ERA report series

14 The data assimilation system and initial performance evaluation of the ECMWF pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C)

Paul Poli, Hans Hersbach, David Tan, Dick Dee, Jean-Noël Thépaut, Adrian Simmons, Carole Peubey, Patrick Laloyaux, Takuya Komori, Paul Berrisford, Rossana Dragani, Yannick Trémolet, Elias Hólm, Massimo Bonavita, Lars Isaksen and Mike Fisher

Series: ERA Report Series

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications/

Contact: library@ecmwf.int

European Centre for Medium-Range Weather Forecasts
Shinfield Park, Reading, RG2 9AX, England

Abstract

Within the ERA-CLIM project, ECMWF is producing a pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C). This pilot reanalysis is part of a suite of experiments and complements a model-only integration (ERA-20CM) and a land-surface reanalysis (ERA-20CL). The intention of ERA-20C is not to produce a final 'best-product' state-of-the-art climate dataset. The prime target is the study of the feasibility of reanalysing the century using new data assimilation methods to tackle the problem of observing system changes, assimilating here only surface observations. The data assimilation system in ERA-20C is an Ensemble of Data Assimilations (EDA) of 10 members. The 10 members are forced by HadISST2.1.0.0 ensemble of sea-surface temperature and sea-ice conditions. Each ERA-20C member employs a 24-hour four-dimensional variational (4D-Var) analysis scheme. The input observations are provided by ISPD 3.2.6 and ICOADS 2.5.1. The surface pressure observations are bias-corrected using a variational bias correction within the assimilation. The wind observations above surface oceans are assimilated as we verified that their assimilation improves the quality of the atmospheric circulation representation, especially in the Tropics.

The global background error covariances employed by the 4D-Var are updated automatically every 10 days, using a 90-day sample drawn from the ensemble. The local background errors in vorticity are modulated daily, from the ensemble spread, to represent the error of the day. Over time, the background error horizontal correlations become more narrow and the background error variances become smaller. This is in line with expectations that a loose observing network can only help make large-scale adjustments to an analysis, while a more dense observing network can help capture smaller-scale errors. Between 1900 and 2000, the ensemble surface pressure spread drops from about 1 hPa in well-observed areas (e.g. Europe) and 15 hPa in unobserved areas (e.g. Southern Oceans) to 0.5–5 hPa or less. However, the update in the vertical structures of background errors brings the unintended consequence that the mean temperature analysis increments vary in strength over time with height. This introduces unwanted spurious trends in upper-air temperatures. Also, the assimilation diagnostics suggest that the representation of atmospheric tides is affected by a suboptimal model time-step, which could not be reduced owing to computational costs.

The time variation of the Root Mean Square (RMS) of observation minus background statistics fits reasonably well that of expected departures, which can be computed from the ensemble spread and assumed observation errors. This agreement could be further improved by revising the observation error estimates used in the assimilation. We propose here, for future reanalyses, time-varying observation error estimates for surface pressure and wind, separated by platform type.

In two case studies of extreme events, the European Great Storm of 1987 and the U.S. Great Blizzard of 1899, the meteorological analyses of ERA-20C match those of other sources. The ERA-20C consecutive production streams are found to match in the Northern Hemisphere and Tropics within ± 0.3 hPa in mean sea level pressure (and 1 K in lower tropospheric temperature) in 1980 and 2000, and within ± 1 hPa (and 2.5 K in lower tropospheric temperature) in 1920 and 1940. For upper levels 100–200 hPa the consecutive streams are separated by 2.5–5 K or more, indicating as expected that surface-only observations do not provide sufficient constraints for a realistic product at such high levels and above (stratosphere), and those products should thus be used with caution.

From the elements presented in this report, the long-term climate trends in ERA-20C are incorrect in most places except probably near the surface in the Northern Hemisphere extratropics. However, there are several indications that the representation of meteorological events on a daily basis, including extremes, may be of reasonable quality.

The report concludes with a list of major issues to be addressed in an upcoming rerun of the ERA-20C control experiment.

Contents

1 Introduction

The European reanalysis of global climate observations collaborative project (ERA-CLIM) aims to build a global reanalysis system capable of assimilating observations of all kinds covering the 20th century up to present times. A prime difference with the ERA-Interim reanalysis system (Dee *et al.*, 2011a) is the target time period, which comprises more than the well-observed years since 1979, when in situ and satellite observations are plentiful compared to earlier times. With more than 14,000 users of ERA-Interim registered at ECMWF as of July 2013, ERA-Interim serves applications in many fields. A century-long dataset with recent years of quality similar to that of ERA-Interim would probably meet even more needs, and see new applications being developed in other areas. However, the difficulty in creating such a dataset is that the observation coverage improves by orders of magnitude during the 20th century, while the model and assimilation methodology used in ERA-Interim are inapplicable to earlier years and a poorly observed system without significant developments.

In order to devise a reanalysis system capable of reanalysing the whole century, the ERA-CLIM project plan lays out a series of pilot atmospheric reanalyses. They assimilate atmospheric observations of increasing diversity, from the lowest rank to the highest: 1) from no observations, 2) to surface-only observations, 3) followed by surface and upper-air observations, and 4) finally surface, upper-air, and satellite observations. The lowest rank reanalysis is essentially a model-only integration, constrained only by forcing. It involves upgrading the ECMWF model to accept forcing as specified for example by the Coupled Model Intercomparison Project Phase 5, CMIP5 (World Climate Research Programme, 2011). The second-rank reanalysis is called 'Pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C)'. It serves as a stepping stone, and requires developments in the data assimilation methodology. The difficulty in ERA-20C is to capture the changing observing system quality over time, and make optimal use of observations as they become available in greater numbers.

Given the ERA-CLIM context, the scientific intentions for the ERA-20C dataset are modest. For example, the low horizontal resolution and the deliberate limitation of the observation input to surface only reminds that further, more complete, reanalyses are to come. The intention of ERA-20C is not to create a long-standing product to serve the research community as a 'best product' would do. The ERA-20C dataset, a byproduct of the ERA-20C reanalysis system, shall help assess the performance of a given data assimilation approach for creating a century-long reanalysis of climate and weather, given the observation changes over time.

Another intended application of ERA-20C lies in the serial concept of reanalyses in ERA-CLIM. A reanalysis of low rank can be independently compared with observations that are normally assimilated in a reanalysis of higher rank, and validated against them. For example, the upper-air products of the surface-only reanalysis can be compared with upper-air observations. In this concept, a reanalysis dataset can also serve as a reference for homogenisation work on withheld observations, applying concepts such as introduced by, *e.g.*, Haimberger (2007). The result of said homogenisation can then help improve error characterisation of observations, such as break detection in observation timeseries. The lessons learnt can then benefit reanalyses of higher rank, which assimilate such observations.

On several accounts, the ERA-20C resembles the first four-dimensionally complete multi-variate historical reconstruction of the global weather of the 20th century produced by Compo *et al.* (2011): the 20th Century Reanalysis (20CR). One similarity is that the ERA-20C dataset shall represent the evolution of the global climate over the 20th century, along with estimates of uncertainty. However, without upper-air or satellite observations, the expected accuracy for recent times shall fall short of a reanalysis that assimilates surface and upper-air observations, as well as satellite observations. Compo *et al.* (2006) and Whitaker *et al.* (2009) already evaluated the usefulness and feasibility of a surface-pressure only data assimilation system.

It is still useful to remind the expected accuracy of a Numerical Weather Prediction (NWP) system in such

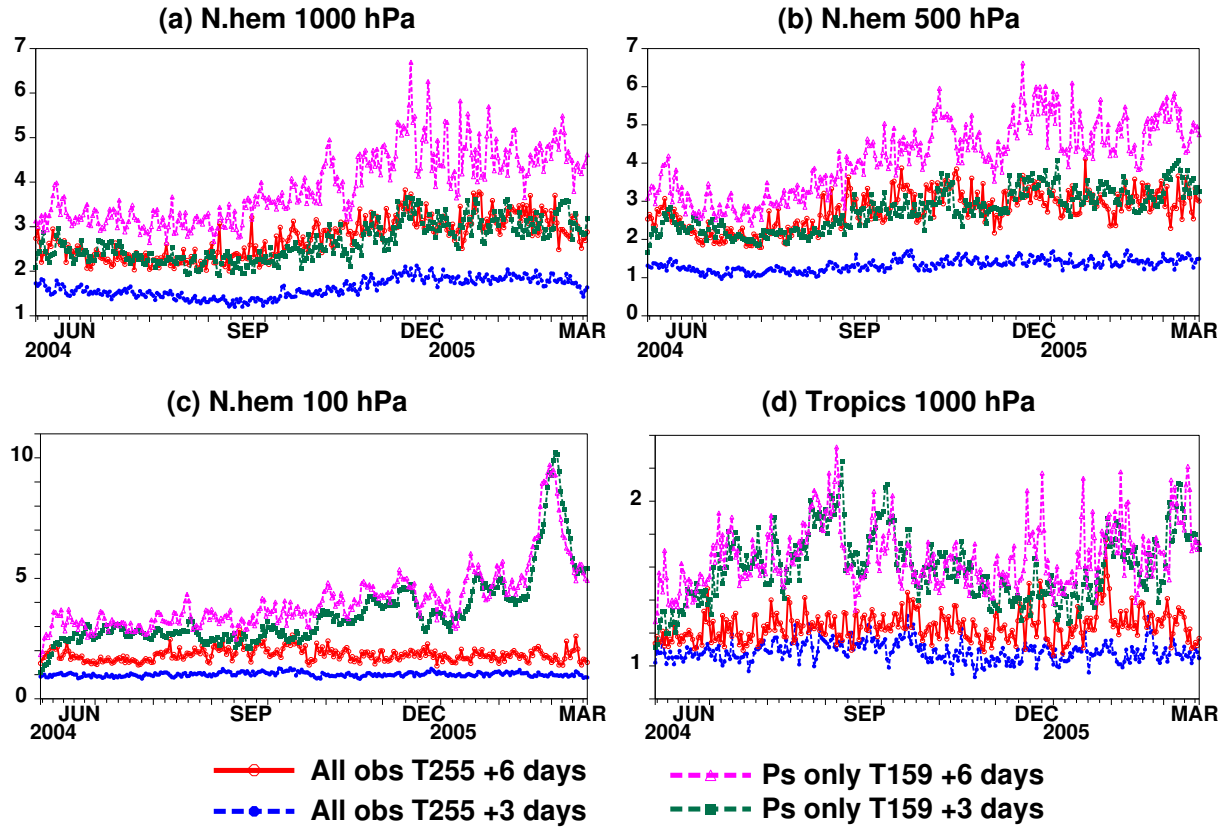

Figure 1: Timeseries of root mean square error of temperature forecasts (in K), as a function of time, for an experiment 'All obs' assimilating surface, upper-air, and satellite observations as in ERA-Interim, and an experiment 'Ps only' assimilating only surface pressure observations. All forecasts are verified against the operational analyses

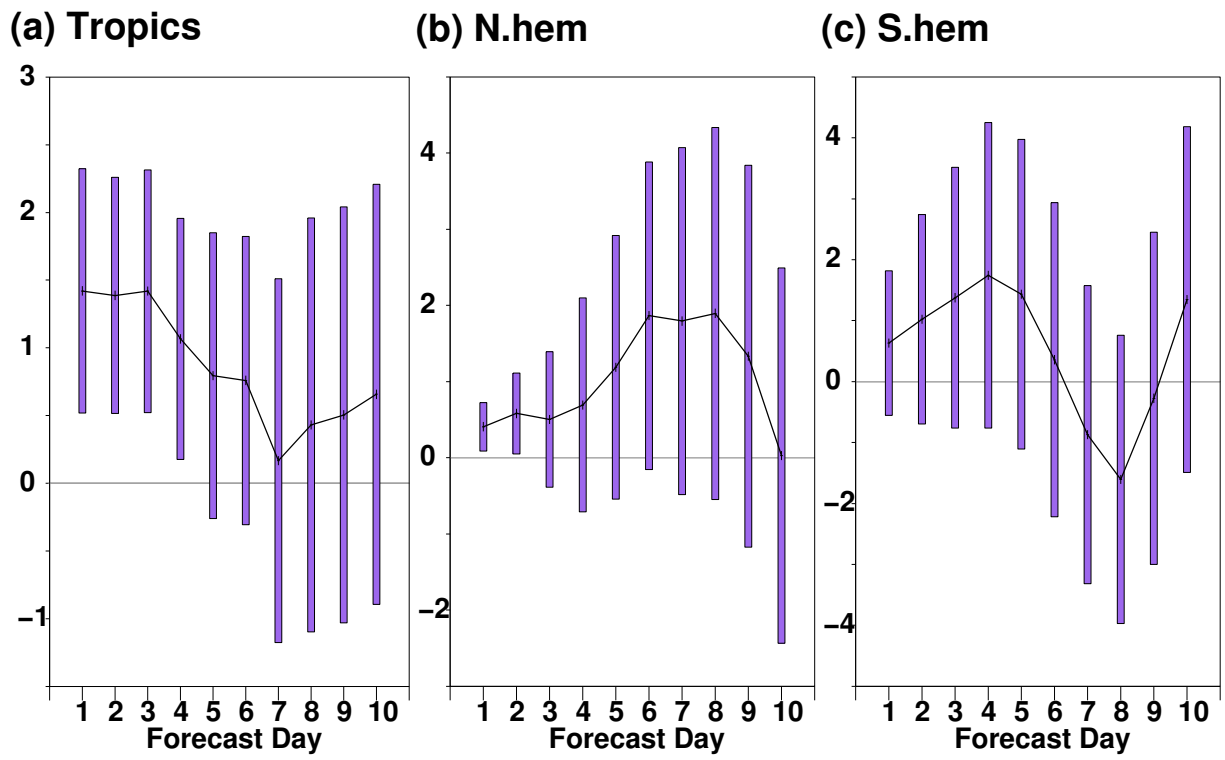

Figure 2: Differences between 500 hPa geopotential height forecast anomaly correlation (in percents) between an experiment that assimilates surface wind observations from ship and buoys and an experiment that does not. Positive (negative) differences indicate a positive (negative) impact of the wind assimilation. The error bars indicate the confidence intervals at 95% statistical significance. Time period is 2 June 2004–10 April 2005 (313 forecasts).

conditions. The baseline experiment for this discussion resembles ERA-Interim, in that it assimilates all available observations, including conventional surface and upper-air, as well as satellite observations, at a horizontal truncation T255 or approximately 80 km horizontal resolution. Over the Northern Hemisphere extratropics (latitudes 20°N–90°N), Figure 1(a) shows that this baseline can issue temperature forecasts whose root mean square error (RMSE) at day 6 is greater than at day 3, illustrating the loss of predictability over time. A second experiment, initialised from the same conditions as the baseline, assimilates, at a lower resolution (horizontal truncation T159), surface pressure observations only. Compared to the baseline, it shows degraded forecast scores at day 3. However, its 3-day forecasts are still better than its 6-day forecasts. This indicates that a data assimilation system with only surface pressure observations retains some predictive skill in the Northern Hemisphere extratropics near the surface at day 3, at about the same accuracy level as a fully observed system would do for day 6. A similar qualitative conclusion is found over the Southern Hemisphere extratropics (not shown). At 500 hPa, Figure 1(b) indicates that the surface pressure-only experiment retains predictive skills. However, at 100 hPa, *i.e.*, above the tropopause, Figure 1(c) shows no predictive skills for the surface pressure-only experiment as 3-day and 6-day forecasts show the same level of error. Of importance, the figure also shows that the surface pressure-only experiment fails to properly capture the sudden stratospheric warming in February 2005. In the Tropics, Figure 1(d) confirms the expectation that the absence of geostrophic balance leads the surface pressure-only experiment to have little predictive skill in this region. From this simple demonstration, a surface pressure-only reanalysis (a) may reproduce correctly the daily evolution of tropospheric meteorology in the mid- and high-latitudes, but (b) may be limited in representing realistically the daily evolutions in the tropical troposphere and the stratosphere without additional observational constraints.

The observational constraints in the Tropics can be improved by considering also atmospheric wind observa-

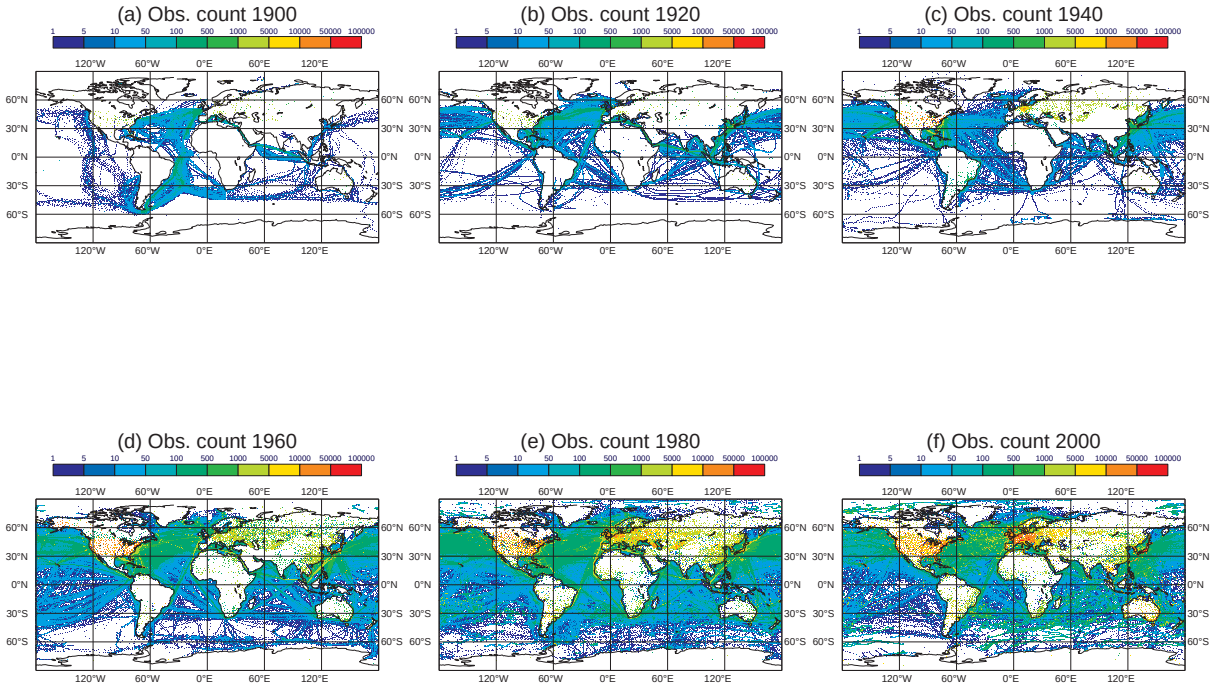

Figure 3: Maps of surface pressure observation count (combining ISPD 3.2.6 and ICOADS 2.5.1), for selected years, in 1° latitude × 1° longitude bins

tions near the surface above oceans. To demonstrate this, the surface pressure observations-only experiment shown in Figure 1 is repeated with the addition of surface wind observations over oceans. The wind observations over land and not assimilated because of representativeness issues (same practice as ECMWF operations). The number of additional observations is very small: about 1,300 (400 and 120) daily wind reports in the Northern Hemisphere extratropics (respectively: the Tropics and the Southern Hemisphere extratropics), compared to 42,000 (5,000 and 5,000) daily surface pressure observations in the Northern Hemisphere extratropics (respectively: the Tropics and the Southern Hemisphere extratropics). Comparing the two experiments, Figure 2(a) suggests a positive impact of the surface wind observations on the 500 hPa geopotential height in the Tropics. This impact is significant at 95% throughout most of the forecast range. This indicates the benefits of the assimilation of atmospheric winds above oceans on the representation of the tropical circulation. Furthermore, a positive impact is also found in the first few days of forecasts in the Northern and Southern Hemispheres extratropics (Figures 2(b),(c)). Note, in the Southern Hemisphere extratropics, results are insignificant due to the small number of additional observations (on average 120 wind reports per day). As a consequence it was decided that ERA-20C should assimilate also near-surface atmospheric winds from oceanic surface reports.

