The EU-FP7 ERA-CLIM2 project contribution to advancing science and production of Earth-system climate reanalyses

Roberto Buizza, Stefan Brönnimann, Leopold Haimberger, Patrick Laloyaux, Matthew J. Martin, Manuel Fuentes, Magdalena Alonso-Balmaseda, Andrea Becker, Michael Blaschek, Per Dahlgren, et al.

The EU-FP7 ERA-CLIM2 project contribution to advancing science and production of Earth-system climate reanalyses

Roberto Buizza (1), Stefan Brönnimann (4), Leopold Haimberger (5), Patrick Laloyaux (1), Matthew J. Martin (2), Manuel Fuentes (1), Magdalena Alonso-Balmaseda (1), Andreas Becker (10), Michael Blaschek (5), Per Dahlgren (1), Eric de Boisseson (1), Dick Dee (1), Marie Doutriaux-Boucher (3), Feng Xiangbo (14), Viju John (2,3), Keith Haines (14), Sylvie Jourdain (9), Yuki Kosaka (1), Daniel Lea (15), Florian Lemarié (15), Michael Mayer (5), Palmira Messina (16), Coralie Perruche (8), Philippe Peylin (16), Jounie Pullainen (13), Nick Rayner (2), Elke Rustemeier (10), Dinand Schepers (1), Roger Saunders (2), Jörg Schulz (3), Alexander Sterin (7), Sebastian Stichelberger (5), Andrea Storto (12), Charles-Emmanuel Testut (8), Maria- Antóonia Valente (6), Arthur Vidard (15), Nicolas Vuichard (16), Anthony Weaver (11), James While (2), Markus Ziese (10)

1) European Centre for Medium-Range Weather Forecasts (ECMWF; Europe);
2) Met Office, UK (UKMO; UK);
3) European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT; Europe);
4) University of Bern (Switzerland);
5) University of Vienna (Austria);
6) Instituto Dom Luiz Faculdade de Ciências da Universidade de Lisboa (Portugal);
7) Russian Research Institute of Hydrometeorological Information (RIHMI; Russia);
8) Mercator Ocean Societé Civile (MERCATOR; France);

9) Météo-France (MF; France);
10) Deutscher Wetterdiens (DWD; Germany);
11) Centre European de Recherche et de Formation Avancée en Calcul Scientifique (CERFACS; France);
12) Centro Euro-Mediterraneo Sui Cambiamenti Climatici (CMCC; Italy);
13) Ilmatieteen Laitos (FMI; Finland);
14) The University of Reading (UK);
15) Institut National de Recherche en Informatique et en Automatique (INRIA; France);
16) Laboratoire des Sciences du Climat et de l'Environnement (LSCE) France.

Corresponding author: Roberto Buizza, European Centre for Medium-Range Weather Forecasts (ECMWF), Shinfield Park, Reading, RG2-9AX, U.K..

Submitted to the Bulletin of the American Meteorological Society (BAMS), 7 August 2017.

Capsule Summary

The main goals and activities of the ERA-CLIM2 project are presented, and some of its key results, included the first ensemble of coupled reanalysis of the 20th century, are discussed.

Abstract

ERA-CLIM2 is a European Union Seventh Framework Project started in January 2014. It aims to produce coupled reanalyses, which are physically consistent data sets describing the evolution of the global atmosphere, ocean, land-surface, cryosphere and the carbon cycle. ERA-CLIM2 has contributed to advancing the capacity for producing state-of-the-art climate reanalyses that extend back to the early 20th century. It has led to the generation of the first ensemble of coupled ocean, sea-ice, land and atmosphere reanalyses of the 20th century. The project has funded work to rescue and prepare observations, and to advance the data-assimilation systems required to generate operational reanalyses, such as the ones planned by the European Union Copernicus Climate Change Service. This paper summarizes the main goals of the project, discusses some of its main areas of activities, and presents some of its key results.

1. The ERA-CLIM2 project

ERA-CLIM2 (European Reanalysis of Global Climate Observations 2) is a 4-year research project funded by the European Union Seventh Framework Program (EU FP7; Grant Agreement No. 607029; see Appendix for a list of the Consortium's Institutions), that started on the 1st of January 2014 following the successful completion of its predecessor project ERA-CLIM. The research initiated in these two projects underpins a concerted effort in Europe to build the information infrastructure needed to support climate monitoring, climate research and climate services, based on the best available science and observations. ERA-CLIM2 is one of several collaborative research projects designated by the European Commission as precursors to the EU's Copernicus Climate Change Service (C3S). Indeed, ERA-CLIM2 activities on data rescue, satellite data rescue and reprocessing, coupled data assimilation and reanalysis production have had a strong impact on the design and implementation of the C3S. Furthermore, activities performed in ERA-CLIM2 and aiming at improving the assimilation of observations at inter-medium boundaries, i.e. sea surface temperature and sea-ice concentration observations, will certainly benefit ocean and sea-ice reanalyses and operational marine forecast activities performed within the EU's Copernicus Marine Environment Monitoring Service (CMEMS).

The aim of the ERA-CLIM2 project is to enable production of state-of-the-art reanalyses of the coupled climate system for the 20th century, following earlier work on coupled and long-term climate reanalysis at the National Center for Atmospheric Research and the National Centers for Environmental Predictions (NCAR and NCEP; *Saha et al.*, 2010, *Compo et al.*, 2011), and the European Centre for Medium-Range Weather Forecasts (ECMWF) strategy described in a previous BAMS article (*Dee et al.*, 2014). Climate reanalyses are physically consistent data sets derived from observations that document the recent evolution of the global atmosphere, ocean, land-surface, cryosphere and the carbon cycle. Reanalysis data are generated by a sequential process called data assimilation, which combines first-guess estimates defined by short model forecasts with vast amounts of data from a range of observing platforms (surface, upper-air, satellites). Climate reanalyses provide long (multi-decadal) time series of gridded estimates for many different climate variables, which are used to study past weather events, estimate climatologies, monitor climate change, and supply crucial data on climate needed for science and applications.

Reanalysis is a complex activity that requires large computational resources, access to observations from many providers, and expertise in multiple disciplines. During the past three decades, successive reanalyses of the global atmosphere have been produced by NCAR and NCEP (*Kalnay et al.,* 2006, *Saha et al.,* 2010), the Japan Meteorological Agency (JMA; *Onogi et al.,* 2007), and the National Aeronautics and Space Administration (NASA; *Schubert et al.,* 1993, *Rienecker et al.,* 2011), in addition to ECMWF. A global reanalysis extending back to the late nineteenth century was first produced by NOAA in collaboration with the Cooperative Institute for Research in Environmental Sciences (CIRES), using only surface pressure observations and prior estimates of sea-surface temperature (SST) and sea-ice distributions to avoid the effects that large changes in the observing system could have had on the reanalysis (*Compo et al.,* 2011).

