# CERA-20C: An Earth system approach to climate reanalysis

**PATRICK LALOYAUX,
ERIC DE BOISSÉSON, PER DAHLGREN**

ECMWF has completed the production of a new global 20th-century reanalysis which aims to reconstruct the past weather and climate of the Earth system including the atmosphere, ocean, land, waves and sea ice. This coupled climate reanalysis, called CERA-20C, is part of the EU-funded ERA-CLIM2 project, which builds on the ERA-CLIM project. The latter produced ERA-20C, a 20th-century reanalysis for the atmosphere, land and waves only (*Poli et al.*, 2016).

First results show that CERA-20C improves on the representation of atmosphere–ocean heat fluxes and of mean sea level pressure compared to previous reanalyses. At the same time, there are undesirable discontinuities in ocean heat content and an excessive accumulation of Arctic sea ice.

To account for errors in the observational record as well as model error, CERA-20C provides a ten-member ensemble of climate reconstructions. As expected, the spread of the ensemble decreases over time as the observational record improves. However, verification for the year 2005 suggests that the spread should be larger to give a better indication of the confidence we can have in the reanalysis data.

## History of reanalysis at ECMWF

Since its creation in 1975, ECMWF has been a key player in the production of reanalyses, which provide a numerical description of the recent climate by combining models with observations. The initial focus was on producing atmospheric reanalyses covering the modern observing period, from 1979. The first of these, FGGE, was produced in the 1980s, followed by ERA-15, ERA-40 and ERA-Interim. The next reanalysis in this series, ERA5, is now in production after many years of research and a great deal of technical preparation (*Hersbach & Dee*, 2016). ECMWF is also producing ocean reanalyses. In 2016, the ORAS4 system was replaced by ORAS5, which incorporates the latest improvements in ocean models, data assimilation methods and forcing fluxes.

The various reanalysis products have proven to be an important resource for weather and climate-related research as well as societal applications at large. Reanalyses also support numerical weather prediction since they can be used for the initialisation of reforecasts, the calibration of ensemble forecasts and model validation and verification. Reanalyses make it possible to study the inter-annual variability of forecast skill and to test new model versions on past severe weather cases. ERA-Interim and ORAS5 are the current operational reanalyses at ECMWF for the atmosphere and the ocean, respectively. They are created via an unchanging frozen data assimilation system and model, which ingest all available observations to provide the best state estimate over the target period.

Extending these reanalyses further back in time is a tremendous scientific challenge as the observing system is very sparse before the availability of satellite data from the 1970s onwards, and especially before the arrival of radiosonde measurements in the 1930s. To tackle the unavoidable issue of the ever-changing observational network, the ERA-CLIM project has developed a whitelisting approach to data selection for reanalyses covering the whole 20th century. Instead of assimilating the full observing system at any time, only observations with a good spatial and temporal coverage over the entire century are used. Modern data assimilation systems are able to faithfully reconstruct the large-scale tropospheric circulation from surface pressure observations only. The quality of such reconstructions does depend on the observation density and will never outperform a system using all sources of observations, including satellite and upper-air measurements. However, by using a more consistent observational network the whitelisting approach avoids artificial variability and spurious trends generated by the introduction of new instruments. This makes it possible to extend climate reconstructions further back in time to cover a period of 100 years or more with a focus on low-

| Dataset | Period | Type | Ensemble | Atmosphere | Land | Waves | Ocean | Sea ice |
|---|---|---|---|---|---|---|---|---|
| ERA-Interim | 1979-present | Full observing system | | ✓ | ✓ | ✓ | | |
| ORAS5 | 1975-present | Full observing system | ✓ | | | | ✓ | ✓ |
| ERA-20C | 1900-2010 | Selected observing system | | ✓ | ✓ | ✓ | | |
| ORA-20C | 1900-2010 | Selected observing system | ✓ | | | | ✓ | ✓ |
| CERA-20C | 1901-2010 | Selected observing system | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

**Table 1** List of selected ECMWF reanalysis datasets showing the period covered, the observing system used and the different Earth system components included in the climate reconstruction.

frequency climate variability. The overall aim is to improve our ability to produce consistent reanalyses of the climate system, reaching back in time as far as possible given the available instrumental record.