The ERA-20C reanalysis uses observations provided by the International Surface Pressure Databank (ISPD) version 3.2.6 (Yin *et al.*, 2008) and the International Comprehensive Ocean Atmosphere Data Set (ICOADS) version 2.5.1 (Woodruff *et al.*, 2011). Note that ISPD incorporates tropical cyclone best track data from the International Best Track Archive for Climate Stewardship, IBTrACS (Knapp *et al.*, 2010). As expected from the observing system changes, Compo *et al.* (2011) demonstrated that the reanalysis quality improves over time with greater number of observations in ISPD. In ERA-20C, Figure 3 shows, for selected years, the coverage in surface pressure observations for several years. In the early part of the century, the pressure observations, especially over land, are mostly located in the Northern Hemisphere. The Tropics are fairly well observed thanks

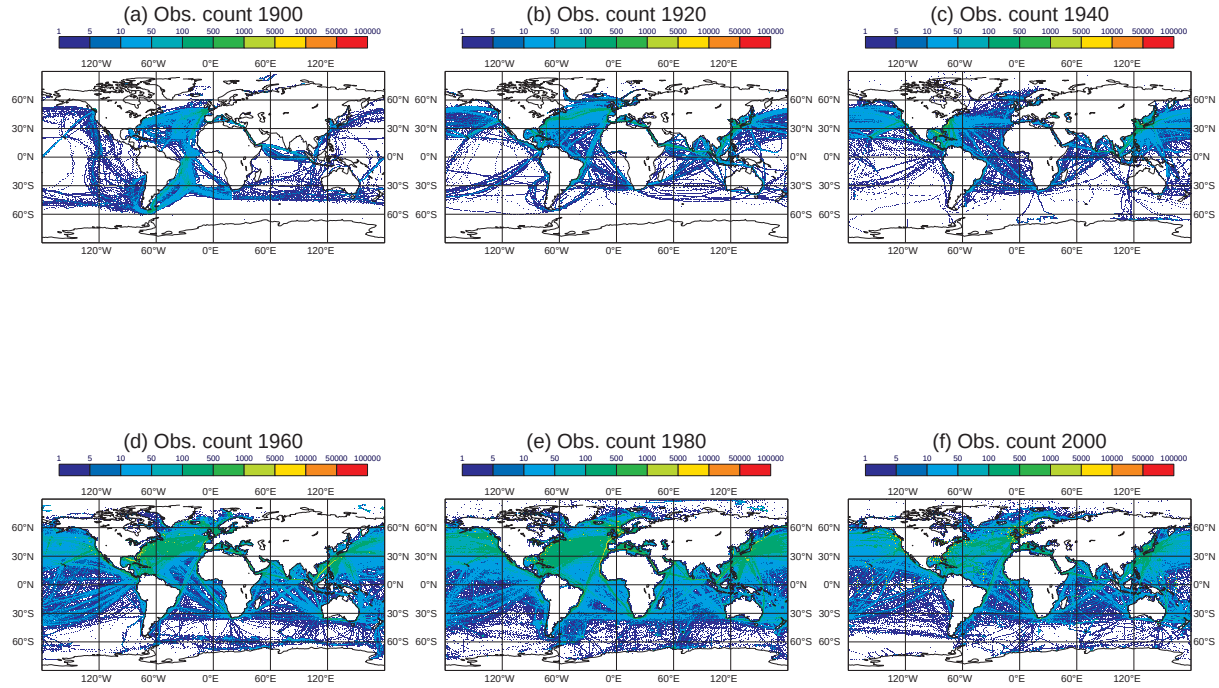

Figure 4: Maps of surface wind vector observation count (from ICOADS 2.5.1), for selected years, in 1° latitude × 1° longitude bins

to shipping routes, although the absence of geostrophic balance at the Equator means that the pressure observations have limited impact in that region on reanalysis quality, as shown earlier. The Pacific and the Southern Oceans remain for the most part unobserved for years. Even in recent times, several remote continental areas remain still unobserved (*e.g.*, Sahara, Amazon region) by in situ pressure sensors.

In terms of wind observations, we remind that we only use observations over ocean. This is convenient as there is yet no equivalent database to ISPD or ICOADS which would include wind observations over land. Figure 4 shows that the ocean coverage varies with human activities throughout the century. Ship traffic to the polar regions in early years point to whaling campaigns, while the hubs of the Suez and Panama canals indicate prominent shipping routes. Figure 4 illustrates somehow shifts in global balance of power throughout the 20th century. In the year 1900, Europe appears as the main hub of all shipping routes. Around World War II, America emerges as the hub of most shipping routes. At the beginning of the 21st century, South-East Asia appears as an important hub, well covered by observations over oceans. New, Arctic, shipping routes appear around the years 1980s. In parallel, the Antarctic, yet an unexplored continent in the year 1900, is routinely approached by ships in the year 2000, suggesting a growing appetite from human activity in this region of the world. This view of prime interest for the Historian, may be considered as a side interest to the Meteorologist. It is yet a prime concern to the Climatologist. In a nutshell, our knowledge of the past improves over time, and covers primarily affluent regions.

The time variations in observation quantity are repeatedly portrayed as fundamentally limiting reanalyses from drawing reliable long-term trends (*e.g.*, Thorne and Vose, 2010). However, the Figures 3 and 4 suggest that any initiative to analyse past observations is faced with the same problem, regardless of the method (*e.g.*, simple climatological report averaging, or more complex approach). By going back to the individual, original observations, reanalysis exposes itself more to the problem of time-varying observation quantity and quality, and

heterogeneity in space. At the same time, reanalysis also gives a framework for treating the problem properly (Dee *et al.*, 2011b), by accounting explicitely for uncertainties. To this end, data assimilation theory offers several algorithmic solutions. Compo *et al.* (2006) first studied the feasibility of reanalysing the weather of the 20th century using only surface pressure observations and boundary sea-surface conditions. Whitaker *et al.* (2009) compared for this application the performance of a four-dimensional variational (4D-Var) assimilation scheme and an ensemble-based assimilation scheme. They concluded on the superiority of both schemes over a three-dimensional variational (3D-Var) assimilation scheme. However, there is a major difference between the ensemble-based and 4D-Var-based approaches in the treatment of the time-varying quality, called background error by data assimilators. In the ensemble-based data assimilation system, the time-varying quality is accounted for by the ensemble spread, which evolves with the observation coverage. However, as noted as one of the key points in the comparison conducted by Whitaker *et al.*(2009), a 4D-Var analysis system does not produce on its own an estimate of analysis uncertainty. This means that the uncertainty in the background (starting point to the next assimilation) cannot be updated by the analysis. Consequently, a 4D-Var system such as routinely employed at ECMWF for NWP operations requires a prior representation of background error covariances. While variances can be adjusted on a daily basis using an ensemble of 4D-Var solutions (Raynaud *et al.*, 2009), the full covariance estimation is only occasionally updated, using dedicated, off-line, procedures. Consequently, developments had to be made for ERA-20C to achieve a self-updating time-varying evolution of the background error covariances. Also of interest for future NWP improvements at ECMWF, ERA-20C offers a testbed, in a poorly observed data assimilation system, for this improved concept of self-updating background errors.

This paper describes the data assimilation components in ERA-20C, along with an initial view on their impact on the ERA-20C production started on 18 December 2012 (Dee, 2013) and recommendations for fixing the problems found.

The outline of this document is as follows. **Section 2** summarises the major components of the ERA-20C reanalysis system and production. **Section 3** presents the observation quality control steps, observation errors assumed in ERA-20C, and suggested refinements in such estimates to be used in future reanalyses. **Section 4** details how the background errors are derived and evolved in ERA-20C, and what diagnostics are available in the output datasets as proxy for the reanalysis uncertainty estimate. **Section 5** is a preliminary evaluation of the assimilation performance, including forecast scores, observation fits, a few special cases of meteorological interest, the quality of stream boundaries, comparison with ERA-Interim, and initial assessment of trend quality. **Section 6** presents conclusions and issues to be fixed in an upcoming rerun of the ERA-20C control experiment.

2 Brief description of the ERA-20C reanalysis system components and production

2.1 Input datasets

The time-varying input consists of observations and sea-surface boundary conditions. The observations are provided by two datasets. The ISPD 3.2.6 dataset provides atmospheric surface pressure observations. The ICOADS 2.5.1 dataset also provides atmospheric surface pressure observations, though only above oceans, but includes in addition other geophysical parameters above oceans, such as atmospheric and oceanic temperatures, and atmospheric near-surface winds. After acquisition from original sources, these observations are archived in the ECMWF Observation Feedback Archive, OFA (Kuchta, 2009). In the ERA-20C production, the observations are extracted from the OFA and converted to the format expected by the assimilation, the so-called Observation Data Base, ODB (Saarinen, 2004).

To help with future validation work, conventional observations in BUFR format from the ERA-40 and operational archives are also used in input. Note, these observations are blacklisted to prevent using them in the assimilation.

The sea-surface boundary conditions are provided by the HadISST2.1.0.0 dataset (Rayner *et al.*, *in preparation*). It is an ensemble of 10 reconstructions of sea-surface temperature and ice conditions between the years 1899 and 2010.

2.2 Output datasets: products

The time-varying output includes two types of products. The first product consists of fields that describe the four-dimensional structure of the atmosphere, the land surface, including limited information about sub-surface upper layers, and the ocean waves, for various geophysical parameters. In ERA-20C, the output fields are archived every 3 hours, at the full horizontal and vertical resolutions. This represents about 200 thousand fields for every day of reanalysis.

The second product is the so-called observation feedback, stored on the OFA. It contains the observations used as input, as well as quantitative information added by the assimilation, such as quality control flags (usage indication), departures before and after assimilation, and bias correction when relevant, for each individual observation.

2.3 Model components

The atmosphere, ocean wave, and land surface model configurations are as used by ECMWF operations as of December 2012, described in the documentation of IFS version cycle 38R1 (ECMWF, 2013), with differences described hereafter. Modifications were made to the interfaces to allow additional forcing input intended to help realistically integrate the model throughout the century (Hersbach *et al.*, *in preparation*). The vertical resolution is identical to ECMWF Operations as of 2012, with 91 vertical model levels between the surface and 0.01 hPa or about 80 km altitude. The horizontal resolution is reduced to make the production computationally affordable, from spectral triangular truncation T1279 or approximately 16 km in ECMWF operations to T159 or approximately 125 km in ERA-20C. The model and radiation time-steps are both 1 hour. The hourly radiation calls are found to improve the representation of atmospheric tides. Early experimentation also suggested that the 1 hour time-step poses problems to properly represent atmospheric tides; however, the proper solution of doubling the model time-step, would effectively result in doubling the cost of the ERA-20C production, an impossible option given the project constraints. The radiation is called at the same horizontal resolution as the atmospheric model, whereas in the operational configuration the radiation runs at a lower horizontal resolution than the atmospheric model.

2.4 Data assimilation

The data assimilation system is an Ensemble of Data Assimilations (EDA) (Isaksen *et al.*, 2010). The ERA-20C configuration uses 10 members in total: 9 perturbed members and 1 unperturbed member or control. The land surface and ocean wave states are not analysed in the sense that they are not updated with the help of observations. For the atmosphere, the horizontal spectral triangular truncation in the assimilation is T95, or approximately 210 km. Each member employs a 4D-Var analysis scheme, with two minimisations. The 4D-Var analysis window is 24-hour-long, instead of 12 hours in ERA-Interim (Dee *et al.*, 2011a) and ECMWF Operations. These window configurations are shown in Figure 5. The ERA-20C analysis window starts at

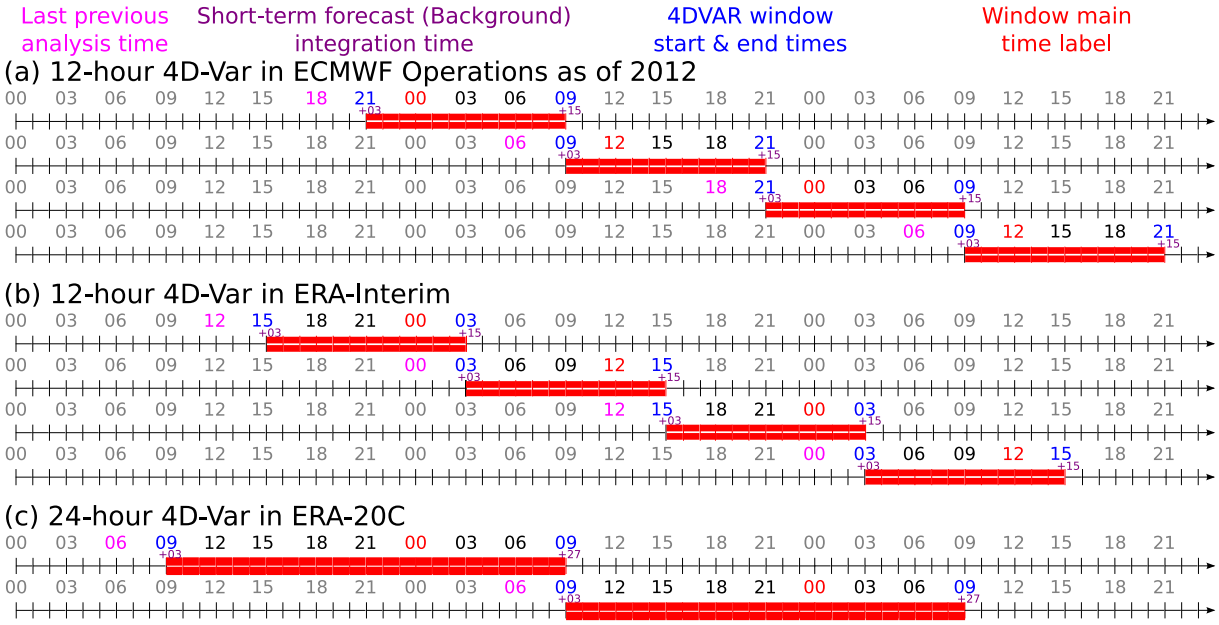

Figure 5: Configuration of the 4D-Var assimilation analysis windows used by (a) ECMWF Operations, (b) ERA-Interim, and (c) ERA-20C

9:00:01 UTC on any given day, and extends until 9:00:00 UTC the following day. In this window, observations are considered in 25 time-slots of 1 hour each (except for the first and last ones, 30-minute long). Each 4D-Var produces analysis estimates valid at 9 UTC, 12 UTC, *etc*... until 6 UTC the following day.

In the strong-constraint 4D-Var framework, the increment in control space is determined at the beginning of the window by computation of the initial state which gives the best fit given all constraints. The background state used as prior estimate is hence defined at the beginning of the window, to be valid at 9 UTC. It is formed as a 3-hour forecast issued from the latest prior analysed state (6 UTC on that same day). Consequently, the background error covariances used in the analysis scheme shall characterize the errors in these forecast states.

Another consequence of the extended analysis window is the necessity to activate the digital filtering of increments in temperature and vorticity. These effectively act to prevent analysis increments that would normally result in large unrealistic growing oscillations within the 24-hour window (Poli and Tan, 2012). If uncontrolled, such unrealistic growing oscillations are usually not a problem with a short analysis window, up to 12 hours, but lead more often to model explosion when integrated for 24 hours. Also, the digital filtering of increments in temperature and vorticity is found in ERA-20C to improve forecast scores (*ibid.*).

2.5 Bias correction

A variational bias correction (VarBC) scheme (Derber and Wu, 1998; Dee, 2004) is employed within the IFS 4D-Var to correct for biases in surface pressure observations. This scheme was originally developed specifically for ERA-20C (Poli, 2011). However, the basic tenets are still relevant for potential NWP applications. In this scheme, the bias assigned to each individual station is updated every time an observation from said station is assimilated in the 4D-Var. In ERA-20C, the observation biases are EDA-member-dependent. Initial experimentation with the scheme focused on finding optimal global settings for observation bias background error. The specification of this error, in terms of observation count equivalent, determines how many observations are needed for the bias estimate to completely absorb the mean difference seen in the observation minus

Stream number	Experiment identifier	Start date	End date
1	1726	1 January 1899	31 December 1920
2	1727	1 January 1919	31 December 1940
3	1728	1 January 1939	31 December 1960
4	1729	1 January 1959	31 December 1980
5	1730	1 January 1979	31 December 2000
6	1731	1 January 1999	31 December 2010

Table 1: Production streams employed for ERA-20C. Experiment identifier refers to the Meteorological Archive and Retrieval System (MARS)

background departures. In short, this quantity determines the reaction speed of the variational bias correction to a sharp change in the observation bias.

However, it was quickly realised that no global settings could satisfactorily handle the two very different problems found in reanalysis: 1) stations with timeseries breaks and 2) stations with stable biases. The former problem requires a fast update, while in the latter problem it is more desirable to adopt a fixed, or very slow evolution, in order to avoid absorbing as observation bias the background errors that occur on seasonal or synoptic time-scales. The solution adopted was to apply the concepts first developed by Haimberger (2007) to the problem of surface pressure bias detection (Hersbach, *in preparation*). The starting point was to generate observation feedback from 20CR, by comparing the observations to 20CR fields, or to use directly the 20CR observation feedback when available (namely for data in ISPDv3.2.6 that were already present in ISPDv2.2). Using then observation minus 20CR background-equivalent departures, a standard normal homogeneity test (SNHT) is applied to each observation timeseries. The results, time-varying values of likelihood of observation timeseries breaks, are archived alongside the observation as estimates of bias volatility. Finally, during ERA-20C production, the VarBC adjusts its reaction speed for each station on a daily basis, depending on the probability of break determined *a priori* as so-called bias volatility. A high probability indicates that the VarBC is to react fast, in order to rapidly estimate a new observation bias. A low probability indicates that the VarBC is to react very slow and essentially maintain its prior estimate. In the latter case, fluctuations in the observation minus ERA-20C background are then mapped directly onto analysis increments.

2.6 Production

The ERA-20C production, initiated on 18 December 2013, employs six parallel runs, or streams, shown in Table 1. The streams are separated by 20 years, and usually cover 22 years (except for the last one). In each stream, the first year allows to spin-up the system, and the last year (except for the last stream) is to enable to compare the quality of the overlap with the following stream, so as to estimate the magnitude of any possible boundary discontinuity. The layout of this production schedule is shown in Figure 6.

Each stream is initialised as follows. The initial conditions for each stream are taken from an ERA-Interim state valid on 1 January 1989. The bias correction is kept very small (near-zero) for the first six months. Without such precaution we find that the bias correction absorbs the large errors in the initial conditions. This is in line with the approach in 20CR (Gil Compo, *pers. comm.*). The background errors are initially kept constant for the first 3.5 months, before allowing for an automatic update (see section 4).

Overall each production stream advances at an average rate that varies between 20 and 40 days of reanalysis per elapsed day of production. These numbers are valid when each 4D-Var analysis is run on 2 nodes, *i.e.*, each production stream employs then 20 nodes (out of 1400 nodes installed on ECMWF high-performance comput-

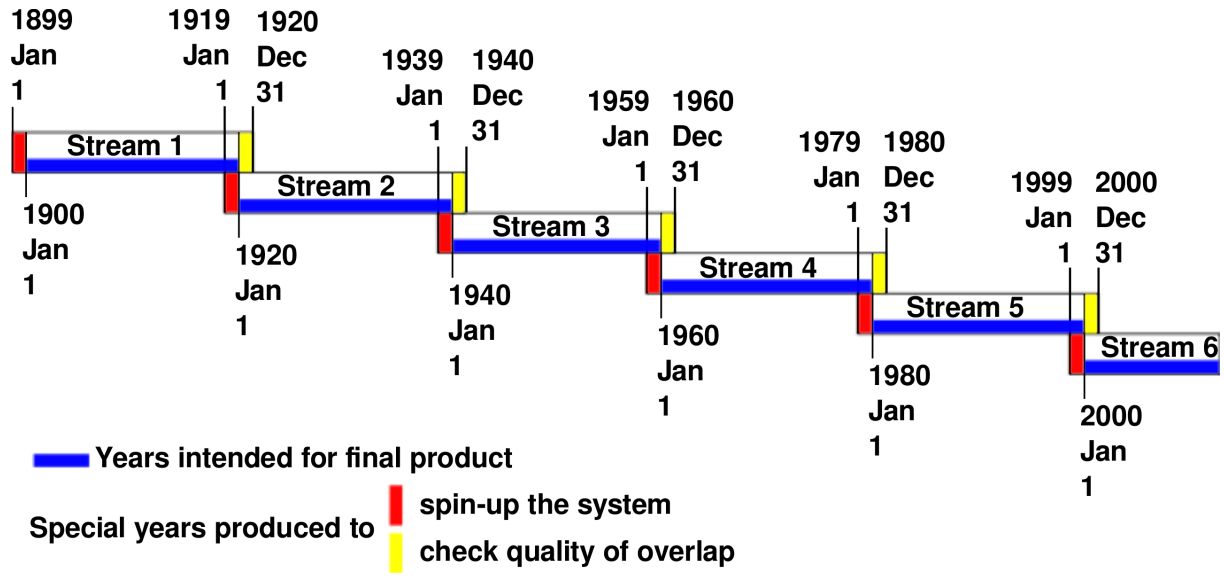

Figure 6: Schematic layout of the ERA-20C production streams

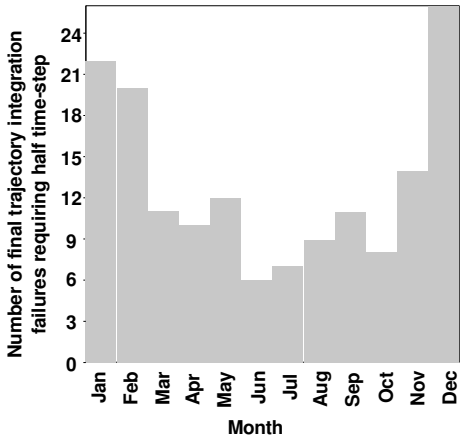

Figure 7: Monthly occurrences of final trajectory failures in ERA-20C, triggering automatic re-run with reduced time-step

ing as of early 2013). When the number of employed nodes is occasionally doubled, the rate of production can peak to 80 days of reanalysis per elapsed day of production. The variability in production speed is only partially related to high-performance computing system availability. The production speed is mostly affected, at irregular intervals, by disruptions caused by the inevitable upgrades that occur on a high-performance computing system mostly used for operational NWP targeted at continuous performance improvement.

