CORE-CLIMAX

Climate Data Record Assessment

Instruction Manual

Doc.No. : CC/EUM/MAN/13/002

Issue : v3

Date : 3 December 20132013

EUMETSAT

EUMETSAT Allee 1, D-64295 Darmstadt, Germany

Tel: +49 6151 807-7

Fax: +49 6151 807 555

http://www.eumetsat.int

Table of Contents

1 INTRODUCTION

1.1 Purpose and Scope

The purpose of this user manual is to provide instruction on how to use the CORE-CLIMAX System Maturity Matrix (SMM) to assess the maturity of Climate Data Records (CDR). In addition instruction is provided how to use the newly developed and still experimental Application Performance Matrix (APM) to assess the quality of a CDR for a specific application. The document describes the concept for both tools and gives specific instructions to fill the SMM with scores.

1.2 Document Structure

The document is structured in the following way:

Section 1: Introduction

Section 2: Overview

Section 3: Data Record Description Template

Section 4: System Maturity Matrix (SMM)

Section 5: Application Performance Matrix (APM)

Appendix A: CORE-CLIMAX Data Set Description

1.3 Reference Documents

[RD.1] Wilson, J., M. Dowell and A. Belward (2010): European capacity for monitoring and assimilating space based climate observations – Status and prospects. JRC Scientific and Technical Report, *EUR 24273 EN*, 46 pp., DOI: 10.2788/70393.

[RD.2] Dowell, M., P. Lecomte, R. Husband, J. Schulz, T. Mohr, Y. Tahara, R. Eckman, E. Lindstrom, C. Wooldridge, S. Hilding, J. J. Bates, B. Ryan, J. Lafeuille, and S. Bojinski (2013): Strategy towards and architecture for climate monitoring from space. 39 pp., [available from: www.ceos.org, www.wmo.int/sat, www.cgms-info.org].

[RD.3] Bates, J. J. and J. L. Privette, (2012), A maturity model for assessing the completeness of climate data records, Eos Trans. AGU, 93(44), 441.

[RD.4] NOAA CDR Program Coding standards, 23 pp. [Available at http://www1.ncdc.noaa.gov/pub/data/sds/cdr-general-programming-standards.pdf].

[RD.5] BIPM, 2008: International vocabulary of metrology – Basic and general concepts and associated terms (VIM), BIPM, JCGM 200:2008 (http://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2008.pdf).

[RD.6] Joppa Lucas N.G. McInerny, R.Harper, L. Salido, K. Takeda, K. O'Hara, D. Gavaghan, S. Emmott, 2013, Troubling Trends in Scientific Software Use, Science, Vol. 340 no. 6134, pp. 814-815, DOI: 10.1126/science.1231535).

[RD.7] GCOS-154, 2011: Systematic Observation requirements for Satellite-Based Products for Climate – 2011 Update, 139 pp.

[RD.8] Leroy, S. S., J. G. Anderson and G. Ohring, 2008: Climate Signal Detection Times and Constraints on Climate Benchmark Accuracy Requirements. *J. Climate*, **21**, 841-846, DOI: 10.1175/2007JCLI1946.1.

2 OVERVIEW

Developing ECV climate data records poses many challenges because of the varied use of climate data, the complexities of data record generation, and the difficulties in sustaining the activities over extended periods of time. Therefore it is essential to assess the capability of the existing climate data record development activities to ensure the prolonged generation of high quality ECV climate data records so that they can help to produce the underpinning science that supports decisions on mitigation and adaptation for a changing Earth climate.

In preparation of the Copernicus Climate Change Service an assessment of the needs for full access to standardised climate change data is mandatory. The European Joint Research Centre conducted a workshop 2009 that did an ad hoc analysis of the European capacity on the means to provide these data and how Copernicus Services can effectively contribute to providing these data. The report by Wilson et al. [RD.1] is summarising the results of this workshop that identified 44 GCOS ECVs as the minimum set of standardised climate data that EC should be considering. This workshop did also a first attempt to analyse the capacity according to maturity, differentiating between sustained operational capacity and non-operational funded repetitive capacity and additional infrastructure needs in order to fill gaps identified.

The report by Dowell et al. [RD.2] lines out a high level strategy for an architecture for climate monitoring from space that considers the whole value adding chain from making measurements to the development of policy and decision making. This report details two usage scenarios for such architecture:

- The promotion of a common understanding of the implementation implications of meeting the various climate monitoring requirements, and
- To support an assessment of the degree to which the current and planned systems that provide measurements from which climate data records are generated meet the requirements, and the generation of an action plan to address any identified shortfalls/gaps.

Essential for the second usage scenario is to assess what exists, what the degree of completeness and sustainability of the existing is, what quality the existing has and what is planned/committed for the future. The group of authors of the Dowell et al. [RD.2] report and the CEOS Working Group Climate together with WMO established the so called GCOS ECV inventory (ecv-inventory.com) for climate data records derived from satellite measurements. Currently, the inventory consists of approximately 220 entries provided by space agencies around the world and provides a first basis for an analysis of the existing data records. Because the first call to populate the inventory was only directed to space agencies the current inventory holding is not complete and further work is needed to cover all relevant data records. In addition an analysis of the ‘fit for purpose’ of the data records needs to be done.

To support the international activities described above and the establishment of the Copernicus Climate Change Services one major objective of the CORE-CLIMAX project is to systematically assess the capacity of ongoing European activities in the area of generation and provision of climate data records. With respect to a Copernicus Climate Change Service also the role of in situ data and model-based reanalysis needs to be considered.

For an assessment of the European capacity in the most objective way possible we need tools that provide a basis for information preservation, expectations, and a metric for progress to completeness. The maturity matrix approach proposed by Bates and Privette [RD.3] offers a systematic mean to assess if the data record generation follows best practises in the areas science, information preservation and usage of the data. Some example uses of the matrix maturity are the assessments of data records developed in the NOAA Climate Data Record program and in the 2nd phase of SCOPE-CM to measure progress in the projects. For both these cases, maturity assessments were first done as self assessments. External assessments could be done in a form of audit.

The CORE-CLIMAX project's proposition is based on [RD.3], but extending the model to more general so that it can be applied not only for satellite data sets, but for all climate data records (*in situ*, combined satellite and *in situ*, reanalyses). The project discussed its adapted approach with many leading initiatives in Europe such as the EUMETSAT network of Satellite Application Facilities (SAF) and the ESA Climate Change Initiative but also internationally with WMO, the CEOS WG Climate, NOAA and USGS.

Basically, three different aspects of our capacity to generate data records need to be considered:

- Scientific, engineering and information preservation practises;
- Usage of products including feedback and update mechanisms ;
- Quality of products with respect to applications.

Assessing if data record generation follows best practises provides an internal view on strengths and weaknesses of the processes to generate, preserve and improve climate data records for agencies and each individual data record provider. It also provides a general information to the community concerning the status of individual data records as well as collective information on the state of all existing records, highlighting areas for development and improvement. The assessment of quality of products is facilitating an external view on data records trying to answer the most important user question: Is the quality good enough for my application?

The CORE-CLIMAX project defined three major elements for its capacity assessment:

- Data record descriptions that contain technical specifications and also information on quality, e.g., links to further documentation and/or inventories such as the CGMS-CEOS-WMO inventory (see Section 3);
- A System Maturity Matrix (SMM) that evaluates if the production of a data record follows best practices for science, engineering, information preservation and facilitation of usage (see Section 4), and;
- A new so called Application Performance Matrix (APM) that attempts to evaluate the performance of an ECV CDR with respect to a specific application (see Section 5). To be able to apply the APM, user requirements for each application are needed to compare the actual technical specifications and validation results to them.

The three elements of the capacity assessment are designed to be independent of each other and represent means to support an assessment but do not provide the assessment results per se. The SMM is designed to principally be used without considering specific applications. With this the SMM does not depend on user requirements for specific applications and their change over time. In contrast the APM facilitates a comparison of the real technical features of a data record and results of validation and other data quality assessment activities to user

requirements for an application. It basically provides summary information on how close a specific data record is at fulfilling the requirements of a specific application. The APM was developed because the need of giving advice to data users what data record can be used for what application. This need is manifested for instance in the huge amount of information provided on validation of data records that is unlikely to be processed by institutions that want to use the data records. The APM is intended to support institutions in making choices among different existing data records without the need to assess the full documentation of all potential data records. However, it shall be noted that the APM is a new tool that will be used for the first time in the CORE-CLIMAX capacity assessment workshop. Thus, it is expected that the tool will be further adapted or even proved to be not needed.

3 DATA RECORD DESCRIPTION TEMPLATE

In Appendix A you can find the data record description template that shall be filled for each individual data record that enters the CORE-CLIMAX assessment. The template is structured very similar to the template used for data sets entering the Climate Model Inter-comparison Project (CMIP) exercise. Only the part on the applications has been extended as the usage of most climate data records goes beyond the climate model comparison. Keeping these templates very similar was done purposefully to support the usage of the assessed data records in the CMIP-6 exercise with preparations being started during 2014.

The Data Set Description Template contains advice on how to fill the individual sections. The overall aim is that these descriptions do not extend to more than five pages.

4 CORE-CLIMAX SYSTEM MATURITY MATRIX

The SMM is a tool to assess the system maturity of a CDR. SMM basically assesses whether CDR generation procedures have been compliant with best practices developed and accumulated by the scientific and engineering communities. The concept behind the CORE-CLIMAX system maturity matrix can be best illustrated as shown in Figure 1.