In this context, ECMWF has produced the uncoupled atmospheric reanalysis ERA-20C, which covers the period January 1900 to December 2010. ERA-20C assimilates only conventional observations of surface pressure and marine wind, obtained from well-established climate data collections. Model forcings are specified from CMIP5 (Coupled Model Intercomparison Project Phase 5) recommendations to obtain an appropriate climate reconstruction. The atmospheric lower boundary conditions are prescribed using the UK Met Office Hadley Centre's HadISST2 monthly analysis product for sea-surface temperature and sea ice.

ERA-20C has delivered three-hourly products describing the spatial and temporal evolution of the atmosphere, land surface and waves. Following a similar approach, the ORA-20C reanalysis reconstructs the ocean and sea-ice state over the 20th century. Temperature and salinity profiles are assimilated into the ocean model, which is also constrained by fluxes from ERA-20C and a sea-surface temperature relaxation towards the HadISST2 product. Table 1 summarises the main features of some of the reanalyses produced at ECMWF.

For researchers wondering which reanalysis is the most appropriate for their project, it is important to understand the fundamental differences between the different types of reanalysis produced at ECMWF. Users interested in a temperature reference in the stratosphere for the assessment of a satellite in the 1980s might for example want to use ERA-Interim or ERA5, while users interested in precipitation anomaly trends in Europe over the last century may wish to opt for ERA-20C or CERA-20C.

ECMWF has also released a model-only integration of the 20th century, known as ERA-20CM, which includes no data assimilation. It is an ensemble of ten atmospheric model integrations which can be used to determine the systematic climate errors of the model.

## CERA-20C – a new approach

ECMWF's Roadmap to 2025, which summarises the Centre's new ten-year Strategy, highlights that, "As forecasts progress towards coupled modelling, interactions between the different components need to be fully taken into account, not only during the forecast but also for the definition of the initial conditions of the forecasts."

In this context, the reanalysis capabilities developed in the ERA-CLIM project have been extended to the ocean and sea-ice components in the ERA-CLIM2 project. A new assimilation system (CERA) has been developed to simultaneously ingest atmospheric and ocean observations in the coupled Earth system model used for ECMWF's ensemble forecasts (Box A). 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.

### Description of the coupled assimilation system  **A**

The CERA system is based on a variational method with a common 24-hour assimilation window shared by the atmospheric and ocean components. The coupled model is introduced at the outer-loop level 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 (*Laloyaux & Dee*, 2015). This means that air–sea interactions are taken into account when observation misfits are computed and when the increments are applied to the initial condition. In this context, ocean observations can have a direct impact on the atmospheric analysis and, conversely, atmospheric observations can have an immediate impact on the analysed state of the ocean.

For the CERA-20C datasets, the resolution of the atmospheric model is set to TL159L91 (IFS version 41r2), which corresponds to a 1.125° horizontal grid (125 km) with 91 vertical levels going up to 0.1 hPa. The ocean model (NEMO version 3.4) uses the ORCA1 grid, which has approximately 1° horizontal resolution with meridional refinement at the equator. There are 42 vertical ocean levels with a first-layer thickness of 10 m.

A new version of the CERA system is under development based on the coupled model at higher resolution, which will include a quarter-of-a-degree ocean model. This new coupled assimilation system will be used to produce a short reanalysis over a recent period. It will assimilate all types of observations including upper air, satellite, wave, land surface and sea-ice measurements.

CERA-20C is the first ten-member ensemble of coupled climate reanalyses of the 20th century (production details are given in Box B). 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 air–sea interface is relaxed towards the sea-surface temperature from the HadISST2 monthly product to avoid model drift while enabling the simulation of coupled processes. No data assimilation is performed in the land, wave and sea-ice components, but the use of the coupled model ensures a dynamically consistent Earth system estimate at any time.