ERA-CLIM2 has contributed to advancing reanalysis science and development in four main areas:

i. *Observation data rescue and post-processing*: activities under this theme include a large effort on data rescue for historic in-situ weather observations around the world, and substantial work on the reprocessing of satellite climate data records and enabling the use of historical satellite data for reanalysis;

ii. *Data assimilation methods*: activities under this theme aim to progress the development and testing of ‘coupled assimilation methods’, capable of including observations from different Earth-system components (land surface, ocean, sea-ice, atmosphere, chemical components, …) to produce a more consistent estimate of the Earth-system evolution, especially at the surface;

iii. *Reanalysis production*: activities under this theme aim to generate the innovative reanalysis data-sets, such as the first European coupled ocean-land-atmosphere reanalysis of the 20th century, and to provide access to the reanalysis data;

iv. *Evaluation and uncertainty estimation*: activities under this theme aim to assess the reanalyses’ quality and how products differ from previous uncoupled products, and to develop methods for estimating uncertainty in reanalyses.

Hereafter, we will discuss briefly some of these activities.

2. Reanalysis as a tool to monitor the climate

The Earth’s climate has traditionally been studied by statistical analysis of observations of particular weather elements such as pressure, temperature, wind and rainfall. These meteorological observations are temporally and spatially incomplete and are often presented in terms of long-term averages to identify evidence of climate change. Reanalyses provide a more complete source of data to understand and monitor the climate. In a reanalysis, the weather observations collected in past decades are fed into a modern forecasting system designed to assimilate observations from existing and planned platforms (e.g. forthcoming satellite instruments), which provides a physically consistent description of the Earth system. By constantly correcting a model simulation towards the observations, a reanalysis combines the advantages of the model first guess and the available data (further details in section 4). Therefore, reanalyses are physically consistent, spatially complete and encompass many variables for which observations are not always available. Irregular and intermittent observation sampling throughout the reanalysis period, especially during early periods, might however prevent the reanalysis dataset from showing temporal homogeneity at both global and regional scale (e.g. *Ferguson and Villarini*, 2012; see also Section 7).

Since its creation in 1975, ECMWF has been a key player in the production of reanalyses. The initial focus was on producing atmospheric reanalyses covering the modern observing period, from 1979. The first of these reanalyses, FGGE (First GARP Global Experiment, where GARP stands for Global Atmospheric Research Program), was produced in the 1980s at the Geophysical Fluid Laboratories (GFDL; *Ploshay et al.*, 1992), followed by ERA-15 (European Reanalysis, version 15-years), ERA-40 (*Uppala et al.*, 2005) and ERA-Interim (*Dee et al.*, 2011). The next reanalysis in this series, ERA5, is now in production after many years of research and development.

Generating a reanalysis for climate monitoring is very challenging because it needs to be extended further back in time when the observing system was very sparse, especially before the availability of satellite data from the 1970s onwards, and even more so before the arrival of radiosonde measurements in the 1930s. To tackle the unavoidable issue of the ever-changing observational network, the European Union started the ERA-CLIM (European Reanalysis of Global Climate Observations) project, to investigate data selection for reanalyses covering the whole 20th century. As part of the ERA-CLIM project, ECMWF produced the uncoupled atmospheric reanalysis ERA-20C, which covers the period January 1900 to December 2010 (*Poli et al.*, 2016). The ERA-20C reanalysis assimilated only conventional observations of surface pressure and marine wind, obtained from well-established climate data collections. ERA-20C delivered three-hourly products describing the spatial and temporal evolution of the atmosphere, land surface and waves.

As part of the FP7 ERA-CLIM2 project, the reanalysis capabilities developed in the ERA-CLIM project have been extended to the ocean and sea-ice components. A new assimilation system (CERA, the Coupled European Reanalysis system) has been developed to simultaneously ingest atmospheric and ocean observations in the coupled Earth system model used for ECMWF's ensemble forecasts (*Laloyaux et al.*, 2016, 2017). This approach accounts for interactions between the atmosphere and the ocean during the assimilation process and has the potential to generate a more balanced and consistent Earth system climate reconstruction (Figure 1). CERA has also been found to reproduce better the observed negative SST-precipitation relationships on monthly timescales due to the resolving of atmospheric feedbacks on SST (Figure 2). Efforts are being made to investigate whether this improvement will improve the prediction of final precipitation (*Feng et al.*, 2017b). One of the key deliverables of the ERA-CLIM2 project is CERA-20C, the first ten-member ensemble of coupled climate reanalyses of the 20th century. It is based on the CERA system, which assimilates only surface pressure and marine wind observations as well as ocean temperature and salinity profiles. The data are openly available from the ECMWF platform http://apps.ecmwf.int/datasets/.

There is now a strong need for detailed information of CO2 fluxes and carbon pools from the climate modelling community who want to understand and quantify the carbon cycle at global and regional scales, and from policy makers and citizens who want to take well-informed decisions on CO2 emissions at regional and local scales. For this reason, ERA-CLIM2 is also producing associated global reanalyses of carbon fluxes and stocks using terrestrial biosphere and ocean biogeochemistry models, which are forced by the CERA-20C reanalysis.

A new version of the CERA system has also been developed based on a higher resolution coupled model with the full observing system. This system is now used to produce the

CERA-SAT (the Coupled European ReAnalysis of the SATellite era) reanalysis which covers the period 2008-2016.

The availability of atmospheric, oceanic and coupled reanalyses allows for new advanced coupled diagnostics of the global energy cycle. *Mayer et al.* (2017) improved the classic method of evaluating the vertically integrated atmospheric energy budget such that it is independent of reference temperature and consistent with diagnosed ocean heat budgets. *Pietschnig et al.* (2017) demonstrated the value of comparing the net energy transport through all major Arctic Ocean gateways from ocean reanalyses with independent mass-consistent transport estimates from instrumented mooring observations.

3. Observation data rescue

Reanalyses efforts strongly depend on observations of the atmosphere, land surface, the ocean and cryosphere. Observations are assimilated into a coupled general circulation model in order to produce the reanalysis, but observations are also used in several other steps. They are used to constrain the boundary conditions, to calibrate certain relations and to validate the final product. Particularly when going back in time, not all observations are available. A large fraction of historical meteorological observations has never been digitised because the data have thus far not been considered valuable. Even in the rather recent past, the availability of satellite products (and the computer code to read and process these data) is an issue that needs to be addressed.

Major efforts were therefore undertaken in ERA-CLIM2 to collect and make available observations for reanalyses (*Brönnimann et al.*, 2017). Such an undertaking requires a much broader vision than the production of one reanalysis. Availability of historical observations becomes a legacy, and producing reanalyses or other data products must be seen as a continuous effort.

Within ERA-CLIM2, millions of radiosonde, pilot balloon and other ascending instruments' profiles were digitised, which allows describing the third dimension of the atmosphere back to the early 20th century. Although the radiosonde data were not yet incorporated into the CERA-20C reanalysis produced in ERA-CLIM2, they were used for the validation of reanalysis products. A test reanalysis that included the historical upper-air data and demonstrated the potential benefits was performed for the period 1939-1967 (*Hersbach et al.,* 2017). By the end of the project, all digitized upper air data will be available in assimilation-ready format with bias adjustments for radiosonde temperatures (*Haimberger et al.*, 2012) extending back to 1939. Future reanalysis efforts will be able to incorporate this vast amount of upper-air data and will thus build on the ERA-CLIM2 efforts.