During production, little manual intervention is required in the main analysis part thanks to an automatic procedure devised specifically to detect model explosions. These explosions occur nearly only in the last model non-linear integration (trajectory) after addition of the analysis increments throughout the analysis window. When such explosions are detected in any member of a production stream, the stream is automatically stopped and the date of the stream is set back to the previous day, with instructions to use a reduced model time-step (30 minutes instead of 60 minutes) for two days (*i.e.*, the day before the problem is encountered, and the day of the problem). This simple but robust automatic procedure is found to work satisfactorily, without the need to further halve the time-step to 15 minutes. As these problems are recorded as they happen, it is possible afterwards to analyse the occurrences of such problems. We find that these occur approximately randomly from year

to year. However, Figure 7 shows that most failures occur during the winter months of either hemisphere. This is consistent with findings by Poli and Tan (2012) who reported that local large analysis increments (generated by an analysis with a poorly observed system, 24-hour 4D-Var and 1-hour time-step) were sometimes found in areas of sharp sea-land transitions with ice cover: in Alaska or Siberia, but also next to the Ross Sea.

Problems requiring manual interventions can be related to file-systems being unavailable due to maintenance, or instabilities in the main tape archive (ECFS) which sometimes returns errors when files cannot be archived as expected. Other aspects requiring intervention are in the post-processing and conversion of observation data to the format expected for the OFA: for some BUFR observations (which are not assimilated, but simply used for comparison), at some dates, the conversion program fails and manual intervention is needed to remove the problematic data from the conversion.

3 Observations handling

3.1 Observation quality control

The observation quality control in ERA-20C is a four-step process. It is generally described in the IFS documentation for version cycle 38R1 (ECMWF, 2013). We highlight here the elements relevant to the observations assimilated in ERA-20C. First, a so-called blacklist process rejects observations according to predefined rules. In ERA-20C, the rules are as follows. All observations other than ISPD 3.2.6 or ICOADS 2.5.1 are rejected (This rejects BUFR observations from the ERA-40 and operational observation archives). Observations of geophysical variables other than surface pressure, surface geopotential, or wind, are rejected. This step also removes observations reported exactly at location exactly 0° latitude and 0° longitude, except for buoy observations because there is one buoy from the Pirata network at that location. The blacklist also rejects observations of wind above the land surface or near the coastlines. Because of the coarse horizontal resolution of the assimilation, such observations could be interpreted as over land and assigned an incorrect surface roughness. Observations of wind reported from latitudes 30°N–90°N at a station altitude that is distant by more than 50 meters height from the model surface are also rejected, because these observations are then located in closed seas in mountainous regions, posing a representativeness problem. At latitudes 90°S–30°N, where there are typically much fewer observations and closed seas, the criterion is less conservative and the maximum allowed height difference is 200 meters. Observations of surface pressure reported at heights more than 800 meters below the model surface are rejected. Likewise, observations of surface geopotential reported at pressure levels 100 hPa greater than the model surface pressure are rejected.

A second step assigns preferences so as to remove redundant mass information within a multi-variate report. The rules, as employed for operational NWP, are as follows: a surface pressure observation is preferred to a surface geopotential observation and a pressure observation reported at the station level is preferred to a pressure observation reported at sea-level. The first reason for this is that the reporting at station level is closer to the original measurement. The second reason is that there are several methods in use for the reduction of atmospheric pressure at sea level. This cause for disparity between measurements was discussed by WMO experts as early as the 1950s (WMO, 1954; 1964; 1968). However, as of writing, the World Meteorological Organisation only recommends a single practice for stations below 750 m altitude (WMO Commission for Instruments and Methods of Observation Expert Team on Standardisation, 2012). For stations located at altitudes higher than 750 m, the reduction methods are still a matter of regional choices. The resulting differences in mean-sea-level-pressure reduction can reach a few hectoPascals for high-altitude stations. This absence of standard practice means that the single observation operator used to compute atmospheric pressure at sea-level in the data assimilation system is sub-optimal in many cases, and introduces additional error. Consequently,

whenever possible, the surface pressure is preferred to the pressure reduced at sea-level.

A third step is the so-called background or first-guess check. It compares the observation to the first-guess computed from the short forecast initialised from the previous analysis and integrated between 3 hours and 27 hours, to cover the entire 4D-Var window (see Figure 5). Observations with departures larger than a threshold are rejected. With the introduction of the Huber norm in the variational quality control (Tavolato and Isaksen, 2010), the threshold is now extremely large, about eighteen times the standard deviation of the expected error. The expected error is the square root of the sum of the squares of background error and observation error. Since the background error varies over time and space, the expected error also varies over time and space, and the rejection threshold also varies. For surface pressure observations, the rejection thresholds decrease from regional maxima (minima) of 310 hPa (20 hPa) in 1901 down to regional maxima of 175 hPa (17 hPa) in 2008. For wind observations, the rejection thresholds are also similarly extremely permissive, but an added criterion removes observations reporting or in regions of wind speed greater than 35 m/s. Overall, less than 1% of the observations are rejected by the first-guess check.

Finally, a variational quality control (Andersson and Järvinen, 1999) using the Huber norm (Tavolato and Isaksen, 2010) rejects observations that cannot be fitted within reasonable limits by the first minimisation. About 2% (14%) of the surface pressure (near-surface wind, respectively) observations are typically rejected by this step in the early years of ERA-20C. We find that this rejection rate increases as the observation time is located towards the end of the analysis window. Any observation-related issue would be constant over the window. Consequently, this behaviour points to yet further possible improvements in the data assimilation; in particular here, we use a strong-constraint 4D-Var without representation of model error (a weak-constraint formulation would allow for gradually increasing model errors within the window). However, in the early 2000s, when the system quality improves with more observations, fewer observations are rejected at this step, with less than 1% surface pressure observations, for example.

In ERA-20C it turns out that the first-guess check limits are too generous and let sometimes obviously wrong observations enter the analysis. The worst occurrence is found with surface pressure observations from the island Tristan da Cunha located at an altitude of 51 m above sea-level, in the Southern Atlantic Ocean, latitude 37.05°S longitude 12.32°W. In ISPD 3.2.6 this station comes from collection number 3013 (‘South African Weather Service Stations (South Africa) 1850-2003’). Surface pressure observations from this station are found for the first time in January 1954, in the range between 719 hPa and 829 hPa, with an average of 819 hPa. In the following weeks, these observations always fail the first-guess check because the background departures, around -190 hPa, exceed the 18-sigma limit which fluctuates between 50 hPa and 150 hPa. However, around mid-April 1954, the local background error at this location increases so that the 18-sigma first-guess check limit exceeds 200 hPa. The day this happens, the observations from Tristan da Cunha pass the first-guess check and are accepted in the first minimisation. The first few days that this happens, the neighboring observations from ships cause the first minimisation to fail to fit the abnormally low pressure observations from Tristan da Cunha, after which point the observations are rejected by the VarQC. Yet, this rejection happens too late because the bad observations that were assimilated in the first minimisation pull the analysis state to a much lower surface pressure, which does not get filled in completely by the forecast model in 24 hours. Worse, after a few analysis cycles, the bad surface pressure observations manage to pull the analysis state closer to around 800 hPa at the island location: the bad data are then completely assimilated, and all the neighboring observations from ships are rejected. After that point, the subsequent months see an unrealistic low centred on Tristan da Cunha, at around 800 hPa. As this problem was spotted during production of ERA-20C stream number 3, the observations from this station were manually blacklisted and the corresponding production stream re-set to 1954. However, this was only done because the problem was so serious that it affected monitoring statistics for large areas. It is now clear that many similar problems must be present throughout the century, albeit of smaller amplitude. After the fact, it now appears that the 18-sigma limit was determined from modern time period with a modern

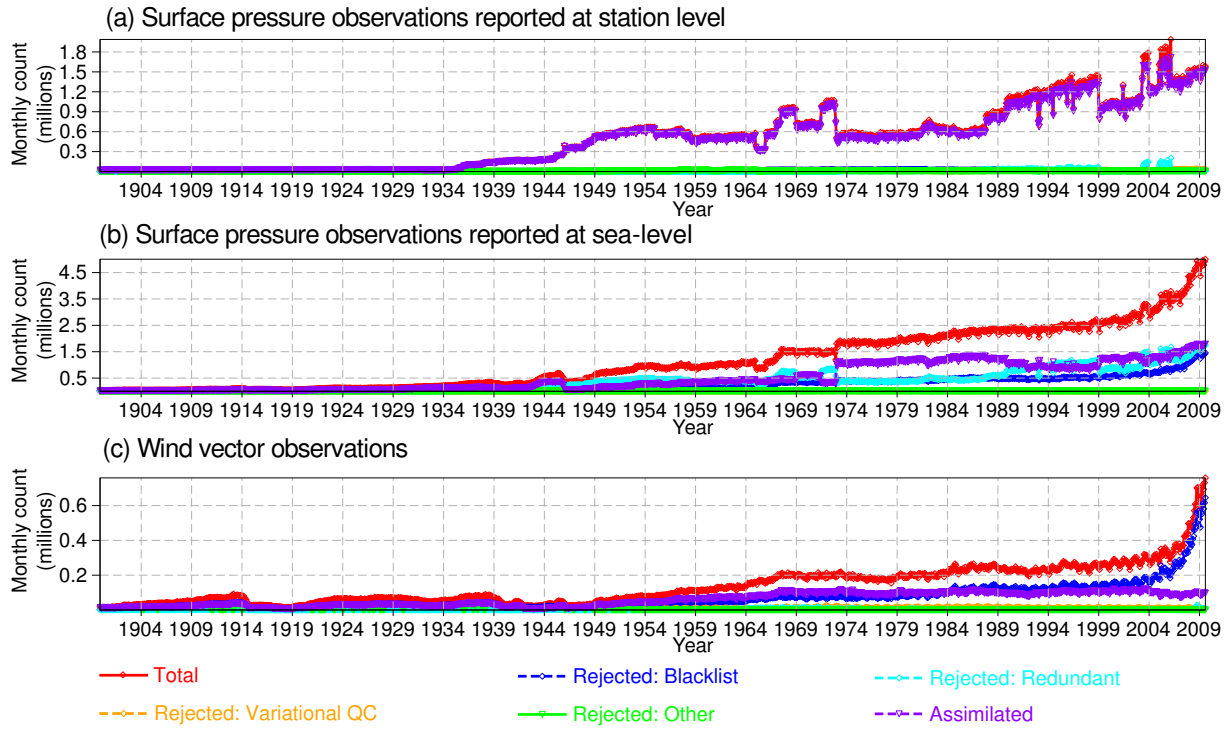

Figure 8: Monthly count of observations of the geophysical observations used in ERA-20C, showing impact of the various observation quality control steps described in the text

observing system and when the background errors are smaller than 1 hPa. The problem is that this limit was determined from statistics of observation minus background departures normalised by sigma defined as the observation error, rather than defined as used by the assimilation (*i.e.*, square root of the sum of the squares of background and observation errors). In the light of this, it now appears that the first-guess rejection limit should have been 10 sigmas, rather than 18 sigmas as used by ECMWF NWP operations. Another problem understood *a posteriori* is that the expected error does not vary over time during the assimilation window, and may need to be dynamic for long 4D-Var windows.

Figure 8(a) shows timeseries of the monthly observation counts for surface pressure observations reported at station level. Out of the total available, from ISPD 3.2.6 and ICOADS 2.5.1, most of the observations are used. For observations reported at sea-level (Figure 8(b)), most of the data rejections relate to the redundancy check, because these may be duplicated with surface pressure observations reported at station level, which are preferred. Because of the generous limits as explained above, the rejections due to first-guess and variational quality controls are too small to appear different from the zero line. Figure 8(c) shows the observation counts of wind vectors. About half of the winds are rejected by the blacklist step, most likely because a large fraction of maritime traffic remains fairly close to coast. Otherwise, as for surface pressure, most of the wind observations that pass the blacklist quality control stage are assimilated.

Figure 9(a) shows timeseries of assimilated surface pressure observation yearly counts, by report type. The predominance of observations from land stations (*i.e.*, ISPD 3.2.6) is visible in the recent years. The figure indicates clearly the importance of ship observations (*i.e.*, ICOADS 2.5.1) in the earlier years. The number of tropical cyclone best-track archive data (IBTrACS) increases over time during the first half of the century, and appears nearly stable only the second part of the century. This remark is important for future potential interpretation of the trends in tropical cyclone activity in ERA-20C. Meteorological vessel observations are found during the post-World War II era, until these types of dedicated ships were retired owing to the deployment of

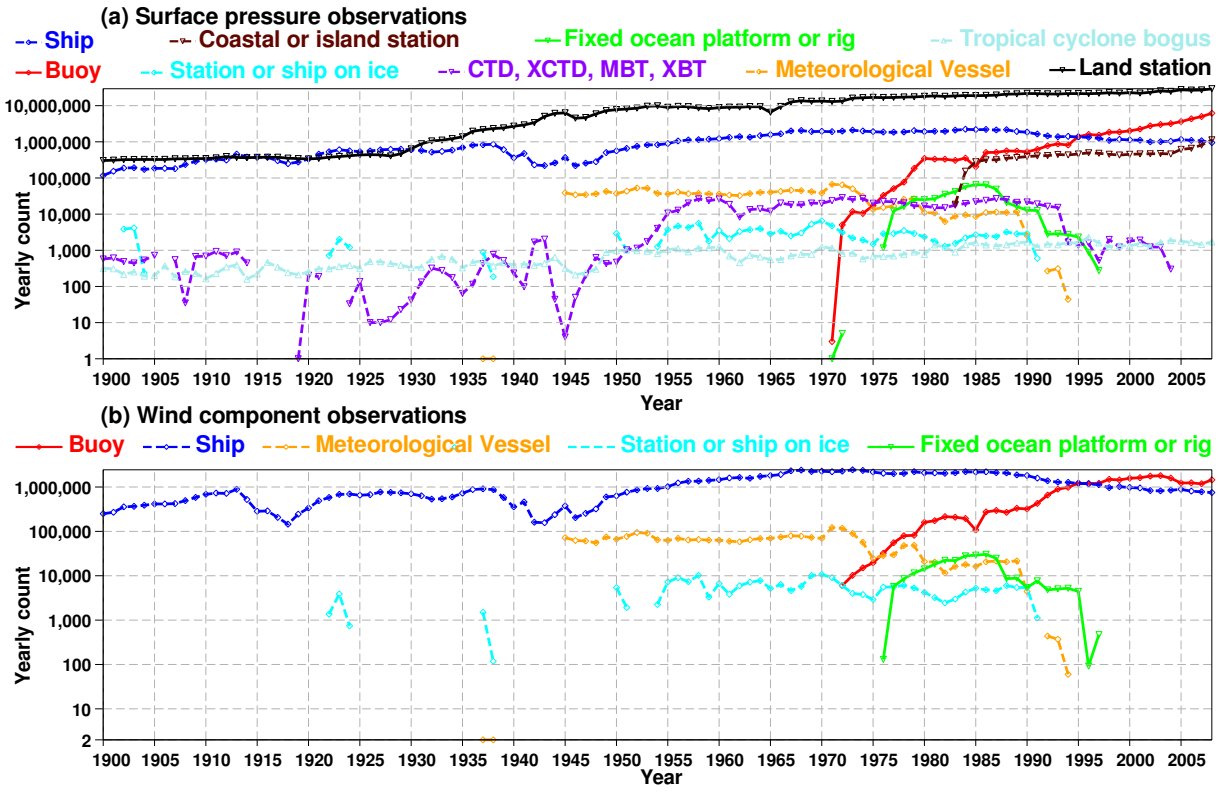

Figure 9: Yearly count of observations of the geophysical observations assimilated in ERA-20C, by report type. Only complete years are represented

satellite observing systems offering global observations. Figure 9(b) shows that, for the most part, assimilated wind observations come from ship. Only in the last stream, buoy observations (from drifters) exceed the data provision from ship.

3.2 Observation errors assigned a priori

The standard deviation of the observation error, hereafter simply called observation error, is specified as follows. For surface pressures, the observation-type-dependent specification is expressed in terms of air mass and hence varies also with the actual observed pressure. It amounts approximately to about 1.1 hPa for observations reported above land, 0.9 hPa for buoy observations, 1.5 hPa for ship observations, and 1.6 hPa for tropical cyclone bogus from IBTrACS. Observations of wind, which is assimilated according to its zonal and meridional components, are assigned component errors of 1.5 m/s for ship and 1.3 m/s for buoys (we remind that wind observations over land are not assimilated). Note that these estimates include estimates in observation timing error and observation location error. Such location errors are very likely greater for ship observations in the early part of the century. The same comment may apply to buoys until the mid-1990s until the widespread use of Global Positioning System (GPS) for the determination of geographic coordinates.

The observation error estimate is adjusted within the assimilation by the Huber norm (Tavolato and Isaksen, 2010). Potentially 'bad' data are not given the same weight as obviously 'good' data. This formulation allows an impact from sets of observations with consistent large departures (*e.g.*, several clustered observations of a deep low not found in the background). In ERA-20C, this minimises the impact of 'bad' data, located too far away from the first-guess estimate.

3.3 Observation errors assessed a posteriori

After assimilation, it is possible to estimate the observation error, using the method proposed by Desroziers *et al.* (2005). This method assumes that the background and the observation errors present different correlation structures. Figure 10(a) shows observation error estimates using this method for surface pressure. For land station observations, the plot suggests that surface pressure observation quality improves over the century, from 1.6 hPa in 1900 to about 0.8 hPa in the years 2000s (compare this with the constant assumed estimate of 1.1 hPa). The surface pressure land observation quality improvement is especially visible in the first two streams, pre-World War II. A mixture of factors may explain this improvement, such as generally improved instrumentation and site installations, more regular calibration practices and facilities, and also probably more accurate reporting of the observation time. Surface pressure observations from buoys appear to be of similar quality as from land stations in the recent times, but of worse quality in the early years of introduction (in the 1970s); the location error may play a large role in this total error estimate.

The surface pressure observations from ship also appear to improve in quality, from 2 hPa down to 1.2 hPa in recent years (compare this with the constant assumed estimate of 1.5 hPa). The error estimates of surface pressure observations in ICOADS 2.5.1 from Ocean bottle and Conductivity Temperature Depth (CTD), eXpandable CTD (XCTD), Mechanical or digital or micro Bathythermograph (MBT), Expandable Bathythermograph (XBT) are based on small yearly samples most of the time (see Figure 9), and many of those estimates are not shown because they are based on fewer than 100 observations in a given year. Nevertheless, the error estimates for CTD/XCTD/MBT/XBT shown here are substantially larger than for other report types and would need to be understood in future repeats of similar reanalysis. Tropical cyclone bogus observations appear with about near-constant errors throughout the century, around 4–5 hPa, although this estimate may be affected by the poor horizontal resolution of ERA-20C. The quality of ship observations appears to be degraded during World War II, either probably of less frequent recalibrations and/or limited ability to conduct observations at the expected

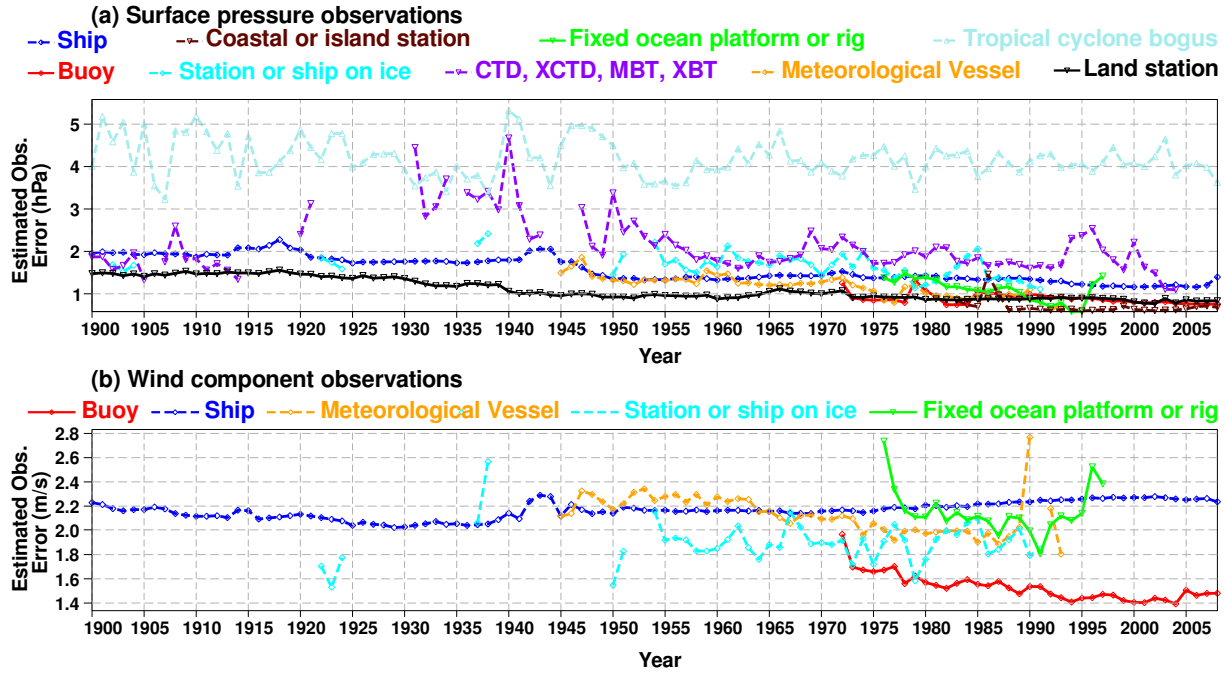

Figure 10: Estimate of observation error for the geophysical observations assimilated in ERA-20C, by report type, computed on a yearly basis (estimates based on 100 observations or less in any given year are omitted; see Figure 9 for the corresponding yearly observation counts)

time. In the 1980s, the quality of surface pressure observations from meteorological vessels approaches that of land station measurements. This could be explained by a better known position of the ship, and/or more regular observing times closer to the schedule time, and/or more regularly recalibrated instrumentation and trained onboard personnel.