## Impact of ocean coupling

### Heat fluxes

One of the benefits expected from a coupled assimilation system is a more consistent treatment of the air–sea interface. When decoupled, the ocean and atmospheric systems use boundary conditions that do not take into account ocean–atmosphere feedbacks.

In ERA-20C, the atmospheric lower boundary conditions come from the HadISST2 sea-surface temperature and sea-ice product, which does not contain any information about the ocean dynamics. In ORA-20C, the ocean is forced by the ERA-20C fields, which are fixed and cannot adjust
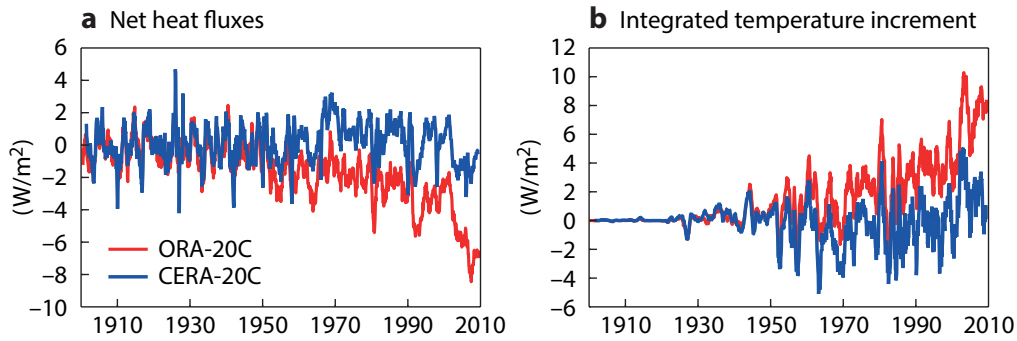
**a** Net heat fluxes

**b** Integrated temperature increment

**Figure 1** Time series of CERA-20C and ORA-20C control member values of (a) the global average of net air–sea heat fluxes and (b) the integrated temperature increment over the ocean.

to the ocean model behaviour. The long-term heat fluxes received by the ocean therefore suffer from inconsistencies at the air–sea interface. The resulting net heat fluxes over the ocean in ORA-20C show a negative trend from the 1940s onwards that tends to cool the ocean. To keep the ocean model close to the observed state, the ocean data assimilation has to compensate with a growing positive temperature increment (Figure 1).

In CERA-20C, the ocean and the atmosphere communicate every hour through the air–sea coupling at the outer-loop level of the variational method. Changes in the state of the atmosphere directly impact the ocean properties and vice versa. The combination of the coupled data assimilation and improvements in the atmospheric data assimilation corrects for the spurious trend in the net heat fluxes received by the ocean seen in ORA-20C. On average, heat flux and ocean temperature increments in CERA-20C oscillate around 0 W/m$^2$, suggesting a more balanced system.

### CERA-20C production and output **B**

The production of climate reanalyses requires large computing and archiving resources. To produce the CERA-20C dataset in a reasonable period of time, the period 1900–2010 was divided into 14 different streams of 10 years. Each production stream was initialised from the uncoupled reanalyses ERA-20C and ORA-20C. The first two years of each production stream were used for spin-up (only one-year of spin-up for the first stream) to produce the final climate dataset for the period 1901–2010.

The computation footprint of CERA-20C on ECMWF's high-performance computing facility is significant, with seven months of production using 20,000 cores, which represents 5% of the total resources. 500,000 variational problems had to be solved processing up to 5,000,000 observations, at a pace of one every 30 seconds.

