Final Publishable Summary Report for ERA-CLIM

Dick Dee
ERA-CLIM Coordinator
28 February 2014

On behalf of the ERA-CLIM Consortium:

European Centre for Medium-Range Weather Forecasts (**ECMWF**)

Met Office, United Kingdom (**UKMET**)

Universität Wien, Austria (**UNIVIE**)

Universität Bern, Switzerland (**UBERN**)

All-Russian Research Institute for Hydrological Information, Russia (**RIHMI**)

Fundação da Faculdade de Ciências da Universidade de Lisboa, Portugal (**FFCUL**)

European Organisation for the Exploitation of Meteorological Satellites (**EUMST**)

Météo-France, France (**METFR**)

Universidad del Pacífico, Chile (**UPAC**)

Table of Contents

Executive summary

This report summarizes the scientific outcomes of the collaborative research project entitled "European reanalysis of global climate observations" (ERA-CLIM), carried out within the Framework Programme 7 of the European Union under Grant Agreement No. 265229.

The project set out to develop observational datasets suitable for global climate reanalysis, with a focus on the past 100 years. These include atmospheric, terrestrial and oceanic observations from a variety of sources, global reanalyses of the observations, and various uncertainty indicators needed for climate applications.

In-situ data rescue efforts conducted within the project have led to a large improvement of the digital instrumental record for the early 20th century. The number of pre-1957 weather observations from kites and early radiosondes available for climate reanalysis has more than doubled as a result of these efforts.

ERA-CLIM has provided initial impetus for an important international activity in satellite data rescue by investigating the availability of data records from pre-operational satellites and their potential use in future climate reanalyses. This has led to a first comprehensive survey of early satellite data for reanalysis, including priorities for data rescue and information about next steps.

Several satellite data records important for climate reanalysis have been reprocessed during the project. These activities have helped develop and solidify a comprehensive programme for reprocessing of satellite data holdings at the Organisation for the Exploitation of Meteorological Satellites (EUMST).

The UK Met Office Hadley Centre (UKMET) has developed improved global gridded estimates of the evolution of sea-surface temperature and sea-ice concentration during the 20th century, together with meaningful information about uncertainties in these estimates (HadISST.2). In addition, important new data collections containing high-quality sub-surface and surface ocean observations have been developed (EN4, HadIOD).

Using many of the observational data sets developed in ERA-CLIM, a set of climate model integrations (ERA-20CM) and several pilot reanalyses for the 20th century (ERA-20C, ERA-20CL) have been produced at ECMWF. These datasets, with a combined volume of hundreds of terabytes, will be made freely available via the internet during 2014. A first test reanalysis for the period 1939-1949 has been conducted using newly recovered upper-air observations from the project.

Several project activities have contributed to the estimation and reduction of uncertainties, in climate observations as well as gridded products derived from them. A permanent Observation Feedback Archive facility has been developed by

the project to allow users easy access to the input observations used in climate reanalyses, and to information about uncertainties associated with the observations.

The impact on science of the global climate data produced in ERA-CLIM will be substantial. The 20th-century reanalysis data sets (ERA-20CM, ERA-20C, ERA-20CL) will be made available on the internet in stages during 2014.

A follow-up project (ERA-CLIM2, Grant Agreement No. 607029) will continue to strengthen data rescue, data reprocessing, and climate reanalysis. ERA-CLIM2 will introduce an added focus on development of consistent reanalyses of the coupled Earth system, including atmosphere, ocean, land and carbon. The combined results of the two successive ERA-CLIM projects will provide key resources and capabilities needed to implement Copernicus Climate Change Services.

Summary description of project context and objectives

European Reanalysis of Global Climate Observations (ERA-CLIM)

ERA-CLIM is a collaborative research project involving 9 partners, funded by the European Commission for a three-year period beginning January 2011. The overall goal of the project is to prepare input data and assimilation systems for a future global atmospheric reanalysis of the 20th century. This involves recovery and digitization of early meteorological observations, reprocessing and recalibration of radiance observations from satellites, and preparation of climate-quality atmospheric forcing data and boundary conditions. These input data sets are used in several pilot reanalyses, including an exploratory climate reanalysis of the 20th century based on surface observations only; a corresponding high-resolution reanalysis of land-surface parameters, and a new atmospheric reanalysis of the satellite era from 1979 to present. Access to the ERA-CLIM pilot reanalyses, and to all observations used to produce them, will be provided via the internet.

Key objectives

- Improve the available observational record for the early 20th century
- Prepare data sets and assimilation tools needed for global reanalysis
- Provide information about data quality by means of pilot reanalyses
- Develop an Observation Feedback Archive facility for users
- Assess and reduce uncertainties in reanalysis data

Policy relevance

The project addresses policy objectives underlying the Commission’s Copernicus programme, by developing core information products for future Copernicus Climate Change Services. These include global atmospheric reanalyses to support climate monitoring, to serve as boundary conditions for high-resolution regional reanalyses and other specialized products, and as benchmark datasets for calibration and validation of climate models.

Description of the main S&T results/foregrounds

In-situ data recovery: Imaging, digitization and quality control

The project delivered substantial numbers of pre-1957 upper-air and surface weather observations that have not before been available in digital form suitable for climate analysis. A total of more than 2.5×10^6 days of station records were preserved on 450,000 images. Approximately half were converted into digital records to prepare new input datasets for reanalysis. The data are distributed over large parts of the globe, especially poorly observed and climate-sensitive regions such as the Tropics, polar regions and oceans. Early 20th-century upper-air data from Europe and the US were included as well. Total numbers of digitized/inventoried records, i.e. time series of meteorological data at fixed stations or from moving observational platforms, are 80/214 (surface), 735/1,783 (upper-air), and 61/101 (moving upper-air, i.e. data from ships etc.).

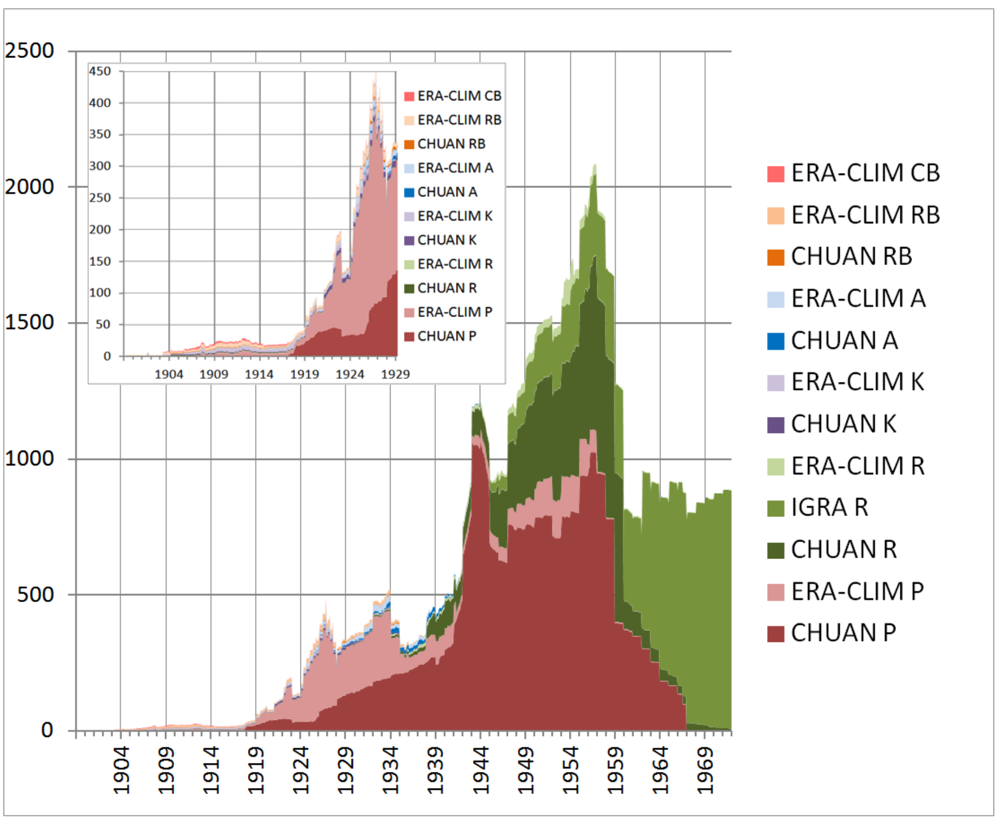


Figure 1: *Number of inventoried ERA-CLIM upper-air records, available CHUAN upper-air records (without merged IGRA records), and IGRA radiosonde records as function of time. Multiple measurement platforms are multi-counted. CB: captive balloon, RB: registering balloon, A: aircraft, K: kite, R: radiosonde, P: pilot balloon.*

ERA-CLIM had access to archives that were previously inaccessible to the international scientific community. A comparison with the existing CHUAN (Stickler et al. 2010) and IGRA (Durre et al. 2006) upper-air data sets are presented in Figs. 1 and 2. The largest part of new station days consists of regular surface data and pilot balloon wind observations. After 1938, radiosonde observations also significantly contribute. Their largest contribution occurs during the period 1947-56 and during the early radiosonde era before 1938. The quantity of moving platform upper-air data may be more modest, although, often originating from oceanic regions that are not covered by other data sources.

Until the late 1930s, most daytime (06-18 UTC) ascents did not reach altitudes higher than 5 km above sea level. In the mid to late 1940s, already 15-25% of these ascents reached altitudes of at least 8 km. During the same period, the contribution of ascents reaching more than 13 km became significant. From about 1950 on, more than 15% of the day-time sondes ascended higher than 20 km, and after 1955, a small number of records includes ascents higher than 30 km.

The quality control (QC) consisted of flagging of suspicious values during the digitization process, checking these values afterwards with the help of the digital images, and range checks, among other methods. Departures from the ERA-CLIM E20C surface-only reanalysis (ERA-20C; Poli et al. 2013), as described below in more detail, were used as well for this purpose. The QC applied to the complete upper-air data is described in more detail in Stickler et al. (2014b).

