ECMWF Feature article

from Newsletter Number 144 – Summer 2015

METEOROLOGY

CERA: A coupled data assimilation system for climate reanalysis

cosmin4000/iStock/Thinkstock

www.ecmwf.int/en/about/news-centre/media-resources

doi:10.21957/0dx9o7bg

This article appeared in the Meteorology section of ECMWF Newsletter No. 144 – Summer 2015, pp. 15–20.

CERA: A coupled data assimilation system for climate reanalysis

Patrick Laloyaux, Dick Dee

Climate reanalysis combines information from past meteorological observations with modern forecast models, using data assimilation techniques originally developed for numerical weather prediction. The resulting reanalysis datasets provide a comprehensive and coherent record of Essential Climate Variables over an extended period of time.

ECMWF recently completed the production of ERA-20C, an atmospheric reanalysis spanning the entire 20th century, based on observations of surface pressure and marine winds (Newsletter 141, p. 9). There are now plans for a new 20th century reanalysis, called CERA-20C, in which ocean and atmospheric observations are assimilated simultaneously into a coupled atmosphere–ocean model.

Interest in coupled data assimilation is growing. Several research groups and national weather services have begun to use coupled models for forecasting and are developing different ways to initialise the forecasts with observations. The innovative approach developed at ECMWF for the purpose of climate reanalysis is at the cutting edge of data assimilation research. As described in this article, the CERA system has been found to lead to a better fit with observations, an improved use of near-surface measurements and smaller initialisation shock effects.

From ERA-20C to CERA-20C

ERA-20C is the first major reanalysis product of an ambitious collaborative research and development programme in climate reanalysis, led by ECMWF and involving many institutions in our Member States and elsewhere. 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.

The European Commission has provided substantial financial support for climate reanalysis in the form of two successive collaborative research grants, starting with ERA-CLIM in 2011 and followed by ERA-CLIM2 in 2014. An important new goal for ERA-CLIM2 is to develop a second 20th-century reanalysis that uses a fully coupled atmosphere–ocean model. The tentative name for this coupled climate reanalysis, which we hope to complete by the end of 2016, is CERA-20C.

Technical development of the CERA system began in 2013 at ECMWF as part of a two-year project on coupled atmosphere–ocean data assimilation funded by the European Space Agency (ESA). The system is built around the same coupled atmosphere–ocean model used for ECMWF's ensemble forecasts (ENS), i.e. the Integrated Forecasting System (IFS) (atmosphere, land surface, waves) coupled with NEMO (ocean, sea ice). However, the data assimilation scheme implemented in the CERA system is new and fundamentally changes the way observations are used.

Stronger coupling

In the current operational configuration of the ENS, observations for the atmospheric component of the model are assimilated separately from ocean observations. On the one hand, the IFS model is used to produce the analysis for the atmosphere, waves and land. This analysis is commonly referred to as 'weakly coupled' as the IFS model computes model-observation misfits for the three components but relies on separate analysis schemes to correct them. Information from ocean observations can only affect atmospheric state estimates after forward integration of the coupled forecast model. On the other hand, the ocean analysis based on the NEMO model is influenced by the atmospheric analysis, allowing for some interaction between the atmosphere and the ocean. Coupled forecasts are initialised by combining the two analyses even though they may not be fully consistent with each other.

The CERA system introduces a stronger coupling between ocean and atmosphere in the analysis step, so that 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. Introducing a coupled analysis step makes it possible to obtain more consistent fluxes at the atmosphere–ocean interface and potentially to make better use of near-surface observations and to reduce initialisation shocks in coupled forecasting.