The evolution of the global weather for the period 1901–2010 is represented by a ten-member ensemble of 3-hourly estimates for ocean, surface and upper-air parameters. This represents more than 1,600 terabytes of data that had to be archived in ECMWF's MARS archiving system.
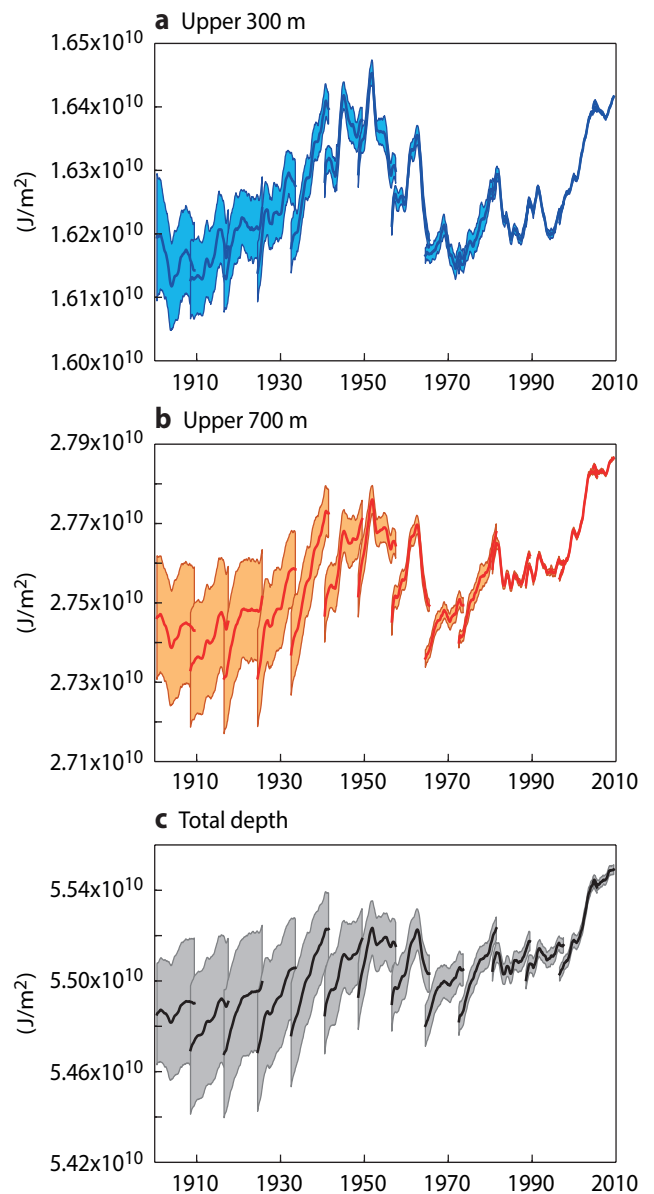
**a** Upper 300 m

**b** Upper 700 m

**c** Total depth

**Figure 2** Time series of the global average ocean heat content in the CERA-20C ensemble for (a) the first 300 metres, (b) the first 700 metres and (c) the whole water column. The solid lines are the ensemble mean and the shading shows the ensemble spread.

### Heat content

The evolution of ocean heat content over the 20th century is of particular interest as it has been identified in several studies as an indicator of ocean heat uptake, a process that is relevant to climate studies.

In CERA-20C, time series of heat content show discontinuities between streams resulting from the model drift from its initial state (Figure 2). The model drift reflects the fact that the initial conditions from ERA-20C and ORA-20C are inconsistent with the coupled model's natural state. In the early 20th century, when the uncertainty in the state of the ocean is high and the ocean model is poorly constrained by observations, the ocean component of CERA-20C drifts towards its preferred state. As the observing system grows, the uncertainty and the drift are reduced. The relatively well-observed upper ocean adjusts faster than the ocean interior, where the timescales of ocean processes are particularly slow and the observational constraints are very small. Further work is needed to understand and reduce the model drift so that the initial conditions and the ocean model behaviour are more realistic in poorly observed periods and areas.

### Sea ice

Ocean–sea ice interactions through the LIM2 model have only recently been included in ECMWF's coupled model. ORA-20C provides a first record of sea-ice conditions for the 20th century in ocean-only mode while CERA-20C is the first application allowing these interactions in coupled mode on an interannual timescale.

