DESIGN OF SUPPORT INFRASTURCURE FOR CDR QUALITY ASSESSMENT IN A REANALYSIS ENVIRONMENT

EU FP7
Core-Climax
Deliverable
D4.42

ABSTRACT— Reanalysis-based assessment of observational CDRs depends on substantial support infrastructure. The purpose of this document is to explain the main elements in a way that is accessible to CDR providers, to assist them in making appropriate preparations for receiving maximum benefit from the feedback that the reanalysis environment can provide.

Document history

Contributors	Organization	Action	Date
David Tan	ECMWF	Draft outline created	June 2014
David Tan & Paul Poli	ECMWF	First circulated draft	18 October 2013
David Tan	ECMWF	Version 1.0	31 October 2013

Table of Contents

Design of support infrastructure for CDR quality assessment in a reanalysis environment

I. Introduction

A key strength of the reanalysis environment is that it provides many opportunities for assessing the quality of observational climate data records (CDRs). These opportunities arise from the data assimilation techniques that are central to the reanalysis process (and which are developed in close collaboration with colleagues working in numerical weather prediction, NWP).

Use of the reanalysis environment for quality assessment of observational CDRs is set to grow significantly in the coming decades. New and more varied CDRs are becoming available from Earth observation programmes gathering present-day observations, and also from efforts to generate historical CDRs (through a combination of data recovery and reprocessing). These CDRs contain information on an increasing range of Essential Climate Variables, ranging from traditional atmospheric parameters (e.g. temperature, wind, precipitation) to atmospheric composition (aerosol, long-lived & reactive gases) to other properties of the Earth-system (soil moisture, sea-ice, ocean colour). Much of the activity is aligned with the Global Framework for Climate Services and the European Union's Copernicus Programme for Earth Observation (and particularly its Climate Change service). Use of the reanalysis environment for CDR assessment is not limited to Thematic CDRs (TCDRs), for which the product parameters are geophysical parameters derived from in-situ measurements or from remotely-sensed Fundamental CDRs (FCDRs, for which the product parameters are quantities more closely related to the sensing instrument, e.g. calibrated radiances, brightness temperatures): given appropriate auxiliary tools, such as observation operators, the assessment is possible for the FCDRs themselves.

Reanalysis-based assessment of observational CDRs depends on substantial support infrastructure. The purpose of this document is to explain the main elements in a way that is accessible to CDR producers, to assist them in making appropriate preparations for receiving maximum benefit from the feedback that the reanalysis environment can provide. Full benefit is only possible when the observational CDRs are supplemented with appropriate documentation, metadata and auxiliary tools. It is sensible to distinguish two distinct but related needs:

Section II: infrastructure needed to assess CDR quality by comparison with existing reanalyses,

Section III: infrastructure needed to assess CDR quality by assimilation into a new reanalysis.

The distinction reinforces the recognition that a prerequisite for assimilation into a new reanalysis (Section III) is that the CDR has already demonstrated promising quality under other measures - for example through diagnostics derived from Section II or other "off-line" assessments.

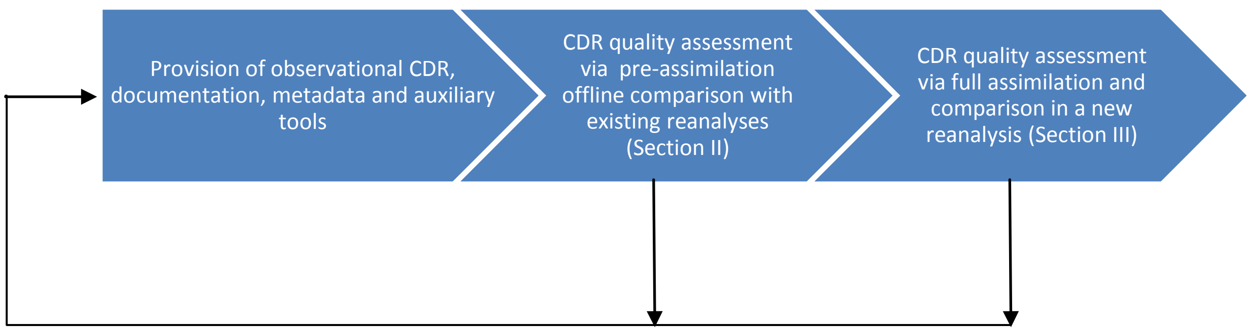

Figure 1: Progress of an observational CDR to quality assessment via full assimilation in a new reanalysis. Progress is contingent upon satisfactory assessment in pre-assimilation/offline comparisons. The arrowed lines represent the feedback loop for stimulating improved CDRs and/or supplementary materials (documentation, metadata, auxiliary tools etc).

It is important to recognize that the progress of an observational CDR through quality assessment in the reanalysis environment (Figure 1) is invariably an iterative process. The feedback loops, depicted by the arrowed lines, serve to resolve issues arising within the assessment steps. Several different types of feedback arise, ranging from (i) quantitative information about observation-reanalysis differences, (ii) requests for new observational CDRs, and (iii) requests for additional documentation, metadata and auxiliary tools.