 doi:10.21957/0dx9o7bg

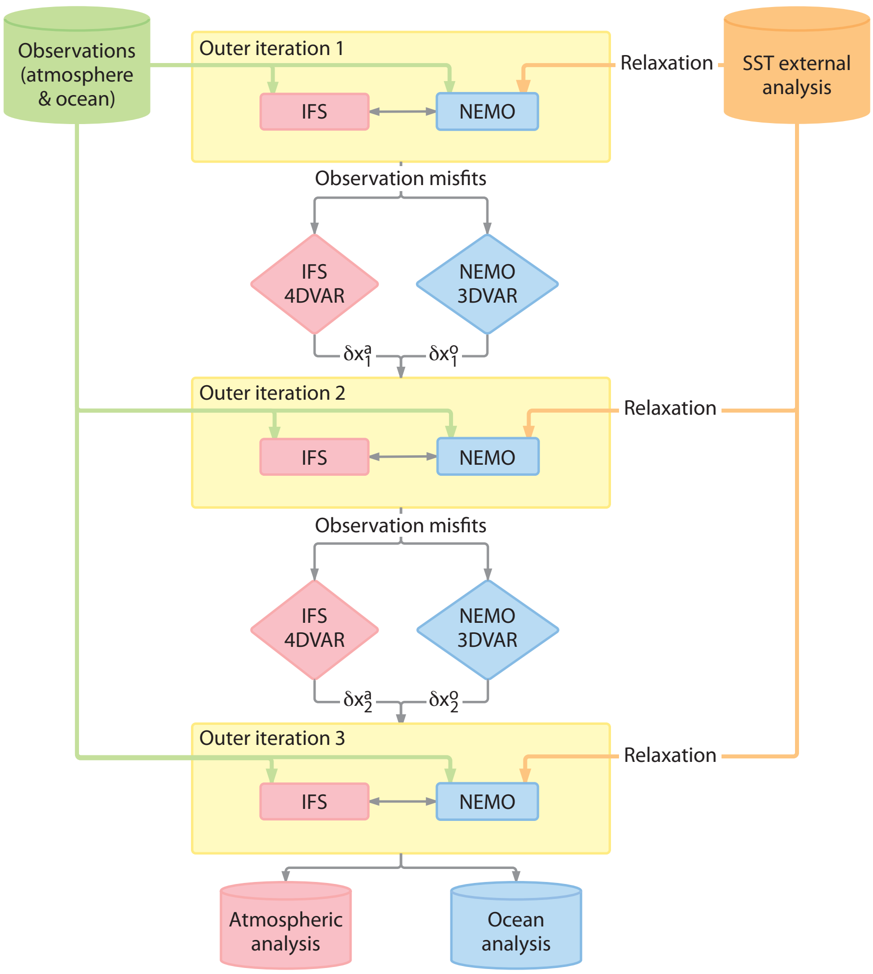

Figure 1 Schematic diagram of the CERA system.

The initial configuration of the CERA system uses a low-resolution version of the coupled model (T159L91 for IFS and ORCA1Z42 for NEMO). The computation of the ocean and atmospheric analyses is based on an incremental (iterative) variational approach with a common 24-hour assimilation window shared by the two components. Figure 1 represents one assimilation cycle of the CERA system, which consists of a pair of nested loops that run the coupled model and compute corrections to the initial coupled model state.

The CERA system uses two outer iterations, which is the default for the uncoupled IFS data assimilation system at low resolution. At the beginning of the first outer iteration, a 24-hour coupled model integration produces a first guess and computes the misfit with ocean and atmospheric observations (outer iteration 1 in Figure 1). The ocean and atmospheric increments are then produced separately by two different inner loops that solve a linearised version of the variational formulation. The computation of the atmospheric increment is based on a 4DVAR approach using the tangent linear and adjoint of a simplified version of the atmospheric model at lower horizontal resolution (T95L91). The ocean increment is computed by the NEMOVAR system with its 3DVAR FGAT configuration, which does not require tangent linear or adjoint models. Each method uses its own background error covariance model, which means that correlations between the ocean and the atmosphere are not explicitly represented. The production of the two increments is represented in Figure 1 by the top IFS-4DVAR and NEMO-3DVAR diamonds. A Direct Initialisation (DI) technique updates the current coupled estimate by adding the increments to the initial coupled state defined at the beginning of the assimilation window.