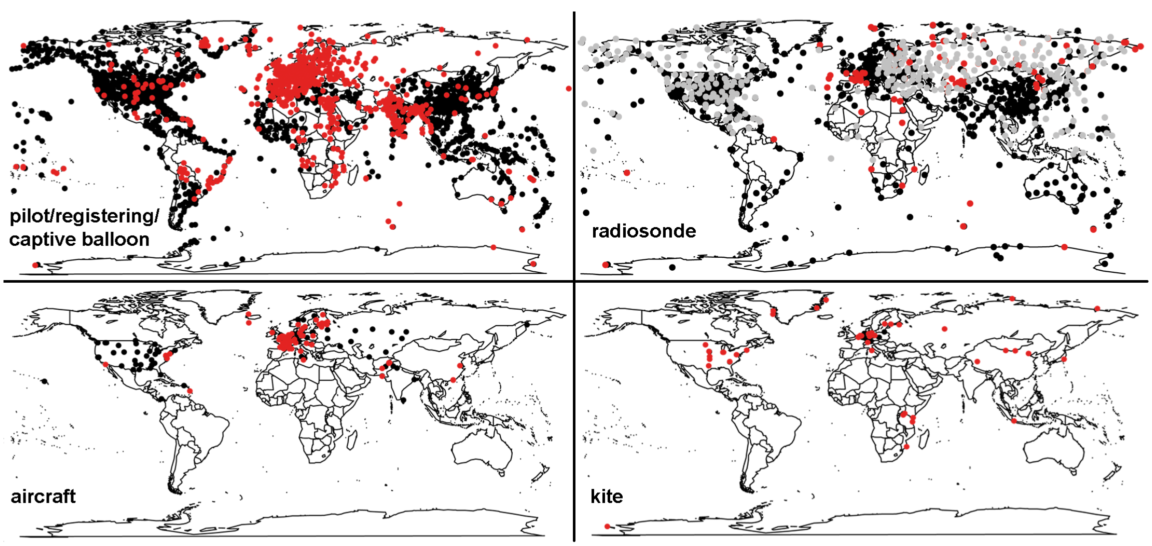


Figure 2: *Global distributions of all inventoried ERA-CLIM upper-air stations (red), and additional available CHUAN upper-air stations (black). Different measurement platforms are shown separately in four panels. The radiosonde panel displays the pre-1958 IGRA stations (grey), together with the additional CHUAN and ERA-CLIM stations.*

The collaboration in ERA-CLIM allowed for an intense exchange and knowledge transfer between the partner institutions with respect to data rescue techniques such as imaging, job handling (for which a web tool was developed), and experiences with optical character recognition software and QC tests. A large, albeit still incomplete, catalogue of available historical data sources was developed and made available in the form of a web-based meta-database, which can serve as a starting point for further data rescue projects.

Stickler et al. (2014a) give an overview of the ERA-CLIM data rescue activities, the recovered data and the quality control procedures used. The upper-air data are further described in Stickler et al. 2014b and will be included in the next version of the CHUAN data set (Stickler et al. 2010).

Satellite data rescue for reanalysis

Three ERA-CLIM partners (UK Met Office, ECMWF, and EUMETSAT) have investigated the availability of historic satellite data records and their potential for use in future atmospheric reanalyses. This led to a first comprehensive survey of early satellite data for reanalysis, including priorities for data rescue and information about next steps. A selection of these early sensors is shown in Table 1.

Drawing from this initial survey, and including also modern satellite instrumental records, a proposal for the use of satellite data in future reanalyses was drawn up. Covering as many as 190 satellite datasets, it was structured as follows: 23 legacy datasets to be assimilated, 38 legacy datasets to be assimilated with improved observation operators, 88 reprocessed datasets to be assimilated, 8 reprocessed datasets to be monitored, 6 early satellite datasets to be tested for assimilation before seeking reprocessing/recalibration, and 27 early satellite datasets to be tested for monitoring (assimilated passively) before seeking reprocessing/recalibration.

The ERA-CLIM survey of early satellite records for reanalysis has raised international awareness of the importance of satellite data rescue and has prompted new projects. For example, the SMMR dataset identified as highly important was the subject of a search by EUMETSAT and its network of satellite application facilities, and resulted in a positive outcome, with a copy of the raw radiance dataset being located. In addition, scientific developments were initiated to tackle the problem of radiative transfer for the PMR instrument, the spectral response function of Nimbus-6 HIRS was estimated from paper records, and the reprocessing of AVHRR on polar orbiters since 1982 has been completed.

Instru-ment	Characteristics	Immediate concern	Recom-mendation
NEMS	Microwave spectrometer, with two water vapour channels near 22 GHz (5 mm) and three channels near 59 GHz (10 mm), spatial resolution 180 km at nadir	Nadir-viewing only, data on microfiche	Reject for now
SCAMS	Microwave spectrometer, with one water vapour channel near 22 GHz (5 mm), three channels near 59 GHz (10 mm), one window channel, spatial resolution 150 km at nadir	Data recovery in process by NSSDC.	Consider for assimilation
SSM/T	Microwave temperature sounders precursors to AMSU-A and AMSU-B but with bigger fields-of-view. Met Office preparing a homogenized data for ERA-CLIM.	RT forward model needed for SSM/T	Assimilate
SMMR	Microwave radiometer, ten channels: dual-polarization measurements at 6.63, 10.69, 18.0, 21.0, and 37.0 GHz, spatial resolution 150 km at nadir	Raw radiance data not found	Keep looking for data
SSH	Discrete filter radiometer, six channels in the 15 micron CO_2 band, one window channel, eight water vapour channel in the 22–30 micron band, one channel in the 10 micron ozone band	Data lost forever?	Keep looking for data
HIRS on Nimbus-6	Discrete filter radiometer, seven channels in the 15 micron CO_2 band, two window channels, two water vapour channels, five channels in the 4.3 micron band, spatial resolution 25 km at nadir	Data recovery in process by NSSDC. Digital version of the SRF not found.	Assimilate
SCR	Radiometer observing through a pressurized optical cell, six channels in the 15 micron CO_2 band, spatial resolution 112–160 km at nadir (Nimbus-5: eight channels in the 15 micron CO_2 band, three window channels, one water vapour channel at 18.6 microns, spatial resolution 30 km at nadir)	RT coefficients challenging	Validate
PMR	Radiometer observing through a pressurized optical cell	RT coefficients challenging	Assimilate
HRIR	Visible and infrared imager, 8 km spatial resolution at nadir, 3.5–4 micron channel (and also 0.7–1.3 for Nimbus-3)	Digital version of SRF not found	Validate
MRIR	Infrared imager, five channels including a water vapour channel in the 6.7 micron band	Digital version of SRF not found	Validate
THIR	Infrared imager, one window channel and one water vapour channel in the 6.7 micron band	Only JPEG images available, raw radiance data lost forever?	Keep looking for data
IRIS	Michelson interferometer, covering 5–20 microns with 5 cm^{-1} normalized apodized spectral resolution (Nimbus-4: 6.25–25 microns, 2.8 cm^{-1} resolution), nadir spatial resolution 144 km	Short time period, calibration biases	Validate
SIRS	Grating spectrometer, covering 11–15 microns (Nimbus-4: 11–36 microns), nadir spatial resolution 220 km	Narrow swath (up to 12 degrees only from nadir)	Consider for assimilation
AVHRR	Imager on polar orbiters, atmospheric motion vector (wind) retrievals at the poles. EUMETSAT and CIMSS working on reprocessing.	Reprocessing not complete yet	Assimilate
SeaSat	First scatterometer ever. Suspicious end-of-life.	Very short dataset (97 days)	Validate
NSCAT	Scatterometer from U.S.	Short dataset (9 months)	Assimilate

Table 1: *Selection of early satellite instruments with significant potential impact on the reanalysed record, along with recommendations for future use (excerpt from a poster by Poli, Saunders, John, and Kelly, presented at the International TOVS Study Conference 2012).*

Consistent reprocessing of satellite data records

As was the case with the in-situ data recovery work performed in ERA-CLIM, the project has kick-started various European activities on reprocessing European satellite data records. In ERA-CLIM, EUMETSAT embarked on a full-scale reprocessing of numerous of its missions. The following important datasets were delivered: atmospheric motion vector winds, clear-sky radiances, and all-sky radiances from the European geostationary satellites Meteosat-8 and Meteosat-9 for years 2004 to 2008 (a total of 215 Gb), as well as atmospheric motion vector and ocean backscatter from the European polar-orbiting satellite Metop-A (covering years 2007 to 2013).

These data are now available from the EUMETSAT central archiving facility and have also been transferred to ECMWF for use in future reanalyses. The planning and implementation of the reprocessing systems and methodologies at EUMETSAT have laid the technical foundation for sustaining these activities in the longer term. ERA-CLIM involvement has helped to raise the profile of reprocessing activities within EUMETSAT. In addition, the implementation of satellite data reprocessing systems in a sustainable infrastructure has helped build the European capacity for meeting its international obligations with regard to provision of fundamental climate data records.

Several adjustments have had to be made during the project to account for insufficient computing resources and various difficulties associated with installation of new hardware in the EUMETSAT operational environment. Consequently, the planned reprocessing of satellite data records from Global Positioning System radio occultation and of a satellite data record of ozone from Metop-A has been postponed. In compensation the reprocessing of two other Metop-A satellite data records mentioned above were extended in time, adding the most recent years onto the original plan. In addition, reprocessing of all-sky radiances was conducted even though this was not envisioned at the time the project was committed.