Regarding wind observations, Figure 10(b) indicates also the better quality of dedicated meteorological vessels as compared to regular ship observations in the 1980s. However, one must note the great disparity between the quality of observations labelled as coming from ships, as some of them also come from trained and regularly serviced facilities (*e.g.*, national navy). The wind observation errors from ship appear to increase slightly in the recent years, in excess of 2.2 m/s, compared to usually less than 2.0 m/s over the century. One possible explanation is the larger size of vessels that roam the seas today. Larger ships are more likely to influence the immediately surrounding wind circulation than smaller ones. Since the 1980s, the near-surface atmospheric wind observations with the best quality appear to be provided by buoys, with wind component errors estimated at about 1.4 m/s.

Overall, these preliminary results suggest that the constant observation error specifications ought to be time-varying in future reanalyses, and refined in more detail to separate meteorological vessels from other ships. These initial findings should be confirmed by cross-checking with independent evaluations, *e.g.*, triple collocation.

4 Background errors

Compo *et al.* (2011) demonstrated that the spread of their reanalysis ensemble fluctuates significantly with the observation availability throughout the century. Seen today, improvements in observation coverage over the

century result from expansions in observing networks at the time, but also from current data recovery efforts which put into digital form entire sets of observations from original paper or microfilm records. As seen in Figure 3, the evolutions in observation availability tend to affect wide areas; they are regional in nature and not spatially random. They consequently need to be accommodated by the assimilation system. In 20CR, this is achieved by the ensemble-based assimilation system. The ensemble spread, fluctuating with the observation coverage, is directly used as an approximation of the background error to filter the observations in the analysis step. However, in a variational system such as that employed in ERA-20C, specific developments had to be made to achieve such time-varying evolution of the background error covariances. We first indicate below how the background errors are specified in ERA-20C, and how this differs from ECMWF Operations as of December 2012 (Fisher 2003, Fisher 2004). We then illustrate how the EDA information is used to estimate the background errors. We also present initial results from the first years of ERA-20C production.

4.1 Separation between local error standard deviations and global error covariances

We use hereafter the following terminology: **'background errors'** designate background error standard deviations, and **'background error covariances'** refer to the product of average global background error variances with background error correlations.

The ability to specify daily space-varying background errors was introduced in ECMWF Operations in 2009. The spread between forecast members in the EDA is used to modulate the map of background errors of vorticity. Note that this only affects regional patterns of background errors of vorticity, and not the global average of vorticity background errors, which are specified in the background error covariances as defined above. The background errors of all other variables, including the logarithm of surface pressure, are affected by the flow-dependent contribution from vorticity through the nonlinear balance and omega equation plus hydrostatic relationships. However, while the ECMWF Operations adjusts the map of background errors used by the high-resolution deterministic system, the ECMWF Operational EDA uses near-globally constant background errors to perform its own analysis. The ERA-20C EDA is different, in that it uses its own vorticity spread to consistently update the vorticity background errors that are used at the next analysis cycle.

The background error covariances are only updated by the ECMWF Operations when there are significant model changes (*e.g.*, resolution upgrade) or significant changes in the data input and/or assimilation. In ERA-20C, in order to achieve time-variations of the background errors, background error covariances are updated every 10 days of reanalysis.

4.2 Derivation of background errors from the EDA

We retrace now the computation of background errors in ERA-20C. To this end, we first remind the setup of the EDA. The control and perturbed members are each cycled in time. The perturbed members use a model and assimilation system with theoretically doubled errors (Bonavita *et al.*, 2010). The intention is that the spread of the difference between the perturbed members and the control gives an estimate of the total error in the control state (Isaksen *et al.*, 2010). To achieve this, the individual error sources are represented as follows. The model error is represented by stochastic physics (Palmer *et al.*, 2009). Observation errors are represented by adding perturbations to the observations. As for the forcing errors, in the ECMWF Operational EDA, perturbations are added to the sea-surface temperature field. However, ERA-20C uses an ensemble of sea-surface conditions provided by the Met Office (HadISST.2.1.0.0 by N. Rayner and collaborators, dataset yet to be published), without additional perturbations.

We now consider the year 1900, which is the second production year of the first ERA-20C stream. Figure 11(a)

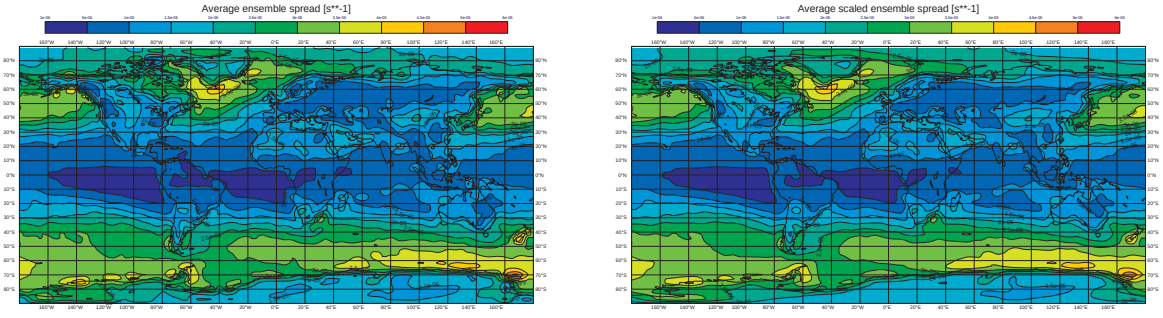

Figure 11: Left: (a) Average background ensemble spread of vorticity at model level 89 (i.e., near the surface) for the year 1900. Right: (b) Average background scaled ensemble spread of vorticity at model level 89 for the year 1900

shows the average background ensemble spread of the EDA in vorticity near the surface, at model level number 89, for that year. As explained above, the backgrounds are defined as forecasts initialised at 6 UTC, from +3 hours to +27 hours. On any given day of the year 1900, $i \in [\![1, N]\!]$ with N =365, the spread in vorticity between the members, or vorticity ensemble spread, is noted σ_{v_i}. The average vorticity ensemble spread over the year, noted $\mathbf{S}_v$, is computed as follows:

$$\mathbf{S}_v = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\sigma_{v_i}^2} \tag{1}$$

The map is computed using a grid of 5° latitude × 5° longitude. This gridding is used throughout section 4 for consistency. The map shows lower standard deviation over the Tropics (as expected because of lower natural variability to begin with). Comparing with Figure 3(a) which shows the observation coverage in 1900, we also note that the vorticity spread is lower where observations are consistently present (North America, Europe). The Southern Oceans and the Arctic, mostly unobserved back in those days, show the largest standard deviations. The storm track eastern of the U.S. East Coast shows standard deviations that are locally larger than compared to the western Atlantic Ocean.

The average background ensemble spread shown above could be considered a true estimate of the background error for the parameter considered (vorticity at model level 89) if all the error sources in the EDA were properly represented. However, because one cannot reasonably expect all these errors to be exactly represented, a calibration and smoothing over a number of days are applied to form a so-called scaled ensemble spread. For further details about scaled ensemble spread see the works of Bonavita *et al.* (2011), Bonavita *et al.* (2012), and Bonavita (2011). In the ECMWF Operational EDA, the calibration is determined every day using the high-resolution deterministic analysis as a reference. However, in ERA-20C, because of the absence of verifying analyses of sufficient quality for the whole century, the calibration is assessed once for a recent time period (June-August 2004), and applied throughout the century.

The average background scaled ensemble spread of vorticity at model level 89, noted $\mathbf{S}_v^s$, is shown in Figure 11(b). This map is computed using a formula similar to (1), except that each input field is the scaled ensemble spread $\sigma_{v_i}^s$. This map very much looks like Figure 11.

As explained in section 4.1, the maps of vorticity scaled ensemble spread are used to modulate locally the global-mean profiles of vorticity background error standard deviations read from the global covariances. To illustrate the lack of horizontal variability in these covariances, Figure 12 shows a map of the background error standard deviations found in the global covariances, for a selected date (1 February 1900). This map is generated with the help of an off-line diagnostics tool applied outside the assimilation system. As a first approximation,

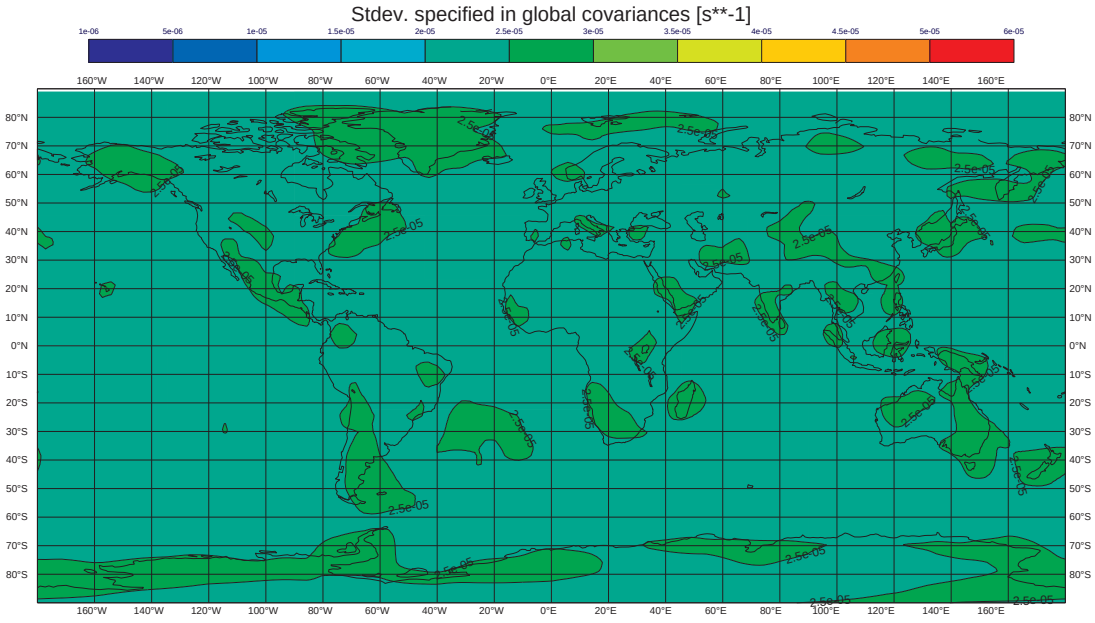

Figure 12: Background errors of vorticity at model level 89 specified by the global covariances, before modulation by the EDA spread, for the year 1900

the information contained in this map can be summarised by a global scalar. Such background errors would have been those employed by a variational analysis scheme in the absence of an ensemble. The lack of hemispheric differences justifies the importance of modulating these global covariances by an ensemble-driven map of spread to represent differences in background quality that results from the differential observational coverage as shown in Figure 3. For vorticity at model level 89, we note these scalar values as s_{v_i}. Following the procedure described in IFS documentation (ECMWF, 2013) we simulate how the field of background errors $\mathbf{S}_v^b$ is formed:

$$\mathbf{S}_v^b = R \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left[\frac{s_{v_i} \sigma_{v_i}^s}{\sqrt{\overline{\left(\sigma_{v_i}^s\right)^2}}} \right]^2} \tag{2}$$

where the overbar is an operator that computes the global average of a field.

The resulting field $\mathbf{S}_v^b$, shown in Figure 13, is an estimate of the actual background error standard deviation used by the assimilation for vorticity at model level 89. It retains the main geographical features of the scaled ensemble spread field shown in Figure 11(b), except the overall level has been adjusted.

Note, all these calculations are conducted off-line, outside the actual data assimilation system. They do not necessarily represent what is used by the assimilation. However, estimates of background error standard deviations are generated inside the assimilation by a randomisation method (Fisher and Courtier, 1995), for the control variables. In our configuration, such on-line estimates are computed at the end of the analysis window (24-hour-long here). They hence do not describe exactly the background errors used to characterize the background at the beginning of the analysis window, but they are the closest on-line diagnostics available. They are shown in Figure 14, for vorticity at model level 89. A formula similar to (1) is used to compute the average over N fields of daily on-line diagnostics of background error fields. Comparing Figure 13 and Figure 14, we can see very similar features, the latter appearing as a smoother version of the former. These differences can be expected as the randomisation will never render all the details as a limited number of vectors are employed.

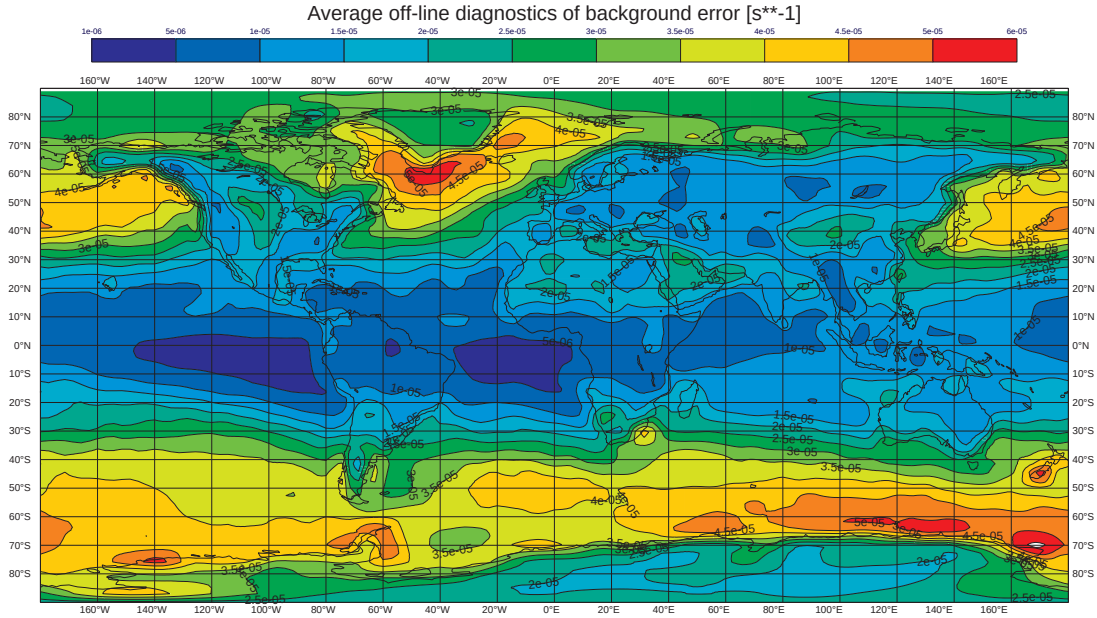

Figure 13: Off-line estimate of the average background error of vorticity at model level 89 used by the assimilation, for the year 1900

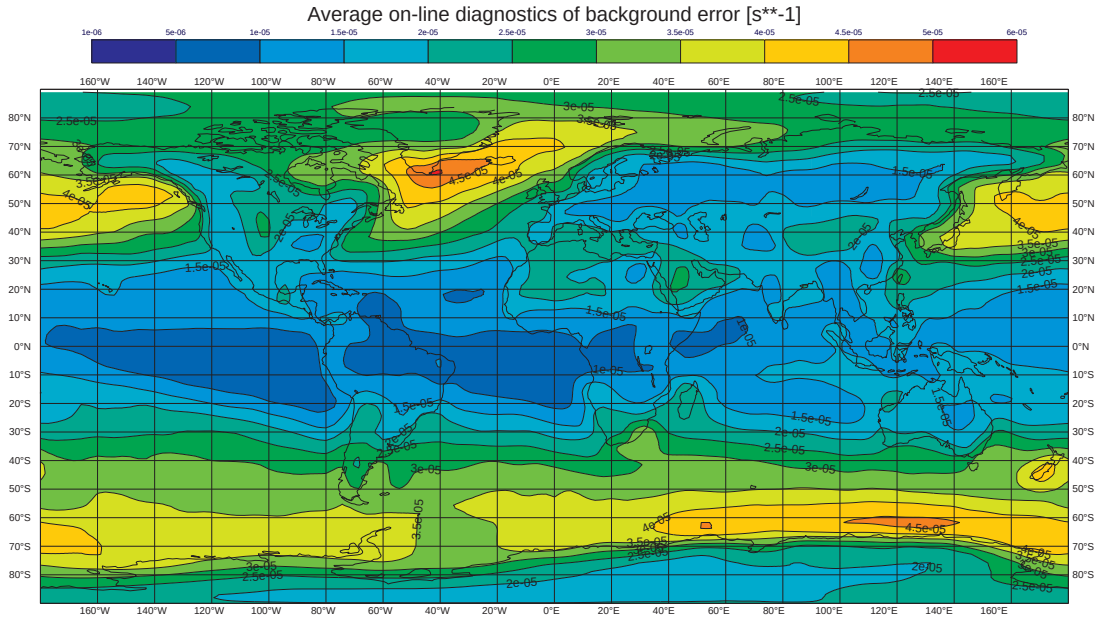

Figure 14: On-line diagnostics of the average background error of vorticity at model level 89 used by the assimilation, valid at the end of the 24-hour analysis window, for the year 1900

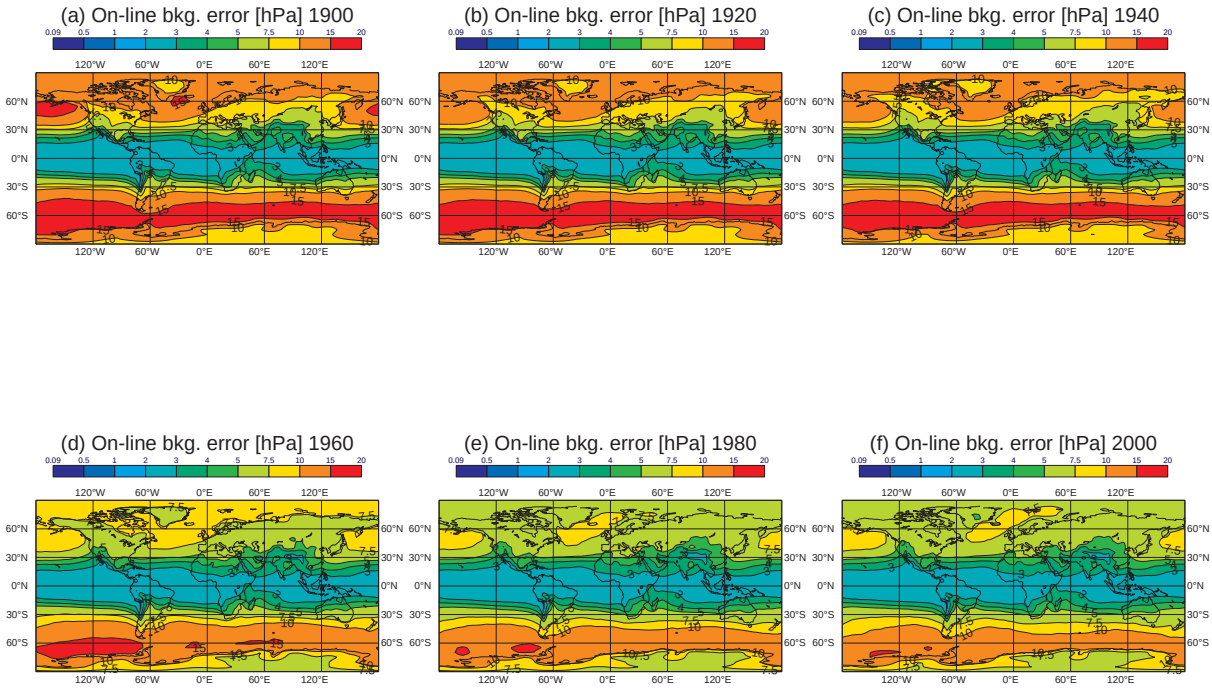

Figure 15: On-line diagnostics of the average background error of surface pressure used by the assimilation for various years, valid at the end of the 24-hour analysis window

This is especially true with the ERA-20C settings, where the number of random vectors (Fisher and Courtier, 1995) is reduced to 18 as compared to the operational default of 50, so as to save computing power. Also, as expected, the errors shown in Figure 13 are a bit smaller than those shown in Figure 14, because the former apply to the background at the beginning of the analysis window, while the latter apply to the background at the end of the analysis window (24 hours later).