The second outer iteration starts by integrating the coupled model from the new initial condition (outer iteration 2 in Figure 1). During this second 24-hour integration, ocean fields are calculated partly on the basis of fluxes and wind stresses that have been affected by the atmospheric increment computed in the first outer iteration, while atmospheric fields are calculated partly on the basis of sea-surface temperatures (SST) and surface currents that have been altered by the ocean increment. The second atmospheric and ocean increments are computed using the new available observation misfits (bottom IFS-4DVAR and NEMO-3DVAR diamonds in Figure 1) and the initial condition is updated accordingly. Finally, the ocean

doi:10.21957/0dx9o7bg

and atmospheric analyses are produced by the last coupled model integration, which ensures the computation of a dynamically consistent coupled state (outer iteration 3 in Figure 1). Rather than assimilating SST observational data directly, the CERA system uses a gridded SST analysis product during the coupled model integrations to strongly constrain the upper level ocean temperature. In practice this is implemented by a nudging scheme with a timescale of two to three days.

There are two fundamental differences in the design of the CERA system compared to the uncoupled operational ECMWF data assimilation system. The first difference is the use of a coupled model in the variational method and in the forecast that carries the analysis forward in time. The second difference is the treatment of the atmospheric boundary condition, which evolves dynamically in the coupled model by receiving the SST fields computed in the NEMO model through the coupled framework.

Benefits of a coupled analysis

A comparative study has been carried out to assess the differences between the analyses produced by CERA on the one hand and by an uncoupled system (UNCPL) based on the uncoupled operational ECMWF scheme on the other. To isolate the effects of using the coupled system, UNCPL was run using the same resolution, model cycle and 24-hour assimilation window as CERA, and both systems assimilated the same ocean and atmospheric observational datasets with the same number of outer and inner iterations. Increments were applied by direct initialisation in both experiments. Finally, the SST relaxation performed in the NEMO model of the CERA and UNCPL systems used the same SST analysis product with the same relaxation coefficient. Comparing the behaviour of the two systems serves to highlight some potential benefits of the CERA system: producing a consistent coupled ocean–atmosphere analysis; making better use of near-surface measurements; and reducing initialisation shocks.

Better fit with observations

A first comparative study assesses the differences between the atmospheric analyses produced by the CERA and the UNCPL systems. As differences between both systems are expected to be located in the lowest layers of the troposphere above the ocean–atmosphere interface, the comparison focuses on heights of up to 700 hPa over sea. Figure 2 represents the spatial distribution of the conventional temperature observations assimilated by the CERA system between the Earth's surface and 700 hPa and located over sea or near coasts, for September 2010. The temperature observations are measured by aircraft taking off or landing from airports along coasts and by radiosondes launched from small islands or ships.

Figure 3 shows the background temperature root-mean-square error (RMSE) difference between the CERA and UNCPL systems with respect to the selected conventional temperature observations, for September 2010. Background RMSE is used as an indication of the quality of the analysis on which the background is based. The CERA system used about the same number of observations as the UNCPL system. The observations used to calculate the RMSE shown at 500 hPa include all the observations used above 500 hPa. The negative differences at 1,000 hPa mean that the CERA background RMSE is smaller at this level. The error bars provide an indication of the significance of the results. CERA's better performance could be a result of using a coupled model to compute the first-guess estimate and the background departure vectors. By contrast, the UNCPL system uses persisted SSTs to constrain its atmospheric surface boundary condition in the first-guess computation.