Surface pressure and mean-sea-level-pressure data were also digitised for several countries in both the Northern and Southern Hemisphere (NH and SH), some with sparse observation networks, and sent to the International Surface Pressure Databank (ISPD). These will be assimilated in forthcoming ISPD versions to be used in future reanalyses inputs. Many other land surface, daily and sub-daily, observations of temperature, relative humidity, surface wind, cloud cover, precipitation, evaporation and sunshine duration were rescued, subjected to Quality Control procedures and can be used for reanalyses comparison purposes (*Brönnimann et al.,* 2017).

Snow is an important component of the climate system, at the interface of the land-surface, vegetation and atmosphere. It is highly relevant for various fields such as ecology, water resources, transport, and tourism. By digitising large amounts of historical snow data and combining this with satellite products, ERA-CLIM2 generated snow products for various uses, including (but not limited to) reanalyses. Snow courses are specified paths of a few km length around a location: along this path, different snow properties are measured regularly. Thanks to ERA-CLIM2, snow course about 1.2 million snow course observations have been collected from close to 400 stations compiled and made available.

Satellite data are the backbone of today's reanalysis data sets, but they became only available in the mid-1960s and were built for the purpose of weather monitoring. The use in reanalysis requires a quantification and correction of long term effects due to systematic differences between satellite instruments of the same kind and changes in the characteristics of satellites and performance of sensors during their operational lifetime in space. A reprocessing of the data that applies the corrections is fundamental to serve the generation of physically consistent data records of geophysical variables by reanalysis. In ERA-CLIM2, efforts were put into re-processing of satellite data of infrared and microwave radiances from geostationary imagers and microwave sounders, radio occultation bending angle profiles for several satellites and atmospheric motion vectors from different instruments in geostationary and polar orbit. In addition, as part of satellite data rescue activities (*Poli et al., 2017*) radiative transfer calculations for some early satellite instruments were enabled to allow for monitoring and assimilation of the satellite measurements using the circulation model which also supports the characterisation and correction of instrument issues. Indeed, a more comprehensive use of early satellite data is expected to improve future reanalyses.

4. Data assimilation methods for reanalysis

The ERA-CLIM2 reanalyses have been produced using a state-of-the-art data assimilation system, capable of combining observations in the atmosphere and ocean with a coupled ocean-land-atmosphere model. Part of the work within ERA-CLIM2 was devoted to improve such a data-assimilation system, testing also new methods and ideas that could be used in future reanalysis productions.

A crucial part of the coupled climate system is the interface between the ocean/sea-ice and atmosphere (*Feng et al.,* 2017). In the existing CERA system, the sea surface temperature is constrained to follow a global observational analysis, such as HadISST2 (*Titchner and Rayner* 2014) or OSTIA (*Donlon et al.* 2012), which are calculated externally. Enabling the next CERA system to assimilate directly the wealth of high quality satellite SST observations should allow the system to combine them with the temperature profile data and the coupled model background in a more consistent manner, thereby improving the accuracy of the reanalysis. Care has to be taken to deal properly with biases in the satellite data that could otherwise introduce spurious trends. In addition, the in situ ocean data are sparse, particularly in the early part of the 20th Century.

Figure 3 shows an example of the results from using a new method for effectively spreading sparse observational information in the data assimilation. Developments have also been made to improve the assimilation of sea-ice concentration during the satellite era. Sea-ice concentration estimates have error characteristics which make it difficult to assimilate in most data assimilation algorithms so techniques have been tested which transform the sea-ice concentration into a form which has Gaussian errors.

The ocean data assimilation system used in CERA is a state-of-the-art three-dimensional variational system called NEMOVAR (Nucleus for European Modelling of the Ocean Variational assimilation system). It is now a very flexible framework for assimilating data into the NEMO ocean model, and includes developments to allow the ocean data to be assimilated using sophisticated techniques similar to those used in the atmosphere (*Weaver et al.,* 2016). The system can now be configured to use information from an ensemble of model runs to improve the assimilation, which is a more sophisticated system than the algorithm used in the production of the ERA-CLIM2 reanalyses. In the present CERA system, the ocean bias correct scheme is not applied during the ocean data assimilation as it normally needs to be estimated from a priori run. However, now with CERA-20C being finished, it becomes practically possible to implement this scheme throughout the whole 20th century. Implementing the bias correction scheme will avoid the ocean model to produce a spurious reaction to adjust the imbalance between the ocean and atmosphere initial conditions. Efforts are also being made to investigate if the implementation will benefit the atmosphere analysis. Furthermore, a four-dimensional version of the ocean assimilation system has been developed and is under testing.

Currently the coupled model is introduced at the outer-loop level in the CERA system by coupling ECMWF's Integrated Forecasting System (IFS) for the atmosphere, land and waves to the NEMO model for the ocean and to the LIM2 model for sea ice. This means that air–sea interactions are continuously taken into account when observation misfits are computed and when the increments are applied to the initial condition. This allows feedback between the ocean and atmosphere models (*Laloyaux et al.* 2016, 2017). Investigations into whether the atmosphere observations could be used to directly correct the ocean model, and vice versa, have been carried out in ERA-CLIM2 (*Storto et al.,* 2017; *Pellerej et al.,* 2016). This technique, known as "strongly coupled data assimilation", could allow even better use of sparse observations. There are many open research questions in the development of strongly coupled data assimilation, so ERA-CLIM2 partners were involved in the organisation of a Coupled Data Assimilation workshop in October 2016, sponsored by WMO, to discuss these with the wider international research community.

Land and ocean carbon reanalyses are also being produced as part of ERA-CLIM2 and the techniques used to produce them are being developed for use in future reanalyses. Various new data streams have been tested for improving the accuracy of the land carbon reanalysis both through improved state-estimation, and through improving knowledge of the model parameters (*Peylin et al.,* 2016). The methods used to couple ocean biogeochemical models to the physical ocean/atmosphere reanalysis system have also been investigated in order to provide improved information about the ocean carbon cycle.

5. Reanalysis and society

ERA-CLIM2 demonstrated that reanalyses can be used to understand past events and provide valuable information to present day society. A demonstration case was discussed by *Brugnara et al.,* (2017), who examined the weather conditions in December 1916, in the middle of the First World War, when a massive snow fall event in the Southern Alps triggered countless avalanches, which killed thousands of soldiers and civilians. This event was studied using dynamical downscaling of the earlier ERA-20C reanalysis (the uncoupled reanalysis produced by the FP7 project ERA-CLIM, the precursor of ERA-CLIM2) in combination with historical observations. By looking at reanalysis data (Figure 4), the atmospheric conditions that led to such catastrophic events could be understood: a blocking flow situation, moisture transport from the warm Mediterranean Sea towards the Alps, and a rapidly rising snow line, leading to a dangerous “rain-on-snow” situation. Historical events that are captured by the ERA-CLIM reanalyses can thus inform present day risk management.

6. A new approach: ensembles of reanalyses to estimate confidence

The accuracy of any physical measurement is limited. Furthermore, the spatial resolution of measurements and of assimilated gridded data is limited, and thus there may be deviations of the grid point values from the true values. Confidence (or uncertainty) is best described by the probability distribution of these deviations. The distribution itself can be characterized by a few parameters such as standard deviation, or by a limited-size ensemble of realizations drawn from the distribution, from which the user may derive statistics. The latter approach consumes much more data storage but also leaves more choices for informed users.