(a) (b)

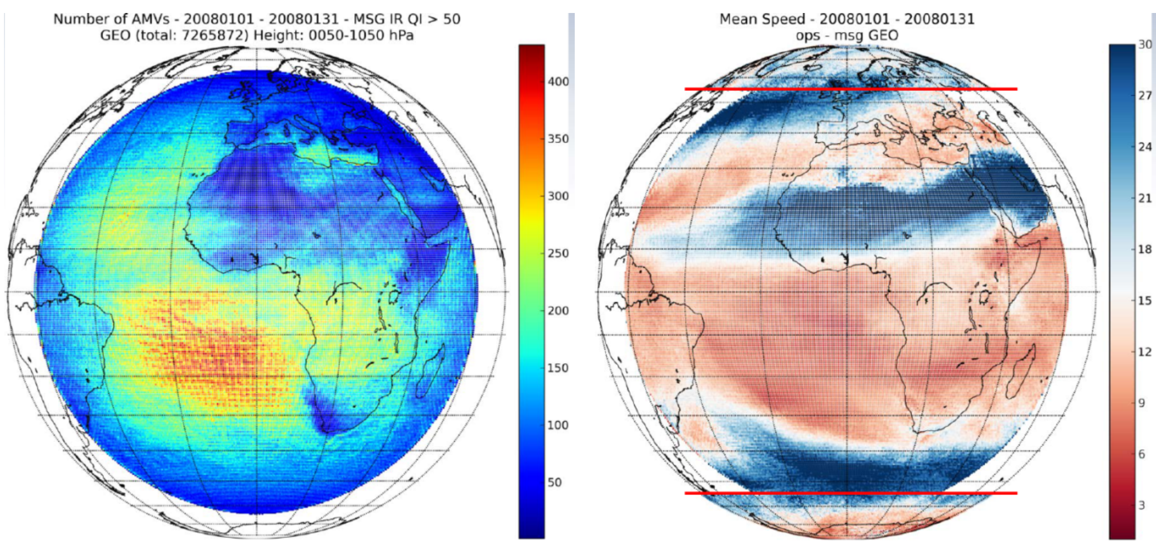


Figure 3: *(a) Geographical coverage in atmospheric motion vector winds achieved by a single geostationary satellite for January 2008. The color shading shows the data yield over the month. (b) Increased geographical coverage with the addition of the AVHRR from Metop-A, covering the regions located poleward of the red horizontal lines (the color shading shows mean observed wind speed -- note the absence of discontinuity across the red lines)*

Global estimates of sea-surface conditions for the 20th century

The UK Met Office Hadley Centre produced two successive versions of a new set of global gridded estimates of sea-surface temperature and sea-ice concentration.

The HadISST.2.0.0.0 product. This preliminary HadISST.2 product comprises two datasets, both provided as globally complete daily sea-surface temperature (SST) and sea ice concentration (SIC) fields at 0.25°x0.25° resolution. One set for the period 1899-2010 is based on 1°x1° monthly fields of sea-surface temperature, and one for the period 1961-2010 is based on a 1°x1° five-day (pentad) analysis. The observations used for the gridded SST estimates are: in-situ observations from ICOADS and from the Met Office observational data base; SST retrievals from AVHRR Pathfinder v5 data (1985-2007); and SST retrievals from the ATSR2 and AATSR (1995-2011) METEO products. In-situ data are an update of Kennedy et al. (2011) with bias adjustments to account for changes in measurement method. For HadISST.2, the data were quality controlled and gridded as anomalies relative to a 0.25° daily climatology. A realization of the Kennedy et al. (2011) bias adjustments that was consistent with the ATSR series was used. The AVHRR data have been adjusted for biases due to aerosol contamination and diurnal drift, estimated by comparison with coincident ATSR and in situ observations and measurements of aerosol optical depth.

The gridded SST estimates were obtained using a 2-step interpolation procedure. In the first step, an iterative process (Ilin and Kaplan 2009) was used to produce a large-scale interpolation using 71 empirical orthogonal functions. The resulting principal component series was smoothed using a Kalman filter to ensure month-to-month (or pentad-to-pentad) continuity and a single sample was drawn from the posterior distribution. The residuals from the large-scale reconstruction were then analysed using a non-stationary local covariance function (Karspeck et al. 2012). The resulting fields were then blended with samples drawn from the local covariance function according to the uncertainty in the local analysis. This procedure produces fields that have consistent variability throughout the entire record.

Data sources used for the estimation of SIC are as described in Rayner et al. (2003). An additional adjustment has been applied to Arctic sea-ice concentrations to remove a bias in the HadISST1.1 SICs after 1995.

The analysed SST fields were interpolated to 0.25° daily resolution and then blended with SSTs estimated from the SICs. The analysed SST fields were trimmed back and the gap between the analysed SST field and the marginal ice zone SSTs was filled such that the Laplacian of the SST field matched the Laplacian of the SST climatology.

The HadISST.2.1.0.0 product. HadISST.2.1.0.0 differs from HadISST.2.0.0.0 in a number of particulars. Estimates of sea-ice concentration were improved by using satellite input that has been homogenised using information from modern sea ice charts. Early charts showing only sea-ice extent information were reprocessed to provide a more consistent representation of ice concentration within the ice pack. The net effect of these changes is an increase in amount of ice in HadISST.2.1.0.0 over previous estimates.

SST estimates were improved by making use of the ATSR Reprocessing for Climate (ARC) data set, which is based on a consistent reprocessing of SST retrievals from the Along Track Scanning Radiometer (ATSR) instruments. The new ARC data set has excellent accuracy and stability. The ATSR2 and AATSR ARC data were used to improve the AVHRR aerosol and time-of-day bias adjustments. They were also used to refine the estimates of in-situ SST biases used in creating the gridded SST product. This was achieved by creating a large ensemble of in-situ data sets and selecting those members that matched most closely with the ARC series.

Finally, interpolation from monthly (or 5-day) resolution to daily resolution was done using cubic splines, resulting in a more consistent variance throughout the month than the linear interpolation used in HadISST.2.0.0.0.

Three papers (Titchner et al. 2014, Kennedy et al. 2014), Rayner et al. 2014) describe the creation of the HadISST.2.1.0.0 and HadISST.2.0.0.0 data sets in full detail.

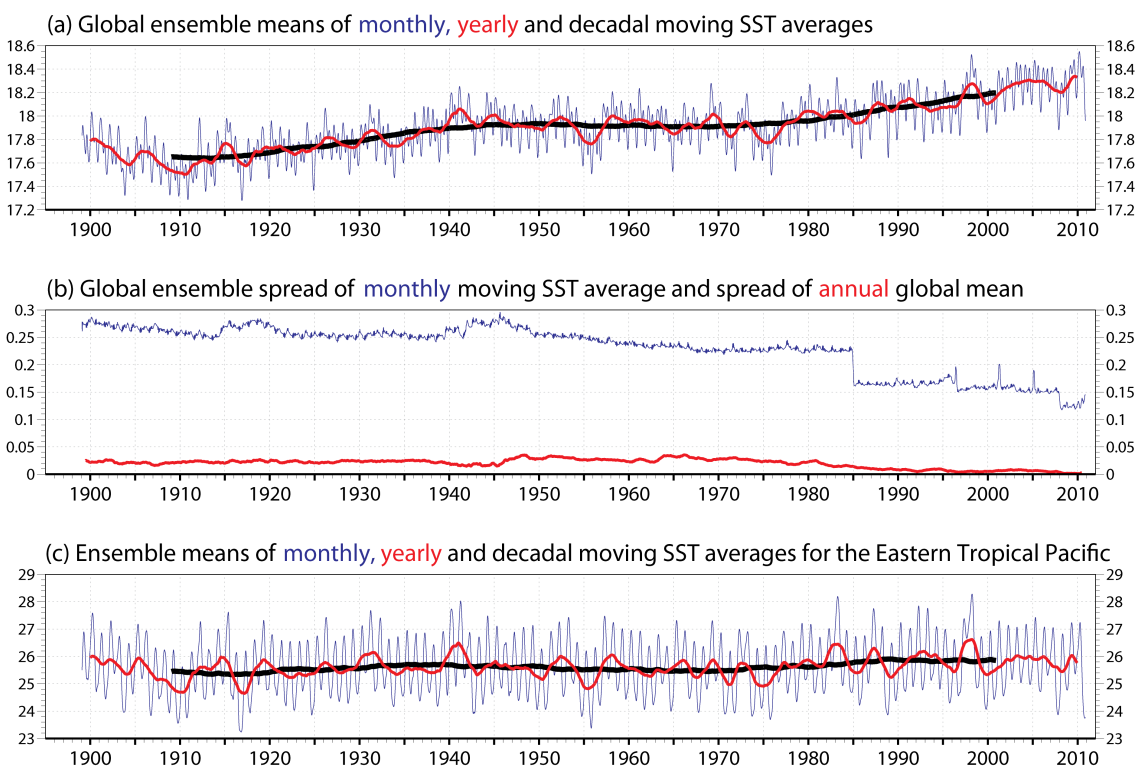


***Figure 4**: Time-series of HadISST.2.0.0.0 data for (a) global average of the SST ensemble mean (blue: monthly, red: yearly, black: decadal moving average) (b) Global average of the SST ensemble spread (blue: monthly moving average) and spread of the annual global mean SST (red) (c) as (a) but limited to the Eastern Tropical Pacific.*

Climate model simulations for the 20th century

Model simulations of the past, using atmospheric forcing and boundary conditions estimated from observations, provide an important tool for understanding and estimating climate change. Depending on the quality of the model input, in particular regarding SST and volcanic aerosols, the model-based climate simulations are well able to represent some components of large-scale and low-frequency variability even without assimilating meteorological observations. Reanalysis of synoptic observations can add a more detailed representation of the past behavior of the atmospheric circulation, including the weather. However, as a result of biases in models and observations compounded with changes in the observational coverage, it is difficult to assimilate observations without introducing false climate signals.