II. Assessing CDR data quality by comparison with existing reanalyses

In this section we describe the infrastructure needed for assessing CDR quality by comparison with existing reanalyses. Such "off-line" assessment is invariably a prerequisite for assessing CDR quality by assimilation into a new reanalysis (Section III).

A key source of information for identifying the infrastructure needs described here has been the experience gained through the Core-Climax programme of visits that facilitated direct contact between ECMWF and a diverse range of CDR providers from both the in-situ and satellite-based communities. A second, complementary, source of information has been the experience of ECMWF's Renanalysis section in the Climate Modelling Users Group (CMUG) of the ESA Climate Change Initiative (CCI). The recurring/prominent issues that have shaped the designed infrastructure are as follows:

- quantitative feedback at the observation datum level is desired - addressed by projecting reanalyses into the CDR space
- meaningful comparison and statistical analysis requires appropriate clustering of similar observations, segregation of dissimilar observations, and good characterization of fluctuations in quality (especially lack of temporal consistency) - addressed by access to appropriate metadata
- there is value in making the quantitative feedback available to third parties - addressed by incorporating a "3rd party feedback loop"

Realizing and exploiting the infrastructure described here will benefit from improved global co-ordination and capacity-building in a number of areas - discussion of these is deferred until Sections IV/V.

Figure 2 depicts a systematic process for projecting an existing reanalysis into the space observed by a CDR, followed by comparison, assessment and feedback to the CDR provider. The projection (sometimes referred to as "forward simulation" or computation of "observation-equivalents") depends on acquisition of three principal inputs: (i) CDR observational data and associated metadata, (ii) a suitable reanalysis, and (iii) tools for reading the CDR and for applying an "observation operator" to the reanalysis. The projected reanalysis, when combined with the observational CDR, constitutes "Observational Feedback Data" that forms the quantitative basis for the CDR assessment.

As mentioned in the Introduction, the reanalysis environment is increasingly being used for assessing CDRs pertaining to a wide range of earth-system parameters. While the process depicted in Figure 2 is formulated to be generic and applicable in principle to all CDRs, implementation for specific CDRs may still require customization. We nevertheless advocate the use of common methodologies wherever possible, in order to maximize efficiency through use of common tools. We have already indicated that off-line assessment of a CDR may lead to on-line assessment resulting in further feedback on the same CDR (as formulated in the next Section). Thus, the ability to add feedback incrementally is also valuable.

In this regard, we recommend a database approach for archiving the Observation Feedback Data - see Appendix 1. Such an approach is proving effective for the CDRs we have been able to consider through the set of Core-Climax visits to/from providers of historical in-situ upper-air data, satellite-based TCDRs for wind velocity, and satellite-based FCDRs for calibrated radiances containing temperature information.

It is worth noting that Observation Feedback generated and archived for one reanalysis and one CDR is re-usable for other CDRs that are sufficiently similar to the first. Such situations are likely to arise when CDR reprocessing activities generate new CDRs that have minimal changes to the observation geolocations and in which the major changes are in the estimates of the observed parameter.