Creation of a climate data record is anchored on a number of assumptions and approximations, and thus is associated with significantly large uncertainties. This is mainly because the observing systems were designed to measure weather, but not for monitoring climate. Unless these assumptions and approximations are well understood and associated uncertainties are well characterized it is quite possible to misinterpret results of scientific analyses using these data records. Therefore uncertainty characterisation is a key area where CDRs need to achieve high levels of maturity.

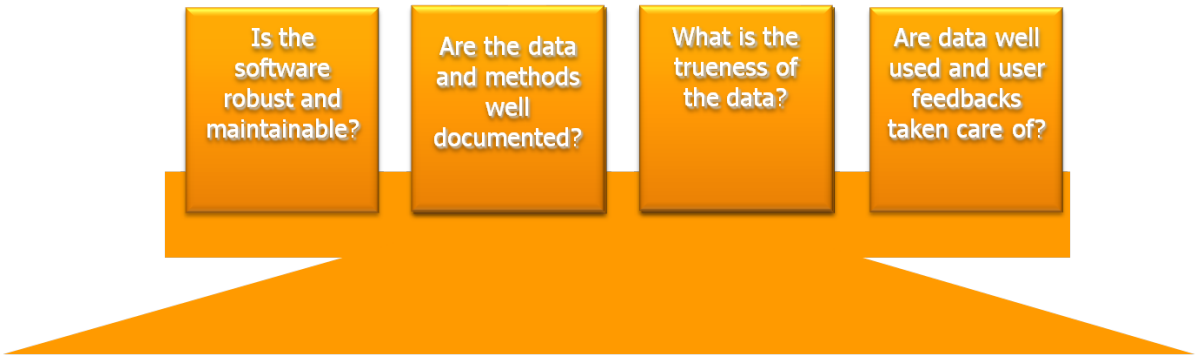
Stable and easily maintainable software is one of the essential components of successful CDRs. It should be easy to diagnose deficiencies, to make changes to the software, and to test the software after modification. Non-maintainable software can result in unexpected increase in the production cost of data sets. The metadata, especially describing the input raw data are essential because development of a CDR is often an evolutionary process and repeated reprocessing of the input dataset is necessary. This also demands the archival of the raw data for reprocessing. CDRs shall be archived in a way that allows easy access to the users with varying requirements and skills. Therefore it demands less complicated file structures and provisions for read and analyses (e.g., sub-setting, plotting) software. Availability of comprehensive descriptions of technical and scientific aspects of the production chain is another essential characteristic of a mature CDR.

Above all the most important maturity characteristic of a successful CDR is the acceptance and usage by the user community and whether there are mechanisms to receive and incorporate feedbacks from the user community.

There are 6 major categories where assessments are made:

1. Software readiness
2. Metadata
3. User documentation
4. Uncertainty characterisation
5. Public access, feedback, and update
6. Usage

For each of these categories the assessment will assign a score from 1 to 6 that reflects the maturity of the CDR with respect to a specific category. An overall score, e.g., an arithmetic mean for a CDR might be computed out of the six categories but it is not considered to be very useful.



Software readiness	Metadata	User documentation	Uncertainty Characterization	Public Access, Feedback and Update	Usage
Are the codes compliant with standards, stable, portable and reproducible?	Do the metadata meets international standards, and allows provenance tracking?	Are the formal documents and peer-reviewed papers up-to-date and public?	Are the uncertainties assessed systematically in a standard manner?	Are the data, source code, and documents publicly available and regularly updated?	Are the data widely used in the scientific, and decision and policy making communities?

Figure 1: Illustration of CORE-CLIMAX system maturity matrix (SMM).

The maturity is also considered in three broad categories that give information on the grade of sustainment of the CDR generation process. The nomenclature for these broad categories has been imported from NOAA and follows [RD.3]:

- Maturity scores 1 and 2 establish Research Capability (RC): All aspects of the CDR are still under development and with the PI most likely in projects.
- Maturity scores 3 and 4 establish an Initial Operations Capability (IOC): At this stage the CDR and associated material are available to the user community. The CDR has reached a status where its usefulness is completely demonstrated and decisions need to be made to sustain its maintenance and further development. At this stage so called

transitions of CDR generation capabilities from research units to more operational oriented units are happening. Good examples for this are the import of the HOAPS data record (www.hoaps.org) from the Max-Planck Institute in Hamburg into the EUMETSAT Climate Monitoring Satellite Application Facility or the transition of the well known International Satellite Cloud Climatology Project (ISCCP) from NASA into NOAA-NCDC.

- Maturity scores 5 and 6: Full Operations Capability (FOC): At this stage the production of the CDR has been transitioned into operational environments, e.g., the whole processing process is under configuration management, fully automated and performance is monitored. The production chain meets the goal of acquiring capabilities to provide uninterrupted and indefinite data provision for climate monitoring. The data provider, e.g., a space agency takes complete responsibility for the maintenance and also further development of the CDR. The specific development activities still are performed by scientists within or external to the responsible agency or both. A current example for a full operations capability is the EUMETSAT CM-SAF.

The major categories of the SMM shown in Figure 2 are subdivided into several minor categories and assessment scores are assigned based on scores in these minor categories. Currently, the minimum score of the minor categories are taken as the score for a major category. The motivation for taking the minimum score is given by the fact that this score is informing about completeness of a major category. It directly points to an area for improvement. The minimum score is accompanied by the mean value of the minor categories scores. Differences of the minimum and mean score give an indication if the scores of the minor categories are close together or if some minor categories have a very high score and only one has a low score. It should be noted that the numbers need anyway an interpretation per assessed data record because the circumstances under which the data records were created hugely differ for satellite, in situ data records and reanalysis.

In the following subsections we provide instructions on how to assign scores to each of the minor categories. The minor categories sometimes include categories that cannot easily be assessed by an external assessor without asking the provider of the data which could be done in a formal audit type assessment but not in the planned CORE-CLIMAX capacity assessment. The major and minor categories that are envisaged to only be used in the self assessment are labelled in the following instruction sections.

The SMM is provided as a multi-level Excel file where the scores shall be provided in the pages associated with the minor categories. These scores are then automatically used to compute the score for the major category. If a minor category is not filled a maturity of 1 will be set. There is one exception which is in the category Usage. In this category the usage of a data record is considered for applications in research and decision making. Which columns are taken into account depends on the intention of the data record. For instance, if the description is only pointing to use in research only that category shall be used to compute the overall maturity.

It is planned to replace the Excel file with a web based tool by the time of the workshop. The CORE-CLIMAX project will inform all workshop participants about its availability.

It is very important to use a unique CDR name and identification number (version) when the SMM is filled to assess a CDR. This shall match the name and identification information on the data set description form. Also a provision of the assessment date to follow the evolution in maturity of a particular CDR is very important.

4.1 Software Readiness

In this major category there are four minor categories: **These are mainly meant to be for self-assessment because the information is rarely publicly available.** All minor categories can be assessed internally and maybe externally but only by asking the data provider directly. However, some indications about software can be found by looking at the availability of software installation/user manual and/or programming guidelines on web pages presenting whole programs, e.g., http://www.ncdc.noaa.gov/cdr/guidelines.html.

The software readiness category provides information on the maintainability of software used to generate the data record. All software used to manipulate the data to its distributed product should be assessed. High maturity is equivalent to a system that doesn't depend on specific individuals that know the software since its origin. Software becomes more easily understandable if the programming follows standards and the installation and usage is documented. Software is also maintainable if it can be ported to other systems which either can be an installation at another place or the need to run it on a new computer system at the same place.

Maturity	SOFTWARE READINESS	METADATA	USER DOCUMENTATION	UNCERTAINTY CHARACTERISATION	PUBLIC ACCESS, FEEDBACK, UPDATE	USAGE
1	Conceptual development	None	Limited scientific description of the methodology available from PI	None	Restricted availability from PI	None
2	Research grade code	Research grade	Comprehensive scientific description of the methodology, report on limited validation, and limited product user guide available from PI; paper on methodology is sumitted for peer-review	Standard uncertainty nomenclature is idenitified or defined; limited validation done; limited information on uncertainty available	Data avaliable from PI, feedback through scientific exchange, irregular updates by PI	Research: Benefits for applications identified DSS: Potential benefits identified
3	Research code with partially applied standards; code contains header and comments, and a README file; PI affirms portability, numerical reproducibility and no security problems	Standards defined or identified; sufficient to use and understand the data and extract discovery metadata	Score 2 + paper on methodology published; comprehensive validation report available from PI and a paper on validation is submitted; comprehensive user guide is available from PI; Limited description of operations cocept available from PI	Score 2 + standard nomenclature applied; validation extended to full product data coverage, comprehensive information on uncertainty available; methods for automated monitoring defined	Data and documentation publically available from PI, feedback through scientifc exchange, irregular updates by PI	Research: Benefits for applications demonstrated. DSS: Use occuring and benefits emerging
4	Score 3 + draft software installation/user manual available; 3rd party affirms portability and numerical reproducibility; passes data providers security review	Score 3 + standards systematically applied; meets international standards for the data set; enhanced discovery metadata; limited location level metadata	Score 3 + comprehensive scientific description available from data provider; report on inter comparison available from PI; paper on validation published; user guide available from data provider; comprehensive description of operations concept available from PI	Score 3 + procedures to establish SI traceability are defined; (inter)comparison against corresponding CDRs (other methods, models, etc); quantitative estimates of uncertainty provided within the product characterising more or less uncertain data points; automated monitoring partially implemented	Data record and documentation available from data provider and under data provider's version control; Data provider establishes feedback mechanism; regular updates by PI	Score 3 + Research: Citations on product usage in occurring DSS: societal and economical benefits discussed
5	Score 4 + operational code following standards, actions to achieve full compliance are defined; software installation/user manual complete; 3rd party installs the code operationally	Score 4+ fully compliant with standards; complete discovery metadata; complete location level metadata	Score 4 + comprehensive scientific description maintained by data provider; report on data assessment results exists; user guide is regularly updated with updates on product and validation; description on practical implementation is available from data provider	Score 4 + SI traceability partly established; data provider participated in one inter-national data assessment; comprehensive validation of the quantitative uncertainty estimates; automated quality monitoring fully implemented (all production levels)	Score 4 + soure code archived by Data Provider; feedback mechanism and international data quality assessment are considered in periodic data record updates by Data Provider	Score 4+ Research: product becomes reference for certain applications DSS: Societal and economic benefits are demonstrated
6	Score 5 + fully compliant with standards; Turnkey System	Score 5 + regularly updated	Score 5 + journal papers on product updates are and more comprehensive validation and validation of quantitative uncertainty estimates are published; operations concept regularly updated	Score 5 + SI traceability established; data provider participated in multiple inter-national data assessment and incorporating feedbacks into the product development cycle; temporal and spatial error covariance quantified; Automated monitoring in place with results fed back to other accessible information, e.g. meta data or documentation	Score 5 + source code available to the public and capability for continuous data provisions established (ICDR)	Score 5 + Research: Product and its applications becomes references in multiple research field DSS: Influence on decision and policy making demonstrated

