Seventh Framework Programme
Theme 6

Project: 607193 UERRA

Full project title:
Uncertainties in **E**nsembles of **R**egional **R**e-**A**nalyses

Deliverable D9.3
Lessons learned

WP no:	9
WP leader:	KNMI
Lead beneficiary for deliverable :	SMHI, KNMI
Name of <u>author</u>/contributors:	Albert Klein Tank, Per Undén
Nature:	OTHER
Dissemination level:	PU
Deliverable month:	24
Submission date: September 24, 2016	Version nr: 1

Common lessons learned relevant for the development of the C3S

ERA-CLIM2, UERRA, QA4ECV, CLIPC and EUCLEIA are acronyms for five FP7 projects from the 2013 Space Call of FP7 that all share the common objective to prepare for the Copernicus Climate Change Service (C3S). The capabilities being developed through these five research initiatives form the scientific and technological foundation for the C3S. As part of the overarching coordination among these five projects, each project has described the lessons learned relevant for the development of the operational services in the C3S. These are provided in Sections 2 to 6.

The overarching lessons that the five projects have in common are summarized below. Because ECMWF (who operates the C3S on behalf of the European Union) has started to bring together expertise from across Europe to deliver the service, it is noted that some of the lessons directly translate to parts of the service, whereas direct links are less clear for other lessons (indicated by *). These may warrant special attention in future.

1. Data rescue must be seen as an ongoing effort which requires a permanent infrastructure and knowledge base in order to access the observational data not used so far (partly because of data policies).
2. A commonly agreed metadata base is required which links strongly to the international community working with historical earth-system observations.
3. Key people with data assimilation expertise and communication skills are often the bottle neck for timely delivery of complex coupled and/or regional reanalyses systems. *
4. Actual reanalysis and ECV production turns out to more demanding than anticipated. *
5. Regional reanalysis at the European scale is relatively new and requires significant support because it cannot lean on existing pan-European operational systems.
6. Long term biases in reanalysis datasets (due to model parametrization and changes in observation coverage) need to be monitored and compared with other observational references because they may introduce spurious climate trends.
7. Intensive communication and engagement with users is required to decide on priorities concerning physical parameters to be archived, observations to compare to, scales and skill scores to investigate. *
8. There is a need for products to be tailored to the varied needs of different user groups for such products to be useful in their decision making; an information portal needs to speak to all types of users.
9. The huge differences in approach and methodology between the many disciplines need to be brought together to create a credible climate service. *
10. It is vital for users to have a clear understanding of the scientific uncertainties involved and of the robustness of the results relevant to the estimation of low-frequency variability and climate trends. *
11. Interactive traceability chains are a useful and efficient tool for data users to obtain confidence in the quality, robustness, and limitations of satellite ECV data records.

12. It is important to maintain close ties to relevant international (ESA, EUMETSAT, NASA, JAXA) and national space agencies to ensure the availability of the most recent state-of-science satellite level-1 data needed for generation of ECV records.
13. There is a need for a continuous stream of research funding in parallel of the operational C3S to guarantee that the services remain state-of-the-art over time. *

PROJECT DELIVERABLE

Project	ERA-CLIM2
Grant Agreement	607029
Work package 9:	Global 20th century reanalysis
Deliverable D9.3:	Common Lessons Learned
Type:	Report
Author:	Roberto Buizza (ECMWF)
Reviewers:	ERA-CLIM2 WP Leaders
Dissemination level	PU
Delivered:	21 April 2016

Common lessons learned relevant for the development of the Copernicus Climate Change Service

Table of contents:

Executive Summary

This report is the ERA-CLIM2 contribution to an overarching Work Package (common to five FP7 Space Calls) that bridges across similar projects. It discusses ***Lessons Learned (LL)*** from the ERA-CLIM2 project that could be relevant also for the other and/or future similar projects, and possibly also for the operational activities under development within the Copernicus Climate Change Services. More specifically, **the report discusses 11 LLs from ERA-CLIM2 project** covering five aspects of reanalysis' activities: Earth-system observations (including data rescue), data assimilation coupling methods, reanalysis production, quantification of uncertainties and funding.

1. Introduction

The understanding of climate change is highly dependent on the available global satellite and conventional observational data in the atmosphere, the land and the ocean and sea-ice, and the development of coupled ocean-land-atmospheric models and assimilation systems that can ingest these data. A continuous cycle of research and development, operational production, diagnostic and evaluation is required to refine further reanalyses, so that they can provide a better, closer-to-reality image of the time evolution of the Earth system.

Within this context, the ERA-CLIM2 project focuses on the production and the assessment of multi-decadal reanalyses of the earth climate system including atmosphere, land, ocean and cryosphere. These reanalyses are based on the coupled ECMWF model that includes currently the IFS atmospheric model, the WAM wave model, the LIM2 sea-ice model and the NEMO ocean model.

In the European Commission FP7 Space Call (9.2) Projects (UERRA, ERA-CLIM2 and QA4ECV, EUCLEIA and CLIPC), there is an overarching Work Package (WP9 or 8) that bridges across the five projects and collates common information from the projects in order to communicate to the outside world in a coordinated way. A deliverable in this Common Overarching WP is a report about the common ***Lessons Learned***, where 'common' means relevant to all projects and to the operational production of reanalysis within Copernicus activities.

This report discusses 11 ***Lessons Learned (LL)*** from the ERA-CLIM2 project that are thought to be relevant also for the other projects and Copernicus activities, covering the four main areas of work of ERA-CLIM2 (sections 1-4) and funding (section 5):

- **Earth-system observations**: data must provide an accurate representation of observed low-frequency variability and trends, and Reanalyses must make use of as many input observations suitable for climate purposes as possible;
- **Scientific developments and data assimilation coupling methods**: reanalyses must include better representations of interactions and feedbacks in the coupled climate system (atmosphere-ocean-land-cryosphere);
- **Reanalysis' production**: reanalyses must be extended further back in time to provide a sufficiently long record for climate studies and climate model validation;
- **Quantification of uncertainties**: users must be provided with useful information about uncertainties relevant to the estimation of low-frequency variability and trends.
- **Funding**: there is a need for a continuous stream to sustain a cycle of data rescue, R&D, production and evaluation.