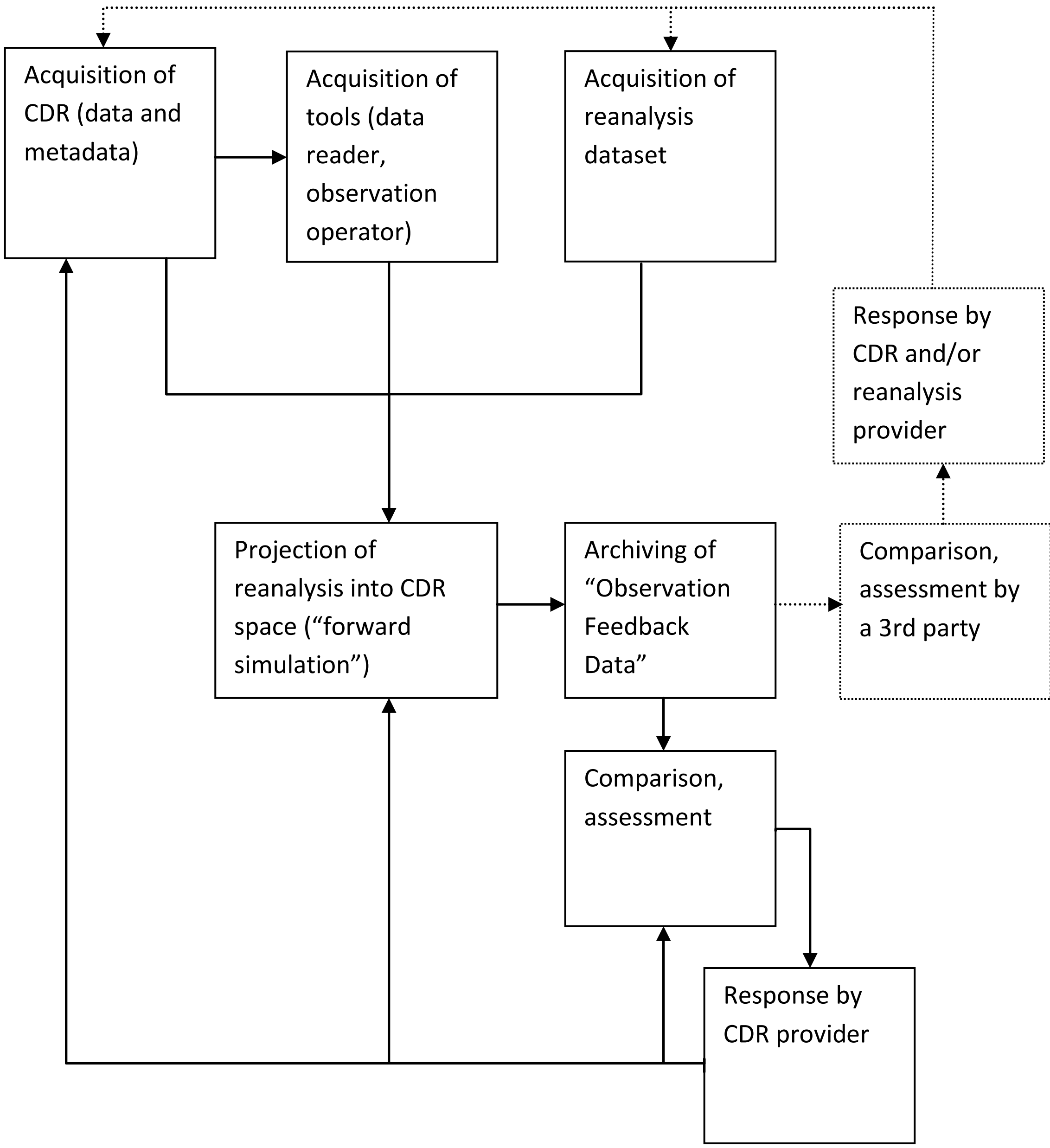

Figure 2: Process for comparing a CDR with a pre-existing reanalysis ("pre-assimilation, off-line assessment"). Projection ("forward simulation") typically involves a collocation step.

The comparison/assessment step generally involves a combination of qualitative and quantitative evaluation of differences between the CDR values and their projected reanalysis counterparts. The evaluations are invariably statistical in nature and seek to characterize and understand the reasons for the differences. They thus involve subsetting/aggregating/filtering the differences according to various dimensions. The importance of temporal consistency means that timeseries analysis features prominently, particularly detection of breakpoints (sudden jumps) and/or artificial drifts. Differences can arise for many reasons and are not always associated with deficiencies in the CDR. Careful consideration must be given to the contributions from deficiencies in the reanalysis and the observation operator. Consequently, the comparison/assessment step needs substantial interaction between all parties. Most of this interaction is more suited to normal scientific dialogue/exchange and a process of informal feedback and response, prior to formal feedback in the form of assessment documents which should be reserved for consolidated conclusions. (It can nonetheless be helpful for the formal feedback to document the lessons learned from the informal feedback.)

Experience has shown that interaction is dominated by a number of issues that arise repeatedly. Among these are:

<u>CDR acquisition issues:</u>

- queries about CDR content: CDR documentation may need clarification
- unintended instances of missing data: may stimulate more rigorous quality control procedures within CDR generation

<u>Reanalysis projection issues:</u>

- interfacing observation operator with reanalysis data: can be a substantial effort in the absence of heritage

<u>Comparison issues:</u>

- developing appropriate labelling of data for archiving/book-keeping to facilitate scientific assessment and characterization of the data

Adopting a process of continual improvement offers good prospects for decreasing the frequency/severity of such issues, thereby reducing the overall time required to assess a CDR. It is often desirable for detailed scientific analysis/feedback to be preceded by technical feedback on a subset of the CDR. This is likely to be advisable when there is little heritage for resolving interfacing issues or for establishing the suitability of the CDR metadata. When CDR generation is computationally expensive and/or time-consuming, there may be merit in making such technical feedback a part of CDR prototyping prior to full-scale production.

Turning to the formal documented feedback, we offer below some guidance on the content but stop short of being prescriptive about the scope and level of detail that should be included. We consider it good practice to clarify mis-understandings quickly and with minimal formality where possible. This could be realized by a two-stage reporting loop as follows:

The 1st feedback report would address issues such as

- spatio-temporal coverage
- range checks on the data and metadata
- statistical summaries (trends, anomalies, mean differences, ...)
- first set of scientific/technical questions to CDR provider

Upon receipt of answers from CDR provider, the assessment would then proceed from a better-informed basis, reducing the risk of incorrect interpretations.