Figure 2: Top Level CORE-CLIMAX Maturity Matrix showing the major categories to be explored during the assessment.

4.1.1 Coding Standards

Coding standards are a set of conventions/rules specific for a coding language which describes style, practises and methods that greatly reduce the probability of introducing bugs. This is especially important in a team environment or group collaboration (which is generally the case for CDR development) so that uniform coding standards are used and reducing oversight errors and saving time for code reviews. It is assuring the maintainability of the code at reasonable cost. There are ISO standards available for software coding, but it is also common to follow organisational standards such as the NOAA CDR Programme coding standards [RD-4].

Table 1: The 6 maturity scores in sub-category Coding standards

Score	Description
1	No coding standard or guidance identified or defined
2	Coding standard or guidance is identified or defined, but not applied
3	Score 2 + standards are partially applied and some compliance results are available
4	Score 3 + compliance is systematically checked in all code, but not yet compliant to the standards.
5	Score 4 + standards are systematically applied in all code and compliance is systematically checked in all code. Code is not fully compliant to the standards. Improvement actions to achieve full compliance are defined.
6	Score 5 + code is fully compliant with standards.

Score 2: *Standard identified/defined means that the data record producer has identified or defined the standards to be used but has not applied it. The information about this most often can be found in software description documents or programming guidelines available from web pages or by asking the data provider;*

Score 3: *This means that the data provider has started to apply the standards and implemented procedures to check the compliance. This information may be available by asking the data provider;*

Score 4: *Score 3 + procedures are systematically applied to check the compliance and the results are often available as internal reports;*

Score 5: *Data provider has identified departures from the standards and actions are planned to achieve full compliance;*

Score 6: *At this stage the software shall be fully compliant with its description and the documented standard. This includes procedures to check the compliance and the results of the tests conducted.*

4.1.2 Software Documentation

Software Documentation – here one needs to assess whether the code is documented with proper header, change history, and comments describing the processes, whether README file is up-to-date, there is documentation available which describe design and overview of the software, and there is software installation and user manual available.

Table 2: The 6 maturity scores in sub-category ***Software documentation***

Score	Description
1	No documentation
2	Minimal documentation
3	Header and process description (comments) in the code, README complete
4	Score 3 + a draft Software Installation/User Manual
5	Score 4 + enhanced process descriptions throughout the code; software installation/user manual complete
6	As in score 5

The assessment can be made, for example, as below:

Score 2: *There are header and limited comments in the code and installation instructions available, but no other documentation is available;*

Score 3: *README file should at least contain information on "Configuration instructions", "Installation instructions", "Operating instructions", "Copyright and licensing", "Contact information", etc.;*

Score 4: *Score 3 + Software User Manual should at least contain information on software concept and design and providing instructions for installing and using the software;*

Score 5: *Code is very well documented and installation/user manual is complete and available on data provider's web page;*

Score 6: *Not used.*

4.1.3 Portability and Numerical Reproducibility

Portability and numerical reproducibility is the usability of the software in different environments (different computing platforms such as Linux, Solaris, Mac OS, Windows etc. and different compilers such Intel, IBM, GNU, Portland, etc) and the results are numerically reproducible. It is important for migrating software from old to new computer system and from one place to another.

Table 3: *The 6 maturity scores in sub-category* ***Numerical reproducibility and portability***

Score	Description
1	Not evaluated
2	PI affirms reproducibility under identical conditions
3	PI affirms reproducibility and portability
4	3rd party affirms reproducibility and portability
5	Score 4 + 3rd party can install the code operationally
6	Score 5 + Turnkey system

The assessment can be made, for example, as below:

Score 1: *Not evaluated means this has not been considered at all;*

Score 2: *PI affirms that the software reproduces results when run on same platform with same input and same compiler, but for different runs. This information can be obtained by asking the data provider;*

Score 3: *The software produces numerically reproducible results on different computing platforms (such as Linux, Solaris, Mac OS, Windows etc.), and/or with different compilers (such Intel, IBM, GNU, Portland, etc);*

Score 4: *Score 3 + 3rd party can install the code operational with minimal manual efforts. Runs reveal that the output is numerically reproducible (within machine rounding errors). This information shall be found in software description documents available from data provider's web pages;*

Score 5: *Score 4 + the code is already used by 3rd party in operational environment under configuration control. This shall be described in the software installation/user manual;*

Score 6: *Turnkey is software that is designed, supplied, built or completely installed and ready to operate. The term implies that the end user just has to turn a key and start using the software, e.g., Linux OS. This shall be described in the software user manual.*

4.1.4 Security

Security is associated with software contents that either have the potential to destroy files and complete environments or are related to file transfer between compute environments. Both should not be contained in software. The security category also checks whether the file system can be accessed from outside and may hamper the integrity of the data generation environment.

Table 4: The 6 maturity scores in sub-category **Security**

Score	Description
1	Not evaluated
2	PI affirms no security problems
3	Submitted for data provider's security review
4	Passes data provider's security review
5	Continues to pass the data provider's review
6	As in score 5

The assessment can be made, for example, as below:

Score 1: *Not evaluated at this stage means that software security issues have not been considered;*

Score 2: *PI has done the testing for security issues in the code and found none. This information can be obtained by asking the PI;*

Score 3: *This information can be obtained by asking the PI. This is a necessary step before porting the software from a research environment to an operational environment;*

Score 4: *This means the software has passed data provider's quality assurance and security tests. Information on this shall be obtained from software installation/user manual;*

Score 5: *Data provider does security assessment whenever there is a software update and the results shall be available from updated software installation/user manual;*

Score 6: *Not used.*

4.2 Metadata

Metadata is 'data' about data and data providers are responsible for providing metadata. Metadata shall be standardised, as complete as possible and adequate. In this category the maturity is assessed using three minor categories that consider the standards used, the meta data at the collection level, i.e., valid for the complete data record and at file level, i.e., valid for the data at a specific granularity.

4.2.1 Standards

Standards – It is considered to be good practise to follow international standards such as ISO-19115 (http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020), CF (Climate and Forecast) http://cf-pcmdi.llnl.gov/), or organisational such as NOAA/NCDC, or the Marine Environmental Data and Information Network (MEDIN, http://www.oceannet.org/).

Table 5: The 6 maturity scores in sub-category **Standards**.

Score	Description
1	No standard considered
2	No standard considered
3	Metadata standards identified and/or defined but not systematically applied,
4	Score 3 + standards systematically applied at file level and collection level by data provider. Meets international standards for the dataset
5	Score 4 + meta data standard compliance systematically checked by the data provider.
6	Score 5

Notes: It is likely that this minor category can only be assessed internally. An external assessment can be made by asking the data provider directly. However, signs for used standards can be found by looking at the data record documentation and/or at a sample data file.

The assessment can be made as follows:

Score 3: Standard identified/defined means that the data record producer has identified or defined the standard to be used but has not applied it. The information about this most often can be found in Format description documents available from web pages or from statements on web pages;

Score 4: A systematic application requires that you can find it in every file of the data product and descriptions;

Score 5: This means that the data provider has implemented procedures to check the metadata contents;

Score 6: not used.

4.2.2 Collection Level

Collection Level – these are attributes that apply across the whole of the data set, such as digital object identifier, processing methods (e.g., same algorithm versions), general space and time extents, creator and custodian, references, processing history. Discovery metadata is part of this, which is a list of information that allows other people to find out what the data set contains, where it was collected and where and how the data record is provided.