Some issues in the settings of the sea-ice coupling to the atmosphere were found in CERA-20C. They translate into a lack of summer melting, leading to the accumulation of Arctic sea ice over the years. The sea-ice thickness in CERA-20C is over 5 metres in most of the Arctic basin, more than twice the expected average of 2 to 2.5 metres, as seen in ORA-20C (Figure 3a,b). A major impact on sea-ice extent is avoided thanks to the relaxation applied at the air–sea interface. A new configuration for the coupling between sea ice and the atmosphere has been developed and tested. These coupled model experiments show a more realistic behaviour closer to the ocean-only mode (Figure 3c). Sea-ice interactions with the ocean and the atmosphere are highly sensitive processes and will need to be monitored carefully for future reanalysis.

## Mean sea level pressure

New climate reanalyses need to be produced periodically to benefit from the latest updates in the models and data assimilation systems developed for numerical weather prediction. The scientific community and dataset users also provide feedback and raise important issues which need to be addressed in future reanalyses. This is why reanalysis is an ongoing activity that should never be regarded as completed.

Scientists have identified an issue in the general circulation in the southern hemisphere in ERA-20C and in the 20th-century atmospheric reanalysis produced by the US National Oceanic and Atmospheric Administration (20CR). The time series of mean sea level pressure (MSLP) decreases significantly between 1900 and 1950 over the Antarctic region, leading to a substantial strengthening of the polar vortex in the first half of the 20th century in these reanalyses (Figure 4). The development of CERA-20C, which is based on ERA-20C infrastructure, provided an opportunity to address this spurious climate drift.

In any data assimilation scheme, the specification of observation and background errors is crucial as it defines the weights used to blend together the information from the measurements and from the model. Specifying the observation error over the entire 20th century is not straightforward since measurement processes and representativeness errors are not well known. For this reason, it was decided to keep observation errors constant in ERA-20C (1.1 hPa for surface pressure and 1.5 hPa for mean sea level pressure). It has been found that those values are too small at the beginning of the century, giving too much weight to the observations. The assimilation adjusts the flow where it is the least constrained to improve the fit to observations. This produces large positive MSLP
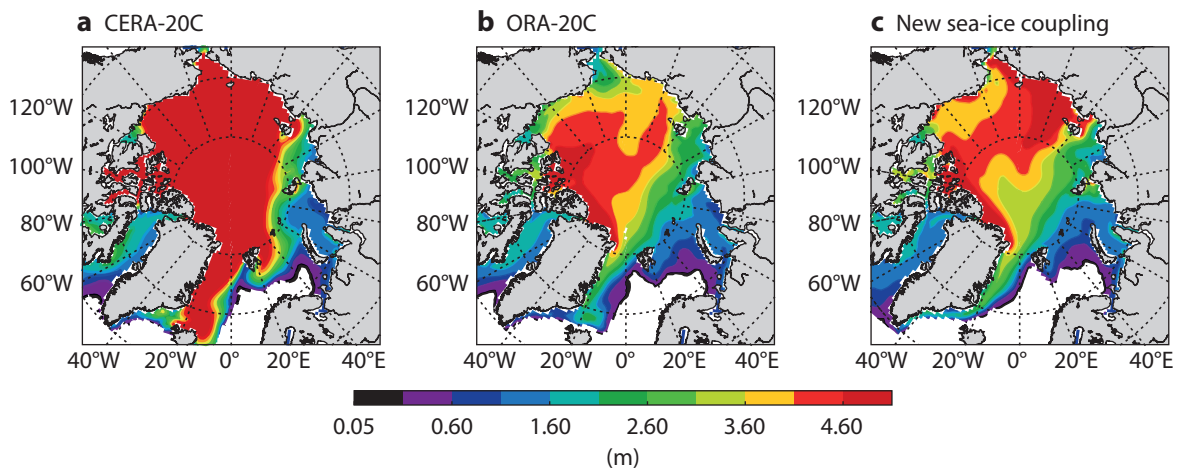