4.3 Maps of background errors throughout the 20th century

Following from the investigations above, we now show maps of on-line estimates of the background errors employed by ERA-20C for geophysical variables with corresponding assimilated observations: surface pressure and 10-meter wind.

Figure 15(a) shows the average background errors used for surface pressure, for the same time period as earlier. Also shown is the second year in each of the five other streams in ERA-20C (years 1920, 1940, 1960, 1980, and 2000). The evolution over time is the product of the improvement in the surface observing network and reduced spread in the sea-surface conditions, which is also related to improvement in observation coverage.

As mentioned earlier, the ensemble spread in all control variables is used to recompute global covariances of background errors of all control variables every 10 days. These are computed from a large number of forecast states, drawn from the EDA members over the past 90 days. On a daily basis, the EDA adjusts the background errors using the ensemble spread, but only from vorticity. The ensemble spread of the other variables (*e.g.*, surface pressure, temperature) are not used for this daily modulation. The background errors of these geophysical variables still get adjusted by vorticity, but only through balance relationships. This implies that the background errors used for surface pressure are not necessarily consistent with the ensemble spread of

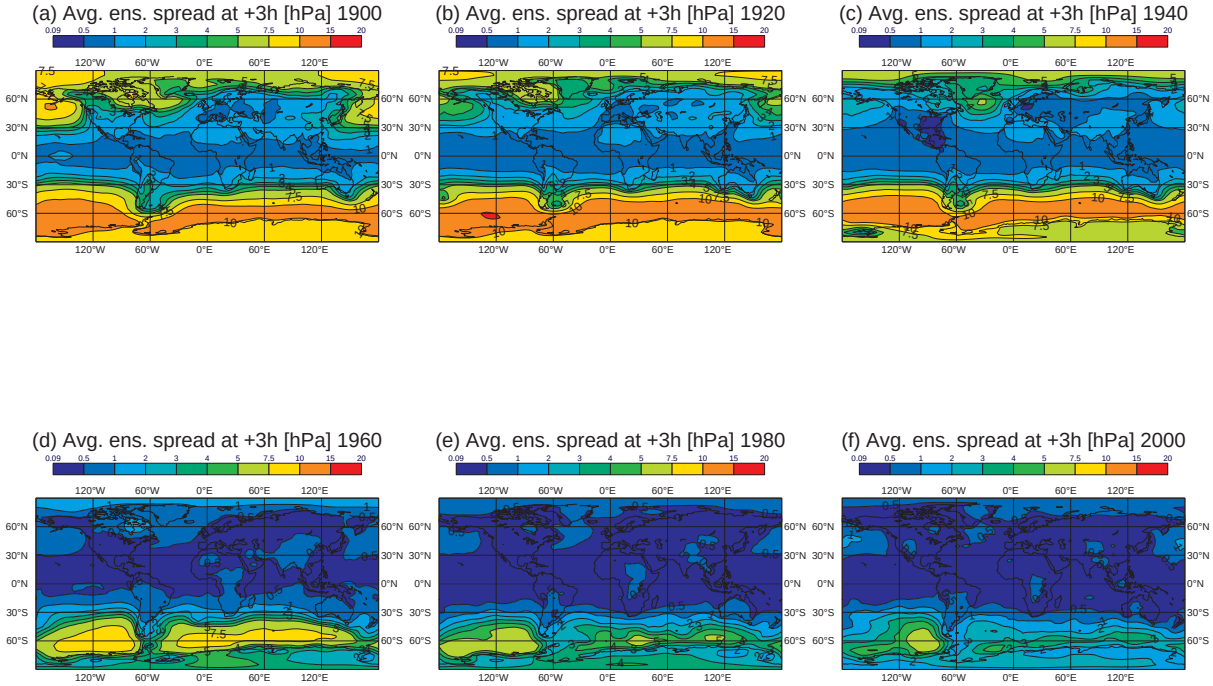

Figure 16: Average background surface pressure ensemble spread, at the onset of the analysis window

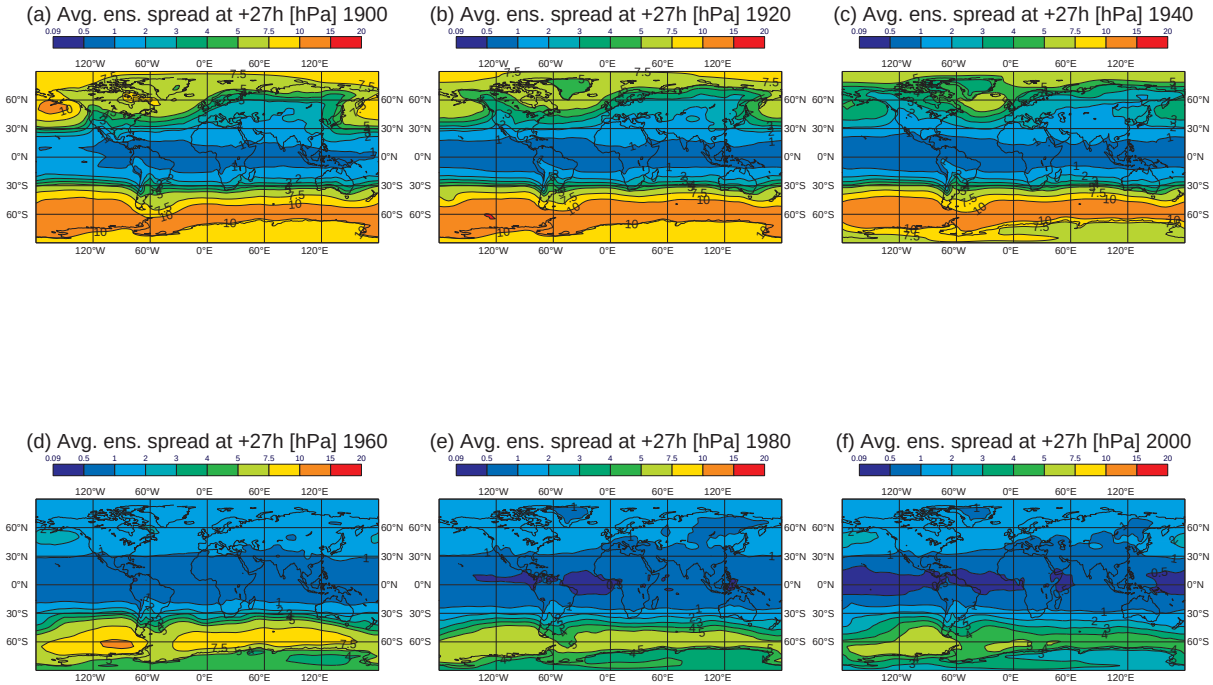

Figure 17: Same as Figure 16 but at the end of the analysis window

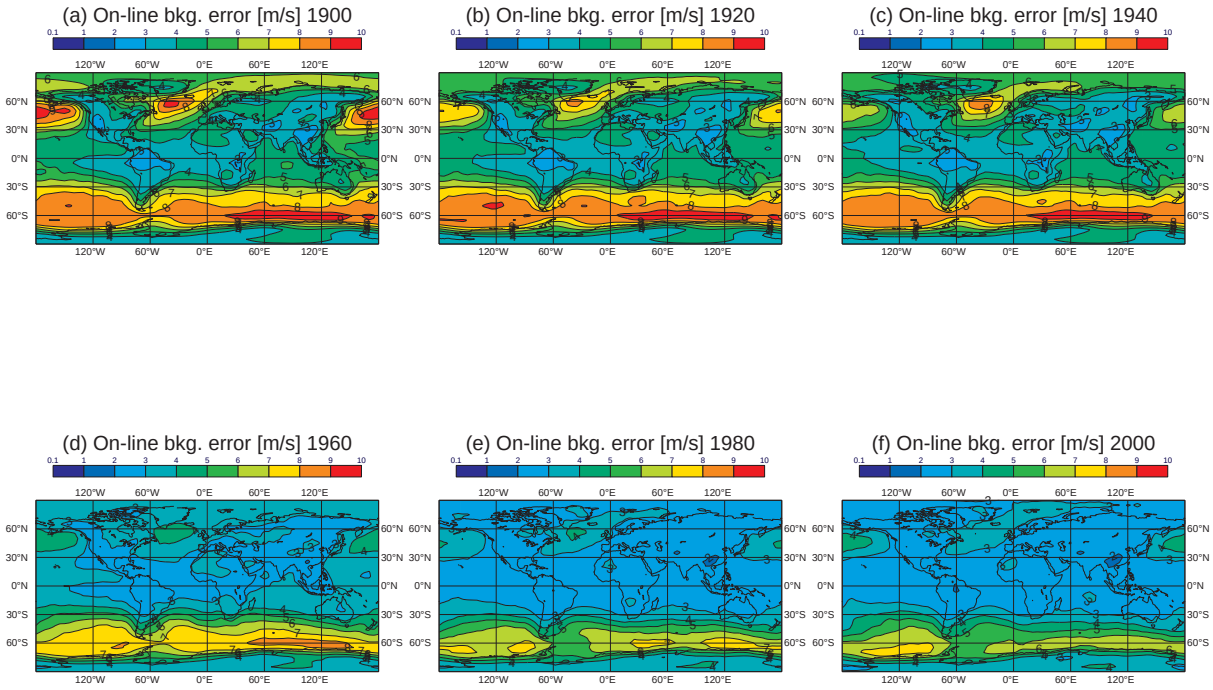

Figure 18: On-line diagnostics of the average background error of near-surface zonal wind, valid at the end of the 24-hour analysis window

surface pressure. Figure 16 shows the average ensemble spread of surface pressure at step 3 hours (onset of the analysis window). At the beginning of the century, the ensemble spread at step 3 hours ranges from around 1 hPa in the Tropics and Europe to 15 hPa in the Southern Oceans and Antarctica. At the end of the century, this spread collapses to under 1 hPa for most of the globe, except the Southern Oceans, where it remains above 3 hPa (the largest spread is observed in the Southern Pacific). As expected, Figure 17 shows that the spread is larger at the end of the analysis window (at step 27 hours), owing to error growth (model errors, and errors in initial conditions). The disagreement between the surface pressure ensemble spread and the surface pressure background errors is explained by the incomplete cycling of spread information in the EDA, whereby only the vorticity spread is cycled as explained above. Nevertheless, we observe that surface pressure background errors (Figure 15) show similar geographical variations as maps of surface pressure spread (Figure 17). The overall disagreement is really about the absolute level of error: the background errors appear too large as compared to the ensemble spread. For example, the background error used over the Southern Oceans in the early years in between 15 and 20 hPa, whereas the ensemble spread is between 10 and 15 hPa. However, we remind that the ensemble spread is not exactly equal to background error, and that in the absence of sufficient amounts of observations to regularly carry out a spread-skill calibration, these background errors represent a first estimate which should be revisited in future reanalyses in the light of the overall evaluation of the ERA-20C final product.

For completeness, Figure 18 shows the average background errors used for zonal wind at model level 91 (lowest model level). In 1900, they range from 2–4 m/s in observed areas to 7–10 m/s in the Atlantic storm track and the Southern latitudes. In 2000, they reduce to under 2 m/s globally, with the greatest uncertainties in the Southern Pacific, around 5–7 m/s.

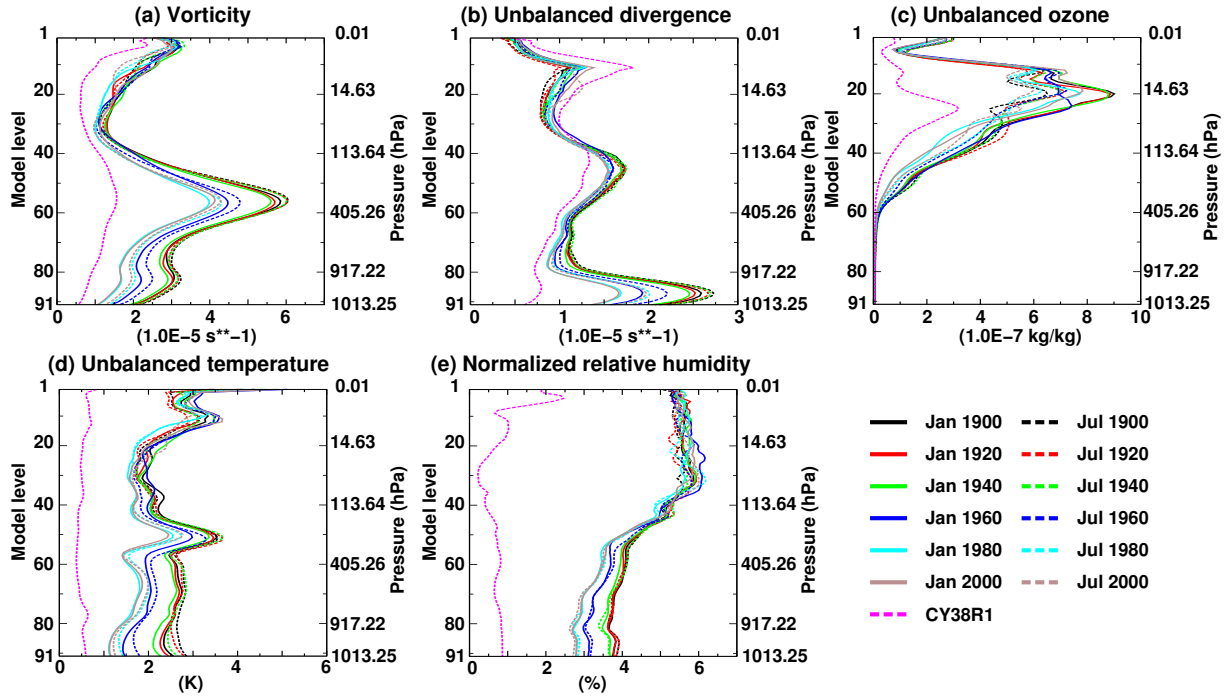

Figure 19: Vertical profiles of background errors for (a) vorticity, (b) unbalanced divergence background errors, (c) unbalanced ozone, (d) unbalanced temperature background errors, and (e) normalised relative humidity. On each plot, the left-hand-side vertical axis indicates model level number. The right-hand-side vertical axis indicates the corresponding pressure level for a surface pressure of 1013.25 hPa. CY38R1 refers to the default settings for a current-day experiment, when all available observations, including radiosondes and satellites observations, are assimilated in an experiment at horizontal resolution T95

4.4 Profiles of background errors throughout the century

Figure 19(a) ((b), (c), (d), and (e)) shows the global-mean profile of background errors for vorticity (respectively: unbalanced divergence, unbalanced ozone, unbalanced temperature, and normalised relative humidity). Several selected months in ERA-20C are shown. The plot also displays the default specifications for IFS cycle 38R1, for comparison. These apply to the IFS numerical weather prediction system with the current observing system (*i.e.*, including radiosondes and satellites). Note that the results shown in these plots do not describe the background errors used by ECMWF Operations where analyses are conducted at a higher resolution than considered here (spectral triangular truncation T95). When all observations are available, CY38R1 shows that background errors of unbalanced temperature are globally around 0.5–1 K throughout the vertical. In ERA-20C, these values increase to 2–4 K in the stratosphere. In the troposphere, lower errors are found in the more recent times, down to 1–2 K in year 2000. The improvement over the years is thus visible on this plot. We also note seasonal effects, with larger errors in the summer months of the Northern Hemisphere as compared to the winter months. As expected, the background errors only change over time at model levels in the troposphere (pressure levels greater than 100 hPa, or model level numbers greater than 40). The stratospheric model levels (model level number lower than 40) feature constant errors throughout the century. This finding is in line with Figure 1, namely that a surface-pressure only reanalysis has essentially no observational information in the stratosphere.

Figures 19(a,d) suggests three clusters in terms of background errors: the years 1900, 1920, and 1940 appear of similar quality; the years 1980 and 2000 feature lower errors; the year 1960 is an intermediate step. This clustering also matches with maps in Figure 18, which can be grouped as follows: (a,b,c), (d), (e,f). The year boundaries for the changes of quality shall be confirmed when the complete ERA-20C record is available.

4.5 Horizontal background error correlations throughout the century

Figure 20(a) ((b), (c), (d), and (e)) shows background error horizontal correlations for vorticity at model level 89 (respectively: unbalanced divergence at model level 89, unbalanced ozone at model level 7, unbalanced temperature at model level 89, and normalised relative humidity at model level 89). The structure functions appear to be more localised in the operational configuration (CY38R1) where all observations are assimilated. The seasonal dependence appears to be small. However, over time, one can observe that the correlation distances for background error are getting shorter. Consequently, the observations in the year 2000 must have a shorter radius of influence on the analysis than in the year 1900. This is consistent with a background that has longer wavelengths that are better known at the end of the century, thereby allowing for more small-scale adjustments to fit the observations. In other terms, the variational assimilation reacts to the observation coverage as follows. In the beginning of the century, a poor observation coverage means that the observations are used to constrain the large-scales features, with little emphasis on small-scale increments. At the end of the century, a dense observation coverage allows to extract small-scale features from the observations.

4.6 Diagnostics in the observation feedback

The sections above described diagnostics of the background errors, with the on-line diagnostics fields intended to be published and accessible to users. These diagnostics are in the space of the model control variables. Future users of the ECMWF Observation Feedback Archive (OFA) will also be able to access similar information in observation space. For that purpose, the ERA-20C observation feedback stores estimates of the scaled ensemble spread at the beginning of the analysis window, in observation space. For example, the average surface pressure background errors from the OFA shown in Figure 21(a) are essentially the same quantities as

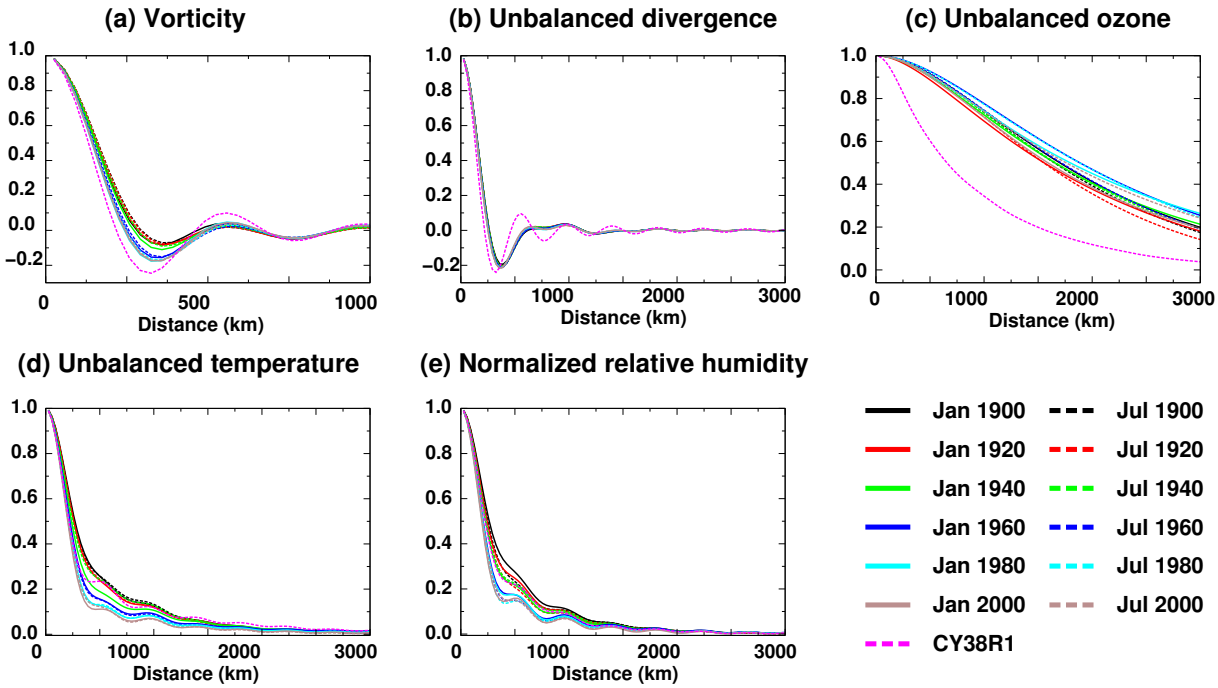

Figure 20: Background error horizontal correlations for (a) vorticity at model level 89, (b) unbalanced divergence at model level 89, (c) unbalanced ozone at model level 7 (about 50 hPa), (d) unbalanced temperature at model level 89, and (e) normalised relative humidity at model level 89.