With reanalysis products getting more mature, it is now possible to quantify their accuracy and assign to the reanalysis data a confidence level. In ERA-CLIM2, this has been achieved by applying an ensemble approach, based on several complete realizations of all quantities.

Both the 20th century reanalysis CERA-20C, and the upcoming coupled reanalysis of the satellite era, CERA-SAT, consist of ten realizations, run in parallel. The ten members can be used to estimate a range of possible states for all the reanalysis output variables. They have been generated using an Ensemble of Data Assimilation (EDA) in the atmosphere (*Isaksen et al.* 2010; *Bonavita et al.* 2016) and with perturbing the positions of in-situ ocean observations, the air-sea fluxes and the SST following ECMWF's Ocean ReAnalysis System-5 (ORAS5) (*Zuo et al.* 2016). The SST perturbations lead to variations in important output parameters such as surface precipitation. Figure 5 shows how CERA-20C precipitation has improved over the earlier 20th century uncoupled reanalysis ERA-20C (*Poli et al.,* 2016) over less well observed land areas, which exhibited a strong under-forecast at the higher quantiles in Africa and at the Monsoon areas for the 0.9 quantiles (*Rustemeier et al.,* 2017). Figure 5 also shows how the calculated precipitation varies within the ensemble. For example, for strong precipitation episodes over Africa (upper right of panel a) the spread has been increased by about 20%, making the ensemble more reliable.

7. CERA-20C: the first European 110-year coupled ocean-land-atmosphere reanalysis

The accurate representation of variability on inter-annual and decadal time scales is a requirement for climate applications of reanalysis data, such as reconstructing the time evolution of the atmosphere surface temperature and of the ocean heat content (Figure 6). Climate signals in reanalyses are inevitably affected by changes in the global observing system and by the presence of time-varying biases in models and observations. To build confidence in climate change information derived from reanalyses, it is important that information about the data assimilation methodology, the forecast model, and the input observations are made available. It is also necessary to compare results based on reanalyses (ECMWF CERA-20C, NOAA 20CR, ECMWF ORA-20C), with results obtained using more traditional, observation-only climate datasets (CRUTEM4 two-meter temperature, GPCC precipitation data, EN4 ocean temperature), and to test whether the climate signals in CERA-20C are robust to different analysis methodologies. CERA-20C data are been made freely available from the ECMWF web site (see: https://www.ecmwf.int/en/research/climate-reanalysis/cera-20c) precisely to favour these comparisons.

8. Summary and conclusions

ERA-CLIM2 is a European Union Seventh Framework project started in January 2014, which involves 17 organisations (see Appendix A). It aims to produce coupled reanalyses, i.e. physically consistent data sets describing the evolution of the global atmosphere, ocean, land-surface, cryosphere and the carbon cycle. The main contributions of the ERA-CLIM2 project to climate science have been to rescue and re-process past conventional and satellite data, improve the capacity for producing state-of-the-art climate reanalyses that extend back to the early 20^{th} century, along with uncertainties, and generate unique and extremely valuable data sets. One of the main deliverables of ERA-CLIM2 has been CERA-20C, the first European coupled reanalysis of the 20th century. CERA-20C is now being used to generate a land (water and energy) and a carbon (land and ocean) component (CERA-20C/Land and CERA-20C/Carbon). At the time of writing, the production of the CERA-SAT reanalysis has started: the aim is to complete the period 2008-to-date by the end of the project (December 2017). Thanks to ERA-CLIM2, many older data have been rescued and post-processed, and are delivered to relevant database providers so that they can be used in future reanalysis. Furthermore, new assimilation methods (e.g. use of a stronger coupling method between the ocean and the atmosphere and the direct assimilation of SST data) developed and tested within the project are planned to be integrated and used in future reanalysis productions.

Understanding climate change is highly dependent on the availability of 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 activities in data-assimilation, of data rescue and observation re-processing, production and diagnosis and evaluation is required to improve future reanalyses, so that they can provide a better, closer-to-reality image of the time evolution of the Earth system.

Here are two examples of why we need a continuous stream of investments in the two areas mentioned above:

- Within ERA-CLIM2 millions of new observation records, mostly made before the International Geophysical Year 1958, have been discovered, rescued, digitized and prepared to be inserted in appropriate data sets so that they can be used in future reanalysis production. As this work has been progressing, new data are discovered, but unfortunately due to lack of resources and time (the project will finish at the end of 2017), they will not be rescued, digitized, quality controlled and prepared to be inserted in the data sets.

- As part of the ERA-CLIM2 work, the possibility to assimilate directly sea-surface temperature observations has been explored and tested in prototype systems. Furthermore, the possibility to use ensemble methods to estimate flow-dependent background error statistics within the ocean has been developed and tested. Neither of these advances could be included in the current ERA-CLIM2 reanalyses, since tested software was not ready in time for production.

Preliminary assimilation experiments have shown that the amount and quality of those data justify a full reanalysis, using earlier satellite observations and all conventional (surface and upper-air) data, back to the early 20th century (*Hersbach et al.*, 2017). Such a reanalysis would realize the potential of the data collected and would lead to a much better description of the climate evolution over the last century. However, the only way to be able to continue these essential activities would be to fund them either through a new stream of European projects, or directly as part of the European Union Copernicus Climate Change Service (C3S) activities.

9. Appendix A: the ERA-CLIM2 Consortium

The ERA-CLIM2 Consortium included 17 organisations:

1) European Centre for Medium-Range Weather Forecasts (ECMWF; Europe);
2) Met Office, UK (UKMO; UK);
3) European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT; Europe);
4) University of Bern (Switzerland);
5) University of Vienna (Austria);
6) Instituto Dom Luiz Faculdade de Ciências da Universidade de Lisboa (Portugal) ;
7) Russian Research Institute of Hydrometeorological Information (RIHMI; Russia);
8) Mercator Ocean Societé Civile (MERCATOR; France);
9) Météo-France (MF; France);
10) Deutscher Wetterdiens (DWD; Germany);
11) Centre European de Recherche et de Formation Avancée en Calcul Scientifique (CERFACS; France) ;
12) Centro Euro-Mediterraneo Sui Cambiamenti Climatici (CMCC; Italy);
13) Ilmatieteen Laitos (FMI; Finland);
14) Universidad del Pacifico (Chile);
15) The University of Reading (UK);
16) Institut National de Recherche en Informatique et en Automatique (INRIA; France) ;
17) Université de Versailles Saint-Quentin-en-Yvelines (France).

10. References

- Brönnimann, S., Allan, R., Buizza, R., Bulygina, O., Dahlgren, P., Dee, D., Gomes, P., Jourdain, S., Haimberger, L., Hersbach, H., Poli, P., Pulliainen, J., Rayner, N., Schulz, J., Sterin, A., Stickler, A., Valente, MM. A., Ventura, M. C., Wilkinson, C., 2017: Observations for reanalysis. *Bull. Amer. Met. Soc.*, in preparation.

- Brugnara, Y., Brönnimann, S., Zamuriano, M., Schild, J., Roher, C., Segesser, D.M., 2017: Reanalysis sheds light on 1916 avalanche disaster. *ECMWF Newsletter*, **151**, 28-34 (doi:10.21957/h9b197).