**Figure 3** Arctic sea-ice thickness in March 1932 from (a) CERA-20C, (b) ORA-20C and (c) a coupled model run with new sea-ice coupling.
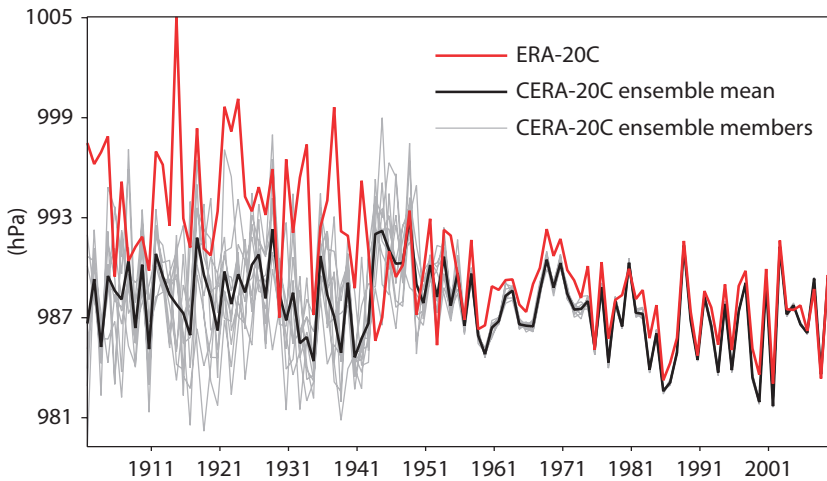
**Figure 4** Time series of mean sea level pressure (MSLP) for the latitudes 90°S–60°S averaged over the period September–November each year.

increments over the Antarctic region resulting in too high MSLP values in the analysis (Figure 5). This behaviour disappears when the observing system becomes denser in the 1950s, with the first SYNOP stations in Antarctica and more observations from ships in the Antarctic Circle.

The Desroziers diagnostic (*Desroziers et al.*, 2005) has been computed on the ERA-20C observation feedback statistics to estimate the a posteriori observation errors. The results show that a time-varying error would be more realistic. The surface pressure error should vary from 1.6 hPa in 1900 to 0.8 hPa in 2010, while the MSLP error should vary from 2.0 hPa in 1900 to 1.2 hPa in 2010.

For the production of CERA-20C, observation errors were adjusted in accordance with the results of the Desroziers

diagnostic. A larger observation error at the beginning of the century causes the assimilation to fit the model data slightly less closely to the observations and prevents large increments over the Antarctic region. As a result, the CERA-20C ensemble mean looks more realistic with better consistency in the climate trends. The larger ensemble spread at the beginning of the century reflects the larger uncertainties in the climate reconstruction as the region is poorly observed before the 1950s.