The first pilot study in ERA-CLIM embodies an ensemble of climate model integrations without data assimilation. Ten atmospheric model integrations for the years 1899 to 2009 were produced at ECMWF. Horizontal spectral resolution is T159 (about 125 km in grid-point space), using 91 levels in the vertical from the surface up to 1 Pa, and a time step of one hour. This ensemble, denoted by ERA-20CM, forms the first step toward a 20th century reanalysis within ERA-CLIM, which is described in more detail later in this document. Sea-surface temperature and sea-ice concentrations are prescribed by an ensemble of realizations (HadISST2), produced by the Met Office Hadley Centre within ERA-CLIM, as described above. Variations among these realizations reflect uncertainties in the available observational sources on which this product is based. Forcing data used in the model radiation scheme, associated with solar forcing, greenhouse gases, ozone and aerosols, follow CMIP5 recommendations. Any effect of uncertainties in these data is neglected. Both the ocean-surface and radiative forcing data incorporate information about the long-term evolution of climate trends in the 20th century, including the occurrence of major events, such as the El Nino-Southern Oscillations and volcanic eruptions.

Although ERA-20CM is not able to reproduce actual meteorological behaviour on synoptic time scales, it can provide a good reference for the low-frequency variability of the atmosphere for the 20th century, and is well suited to project global warming and major events onto other geophysical parameters that are not directly provided in the forcing data.

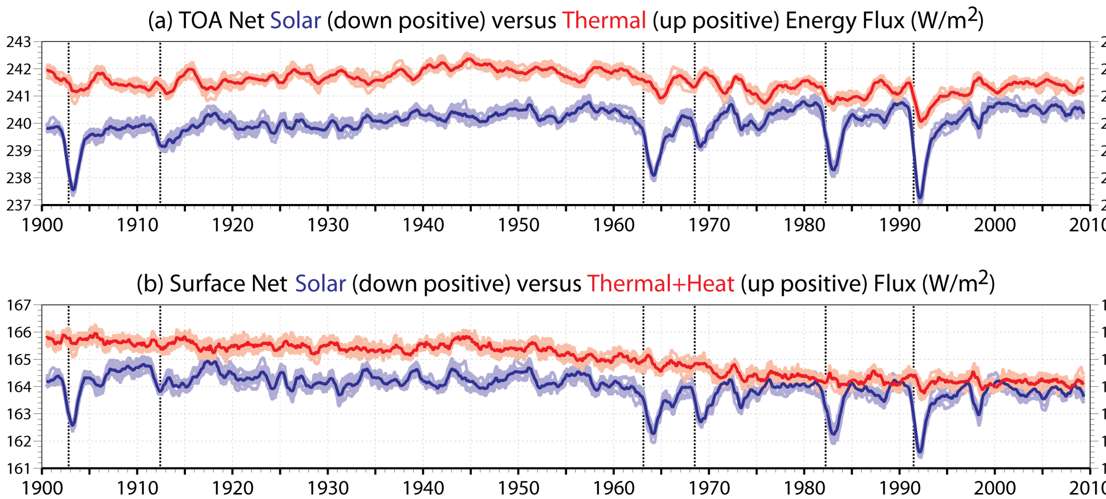


Figure 5: *Evolution of the one-year moving average of energy budgets in ERA-20CM for (panel a) TOA net solar radiation (blue, downward positive) and TOA net thermal radiation (red, upward positive), (panel b) surface net solar radiation (blue, downward positive) and the sum of surface net upward thermal radiation, latent heat and sensible heat (red, upward positive). Dark colors represent the ensemble mean, light colors the individual ERA-20CM members. The vertical black solid lines indicate the eruption dates (in chronological order) of Santa Maria, Novarupta, Mount Agung, Fernandina Island, El Chichón, and Pinatubo.*

The effect of volcanic eruptions on the net solar radiation appears very clear (Fig. 5). For all major eruptions during the 20th century (indicated by the solid vertical lines), a sharp dip up to several W/m^2 is observed at the top of the atmosphere (TOA), which is the result of an increased reflection and absorption by stratospheric aerosols. These dips are followed by a much weaker response in net thermal radiation (TOA and surface) and the net change can have a small temporary effect on the average global surface temperature. Although the values of net energy fluxes appear to be inaccurate at the TOA and surface, the increase in downward net energy flux of about 1 W/m^2 from the 1960s is correct.

Model behavior is found to be good in the lower troposphere. However, ERA-20CM does exhibit some clear biases elsewhere. Compared to ERA-Interim, ERA-20CM is biased cold and dry in the higher troposphere and around the tropopause, while it is biased warm in the upper part of the stratosphere (panel (a) of Fig. 6). The quasi-biennial oscillation is represented, however due to the lack of observations and sub-optimal parametrisation, its phase and period are generally incorrect. The westerly phase of the semi-annual oscillation is far too weak.

Despite these model biases, trends in ERA-20CM are generally found to be realistic. Features of note are a warming maximum in the tropical upper troposphere and widespread stratospheric cooling that increases with height (panel (b) in Fig.6). An enhanced near-surface warming is apparent over the Arctic, most likely the result of receding sea ice. The upper parts of the jets have slightly intensified and the upper troposphere has become wetter.

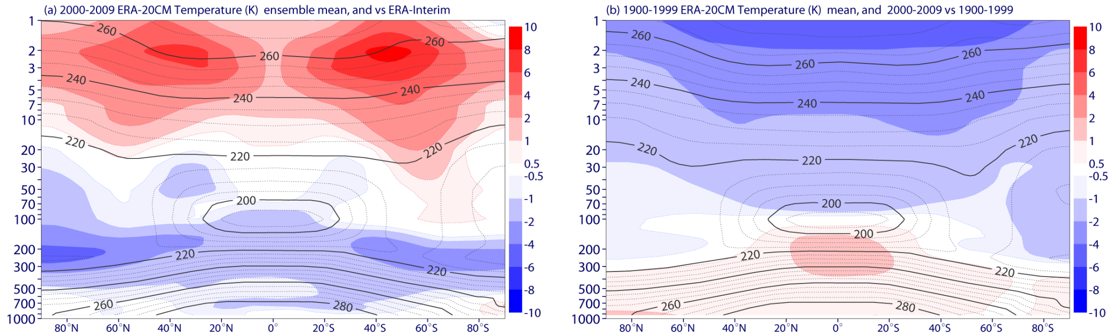


Figure 6: *2000 to 2009 ERA-20CM ensemble mean (contours), its difference with ERA-Interim (colours), 1900-1999 ERA-20CM ensemble mean and the 2000 to 2009 trend (colours) for left and right panels, respectively, for temperature.*

Two-metre temperature follows trends in the prescribed HadISST2 forcing, with an amplified climate signal over land in recent years. Overall, the temperature rise over land is in fair agreement with the CRUTEM4 observational product (Fig 7). Over the last two decades the warming over land exceeds the warming over sea, which is consistent with models participating in the CMIP5 project, as well with the ECMWF ERA-Interim reanalysis.

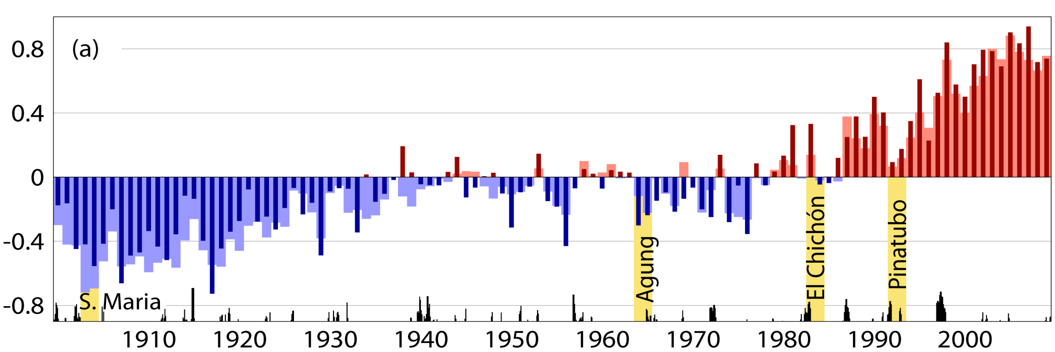


Figure 7: *Annual-mean temperature anomalies (K) relative to 1961-1990 for the ERA 20CM ensemble mean (broader, lighter-coloured bars) and CRUTEM4 (version 2.0.0 from www.metoffice.gov.uk/hadobs; narrower, darker-coloured bars) for area weighted averages taken over all grid boxes for which CRUTEM4 has values. Timings of El Niño events and volcanic eruptions are indicated as described in Hersbach et.al 2013.*

ERA-20CM also reproduces important aspects of inter-annual variability, with a tendency for relatively cool years following the major volcanic eruptions and warm years following El Niño SST maxima. Low frequency variability of total column water vapour also follows trends in SST (directly over the ocean and indirectly over land). The increase of evaporation over the last decades predominantly rains out over the ocean, leading to a decrease in relative humidity over land.

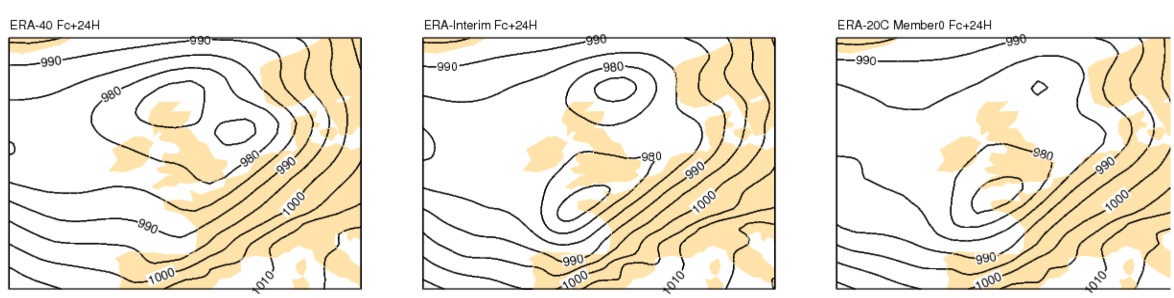
The results of ERA-20CM are freely available from ECMWF, embracing monthly-mean fields for many geophysical parameters, and synoptic fields for a small, essential subset.