2. Earth-system observations

2.1 Historical observations

LL #1 – Establish and maintain a commonly agreed metadata base, and establish and maintain strong link with the international community working with historical observations.

Based on the ERA-CLIM2 experience, the following points should be considered to be able to improve further reanalyses:

- It is important to establish and maintain a commonly agreed metadata base that contains the level of stations with respect to data rescue (global registry developed within ERA-CLIM/ERA-CLIM2), but also on available data (ISTI data bank, ISPD, see below);
- It is important to stablish and maintain strong links with other FP7 projects in order to be better informed about update cycles, other projects, and community initiatives; ERA-CLIM2 has e.g. established very good links with UERRA;
- Community initiatives, which are 'accepted' as playing a coordinative role, should be sustained: Atmospheric Circulation Reconstructions over the Earth (ACRE), International Surface Temperature Initiative (ISTI), the EarthTemp network, and they also provide links to other organizations (e.g., International Environmental Data Rescue Organization (IEDRO));

On this later point, ACRE does annual workshops with good ERA-CLIM2 (and UERRA) participation, Earth Temp also does workshops; ISTI does regular phone conferences with good ERA-CLIM2 participation. ERA-CLIM2 is organising a conference on historical data for reanalysis on 22 June 2016 in Maynooth, Ireland, linked to the annual ACRE workshop.

2.2 Recovery of observations for reanalyses (data rescue)

LL #2 – Data rescue must be seen as an ongoing effort; it is essential to maintain a data rescue infrastructure; there are still that have been rescued within projects not well known by the scientific community, which are not used by the scientific community.

Based on the ERA-CLIM2 experience:

- It is important to maintain a metadata base on data rescue projects (I-DARE);
- It is essential to maintain a data rescue infrastructure, which encompasses important know-how and networks such as links with archives, historians, metadata bases, and data recovery expertise;
- The possibility of establishing feedbacks between the stations metadata base content and databanks (ISTI, ISPD, etc.) in what regards errors detected in historical observations locations or other errors, should be considered; location errors have already been detected during the inclusion of the ISPD databank inventories in the ERA-CLIM2 metadata base;
- There are existing data that are still not used, 'missing' (or not well established) in reanalyses: for example, data rescued within several development projects that perform data rescue work, e.g. the Pilot Program for Climate Resilience (PPCR) of the Climate Investment Funds, implemented by the World Bank Group and Development Banks; these projects are often closer to the local authorities but too far away from the scientific communities.

Data rescue must be seen as an ongoing effort. Changes in the demand (e.g., for analysing past extreme events with respect to changes in extremes) and changes in the technical possibilities (e.g., developments in data assimilation) lead to a re-evaluation of historical data. Such changes will as well occur in the future: Demand will change (e.g., focus of solar and wind data for planning renewable energy generation, or stronger focus on climate impact data), technical possibilities will change (e.g. assimilation of wind over land or cloud motion – variables that have not so far been digitised). In such an instance, one will go back to the archive and will profit from an established procedure, infrastructure, metadata base, and imaged documents. It is therefore important to make sure that the infrastructure and expertise built up in this project will be made available to others.

3. Scientific developments and data assimilation coupling methods

LL #3 - When planning for the development of the data assimilation system to be used in reanalysis production, take into account the potential disruptive impact of the turnover of key people.

When ERA-CLIM2 was planned and the project submitted, it was thought that some key people, who had a very good knowledge of the coupled assimilation system that was going to be used in the reanalysis production, could be employed to do the work. Unfortunately, this was not the case because of substantial personnel changes at ECMWF, partly due to the establishment of the Copernicus Climate Change Service (C3S), and partly because some key people left ECMWF. The net result was again severe delays in the development of the coupled assimilation system.

LL #4 - To make best use of developments aimed at future reanalysis production, allow time for initial testing in the existing (in this case CERA) frameworks.

New developments to the coupled data assimilation system are being made at various centres during ERA-CLIM2. Even when these centres are developing the code base used at ECMWF, testing and integration of new developments into the CERA system requires significant resources associated with it, particularly at ECMWF. Reanalysis production during ERA-CLIM2 cannot include all developments made during the project and many developments are aimed at future coupled reanalysis production. Initial testing and integration of these developments using the CERA system would allow an assessment of which developments should be included in future production runs, and would ensure that developments are transferred to ECMWF within the project. There is still the requirement for continued funding in order to make sure that R&D carried out during ERA-CLIM2 can be properly pulled-through into future production reanalyses (see the last section of this report).

LL #5 - Allow for the significant amount of communication and learning required from the central reanalysis production group in terms of the increasing number of earth system components of coupled reanalysis systems.

Coupled reanalysis systems are becoming ever more complex, including not only physical (atmosphere, land, ocean, sea-ice, waves) components but also carbon and biogeochemical models and data assimilation systems. Development of these systems is carried out by a number of external groups who have built up significant expertise and experience of these various systems. Implementing aspects of these additional components requires significant amounts of communication between the central reanalysis production group at ECMWF and the external groups.

4. Reanalysis' production

LL #6 - When planning for a reanalysis production, take into account the complexity of the data assimilation system, whether new aspects are being included, and whether there are going to be changes in the computing environment that could lead to delays.