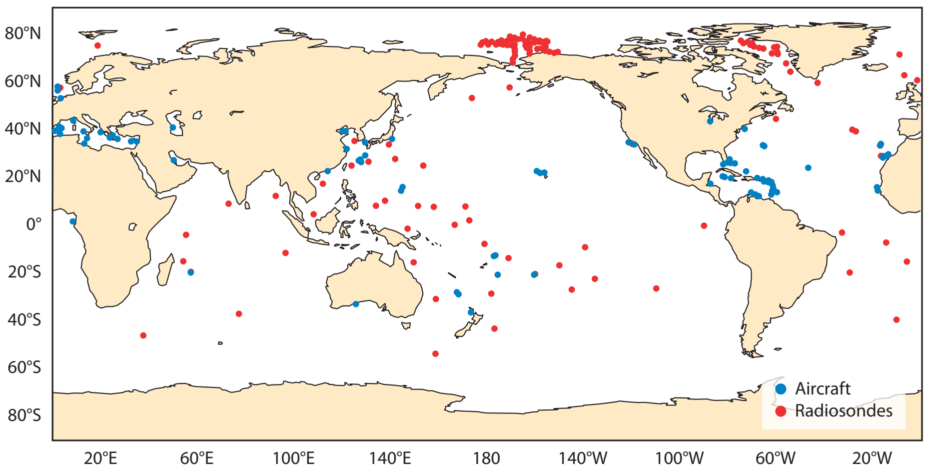

Figure 2 Location of conventional temperature observations by aircraft and radiosondes assimilated by the CERA system between the Earth's surface and 700 hPa over sea and near coasts in September 2010.

 doi:10.21957/0dx9o7bg

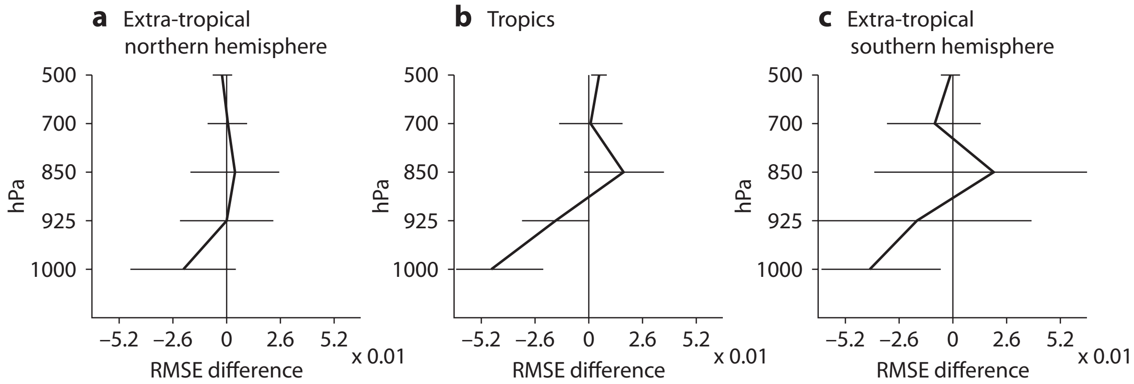

Figure 3 Vertical profiles of the background temperature RMSE difference between the CERA and the UNCPL systems with respect to the selected conventional temperature observations for September 2010. A negative difference means that the CERA background RMSE is smaller. The horizontal lines are error bars representing 95% confidence intervals.

Assimilation of near-surface observations

The CERA system is designed to make better use of near-surface measurements. This is because, in a coupled system such as CERA, any adjustment due to observations near the surface should have an impact on both atmospheric and oceanic variables. To assess the benefits greater coupling brings, we carried out Observing System Experiments (OSEs) in which the CERA and UNCPL systems were run over the period September–November 2013 with and without scatterometer near-surface wind measurements from ASCAT-A, ASCAT-B and OSCAT satellite instruments. In particular, we studied the impact of scatterometer measurements on the quality of the analysis during cyclone Phailin, which formed on 4 October 2013 over the Bay of Bengal and dissipated on 14 October 2013. Tropical cyclones are coupled phenomena with strong interactions between atmospheric wind and ocean temperature and the effect of scatterometer assimilation can be expected to be accentuated during such severe weather events.

The impact of scatterometer data on the analysis of ocean temperature was assessed with respect to conventional observations measured by one Argo buoy. The Argo observing system consists of a fleet of approximately 3,700 drifting probes deployed worldwide. In most cases probes drift at a depth of 1,000 metres and, every 10 days, by changing their buoyancy, dive to a depth of 2,000 metres and then move to the surface while measuring salinity and temperature profiles. The Indian National Centre for Ocean Information Services (INCOIS) operates several Argo probes in the Bay of Bengal. It has initiated a project to monitor upcoming severe weather conditions and to achieve a higher temporal resolution of temperature and salinity profiles for the upper ocean.

