between the extreme quantity values of a nominal indication. The interval difference between the highest and the lowest of a set of values

reading - value observed and recorded at the time of measurement

rectangular distribution - distribution of values with equal likelihood of falling anywhere within a range

repeatability (of an instrument or of measurement results) - condition of measurement, out of a set of conditions that includes the same measurement procedure, same operators, same measuring system, same operating conditions and same location, and replicate measurements on the same or similar objects over a short period of time. The closeness of the agreement between repeated measurements of the same property under the same conditions.

reproducibility (of an instrument or of measurement results) – condition of measurement, out of a set of conditions that includes different locations, operators, measuring systems, and replicate measurements on the same or similar objects. The closeness of the agreement between measurements of the same property carried out under changed conditions of measurement (e.g. by a different operator or a different method, or at a different time)

resolution - smallest change in a quantity being measured that causes a perceptible change in the corresponding indication. (e.g. a change of one (1) in the last place of a digital display)

result (of a measurement) - set of quantity values being attributed to a measurand together with any other available relevant information. The value obtained from a measurement, either before or after correction or averaging

sensitivity - quotient of the change in an indication of a measuring system and the corresponding change in a value of a quantity being measured. The change in response (of an instrument) divided by the corresponding change in the stimulus

standard deviation - a measure of the spread of a set of results, describing how values typically differ from the average of the set. Where it is not possible to obtain an infinite set of results (in practice it never is) we instead use the estimated standard deviation.

standard uncertainty - measurement uncertainty expressed as a standard deviation.

systematic error – component of measurement error that in replicate measurements remains constant or varies in a predictable manner. A bias or offset (either positive or negative) from the correct value

true value – quantity value consistent with the definition of a quantity, i.e. the value that would be obtained by a perfect measurement

Type A evaluation of uncertainty - evaluation of a component of measurement uncertainty by a statistical analysis of measured quantity values obtained under defined measurement conditions.

Type B evaluation of uncertainty - evaluation of a component of measurement uncertainty determined by means other than a Type A evaluation of measurement uncertainty

uncertainty budget - statement of a measurement uncertainty, of the components of that measurement uncertainty, and of their calculation and combination

uncertainty of measurement - non-negative parameter describing the dispersion of the quantity values being attributed to a measurand. Alternatively described as a quantity representing the doubt in result of a measurement.

uniform distribution - distribution of values with equal likelihood of falling anywhere within a range

validation - the process of assessing, by independent means, the quality of the data products derived from the system outputs

C Appendix on Example process chains

C.1 Worked example: GRUAN RS92 radiosonde temperature product

As part of GAIA-CLIM this approach[29] was used on the GCOS Reference Upper Air Network (GRUAN) which provides, among other data products, atmospheric temperature profiles from regularly launched Vaisala RS92 radiosondes. This design of radiosonde is widely used and measures temperature with a capacitive temperature sensor. While the RS92 is often used with the manufacturers data processing software GRUAN apply own data processing which includes bias corrections. The GRUAN data products include the associated measurement uncertainty, for the RS92 products the data processing and uncertainty calculation is described in the paper[30].