- Compo, G.P., Whitaker, J.S., Sardeshmukh, P.D., Matsui, N., Allan, R.J., Yin, X., Gleason Jr, B.E., Vose, R.S., Rutledge, G., Bessemoulin P., Brönnimann, S., Brunet, M., Crouthamel, R.I., Grant, A.N., Groisman, P.Y., Jones, P.D., Kruk, M.C., Kruger, A.C., Marshall, G.J., Maugeri, M., Mok, H.Y., Nordli, Ø., Ross, T.F., Trigo, R.M., Wang, X.L., Woodruff, S.D., and Worley, S.J., 2011: The Twentieth Century Reanalysis Project. *Q. J. R. Meteorol. Soc.*, **137**, 1-28.

- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F., 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.*, **137**, 553–597 (doi:10.1002/qj.828).

- Dee, D. P., Balmaseda, M. A., Balsamo, G., Engelen, R., Simmons, A. J. and J.-N. Thépaut, 2014: Toward a consistent reanalysis of the climate system. *Bull. Amer. Meteor. Soc.*, **95**, 1235-1248 (doi:10.1175/BAMS-D-13-00043.1).

- Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., & Wimmer, W. (2012). The operational sea surface temperature and sea ice analysis (OSTIA) system. *Remote Sensing of Environment*, **116**, 140-158.

- Feng, X. and K. Haines, 2017. Uncertainties of sea surface and air temperature in the CERA-20C coupled reanalysis ensemble. *Q. J. R. Meteorol. Soc.*, submitted.

- Ferguson, C. R., Villarini, G., 2012: Detecting inhomogeneities in the Twentieth Century Reanalysis over the central United States. *J. Geophys. Res.*, **117**, 2156-2202.

- Hersbach, H., S. Brönnimann, L. Haimberger, M. Mayer, L. Villiger, J. Comeaux, A. Simmons, D. Dee, S. Jourdain, C. Peubey, P. Poli, N. Rayner, A. Sterin, A. Stickler, M. A. Valente, S. Worley, 2017: The potential value of early (1939-1967) upper-air data in atmospheric climate reanalysis. *Quart. J. Roy. Meteorol. Soc.*, **143**, 1185-1196 (doi:10.1002/qj.3020).

- Kalnay E, Kanamitsu M, Kirtler R, Collins W, Deaven D, Gandin L, Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M.,Ebisuzaki, W., Higgins W., Janowiak J., Mo, K.C., Ropelewski C., Wang J., Leetma, A., Reynolds, R., Jenne, R., Joseph, D., 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteorol. Soc.*, **77**, 437 – 471.

- Laloyaux, P., Balmaseda, M., Dee, D., Mogensen, K. and Janssen, P, 2016: A coupled data assimilation system for climate reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 142, 65-78.

- Laloyaux, P., de Boisseson, E. and Dahlgren, P, 2017: CERA-20C: An Earth system approach to climate reanalysis. *ECMWF Newsletter*, **150**, 25-30).

- Mayer, M., L. Haimberger, J. M. Edwards, P Hyder, 2017: Towards consistent diagnostics of the coupled atmosphere and ocean energy budgets. *J. Climate*, accepted for publication.

- Onogi K., Tsutsui J., Koide H., Sakamoto, M., Kobayashi, S., Hatsushika, H., Matsumoto, T., Yamazaki, N., Kamahori, H., Takahashi, K., Kadokura, S., Wada, K., Kato, K., Oyama, R., Ose, T., Mannoji, N. and Taira R., 2007: The JRA-25 Reanalysis. *J. Meteor. Soc. Japan*, **85**, 369 – 432.

- Pellerej, R., A. Vidard, and F. Lemarie. Toward variational data assimilation for coupled models: first experiments on a diffusion problem. *Arima, CARI 2016* (preprint), 2016. [Available online at https://hal.archives-ouvertes.fr/hal-01337743].

- Peylin, P., Bacour, C., MacBean, N., Leonard, S., Rayner, P. J., Kuppel, S., Koffi, E. N., Kane, A., Maignan, F., Chevallier, F., Ciais, P., and Prunet, P., 2016: A new stepwise carbon cycle data assimilation system using multiple data streams to constrain the simulated land surface carbon cycle, *Geosci. Model Dev.*, **9**, 3321-3346 (doi: 10.5194/gmd-9-3321-2016).

- Pietschnig, M., M. Mayer, T. Tsubouchi, A. Storto, L. Haimberger, 2017: Comparing reanalysis-based volume and temperature transports through Arctic Gateways with mooring-derived estimates. *J. Geophys. Res. Oceans*, under review.

- Ploshay, J. J., Stern, W. F., and Miyakoda, K, 1992: FGGE reanalysis at GFDL. *Mon. Wea. Rev.*, **120**, 2083-2108.

- Poli, P., D. Dee, R. Saunders, V. O. John, P. Rayer, J. Schulz, K. Holmlund, D. Coppens, D. Klaes, J. E. Johnson, A. E. Esfandiari, I. V. Gerasimov, E. B. Zamkoff, A. F. Al-Jazrawi, D. Santek, M. Albani, P. Brunel, K. Fennig, M. Schröder, S. Kobayashi, D. Oertel, W.

Döhler, D. Spänkuch, and S. Bojinski, 2017: Recent Advances in Satellite Data Rescue. *Bull. Amer. Meteorol. Soc.*, **91**, 1471-1484.

- Poli, P., H. Hersbach, D.P. Dee, P. Berrisford, A.J. Simmons, F. Vitart, P. Laloyaux, D.G. Tan, C. Peubey, J. Thépaut, Y. Trémolet, E.V. Hólm, M. Bonavita, L. Isaksen, and M. Fisher (2016) ERA-20C: An Atmospheric Reanalysis of the Twentieth Century. *J. Climate*, **29**, 4083–4097 (DOI:10.1175/JCLI-D-15-0556.1).

- Rienecker, M.M., Suarez, M.J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M.G., Schubert, S.D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu. W., Joiner, J., Koster, R.D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C.R., Reichle, R., Robertson, F.R., Ruddick, A.G., Sienkiewicz, M., and Woollen, J., 2011: MERRA – NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624-3648 (DOI: 10.1175/JCLI-D-11-00015.1).

- Rustemeier, E., Ziese, M., Meyer-Christoffer, A., Schneider, U., Finger, P., Becker, A. (2017): Uncertainty assessment of the ERA-20C reanalysis based on the monthly in-situ precipitation analyses of the Global Precipitation Climatology Centre. In prep for submission to *J. Hydrometeor*.

- Saha S., Moorthi S., Pan H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Hou, Y.-T., Chuang, H.-Y., Juang, H.-M.H., Sela, J., Iredell, M., Treadon, R., Kleist, D., van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., van den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen Y., Han, Y., Cucurull, L., Reynolds, R.W., Rutledge, G., Goldberg, M., 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteorol. Soc*., **91**, 1015-1057.

- Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Ziese, M. (2015): GPCC Full Data Reanalysis Version 7.0 at 1.0°: Monthly land-surface precipitation from rain-gauges built on GTS-based and historic data. 2015. (DOI: 10.5676/DWD_GPCC/FD_M_V7_100).

- Schubert, S.D., Rood, R., and Pfaendtner, J., 1993: An assimilated dataset for Earth science applications. *Bull. Am. Meteorol. Soc*., **74**, 2331-2342.

