AN OVERVIEW OF EUROPEAN EFFORTS IN GENERATING CLIMATE DATA RECORDS


The CORE-CLIMAX project has delivered methods and assessments of the capability to provide climate data records, processes for deriving and validating these records, and opportunities to feed back the lessons learned from reanalysis.

The Coordinating Earth Observation Data Validation for Reanalysis for Climate Services project (CORE-CLIMAX; see the appendix for a list of the key acronyms used in this paper) was conducted in support of the establishment of the C3S and other international activities, such as the formulation of the architecture for climate monitoring from space conducted by major space agencies.

A major objective of CORE-CLIMAX was to systematically assess the capacity of ongoing European activities in the area of generation and provision of CDRs. With respect to C3S, the roles of in situ data and model-based reanalysis data need to be considered as well. As the in situ datasets are generally under national responsibility, the CORE-CLIMAX assessment did not cover all in situ data records generated...
The SMM and APM

C3S. CORE-CLIMAX provided means that support provision of CDRs to support the establishment of the sustainability, and what is planned for the future.

periods of time. A quality-controlled process is needed difficulties in sustaining the activities over extended

ASSESSMENT OF EUROPEAN CAPABIL

ARetated on the basis of six maturity levels. An example of the SMM is shown in Fig. 1.

A PM (SGD) is developed that attempts to evaluate the performance of an ECB-E CD R with respect to the user requirements of a specific application, utilizing information provided by the technical specifications of a data record, as well as by the public sources, such as data record quality assessments performed during GEWEX (e.g., Stubenrauch et al. 2013). The most important technical specifications evaluated include quantitative information on i) coverage, ii) sampling, iii) uncertainty characterization, and iv) usability. Analysis

Assessment exercise and results. The SMM and APM were tested during a capacity assessment workshop on 21–23 January 2014 at EUMETSAT headquarters in Darmstadt, Germany, where some 40 participants represented a diverse set of European CDR producers and stakeholders.

on 21–23 January 2014 at EUMETSAT headquarters in Darmstadt, Germany, where some 40 participants represented a diverse set of European CDR producers and stakeholders.

Fig. 1. Example of a CORE-CLIMAX SMM. The colors mark the maturity scores achieved. Maturity scores 1 and 2 denote research capability (i.e., many aspects of the CDR are still under development). Maturity scores 3 and 4 mark initial operations capability, which means the CDR is available to the user community, its usefulness is demonstrated, and decisions need to be made to sustain its maintenance and further development. Maturity scores 5 and 6 indicate full operations capability, where the production of the CDR is in a mature operational environment.

1

by European providers. However, the inclusion of some of the best-known in situ data records in the assessment covered the value of the tools developed in CORE-CLIMAX and encourages more nationally managed data records to be included in a future assessment.

Another objective of CORE-CLIMAX was to coordi- nate the identification of available physical mea- surements, which can be reconciled with previously existing data records, to form long time series. With the GCOS, Copernicus, and ESA CCI projects, along with EUMETSAT and its SAF network, coordination took place with specific efforts to be undertaken by new Copernicus projects to further upgrade their product catalogs to include climate-relevant valida- tion and lay the observational basis for service activities. CORE-CLIMAX identified the integration of CDRs into the reanalysis chain by proposing a feedback mechanism ensuring that the results of the reanalysis process get appropriately reflected into updates of the CDRs and proposed and evaluated methods for intercomparing different reanalyses.

The present paper is structured to report on the four main achievements of the project: 1) synthesize European activities in generating ECB-E CDRs 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; 2) propose a quality-controlled process for evaluating the accuracy of ECB-E CDRs and rate the quality of products with respect to applications, ensuring that the results of the reanalysis process get appropriately reflected into updates of the CDRs, and 4) propose a process for comparing reanalyses.

ASSESSMENT OF EUROPEAN CAPABIL

ITY TO PROVIDE ECB CLIMATE DATA

Fig. 1. Example of a CORE-CLIMAX SMM. The colors mark the maturity scores achieved. Maturity scores 1 and 2 denote research capability (i.e., many aspects of the CDR are still under development). Maturity scores 3 and 4 mark initial operations capability, which means the CDR is available to the user community, its usefulness is demonstrated, and decisions need to be made to sustain its maintenance and further development. Maturity scores 5 and 6 indicate full operations capability, where the production of the CDR is in a mature operational environment.
Fig. 2. Decomposition of the logical view of the architecture for climate monitoring from space (Dowell et al. 2013).

Aspects. The details of this structured process for deriving CDRs are given in Schulz (2015).