Atmospheric reanalysis using surface observations (1900-2010)

ERA-CLIM produced a global reanalysis of the atmosphere, land, and ocean waves, referred to as ERA-20C. It employed the ECMWF model and data assimilation system at 125 km global resolution with 91 vertical levels extending from the surface up to 0.01 hPa (approximately 80 km). The assimilated data consist of surface pressure and marine wind observations, including many observations produced by ERA-CLIM data rescue activities.

The production of ERA-20C was carried out in two stages. A first set of experiments were conducted to establish the expected uncertainties in the climate reconstructions with such a reanalysis system, by employing an ensemble of 10 realizations, and to identify data problems (see detailed report by Poli et al., 2013). Each of the 10 members of the ensemble used a different global estimate of sea-surface temperature and sea-ice concentrations provided by the Met Office within ERA-CLIM. Other sources of uncertainties were included, such as the model's physics and the imperfect observations. Production of this ensemble mobilized about 7% of ECMWF high-performance computing for an 8-month duration between end 2012 and mid-2013. It was split in six streams, which each advanced at about 30 days/day. Total data volume produced was about 700 Tb.

Investigation of the output suggested a higher-than-expected accuracy in terms of weather events. For example Figure 8 shows a map of the meteorological situation forecast and analysed by ERA-20C for the European Great Storm of October 1987, compared with those obtained from earlier comprehensive reanalyses based on upper-air and satellite observations, such as ERA-40 and ERA-Interim. A forecast based on the ERA-20C reanalysis, based on surface observations only, is better than the forecast based on the earlier ERA-40 reanalysis. This improvement illustrates the advances in data assimilation that have taken place in recent years.



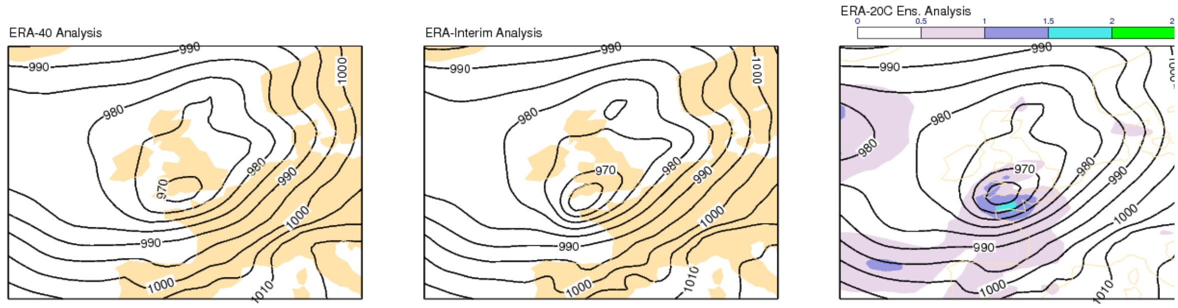


Figure 8: *Maps of predicted (24 hours ahead) and analysed mean sea-level pressures over Europe for 16 October 1987 (00 UTC), in ERA-40, in ERA-Interim, and in the ERA-20C ensemble experiment (with overlaid ensemble spread to show areas of greater uncertainties).*

In terms of producing realistic uncertainty estimates, the first ERA-20C experiments also demonstrated that the ensemble spread achieved for observed variables was consistent with departures with respect to assimilated observations. Figure 9 shows that the temporal evolution of root-mean-square (RMS) pressure observation departures agrees fairly well with in the predicted accumulated observation errors and reanalysis state uncertainties as estimated by the ensemble spread.

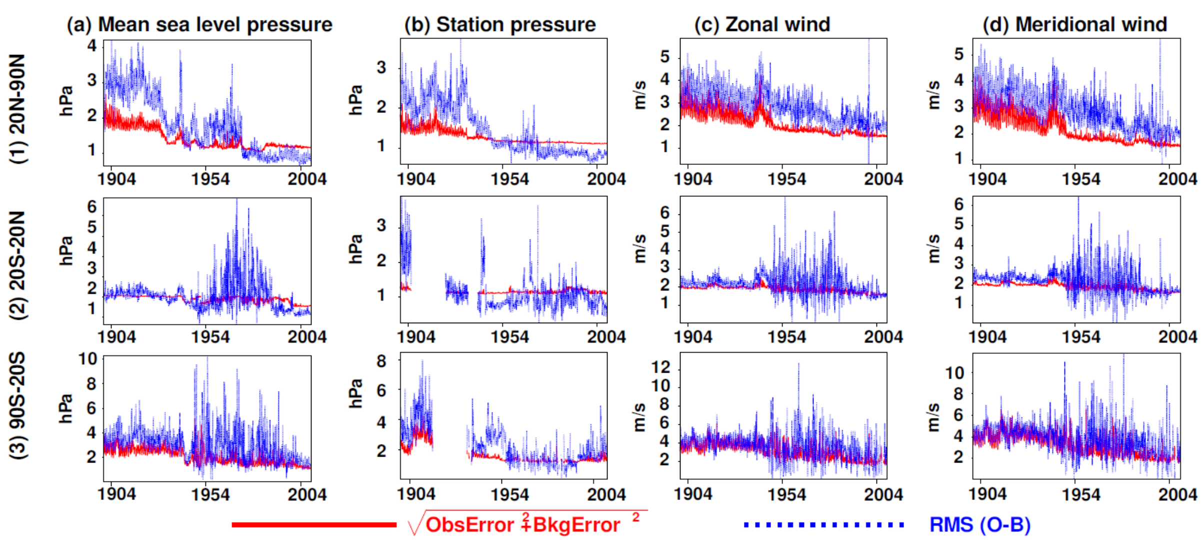


Figure 9: *Validation of the error estimates: comparison between (in blue) the root-mean-square of the reanalysis minus observation departures and (in red) the estimates produced by adding the assumed observation errors to the estimated reanalysis state uncertainties (estimated by the ensemble spread)*

These first reanalysis experiments also helped identify a large number of problematic observations, e.g. due to stuck sensors, miscoding of reports, etc. In addition, investigation of the reanalysed output highlighted two key problems in the experiments, namely an substantial bias in the representation of atmospheric pressure tides and an unwanted spurious signal in the upper-air temperatures imposed by an unrealistic imprint of surface data in the high regions of the atmosphere.

To address these problems, a second set of experiments was conducted. This set of experiments made use of the state uncertainty estimates generated in the first set. The atmospheric tidal bias was reduced by using a smaller model time-step (30 minutes instead of 60 minutes previously). An algorithm was devised to detect stations reporting suspicious time-series with time invariant (or constant) pressures. Because the computing cost was essentially a tenth of the first set of experiments, this second production could be completed in 7 weeks only. It was split in twenty-two streams, each advancing at about 50 days/day. Initial investigation suggests that the dataset thus produced preserves the known large-scale climate trends -- on average: near-surface warming and cooling at stratospheric levels -- while adding the daily realism brought by the assimilation of surface observations and already noted with the first ensemble production. Figure 10 shows the time-series of globally averaged temperature anomalies at selected altitudes in the two sets of experiments; the problems found at the higher levels are corrected in the final production.

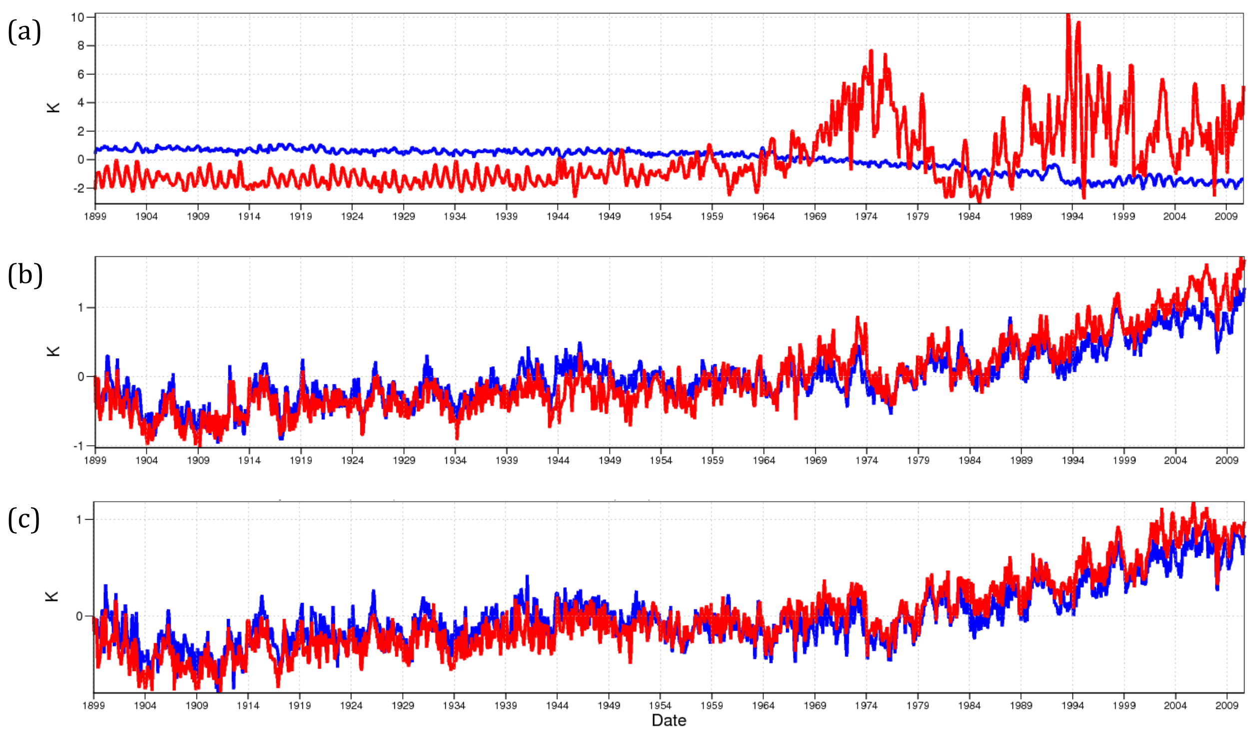


Figure 10: *Time-series of temperature anomalies (relative to the mean over the period 1900-2010) in the first set of ERA-20C experiments (red) and in the final production (blue), for three model levels, (c) near the surface, (b) near 700 hPa (or about 3 km altitude), and (a) near 50 hPa (or about 20 km altitude).*

The dataset produced in the second set of experiments, and its accompanying climate summaries (monthly means) shall become the final ERA-20C product. It will be copied onto disc for fast-access and delivery to the public, with a release planned around mid-2014.

