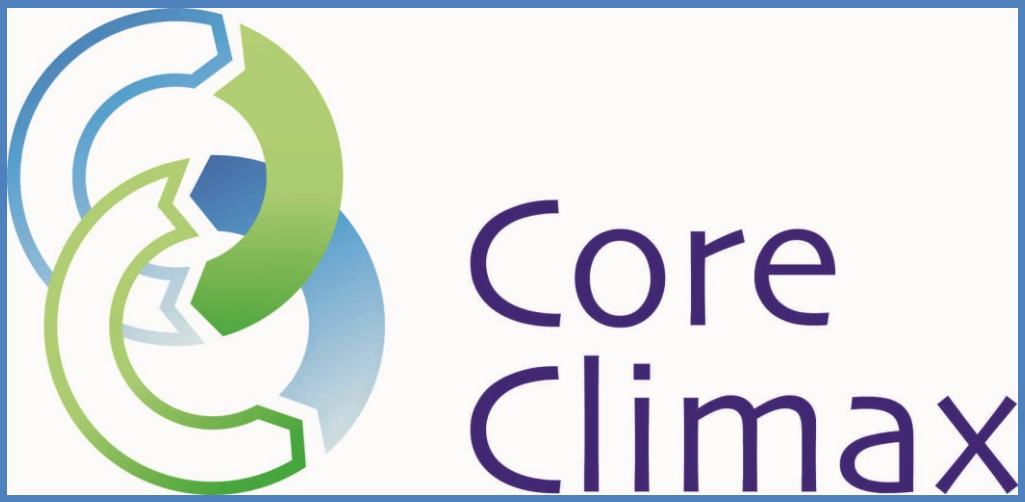
EU FP7
Core-Climax
Deliverable
D4.41



PROCEDURE FOR FEEDING BACK IMPROVED ANCILLARY DATA TO ASSIST CLIMATE DATA RECORDS UPDATES

ABSTRACT—The generation of Essential Climate Variable (ECV) in the form of Climate Data Records (CDRs) from observing systems "raw data" uses ancillary data as a representation of environmental components that cannot be estimated from the observed data alone. The present document first reviews the procedures currently in place for providing CDR generation with ancillary data from reanalysis products. In a second part, the limitations of the current approaches are pointed out. Finally, the document proposes a range of measures that could improve this generation process by using improved and consistent data from reanalysis products.

Document history

Contributors	Organization	Action	Date
Paul Poli	ECMWF	Draft outline created	January 2013
Paul Poli	ECMWF	Detailed outline finalized	June 2013
David Tan	ECMWF	Contents draft produced	October 2013
David Tan	ECMWF	Contents first version finalized	November 2013
Viju John	EUMETSAT	Comments	November 2013
Joerg Schulz	EUMETSAT	Further comments (incl. separate L1 and L2 user communities)	December 2013
Paul Poli	ECMWF	Integration and final version	December 2013

Table of Contents

I. Introduction

The Global Framework for Climate Services recognizes the need to deliver climate information/services to a diverse range of users spanning multiple societal sectors. This is a complex undertaking that involves acquisition/preservation of "raw data", subsequent processing to generate "higher level products", namely datasets meeting the quality criteria associated with Essential Climate Variable (ECV) products and Climate Data Records (CDRs), and further activities to provide climate services (including interpretation of data and their uncertainties) to users.

Figure 1 attempts to summarize the chain of processes that play a role in such delivery. It is worth noting that acquisition of raw data is typically a multi-decadal endeavour, and that rescue/preservation/re-processing/stewardship of historical data remain vitally important activities. For example, some of the temperature, pressure, and wind measurements collected by ship observers in the 1800s have only been put to use in climate databases in recent years, after efforts to rescue and digitize the original hand-written records. Another example is given by several satellite datasets originally acquired not for climate applications but rather for numerical weather prediction. The original satellite data were acquired using a succession of instruments over several decades and are often not directly suitable for climate applications; however, the measurements characteristics are now sufficiently well understood for the data to be re-processed and/or assimilated in climate reanalyses which are updated daily and distributed to thousands of users. The great spread between these two examples illustrates

that the processes to generate climate data records vary significantly and are not fully reflected in a single example.