Figure 2 shows the decomposition at the first level. Boxes A2, A3, and A8 represent the three different activities that create data records for different applications. Box A3 shows the classical approach to creating long-term CDRs, which are mostly used to study decadal variability and trends (box A7).

The structure of this logical view has the major advantage that for each of the boxes, responsible organizations can be identified. For instance, the creation and maintenance of CDRs (functions A2 and A3) involve many activities including the creation of FCDRs that are best performed by space agencies operating the specific satellite sensor. In contrast, the generation of higher-level climate information records, such as climate indices, or the number of storms that make landfall in a certain region, often need the combination of both CDRs originating from spaceborne and ground-based systems, as well as modeling components (e.g., through some form of reanalysis). Thus, such activity might be best placed in an organization that combines several sources of information. This includes reanalysis centers, but also climate service centers, and meteorological and environmental agencies.

The actual creation and maintenance of CDRs (function A3) can be further decomposed (as shown in Fig. 3) into the following components: sensor corrections and geo-localization, calibration and inter-calibration, conversion to a geophysical parameter, validation versus requirements, archive and access to data, and a user feedback mechanism (together with the documentation and review processes). This view contains many subcategories of the maturity matrix that describe the best practice for the indicated activities. Detailed descriptions of these lower-level processes can be found in Schulz (2015).

A HARMONIZED APPROACH FOR CDR VALIDATION. The generation of ECVs needs to put strong emphasis on the generation of fully described error-characterized and consistent data products. For example, generation of many ECVs requires ancillary information about the state of the atmosphere (e.g., cloud screening for SST, atmospheric correction for spaceborne sensors). As such, the consistency between the various ECV products (e.g., cloud or sea ice masks) extends to ensuring the consistency in the approaches of CDR generation. The validation process, as an integral component of the CDR processing chain, is critical to ensuring such consistency.

A need for a generic validation strategy. The WGCV defines validation as the process of assessing, by independent means, the quality of the data products derived from satellite observations. This so-called product validation ensures that the quality of the products is properly assessed, through quantification of the uncertainties in both the data itself and the measurement system deployed for generating the data. Hence, validation can be considered an end-to-end process that encompasses the entire system, from sensor to product.

As the purpose of calibration and validation is to provide confidence in the quality of the FCDRs/TCDRs and their uncertainties, information is required on how the calibration and validation activities are carried out, at each stage of the data process chain. To establish the traceability of a validation process, one first needs to document each step of the validation process. After an analysis of current validation practices in Europe for space-based ECV CDRs, a generic validation strategy was developed in CORE-CLIMAX (Zeng et al. 2014a).

Starting with documenting the generation of so-called reference datasets, the generic process then requires the assessment of the independence level of the product with respect to such datasets (see Fig. 4). Using the quality-controlled reference datasets, the data producer can make a self-assessment. A complete validation process should also consider the internal consistency of the validated CDRs (e.g., comparing the physically interrelated variables). Ideally, independent assessments should be carried out, subject to external review. Finally, qualification of the independence levels of the independent assessment and the external review are required. The last step consists of allowing future repeats, by sustaining the established validation facilities and procedures. This repeat step implies achieving an operational level of service, so that validation activities and data releases are regularly/routinely implemented.

Importance of consistency validation. The CDS of C3S plans to include ECV CDRs and associated uncertainties, information is required on how the calibration and validation activities are carried out, at each stage of the data process chain. To establish the traceability of a validation process, one first needs to document each step of the validation process. After an analysis of current validation practices in Europe for space-based ECV CDRs, a generic validation strategy was developed in CORE-CLIMAX (Zeng et al. 2014a).

Starting with documenting the generation of so-called reference datasets, the generic process then requires the assessment of the independence level of the product with respect to such datasets (see Fig. 4). Using the quality-controlled reference datasets, the data producer can make a self-assessment. A complete validation process should also consider the internal consistency of the validated CDRs (e.g., comparing the physically interrelated variables). Ideally, independent assessments should be carried out, subject to external review. Finally, qualification of the independence levels of the independent assessment and the external review are required. The last step consists of allowing future repeats, by sustaining the established validation facilities and procedures. This repeat step implies achieving an operational level of service, so that validation activities and data releases are regularly/routinely implemented.

Importance of consistency validation. The CDS of C3S plans to include ECV CDRs and associated uncertainties, information is required on how the calibration and validation activities are carried out, at each stage of the data process chain. To establish the traceability of a validation process, one first needs to document each step of the validation process. After an analysis of current validation practices in Europe for space-based ECV CDRs, a generic validation strategy was developed in CORE-CLIMAX (Zeng et al. 2014a).