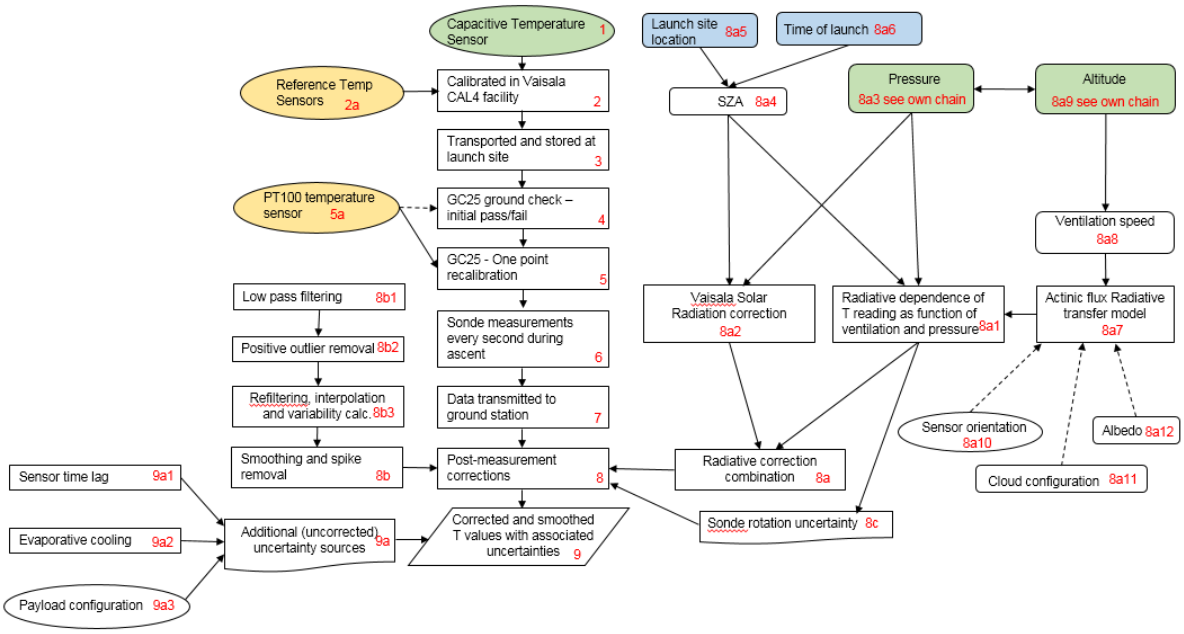


Figure 5 The traceability chain for the GRUAN RS92 temperature data product

This uncertainty calculation was used as the basis for GAIA-CLIM example documents using the method outlined above. First step was to identify the different elements that contribute to uncertainty. These elements include instrument properties, data processing methods and measurement conditions, the elements are then arranged these into a traceability chain. The traceability chain for the GRUAN RS92 temperature data product is shown in Figure 5. This includes a main chain starting at the initial measurement going to the final data product, and sub chains such as the GRUAN radiative heating correction. As you move up the chain the final data product the total uncertainty at each step includes the uncertainty of all steps below it in the chain and all sub-chains that attach to an element.

When the different elements have been identified and arranged in the traceability chain the elements are then described individually using the table described before and shown in Table 5.

[29] 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, doi:10.5194/amt-3-1217-2010, 2010. http://www.atmos-meas-tech.net/3/1217/2010/amt-3-1217-2010.html
[30] Dirksen, R. J., Sommer, M., Immler, F. J., Hurst, D. F., Kivi, R., and Vömel, H.: Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde, Atmos. Meas. Tech., 7, 4463-4490, doi:10.5194/amt-7-4463-2014, 2014. http://www.atmos-meas-tech.net/7/4463/2014/amt-7-4463-2014.pdf

Table 5 An example element table from the GRUAN RS92 temperature data product traceability chain

Information / data	Type / value / equation	Notes / description
Name of effect	Capacitive temperature sensor	
Contribution identifier	1, statistical uncertainty $u_{u(T)}$	
Measurement equation parameter(s) subject to effect	Temperature	
Contribution subject to effect (final product or sub-tree intermediate product)	Temperature	
Time correlation extent & form	None	Random over ascent
Other (non-time) correlation extent & form	None	Random over ascent
Uncertainty PDF shape	Normal	
Uncertainty & units	±0.5 K (2σ) [accuracy] & 0.1-0.15 K (1σ) in the trop., rising to 0.5 K (1σ) at 10 hPa [statistical unc.]	
Sensitivity coefficient	1	
Correlation(s) between affected parameters	None	
Element/step common for all sites/users?	Yes	
Traceable to …	Accuracy to 2	Calibration in Vaisala CAL4 facility
Validation	Inter-comparison studies.	

This is the statistical uncertainty of the capacitive sensor as determined by the manufacturer and the element table was filled out as follows:

Name of effect – The same as is used in the traceability chain.

Contribution identifier – includes the identifier used in the traceability chain as well as the identifier used in the uncertainty calculations described in Dirksen et al. 2014.

Measurement equation parameter(s) subject to effect – The measurement equation is for the temperature data product.

Contribution subject to effect – This element is part of the main chain this is the final product, this would be different for where the element is part of a sub chain. For example element 8a4, the solar zenith angle, has the Actinic flux radiative transfer model as its contribution subject to effect.

Time correlation extent & form – This element is a random statistical uncertainty and so has no correlation. Elements identified with some form of time correlation includes element 8a8, the ventilation speed, which was systematically correlated over an ascent.

Uncertainty PDF shape – This is a random statistical uncertainty which follows a normal distribution.

Uncertainty & units – the uncertainty and units here were determined from the uncertainties in the GRUAN data products. In some cases, the element descriptions were accompanied by figures showing how the uncertainty can vary over time or over an ascent.

Sensitivity coefficient – This uncertainty is included in the total uncertainty as part of a sum in quadrature. Sensitivity coefficients can be different from 1 when the uncertainty uses a different unit from the final data product. In some cases uncertainties were identified which were not used in the total uncertainty, in which case a sensitivity coefficient of 0 was used.

Correlation(s) between affected parameters – This uncertainty has no correlation with affected parameters. An example of an element which does have correlation is element 8a7, the actinic flux model, which is correlated with 8a8, the solar zenith angle.

Element/step common for all sites/users – This element is common to all users, as are most elements of the GRUAN RS92 temperature uncertainty. Elements which might not be the same for all users could include different ground check methods, which could vary by site.