Table 6**:* The 6 maturity scores in sub-category ***Collection Level

Score	Description
1	None
2	Limited
3	Sufficient to use and understand the data independent of external assistance; Sufficient for data provider to extract discovery metadata from meta data repositories
4	Score 3 + Enhanced discovery metadata
5	Score 4 + Complete discovery metadata meets international standards
6	Score 5 + Regularly updated

The assessment can be made as below:

Score 1: Data files have no global attributes;

Score 2: Only attributes like space and time coverage, custodian of data are provided, but no information on measurement/processing methods or history are available;

Score 3: All relevant information on processing (for example retrieval input radiance data version and provenance) and for general understanding the data (such as references and comments). Also contains information on how to extract discovery metadata from repositories;

Score 4: Score 3 + more information on discovery metadata (for example, how to obtain raw data (level 0 in case of satellites) and the necessary information to process those data);

Score 5: Score 4 + all the available information on the data are provided with the data using a defined standard;

Score 6: Score 5 + Updates are provided whenever new metadata become available. For example, information on events impacting the quality of the data record (e.g., information provided at http://www.oso.noaa.gov/poesstatus/), or the addition of commentary metadata such as publications written about the data record.

4.2.3 File Level

***File level** attributes* are those specific to the granularity of the data and vary with each measurement entity.

Table 7: The 6 maturity scores in sub-category ***File Level***

Score	Description
1	None
2	Limited
3	Sufficient to use and understand the data independent of external assistance
4	Score 3 + Limited location (pixel, station, grid-point, etc.) level metadata
5	Score 4 + Complete location (pixel, station, grid-point, etc.) level metadata
6	Score 5

The assessment can be made as follows:

Score 1: Data files contain no variable attributes;

Score 2: Data geographical coordinates are described and data units are provided;

Score 3: The data files are provided with data geographical coordinates, units, valid range, and missing and/or fill values;

Score 4: Score 3 + coordinate bounds are provided. There is some location level (i.e., station level for an in situ data set, pixel level for a swath level satellite data, grid point level for a gridded in situ or satellite and reanalysis data) information available in the data files. An example for location level metadata is surface type;

Score 5: Score 4+ additional location level metadata such as level of confidence in the retrieval for each data location is provided;

Score 6: Not used.

4.3 User documentation

Documentation is essential for the effective use and understanding of a data record. There are four minor categories to assess the completeness of user documentation.

4.3.1 Formal Description of Scientific Methodology

Formal description of scientific methodology refers to description of the physical basis of measurements, processing of the raw data to higher levels (in case of satellite data this involves geo-location, calibration, inter-calibration, retrieval methods, and space-time averaging methods). For station based data records this can be descriptions of data filtering, corrections, aggregation procedures, etc. For reanalysis this would include the description of data assimilation techniques, the physical model used, etc. An example of a formal description is an Algorithm Theoretical Baseline Document (ATBD) as e.g., provided for a satellite retrieval algorithm. As such documents are most often subject of an agency internal review process it is required to also have a peer reviewed publication(s) on the methodology to increase the maturity.

Table 8: The 6 maturity scores in sub-category ***Formal description of scientific methodology***

Score	Description
1	Limited scientific description of methodology available from PI
2	Comprehensive scientific description available from PI and Journal paper on methodology submitted
3	Score 2 + Journal paper on methodology published
4	Score 3 + Comprehensive scientific description available from Data Provider
5	Score 4 + Comprehensive scientific description maintained by data provider
6	Score 5 + Journal papers on product updates published

EXAMPLE: Satellite retrieval algorithm:

Score 1: Draft of ATBD for the retrieval algorithm is available, e.g.in the Internet. To assess it one would search the web pages on the data record for an ATBD;

Score 2: Complete version of ATBD(s) is available which includes all the steps which were used to produce the data set from basic measurements to the final product. The method is also summarised and submitted to a relevant journal for publication. The latter can be hard to assess from outside but often submitted papers appear on web pages of existing data records;

Score 3: In addition to Score 2 a journal paper is available which can be checked using tools such as Web of Science;

Score 4: ATBD is available from the data provider, e.g., if a data record is transferred from a research group (which is part of sustaining the data record measured by the maturity) to an operational/research agency that takes responsibility for production and/or distribution of the data record the documentation of the methodology shall appear on the data provider's web site. It is assumed that the documents have passed each agencies internal review processes before they appear. To assess this, one needs to browse the data provider's web site.

Score 5: This score is related to updates of the documentation following updates of the data record (see Public Access, Feedback and Update). A sign for maintenance is if the ATBD has proper document version numbering and is referring to a specific version of the data record;

Score 6: The ultimate score in this example is that each update in the retrieval algorithm is also published in peer reviewed literature, i.e., accepted by the community through the anonymous review process developed by the community.

Note: In case of in situ data sets or reanalyses ATBD may not be the name of the document. In that case measurement manual, post-processing manual, model descriptions or other

technical reports can have the same functionality as the ATBD. It is however required that a description of the method is available to the public.

4.3.2 Formal Validation Report

A ***Formal validation report*** contains details on the validation activities that have been done to assess the fidelity of the data record. It describes uncertainty characteristics of the data record found through the application of uncertainty analysis (see section on Uncertainty Characterisation), and provides all relevant references.

Table 9: The 6 maturity scores in sub-category ***Formal validation report***

Score	Description
1	None
2	Report on limited validation available from PI
3	Report on comprehensive validation available from PI; Paper on product validation submitted
4	Report on inter-comparison to other CDRs, etc. Available from PI and data Provider; Journal paper on product validation published
5	Score 4 + Report on data assessment results exists
6	Score 5+ Journal papers more comprehensive validation, e.g., error covariance, validation of qualitative uncertainty estimates published

EXAMPLE: Satellite retrieval of temperature profiles

Score 1: No validation is done and hence no report;

Score 2: Report on limited validation done using sounding data from a few stations is available by directly asking the PI or from PI's web pages;

Score 3: Detailed report on validation using radiosonde profiles with global representativeness in space and time. Quality controlled radiosonde data such as IGRA or data from reference upper air stations such as GRUAN has been used for validation. PI has also submitted an article on the product and its validation to publish in a peer-review journal. In most cases the report and the submitted article can be found on PI's web pages or it can be obtained by asking the PI;

Score 4: Reports on inter-comparisons to other satellite derived temperature profile data sets are available at this stage both from PI and the data provider. Article submitted on validation is now published and is available from PI/data provider's web page and listed in e.g., Web of Science;

Score 5: The data record has appeared in assessment reports such as from GEWEX;

Score 6: *More papers on uncertainty characterisation are published and data set developer/provider maintains up-to-date information on uncertainty in their data records. Below we give two examples:*

Remote sensing systems maintain a webpage for describing uncertainty in their upper air temperature data set:

http://www.remss.com/measurements/upper-air-temperature#Uncertainty

Met Office Hadley Centre maintains a webpage for describing uncertainties in their sea surface temperature data set:

http://www.metoffice.gov.uk/hadobs/hadsst3/uncertainty.html

Both pages contain a comprehensive list of peer-reviewed publications documenting uncertainties in these data sets.

4.3.3 Formal Product User Guide (PUG)

Formal product user guide (PUG) – This document contains definition of the data set, requirements considered while developing the data set, overview of input data and methods, general quality remarks, validation methods and estimated uncertainty in the data, strength and weakness of the data, format and content description, references, and contact details.

Table 10: The 6 maturity scores in sub-category **Formal PUG**

Score	Description
1	None
2	Limited product user guide available from PI
3	Comprehensive User Guide available from PI
4	Score 3 + available from data provider
5	Score 4 + regularly updated by data provider with product updates and/or new validation results.
6	Score 5

The assessment can be made as follows:

Score 1: *PI has not written a user guide yet;*

Score 2: *A draft user guide may be available from PI directly or from PI's web pages;*

Score 3: *A complete and reviewed (for example by the data provider) user guide is available from PI's webpage. At this stage the user guide shall contain all details given in the above paragraph;*

Score 4: *Score 3 + user guide is available from data provider's web page as well;*

Score 5: *Updated user guide is available from data provider's web page. A sign of updating is increasing version numbering. This is related to updates in the data record itself;*

Score 6: *Not used.*

4.3.4 Formal Description of Operations Concept

Formal description of operations concept – In general, description of operations concept will include the following:

- Statement of the goals and objectives of the system;
- Limitations and constraints affecting the system;
- Clear statement of responsibilities and authorities delegated;
- Processes for initiating, developing, maintaining, and retiring the system.

It should relate a narrative of the process to be followed in implementing and operating a system that produces the data record under consideration. The existence and usage of such a document is increasing the maturity of the process because it makes it more independent of the individuals implementing and operating a system. It also enables studying impacts of planned changes to an existing system.

This category is most likely be considered in the self assessment alone. Operations concept documents are mainly for internal usage and relatively uninteresting for a data record user. However, very often flow charts indicating the major elements of the processing are part of such a document and can support a user in a better understanding how a data record was produced.

Table 11: The 6 maturity scores in sub-category **Formal** d**escription of operations concept**

Score	Description
1	None
2	None
3	Limited description of operations concept available
4	Comprehensive description of operations concept available
5	Operations concept and description of practical implementation available
6	Score 5 + Operations concept regularly updated

The assessment can be made as follows:

Score 1: *At the level of a research capability an operations concept is not needed. It is rather expected that only a few people interact to generate a data record;*

Score 2: *Same as for Score 1;*

Score 3: Draft of the document (which could be just a flow chart) is available. Sometimes such information can be found on web pages of the PI;

Score 4: A comprehensive description of the operations concept is available;

Score 5: Operations concept inclusive of the description of the practical implementation is available;

Score 6: Score 5 + Data provider updates the concept documents whenever there is an update in the operations concept.

4.4 Uncertainty Characterisation

The category Uncertainty Characterisation assesses the practises used to characterise and represent uncertainty in a data record. Four minor categories are considered that try to encompass standards used, the validation process, how uncertainty is quantified and if an automated quality monitoring is implemented that increases the efficiency of production and validation.