Starting with documenting the generation of so-called reference datasets, the generic process then requires the assessment of the independence level of the product with respect to such datasets (see Fig. 4). Using the quality-controlled reference datasets, the data producer can make a self-assessment. A complete validation process should also consider the internal consistency of the validated CDRs (e.g., comparing the physically interrelated variables). Ideally, independent assessments should be carried out, subject to external review. Finally, qualification of the independence levels of the independent assessment and the external review are required. The last step consists of allowing future repeats, by sustaining the established validation facilities and procedures. This repeat step implies achieving an operational level of service, so that validation activities and data releases are regularly/routinely implemented.

Importance of consistency validation. The CDS of C3S plans to include ECV CDRs and associated uncertainties, information is required on how the calibration and validation activities are carried out, at each stage of the data process chain. To establish the traceability of a validation process, one first needs to document each step of the validation process. After an analysis of current validation practices in Europe for space-based ECV CDRs, a generic validation strategy was developed in CORE-CLIMAX (Zeng et al. 2014a).

Starting with documenting the generation of so-called reference datasets, the generic process then requires the assessment of the independence level of the product with respect to such datasets (see Fig. 4). Using the quality-controlled reference datasets, the data producer can make a self-assessment. A complete validation process should also consider the internal consistency of the validated CDRs (e.g., comparing the physically interrelated variables). Ideally, independent assessments should be carried out, subject to external review. Finally, qualification of the independence levels of the independent assessment and the external review are required. The last step consists of allowing future repeats, by sustaining the established validation facilities and procedures. This repeat step implies achieving an operational level of service, so that validation activities and data releases are regularly/routinely implemented.
publicly accessible to facilitate openness, repeatability, transparency, and scrutiny. An example of applying this generic validation strategy in the Yangtze River basin (Huang et al. 2015; Zeng et al. 2014b) is given in the sidebar.

**Reanalysis Feedback Mechanism to CDR Updates.** Procedure for feeding back improved ancillary data to assist CDR updates. The generation of ECV in the form of CDRs from observing system’s “raw data” uses ancillary data as a representation of environmental components that cannot be estimated from the observed data alone. Such ancillary data often come from reanalysis datasets. After first reviewing the procedures currently in place for providing CDR generation with ancillary data from reanalysis products, the limitations of the current approaches were identified (Poli et al. 2013). Several measures were further proposed to enhance this generation process, in particular by using consistent data from reanalysis products.

Support infrastructure for CDR quality assessment in a reanalysis environment. Use of the reanalysis environment for quality assessment of observational CDRs is set to grow significantly in the coming decades, as more CDRs are becoming available from Earth observation programs gathering present-day observations, and also from efforts to generate historical CDRs. The reanalysis environment provides many opportunities for assessing the quality of observational CDRs, for example, by analyzing a posteriori the information produced by the data assimilation procedure that plays a key role in the reanalysis process. However, a reanalysis-based assessment of observational CDRs depends on having a substantial support infrastructure. The main elements that are accessible to CDR providers, to assist them in making appropriate preparations for receiving maximum benefit from the reanalysis feedback, have been outlined by Tan and Poli (2013). Full benefit is only possible when the observational CDRs are supplemented with appropriate documentation, metadata, and auxiliary tools.

The infrastructure is framed toward iterative assessments, designed to improve CDRs via provision of feedback on the reanalysis that addresses noted deficiencies. Such feedback takes two forms: (i) analytical feedback in the form of assessment reports. A distinction is made between offline assessments (projecting an existing reanalysis into the CDR physical space) and online assessments (ingesting the CDR in the assimilation component of the reanalysis system). Satisfactory offline assessments are typically a prerequisite, together with additional data assimilation developments, for online assessments.

Identifying these infrastructure needs was facilitated by two sources of information: the CORE-CLIMAX visiting program, and the CMUG of the ESA CCI, both of which enabled direct contact with a diverse range of CDR providers, from both the in situ and the satellite-based communities. Some of the overarching considerations are as follows: i) Quantitative feedback at the observation datum level is desired; this can be addressed by projecting reanalysis into the CDR space in both offline and online assessments. ii) Meaningful comparison and statistical analysis requires appropriate clustering of similar observations, segregation of dissimilar observations, and good characterization of fluctuations in quality (especially a lack of temporal consistency); this can be addressed by access to appropriate metadata. iii) Feedback should be provided incrementally and the quantitative feedback should be extensible; this can be addressed by multiple feedback loops and a database approach. iv) There is value in making the quantitative feedback available to third parties; this can be addressed by incorporating a “third-party feedback loop.”