High-resolution land-surface reanalysis (1900-2010)

Using methodology described in Balsamo et al. (2012), a high-resolution (25km) global reanalysis of the land surface for the period 1900-2010 is currently in production. This dataset, referred to as ERA-20CL, uses meteorological forcing (temperature, humidity, precipitation) from the ERA-20C reanalysis. The use of high-resolution information about orography and other properties of the land surface allows for a more accurate representation of land-surface processes, as illustrated in Figure 11. Estimates of soil moisture from ERA-20CL are superior to those from ERA-20C as a result of more detailed soil texture information. Meteorological forcing from ERA-20C is adjusted based on local orography and temperature lapse rates. In addition, the land surface model used in ERA-20CL enables larger dynamic changes in soil moisture than the model used in ERA-Interim.

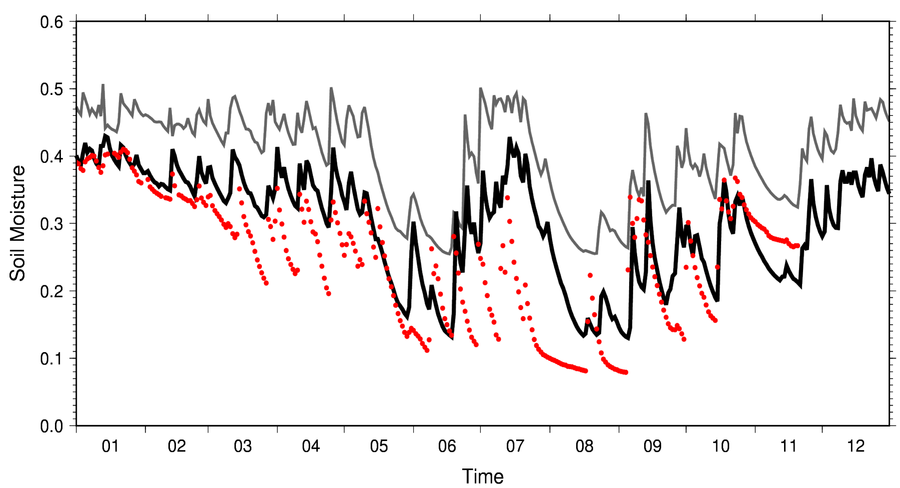


Figure 11: *Time series of observed soil moisture at the location of Uapb-lonoke Farm, AR, USA (34.85N, 91.88W) in 2007 (red dots), collocated ERA-20C (gray line) and ERA-20CL (black line).*

Production of ERA-20CL is scheduled to complete in March 2014; the dataset will be released to the public by summer 2014.

The Observation Feedback Archive (OFA) facility

An internet-based Observation Feedback Archive (OFA) has been developed to allow users of reanalysis products direct access to all input observations used in the reanalysis. The OFA provides an intelligent and user-friendly interface via a web browser, to the observations themselves but also to supplementary information generated by reanalysis, such as quality indicators, bias estimates, and estimates of observation uncertainty.

The OFA currently contains the ISPD and ICOADS data collections used as input to the first ERA-CLIM pilot reanalysis of the 20th century (ERA-20C). Feedback data from ERA-20C, i.e. all observations used with quality control information, analysis fit to observations, bias estimates and error standard deviations, will be published on the OFA in 2014.

The OFA relies on a sophisticated technical infrastructure surrounding ECMWF's Meteorological Archiving and Retrieval System (MARS). Substantial technical work was done in ERA-CLIM to create the OFA web interface and to enhance the technical infrastructure to allow permanent archiving and efficient retrieval of observation feedback information. The OFA is a permanent facility that will continue to be developed in support of future reanalysis projects.

Additional activities to prepare for future reanalyses of the 20th century

Experimental reanalysis using early upper-air weather observations. An ERA-CLIM pilot reanalysis was produced to explore the usage and impact of early upper-air observations. These data embrace the newly digitized data within ERA-CLIM, as described above, plus existing archives such as the CHUAN data set (Stickler et. al. 2010) and NCAR holdings. They include profiles of temperature, wind and humidity, recorded on either height or pressure levels.

Reanalysis of upper-air weather observations prior to the International Geophysical Year (1957-1958) is still a novelty. For this pilot reanalysis, denoted by ERA-PreSAT, the period of 1939 to 1949 was chosen to illustrate the added value of increased upper-air observations nearing the conclusion of the Second World War. In addition to all surface data as assimilated in ERA-20C, ERA-PreSAT has used upper-air temperature and wind data. Humidity was passively monitored, for now. Besides the usage of these upper-air data, the setup of ERA-PreSAT was equivalent to that of ERA-20C.

Only one member was integrated, based on SST and SIC estimates from HadISST.2 and CMIP5 radiative forcing data as used in the ERA-20C control. Identical confidence-levels on the model background were used in the data assimilation. The identical setup of the reanalysis experiments (including spatial resolution) allows for a clean assessment of the impact of the upper-air observations.

The positive impact of using upper-air data is illustrated in Table 2, which displays statistics for upper-air temperature during winter 1940-1941. In this period upper-air wind observations were much more abundant (452,784 versus 16,884 for temperature). However, the combined impact of wind and temperature profiles is evident. Prior to assimilation the deviation from the ERA-PreSAT collocated model temperature is smaller than it is for ERA-20C. This illustrates the positive effect of previously used upper-air wind and temperature data, which is a healthy indication for an improved quality of the ERA-PreSAT upper-air model state. After assimilation, the comparison is, naturally, much better than it is for ERA-20C, where this data was not used. For ERA-20C the upper-air observations serve as an independent verification; the small decrease in standard deviation shows the positive impact of the assimilation of pressure and wind data at the surface.

	Upper air temperature differences relative to observations [K]			
	Model prediction (background)		After assimilation (analysis)	
	Bias	STDV	Bias	STDV
ERA-20C	1.44	5.26	1.20	5.16
ERA-PreSAT	0.37	3.90	0.11	1.87

Table 2: *Historical upper-air temperature (Kelvin) for data between December 1940 and February 1941 (16,884 in total) compared to the collocated model equivalent before assimilation (background) and after assimilation (analysis) for the ERA-20C reanalysis using surface observations only and the ERA-PreSAT reanalysis using in addition upper-air data. Statistics are for all available data, i.e., including outliers and low quality data that were rejected by the ERA-PreSAT quality control system.*

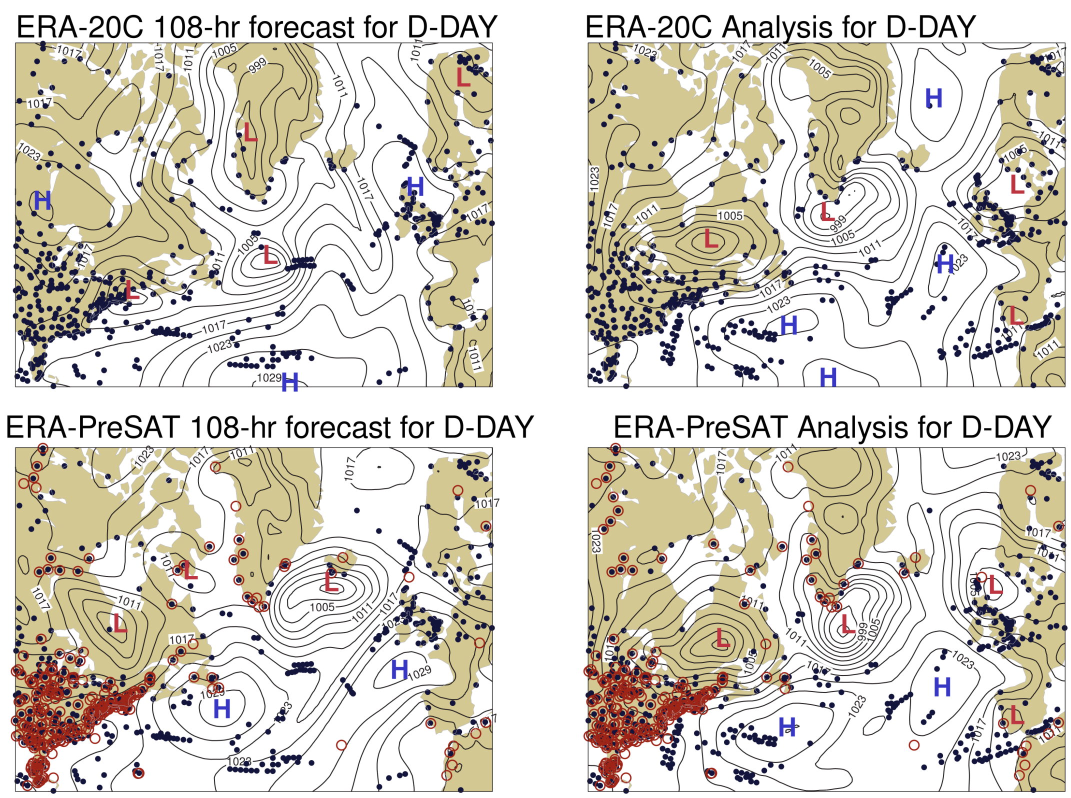


Figure 12: *Forecast from the 2nd of June 1944 (right) and analysis, all verifying on 12 UTC 6 June 1944 (D-Day) for ERA-20C (top) using surface observations only (black dots), and ERA-PreSAT (bottom) using in addition upper-air data (red circles; mainly pilot balloon wind). Coverage corresponds to data used by the analysis for the 2nd June (left) and 6th June (right), respectively.*