4.4.1 Standards

There are no international standards as such available for uncertainty characterisation. However, there is a compelling need for this. Organisational and program standards are sometimes available (e.g., NOAA CDRP). There are basically two areas where standards play an important role:

- Uncertainty nomenclature which should follow established definitions as such provided by metrological institutions. A standard uncertainty nomenclature is for instance provided in [RD.5];
- SI traceability that is the property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties.

The first bullet is indicating that emphasis should be put on the usage of existing and correct definition of uncertainty measures to make results from validation studies concerning the same ECV more comparable.

To support a claim of traceability, the provider of a measurement result or value of a standard must document the measurement process or system used to establish the claim and provide a description of the chain of comparisons that were used to establish a connection to a particular stated reference. For satellite data records the second bullet is practically indicating that uncertainty arising from systematic and random effects in the measurements shall be provided for each step of the product generation, for example, pre-launch and post-launch calibrations as well as inter-calibration of instruments, retrieval, sampling, and aggregation steps. In the end it shall be related to reference data such as laboratory measurements, those from reference measurements such as the Global Reference Upper Air Network (GRUAN) or data from high spectral resolution and stable space-based instruments such as AIRS/IASI may be used to characterise uncertainties. As absolute references are not readily available measurements may be taken as reference if their accuracy is about one order of magnitude better compared to the measurement that is assessed.

For in situ data traceability can be established by calibrating networks of measurement devices by comparing the instruments with the in laboratory reference instrument or through measurement device inter-comparison activities.

It is acknowledged that for reanalysis systems SI traceability is very hard to be established. However, it can be assessed if the quality of input data to assimilation systems is characterised in a traceable manner and also if provided estimates of uncertainty are used in the data assimilation process or other usage of data, e.g., as boundary condition in ensemble model runs.

The 'SI' element of the traceable means that any unit used shall be traceable back to the seven well-defined base units of the SI system which are the metre, the kilogram, the second, the ampere, the Kelvin, the mole, and the candela.

Table 12: *The 6 maturity scores in sub-category* ***Standards***

Score	Description
1	None
2	Standard uncertainty nomenclature is identified or defined
3	Score 2 + Standard uncertainty nomenclature is applied
4	Score 3 + Procedures to establish SI traceability are defined
5	Score 4 + SI traceability partly established
6	Score 5 + SI traceability established

The assessment can be made as follows:

Score 1: Nothing has been done in early stages of development;

Score 2: The data provider states in the documentation or on web pages which nomenclature is used but no consistent application of it can be verified;

Score 3: Score 2 + the application of the nomenclature is evident from documents such as validation reports and user guides;

Score 4: Score 3 + a document exists that describe how traceable comparison chains to a specified reference will be established;

Score 5: Score 4 + the steps in the afore mentioned document are implemented as far as possible. It is known that in particular for satellite measurements no real references in space are existing but if an unbroken chain of comparisons to the best available instrument is established Score 5 can be assigned;

Score 6: Score 5 + the traceability is fully established. Maybe no existing data record will reach Score 6 until real reference measurements in space are provided but by not achieving it the need remains always documented.

Note: The maturity levels start with the nomenclature and finishes with the Si traceability because this presents the logical order in which a system to quantify and present uncertainty would be build.

4.4.2 Validation

Validation - This minor category evaluates the extent to which the product has been validated to provide uncertainty estimates.

Table 13*: The 6 maturity scores in sub-category* ***Validation***

Score	Description
1	None
2	Validation using external reference data done for limited locations and times
3	Validation using external reference data done for global and temporal representative locations and times
4	Score 3 + (Inter)comparison against corresponding CDRs (other methods, models, etc)
5	Score 4 + data provider participated in one inter-national data assessment
6	Score 4 + data provider participated in multiple inter-national data assessment and incorporating feedbacks into the product development cycle

The assessment can be made as follows:

Score 1: *New product and no validation activity has been performed;*

Score 2: *The product is validated only for a few locations or short periods. For example, a temperature profile data set derived from satellite measurements is validated only for a few radiosonde stations such as ARM stations;*

Score 3: *Following the same example here validation is done with a global radiosonde data which is quality controlled (for example, IGRA data set) or GRUAN stations;*

Score 4: *Score 3 + comparisons are made with other satellite derived temperature products using different retrieval technique and/or re-analyses data sets;*

Score 5: *Data provider participated in an international data quality assessment. For example, GEWEX did assessments for cloud properties and radiation fluxes where a team produces multi data record comparison results that are reviewed by an independent panel;*

Score 6: *Data provider participated regularly in more than one data quality assessment and results resulting in improvement of the data record.*

4.4.3 Uncertainty Quantification

Uncertainty quantification - This minor category evaluates the extent to which uncertainties have been quantified.

Table 14: *The 6 maturity scores in sub-category* ***Uncertainty quantification***

Score	Description
1	None
2	Limited information on uncertainty arising from systematic and random effects in the measurement
3	Comprehensive information on uncertainty arising from systematic and random effects in the measurement
4	Score 3 + quantitative estimates of uncertainty provided within the product characterising more or less uncertain data points
5	Score 4 + temporal and spatial error covariance quantified
6	Score 5 + comprehensive validation of the quantitative uncertainty estimates and error covariance

The assessment can be made as follows:

Score 1: *No validation and therefore no uncertainty quantification.*

Score 2: *Only limited information on uncertainty because of limited validation;*

Score 3: *Comprehensive information is available so that nature of uncertainty is well understood, for example, whether uncertainty is varying depending upon geographic region, state, and instrument geometry. Uncertainties are estimated for each step of the production, for example, the uncertainty contributions in temperature profile data set from radiometric noise in the input satellite measurements, radiative transfer modelling and retrieval errors, sampling errors (e.g., non-availability of data in the presence of clouds or precipitation), smoothing errors due to insufficient horizontal and vertical resolutions of the instruments;*

Score 4: *Score 3 + quantitative comprehensive information described in Score 3 is available for each data point;*

Score 5: *Score 4 + the spatial and temporal error covariance quantified;*

Score 6: *Score 5 + the uncertainty estimates are validated using superior quality data sets (e.g., data set assessment activities).*

Note: A very detailed description of uncertainty in a SST data set is provided at http://www.metoffice.gov.uk/hadobs/hadsst3/uncertainty.html.

4.4.4 Automated Quality Monitoring

Automated quality monitoring is the monitoring of data quality while processing the data. Automated quality monitoring helps to assess, during the processing, major issues that may occur in a newly processed data record. It may lead to a stop and restart of processing activities if errors are detected. In that sense it can save significant resources in very large processing endeavours and is a clear sign for a mature processing system. Automatic quality monitoring a couple of steps such as defining a metric, procedures, data used in comparisons, setting of thresholds for deviations, and checking the data against them to identify anomalies in the data record.

Table 15: *The 6 maturity scores in sub-category* ***Automated quality monitoring***

Score	Description
1	None
2	None
3	Methods for automated quality monitoring defined
4	Score 3 + automated monitoring partially implemented
5	Score 3 + monitoring fully implemented (all production levels)
6	Score 5 + automated monitoring in place with results fed back to other accessible information, e.g. meta data or documentation

The assessment can be made as follows:

Score 1: *No automated quality monitoring in place.*

Score 2: *No automated quality monitoring in place. It is expected that at the research level no resources are foreseen to invest into automated monitoring tools;*

Score 3: *A metric (e.g., radiometric noise of one or more channels of the instrument used is significantly above specification, number of good retrievals in a grid box is below a threshold value), procedures, data used in comparisons, setting of thresholds for deviations, etc. for automated quality monitoring has been defined;*

Score 4: *Score 3 + the proposed monitoring is partially implemented, e.g., at some product levels or only for input or output data;*

Score 5: *Score 3 + quality monitoring is implemented at all production levels (i.e., inputs, retrieval, and space-time aggregation);*

Score 6: *Score 5 + Results of automated quality monitoring is now reflected in metadata and documentation. For example, the quality monitoring procedures and results are described in ATBD and product user guide.*

4.5 Public Access, Feedback and Update

This category contains four minor categories related to archiving and accessibility of the data record, how feedbacks from user communities are established and whether these feedbacks are used to update the data record.

4.5.1 Access and Archive

Access and archive evaluates the ease of distributing the data, documentation, and source code to users. It also checks the characteristics of the archive so that longer-term preservation is guaranteed. According to Long Term Data Preservation (http://earth.esa.int/gscb/ltdp/) guidelines an archive should keep more than one copy, use different media/technologies, and different locations. Public assess means that the data are available without restrictions, but access may be subject to a fee. Data provider here means organisations such as space agencies, national meteorological centres or research institutes. An institutionalised data provision is considered to be more mature compared to the provision by an individual investigator.

Table 16: *6 maturity scores in sub-category* ***Access and Archive***

Score	Description
1	Data may be available through request to PI
2	Data available through PI
3	Data and documentation archived and available to the public from PI
4	Data and documentation archived and available to the public from Data Provider
5	Score 4 + source code archived by Data Provider
6	Score 5 + source code available to the public from Data Provider

The assessment can be made as follows:

Score 1: *Data record is not ready yet to be given to users and is not archived; it may be available to beta-users for testing. PI is still conducting initial validation of the data product;*

Score 2: *Data record is now ready to be given to users, but not archived yet. Documentations are in draft form. Users can get the data by requesting is from the PI;*

Score 3: *Data record and documentation are readily available from the PI, e.g., on web pages;*

Score 4: *Data record and documentation are transferred from PI's to an institutional maintained archive from which the data are accessible for users;*

Score 5: *The source code is also archived by the data provider, but not publicly available.*

Score 6: The ultimate maturity is reached when the data record, documentation and the source code which has been used to produce the data record are archived, maintained and available to the public. See for example [RD.6] for the need of making codes public and peer-reviewed.