- Storto, A., M. J. Martin, B. Deremble, S. Masina, 2017: Strongly coupled data assimilation experiments with linearized ocean-atmosphere balance relationships. *Mon. Wea. Rev*., submitted.

- Titchner, H. A., & Rayner, N. A. (2014). The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations. *Journal of Geophysical Research: Atmospheres*, **119**(6), 2864-2889.

- Uppala, S. M., P. W. Kallberg, A. J. Simmons, U. Andrae, V. Da Costa Bechtold, M. Fiorino, J. K. Gibson, J. Haseler, A. Hernandez, G. A. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R. P. Allan, E. Andersson, K. Arpe, M. A. Balmaseda, A. C. M. Beljaars, L. Van De Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Hólm, B. J. Hoskins, L. Isaksen, P. A. E. M. Janssen, R. Jenne, A. P. Mcnally, J.-F. Mahfouf, J.-J. Morcrette, N. A. Rayner, R. W. Saunders, P. Simon, A. Sterl, K. E. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo, J. Woollen, (2005): The ERA-40 re-analysis. *Q.J.R. Meteorol. Soc.*, **131**, 2961–3012 (doi:10.1256/qj.04.176).

- Weaver, A.T., Tshimanga J., Piacentini A., 2016. Correlation operators based on an implicitly formulated diffusion equation solved with the Chebyshev iteration. *Q. J. Roy. Meteorol. Soc.*, **142**: 455-471.

11. Figures captions

Figure 1. High-pass filtered sea surface temperature (colour) and wind stress (contour) for ERA-20C (left) and CERA-20C (right) over the period 01/04/73 to 27/03/74. CERA-20C represents the Tropical Instability Waves thanks to the ocean dynamics and the atmosphere is responding accordingly with the surface wind stress sensitive to the ocean TIWs. In ERA-20C, there are no TIWs and wind stress signals.

Figure 2. (a-c) SST-precipitation correlations for their monthly fluctuations in observational data (HadISST2 and GPCP), ERA-20C and CERA-20C (control run), over 1979-2010. Blank areas are where the correlations do not pass the significance test at the 90% confidence level. Note the agreement between observations and CERA-20C in the heavily precipitating regions.

Figure 3. Examples of sea surface temperature observations (left) and temperature profile observations (right) in °C from Jan 1953 (top panels) and Jan 2010 (middle panels). The bottom plot shows the percentage change in error in SST from assimilating the Jan 1953 data using the new version of NEMOVAR.

Figure 4. Results from the dynamical downscaling of ERA-20C: (a) total precipitation (shading) on 13 December 1916 and mean freezing level (grey contours; in m). (b) change in snow depth between 5 and 13 December 1916. Circles represent snow observations, red crosses show the locations of documented major avalanches on this day. The military front line in 1916 is shown as red dotted line.

Figure 5. Quantile-Quantile Plot of monthly regional mean short term precipitation forecasts from ERA-20C (black) and CERA-20C (grey) against Full Data Monthly V7 (Schneider et al., 2015)GPCC precipitation estimates from rain gauges provided by the Global Precipitation Climatology Centre (GPCC). a) Africa, b) India/Monsoon region,, c) South America,. Time interval considered: 1901-2000 on 1° spatial resolution, monthly temporal resolution.

Figure 6. The two time series show the evolution of yearly global-mean anomalies relative to the period 1961–1990 for the two-metre temperature (top) and the upper-ocean heat content in CERA20C.

12. Figures

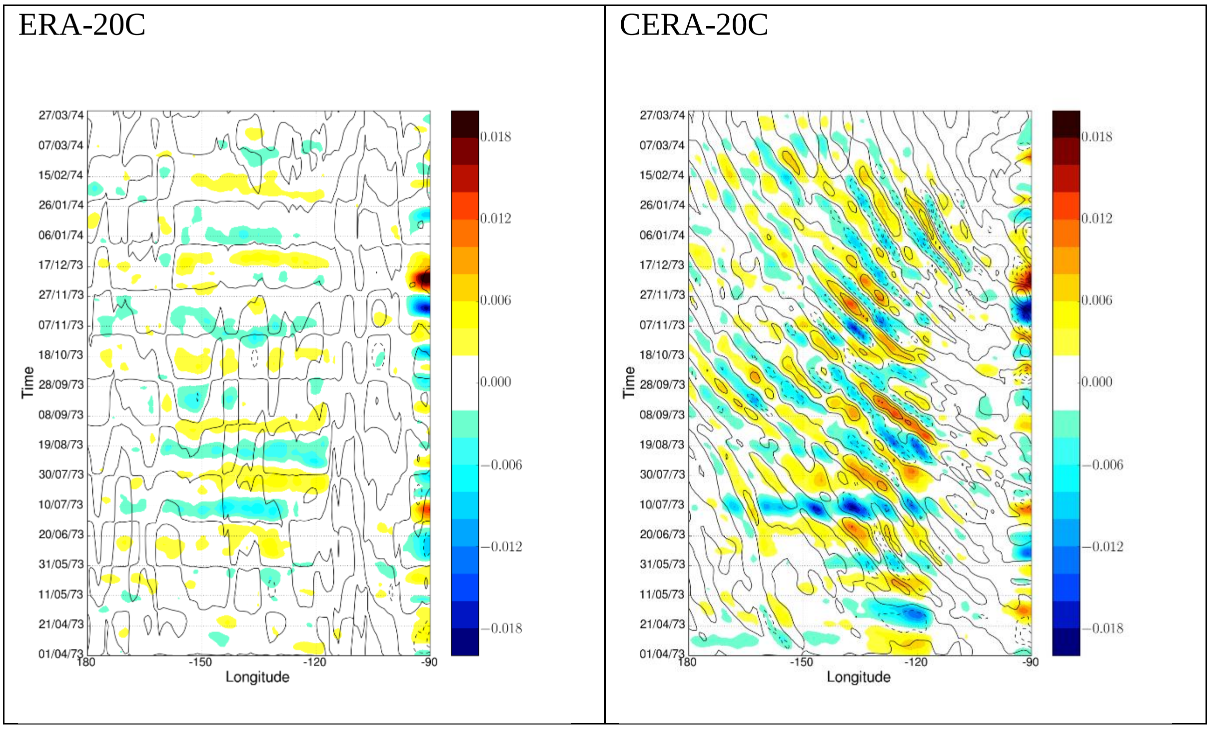

Figure 1. High-pass filtered sea surface temperature (colour) and wind stress (contour) for ERA-20C (left) and CERA-20C (right) over the period 01/04/73 to 27/03/74. CERA-20C represents the Tropical Instability Waves thanks to the ocean dynamics and the atmosphere is responding accordingly with the surface wind stress sensitive to the ocean TIWs. In ERA-20C, there are no TIWs and wind stress signals.