Another illustration of the added value of upper-air data regards the weather situation for the D-Day landing on June 6 1944, illustrated in Fig. 12. The original landing was planned one day earlier, but was delayed due to anticipated adverse weather conditions. A low-pressure system in the Atlantic was bringing strong winds towards the Channel that could set up waves that could trouble the landing crafts. The weather for the 6th of June was expected to be more favorable, due to an evolving ridge diverting the flow to an off-land direction. A weather map made by the allied chief meteorologist James Stagg valid for 13 UTC 6 June (Stagg, 1971), appears to resemble the ERA-PreSAT analysis very well (lower right panel). Both show the three low-pressure systems alternated by the two high-pressure areas more southwards. This pattern is visible in the ERA-PreSAT from the 2nd of June. Although this forecast (lower left panel) may not represent the low pressure close to Scotland, it does anticipate fair weather condition for D-Day. In contrast, the forecast and verifying analysis from ERA-20C (top panels) are less favourable.

Reanalysis experiments for the modern observing period. Many short reanalysis experiments were produced in ERA-CLIM to test the impact of specific input observations and various data assimilation configurations for future reanalyses of the modern observing period. We highlight one such experiment in this report.

The ability of the modern ECMWF data assimilation system was tested to address a major known deficiency found in the ERA-Interim reanalysis: a sudden drying in the reanalysis products that occurred whenever microwave imager satellite data were assimilated. A detailed discussion of this problem can be found in Section 5.2.1 of Dee et al (2011). Figure 13 shows a nearly one-year-long pair of reanalyses, which differ by the assimilation (or not) of the assimilation of the data that caused the problem in ERA-Interim. The difference between the two time-series is still significant, at about 0.10 mm/day of global precipitation, but is much reduced compared to the problem found in ERA-Interim, which was between 0.20 and 0.30 mm/day.

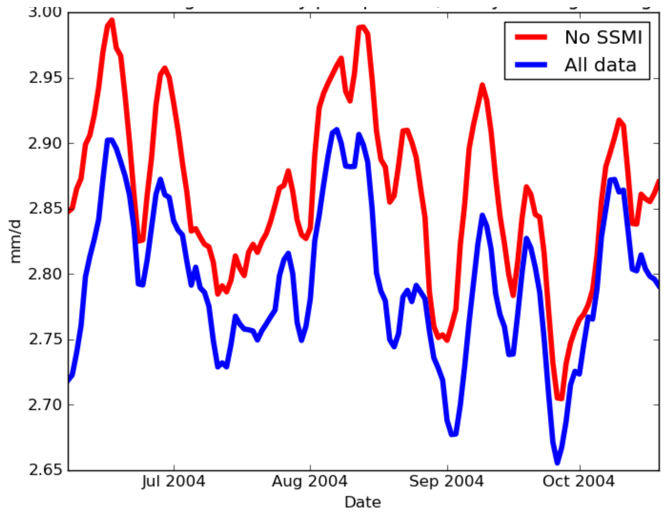


Figure 13: *Time-series of global average of total daily precipitation for two reanalyses of the modern time period with a recent version of the ECMWF data assimilation system (excerpt from results presented at the 4th WCRP International Conference on Reanalyses, 2012).*

Progress in reducing uncertainties in climate observations

Improving the construction of time-series from observations, whatever the method, requires dedicated attention to treat what is commonly referred to as "data problems". Such problems can stem from a variety of factors, including, and not limited to, incorrect time or location of observations, incorrect units used in the reporting, incorrect reference of measurement, non-reported change in instrumentation.

On the temporal aspect, ERA-CLIM has produced an internal report summarising progress on improving the temporal consistency and homogeneity of in-situ surface and upper-air observations. This included work performed at the University of Lisbon on homogenization of time series of Portuguese temperature observations (Morozava and Valente 2012). It also includes work by the University of Vienna to improve the homogeneity of pre-1958 radiosonde temperatures, using data from NOAA's 20th-Century Reanalysis Project as a reference. Figure 14 shows an example of unadjusted (top), correction (middle), and adjusted (bottom) time-series of meridional wind speed at 700 hPa, for a station located in the USA. The method developed by the ERA-CLIM team is able to detect and correct for a change in behaviour in the 1940s. The report also includes work undertaken at ECMWF to allow break detection of data records in ICOADS, which was needed to perform the ERA-CLIM pilot reanalysis of the 20th century.

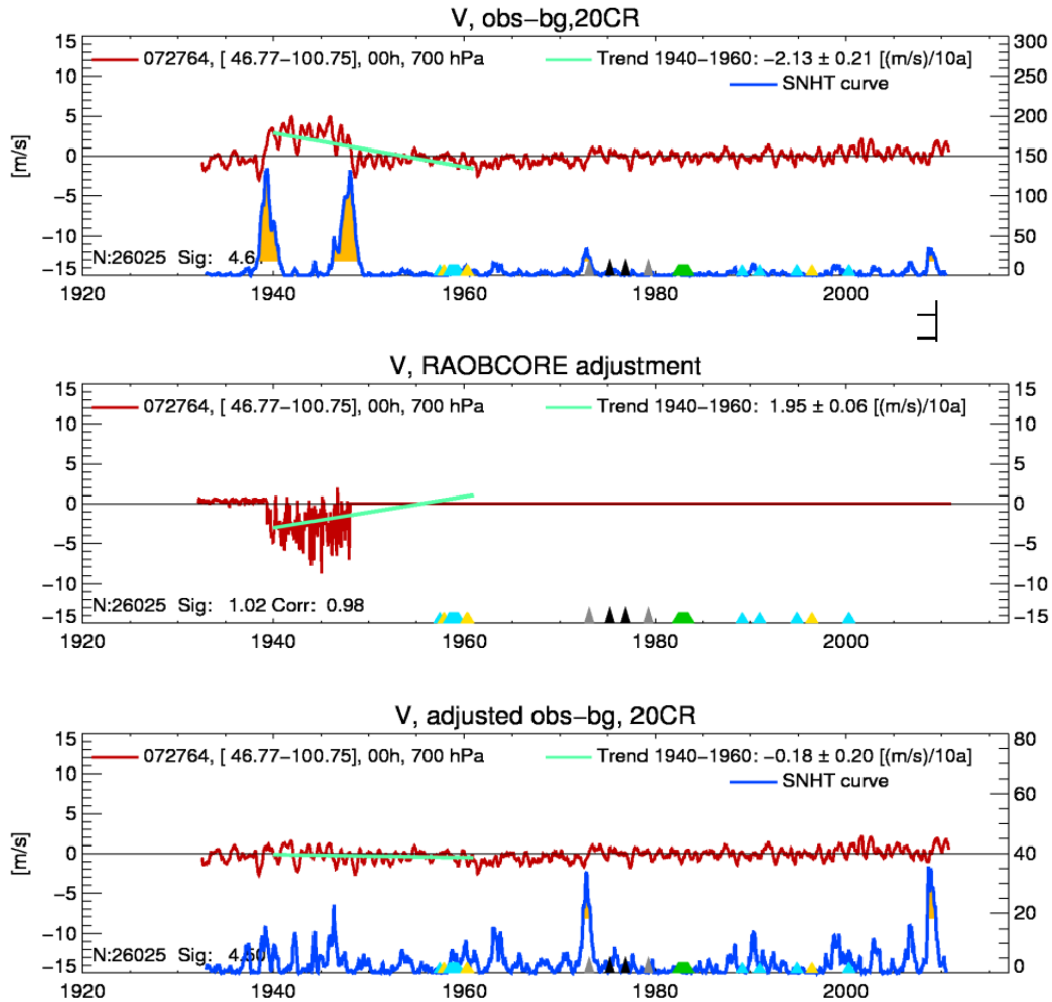


Figure 14: *Time-series of unadjusted (top), correction (middle), and adjusted (bottom) time-series of meridional wind speed at 700 hPa, for a station located in the USA.*

Regarding the use of advanced (variational) methods for automatically estimating and adjusting biases in observations during reanalysis, a first assessment of bias adjustments generated in ERA-20C was reported in Poli et al (2013). A proposal for bias correction and homogenization methodology to be used in a future extended climate reanalysis was also drafted. The proposal addresses in-situ surface, in-situ upper-air, and satellite observations.

On the question of identifying incorrect references, for example for measuring wind direction, a variational scheme was developed to estimate systematic errors at radiosondes launching stations in referencing wind direction observations.
To address problems with reporting incorrect time and location, several iterations took place between the various ERA-CLIM partners to improve the CHUAN dataset and correct mistakes as cross-checks were conducted.

Ocean observations for climate reanalysis

The UK Met Office Hadley Centre, through its participation in ERA-CLIM, has developed several important collections of ocean observations targeted for future climate reanalyses. This includes a new quality-controlled collection of ocean subsurface temperature and salinity profile data for the period 1900 to present, referred to as EN4 (Good et al 2013). It contains all observations from the previous EN3 database, as well as a large number of new measurements (especially pre-1950) from a variety of sources. The data processing and quality control applied to the raw data has been completely revised.