## Ensemble technique

For the first time, CERA-20C provides a ten-member ensemble climate reanalysis for all parameters and levels over the 20th century. Ensemble generation is based on the Ensemble of Data Assimilations (EDA) system developed
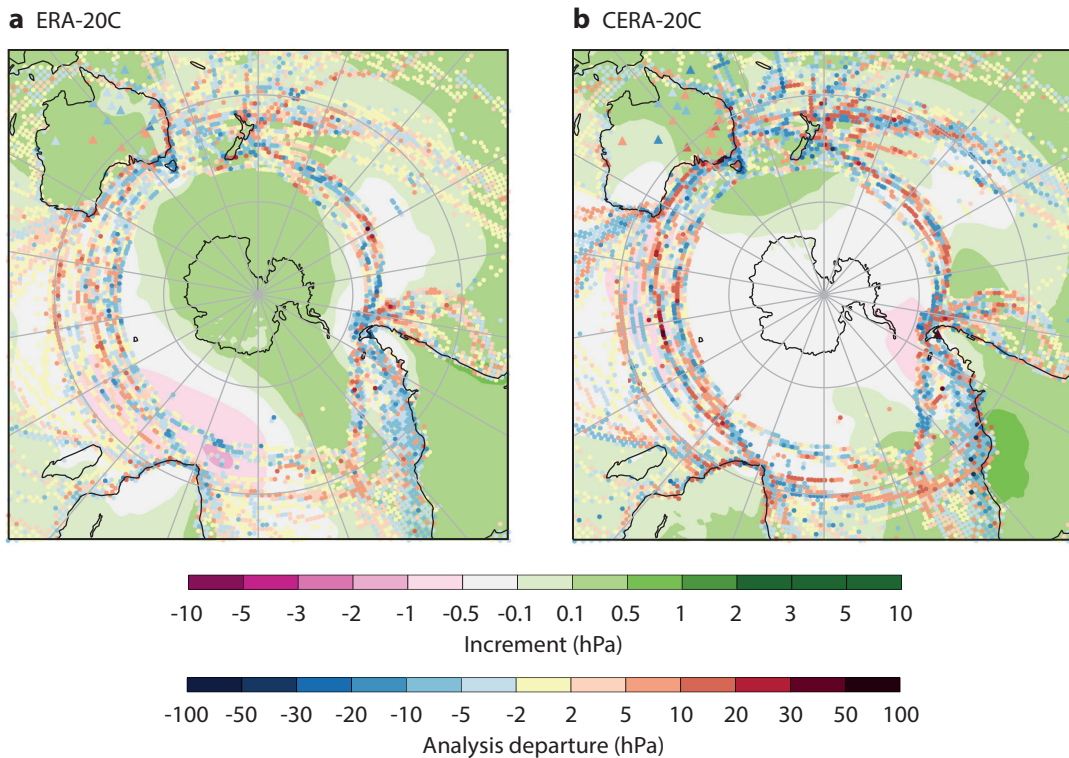


**Figure 5** Mean sea level pressure increments (shading) for the year 1924 in (a) ERA-20C and (b) CERA-20C. Dots and triangles represent the difference between the analysis and mean sea level pressure observations and surface pressure observations, respectively (analysis departures).
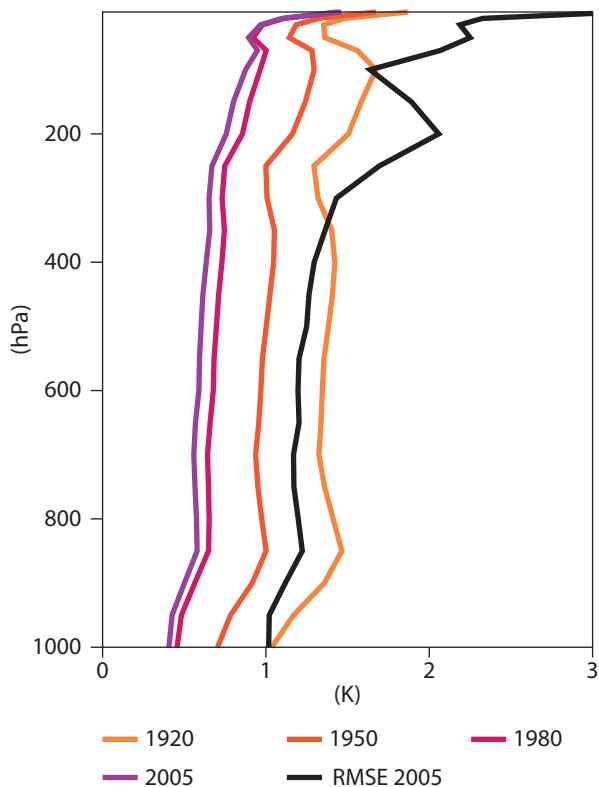
**Figure 6** Vertical profiles of the standard deviation of the CERA-20C ensemble for temperature over 60°S–60°N for the years 1920, 1950, 1980 and 2005. The root-mean-square error (RMSE) of the CERA-20C ensemble mean compared to ERA-Interim in 2005 has been plotted for comparison.

at ECMWF and Météo-France, which explicitly accounts for errors in the observational record and in the forecast model. The information from the ten members is used during the assimilation to compute a flow-dependent background error, which determines how to spread the information from observations in space.