Traceable to – in this case this is traceable to a referenced document:

> Vaisala: CAL4 Calibration machine Traceability and Uncertainty, Technical Document DOC210645, Vaisala DCO210645, 2002.

Validation – the referenced document described intercomparisosn studies carried out by the sensor manufacturer.

When the element tables have been completed, they were then summarised in a table describing the uncertainties value, structure, correlation and traceability level, as determined in Table 6.

An example of the uncertainty summary table is shown in Table 6, this is an excerpt from the uncertainty summary table for the GRUAN RS92 temperature product. For element 1 most of the entries are as shown in the element table. The traceability has been determined to be high because of the calibration method which is traceable to SI. Elements which did not have associated uncertainties were identified as were elements which had uncertainties which were not used in the total uncertainty calculation and feedback on possible improvements in the uncertainty calculation was provided to GRUAN.

Table 6 an excerpt from the GRUAN RS92 temperature data product uncertainty summary table

Element identifier	Contribution name	Uncertainty contribution form	Typical value	Traceability level (L/M/H)	random, structured random, quasi-systematic or systematic?	Correlated to? (Use element identifier)
1	Capacitive sensor	Accuracy statistical uncertainty $\sigma(T)/\sqrt{N'}$	±0.5 K (2σ) ±0.1 K (1σ) in the trop. 0.2-0.5 K (1σ) in strat.	H H	Systematic (over ascent) random	none
2	Vaisala CAL4 calibration repeatability	constant	±0.15K (2σ)	H	random	none
2a	Reference T sensor accuracy	constant	±0.5 K (2σ)	H	systematic	none
3	Transport & storage	constant	0 K	L	systematic	none
4	GC25 ground check pass/fail	Rectangular	0 K	M	systematic	2a
5	GC25 one point re-calibration	constant	±0.16 K (1σ)	H	systematic	2 & 5a
5a	GC25 PT100 T sensor accuracy	constant	±0.15 K (1σ)	H	systematic	none
6	Measurement time frame	N/A	0 K	H	random	none
7	Data transmitted to station	N/A	0 K	H	random	none
8	Post-measurment corrections	Primarily α ΔT (solar radiation correction)	±0.22 K (2σ) in trop. ±0.5 K (2σ) in strat.	M	quasi-systematic	none
8a	Radiation correction	constant	<0.36 K (2σ)	M	systematic	none
8a1	Solar radiation temperature model	constant	<0.2 K (2σ)	M	systematic	none
8a2	Vaisala radiation correction	constant	0 K	M	quasi-systematic	8a1
8a3	Pressure	$\Delta T(I_s, p, v) = a \cdot x^b$ with $x = \frac{I_s}{p \cdot v}$	<0.001 K (1σ) in the trop., rising to ±0.03 K (1σ) in the strat.	M	random	Pressure product & 8a10
8a4	Solar Zenith Angle	constant	0 K	M	Systematic (over ascent)	
8a5	Launch site location	constant	0 K	H	Systematic	
8a6	Time of launch	constant	0 K	H	Systematic (over ascent)	
8a7	Actrinic flux model	constant	0 K	M	quasi-systematic	none
8a8	Ventilation speed	constant	±0.01 K (2σ) in the trop. up to ±0.3 K (2σ) in the strat.	M	quasi-systematic	Altitude product
8a9	Altitude	constant	0 K	M	quasi-systematic	Altitude product
8a10	Sensor orientation	constant	0 K	M	systematic	8a1
8a11	Cloud configuration	constant	0 K	L	Systematic	
8a12	Albedo	$\Delta T \cdot u_c(I_a)/I_a$ where ΔT is the solar radiation correction term from 8a1 and $U_c(I_a)$ = [illegible]	<0.05K (2σ)	M	Systematic (over ascent)	none
8b	Smoothing & spike removal	constant	±0.05K (2σ)	M	quasi-systematic	2
8b1	Low pass filtering	constant	0 K	M	quasi-systematic	2
8b2	Positive outlier removal	constant	±0.05K (2σ)	M	quasi-systematic	2
8b3	Refiltering, interpolation & variability	constant	0 K	M	quasi-systematic	2
8c	Rotating sonde	$2 \cdot \Delta T/\sqrt{3}$ where ΔT is the solar radiation corr.	±0.1 K (2σ) in trop. ±0.4 K (2σ) in strat.	M	quasi-systematic	
9	Additional uncorrected sources	constant	<0.2K (2σ)	M	Systematic (over ascent)	8a5 & 8b
9a	Sensor time lag	constant	< 0.03K (2σ)	M	quasi-systematic	none
9b	Evaporative cooling	constant	<0.2K (2σ)	M	Systematic (over ascent)	none
9c	Payload configuration	constant	0 K	L	Systematic	8a5 & 8b