CERA-20C and CERA-SAT, the two reanalysis that will be delivered by the ERA-CLIM2 project, includes many innovation aspects and are technically very complex. The production of a reanalysis as CERA-20C that covers a very long period (1901-2010) involves running many parallel streams, which overlap for a certain period (say 1-2 years). In the CERA-20C case, this endeavour was made even more challenging by the fact that each stream included a 10-member ensemble of analyses, so that error statistics and analyses' uncertainties could be estimated. The result was that, firstly, the development of the system was very complex from the technical point of view, and, secondly, the production was extremely complex. These challenges had to be addressed while the ECMWF High-Performing Computing (HPC) was being upgraded. The unforeseen late availability of the new HPC at ECMWF caused a slower progress and a delayed start of the reanalysis productions. The net result is that the ERA-CLIM2's deliverables are about 12 months late.

LL #7 - When establishing the production suite(s), identify possible bottlenecks and take actions that could improve efficiencies, and take into account potential negative impact that running many jobs in parallel could have.

The CERA-20C production involved running in parallel 14 streams, each with 10 members. The net result was that when production started, there were 140 parallel 4-dimensional variational assimilation suites running in parallel, each involving several jobs with large Input/Output (I/O). It was only when production started and all streams were launched, that it was realized that the set-up was very inefficient, and would have required at least 12-months to complete CERA-20C. Expert technical work then started, and thanks to optimisation work and the re-design of some of the jobs, and to the allocation of special (higher) priorities to the jobs, we did manage to reduce the production time to about 6 months. This work required also the help of analysts and experts from the ECMWF Computing Department.

LL #8 - When estimating the production time, take 'archiving' into account.

This is another area where, when production started, there were clear inefficiencies due to the shear amount of data that the 140 data-assimilations were generating. As for the point above, this was not foreseen as a problem, since this was the first time that so many streams were run in parallel, with a coupled ocean-land-atmosphere model. Issues were addressed by re-designing some of the scripts.

5. Quantification of uncertainties

LL #9 - In the ERA-CLIM2 proposal it has been foreseen that a large number of historical upper air data in the 1920-1970 period will be collected, however assimilation of these has not been required as a formal deliverable. Only a few test assimilations were foreseen.

In retrospect, the value of such assimilations has been likely underestimated. Luckily, a longer assimilation run (ERA-preSAT) using the digitized upper air data collected so far could be performed for the period 1939-1967. The enormous potential of such an assimilation for data quality control could be shown. However, a necessary rerun of this assimilation could so far not be realized due to lack of human but also computational resources. As a result, many of the newly digitized data have yet to be assimilated and the quality of the feedback for those data that have been assimilated leaves room for improvement. This affects also the development of bias adjustment procedures for upper air data.

LL #10 - Data exchange for quality assessment and assimilation of digitized data has been made unnecessarily complicated because of the use of relatively unstructured or nonstandard data formats.

In retrospect, the ERA-CLIM upper air database could have been delivered in a more standard, easier to read format, such as netCDF, or at least a reading routine that reads the data into a structure in memory could have been provided. The issues related to that have now been mostly solved, but at a relatively high cost of human resources. Quantification of uncertainties is also affected by late development of the Observation Feedback Archive. It exists for quite a variety of data at ECMWF in ODB2 format. However only some of these data sets are not readily accessible from outside ECMWF.

Despite this self-criticism, it should be mentioned that several project partners (FFCUL, UBERN, RIHMI, UNIVIE, ECMWF) are actively using reanalysis feedback data for their observation data quality control. The emphasis on the production of this kind of data has well paid off already.

6. Funding: the need for a continuous stream to sustain a cycle of data rescue, R&D, production and evaluation

LL #11 - There is a need to guarantee a stream of R&D funding to sustain a continuous cycle of 'Data rescue', 'Research and Development', 'Reanalysis production' and 'Evaluation'.

Continuity is required to guarantee that advances in the different areas of work that lead to a reanalysis production, are being then used in follow-on reanalysis productions. Continuity makes it possible to advance further the science behind reanalysis, including the development of observation operators capable to assimilate rescued observations and the upgrading of coupled data assimilation systems, and deliver better and more accurate reanalysis of the Earth-system. Continuity of European Union funding for research projects would guarantee a continuous stream of advances and developments into reanalysis operational productions by the Copernicus Climate Change Service (C3S).

Continuity is required because the assimilation system used for a specific reanalysis production (say the CERA-20C one) needs a long time to be developed and tested (say order 2-4 years), and thus can only be based on advances coming from past projects (in this case, it was built within the ERA-CLIM and CERA projects). Similarly, data rescued within one specific project (say, again, ERA-CLIM2) will not be used in the current reanalysis but only in the future ones, since (to be able to use them in data assimilation) they will require the development of appropriate observation operators. Furthermore, their use will need to be tested properly in research experiments to assess whether they are properly used and they do not insert any spurious signal.

The establishment of a continuous cycle would guarantee that operational reanalyses, such as the ones produced by C3S, would be of increasing accuracy and value, and would include:

- The use of more rescued data in reanalysis production and/or evaluation;
- The use of upgraded coupled assimilation systems that are capable to use more rescued data, and that include further relevant processes (e.g. a sea-surface-temperature analysis component, the assimilation of aerosol optical depth and the inclusion of a flow-dependent uncertainty estimation for the ocean as well as for the atmosphere);
- The production of more accurate reanalyses covering longer, relevant periods.