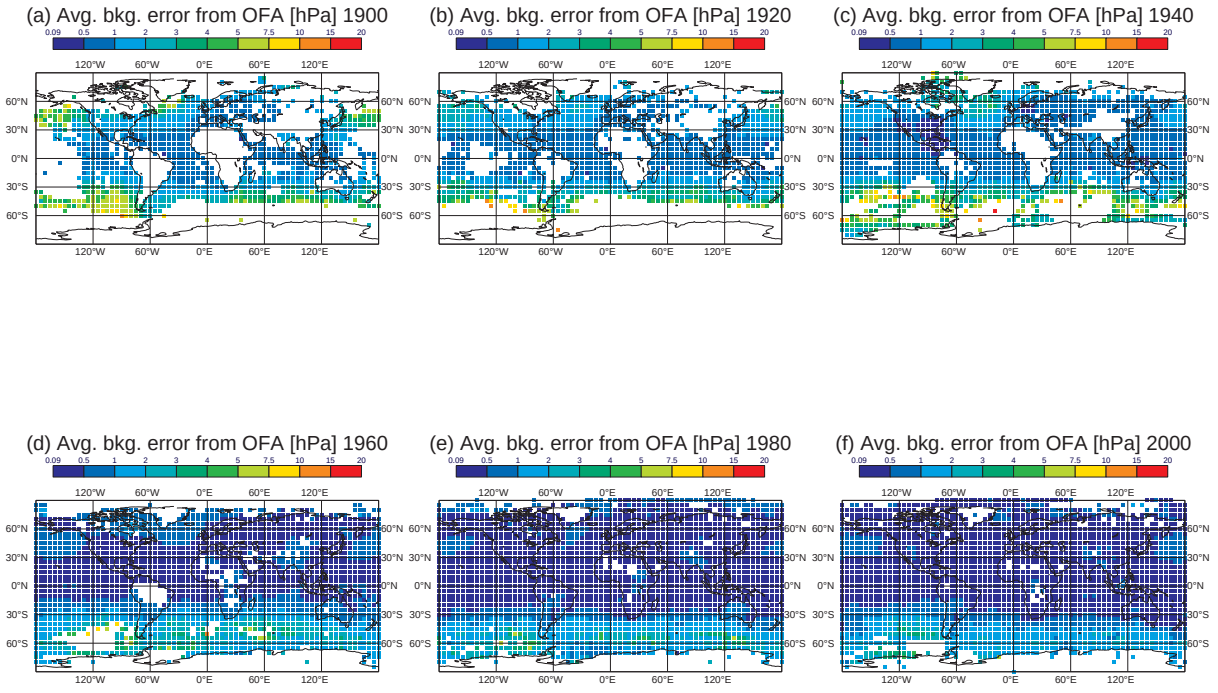

Figure 21: Average background errors of surface pressure, as found in the observation feedback

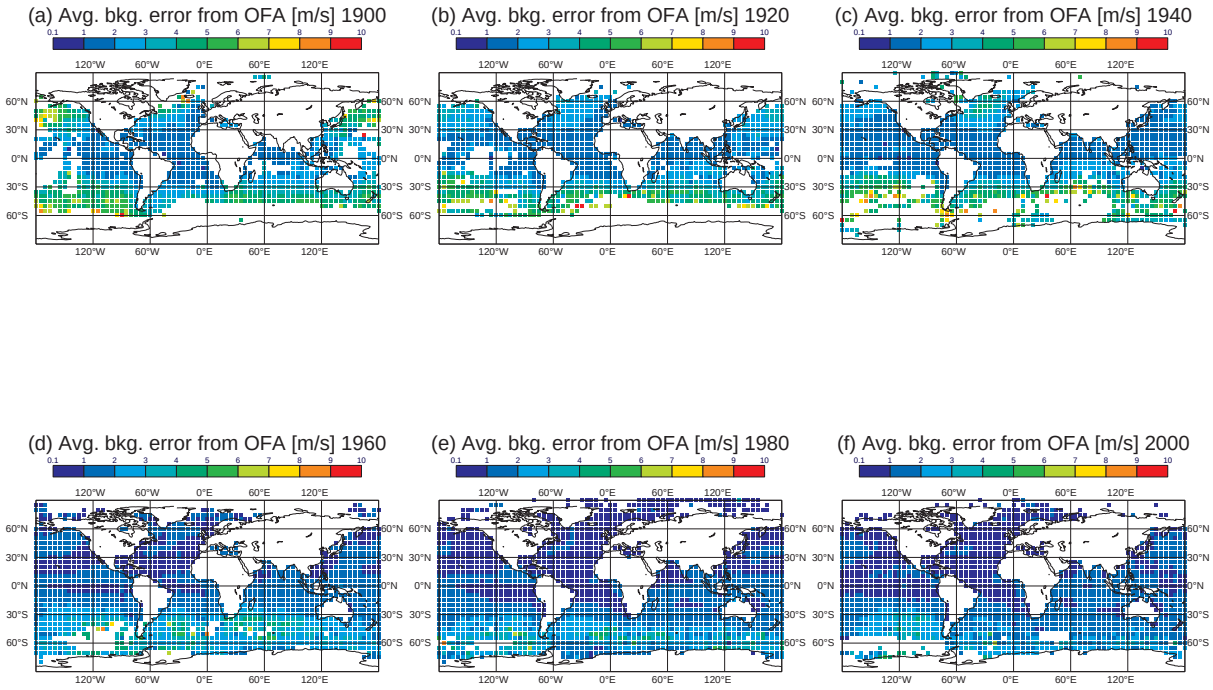

Figure 22: Average background errors of 10-meter zonal wind, as found in the observation feedback

shown in Figure 16, but rescaled by the EDA calibration and sampled at the observation level (*e.g.*, restricted horizontal coverage, and averaging from non-uniform temporal coverage). There are two important caveats when using these quantities. First, these quantities contain the scaled ensemble spread, which is not the exact background error used by the assimilation, but can be considered as a fair approximation as discussed in detail earlier. Second, these quantities are valid at the beginning of the analysis window. They do not describe the background errors for observations located towards the end of the window.

Figure 22 shows the average background errors stored in the OFA, for the 10-meter zonal wind. Again, these really describe the scaled ensemble spread at the beginning of the analysis window. Overall, the errors decrease as the observation coverage improves. In the years 1900, 1920, and 1940, one notes a good correspondence between unobserved areas in terms of surface pressure and wind (Figures 3(a,b,c), Figures 4(a,b,c); the Pacific Ocean and the Southern Oceans), areas of higher ensemble spread (Figures 21(a,b,c), Figures 22(a,b,c)), and areas of higher surface pressure and wind background errors (Figure 15(a,b,c), Figure 18(a,b,c)). Likewise, in the year 2000, the small area void of wind and pressure observations around the Southern Pacific (Figures 3(f), Figures 4(f)) is a region of locally higher surface pressure and wind ensemble spread (Figures 21(f), Figures 22(f)) and background errors (Figure 15(f), Figure 18(f)).

5 Initial evaluation of the assimilation performance

5.1 Forecast scores

The ERA-20C production includes the generation of 10-day forecasts, issued daily from the EDA control member analysis state at 00 UTC. The forecast scores can help quickly assess the instant accuracy level of a reanalysis product, as compared to its siblings. The forecast Root Mean Square Error (RMSE) of ERA-

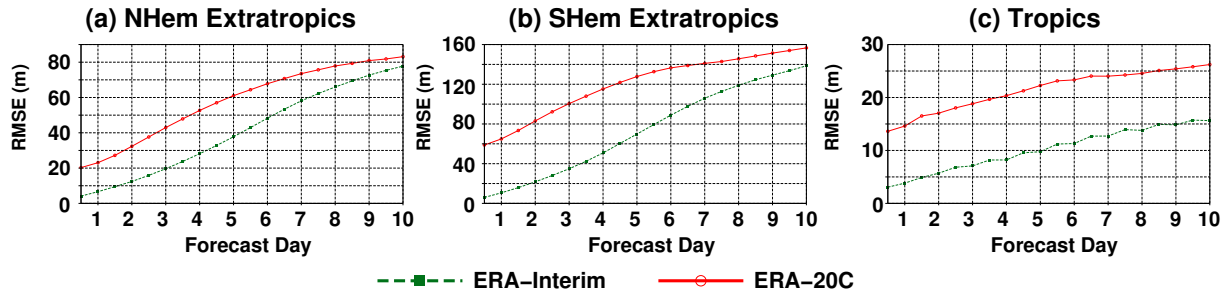

Figure 23: Root Mean Square Error (RMSE) of 500 hPa geopotential height forecasts between 1 July 2004 and 31 August 2004

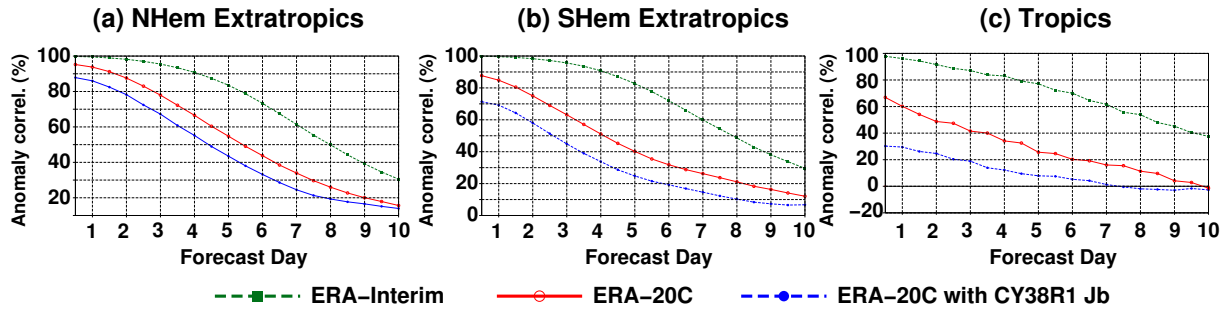

Figure 24: Anomaly correlation of 500 hPa geopotential height forecasts between 1 July 2004 and 31 August 2004

20C is compared with ERA-Interim in Figure 23(a), in the Northern Hemisphere extratropics, for geopotential height at 500 hPa in July and August 2004. The ERA-Interim forecasts are verified against their own analyses, whereas the ERA-20C forecasts are verified against ECMWF operations at the time. The 6-day forecasts of ERA-Interim are of similar accuracy to the 3- to 4-day forecasts of ERA-20C, which confirms the expectations mentioned in the introduction by restricting the assimilated observing system to surface-only. In the Southern Hemisphere extratropics, Figure 23(b) shows that ERA-20C presents much greater errors, around 100 m RMSE at day 3 instead of 45 m in the Southern Hemisphere extratropics. In the Tropics, owing to lower natural variability, Figure 23(c) shows smaller forecasts errors than in the extratropics. The ERA-20C reanalysis shows some forecast skill up to about day 6. Beyond that range, the forecast errors saturate to the model and natural variability.

In terms of anomaly correlation, Figure 24(a) confirms that the 6-day forecasts of ERA-Interim are of similar accuracy to the 3- to 4-day forecasts of ERA-20C. Note, all the anomaly correlations shown here are computed using the ERA-Interim climatology. For information, Figure 24(a) also shows the impact of the automatic adjustment of background errors, by comparing to an experiment which uses the default background errors as used by ECMWF operations (normally for a fully observed system). The impact of the adjustment of the background errors is a gain of 1 day in terms of forecast skill. In the Southern Hemisphere extratropics, Figure 24(b) shows that ERA-20C has generally lower skill. However, because of the nature of the verification metrics, the forecasts there do not appear twice as degraded as compared to the Northern Hemisphere, as they do in terms of RMSE. The 60% anomaly correlation level, usually considered as the limit of usefulness of a forecast, is reached at day 3, instead of day 4 in the Northern Hemisphere (compared to day 7 in both hemispheres for ERA-Interim). One likely explanation is that the ERA-20C climatology differs significantly from the ERA-Interim climatology in the Southern Hemisphere, so that large departures from the climatology give credence to ERA-20C forecasts and analyses. In the Tropics, Figure 24(c) shows that ERA-20C forecasts are usually below 60% anomaly correlation as early as day 1. However, in this region, the importance of using adapted background errors seems to be particularly significant.

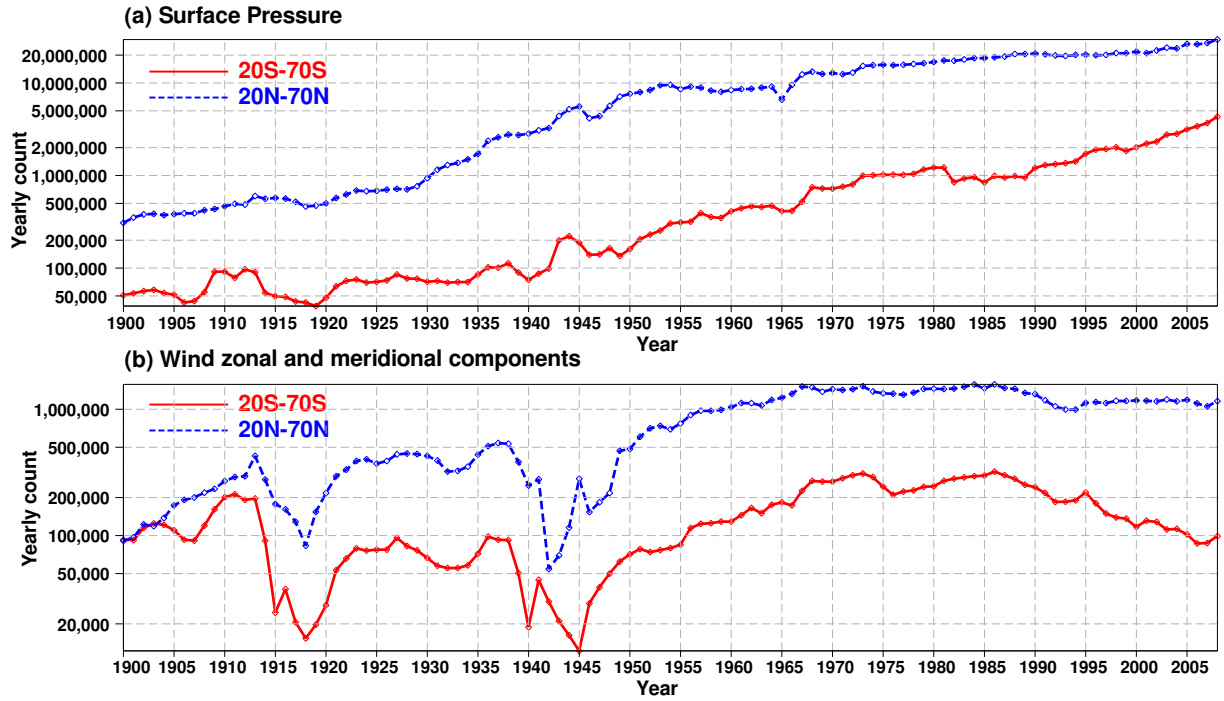

Figure 25: Yearly count of assimilated observations

5.2 Fit to assimilated observations

The number of assimilated observations in the EDA control member, already shown for the globe in Figure 8, is now discussed for the mid-latitude regions, defined here between 20° and 70° latitude in either hemisphere. Figure 25(a) shows that the number of assimilated surface pressure observations increases nearly exponentially during the 20th century, in both hemispheres, by a factor of 50 between the beginning and the end of the whole time period. The number of assimilated wind observations, however, seems to reach a maximum in the 1970s, before decreasing during the years 2000s.

Figure 26(a) shows the fit to assimilated surface pressure observations, for the background and the analysis. The statistics relate to the control member of the EDA. The RMS of background departures decrease over time, with the Southern hemisphere featuring systematically larger errors than the Northern hemisphere. Although the overall evolution may appear as a continuous improvement process, it is interesting to note a possible 'step' effect between World War II and the 1960s when the RMS become significantly smaller, both in terms of surface pressure and wind (Figure 26(b)). This concept of 'step' improvement may indicate that some key error structures get resolved much better, past a certain minimum of global observations. Around the 1960s the RMS of the background departures in the Southern Hemisphere reaches the level of RMS observed in the Northern Hemisphere in 1900, both for surface pressure and wind (4 hPa and 4 m/s, respectively). After that, the RMS of the background departures continues to drop in the Southern Hemisphere until the recent years. At the same time, in the Northern Hemisphere, the RMS of the background departures, below 2 hPa and 3 m/s since 1980, improves at a much reduced pace and nearly levels off, as if some level saturation had been reached. This conclusion, valid in the ERA-20C system, does not mean at all that further observations would not improve our understanding further, but probably rather that model and data assimilation improvements are needed before we can continue to retrieve further meaningful information from additional observations.

The RMS of the analysis departures are generally more stable, around 0.5–1.5 hPa for surface pressure (2 m/s

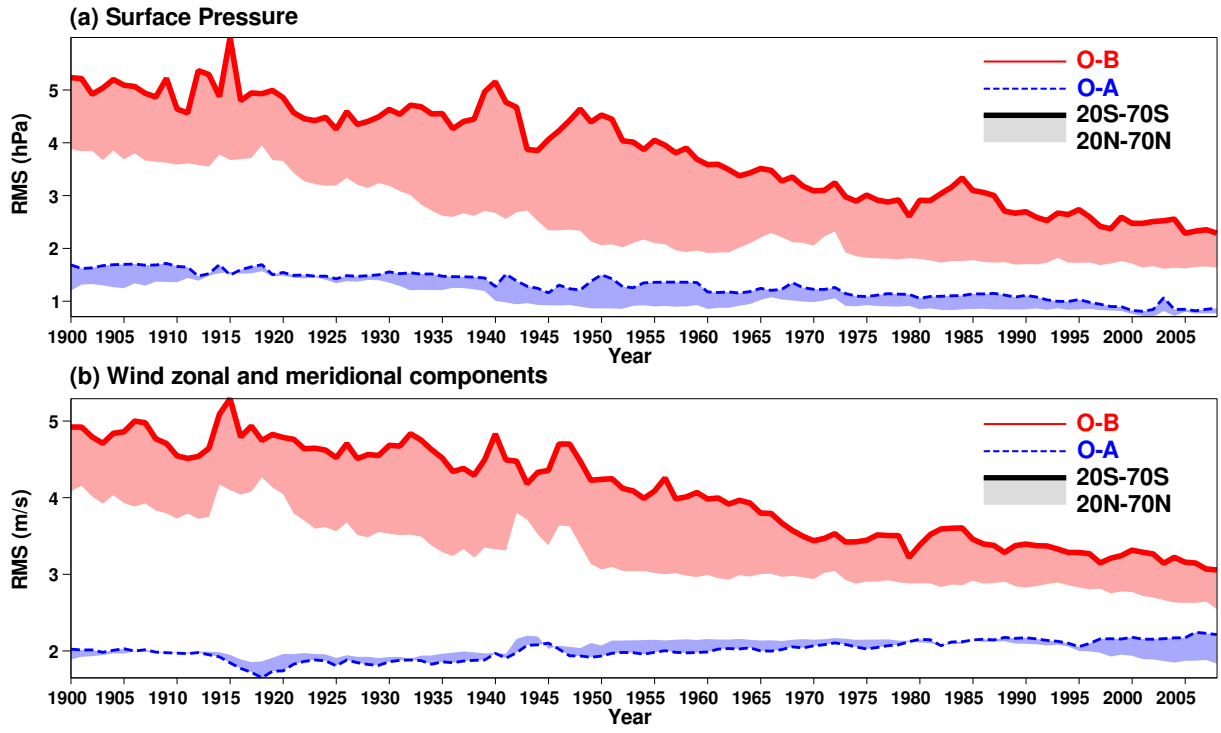

Figure 26: Monthly Root Mean Square (RMS) of observation minus background (O-B) and observation minus analysis (O-A) departures, for assimilated observations, per year. Thick line shows Southern Hemisphere mid-latitudes (20°S–70°S), and thin line shows Northern Hemisphere mid-latitudes (20°N–70°N); the region between the two curves is shaded to show the quantitative gap in quality between the two hemispheres

for wind), in the two regions considered here. As expected for a system which improves over time, the analysis departures also decrease over time for surface pressure. However, for wind, we note that for a good part of the century the analyses fit better the winds observations in the Southern Hemisphere than they do in the Northern Hemisphere. This result is contrary to background fits worse in the Southern Hemisphere than in Northern Hemisphere. This suggests that we over-fit the wind observations in the Southern Hemisphere, and probably also in the Northern Hemisphere. As pointed out before, the wind observation errors should probably have been larger than assumed. This suboptimality may then contribute to the counter-intuitive behaviour of analysis departures slightly increasing over time for wind, while they decrease as expected for pressure.