Procedures for feeding back reanalysis results and plans to CDR producers. Recommendations were made by Gregow et al. (2014) on how reanalysis needs are systematically communicated to CDR producers and effectively actioned within CDR update plans. The principal conclusion is that effective communication of reanalysis needs should involve procedures at two levels: i) the peer-to-peer level, which allows communication of feedback from reanalysis results, particularly the results of CDR quality assessment arising from the reanalysis environment, and ii) the synthesis level, which promotes coordination and improves the efficiency of peer-to-peer level communications. For the latter, two types of synthesis situations may arise: i) coordination of efforts between multiple CDR producers and ii) consolidation of feedback from multiple users for formulating a CDR update plan.

By building the underlying technical solutions (e.g., design and implementation of suitable databases and tools), progress will be reflected by increased maturity ratings of the CORE-CLIMAX SMMS. The CORE-CLIMAX project developed a set of procedures for comparing reanalyses and for comparing reanalyses to assimilated observations and CDRs (Gregow et al. 2014b; Kaiser-Weiss et al. 2015). To do so, five categories of comparisons were identified: i) descriptiBefore adding to the intercomparison of diverse products from the same ECV, which has added value on its own, we would like to emphasize that there is also added value in several products with quantified uncertainties for a range of applications. Of particular interest is the CORE-CLIMAX project’s coordinate approach to the intercomparison of ISO procedures when it comes to quality standards concerning CDR/ECV/reanalysis products (i.e., time

**Conclusions and Recommendations.** The CORE-CLIMAX project aimed at substantiating how Copernicus observations and products (from both in situ and space components) can contribute to climate change analyses. CORE-CLIMAX assessed the European capability to provide ECV CDRs, prepared a structured process to validate CDRs, coordinated a harmonized approach for validating ECV CDRs, identified the integration of CDRs into the reanalysis chain, and formulated a process to compare the results of different reanalysis techniques. With respect to the development of CJS, there is a need for a worldwide survey of the current situation between global reanalysis and regional reanalysis, recommendations were reported in Gregow et al. (2016) on the climate data store, sectoral information system, evaluation and quality control, output reach and dissemination, as well as coordination. Detailed descriptions of all above aspects can be found online (www.coreclimax.eu) and in the online supplemental material (https://doi.org/10.1175/BAMS-D-16-0074.2). In addition to the intercomparison of diverse products from the same ECV, which has added value on its own, we would like to emphasize that there is also added value in several products with quantified uncertainties for a range of applications. Of particular interest is the CORE-CLIMAX project’s coordinate approach to the intercomparison of ISO procedures when it comes to quality standards concerning CDR/ECV/reanalysis products (i.e., time
APPLICATION OF CORE-CLIMAX GENERIC VALIDATION STRATEGY IN THE YANGTZE RIVER BASIN

TWS is achieved by balancing pre- precipitation evaporation, and river runoff from satellite observations and in situ observations. The same is also obtained from ERA-Interim data (Dee et al. 2011). Upon comparing these TWS data to the GRACE observations of storage changes (Landerer and Schwensen 2012), we can spot the potential quality issues between different versions of GRACE observations.

Data. The TWS anomaly and cumulative TWS anomaly are estimated from GPCP precipitation (Adler et al. 2003), surface energy balance system (SEBS)-estimated evapotranspiration (Chen et al. 2014), observed discharge, as well as from ERA-Interim data (Dee et al. 2011), which are compared with GRACE TWS (Landerer and Schwensen 2012) for the upper Yangtze River reach and the whole Yangtze River basin. The river discharge measurements from Yichang station for the period 2001–10 are used for the upper Yangtze reach. The discharge measurements from Datiang for the period 2001–10 are used for the whole Yangtze River basin study. The GPCP precipitation data were obtained online (http://jisao.washington.edu/data/precip/), as were the GRACE data (http://grace.jpl.nasa.gov, data version RL05 (DSTvSCS1401)). The GRACE monthly gridded data represent an equivalent water thickness deviation to the average. The GRACE monthly anomaly and cumulative anomaly over Yichang station (i.e., the upper reach), calculated using an earlier version of the GRACE dataset (RL04 ss201008), are different from the results shown in Fig. S5a,b. For the TWS anomaly, the earlier version of the GRACE dataset has a phase difference of about 10 days, when compared to the current version. For the cumulative TWS anomaly, the earlier version did not show the increase in the monthly cumulative TWS anomaly after 2004, which was expected because of the filling of the Three Gorges Dam reservoir (Fig. S5b, top).