*** ***

THEME SPA.2013.1.1-05

EUCLEIA
(Grant Agreement 607085)

EUropean CLimate and weather Events: Interpretation and Attribution
Deliverable D9.3
Common lessons learned relevant for the development of the Copernicus Climate Change Service

Deliverable Title		*Common lessons learned relevant for the development of the Copernicus Climate Change Service*
Brief Description		*Report on the common lessons learned for the development of the Copernicus Climate Change Service from the five FP7 projects from the 2013 FP7 Space Call.*
WP number		9
Lead Beneficiary		*Peter Stott, Met Office*
Contributors		*Nikos Christidis, Met Office*
Creation Date		8th December, 2015
Version Number		2
Version Date		7th January, 2016
Deliverable Due Date		December, 2016
Actual Delivery Date		
Nature of the Deliverable	R	*R - Report*
		P - Prototype
		D - Demonstrator
		O - Other
Dissemination Level/ Audience		*PU - Public*
	PP	*PP - Restricted to other programme participants, including the Commission services*
		RE - Restricted to a group specified by the consortium, including the Commission services
		CO - Confidential, only for members of the consortium, including the Commission services

Version	**Date**	**Modified by**	**Comments**
1	8/12/15	Peter Stott	First draft
2	7/1/16	Peter Stott	Second draft

Table of Contents

1. Executive Summary

Provide a brief summary of:
- the work performed and the main results achieved so far.

This report summarises the common lessons learned by the five projects funded under the 2013 FP7 Space Call relevant for the development of the Copernicus Climate Change Service. The five projects involved are ERA-CLIM2, UERRA, QA4ECV, CLIPC and EUCLEIA. There is a common website containing links to the websites of the five projects at http://www.clipc.eu/c3s-precursors/c3s-precursors. The five topics were chosen to specifically target preparations for the Copernicus Climate Change Service (C3S). The five projects funded to tackle these projects started in late 2013 and early 2014. Together they aim to develop the scientific and technological foundations for C3S.

The main lessons learned so far under these projects for the development of the Copernicus Climate Change Service are as follows:

1) Active user engagement is crucial to the development of a climate service to support decision making. CLIPC has found that active user engagement not only helps the project to stay in touch with what users want, but also has facilitated communication between different domains by forcing all to communicate in a jargon free language. This has helped to mitigate one of the major problems, the huge differences in approach and methodology between the many disciplines which need to be brought together to create a credible climate service.
2) It is vital to have a clear understanding of the scientific uncertainties involved and a clear communication of the robustness of assessments for decision making. A forthcoming meeting sponsored by CLIPC and co-organised by EUPORIAS, EUCLEIA and QA4ECV will investigate different aspects of uncertainties in climate data, focussing on communication of uncertainties in a manner which develops confidence.
3) There is a need for products to be tailored to the varied needs of different user groups for such products to be useful in their decision making. EUCLEIA has found that there is a clear demand for attribution services across sectors but there are different requirements concerning aspects such as timeliness of products and communication protocols. CLIPC, in seeking to address a spectrum of users from climate scientist, impact scientists, intermediaries and societal end users, has learnt that an information portal needs to speak to all types of users allowing them to navigate through the portal in a natural way without making the specialist versus non-specialist distinction.

2. Project Objectives

With this deliverable, the project has contributed to the achievement of the following objectives (DOW, Section B1.1):

No.	Objective	Yes	No
1	Derive the requirements that targeted user groups (including regional stakeholders, re-insurance Companies, general public/media) have from attribution products and demonstrate the value to these users of the attribution products developed under EUCLEIA.	YES	
2	Develop experimental designs and clear ways of framing attribution studies in such a way that attribution products provide a fair reflection of current evidence on attributable risk.		NO
3	Develop the methodology for representing the level of confidence in attribution results so that attribution products can be trusted to inform decision making.		NO

4	Demonstrate the utility of the attribution system on a set of test cases of European weather extremes.		NO
5	Produce traceable and consistent attribution assessments on European climate and weather extremes on a range of timescales; on a fast-track basis in the immediate aftermath of extreme events, on a seasonal basis to our stakeholder groups, and annually to the BAMS attribution supplement.	YES	

3. Detailed Report

There are five Copernicus Climate Change Service Precursor Projects funded under the 2013 FP7 Space Call. These are ERA-CLIM3 (European Reanalysis of the Global Climate System), UERRA (Uncertainties in Ensembles of Regional Re-Analysis), QA4ECV (Quality Assurance for Essential Climate Variables), CLIPC (A Climate Information Portal for Copernicus) and EUCLEIA (European Climate and Weather Events: Interpretation and Attribution). This report summarises the main common lessons learned relevant for the development of the Copernicus Climate Change Service.

The lessons learned fall into three categories which are discussed below.

3.1 User Engagement

Active user engagement is crucial to the development of a climate service to support decision making. CLIPC has found that active user engagement not only helps the project to stay in touch with what users want, but also has facilitated communication between different domains by forcing all to communicate in a jargon free language. This has helped to mitigate one of the major problems, the huge differences in approach and methodology between the many disciplines which need to be brought together to create a credible climate service.

3.2 Communication of robustness of products and scientific uncertainty

It is vital to have a clear understanding of the scientific uncertainties involved and a clear communication of the robustness of assessments for decision making. A forthcoming meeting sponsored by CLIPC and co-organised by EUPORIAS, EUCLEIA and QA4ECV will investigate different aspects of uncertainties in climate data, focussing on communication of uncertainties in a manner which develops confidence. Further details are as follows.

3.2.1 Forthcoming workshop on communication of workshop.

As a result of common lessons learned so far under the precursor projects, a workshop has been organised to take place in February, 2016.

The meeting will be hosted by GERICS in Hamburg, sponsored by FP7 project CLIPC and co-organised by additional FP7 projects EUPORIAS, EUCLEIA and QA4ECV. The workshop will discuss different aspects of uncertainties in climate data, focussing on communication of uncertainties in a manner which develops confidence. The sub-title of the workshop, “Presenting Uncertainty with Confidence”, is designed to emphasize that uncertainty is part of the scientific result, not a limitation. Contributions, within the context of climate change and its impacts, will be invited the following topics:

- describing quantitative and qualitative uncertainty;
- successful exploitation of data with high uncertainties;
- approaches to communicating value, quality and uncertainty in scientific knowledge, how important soft skills are in the communication (i.e. sharing knowledge, etc.) compared to hard skills (i.e. statistics, etc.).