Note, the results in this document draw from expert visits carried out in the framework of the Core-Climax project.

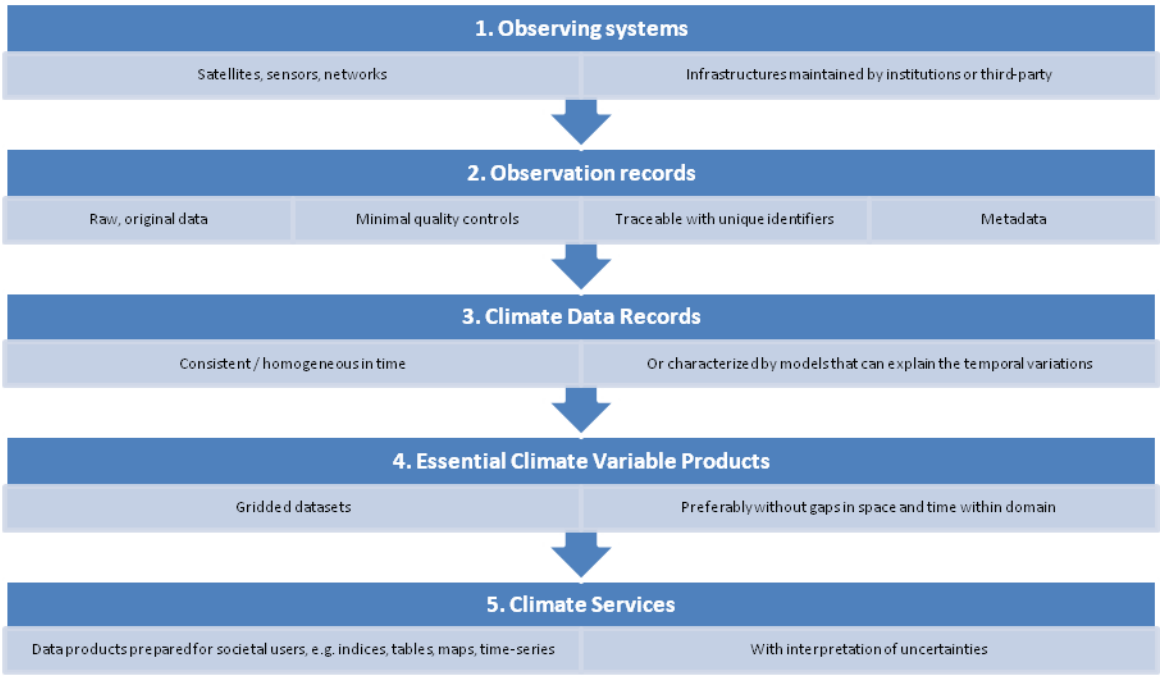


Figure 1: Provision of climate services from observing systems. The arrows represent the processes involved in advancing along the 5 steps. Not shown, but to be kept in mind, is the iterative nature of each process.

In Figure 1, the process to advance from Step 2 to 3 can be detailed further, still retaining general concepts. The input information to this process is made of observation records, but also of instrument characterization and meta-data, as well as ancillary information about components of the environment to which the observed data are sensitive to, but for which the input observed data aren't sufficient to get a complete picture. Relevant examples include:

- Ancillary information may be used to locate properly in situ atmospheric data in the vertical: for example a measurement of temperature with respect to height can be located in pressure if a vertical profile of pressure is provided as ancillary input;

- Ancillary information may be used to assist in the retrieval of geophysical quantities from remote sensing measurements, for example to assess carbon dioxide concentration from an infra-red spectrum, given ancillary temperature information;

- Ancillary information may be used to assist in the quality control of climate data records, for example to flag as unreliable measurements of sea-surface temperature from infra-red sensors in cloudy areas, or to flag as less accurate humidity observations from radiosondes in the case of very low temperatures.

The present document reviews the current procedures (in sections II. and III.) and then (in section IV.) proposes a range of measures that could improve the procedure for providing CDR generation with improved and consistent ancillary data from reanalysis products.