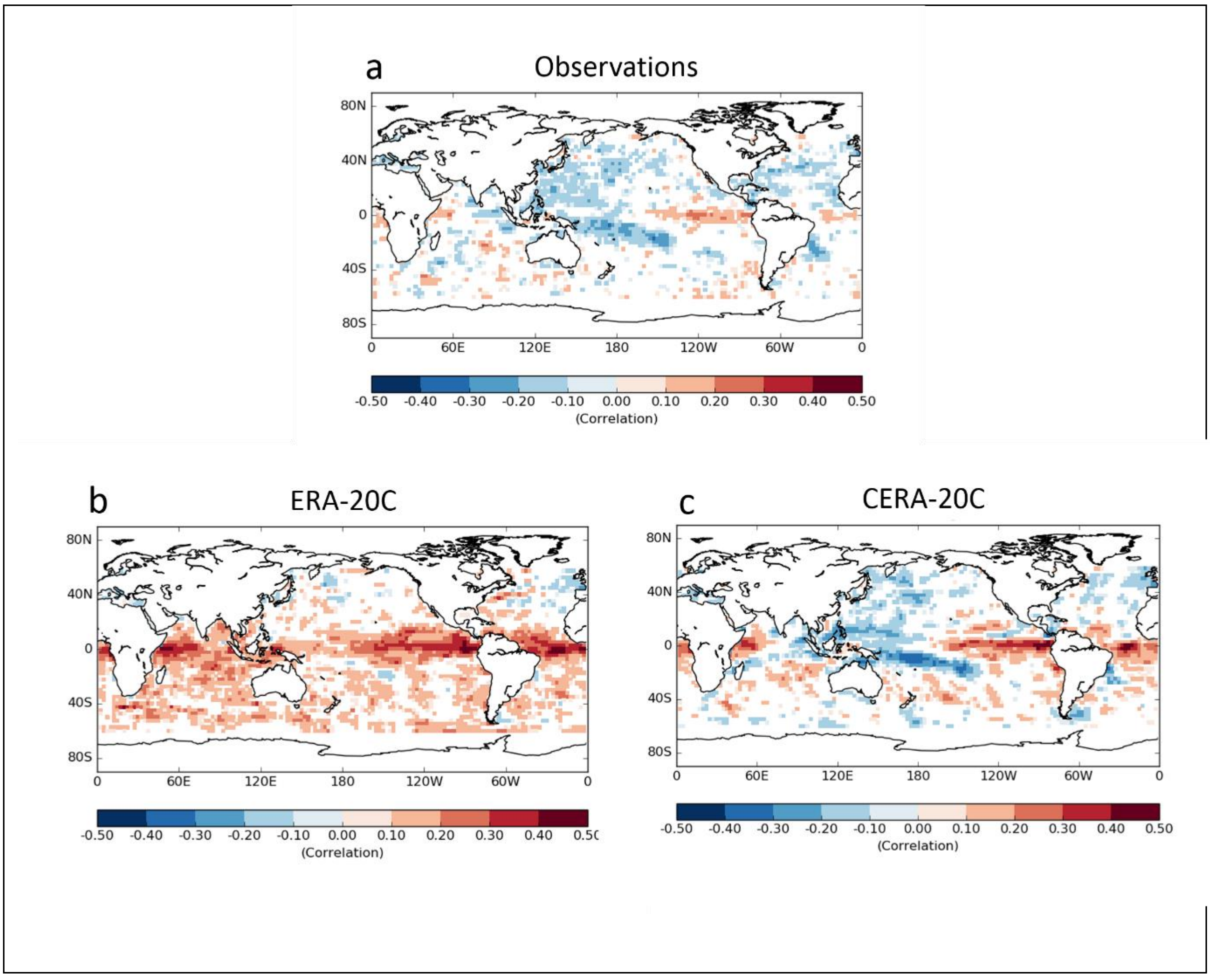

Figure 2. (a-c) SST-precipitation correlations for their monthly fluctuations in observational data (HadISST2 and GPCP), ERA-20C and CERA-20C (control run), over 1979-2010. Blank areas are where the correlations do not pass the significance test at the 90% confidence level. Note the agreement between observations and CERA-20C in the heavily precipitating regions.

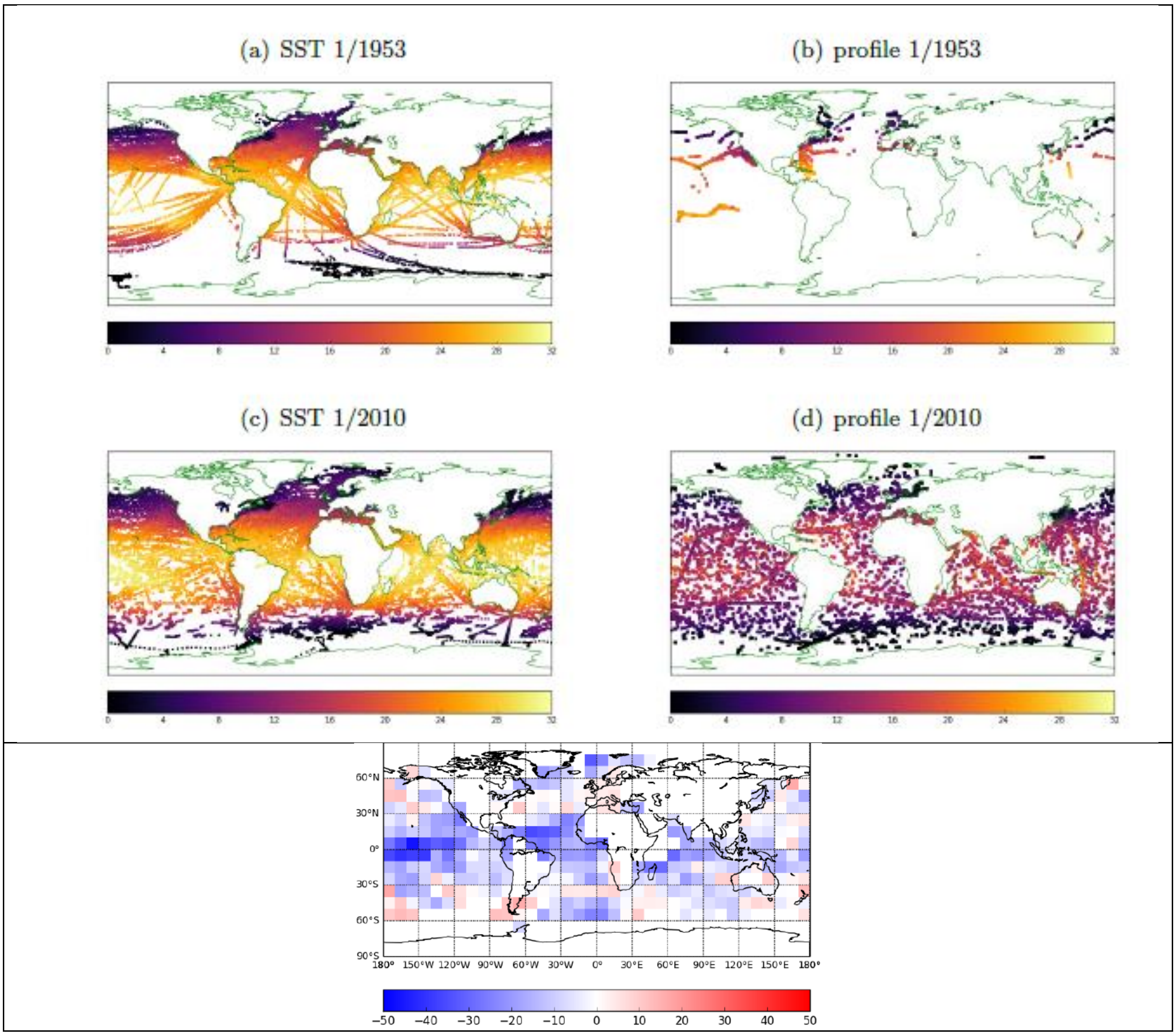

Figure 3. Examples of sea surface temperature observations (left) and temperature profile observations (right) in °C from Jan 1953 (top panels) and Jan 2010 (middle panels). The bottom plot shows the percentage change in error in SST from assimilating the Jan 1953 data using the new version of NEMOVAR.