4.5.2 Version Control

Version control is a measure taken to trace back the different versions of algorithms, software, format, input and ancillary data, and documentation used to generate the data record under consideration. It allows clear statements about when and why changes have been introduced.

Table 17: Six maturity scores in sub-category - **Version control**

Score	Description
1	None
2	Preliminary versioning by PI
3	Versioning by PI
4	Version control institutionalised
5	Fully established version control considering all aspects
6	Not used

The assessment can be made as follows:

Score 1: No version number visible in data record files, metadata or documents;

Score 2: Data record contains some version information;

Score 3: Data record contains version information in meta data and documentation;

Score 4: Data version control is transferred from PI to an institutional maintained archive. This is for instance visible if you can order a version of data record from an archive;

Score 5: Data provider has established full version control for the data record including versions of algorithms, software, format, input and ancillary data, and documentation;

Score 6: The score is not used as there is no further step possible.

4.5.3 User Feedback

User feedback is important for developers and providers of data records to improve quality, accessibility, etc. of a data record. This category is to evaluate whether mechanisms are established to receive, analyse and use user feedbacks. Feedback can reach a data provider in many ways but needs to be organised when it systematically should be used to improve a data record and/or the service around it. In the scientific environment data records are presented and discussed at work shops and conferences. A scientist may takes messages back to his lab and starts to think and realise improvements if resources are available. A higher maturity for gathering feedback is obviously reached when a data record has been institutionalised and the responsible institute has established regular feedback processes may starting with a help desk up to periodical workshops where the feedback is gathered.

Table 18: *6 maturity scores in sub-category* ***User feedback mechanism***

Score	*Description*
1	None
2	PI collects and evaluates feedback from scientific community
3	PI and Data provider collect and evaluate feedback and from scientific community
4	Data provider establishes feedback mechanism such as regular workshops, advisory groups, user help desk, etc. and utilises feedback jointly with PI
5	Established feedback mechanism and international data quality assessment results are considered in periodic data record updates
6	Score 5 + Established feedback mechanism and international data quality assessment results are considered in continuous data provisions (Interim Climate Data Records)

The assessment can be made as follows:

Score 1: *Data record is not used by users yet, hence no feedback;*

Score 2: *Users are directly contacting PI to provide feedback or vice versa. This can be only known by asking the PI directly to assess or by looking for conference contributions about the data record;*

Score 3: *An institutionalised data provider is supporting the Principal Investigator collecting user feedbacks, e.g., the data record was produced as part of a larger programme and the agency organising the programme is also presenting the data record and is multiplying the feedback;*

Score 4: *Data provider has established feedback mechanisms. One can look for help desk support, announcement of annual workshops on a set of data records from one institution, etc.*

Score 5: *This will be reflected in user manual and other documentation on web pages, etc.;*

Score 6: *A sign of this is to check whether interim data records are provided (operational continuation of a climate data record employing the same procedures) and if feedback is also considered for this.*

4.5.4 Updates to Record

Updates to record evaluates if data records are systematically updated or if this is rather done in ad hoc fashion. The latter is an indication that the update very much depends on irregular funding and is not done by a bigger institution that provides the update as part of a service.

Table 19: *6 maturity scores in sub-category* ***Updates to record***

Score	Description
1	None
2	Irregularly by PI following scientific exchange and progress
3	Irregularly by PI following scientific exchange and progress
4	Regularly by PI utilising input from established feedback mechanism
5	Regularly operationally by data provider as dictated by availability of new input data or new methodology following user feedback
6	Score 5 + capability for fast improvements in continuous data provisions established (Interim Climate Data Records)

The assessment can be made as follows:

Score 1: *No update is made;*

Score 2: *This can be seen by occasional increase in the version number for the data set;*

Score 3: *Same as score 2*

Score 4: *This can be seen by increase in the version number for the data set and documentation at reasonable frequency, e.g., every 2-3 years and by announcements of planned new versions on web pages;*

Score 5: *This information will be available in product user guide, the CORE-CLIMAX data set description form or the data providers web pages;*

Score 6: *This information will be available in product user guide or in the CORE-CLIMAX data set description form and will be described in data provider's web pages.*

4.6 Usage

This category contains two minor categories related to the usage of products in research applications and for decision support systems. Under usage in decision support systems we understand the use in applications that directly support decisions, e.g., a NDVI product might be used as background map for clarifying insurance claims for cattle drovers in Africa or a solar irradiance map is directly used for infrastructure planning. In additions all citations in reports such as the Intergovernmental Panel for Climate Change (IPCC) reports that support decisions and policy making on mitigation and adaptation are credible for the decision support section.

The two minor categories allow for a separate assessment of the usage of data records, i.e., the assessment result can state a high maturity for usage in research and a lower or no maturity for decision support systems. For the overall score it is important to know for which application the data record was created. This information shall come from Section 1 of the CORE-CLIMAX Data Record Description Form (see Appendix A). If the description is only pointing to use in research only that category shall be used to compute the overall maturity.

4.6.1 Research

Research applications of a data product can be evaluated by its appearance in publications and citations of such publications.

Table 20: *6 maturity scores in sub-category* ***Research***

Score	Description
1	None
2	Benefits for research applications identified
3	Benefits for research applications demonstrated by publication
4	Score 3 + Citations on product usage occurring
5	Score 4 + product becomes reference for certain applications
6	Score 5 + Product and its applications becomes references in multiple research field

The assessment can be made as follows:

Score 1: *Product is not used yet.*

Score 2: *An available research plan or similar document outlines usage in research applications;*

Score 3: *A peer reviewed publication exists that describes the usage of the product in a research application;*

Score 4: *The peer reviewed publication under score 3 is cited by peer reviewed publications of other applications;*

Score 5: *The product is used as reference in almost all peer reviewed publication for a specific application;*

Score 6: *The product is used as reference in almost all peer reviewed publication for applications in different research fields, e.g., climate modelling and climate system analysis.*

4.6.2 Decision Support System

As described above under usage for ***Decision Support System*** (DSS) any direct use in infrastructure planning or other business areas such as insurance and indirect support, e.g., through citations in IPCC reports, to decision and policy making in political context, e.g., the Europe 2020 growth strategy is accountable for this minor category.

Table 21*: 6 maturity scores in sub-category* ***Decision support system***

Score	*Description*
1	None
2	Potential benefits identified
3	Use occurring and benefits emerging
4	Score 3 + societal and economical benefits discussed
5	Score 4 + societal and economical benefits demonstrated
6	Score 5 + influence on decision (including policy) making demonstrated

The assessment can be made, for example, as below in case of climate change mitigation and adaptation:

Score 1: *Product is not used yet for this application;*

Score 2: *An available report suggesting that the product can be used for certain decision making applications;*

Score 3: *Product has been used in decision making applications. For example used in studies for impact assessments and a report(s) is available (please provide evidence). This should be available at the data provider's side with some evidence on the user side;*

Score 4: *The results of studies in Score 3 are used for mitigation or adaptation planning. For example, a state or national government report on the planning is available which cites the study using the data set;*

Score 5: The results of studies in Score 3are used in mitigation or adaptation and resulted in societal and economical benefits;

Score 6: Used in for example in national and international climate policy making, for example, Kyoto Protocol.

Note: One can also point to the use of a data record in other applications which has economical benefits such as use by an insurance company for decision making or use in a climate service, e.g., the major application areas mentioned in the WMO Global Framework of Climate Services (agriculture and food security, disaster risk reduction, health and water).

5 APPLICATION PERFORMANCE MATRIX

As mentioned in Section 2 the Application Performance Matrix (APM) attempts to evaluate the performance of an ECV CDR with respect to a specific application. The APM was added to the capacity assessment during the discussions about the SMM because it became clear that the SMM cannot answer the question on how good a data record is for a specific application. To be able to assess suitability of a data record with the APM, user requirements for each considered application are needed to compare the actual technical specifications and validation results to them.

The following sub sections describe the current APM and show one example exercised by the CORE-CLIMAX consortium. It should be kept in mind that the APM is a new tool that has never been used before and is considered to be experimental. The usage of the APM in this workshop shall help to analyse its usefulness (or redundancy) and a potential way for further development.

5.1 Scope and Prerequisites

Figure 3 illustrates the concept of the APM posing a set of typical questions (a query) that a user may ask when a data record is being searched for. Whereas questions towards the spatiotemporal coverage may be easy to answer from the technical specifications of a data record, questions towards results of uncertainty analysis are more difficult and a suggestion on the suitability of a data record for an application may need interaction between the application and data record experts. Key for any suggestion for usage based on this is an understanding of the user requirements for an application. For instance GCOS provides useful requirements for its ECVs which can be used as guidelines for suggestions of data records for applications in climate system analysis. However, a detailed analysis of user requirements per application would be useful to enhance the usability of the APM in the future.

The basic principle of the APM is easy as it evaluates how well the data record's technical specifications and accessible validation results, which should be listed in aProduct Specification Table (PST), match the user requirements for the application considered, which should be listed in a User Requirement Table (URT). The (PST) is a database that consists of all relevant details on the climate data record, such as the technical specifications (e.g. period covered, temporal and spatial resolution/sampling, temporal and spatial coverage, etc) and a summary of validation results (e.g. uncertainty arising from systematic and random effects, temporal stability, etc). The PST ideally is part of a climate data record inventory where all PSTs are coming together.