II. Acquisition/use of ancillary data in CDR updates: current practices

Figure 2 shows a schematic stratification of the community that makes use of reanalysis data as ancillary data in their CDR updates. The community is sizeable and diverse in nature, and other stratifications could be relevant in other contexts. The stratification here has been chosen to reflect differences in the practices/procedures of the different sub-communities.

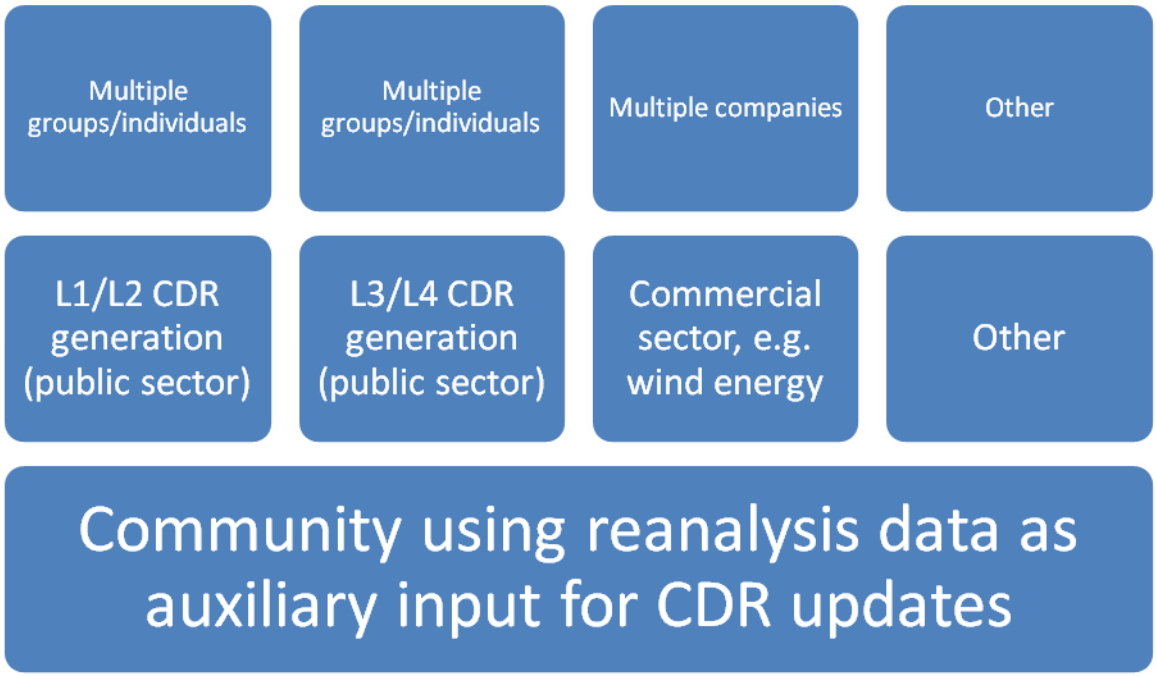


Figure 2: Users of reanalysis data as ancillary input to their processes of CDR generation

Table 1 details how these sub-communities handle different aspects of reanalysis product acquisition/use. Note, in the following, the L1 and L2 CDR generation are split because these are very different groups: the sizes of communities differ (there are many groups involved in L2 work but rather few doing L1), and L1 is rather a specialized activity, typically at space agency level. For the commercial sector, we take the example of the wind power energy sector, which was the topic of one expert visit carried out in the project (European Wind Energy Association, EWEA).

Aspect of CDR generation regarding ancillary reanalysis data	L1 CDR generation	L2 CDR generation	L3/L4 CDR generation	Commercial wind energy
Size of the European community (note: numbers are indicative, mostly for comparison between the communities)	Around 50 groups (counting 1 group per instrument team), 200 individuals	Around 500 groups, 2000 individuals	Around 40 groups, 150 individuals. For the purposes of the current document, Regional reanalysis is treated within this category.	Around 50 companies, typically working within constraints of commercial confideniality
Examples of programmatic coordination at institutional level	Each space agency generates L1 for its own sensors. Few like EUMETSAT have a programmatic coordination for L1 CDRs.	EUMETSAT etc	• ESA CCI (subset of GCOS ECVs) • EURO-4M/UERRA (regional reanalysis)	Not confirmed. The EWEA (European Wind Energy Association, http://www.ewea.org/) “is the voice of the wind industry, actively promoting wind power in Europe and worldwide. It has over 700 members from almost 60 countries making EWEA the world's largest and most powerful wind energy network.” Not all industrial communities are as well organized.