The objectives are:

1. to share information between delegates;
2. to create an authoritative report on best practices, paradigms of success, problems and emerging solutions, particularly with regard:
    - to communicating and exploiting information about uncertainty and data quality;
    - accurate propagation of uncertainty information through the processing chain from environmental measurements or simulations through to policy relevant indicators;
3. to publish the workshop report in a peer review journal.

In order to meet the third objective, a lead author team will be identified well in advance of the workshop, and a series of telephone conferences will be held in advance of the workshop to determine the outline of the report and the associated organisation of breakout groups in the workshop.

The workshop will be limited to around 25 invited delegates. We hope to obtain a broad range of views by inviting delegates from collaborative projects who will be able to represent the breadth of activity in their project team.

The objective of the workshop will be to produce a report for peer review publication which would deal with obstacles and opportunities associated with building confidence in climate services (or building services which deserve confidence), particularly in association with communication of uncertainty.

The meeting will be spread over 3 days, starting and ending at 1pm to allow delegates to, in general, travel to and from the meeting on the first and last day. There will be introductory presentations, breakout groups, and a final plenary session for discussion of the conclusions from the breakout groups. There will be an extra session in the afternoon of the last day for a lead author team.

3.3 Tailoring of Products

There is a need for products to be tailored to the varied needs of different user groups for such products to be useful in their decision making. EUCLEIA has found that there is a clear demand for attribution services across sectors but there are different requirements concerning aspects such as timeliness of products and communication protocols. CLIPC, in seeking to address a spectrum of users from climate scientist, impact scientists, intermediaries and societal end users, has learnt that an information portal needs to speak to all types of users allowing them to navigate through the portal in a natural way without making the specialist versus non-specialist distinction.

4. Lessons Learnt

The main lessons learned have been:

1) Active user engagement is crucial to the development of a climate service to support decision making.
2) It is vital to have a clear understanding of the scientific uncertainties involved and a clear communication of the robustness of assessments for decision making.
3) Products need to be tailored to the varied needs of different user groups for such products to be useful in their decision making.

5. Links Built

Links have been built with CLIPC, EUPORIAS, and QA4ECV in particular through the development of a workshop to discuss different aspects of uncertainties in climate data, focussing on communication of uncertainties.

Seventh Framework Programme
Theme 6 [SPACE]

Project: 607193 UERRA

Full project title:
Uncertainties in **E**nsembles of **R**egional **R**e-**A**nalyses

Deliverable D9.3
Lessons learned

WP no:	9
WP leader:	KNMI
Lead beneficiary for deliverable :	KNMI,SMHI. Met Office, plus more
Name of author/contributors:	Per Undén, plus more
Nature:	Report
Dissemination level:	PU
Deliverable month:	24
Submission date: January xx, 2016	Version nr: 1.1

Common lessons learned relevant for the development of the Copernicus Climate Change Service

Per Undén

Sveriges Meteorologiska och Hydrologiska Institut (SMHI), Norrköping, Sweden

Table of Contents

1. **Introduction**

In the FP7 Space Call (9.2) Projects, UERRA, ERA-CLIM2 and QA4ECV, EUCLEIA (and CLIPC?), there is an overarching Work Package (WP9 or 8) that bridges across the 5 projects and collates common information from the projects in order to communicate to the outside world in a coordinated way.

A deliverable in this Common Overarching WP is a report about the Lessons learned, and in particular Common lessons. They should also have a bearing on the development of the Copernicus Climate Change Services.

The individual Coordinators have summarised their lessons and provided import for this report where the common aspects have been sorted into different areas. Observations availability, reanalysis production from the technical point of view, archiving and data services, quality evaluation are the most important common aspects over most of the projects.

2. **Observations and access to observations**

a. ***Historical observations, availability and data rescue***

In UERRA the Data Rescue and Data Development Activities have been carried out mainly at URV plus a relatively small effort at NMA-Romania, focused on national data. URV and CRU (at UEA) have surveyed both the availability of data in international archives (and particularly MARS at ECMWF which is used by all reanalysis efforts.

There are well known areas around the Mediterranean where there are data voids in space and time, particularly before 1950. What was surprising to see is that there are significant gaps in Northern and Western Europe in the international archives and at ECMWF (MARS). In the 1960s there are very few surface stations from France, Sweden, UK and Norway and low coverage from Poland and Spain.

The Data Rescue has progressed very well due to good staff resources from a University environment. It is possible to combine work on the data with studies and the availability of more than expected resources has enabled a much quicker and bigger number of recovered observations than planned in UERRA (8 instead of 3.7!), when new accessed data in digital format from Catalonia and Slovenia are provided)

b. *Recovery of observations for the reanalyses*

This is particularly serious for the analysis of surface parameters like precipitation and 2m temperature. For countries with missing data, reanalyses will then depend mostly on the background used in the analysis and the precipitation especially will be unreliable.

It turns out that digitisation work is carried out in a number of countries' NMS' but the results may not be finished or that they have not been transferred from national archives to the international data bases. If they were, then there would be large data amounts added to the archives.

Particularly URV, but also NMA-RO to a lesser degree, are digitising quite large amounts, and more than obliged under the UERRA undertakings –especially in the URV case-, but if the other already digitised data can be taken care of and processed, it would dramatically add to the data amounts available for reanalysis.

This process of extracting, collating and transforming the observations from various national (open) data archives is a severely underestimated task, that often has no clear designated responsible (in terms of financed resources).

Other countries show little response to data inquiries or requests. This is generally due to national data policies that are restrictive in a large number of countries. This is not a surprise, however, but those in these 5 Projects and other international bodies must continue to lobby for changes of data policy in favour of contributing to the most accurate climate renalyses that are possible.