The assimilation statistics discussed so far describe the whole set of observations assimilated in the ERA-20C production experiments, regardless of the observation time of the day. However, with a 24-hour analysis window, the observations assimilated at the beginning of the window are compared to a background that is a short forecast integration, whereas the observations assimilated at the end of the window are compared with a background that is a longer forecast integration (see Figure 5). In fact, Figure 26 hides a great variability in terms of actual background departures at the observation level. To expose this complexity, it is instructive to study the background (and analysis) departures as a function of the time elapsed since the beginning of this window. This elapsed time ranges from 0 to 23 hours, where 0 refers to observations collected between 9 and 10 UTC. Figure 27 shows such statistics, binned in elapsed hours, by latitudinal bands in five columns (a–e), for selected years of production. Rows (1,2,3) describe statistics for surface pressure observations, and rows (4,5,6) describe statistics for wind component observations. The spikes in observation counts come from the observation reporting times, which are usually at UTC hours multiples of 3. This panel may also be of interest for users interested in data quality changes over the years. For example, the clustering mentioned earlier in terms of product quality with respect to the quality of the tropospheric circulation appears in Figures 27(b2,c2,d2,e2): years 1900, 1920, and 1940 are of similar quality, and years 1980 and 2000 are of much better quality, whereas the year 1960 appears of intermediate quality. In the Southern Hemisphere high latitudes, we note (as expected) the absence of observations in 1900 and 1920. From the OFA, we find that observations during 1940 over that area belong to ISPD 3.2.6 collection ‘Antarctic Expeditions’ (for about 98%), and ICOADS 2.5.1 collections ‘Antarctic Expeditions: Print./Published (held at Met. Office)’ and ‘US Merchant Marine’ (for the remaining 2%). As expected, Figure 27(a1) shows that the year 2000 is the best observed quantitatively in terms of surface pressure. One can determine from the OFA that many of these observations in 2000 come from the ISPD 3.2.6 collection ‘Federal Climate Complex Integrated Surface Database’.

In the Southern high latitudes, the spike in wind observations in year 1980 in Figure 27(a4) comes mostly from the contribution of ICOADS 2.5.1 collection ‘International Marine (US- or foreign-keyed ship data)’. From Figure 27(a5) we find that this year features one of the smallest background departures for wind over that region, around 4 m/s. This is in line with the map shown in Figure 22 where smaller background errors can be spotted in the Southern Pacific Ocean. This illustrates the importance of the data recovery of observations to improve regionally the quality of the reanalysis product.

The analysis fit to assimilated observations (or residuals), shown in Figure 27 rows (3,6), also exhibit a growing trend over the course of 24-hour analysis window. However, this error growth is much reduced, as compared to the growth observed for background departures. If those residuals are compared to the innovations, the reanalysis quality appears nearly constant during the analysis window.

Figure 28(a1) shows a map of the mean surface pressure background departures, after bias correction, but before assimilation, for all observations in 1900 reported at hours between 9 UTC and 15 UTC, *i.e.*, around 12 UTC. Over the years (from row 1 to 6), the geographical coverage improves, and the magnitude of the mean departures decreases. The columns (a–d) show the statistics binned by observation time. Columns (b–d) show respectively statistics for observation that belong to the 2nd, 3rd, and 4th 6 hour-intervals of the assimilation window (around 18 UTC, 00 UTC, and 06 UTC). A close inspection shows that the columns (a,c) feature mean,

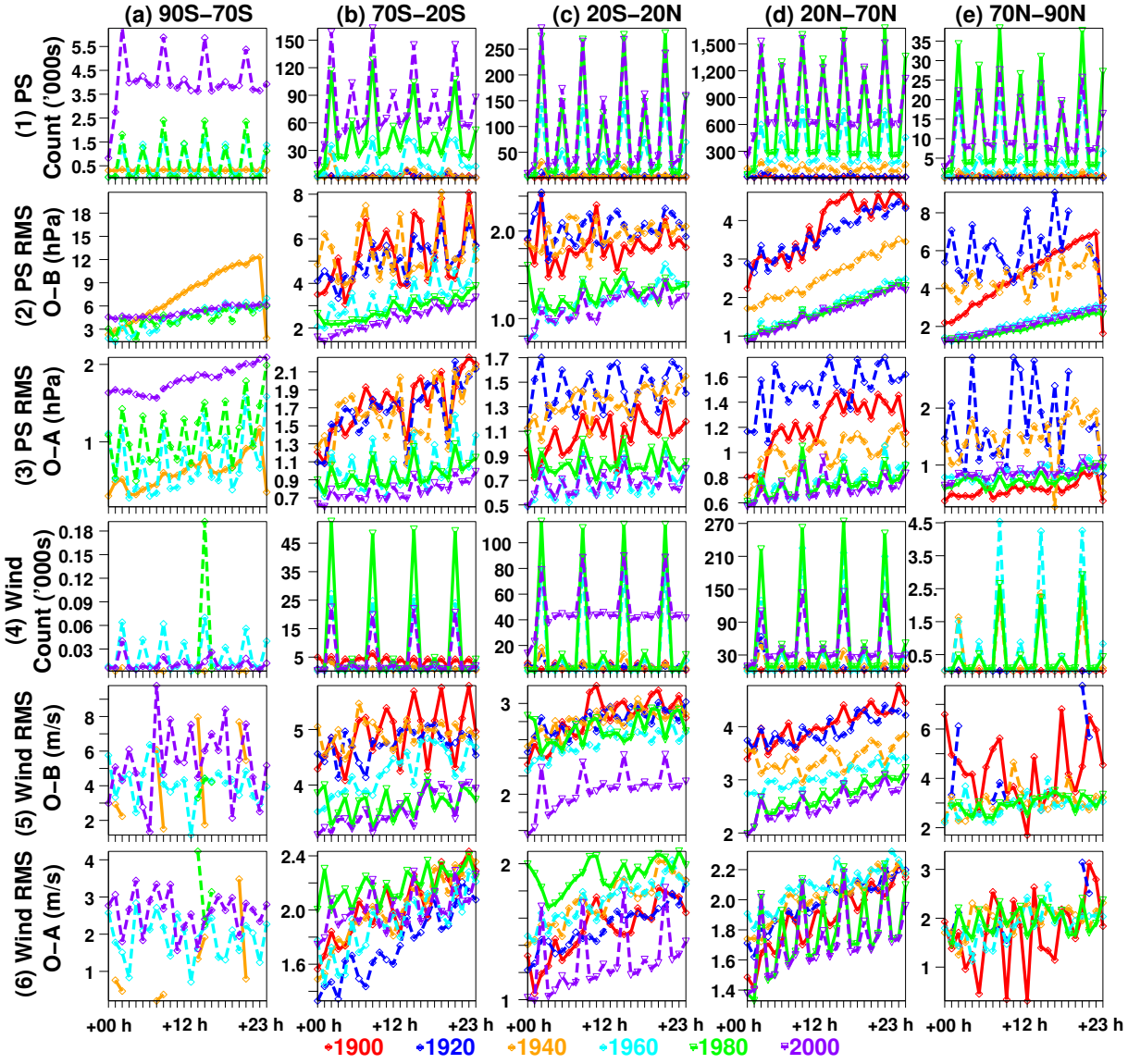

Figure 27: Data count (in thousands) in row (1), Root Mean Square (RMS) of observation minus background (O-B) in row (2), and observation minus analysis (O-A) departures in row (3), for assimilated surface pressure observations, as a function of observation time elapsed with respect to the beginning of the analysis window (from 0 to 23 hours), for selected years. Columns (a,b,c,d,e) show different latitudinal bands. Rows (4,5,6) as rows (1,2,3) except for wind component observations.

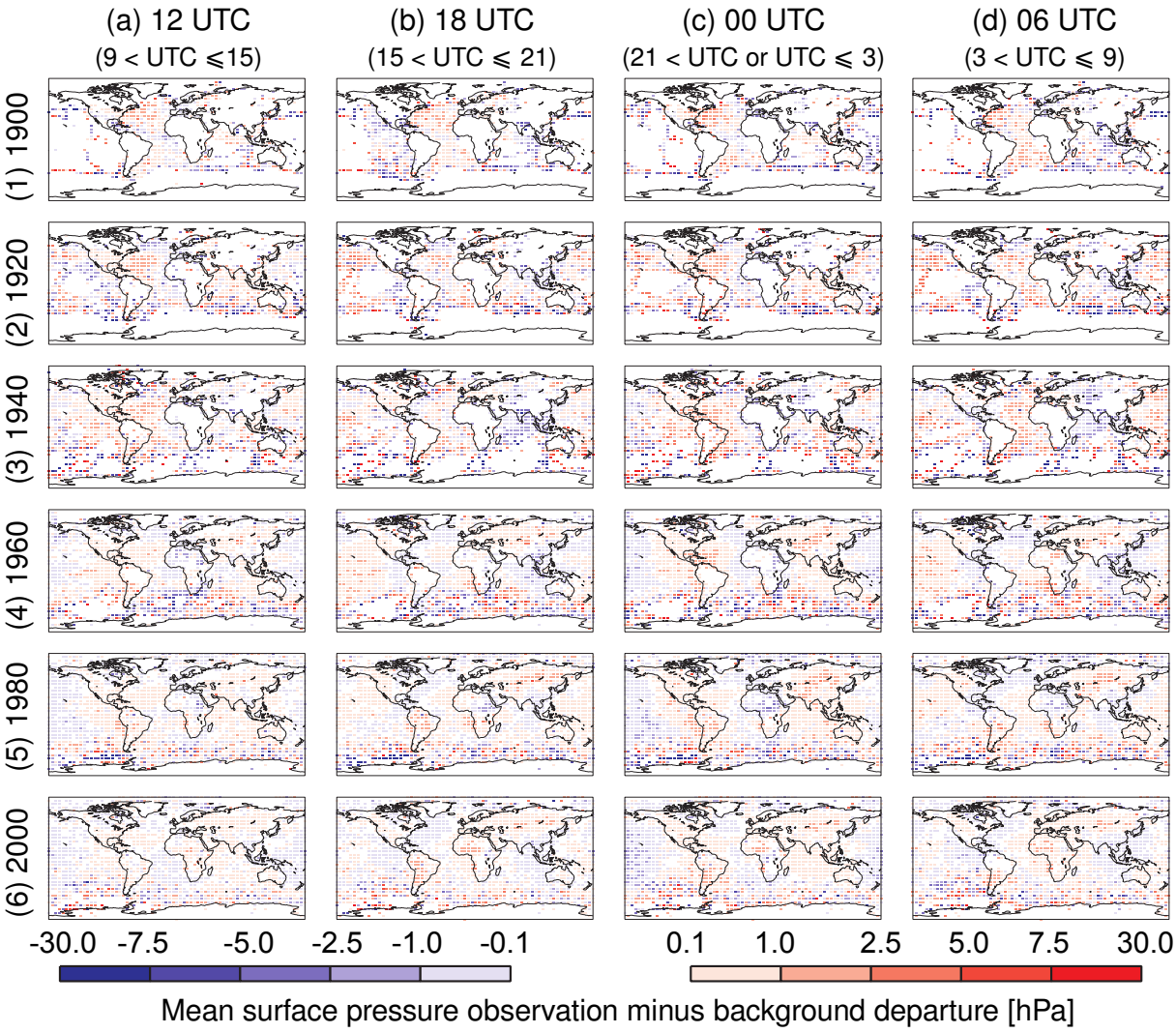

Figure 28: Mean surface pressure observation minus background departures, for selected years (rows 1 to 6), and as a function of time of day (columns (a) to (d), ordered by increasing observation time with respect to the beginning of the assimilation window at 9 UTC)

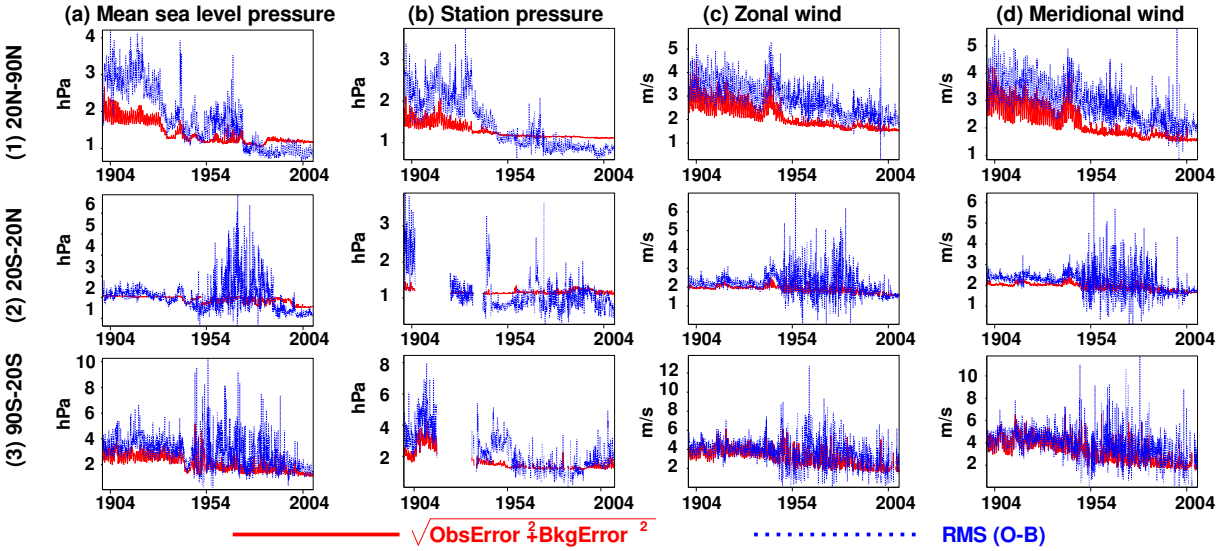

Figure 29: Monthly timeseries of Root Mean Square (RMS) of observation minus background (O-B) departures and predicted departures (from the square root of the sum of assumed observation error squared and background error squared), for assimilated observations of (a) surface pressure reported at mean sea level, (b) surface pressure reported at station, (c) zonal wind, (d) meridional wind. Only observations in the first 90 minutes of each assimilation window are considered (because the diagnostic for background error available in observation space is only valid at the beginning of the assimilation window). Rows (1,2,3) show different latitudinal bands

consistent, patterns of positive departures over Central America, surrounded by patterns of negative departures located over the tropical Pacific and the tropical Atlantic. In columns (b,d), these patterns are less obvious, but appear to be shifted by 90 degrees in longitude, with negative departures over Central America, and usually positive departures over the tropical Pacific. These systematic departures point to a model bias, namely an under-representation of atmospheric tides, discussed in more details later.

The observation feedback contains diagnostics of background errors for all observations assimilated in ERA-20C. These estimates, as discussed earlier, are strictly valid at the beginning of the window. Combined with the assumed observation error (also stored in the observation feedback), one can form statistics of the expected (or predicted) departures at the beginning of the assimilation window. They can then be compared with observed Root Mean Square of background departures at the beginning of the window. Figure 29 shows both sets of quantities, for the first 90 minutes of the assimilation window. For surface pressure observations reported at station level (hence from land stations) in the Northern Hemisphere extratropics, Figure 29(b1) shows that we under-estimate the departures at the beginning of the century. Either the observation and/or the background errors are too small there. Based on the section above about observation errors, our surface pressure observation errors should be larger than assumed at the beginning of the century for land stations, at 1.6 hPa instead of 1.1 hPa. At the end of the century, the departures are over-estimated compared to actual departures. Here again, this is consistent with earlier findings that the observation errors should have been lower than assumed at the end of the century, at 0.8 hPa instead of 1.1 hPa. Similar conclusions can be drawn in the Tropics and to a lesser extent in the Southern Hemisphere extratropics (Figure 29(b2,b3)).

For surface pressure observations reported at sea level (Figure 29(a1–a3), the situation is slightly more complex, because it mixes land and ship observations. The land observations come from stations that only exchange surface pressure at sea level, and not surface pressure at the station level. This situation is more prevalent at the beginning of the century. At that time, the land stations are nearly all exclusively located in the Northern Hemisphere (see map of data coverage in 1900 shown in Figure 3(a)). Over the years, the land stations

gradually report also observations at station level. Consequently, in the Northern Hemisphere extratropics, we find a combination of two elements: the improper representation of land station errors for pressure (too low at the beginning, too high at the end), and the evolution of observation errors for (mostly) ship observations. In the section above about observation errors, we find that the pressure observations from ships are also under-estimated at the beginning of the century (probably around 2.0 hPa instead of 1.5 hPa), and over-estimated at the end of the century (probably around 1.2 hPa instead of 1.5 hPa). The remaining discrepancies in Figure 29(a2,a3) may be explained by suggesting that our background errors are under-estimated in the Tropics in the early years, and over-estimated in the Southern Hemisphere in the end. Note, all this discussion assumes that the background errors reported in the observation feedback are exactly as used by the assimilation, which is not completely correct as discussed in section 4.

For the wind zonal and meridional components in Figure 29(c1–d3)(d1–d3), we find an overall tendency to under-estimate the departures in all areas and at all times. From earlier findings, wind observation (mostly from ships) errors are probably around 2.0 m/s, instead of 1.5 m/s. Yet, these differences, even if in the right direction, are of insufficient magnitude the gap between expected and observed departures. One element to remember is that the background errors are computed at the resolution of the analysis, from forecasts differences all at the same horizontal resolution of 210 km. Nature and actual observations contain small-scale features which are not represented at such scales. At a local site, this source of uncertainty is a representativeness error and should formally be assigned to the observation part in the assimilation scheme. However, on a global scale, these errors probably have similar signatures between sites. The observation error computation used earlier (Desroziers *et al.*, 2005) separates between error sources assuming they have distinct structures. The method may then probably assign this wind representativeness error to a background error. Also we note that such discrepancy is not visible in surface pressure, which by nature represents synoptic scales, probably fairly well captured at the resolution used in ERA-20C. So it could thus appear that the observation errors (notably the representativeness part) are probably severely under-estimated for the winds at the resolution of ERA-20C. We note that more work would be required to substantiate this hypothesis.

A general finding from Figure 29 is that the observation error assignment in ERA-20C is not perfect, but the background errors seem to behave qualitatively satisfactorily across geophysical variables, areas, and over the years. We also note that the background error estimation scheme represents some of the seasonal variability found in the actual departures. However, it fails to capture the magnitude of this winter-to-summer cycle. It may be that the 90-day time span for computing the background errors is too long and also causes a lag. The 90-day sample may need shortening to capture the extremes of each season.

5.3 Analysis increments

The analysis increment is defined as the state difference analysis minus background, both being valid at the same time. In the strong-constraint 4D-Var, the increment is determined at the beginning of the window. One can compute, *a posteriori*, the effective increment at any time of the analysis window. For the present discussion we consider first the increment at the beginning of the analysis window. It is namely the difference between the 9 UTC analysis and the 3-hour forecast issued from the 6 UTC analysis determined by the previous assimilation cycle (see Figure 5).

Figure 30(a) shows the global equal-area mean increment for surface pressure. It is generally within ± 0.02 hPa, and usually slightly negative after 1960, between 0 and 0.01 hPa. The RMS of the increments, in Figure 30(b), slightly decreases over time, presumably because the analysis starts from a background which is already close to the observations. In the years 2000s, we see more variability in the RMS of the increments. This point is yet unexplained.