Aspect of CDR generation regarding ancillary reanalysis data	L1 CDR generation	L2 CDR generation	L3/L4 CDR generation	Commercial wind energy
Typical characteristics of the ancillary information required from reanalysis products	• Forward simulations of instruments employing reanalysis fields. • Mostly used in detection of potential instrument issues and for evaluation of developed corrections	• Range of geophysical parameters (commonly temperature, humidity, wind, cloud properties), space-/time-collocated with the satellite viewing geometry	• As per L2 CDR Generation • Range of geophysical parameters, at the boundaries and/or initial times of the regional reanalysis	• Range of geophysical parameters, with spatio-temporal variability representative of the micro-scale fluctuations experienced at a wind-energy production site
Coordination between institutional programmes	Provided via CGMS in activities such as GSICS and SCOPE-CM	Not confirmed.	Not confirmed.	Not confirmed.
Means of acquisition of ancillary reanalysis data	• Bilateral exchange has been used in the past. This will become unwieldy as L1/L2 reprocessing programmes intensify/proliferate. Pilot projects to		• Semi-automated "public acquisition", with reanalysis providers giving	• Usually automated "public acquisition"

Aspect of CDR generation regarding ancillary reanalysis data	L1 CDR generation	L2 CDR generation	L3/L4 CDR generation	Commercial wind energy
	adopt automated "public acquisition" are underway		scientific/technical support to multiple groups/individuals	
Strengths in current practices for acquisition/use of ancillary reanalysis data	• Usage is at infancy stage; • Knowledgeable about interfacing to reanalysis product formats. • Knowledgable about collocation/sub-setting tools.	• Conversant in the geophysical parameters available in reanalysis products. • Knowledgeable about interfacing to reanalysis product formats. • Knowledgable about collocation/sub-setting tools.	• Some expertise in interfacing to reanalysis product formats and use of collocation/sub-setting tools.	• Use: Expertise in assessing representativeness of reanalysis products to describe a given site • Acquisition: Versatility in adapting to whatever format the reanalysis data are given
Weaknesses in current practices for acquisition/use of ancillary reanalysis data	• Forward simulations are not part of the 'normal' reanalysis	• The wealth of information available in the feedback files are not easily	• Absence of coordination between groups within a single programme; for	Cost of adapting software to a given data format sometimes considered more important than the benefits of improved data quality

Aspect of CDR generation regarding ancillary reanalysis data	L1 CDR generation	L2 CDR generation	L3/L4 CDR generation	Commercial wind energy
	products or the feedback; • The wealth of information available in the feedback files are not easily accessible;	accessible.	example several groups may consider the same satellite instrument but for different ECVs, leading to separate overlapping requests for auxiliary data and potential inconsistencies in subsequent processing.	
Needs for reanalysis data that are currently being met		Global coverage for the satellite-era	Global coverage for the satellite-era and for regional reanalysis boundary/initial conditions	Global coverage for at least the past 20 years.
Needs for reanalysis data that are currently not being met	• More frequent temporal sampling, desirable to have hourly sampling as standard. • Availability of uncertainty estimates.		• Ensembles of global reanalyses, to provide multiple realizations of boundary/initial	• More frequent temporal sampling, desirable to have hourly sampling as standard. • Increased vertical resolution in

Aspect of CDR generation regarding ancillary reanalysis data	L1 CDR generation	L2 CDR generation	L3/L4 CDR generation	Commercial wind energy
	This could be achieved to some extent by availability of multiple realisations.		conditions for ensembles of regional reanalyses • More information about random and systematic uncertainties in reanalysis products • Pooled user knowledge about the strengths and weaknesses of different reanalysis products • Improved technical infrastructure for large data volumes and faster transfer rates (in preference to dedicated data preparation by reanalysis providers)	the altitude range 0-1 km • Timeliness: need to have reanalysis data continued into the present delivered within a few days rather than a few months
Opportunities for	• Provide either	• Provide tools for	• Provide common	• Develop a mechanism for