Figure 4 shows the observations at a depth of 40 metres from the Argo float 2901335 operated by INCOIS during the passage of cyclone Phailin. As this float was located on the track the cyclone was forecast to take, the probe setup was changed by a satellite transmission on 9 October to measure profiles approximately every 3 hours between the surface and a depth of 300 metres. This configuration was kept in place until 15 October. The strong winds produced over the ocean by the tropical cyclone led to cold water from the deep ocean rising to the surface. This generated a negative SST anomaly in the tropical cyclone's wake. The observed cold wake appears on 11 October with a drop in temperature by 3°C. Figure 4 shows the temperature analyses produced by the CERA and UNCPL systems at a depth of 40 metres with and without input from scatterometer observations.

Figure 4a shows that, in the CERA system, the assimilation of scatterometer data (CERA-SCATT) led to a consistent improvement in the temperature estimate, by up to 0.4°C, compared to the data assimilation without scatterometer data (CERA-NOSCATT), producing an analysis closer to observations. This illustrates that the use of the coupled model in the assimilation process can produce dynamical ocean and atmospheric feedbacks during the assimilation process. The wind observations using scatterometers affected the analysed state of the ocean in a way that was consistent with the phenomenon of a cold wake after the passage of a cyclone.

However, the analysis produced by the CERA-SCATT experiment and the Argo observations are far from a perfect match, with differences of up to 2°C. This can partly be explained by the coarse resolution of the ocean model, which has only 42 vertical levels and in which the thickness of each layer near the surface is around 10 metres. In addition, the OSTIA product used to constrain the SST analysis with a relaxation scheme is a daily-mean product. It produces the plateaus in the ocean analysis time series seen in Figure 4, which explains why the analysis does not follow the frequent fluctuations seen in the observations.

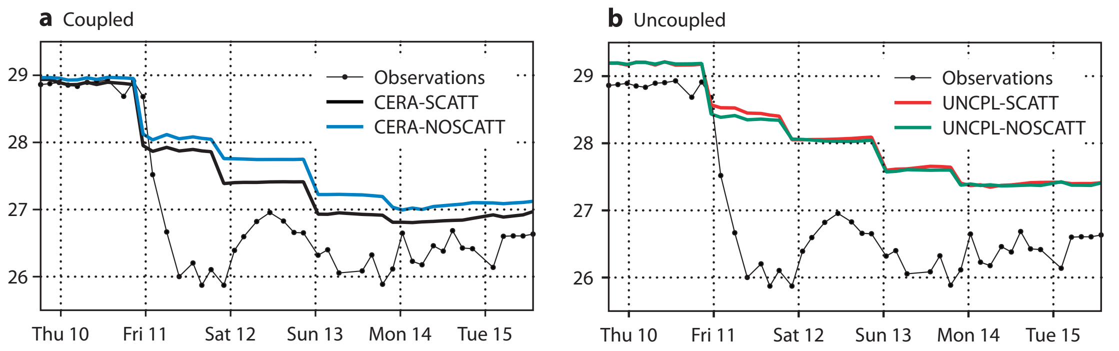

Figure 4 Time series of ocean temperature observations at a depth of 40 metres by the Argo float 2901335 (Observations) with (a) the temperature analyses produced by the CERA system with scatterometer data assimilation (CERA-SCATT) and without scatterometer data assimilation (CERA-NOSCATT); and (b) the temperature analyses produced by the UNCPL system with scatterometer data assimilation (UNCPL-SCATT) and without scatterometer data assimilation (UNCPL-NOSCATT).

In the UNCPL system (Figure 4b), the time series of the analysis with scatterometer observations (UNCPL-SCATT) is very similar to that without scatterometer observations (UNCPL-NOSCATT), showing that scatterometer assimilation had no significant impact on the ocean temperature analysis during the cyclone. A plausible explanation lies in the weaker interaction between atmosphere and ocean in the UNCPL system, where forcings come from instantaneous fields retrieved every 6 hours and accumulated fields over 24 hours, compared to the one-hour coupling frequency in the coupled approach.