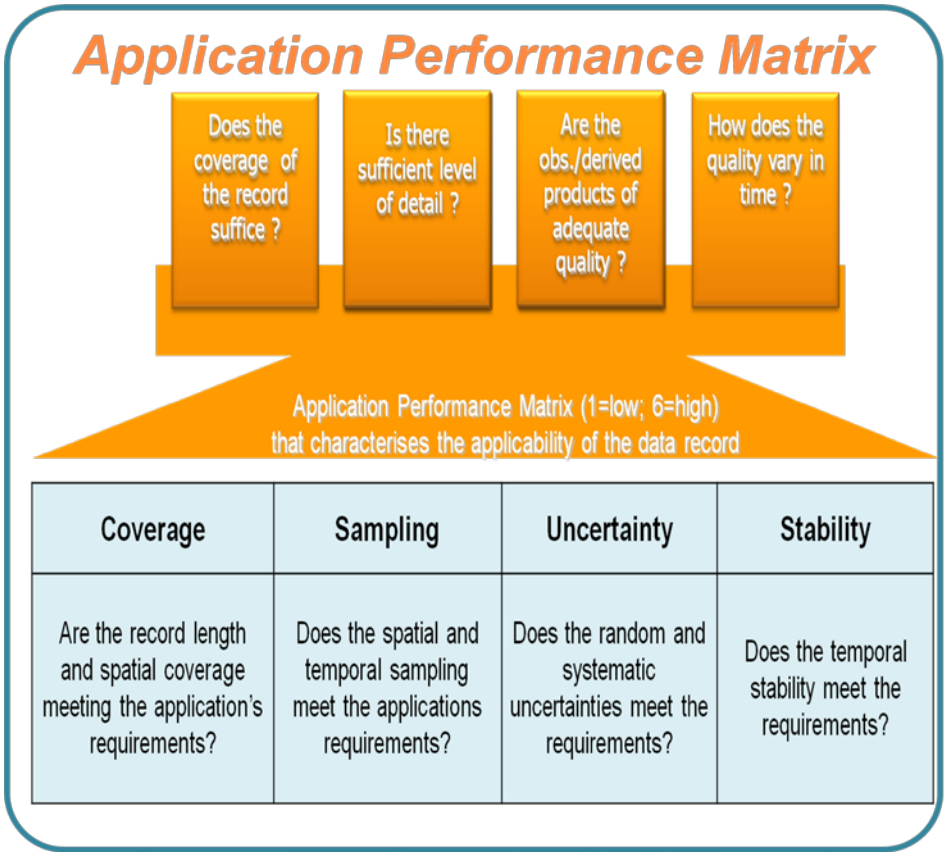


Figure 3: Concept of the new Application Performance Matrix that tries to answer the user question if a data record is suitable for the application in mind.

The User Requirement Table (URT) consists of the user query of requirements expressed in parameters that are provided in the PST. Essentially, the APM evaluation process refers to performing a query on the PST. When such a query is made on several data records simultaneously, the search query result that is returned comprises the APM and a suggestion which data records are most suitable for the application.

In the following we study the APM principle along the example of selecting the better of two data records to be used for the evaluation of a global averaged mid-troposphere temperature trend. The best corresponding GCOS Product is A.3.2 Temperature of deep atmospheric layers. [RD.7] states that the primary benefit of such a data record is the *Monitoring and detection of temperature trends and variability in the troposphere and stratosphere at global and regional scales.*

5.2 APM Scores

For setting up performance scores for the APM we have adopted a common scheme employing three levels of requirements:

- **Threshold** is the minimum requirement to be met to ensure that data record is useful;
- **Target/Breakthrough** is an intermediate level between threshold and optimum/goal which, if achieved, indicates suitability of the data record for the targeted application;

- **Optimum/Goal** is a requirement when reached indicates that the data record is fully suitable for the application.

For each variable a score will be assigned as below:

- **Score 0:** below threshold;
- **Score 1:** matching threshold;
- **Score 2:** matching target/breakthrough;
- **Score 3:** matching optimum/goal.

It is encouraged to develop user requirements having three levels, but if there is only one level as in the current GCOS requirements it is assumed to be the target/breakthrough level. The missing requirement levels are then computed by linearly extrapolating the target requirement in both directions, e.g., by dividing/multiplying it by 2 or by going to the next useful step, e.g. if the target temporal sampling requirement is monthly a useful threshold might be seasonal and a useful optimum is daily.

In order to use the APM for evaluating the performance of a data record for a specific application, two separate pieces of information are necessary – a User Requirement Table (URT) for the application and a Product Specification Table (PST) for the data record(s).

5.3 User Requirements Table

Target user requirements for the GCOS Product A.3.2 available from [RD.7] are given in Table 22. Not mentioned in the GCOS requirements are the needed spatial and temporal coverage or length of record to be used for the trend detection application. As our example is about globally averaged trend the requirement that the data record needs to have global coverage should be kept in mind. It also needs to cover the time of interest.

Table 22: GCOS target user requirement for Product A3.2 Temperature of deep atmospheric layers.

Temporal Resolution	Horizontal Resolution	Vertical Resolution	Accuracy	Stability
Monthly	100 km	5 km	0.2 K	0.02 K/decade

The needed temporal coverage (trend detection time) depends on the relation of accuracy and stability of a data record to the climate change signal, i.e., to the variable that shall be measured. Our example has also been used in [RD.8] and we employ their approach to estimate the needed length of record according. Equation 11 from [RD.8] reads:

$$\Delta t = \left[\frac{12s^2}{m_{est}^2}\sigma_{var}^2\tau_{var}\right]^{1/3}(1+f^2)^{1/3} \quad (1)$$

where Δt is the length of the time series, m_{est} is the estimated trend, $s = m/|\delta m|$ is the signal-to-noise ration of detection with $|\delta m|$ representing the uncertainty of the trend. σ_{var}^2 is an estimate of the natural internal variability of the climate system that is not be associated with the response to a prescribed forcing. $f^2 = (\sigma_{meas}^2\tau_{meas})/(\sigma_{var}^2\tau_{var})$ is the measurement uncertainty factor containing σ_{meas}^2 describing the measurement uncertainty and τ_{var} and

τ_{meas} are correlation times for the natural variability and the measurement uncertainty, respectively. We use the same assumptions made in [RD.8] that are: $\sigma_{var} = 0.18\ K, \tau_{var} = 1.54\ yr, s = 5\ and\ m_{est} = 0.2K/decade$ years for the global average temperature of the 500-hPa surface, consistent with a realistic preindustrial control run of the Met Office's Third Hadley Centre Coupled Ocean–Atmosphere General Circulation Model (HadCM3) [RD.8]. For τ_{meas} we used 2 years as in [RD.8] as assumption for the life time of missions which is on the short side of real mission life times but represents well some of the early NOAA MSU and some of the COSMIC radio Occultation instruments life times. Figure 4 shows results for Δt for three hypothetical data records with three different assumed measurement uncertainties ($\sigma_{meas} = 0.0, 0.06\ and\ 0.2\ K$).

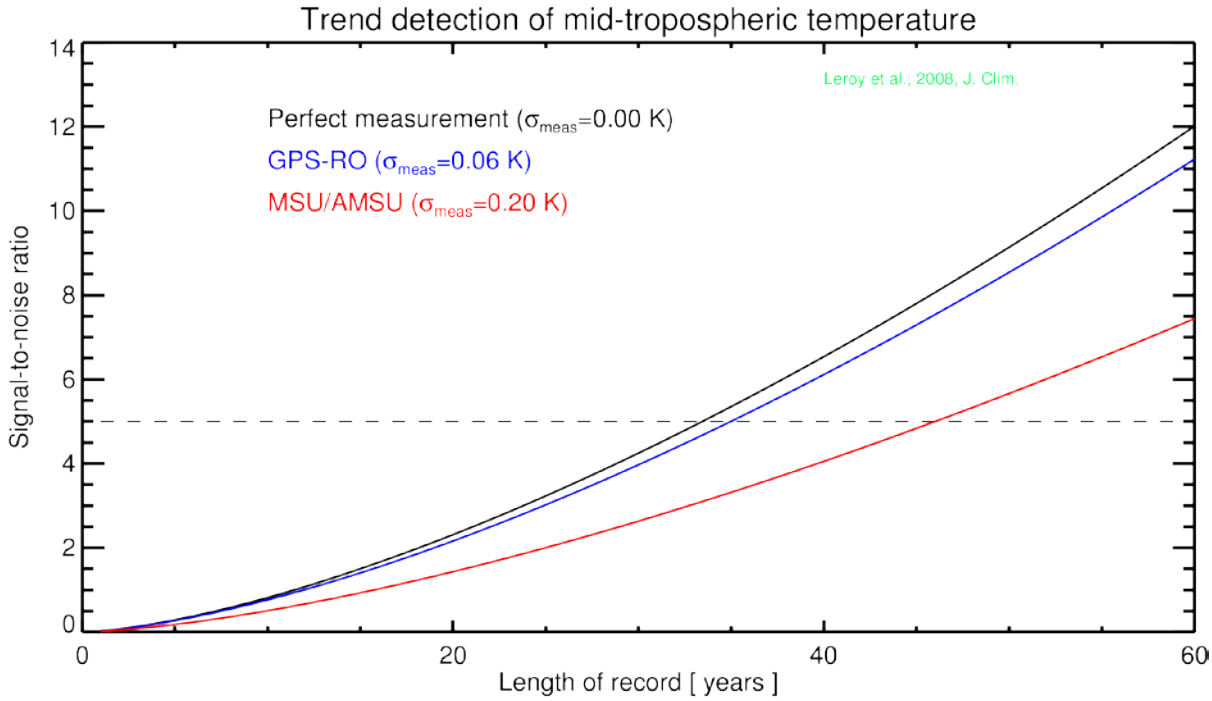


Figure 4: Signal to noise ratio vs. length of record for the temperature trend in the mid-troposphere. The curves represent three hypothetical data records (1) without any uncertainty, (2) radio occultation and (3) microwave sounding.