Aspect of CDR generation regarding ancillary reanalysis data	L1 CDR generation	L2 CDR generation	L3/L4 CDR generation	Commercial wind energy
development	the forward simulations directly or an easy framework for simulating satellite observations from reanalysis outputs. • Provide tools for easy acquisition and analysis of feedback files.	easy acquisition and analysis of reanalysis and feedback files.	tools for scientific preprocessing of reanalysis products, e.g. downscaling and sub-sampling • Develop forums to collect/share user experience • Implement benchmark indicators of reanalysis performance and quality • Leverage on "well-informed" users who are able to advise "less-informed" users	wind-energy community to communicate their evaluations of reanalysis products without compromising commercial confidentiality. One way would be to develop the wind power resource assessment intercomparisons to include several reanalyses and also more diverse observational CDR datasets.

Table 1: Aspects of CDR generation across communities regarding use of reanalysis products as ancillary input

III. Advantages and limitations of the current approaches

During the course of this Core-Climax workpackage, we identified a number of areas of current practice which, in our opinion, raise issues worthy of attention. These are tabulated below. Proposals for how to address the issues are taken up in the next section.

Current practice	Considera-tions	Comments	Issues identified
Some CDR products do not specify which information from which reanalysis was used as ancillary input	Scientific/ Technical	Blame lies partly on the reanalysis producers themselves, for they do not properly identify/tag their products in the metadata	• Need for Algorithm Descriptions and Input Data Specifications • Traceability of reanalysis data being used as ancillary input to CDR generation
CDR generators develop their own handling tools of ancillary data	Scientific/ Technical	Sometimes leads to duplication of effort and sometimes to inconsistency with other CDRs	• Data formats • Need to develop/adopt common toolboxes. Some functionalities exist, e.g. the CDO package, but more are required, and on a sustained basis
Ad-hoc arrangements to access reanalysis data for use as ancillary input to CDR generation	Technical	Ad-hoc arrangements are flexible, but are time-consuming at best and can be showstoppers at worst.	Access to reanalysis data for use as ancillary input to CDR generation
Use of reanalysis data as ancillary input could be better-informed	Scientific	Difficult for CDR generators to take into account the strengths and weaknesses of a particular reanalysis.	Informed use of reanalysis data as ancillary input
Production schedules are not coordinated	Program-matic	Not a problem as long as each process can be iterated, and there is always a “next opportunity” to take onboard or generate a CDR	Better visibility of reanalysis and CDR production schedules would help one another to align their proposals with each other’s application (feeding CDR into reanalysis and vice-versa, feeding reanalysis ancillary data into CDR)

Table 2: Summary of the current practices and issues in feeding ancillary data to CDR generation

IV. Proposed mechanisms to improve procedures for acquisition/use

Following on from the Table of the previous section, we now turn to the issues arising from our review of current practice for feeding back improved ancillary data to assist climate data records updates. These are transferred to the Table below. For each item, we propose a mechanism to address the issue, and propose the groups/individuals who could be approached to take further action.

Issue	Category	Proposed Mechanism	Actions for	Comments
Traceability of reanalysis data being used as ancillary input to CDR generation	Technical	Reanalysis datasets to have DOIs	Reanalysis producers	Urgent need to define in a coordinated way the granularity of the DOI (e.g., per geophysical variable, or per reanalysis)
Data formats	Technical	Provide data in a variety of well-established formats Provide format conversion tools	Reanalysis producers, CDR Generators and/or User Community	
Access to reanalysis data for use as ancillary input to CDR generation	Technical	Further development of "public access", retain bi-lateral exchange where appropriate	Reanalysis producers in consultation with CDR Generators and User Community	
Informed use of reanalysis data as ancillary input	Scientific	Dialogue between reanalysis producers, "scientific" users, and climate service providers.	Reanalysis producers, scientific users, climate service providers	Dialogue to be supported by Quality Assessments.
Coordination of production schedules	Programmatic	Effective communication channels between stakeholders	Reanalysis providers in consultation with CDR Generators and User Community	Stakeholder representatives (CEOS, WCRP, ?)

Table 3: Solutions proposed to improve CDR generation through consistent and well-identified ancillary reanalysis data