In addition, HadIOD, an Integrated Ocean Database of global temperature and salinity observations covering 1900-2010, was developed in ERA-CLIM. HadIOD merges data from EN4 with surface observations from the International Comprehensive Ocean-Atmosphere Data Set (ICOADS). Finally, an ensemble of sub-surface ocean observations was produced to allow users to sample equally plausible observations and thus study the impact of observation errors on their applications.

Potential impact, main dissemination activities and exploitation of results

Impact on climate science

The impact of atmospheric reanalysis on climate science is best grasped by examining the citations of the journal articles that describe previous reanalysis datasets. These are among the highest-cited literature in the geosciences, with (according to Google Scholar at the time of this writing) 1681 articles citing Dee et al 2011 (on the ERA-Interim reanalysis) and 4531 articles citing Uppala et al 2005 (on the earlier ERA-40 reanalysis). The work reported in those articles includes a broad range of subjects in atmospheric science, oceanography, climate science, and many applied fields e.g. in energy, health, environmental science, etc.

International awareness and interest in the ERA-CLIM project has grown proportionally, with numerous users in the science community inquiring about availability and characterization of future data products. The full impact of the outcomes of this project on climate science will occur after the major datasets have been published, later in 2014.

Dissemination of scientific results

Various publications were prepared during the ERA-CLIM project reporting on important outcomes e.g. in data rescue, data analysis, data assimilation, and climate reanalysis. A selection of those publications is included in the list of references below.

An article by the project Coordinator (Dee et al. 2014) outlining the strategy for ERA-CLIM and beyond, targeted for a broad scientific audience, will appear in the widely read Bulletin of the American Meteorological Society in early 2014.

The Coordinator has presented the ERA-CLIM project vision and approach at major international conferences including the WCRP Open Science Conference in Denver (2011), the 4th WCRP International Conference on Reanalyses held in Silver Spring (2012) and the 13th Annual Meeting of the European Meteorological Society in Reading (2013). The Coordinator has also presented numerous seminars and lectures about the project at various institutes in Europe and elsewhere.

A complete list of publications, technical reports, presentations at international conferences and workshops has been submitted separately as part of the reporting required by the Commission.

Dissemination of data products

Several data collections with observations suitable for climate reanalysis have been developed and/or improved in ERA-CLIM:

- An ERA-CLIM extension of the Comprehensive Historic Upper-Air Network (CHUAN), available from the University of Bern and via the PANGAEA data repository at a permanent web location (doi.pangaea.de/10.1594/PANGAEA.821222).
- Additional in-situ surface observations recovered in ERA-CLIM; these are available directly from the ERA-CLIM project partners. These data are also supplied to the National Climate Data Centre at NOAA for inclusion in their data collections
- Several collections of ocean observations (EN4, HadIOD) to be made available via the Met Office Hadley Centre (after publication of a journal article describing each dataset)

Gridded products:

- HadISST.2: Global estimates of sea-surface temperature and sea-ice concentration from 1899-2010. These will become available via the UK Met Office Hadley Centre (after publication of a journal article describing the dataset).
- ERA-20C: Global atmospheric reanalysis from 1900-2010, 3-hourly data at 125 km spatial resolution and 91 vertical levels. The data are being prepared for public release via the internet at ECMWF, expected mid-2014.
- ERA-20CM: Global atmospheric model simulations from 1900-2010, monthly data at 125 km spatial resolution and 91 vertical levels. The data are being prepared for public release via the internet at ECMWF, expected mid-2014.

- ERA-20CL: Global land-surface reanalysis from 1900-2010, 3-hourly data at 25 km spatial resolution. The data are being prepared for public release via the internet at ECMWF, expected mid-2014.

Observation feedback:

- The Observation Feedback Archive (OFA) at ECMWF provides direct access to all input observations used in reanalysis, together with information such as quality indicators, bias estimates, and estimates of observation uncertainty.

ECMWF provides open access to ERA-CLIM reanalysis products and other research datasets at http://apps.ecmwf.int/datasets. This link also points to the Observation Feedback Archive (OFA), which is a permanent facility that will continue to be developed in support of future reanalysis projects.

The Met Office Hadley Centre provides internet access to observations data sets, including those developed in the ERA-CLIM project, via http://www.metoffice.gov.uk/hadobs.

The University of Bern Oeschger Centre for Climate Change Research has created a metadatabase for in-situ data rescue, which includes data inventories for various sources, at http://www.oeschger-data.unibe.ch/metads/

The University of Bern Oeschger Centre for Climate Change Research also maintains a data server for the CHUAN collection at http://www.oeschger.unibe.ch/research/projects/historicalupperair/index_en.html

Preparing for Copernicus Climate Change Services

Reanalysis developments have traditionally been motivated by the need to improve the use of observations for weather forecasting, and to provide datasets for research and education. The scientific user community has derived great value from global reanalyses of the atmosphere, ocean, land surface, and atmospheric composition, as evidenced by numerous publications that refer to these datasets. Increasingly, demand for reanalysis data products is driven by the need for accurate and comprehensive information on climate change and variability.

Different types of reanalyses will be needed to serve both climate science and climate change services in future. These include extended reanalyses spanning as much of the instrumental record as possible, and high-resolution reanalyses of the modern observing period that can be continuously updated in real-time. The ERA-CLIM project has made substantial contributions to the development of both types of reanalyses, and to the improvement of the instrumental observation record in general. The overarching goal is to make the best possible use of observations to serve the interest of society.

Global reanalysis generally, and the ERA-CLIM project specifically, has played an important role in ongoing discussions of a future Copernicus Climate Change Service, e. g. at the GMES and Climate Change conference held in Helsinki (2011). The success of ERA-CLIM has led to proposal for a follow-up ERA-CLIM2 project, which has now been funded by the European Commission until end 2016. ERA-CLIM2 will develop coupled data assimilation capability targeted for production of improved reanalyses of the coupled climate system, following the vision and approaches developed in ERA-CLIM. Further investments are needed to transform reanalysis from the valuable research activity it currently is into a dependable operational service. Resources for this purpose will likely become available through the Copernicus Climate Change Services about to be implemented by the European Commission.

Project public website

The project website (www.era-clim.eu) contains basic information about the project goals, consortium, methods and products. Planning of the project did not fully account for the resources needed to maximize the effectiveness and impact of the web site. Many of the data products developed in ERA-CLIM will become available during 2014, at which time a renewed effort will be made to publicize the value and impact of the ERA-CLIM project outcomes via the web.

References

Balsamo, G., et al., 2012: ”ERA-Interim/Land: A global land-surface reanalysis based on ERA-Interim meteorological forcing.” *ERA Report Series, 13, ECMWF, UK.*

Dee, D. P., et al., 2011: "The ERA-Interim reanalysis: Configuration and performance of the data assimilation system." *Quarterly Journal of the Royal Meteorological Society* 137, 553-597.

Dee, D. P., et al., 2014: "Toward a consistent reanalysis of the climate system." *Bulletin of the American Meteorological Society*, in press.

Durre, I., R. S. Vose, D. B. Wuertz, 2006: “Overview of the Integrated Global Radiosonde Archive”. *J. Climate, 19, 53-68.*

Good, S. A., M. J. Martin and N. A. Rayner, 2013: “EN4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates”. *Journal of Geophysical Research: Oceans,* 118, 6704-6716,

Hersbach, H., et al., 2013: “ERA-20CM: a twentieth century atmospheric model ensemble.” *ERA Report Series, 16, ECMWF, UK.*

Ilin A. and A. Kaplan, 2009: “Bayesian PCA for reconstruction of historical sea surface temperatures”. *Proc. of the Int. Joint Conf. on Neural Networks (IJCNN 2009)*, 1322-1327, Atlanta, USA.

Karspeck, A. R. et al., 2012: "Bayesian modelling and ensemble reconstruction of mid-scale spatial variability in North Atlantic sea-surface temperatures for 1850–2008." *Quarterly Journal of the Royal Meteorological Society* 138, 234-248.

Kennedy, J. J., et al., 2011: "Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization." *Journal of Geophysical Research,* 116, D14.

Kennedy, J. J., et al., 2014: “The Met Office Hadley Centre Sea Ice and Sea-Surface Temperature data set, version 2, part 2: sea-surface temperature analysis. *Journal of Geophysical Research,* in preparation.

Morozova, A. L. and M. A. Valente, 2012: “Homogenization of Portuguese long-term temperature data series: Lisbon, Coimbra and Porto”. *Earth System Science Data Discussions*, 5, 521-584.

Poli, P. et al, 2013: “The data assimilation system and initial performance evaluation of the ECMWF pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C).” *ERA Report Series, 14, ECMWF, UK.*

Rayner, N. A. et al., 2003: “Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century”. *Journal of Geophysical Research*, 108, D14.

Rayner N. A., et al., 2014: “The Met Office Hadley Centre Sea Ice and Sea Surface Temperature data set, version 2, part 3: the combined analysis.” *Journal of Geophysical Research,* in preparation.

Stagg, J. M., 1971: “Forecast for ‘Overlord’”, Norton; First edition.

Stickler, A., et al., 2010: "The Comprehensive Historical Upper-Air Network." *Bulletin of the American Meteorological Society*, in press.

Stickler, A., et al., 2014a: "ERA-CLIM: Historical Surface and Upper-Air Data for Future Reanalyses." *Bulletin of the American Meteorological Society*, in press.

Stickler, A., et al., 2014b: "Description of the ERA-CLIM historical upper-air data." *Earth System Science Data* 6, 29-48.

Titchner, H. A., and N. A. Rayner, 2014: "The Met Office Hadley Centre Sea Ice and Sea-Surface Temperature dataset, version 2, part 1: Sea-ice concentration". *Journal of Geophysical Research*, doi: 10.1002/2013JD020316.

Uppala, S. M., et al., 2005: "The ERA-40 re-analysis." *Quarterly Journal of the Royal Meteorological Society* 131, 2961-3012.