The ensemble technique also aims to provide an indication of the confidence we can have in the data. The ensemble standard deviation is equal to 1.0°C at 1,000 hPa in 1920 and reduces to 0.4°C in 2005 (Figure 6). The ensemble spread in CERA-20C gradually decreases over time, which indicates that we can have more confidence in the data

as more observations become available. The greater observation density means that synoptic weather charts in CERA-20C are much better near the end of the century than at the beginning.

The analysis ensemble spread is supposed to represent the error of the analysis ensemble mean and should ideally be equal to the root-mean-square error (RMSE) of the ensemble mean compared to the true atmospheric state. This has been verified for the year 2005, when ERA-Interim provides a good proxy for the truth as it assimilates all types of observation at a higher resolution. Maps of the ensemble spread and of the RMSE show very similar horizontal structures, which means that the EDA correctly captures where the uncertainties are. However, the RMSE of CERA-20C is about twice as large as the CERA-20C analysis ensemble spread. This global offset between the spread and RMSE can be seen in the vertical profiles (Figure 6). CERA-20C is thus overly confident in the data compared to the actual error. It is important to note that this diagnostic has some limitations as it is based on a proxy of the truth. To improve the uncertainty estimations in CERA-20C, the size of the ensemble and the way the perturbations are generated in the different members when the EDA system is used for climate reanalysis will require further investigation.

## Access to the data and outlook

The multiple production streams have been consolidated excluding the spin-up years to produce the final released climate dataset. An automatic procedure has checked the data for continuity over time and has verified that not a single field is missing. The data server (*http://apps.ecmwf. int/datasets*) provides an interface similar to that of ERA-Interim. Users can select parameters and time periods of interest for download. For large retrievals, scripts are available to download data in batch mode. Users will be able to access the ensemble mean and spread to characterise uncertainties in their own applications.

As indicated in this article, ECMWF has already begun to analyse the data and to identify ways in which future 20th-century global reanalyses can be improved. When users elsewhere obtain the data and provide feedback or publish their findings, they will help improve the way future reanalyses are produced, thereby benefiting climate research and societal applications.

**FURTHER READING**

**de Boisséson**, **E.** & **M. Bamalseda**, 2016 : An ensemble of 20th century ocean reanalyses for providing ocean initial conditions for CERA-20C coupled streams. *ERA report series*, **24**.

**Desroziers**, **G.**, **L. Berre**, **B. Chapnik**, **P. Poli**, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Q.J.R. Meteorol. Soc.*, **133**, 3385–3396, doi:10.1256/qj.05.108.

**Hersbach**, **H.**, **D. Dee**, 2016: ERA5 reanalysis is in production. *ECMWF Newsletter* **No. 147**, 7.

**Laloyaux**, **P.**, **D. Dee**, 2015: CERA: A coupled data assimilation system for climate reanalysis. *ECMWF Newsletter* **No. 144**, 15–20.

**Laloyaux**, **P.**, **M. Balmaseda**, **D. Dee**, **K. Mogensen** & **P. Janssen**, 2016: A coupled data assimilation system for climate reanalysis. *Q. J. R. Meteorol. Soc.*, **142**, 65-78.

**Poli**, **P.**, **H. Hersbach**, **D. Dee**, **P. Berrisford**, **A. Simmons**, **F. Vitart**, **P. Laloyaux**, **D. Tan**, **C. Peubey**, **J.-N. Thépaut**, **Y. Trémolet**, **E. Holm**, **M. Bonavita**, **L. Isaksen** & **M. Fisher**, 2016: ERA-20C: An Atmospheric Reanalysis of the Twentieth Century. *Journal of Climate*, **29**, 4083–4097.