The 2nd feedback report would summarize remaining issues. These are likely to be of a more substantial scientific nature, and could results in suggestions for

- algorithm/code review by CDR provider
- changes to CDR product contents

To conclude this section, we return to the topic of metadata and stress the importance of providing metadata that convey as much information as possible about changes in data quality. A key feature of a climate-quality CDR is its ability to provide metadata that characterizes the CDR's deviations from temporal consistency. Metadata at both the dataset level and the datum level may be needed for this. Provision of such metadata at the earliest possible stage increases the chances of interpreting the CDR correctly in off-line assessment exercises, and also lays the foundation for the development and implementation of observation bias corrections schemes that would hasten the progress of the CDR towards acceptance for full assimilation in a new reanalysis production.

Table 1 summarizes infrastructure needed for the steps in off-line CDR assessment discussed in this section.

Process component	Key considerations	Supporting infrastructure	Comments
Acquisition of CDR (data and metadata)	Data: observed parameter, geolocation coordinates Metadata (datum level): quality indicators, in the form of uncertainty estimates and flags Metadata (dataset level): user guide, scientific and technical documentation of the dataset, existing quality assessments, location of repository for future quality information. Metadata should permit unique identification of observations as well as clustering of similar observations and segregation of dissimilar observations.	Institutional support for CDR provision (data servers, user support services)	More detailed contents described in the Core-Climax Dataset Description Document and the Core-Climax System Maturity Matrix. Common conventions in data flagging is beneficial (e.g. all flags lowered to signify no known problems) Values and/or provenance of auxiliary/background information used in the CDR production should also be specified, could be at datum-level or at metadata level. The FP7 project CharMe is developing a metadata repository.
Acquisition of tools	CDR data reader Observation operator and/or unit convertors	Documentation of software tools. Scientific and technical support for observation operator development (e.g. radiative transfer algorithms).	For some CDRs, unit conversion tools are more appropriate than observation operators, e.g. to convert a CDR for ozone number density into mass mixing ratio. Observation operators often need to be accompanied by instrument metadata to

			compute supporting information (radiative transfer coefficients), possibly requiring interaction with CDR provider or original instrument operator/agency
Acquisition of reanalysis dataset	Do the timespan and geographical extent of the reanalysis cover the whole CDR?	Institutional support for reanalysis provision (data servers, user support services)	A regional reanalysis may be sufficient for some CDRs
Projection of reanalysis into CDR space ("forward simulation")	Typically restricted to projection of the reanalysed "analysis fields". Projection of "background/first-guess fields" is possible if these are archived or can be easily re-computed.		A quantitative form of "observation feedback" at the datum level.
Archiving	Database approach permits (a) generic handling of different observed parameters and (b) incremental feedback.	Database tools for exploration, mining and visualization. Common formats for observation feedback archiving. Support for extending databases incrementally and migrating previous databases	The archived observation feedback is made available to the CDR provider and 3rd parties. Core-Climax is fostering a co-ordinated approach for observation feedback format amongst reanalysis centres.
Comparison, assessment (potentially by 3rd party)	Is the supplied metadata sufficient to permit unique indentification of observations as well as clustering of similar observations and segregation of dissimilar observations?	Statistical analysis tools Community structures to share expertise and assessment information	Development of adequate metadata may need several iterations of the feedback loop. Capacity-building needed to train 3rd parties in interpretation of reanalysis feedback

Response by CDR provider (and/or reanalysis provider in the case of 3rd party assessment)	Does the CDR provider (and/or reanalysis provider) incorporate the assessment feedback within a process that leads to further reprocessing and improved products?	Good communication links, common vocabulary. Retention of expert knowledge about instrument performance and CDR generation algorithms. Institutional support for improving CDRs taking into account the feedback received.	Having an established mechanism to incorporate assessment feedback within a CDR improvement process is a key indicator in the Core-Climax System Maturity Matrix

III. Assessing CDR data quality by assimilation into a new reanalysis

In this section we address the more comprehensive "on-line" assessment of CDR quality that is facilitated by full assimilation of the CDR in the production of reanalysis. Full assimilation is only applied to those CDRs that are deemed fit-for-purpose, for which there are three principal conditions:

- CDR quality must first be pre-established within reasonable margins,
- credible error estimates must be available for the CDR, and
- the observation operator must be integrated within the reanalysis system in a way that links the observed parameter to the degrees of freedom that are adjusted in the data assimilation component of the reanalysis system. The data assimilation schemes of some reanalysis systems additionally require the development/integration of related tangent linear and adjoint operators.

Figure 3 depicts a systematic process for full assimilation of multiple CDRs, followed by comparison, assessment and feedback to the CDR providers. As with the off-line assessment described in the previous section, the process formulated here is informed by lessons learned from Core-Climax supporting visits and from experience within CMUG. It is designed to make maximum re-use of the off-line process formulated in Section II. Data readers and observation operators have been re-directed to integration within the reanalysis system, while reanalysis production supersedes and incorporates projection of the reanalysis into CDR space.

While the designed infrastructure for on-line assessment shares much in common with off-line assessment, there are some substantial additional development issues in the steps leading up to reanalysis production. These are:

- The need to identify duplicated data at the intra-CDR and inter-CDR levels. Duplication at the datum-level must be identified, anticipating that duplicated data may or may not be identical due to differences in processing algorithms or data-gathering techniques. Selection mechanisms must be devised/implemented so that only one version of each duplicated datum is accepted for the assimilation component of the reanalysis system.