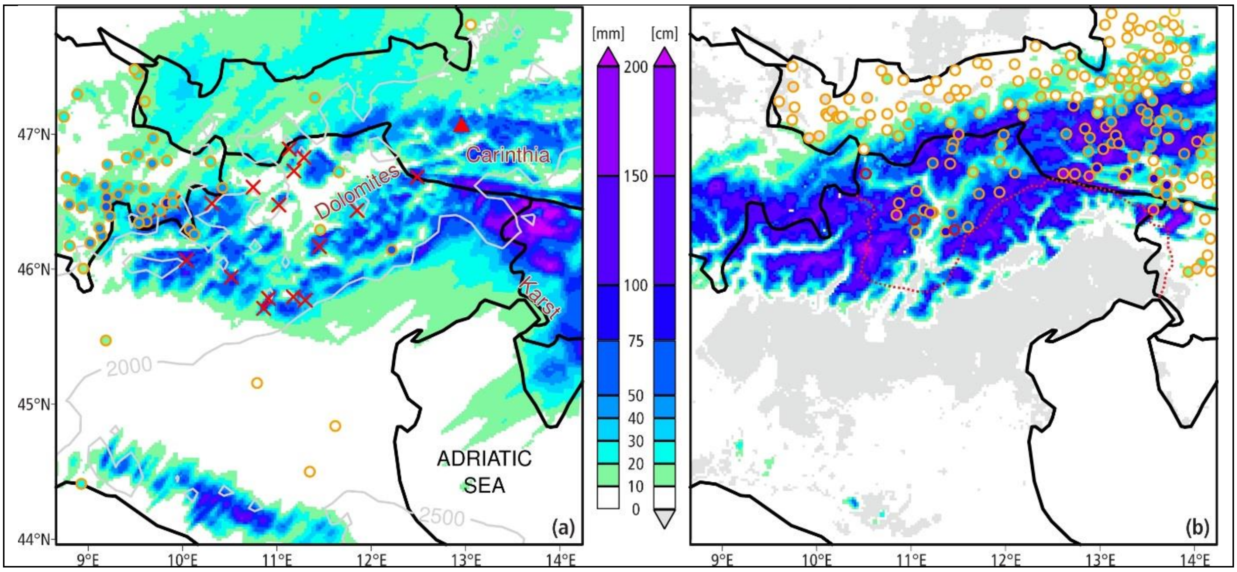

Figure 4. Results from the dynamical downscaling of ERA-20C: (a) total precipitation (shading) on 13 December 1916 and mean freezing level (grey contours; in m). (b) change in snow depth between 5 and 13 December 1916. Circles represent snow observations, red crosses show the locations of documented major avalanches on this day. The military front line in 1916 is shown as red dotted line.

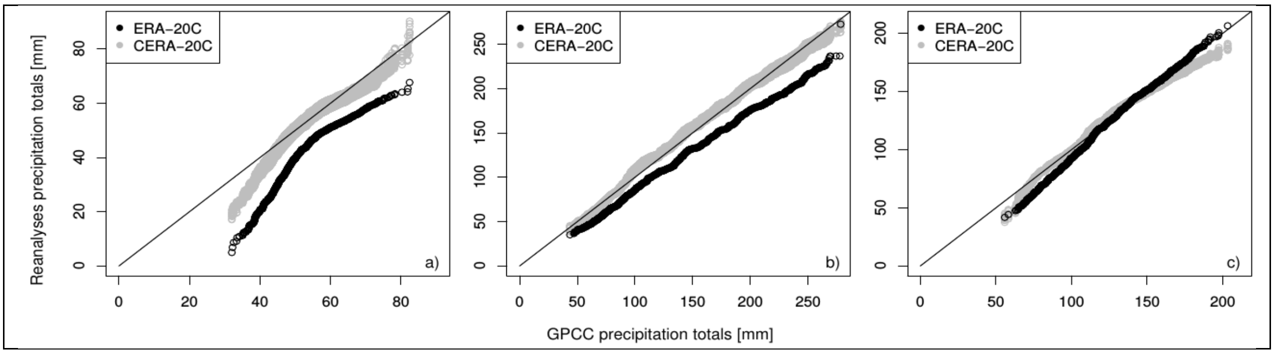

Figure 5. Quantile-Quantile Plot of monthly regional mean short term precipitation forecasts from ERA-20C (black) and CERA-20C (grey) against Full Data Monthly V7 (Schneider et al., 2015)GPCC precipitation estimates from rain gauges provided by the Global Precipitation Climatology Centre (GPCC). a) Africa, b) India/Monsoon region,, c) South America,. Time interval considered: 1901-2000 on 1° spatial resolution, monthly temporal resolution.

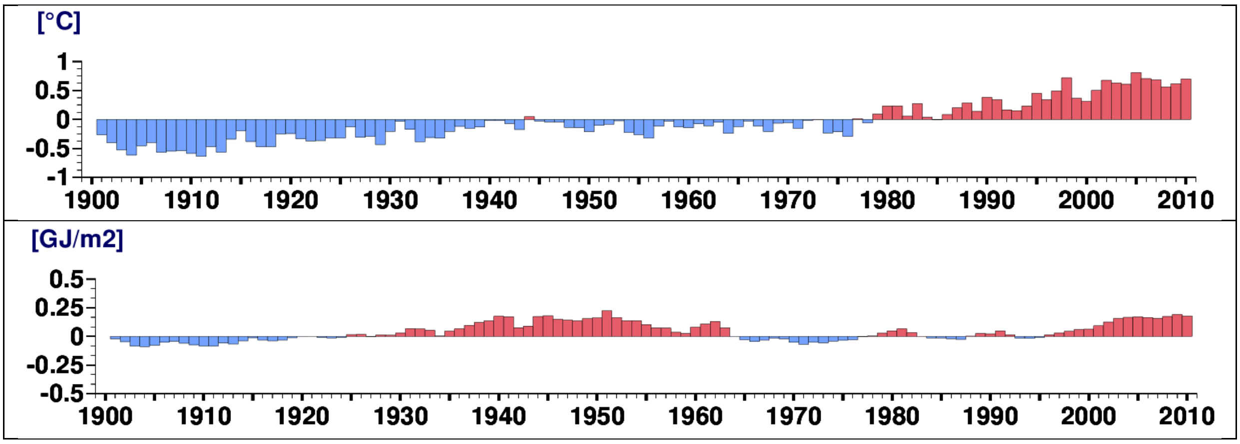

Figure 6. The two time series show the evolution of yearly global-mean anomalies relative to the period 1961–1990 for the two-metre temperature (top) and the upper-ocean heat content in CERA20C.