3. **Setting up and running the Reanalyses**

The work to set up the reanalyses has been much more demanding than anticipated and planned when the applications (at least for UERRA) were written. It is more complex to find and establish the right version of the model and the data assimilation.

a. *Scientific contents of the models*

In many cases the version or configuration of the forecast NPW model or the data assimilation or the ensemble system is not the one that is used operationally and thus requires more or less development and testing. In the case of the regional reanalyses the large area covering Europe and the Atlantic is normally not the one that is of priority for national NWP. There are well functioning global reanalysis or NWP systems from ECMWF or Met Office, Météo-France and DWD while the intermediate area and resolutions of around 10 km is not the focus of the national NWP any more. (They focus on the mesoscale 1-3 km resolution and hence much smaller horizontal domains than in UERRA.

Thus quit a lot of work goes into finding and using the right model switches for the physics in particular, what is optimal for the 10-12 km resolution. They may deserve more attention if they are not used in the national operational NWP context.

Also the data assimilation may behave different in some ways for the large domain or lower resolution bur it is less of a problem so far.

The ensemble system is particularly challenging as the global systems are well developed and working well whereas the regional systems need the global coupling which may or may not be available in a proper form. Moreover, the regional domains are not small so the main perturbations need to be developed or maintained within the area, not just fed on the lateral boundaries.

Reanalysis systems encompass all the components of NWP (observation processing, data assimilation, land surface, forecast model, verification, technical set-up, etc.). This means that the work is often dependent on expert advice from outside the reanalysis teams, and so sometimes must compete with other priorities.

b. ***Technical contents***

The output of the reanalysis is determined by a series of namelist switches and sometimes in a complex way. In order to provide all the required fields for archiving, they all need to be set and in a right combination for the output. This has to be checked also for sensible values to computed (i.e. non-zero to start with).

Sometimes long tests or tests by some user are necessary to spot problems. Then the output fields need to be checked for “normal” behaviour e.g. compared with climate or earlier reanalyses. Where the reanalyses are being run over several decades, it may be necessary to tune the systems separately for different periods. Observation and model background errors may change, in addition to the observing network.

For the efficiency reasons there are also tuning parameters and maybe script and code changes to speed up any inefficient parts of the reanalysis system.

c. ***Experiences of ensemble assimilation***

Three new developments of regional ensemble data assimilation systems are taking place in UERRA: The Met Office Ensemble 4D-VAR, DWD/University of Bonn Localised Ensemble Transform Kalman Filter (LETKF) and multi-physics/model ensemble at Météo-France (with two physics runs from SMHI).
All the methods are a) particular to the UERRA application in that is is for a regional domain (rather than global or meso-scale sub-regional) and at medium resolution of 5-12 km, different from their main NWP configurations and b) new in terms of ensemble generation methods.

Moreover, the large domain regional ensemble systems need lateral forcing from a global ensemble system and with the same number of ensembles as the regional system. This is often hard to achieve, but a reasonable number of ensembles are being produced at ECMWF (20). Still is is a factor to consider and there is not enough experience yet how well they work in the regional systems. Early tests have shown signs of divergence in the ensembles and too large spread which were solved later. The LETKF system at DWD/UBO is new for this application and has both technical and scientific problems at the time of writing. On the other hand, the ensemble nudging assimilation works very well. The scientific work of ensemble reanalyisis at UBO is also carried out with the well proven ensemble nudging assimilation. The multi-model/method system for the surface downscaling reanalysis on the other hand, seems to have too little spread for precipitation but it is reasonable for the temperature.

d. ***Reanalyses experiences and possible problems***

The actual long period reanalyses have not started for most except SMHI, but with a significant delay. Several computational, technical and scientific obstacles had to be surpassed before the “production” could start. It is also crucial to check and diagnose the quality of the output fields (and that they are there at all in the first place of course).

In particular the actual production is more demanding than anticipated and there is a need of daily supervision to check and make sure that the runs are active. NWP monitoring tools aren't always suitable for these long-period runs and so new tools need to be put in place.
On the scientific side, the long term biases need to be monitored and compared with other references and there may be some aspects of the system that introduce biases, usually a component of the model parameterisation. In combination with changes in the observation coverage, this may introduce spurious climate trends. These aspects may or may not be difficult to handle but we need to monitor them and so far the experience is too limited.

4. **Lessons for evaluation**

The evaluation of the uncertainties has been developed without the actual UERRA reanalyses while we have discussed the archiving and before most of the reanalyses have been run.
The discussion about the common output parameters of the models was quite prolonged since it was partly driven by the evaluation WP demands.

Instead the evaluation software was developed using national reanalyses from DWD and national surface and mast data and it has not hampered the progress except that the actual interface programmes and scripts to the final UERRA archives had not been built yet, but source the data from a preliminary archive. Thus evaluation is now restricted to parameters present in the preliminary archive. The communication with users was found helpful to decide on priorities concerning physical parameters, observations to compare to, scales and skill scores to investigate.

The collaboration between the partners, led by DWD and with MET Norway and Meteo-Swiss and others has evolved in a very productive way and a software package is provided on an open github to which it is easy to contribute.

5. **Lessons for the common archive**

It was thought to be relatively straight forward to just pass the “normal” set of model and analysis parameters to the MARS archiving system. However, in UERRA there are 4 partners with 4 different systems and producing many of the same parameters but also many that are more or less different or that don't exist for all systems. It turns out a) that selecting “everything” and from all 4 systems will just overwhelm the archiving system and that there is less value in storing fields that only exist in one system and not the others. b) The archives should be user driven so the WP3 in UERRA (Estimating uncertainties) had a long dialogue with the reanalysis providers in WP2 about which output fields would be needed in UERRA as well as for the user community at large. c) Then the output parameters needed to be clarified and properly described for the purpose of archiving in GRIB2, which none of the participants had done before.