Because the GCOS requirements have one level we expand them for the example as shown in Table 23. As a co requirement to the accuracy the table also includes the estimated lengths of record at which a trend can be detected with a signal-to-noise ratio of 5, 4 and 3. Signal-to-noise levels can be interpreted in a way that s=5 makes a trend detection almost certain, s=4 will work over most places in the world and s=3 is the minimum level where you may try to use the data record. This can be used to put a relatively short record with low uncertainty into the right perspective.

Table 23: Three level expansion of GCOS user requirements for the example also containing estimates for the needed length of record to detect a trend at signal-to-noise ratio of five.

Requirement Level	Temporal Resolution	Horizontal Resolution (km)	Vertical Resolution (km)	Accuracy (K)	Length of record (yrs)	Stability (K/decade)
Threshold	Seasonal	200	10	0.4	TH 46	0.04
					TA 56	
					OP 65	
Target	Monthly	100	5	0.2	TH 33	0.02
					TA 40	
					OP 46	
Optimum	Daily	50	2.5	0.1	TH 27	0.01
					TA 32	
					OP 37	

5.4 Developing a Product Specification Table

Next we set up a Product Specification Table (PST) as shown in Table 24 using technical specifications of two existing data records for mid-troposphere temperature derived from the MSU/AMSU (www.remss.com) instruments series and from Radio Occultation (www.romsaf.org) instruments. For the accuracy requirement we used the hypothetical values from our example. We acknowledge that uncertainty arising from systematic effects in measurement or analysis is not considered here. However, in reality most of the time anomalies from a mean are used for climate analysis and then our negligence is reduced to assuming that the bias is not varying in time. The stability requirement is not further analysed because the assumption on the bias is already leading to perfect stability.

Table 24: PST for the example using two real data records of mid-troposphere temperatures derived from the MSU/AMSU time series and from Radio Occultation data.

Data Record	Spatial coverage	Temporal Sampling	Spatial sampling (km)	Vertical Sampling (km)	Accuracy (K)	Length of record	Stability (K/decade)
MSU/AMSU	70°S - 82.5°N	Monthly	100	5	0.2	1978-present	Not assessed
GPS-RO	Global	Monthly	250	1	0.06	2006-present	Not assessed

5.5 APM Scores for the Example

Deriving the scores for the APM is achieved by comparing the user requirement and specification tables. This is done by simply assessing the closeness to one of the three requirements, e.g., the RO specification for accuracy is closest to the optimum requirement and thus this score is set to 3.

Table 25: APM scores for the data records from Table 24.

Data Record	Temporal Sampling	Spatial Sampling	Vertical Sampling	Accuracy	Length of Record	Stability
MSU/AMSU	2	2	1	2	1	-
GPS-RO	2	0	3	3	0	-

To provide a suggestion to a potential user on which data record to be used now the scores and some overarching requirements need to be interpreted. As our example is rather an artificial quantity we require global coverage which both data records fulfil. In addition a user would also have a requirement on the time period for which the trend should be assessed. It is clear that if this is the 1920s neither data record can be used.

Table 26: Interpretation of APM scores to make a suggestion to a user.

Importance	*Requirement*	*Factor*	*Score*		*Final score (Score x Factor)*		*Winner*
			MSU	**RO**	**MSU**	**RO**	
1	**Accuracy**	**5**	**2**	**3**	**10**	**15**	**RO**
2	**Length of record**	**4**	**1**	**0**	**4**	**0**	**MSU**
3	**Temporal Sampling**	**3**	**2**	**2**	**6**	**6**	**none**
4	**Vertical Sampling**	**2**	**1**	**3**	**2**	**6**	**RO**
5	**Spatial Sampling**	**1**	**2**	**0**	**2**	**0**	**MSU**

One way of rating the data sets would be by simple summing up the scores. In our case this would end up at a sum of eight for both, which is not very meaningful. A more useful way of looking at it may be to assess first if any score is zero, which would indicate that some performance is below threshold and the data record is most likely not suitable for the application. For the example this would lead to the suggestion that the MSU/AMSU data

record should be used whereas the RO data record shows a high potential but is still a too short record for the application.

Another way of rating the data sets would be give weights to the different columns according to their importance for the application. For the evaluation shown in Table 25 one could simply count the number of categories in which a data record is winning. Here a final result would also be 2-2, but summing the Final scores would lead to a suggestion to use the GPS RO data record first.

5.6 Future Realisation

In the future the PST may be attached to an ECV CDR inventory, such as the CGMS-CEOS-WMO. Such an inventory would need the additional function that a user can provide its own URT and the system would automatically generate the APM scores and maybe also the interpretation table. Key to this would be a very good data base that contains all the information and is maintained by one organisation that is provided with the necessary resources.

A further additional guidance element in the inventory could also be the addition of experiences of other users for the same or similar application. This might be facilitated by so called commentary metadata as being explored in the EU FP7 project CHARMe (http://www.charme.org.uk/).

5.7 Usage in CORE-CLIMAX Assessment

As the APM is at a very experimental stage we hope that the CORE-CLIMAX Assessment can support a discussion of the concept and may bring in different and more elaborated ideas. Each participant can try to follow the outlined approach to establish a URT and PST for an application of a data set analysed with the SMM and then try to assess the performance as described. The columns used in Tables 23-26 may differ according to the application but the principles can be retained. In particular the interpretation part is certainly not brought to a final conclusion.

Acknowledgement

We thank ESA CCI, DWD, and CMSAF for testing earlier versions of the maturity matrices and Chris Merchant, University of Reading and John Bates, NOAA/NCDC for useful suggestions.

APPENDIX A CORE-CLIMAX DATA SET DESCRIPTION

(*General Note:* *This data set description shall not become longer than 5 pages per data set described. Please stay to the most important facts and use tables and bullet lists to provide information where appropriate.***)**

(Type Data Set Name and if available digital identifier here):

1 INTENT OF THE DOCUMENT

(*Provide information on what data set is described and for what application(s) it was created. Keep in mind that the information is targeted at users of any level who wish to use the dataset for climate applications. Users may not be expected to be experts for in situ, remote sensing or reanalysis techniques.*)

2 POINT OF CONTACT

(*Please provide a point of contact: Organisation and Contact details (at least a contact name, organisation and e-mail address)*).

3 DATA FIELD DESCRIPTION

(*Provide a link to an existing technical product specification or provide the information in a form of a table in this document. The specification shall at least include variable names and units (eventually including uncertainty estimates that come with the product), length of record, spatial coverage, spatial and temporal sampling.*)

4 DATA ORIGIN

(*Provide a basic description of the methodology used to derive the product including the input data used and the source (provenance) of the data. Also provide a description of data processing methods such as (inter-satellite) calibration, algorithms employed, homogenization applied, mapping and averaging, etc. If the product makes heavily use of NWP and/or climate model data, e.g., as background fields this should be described as well.*

In case of reanalysis data records please indicate what reanalysis system (coupled or single) has been used and name and version of the model(s).)

5 VALIDATION AND UNCERTAINTY ESTIMATE

(*Provide a summary of validation activities performed for the product and provide a summary of systematic and random uncertainty of the product and how these vary with space, time and state (tabulated form appreciated). In particular information on temporal stability of the data which is an indication of whether the data can be used for longer term variability and trend analysis is appreciated.*)

6 CONSIDERATIONS FOR CLIMATE APPLICATIONS

(*Provide information on the applicability of the product for the planned application (stated in section 1) including limitations. In particular observational products applicable for model evaluation should state the different character when compared to model data. For instance*

for satellite-derived products it is important to describe limitations such as validity in specific areas (e.g., ocean or land only), unresolved diurnal cycles or diurnal cycle aliasing due to orbit drifts for polar orbiting satellites, sampling issues such as in the presence of clouds, sensitivity of the instrument, etc and their respective impacts on the application. For in situ measurements or gridded data sets derived from station data limitations due to the representativeness of the data, etc. and their effect for an application shall be provided.)

7 INSTRUMENT OVERVIEW

(*Provide information on the type of instruments (in situ/remote sensing) used to measure the variable provided including the measurement principle (e.g., infrared emission measured with a spectrometer) and give a description of the instrument science objective, capability, measurement principle, satellite and orbit characteristics or observation location and practice for in situ. Provide the strengths and weaknesses of the instrument measurement. If an instrument simulator is available, provide a short description and references later for details.*

In the case of a re-analysis data set only indicate what instrument data relevant to the parameters considered have been assimilated. This can simply be a link to the information.)

8 REFERENCES

(*Provide a complete list of references used in this document and may provide additional reading references on measurement principles, retrievals, modelling, validation, uncertainty characterisation, product, and applications*.)

9 REVISION HISTORY

(*Indicate the version number of this document, the date of writing and who has edited the document.*)