- Some reanalysis systems need observation operators to be supplemented by tangent-linear and adjoint versions. This can be a significant software development task, and automated code generators have often proved inadequate.

- Development of bias correction schemes is often required, requiring significant scientific and technical resources.

- Incorporation of auxiliary reanalysis-model parameters as additional feedback parameters to CDR providers. Here the pressing needs are for a consultation process to establish the CDR-specific requirements, and resources for technical implementation.

Traditionally, the formal output of reanalysis production has been restricted to the archived reanalysis fields (gridded products). Figure 3 shows how such formal output is to be enriched to benefit CDR teams, specifically by making observation feedback available to CDR providers and 3rd parties. The feedback that is possible in off-line assessment (via projection of “reanalysis analysed fields” and sometimes “reanalysis background fields”, see previous section) can be extended to include additional quantitative and qualitative information that is routinely/uniquely available within reanalysis production but not in off-line reanalysis projection. In particular, projection of background fields into CDR space is routinely available at CDR datum level within reanalysis production. Other quantitative information at datum level include observational bias corrections where schemes to estimate these have been implemented (CDR-specific). Qualitative information is also available in the form of datum usage flags, as set by the observation quality control procedures within the reanalysis system. With appropriate infrastructure (requirements consultation process and software development), there is scope for observation feedback to be customized/enhanced with additional datum-level reanalysis-model parameters to assist specific CDR assessments. For example, Core-Climax/ERA-CLIM partners at Eumetsat/MetOffice indicated that their 3rd-party evaluations of SSM/T-2 microwave radiance data (from instruments on the series of DMSP satellites F-11, F-12, F-14, and F-15) would benefit from knowing model parameters such as surface type (land, sea, ice) and the meteorological situation (rain-affected, cloud type, cloud vertical extent).

A further feature of on-line assessment is that observation feedback archived for all ingested CDRs are available for inter-comparison with the feedback for a specific CDR of interest.