Also the output fields are produced by various parts of the systems and are spread in different files where there are also superfluous fields. Scripts to read and extract exactly those fields to be archived are being built.

EUROPEAN COMMISSION

Community research

Reporting period:

Climate Information Portal

CLIPC DELIVERABLE (D -N°: 11.3)

Lesson Learnt

File name: {CLIPC_D11-3_lessonsLearnt_nearFinal}

Dissemination level: PU (public)

Revision table			
Version	**Date**	**Name**	**Comments**
0.1	June 2016	First version	Circulated for discussion and internal review
0.9	July 9th, 2017	Near Final	Shared with over-arching WP.

Abstract

This document gives an overview of lessons learnt during the execution of the CLIPC project, as the CLIPC contribution to the "Lessons Learnt" deliverable of the over-arching coordination work package (D11.3 in the CLIPC Description of Work).

Grant agreement no. 607029

SEVENTH FRAMEWORK PROGRAMME

Release Date	Issue Date

Table of contents

1 Background

The CLIPC project (A Climate Information Platform for Copernicus) is developing a prototype climate information portal and platform for the Copernicus Climate Change Service (C3S). It is one of 5 projects funded at the end of FP7 to contribute to the development of C3S component services. The others are ERA-CLIM2, UERRA, QA4ECV and EUCLIEA. These 5 projects share a joint work-package facilitating coordination between them, and as part of that work this document provides an overview of lessons learnt in the CLIPC project. As context, the CLIPC mission is provided in the next section: further detail about CLIPC can be found on www.clipc.eu, including all the project deliverables (www.clipc.eu/deliverables-and-milestones). Section 3 lists the lessons learnt by work package, and a summary is given in work package 4.

2 The CLIPC mission

- CLIPC will design a platform to provide access to climate information of direct relevance to a wide variety of users, from scientists to policy makers and private sector decision makers;
- The "one-stop-shop" platform will provide data and information on climate and climate impacts, and ensure that the providence of science and policy relevant data products is thoroughly documented;
- Engage with user communities to inform development.

3 Lessons Learnt, by work package

WP2: User Engagement

4. The importance of distinguishing different user categories as each has its own requirements and preferences.

5. There is need for facilitating tailored co-creation processes involving both data providers and different types users to develop products meeting the varied needs of different user categories.
6. Users require different forms of guidance to assist them in navigating through the portal and use it for decision making.

WP3: Portal

- While user requirements may drive initial portal developments, we found out that users need to see live demonstrations and examples before expressing clear choices. Therefore we applied early user consultation and development iterations. For an optimized end result there could even be more rounds (short cycles).
- The challenge is to keep user interfaces simple, and often simple for developers and involved experts is by far not simple enough for (first time) users.
- It is important to direct the right type of users to the right service or user interface, otherwise users get confused and leave the website. Important to provide enough introduction in the website for a user to find out what fits his/her interest.
- You can never give enough user guidance.

WP4 Service Integration

7. Agile approach and mind-set is necessary for co-development (WP2,3,4,6,7,8): there is no one time right: not for the defined indices, not for the developed services. Develop, demonstrate and improve in short cycles works.
8. Regular direct interaction with project partners during specific targeted 'coding sprints' are very productive for integrating services. A coding sprint is a 1-2 day physical meeting, where issues around a specific topic (e.g. combine tool) are identified and directly worked on by involved partners. At least once each six months would be best, but not always feasible.
9. Coming from different scientific communities, getting to the point where all are speaking the same 'language', working together, and all have a shared vision and a realistic implementation plan takes time. This explains the late start of the implementation in the project. A longer project with more time to develop the implementation plan would have many advantages.
10. Showing why and how meta-data is used in the services and tooling helps to convince data providers to be more involved in this subject (often considered a distraction from the primary task of delivering new science).
11. The use of standardized services such as Open Geospatial Consortium (OGC) web services[1], OPeNDAP[2] and user security tokens is essential for easy integration of front-end and back-end services.

[1]Web Map Service (WMS), Web Coverage Service (WCS), WPS (Web Processing Service) and CSW (Catalogue Service for the Web)
[2]Open-source Project for a Network Data Access Protocol, opendap.org

WP5: Data Service Harmonisation

- Well structured meta-data is key to the efficient handling of diverse data types.
- The development of effective data protocols depends on establish a good working relationship between data management specialists and data scientists.
- Making data available through services is a non-trivial chain of activities which needs to involve the data management specialists, service providers and data producers.
- Standards come in a range of varieties and flavours, the trick is to pick the ones that make the job easier and to tailor them, but not too much, to the project needs.
- Making the data available through services does not address any requirements for data preservation.
- It can be difficult to engage data producers in the "tedious" task of agreeing conventions, and applying them. Providing example files helps, as well as giving feedback on early versions of the data.

WP6: Data Harmonisation

The work in WP6 has been focus on three main themes: i) bias correction of climate model data, ii) integrating earth observation and satellite data with climate model data, and iii) developing methods for reducing the size of climate scenario ensembles.