The comparison of the analyses produced by CERA-SCATT and UNCPL-SCATT illustrates the better performance of the coupled assimilation system during the cyclone. The improvement appears to stem from the better use of scatterometer data as well as a smaller analysis departure before the cyclonic event.

Initialisation shocks

A major challenge of coupled ocean–atmosphere forecasting lies in the initialisation, which aims to incorporate information from ocean and atmospheric observations into the model components in an optimal manner. The initialisation method, particularly for relatively short-range coupled forecasts, should ensure that the ocean and atmospheric model components are consistent with one another at the beginning of the forecast, in order to avoid the generation of initialisation shocks. These shocks are imbalances in the vertical fluxes of heat, momentum or freshwater between the atmosphere and ocean initial states. They can occur when there is insufficient communication between the two model components during the calculation of the initial conditions.

Initialisation shock effects must not be confused with forecast errors due to model biases. The shocks that are discussed here are deviations of the forecast from observations that can demonstrably be reduced or eliminated through changes to the initialisation procedure.

To compare the role of initialisation shocks in CERA and UNCPL, two sets of 30 medium-range coupled forecasts spread over April–May 2008, December–January 2008/9 and August–September 2010 were run initialised by the CERA and the UNCPL analyses, respectively. These CERA and UNCPL forecast sets were run with the same coupled model (versions and resolutions) as used in the computation of the CERA and UNCPL analyses to avoid any effects on the forecast caused by using different models. The left-hand panel of Figure 5 shows the RMSE of CERA's 12-hour air temperature forecast at 1,000 hPa relative to the CERA analysis and averaged over all forecast start dates. Land areas are masked out, as the focus is on atmosphere–ocean imbalances. The errors present in the CERA forecasts are the result of biases in the models as well as any imperfections in the CERA initialisation method. The CERA forecasts are taken as a baseline case in the sense that any larger deviations of UNCPL forecasts from their reference analyses should represent a shock imparted by an initialisation procedure that differs from the CERA methodology.

The right-hand panel of Figure 5 represents the RMSE of UNCPL's 12-hour air temperature forecast at 1,000 hPa relative to the UNCPL analysis and averaged over all forecast start dates. Contours show 0.15°C differences between the RMSEs of CERA and UNCPL forecasts, with blue (green) contours marking increased (decreased) RMSE in UNCPL forecasts. UNCPL forecasts show small increases in RMSE in several areas. These are generally areas in which the difference between the SST in the ocean and atmospheric components of the UNCPL analysis is large.