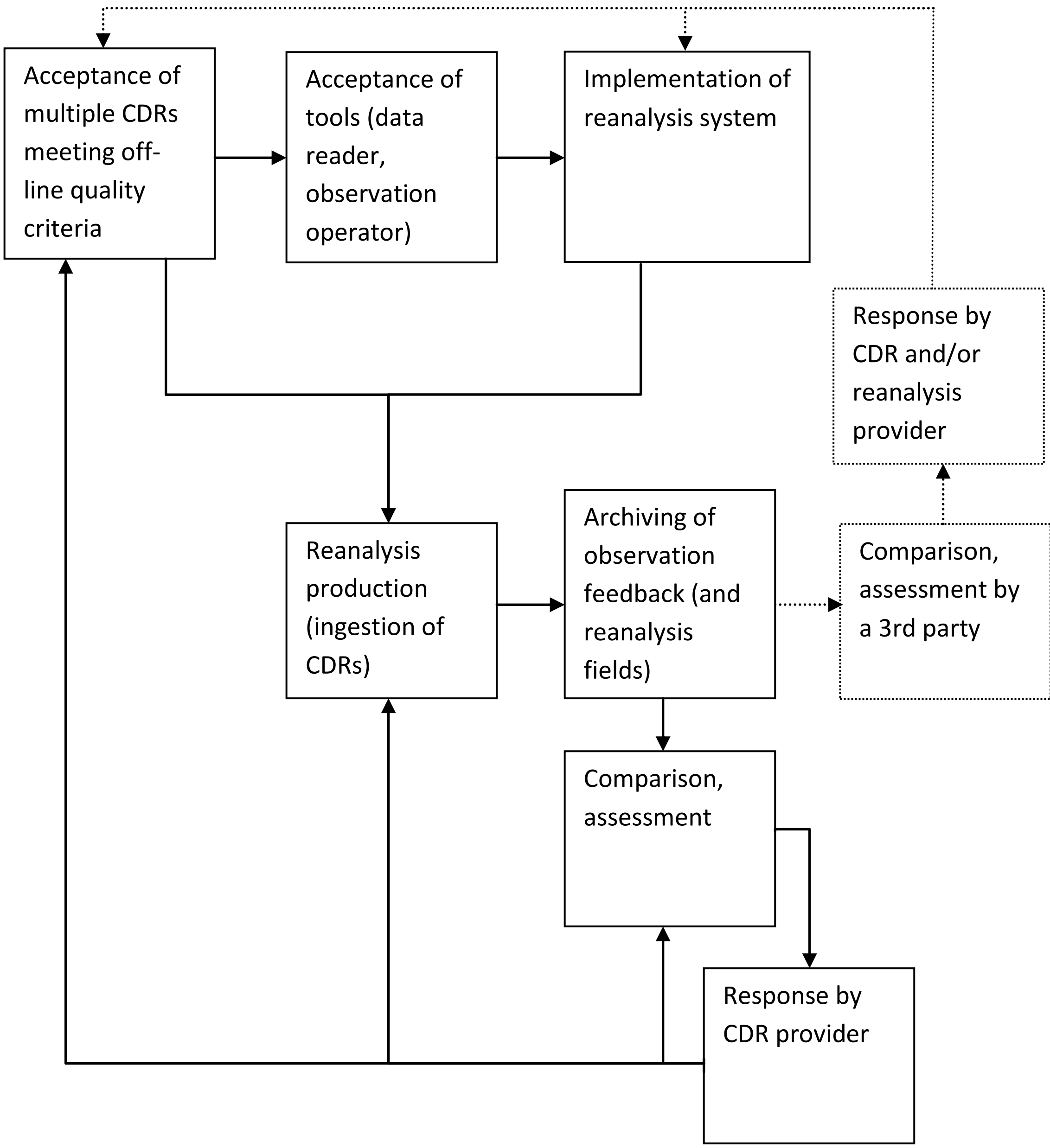

Figure 3: Procedure for CDR assessment via full assimilation in a reanalysis production (on-line assessment)

Table 2 summarizes infrastructure needed for the steps in on-line CDR assessment discussed in this section. For the final three rows, the issues are common with Table 1 and text has been repeated to avoid the inconvenience of cross-referencing.

Process component	Key considerations	Supporting infrastructure	Comments
Acceptance of multiple CDRs meeting off-line quality criteria	CDRs used as inputs to the reanalysis system need to be traceable and transparent Have uncertainty estimates been validated? Are the CDRs consistent at the intra-CDR and inter-CDR levels? Duplicated data (intra-CDR and inter-CDR) must be identified (may or may not be identical) and selection mechanisms devised/implemented	Adoption of Open (Science) Data policies Software development for implementation in the reanalysis system	Excessive inconsistency will render some CDRs unacceptable for on-line ingestion/assessment. If poorly co-ordinated, historical data rescue can lead to duplication of data amongst different data collections
Acceptance of tools	For some reanalysis systems, observation operators may need to be supplemented by tangent-linear and adjoint versions	Scientific and technical support for supplementary tools	Development would benefit from greater access to data assimilation test-beds
Implementation of reanalysis system	Are bias correction schemes required? Do CDR providers require auxiliary parameters from the reanalysis system as additional feedback parameters?	R&D Consultation process to establish the requirements Software implementation	
Reanalysis production (ingestion of CDRs)			Two modes of on-line assessment: (1) production-mode reanalysis intended for

			public release, (2) experimental-mode reanalyses dedicated to assessment of a specific dataset (typically involving production of a "control" reanalysis and one or more "experimental" reanalyses).
Archiving	Database approach permits (a) generic handling of different observed parameters and (b) incremental feedback.	Database tools for exploration, mining and visualization. Common formats for observation feedback archiving. Support for extending databases incrementally and migrating previous databases	The archived observation feedback is made available to the CDR provider and 3rd parties. Core-Climax is fostering a co-ordinated approach for observation feedback format amongst reanalysis centres.
Comparison, assessment (potentially by a 3rd party)	Is the supplied metadata sufficient to permit unique indentification of observations as well as clustering of similar observations and segregation of dissimilar observations?	Statistical analysis tools Community structures to share expertise and assessment information	Development of adequate metadata may need several iterations of the feedback loop. Capacity-building needed to train 3rd parties in interpretation of reanalysis feedback
Response by CDR provider (and/or reanalysis provider in the case of 3rd party assessment)	Does the CDR provider (and/or reanalysis provider) incorporate the assessment feedback within a process that leads to further reprocessing and improved products?	Good communication links, common vocabulary. Retention of expert knowledge about instrument performance and CDR generation algorithms. Institutional support for improving CDRs taking into account the feedback received.	Having an established mechanism to incorporate assessment feedback within a CDR improvement process is a key indicator in the Core-Climax System Maturity Matrix

IV. Tailoring for satellite and conventional CDRs

We believe that the assessment procedures formulated in the previous section are generally applicable to both satellite and conventional (in-situ) CDRs. Nonetheless, specific implementations should have discretion to adapt and/or tailor the procedure according to the particular circumstances of the assessment being undertaken -- when required, this will typically be within the details of the procedure sub-components. For assessment of satellite data that has not previously been used in numerical weather prediction, the development of observation operators is often a substantial task for which several feedback iterations can be required. Rescue and assessment of historical in-situ data is a comparitively recent development, in which there are pressing needs for (a) more comprehensive metadata concerning the instruments deployed and how these changed with time, (b) more co-ordination in data-gathering to identify duplication observations at the intra-CDR and inter-CDR levels.

V. Summary and Conclusions

Use of the reanalysis environment for quality assessment of observational CDRs is set to grow significantly in the coming decades. Because reanalysis-based assessment of observational CDRs depends on substantial support infrastructure, this document has set out to explain the main elements of this infrastructure in a way that is accessible to CDR producers, to assist them in making appropriate preparations for receiving maximum benefit from the feedback that the reanalysis environment can provide. Full benefit is only possible when the observational CDRs are supplemented with appropriate documentation, metadata and auxiliary tools.