- Depending on context and application users may request either bias-corrected or ‘raw’ model data. Bias correction leads to increased realism of data products for certain applications, but the extra processing step makes uncertainty analysis more difficult. Currently, there is no accepted best method for bias correction; it depends on what aspect of climate is in focus. Development of new bias correction methods is currently a very active area. The higher demand the user put on the climate model data the more advanced and complex the correction methods become.
- Earth observation data provide essential information on the state of the climate system, and its recent past (up to a few decades). Typically, the focus is on (near-) real time monitoring, with only a limited number of data sets having long and homogeneous enough time series for climate applications. Thus, in many respect this data complements the data that are derived from climate models. While integration of earth observation data with climate models is a highly interesting idea, there are currently several basic data mismatches that need to be overcome: For climate studies long homogeneous time-series is of paramount importance, and this is often challenging for EO data sets to meet these requirements. Furthermore, there is a conceptual scale mismatch; in the context of EO “high resolution” often means 10’s (or 100’s) of meters, while for regional climate models it means ‘several kilometres’, or for earth system models it means ‘less than 100 km’. However, we expect that the situation improves with time as the EO datasets are extended in time, and thus become relevant for climate studies, as well as new methods and analysis tools are developed. In particular, new reanalysis drawing on development in data assimilation is expected to contribute to this.
- Development of reduced ensembles that still convey the essential climate change information as the full scenario ensemble need to take into account what is actually meant by “the essential climate change information”. This varies from user to user and from application to application. This has led to the development of a method and

tool that will allow the users to produce the reduced ensemble according the their specific needs.

WP7: Impact Indicators

The scientific work in WP7 has been one of 1) reviewing, documenting and providing Tier 1, 2 and 3 impact indicators, 2) guarantee the appropriate provision of metadata describing the indicator and 3) upload the indicators to the portal developers for their inclusion in CLIPC. The main lessons across the three objectives here:

- The phrase "climate impact indicator" conveys different meanings to different communities. By defining "tiers" (Box 1) of data products CLIPC has established a common terminology, but there is still much room for confusion.
- There is considerable redundancy in the availability of Tier 1 indicators (these being supplied by many portals), but the Tier 3 indicators are in short supply.
- Development of necessary metadata standards (including agreed terminology) is time consuming, especially when existing standards are weak.
- Regular technical meetings have been needed to keep the many diverse partners working together
- Google spreadsheets have been useful for monitoring progress within the WP.
- Construction of a matrix to provide guidance on the compatibility of indicators can only minimize the risk of users making "meaningless" combinations, it will not eliminate that risk.

Box 1: The tiers of data products

Tier 0: States of the Climate System

Tier 0 includes the output of climate models, satellite observing systems and surface networks.

Tier 1: Statistics of Climate Data

Tier 1 products are evaluated from the Tier 0 data. This includes statistics such as the "Standardized Precipitation Index". The tier 1 products provide information about the climate state in a format which is useful to users outside climate science.

Tier 2: Impact on the Environment

The tier 2 products include more than information about the climate state. Products such as projected "Flooding Extent" rely on climate data inputs but also on a flood model.

Tier 3: Impact on society

Tier 3 indicators are the most directly relevant to policy makers, but also the hardest to generate.

WP8: Data aggregation and exploration

- The various WP8 tools require easy to understand user interfaces so that users can quickly understand and effectively use them.
- For more complex combination of indicators (e.g. more sophisticated users) it was a good decision to offer an additional, step-by-step tool with advanced functionalities (as part of the 'CLIPC processing tool).
- Carefully selected and documented 'use cases' help users understand the capabilities of the WP8 tools – especially how users can go through a sequence of analytical steps (using several CLIPC tools) in order to achieve a certain result.
- Being able to tap into existing climate databases (e.g. ESGF) and transfer data into

the CLIPC portal offers users an abundance of possible choices when using CLIPC tools. CLIPC can thus function as a toolbox that is layered on top of existing databases.

- WP8 tool development benefited especially from the close cooperation between user consultation (WP2), user interface and portal design (WP4), impact indicator development (WP7) and indicator exploration tool design (WP8). This was most evident in the progress made in joint ‘coding sprints’ that brought together scientists and developers from these work packages.

WP9: Scientific Coordination

- Active user engagement not only helps the project to stay in touch with what users want, but also has facilitated communication between different domains by forcing all to communicate in a jargon free language. This has helped to mitigate one of the major problems, the huge differences in approach and methodology between the many disciplines which need to be brought together to create a credible climate service.

WP 10 Project management and project internal communications

- Regular scheduled telephone calls enable effective communication.
- Having project management teleconferences with the work package leaders to review progress can be tedious, but it does ensure that everyone is aware of what is going on with the other work packages. Regular calls are needed, preferably at a set time each fortnight/month. We started with quarterly calls, which weren’t enough to keep each other informed and up to date. Currently we have fortnightly calls, of an hour length, which seem to be working well.
- Collaborative editing tools are very useful for compiling reports and deliverable documents. The effectiveness of tools depends greatly on the degree of familiarity that people have; managing access controls for bespoke systems causes confusion and delays. Careful use of familiar tools (e.g. google spreadsheets for monitoring progress on deliverables and milestones) can be very effective.
- There is a trade-off to be made between having one single email list for the project (can result in too many mails being sent, which people then have to filter to what’s of interest to them) and multiple email lists (where there is the risk of missing things if you’re not on the right list). Being clear in advance on the correct usage of the email lists can help with this
- Similarly, for situations where the admin/financial aspects of the project are being managed at the partner institutions by different people from the scientific/technical aspects, it’s important to ensure that the list of admin contacts is kept up to date, and that the technical people know that they might have to remind their admin people of admin tasks, if needed.
- We planned to have regular quarterly newsletters produced by the project, but this was dropped in favour of ad hoc news items put on the website, as determining suitable stories for the newsletter was very difficult, especially in the early stages of

the project.

4 Summary

The importance of user interactions in defining services and setting priorities has been clear through many work packages. There is a need for tailoring of services and products for different user groups – the one-stop-shop cannot work on the principal of “one-size fits all”.

Communication was also a significant issue within the consortium as the broad range of backgrounds brought many varied expectations to the CLIPC table. The terminology used to describe data products varies between different specialities. Time and a systematic approach are needed to deal with these issues.

Collaborative tools are necessary, and it takes time to find an approach which suits all partners.

Well defined standards are always welcome, but the process of establishing standards is not well understood and there are many different approaches.

QA4ECV

To be added in Month 48 of the QA4ECV project