Conclusions. The exact explanation for the different results over the upper Yangtze River reach, calculated from the two versions, needs intensive dedicated research. However, the log file for each change made for the production of GRACE data can be accessed easily online (ftp://podac.jpl.nasa.gov/allData/grace/index.jpl), therefore, the first action should be to check if the difference can be attributed to what has been changed during the production of the new version dataset. Another place to look is such information is the Center for Space Research website (www2.car.utexas.edu/grace/RL05.html).

From the above two sources, the text quoted below may explain the possible reasons for the different results over the upper Yangtze River reach, with the two versions of data:

NOTE-I (2012-07-17): We have replaced the CSR-GSM products for the following four months: July 2004; October 2004; March-2005 and February-2006. The GSM products for these four months have been updated with the refined data editing. While the GAC/GAD products for these four months are unchanged, they have been updated as well for consistency. If you downloaded the data products for these four months prior to July 17, 2012, please download the replacement products. No other products previously delivered are so affected.

As reported in the log file, the affected products for the four specific months correspond to the difference in the calculated monthly cumulative TWS anomaly after 2004 between the two versions of the GRACE dataset.

Currently, many CDRs are not yet suitable for use in reanalyses or in climate study applications. The production of consistent high-resolution data records remains a challenge and needs more research urgently. Linking ECVs to essential water variables is a next step for closing the gaps between CDRs and sectoral applications. Also, the closing of other cycles (e.g., carbon and energy) requires more effort.

ACKNOWLEDGMENTS. The CORE-CLIMAX project was financially supported by the European Commission within the Copernicus initiative in its Seventh Framework Programme (FP7 Grant Agreement 313085). The CORE-CLIMAX consortium comprised seven European institutions: Faculty for Geo-Information Science and Earth Observation (ITC), University of Twente; EUMET-SAT; ECMWF; German Weather Service (DWD); VITO, FMI, Météo-France, and two international institutions: the Institute of Tibetan Plateau Research (ITP) and the Cold and Arid Regions Environmental and Engineering
APPENDIX: KEY ACRONYMS USED IN THIS PAPER.

APN Climate Monitoring User Group
BSSN Baseline Surface Radiation Network
C3S Copernicus Climate Change Service
CCI Climate Change Initiative
CDR Climate data record
CDN Climate data store
CEOS Committee on Earth Observation Satellites
CMUG Climate Modeling Users Group
CORE-CLIMAX Coordinating Earth Observation Data Validation for Reanalysis for Climate Services
ECARD European Climate Assessment and Dataset
ECMWF European Centre for Medium-Range Weather Forecasts
ECV Essential climate variable
EQC Evaluation and quality control
ERA-Interim ECMWF interim reanalysis
ESA European Space Agency
EUMETSAT European Organisation for the Exploitation of Meteorological Satellites
FCDR Fundamental climate data record
GLOS Global Climate Observing System
GEWEX Global Energy and Water Cycle Exchanges Project
GPCP Global Precipitation Climatology Centre
GPCC Global Precipitation Climatology Project
GRACE Gravity Recovery and Climate Experiment
HadObD Hadley Centre observational database
ISO International Organization for Standardization
KNMI Koninklijk Nederlands Meteorologisch Instituut
NOAA National Oceanic and Atmospheric Administration
SAF Satellite Application Facility
SMM System maturity matrix
TCDR Thematic climate data record
TWS Terrestrial water storage
UGS U.S. Geological Survey
WGC Working Group on Climate
WGGV Working Group for Calibration/Validation
WMO World Meteorological Organization

REFERENCES


ABSTRACT

The Coordinating Earth Observation Data Validation for Reanalysis for Climate Services project (CORE-CLIMAX) aimed to substantiate how Copernicus observations and products can contribute to climate change analyses. CORE-CLIMAX assessed the European capability to provide climate data records (CDRs) of essential climate variables (ECVs), prepared a structured process to derive CDRs, developed a harmonized approach for validating essential climate variable CDRs, identified the integration of CDRs into the reanalysis chain, and formulated a process to compare the results of different reanalysis techniques. With respect to the Copernicus Climate Change Service (C3S), the systematic application and further development of the CORE-CLIMAX system maturity matrix (SMM) and the spinoff application performance metric (APM) were strongly endorsed to be involved in future implementations of C3S. We concluded that many of the current CDRs are not yet sufficiently mature to be used in reanalysis or applied in climate studies. Thus, the production of consistent high-resolution data records remains a challenge that needs more research urgently. Extending ECVs to close climate cycle budgets (e.g., essential water variables) is a next step linking CDRs to sectoral applications.