The infrastructure is framed in the context of iterative assessment processes that are designed improve CDRs via provision of quantitative feedback at the datum-level as well as analytical feedback in the form of assessment reports. A distinction is made between off-line assessment with an existing reanalysis (Section II, based on projecting the reanalysis into CDR space) and on-line assessment (Section III, based on ingesting the CDR in the assimilation component of the reanalysis system). This is because satisfactory off-line assessment is typically a pre-requisite, together with additional infrastructure development, for on-line assessment.

Identifying the infrastructure needs (described in detail in Tables 1 & 2) has been facilitated by the experience gained through the Core-Climax programme of visits that facilitated direct contact between ECMWF and a diverse range of CDR providers from both the in-situ and satellite-based communities. A second, complementary, source of information has been the experience of ECMWF's Renanalysis section in the Climate Modelling Users Group (CMUG) of the ESA Climate Change Initiative (CCI). We re-iterate some of the over-arching considerations here:

- Quantitative feedback at the observation datum level is desired - addressed by projecting reanalyses into the CDR space in both off-line and on-line assessments
- Meaningful comparison and statistical analysis requires appropriate clustering of similar observations, segregation of dissimilar observations, and good characterization of fluctuations in quality (especially lack of temporal consistency) - addressed by access to appropriate metadata
- Feedback should be provided incrementally and the quantitative feedback should be extensible - addressed by multiple feedback loops and a database approach
- There is value in making the quantitative feedback available to third parties - addressed by incorporating a “3rd party feedback loop” (Figure 2 & 3)

Realizing and exploiting the infrastructure described above will require improved global co-ordination and capacity-building in a number of areas. These include:

- Preserving the existing knowledge-base on CDR quality.
- Training to raise awareness about uncertainties in numerical models, observations, and reanalyses, the importance of traceability, quality indicators, systematic and random errors, validated datasets.
- Achieving the critical mass for generating and interpreting reanalysis feedback; engaging more people in the scientific, technical and software developments
- Co-ordination amongst data rescue initiatives to identify duplicate observations; development and adoption of conventions for unique observation identification
- Compilation of historical metadata, especially when instruments and observing practices have changed many times in the CDR period
- Global inventories of reprocessed data holdings
- Facilitation of feedback dialogue via conference sessions and workshops

VI. Appendix 1: An incremental Observation Feedback Archive approach

The demands for quantitative feedback on observational CDRs that is (a) provided at the datum-level, (b) extensible, and (c) supported by tools applicable for a wide range of CDRs mirror the internal requirements for observation handling within the ECMWF reanalysis system. The database approach previous under development at ECMWF for purely internal use can be adapted for use as an Observation Feedback Archive. Such an archive will typically contain the following categories of information:

A. Observation identification

B. Observation metadata (e.g., platform altitude)

C. Observation data

D. Observation recalibration/reprocessing corrections

E. Pre- or post-assimilation feedback (departures, before/after assimilation, usage flags,…)

These categories are filled using an “incremental” approach to accumulate knowledge from recalibration, reprocessing, and reanalysis.

An example is given below, with a data sample from the SSM/I Fundamental Climate Data Record produced by the CM-SAF, at several steps of the dataset assessment.

VI.A/B/C Original data and metadata, before recalibration/reprocessing

These categories would be filled on the first acquisition of the CDR.

Observation identification (should remain unchanged independently of future reprocessing)								Observation metadata, including spatio-temporal geotagging (may change with future reprocessing)								Observation data (original processing)	
source@hdr	satellite_identifier@sat (WMO Pub. No. 306 table C-5)	satellite_instrument@sat (WMO Pub. No. 306 table C-8)	orbit@radiance (revolution number)	scanline@radiance	scanpos@radiance	vertco_type@body (type of vertical positioning)	vertco_reference_1@body (channel number here)	date@hdr	time@hdr (HHMMSS)	lat@hdr	lon@hdr	stalt@hdr (satellite altitude here)	zenith@sat	azimuth@sat	surface_class@modsurf (type of surface)	Varno@body (type of geophysical variable, ODB table)	obsvalue@body
'CMSAF001', better would be full DOI	241	905	366	missing	1	3 (meaning, data are organized by channel)	1	19870716	000010	5.81 °S	100.25 °E	862.9 km	53.47 °	-59.49 °	5 (meaning, water surface)	119 (meaning, brightness temperature)	212.94 K

VI.D CDR (after recalibration/reprocessing)

The previous table is incremented with information from Observation reprocessing and/or recalibration.

Observation identification		Observation metadata		Observation data (original processing)		*Incremental addition:* Observation reprocessing and/or recalibration
source@hdr	…	date@hdr	…	obsvalue@body	…	obsvalue_cal@body (recalibrated observation, product of several corrections, which may be all listed separately)
'CMSAF001'	…	19870716	…	212.94 K	…	213.59 K

VI.E(i) CDR after matching/comparison to ERA-Interim

The previous table is incremented with information from Pre-assimilation feedback from ERA-Interim, as would be obtained via the procedures of Section 2.