 doi:10.21957/0dx9o7bg

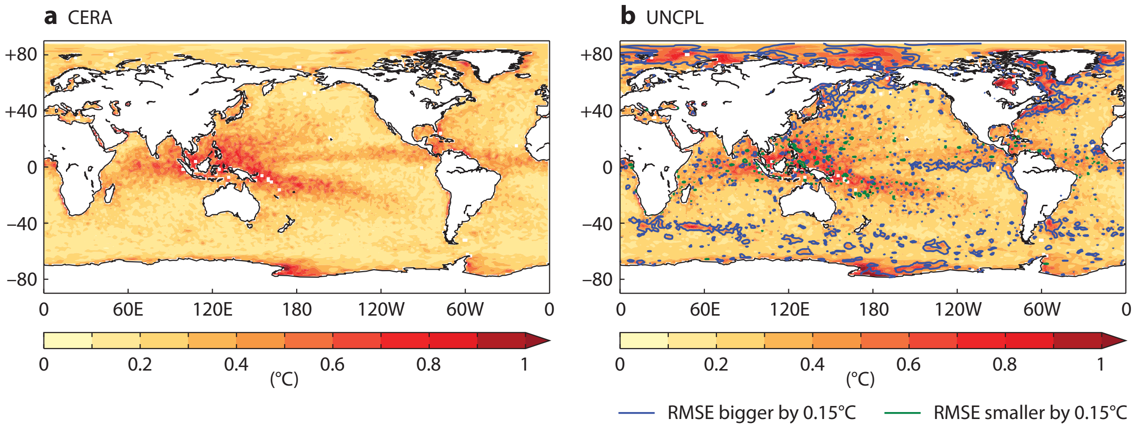

Figure 5 Average RMSE of 30 1,000 hPa temperature forecasts at 12-hour lead times, spread over April–May 2008, December–January 2008/9 and August–September 2010, for (a) CERA and (b) UNCPL , evaluated against their own corresponding analysis. Contours in (b) show 0.15°C RMSE differences between CERA and UNCPL forecasts.

This air temperature shock signal therefore appears to develop primarily due to the change in SST forcing felt by the atmosphere after the transition from the analysis, where the SST is prescribed by the OSTIA product, to the forecast phase, where the SST comes from the ocean component. These air temperature shocks are generally of magnitude 0.2°C or less, but compared to the small baseline RMSE seen in most areas in CERA, they represent substantial error amplifications: the RMSE is increased by 50% or more in the eastern equatorial Pacific, eastern tropical Atlantic, northern Pacific and across most of the Southern Ocean, and it is more than doubled in the Gulf Stream and Arctic regions.

Figure 6 shows the evolution of the RMSE of CERA and UNCPL air temperature forecasts at 1,000 hPa against their own analyses, averaged over the NINO3 region and over all forecast start dates. The larger errors in UNCPL forecasts compared to CERA result from the initial SST discrepancies between the ocean and atmospheric components. The effects of the shock are felt out to at least 10 days' lead time, showing that initialisation shocks can have an impact on short- and medium-range forecasts.

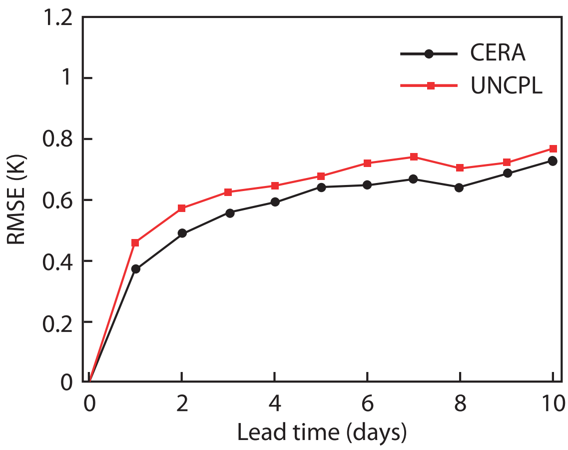

Figure 6 RMSE of thirty 1,000 hPa temperature forecasts spread over April–May 2008, December–January 2008/9 and August–September 2010 and averaged over the NINO3 region, for (a) CERA and (b) UNCPL, evaluated against their own corresponding analysis.

Future plans

The CERA system is now part of the ERA-CLIM2 project, which focuses on the production and assessment of multi-decadal reanalyses of the Earth's climate. The production of an extended climate reanalysis of the 20th century (CERA-20C) will be launched this summer. The use of a coupled model ensures spatial and temporal consistency in the production of long-term climate records.

The CERA system will constrain the coupled model by assimilating only conventional surface pressure and wind observations in the atmosphere, as well as salinity and temperature profiles in the ocean. The air–sea interface will be constrained by an SST relaxation towards the HadISST2 monthly analysis product. Ensemble techniques will be used to compute flow-dependent background error covariances and support uncertainty assessments. All observations used and associated quality feedback information will be provided to users via the ERA-CLIM Observation Feedback Facility. All reanalysis data products will be made available via ECMWF data servers.

The CERA system, originally designed for climate reanalysis, could pave the way for more advanced data assimilation used in weather forecasting. Performance of the coupled system at higher resolution, representation of the diurnal cycle at the air–sea interface and various other technical and scientific challenges will have to be addressed to test this possibility.

 doi:10.21957/0dx9o7bg