Observation identification		Observation metadata		Observation data (original processing)		Observation reprocessing and/or recalibration	*Incremental addition:* Pre-assimilation feedback from ERA-Interim, at the observation date/time and location									
source@hdr	…	date@hdr	…	obsvalue@body	…	obsvalue_cal@body	fgvalue@ei (equivalent B.T. computed from ERA-	emis@ei (emissivity, computed from ERA-Interim)	lsm@ei (land-sea mask, at ERA-Interim resolution)	orography@ei (orography, at ERA-Interim resolution)	seaice@ei (sea-ice fraction, from ERA-Interim)	t2m@ei (2-meter temperature from ERA-Interim)	tsfc@ei (skin temperature, from ERA-Interim)	u10m@ei (10-meter zonal wind, from ERA-Interim)	v10m@ei (10-meter meridional wind, from ERA-Interim)	rainy@qc (rain flag computed from the observations)
'CMSAF001'	…	19870716	…	212.94 K	…	213.59 K	208.92 K	0.57	0.0 (open water)	1.6 m a.m.s.l.	0.0 (no ice)	300.38 K	302.26 K	-7.49 m/s	5.48 m/s	10 (possibly rainy)

VI.E(ii) CDR after matching/comparison to ERA-20C

The previous table is incremented with information from Pre-assimilation feedback from ERA-20C, as would be obtained via the procedures of Section 2.

Observation identification	source@hdr	'CMSAF001'
	…	…
Observation metadata	date@hdr	19870716
	…	…
Observation data (original processing)	obsvalue@body	212.94 K
	…	…
Observation reprocessing and/or recalibration	obsvalue_cal@body	213.59 K
Pre-assimilation feedback from ERA-Interim	fgvalue@ei	208.92 K
	…	…
Incremental addition: Pre-assimilation feedback from ERA-20C	fgvalue@e2oper	209.64 K
	…	…

VI.E(iii) CDR after hypothetical assimilation into a future-generation reanalysis like ERA5

The previous table is incremented with information from Post-assimilation feedback from ERA5, as would be obtained via the procedures of Section 3.

Observation identification		Observation metadata		Observation data (original processing)		Observation repro-cessing and/or recali-bration	Pre-assimilation feedback from ERA-Interim		Pre-assimilation feedback from ERA-20C		*Incremental addition:* Post-assimilation feedback from ERA5		
source@hdr	…	date@hdr	…	obsvalue@body	…	obsvalue_cal @body	fgvalue@ei	…	fgvalue@e2oper	…	fg_depar@era5 (departure before assimilation into ERA5)	an_depar@era5 (departure after assimilation into ERA5)	Datum_status@era5 (was the datum used?)
'CMSAF001'	…	19870716	…	212.94 K	…	213.59 K	208.92 K	…	209.64 K	…	… K	… K	0 / 1

VII. Appendix 2: Template for Assessment Report

Here we provide guidelines for the content of an Assessment Report for a CDR being assessed in a reanalysis environment. Where informal queries have been answered by the CDR provider in the course of the assessment, it is helpful to document this within the Report for the benefit of future users and to inform future updates of the CDR and associated documentation.

1. Introduction

    Provide some motivation for the assessment. Place it in the context of other CDRs/reanalyses.

2. Acquisition of the CDR

    Describe where the CDR and associated documentation was acquired from. Describe any problems encountered and any solutions that may assist future users. Do the data conform the the descriptions in the documentation?

3. Auxiliary tools

    Describe any tools that were used to decode the CDR, including modifications made during conversion from the native CDR format to reanalysis-environment format.

    Describe the observation operators and any issues that arose with them.

4. Acquisition of the reanalysis

    Describe where the reanalysis was acquired from. Describe any pre-processing, including spatio-temporal interpolations. Provide collocation details where these have an impact on representativeness errors.

5. Simulation/ingestion and archiving

    Describe how the projection of the reanalysis into CDR space was performed. Give details of how the results were archived and how they can be accessed.

6. Results of quantitative comparison and assessment

    Illustrate the strengths and weaknesses of the CDR. Depending on the nature of the CDR under assessment, provide a selection of maps, timeseries, profiles, histograms, and tables.
    Is the spatio-temporal coverage as expected? Are the data and metadata with expected ranges? Are the trends, anomalies and mean CDR-reanalysis differences within expected ranges? Is there any indication of temporal inconsistency (e.g. break-points in a timeseries)?

7. Summary of Feedback (to CDR provider, auxiliary tool provider or reanalysis provider)

Consolidate the feedback arising in the previous sections.

Was the documentation sufficient to obtain and use the data?

What were the strengths and weaknesses of the data?

Were anomalies identified? Are there adequate explanations for them?

List any explicit recommendations to the CDR provider.

8. Preliminary responses to Feedback (if applicable)

It is common for results and recommendations to arise at intermediate points during the assessment process, and for these to be discussed with the CDR provider prior to finalizing the Report. In such cases, including responses from the provider helps to inform readers about the level of consensus regarding the interpretation of the results and prospects for recommendations to be actioned in future CDR updates.