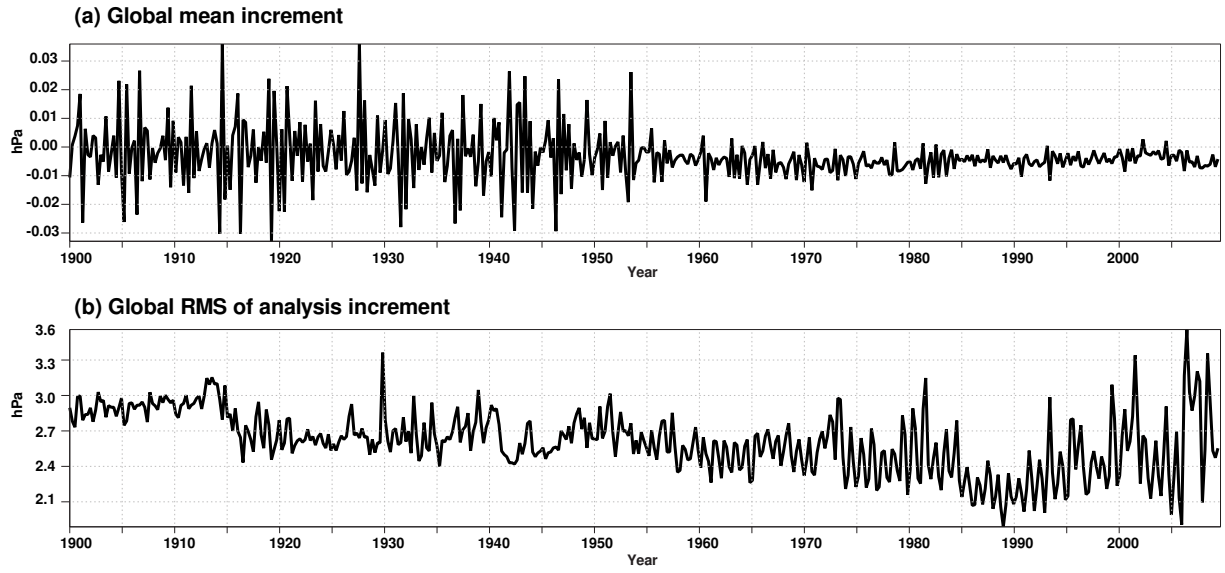

Figure 30: Mean (a) and RMS (b) of the surface pressure analysis increment at the beginning of the analysis window

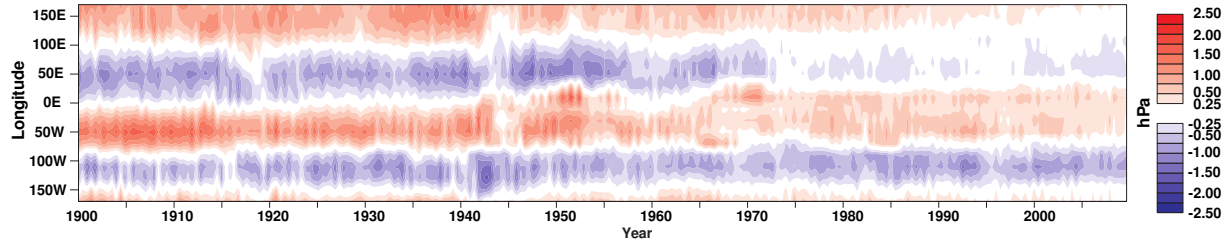

Figure 31: Longitude (y-axis) versus date (x-axis) average surface pressure analysis increment at the beginning of the analysis window, for the latitude band 10°S–10°N

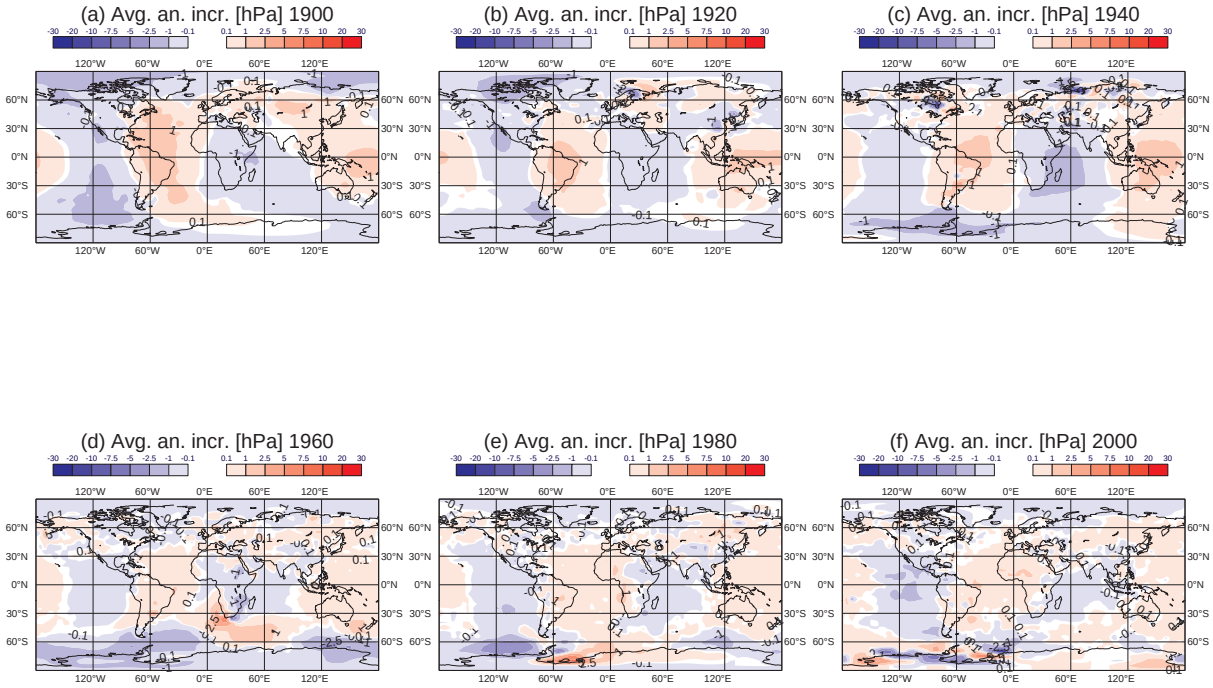

Figure 32: Mean surface pressure analysis increment at the beginning of the analysis window, for selected years

The global statistics hide local effects. Looking at the increments taken at the same time of the day, but averaged over a given latitudinal band, one can visualise the systematic tendency of the analysis, regionally, by plotting as a function of longitude. Figure 31 shows such statistics (longitude on the vertical axis, versus time on the horizontal axis), for the region located around the Equator at latitudes 10°S–10°N. Overall, one can note a wave-2 pattern, reminiscent of Figure 28. As noted earlier, it was known when experimenting with ERA-20C that the 1 hour-time-step employed for the model was insufficient to properly represent the atmospheric tides, and this effect would thus be systematically corrected by the assimilation, with the help of the observations. However, as a growing number of observations is assimilated, the analysis contains a state with less remnants of this insufficient representation of tides, and the background that is issued from this analysis is less biased; consequently, the mean increment gets correspondingly smaller over time. The magnitude of this under-representation of the atmospheric tides is probably around ± 1 hPa at the beginning of the century, and probably around ± 0.7 hPa at the end.

To confirm this conjecture, we evaluate the magnitude of the semi-diurnal tide (S2) in the forecast. It is the sum of the surface pressure forecast fields valid at 0 UTC and 12 UTC, minus the sum of the fields at 6 UTC and 18 UTC, divided by 4. We repeat the same computation for the analysis fields. We then define the S2 error as the difference between the S2 forecast and the S2 analysis. This error, shown in Figure 33, is indeed confirmed to be around 1 hPa at the beginning of the century, and to decrease over time.

For completeness, the maps of increments at all times within the analysis window are shown in Figure 34. The wave-2 pattern of the mean increments moves with areas that are sun-lit throughout the day. This is in line with the hypothesis of insufficient representation of atmospheric tides in the atmospheric model.

Timeseries of the mean vertical temperature increments at the beginning of the analysis window are shown in Figure 35. The increments are usually within ± 0.08 K, or otherwise positive in the mid-troposphere. This

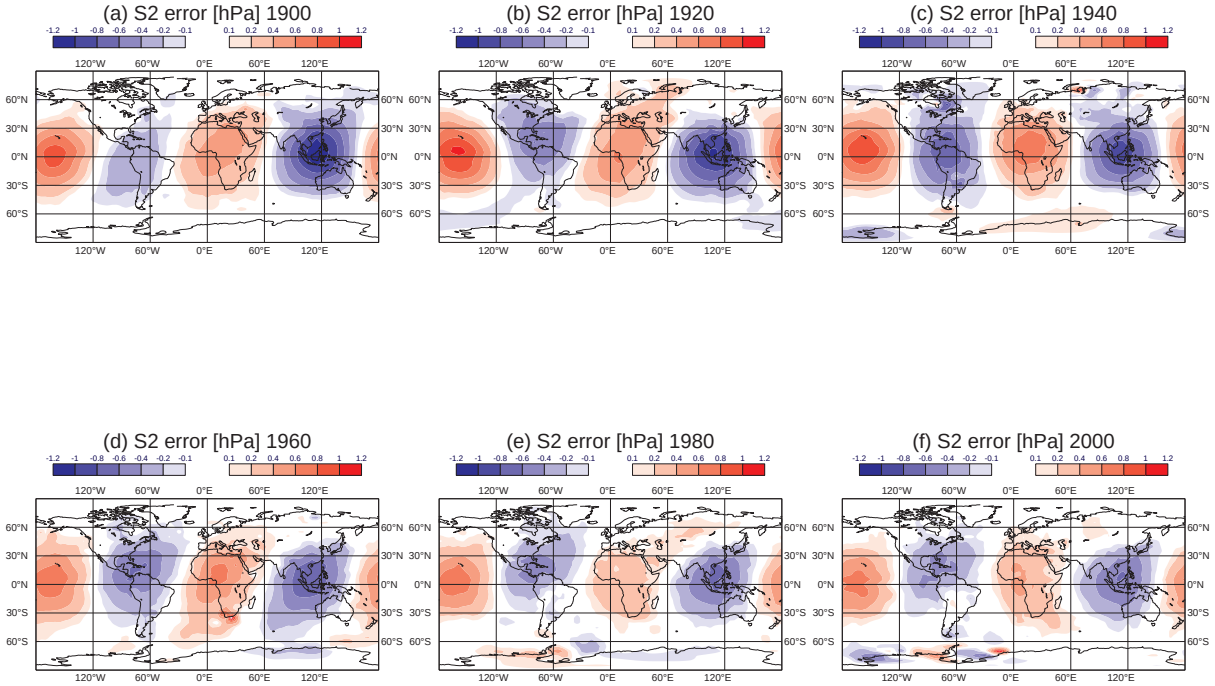

Figure 33: Mean error in semi-diurnal tide (S2) representation of surface pressure in the background, using the analysis as a verification, for selected years

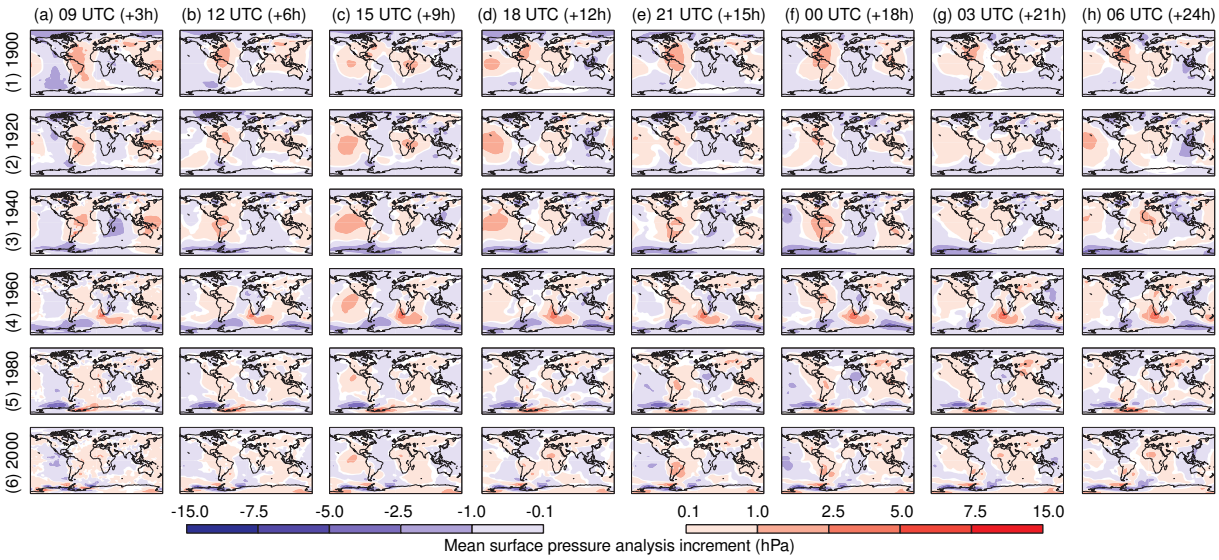

Figure 34: Mean surface pressure analysis increment, for selected years (rows 1 to 6), and as a function of time within the assimilation window (columns (a) to (h), ordered by increasing background integration time, in parentheses)

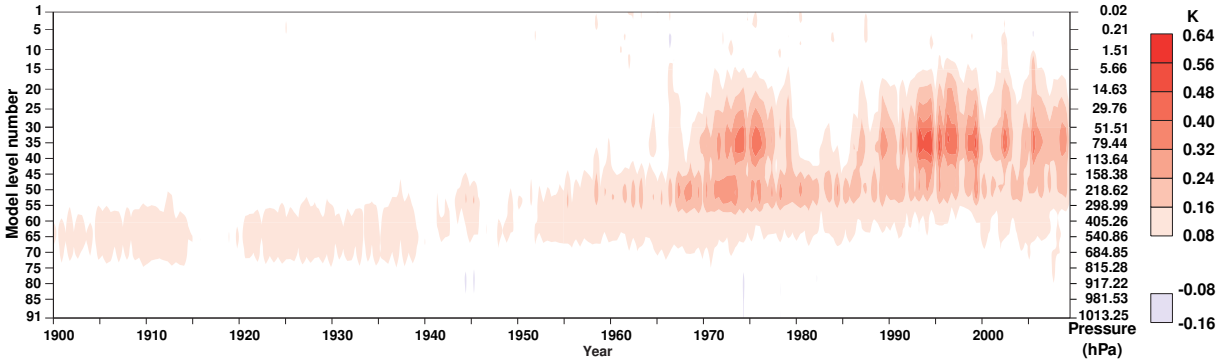

Figure 35: Model level number (y-axis) versus date (x-axis) equal-area global average of the temperature analysis increment at the beginning of the analysis window. Right-hand-side y-axis indicates lower boundary of the corresponding pressure layer, assuming for a surface pressure of 1013.25 hPa

means that the analysis warms slightly the mid-troposphere as compared to the background. This is line with the earlier indication that, globally, the mean pressure in the analysis is slightly smaller than the background surface pressure at 9 UTC. Towards the end of the century, the mean increments become larger, but their location is overall higher in the troposphere, which means their impact in terms of surface pressure is reduced. The displacement of the mean increments is related to changes in the structure functions of the background errors discussed earlier. As the background errors in temperature decrease over time in the lower troposphere (Figure 19(d), model levels 50–91), the data assimilation system, which aims to produce negative increments in surface pressure to fit the observations, cannot warm as much the lower tropospheric temperatures, because of increased background constraints there. Its solution is then to warm the atmosphere higher up, where the background constraints are unchanged (model levels 50 and above), but with a greater intensity, in order to produce the same net effect in terms of surface pressure increment. We thus observe global mean analysis increments which can be as high as 0.5 K in the stratosphere, particularly at the end of the century.

As the analysis increments in temperature displace over time, they affect the trends to a point that may render the final product unfit for the purpose of trend assessment. This vertical propagation of increments over time was not foreseen when the ERA-20C system was developed. It was yet quickly noted as production started, but no immediate solution was available to remediate. In an ensemble-Kalman-filter type of reanalysis (Compo *et al.*, 2011), the vertical localisation of increments limits this effect by preventing any update to the model levels located far away from the surface observations. In a variational system where the analysis solution is found globally, it is more difficult to enforce a vertical separation. This difficulty should be a prime focus of efforts in future developments.

5.4 Bias correction

The surface pressure observations are bias-corrected (Section 2.5). The bias corrections are kept near-zero in the first six months of each production stream. After that, the bias corrections evolve in each ensemble member, so as to effectively absorb, with some lag, the mean difference between observation and background. Figure 36(a) shows that the global average of all corrections is within ± 0.1 hPa. This number is much smaller than the RMS of observation minus background departures, which are in the range 2–4 hPa according to Figure 26(a). The figure also shows that the global estimates of mean bias are nearly continuous at the stream boundary of 1920, but there are visible discontinuities in 1940, and less so in 1980 and 2000.

One additional diagnostic is to consider the global RMS of the bias corrections. The larger the RMS, the more

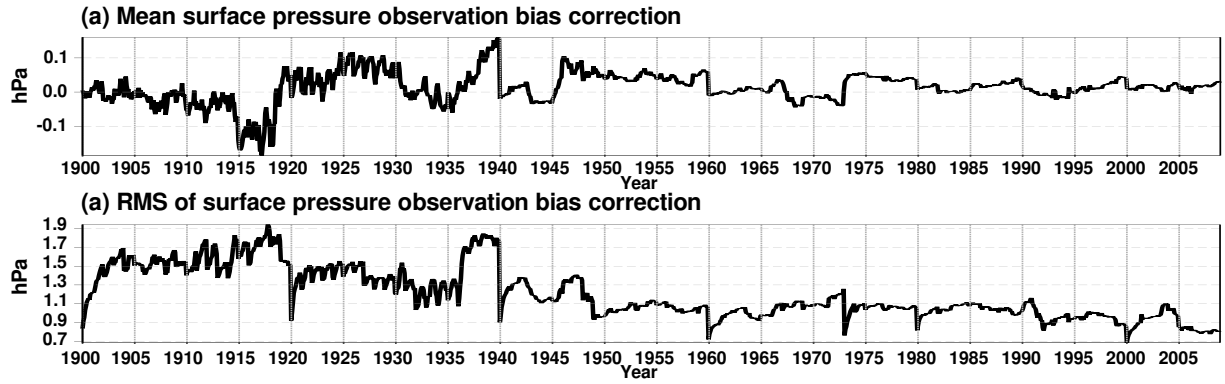

Figure 36: Timeseries of monthly mean observation bias correction of surface pressure

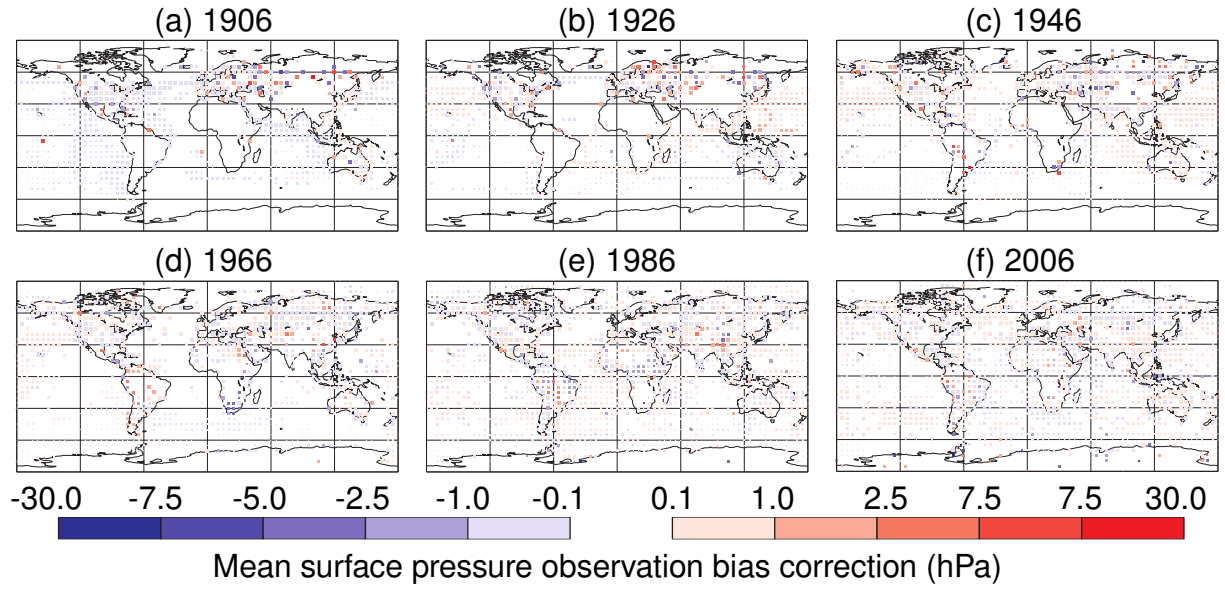

Figure 37: Mean observation bias correction of surface pressure, for selected years. For a fair representation, in each map, the sizes of the symbols reflect the number of observations at a given point (locations marked with dots indicate fewer observations than at the locations marked with larger squares)

work is done by the bias correction scheme, globally. Figure 36(b) shows that the global monthly RMS of the bias corrections is around 1.5 hPa in the first production stream (0.9 hPa in the last production stream). These numbers are much smaller than the global RMS of the analysis increments, suggesting that the surface pressure information extracted by ERA-20C from observations is mapped for the most part into updates of the analysis state, whereas a smaller portion is mapped into updates of the bias correction. Of importance, the convergence of the bias correction estimates seems to only happen between 2 and 4 years after the beginning of each run. Note, the production plan, laid out earlier, envisaged a 2-year overlap between the streams, with only 1 year of spin-up. The preliminary results discussed here suggest that it would be advisable to reduce the response time of the surface pressure variational bias correction in a repeat of ERA-20C.

Because it appears that the bias corrections are not yet stable in the first production years of each stream, we now consider a series of years more into the production. Figure 37(a) shows a map of the mean surface pressure bias corrections for the year 1906. Because the amount of observations can vary considerably from one grid-point to the next, some points could appear unfairly represented. In order to avoid this, we use different symbol

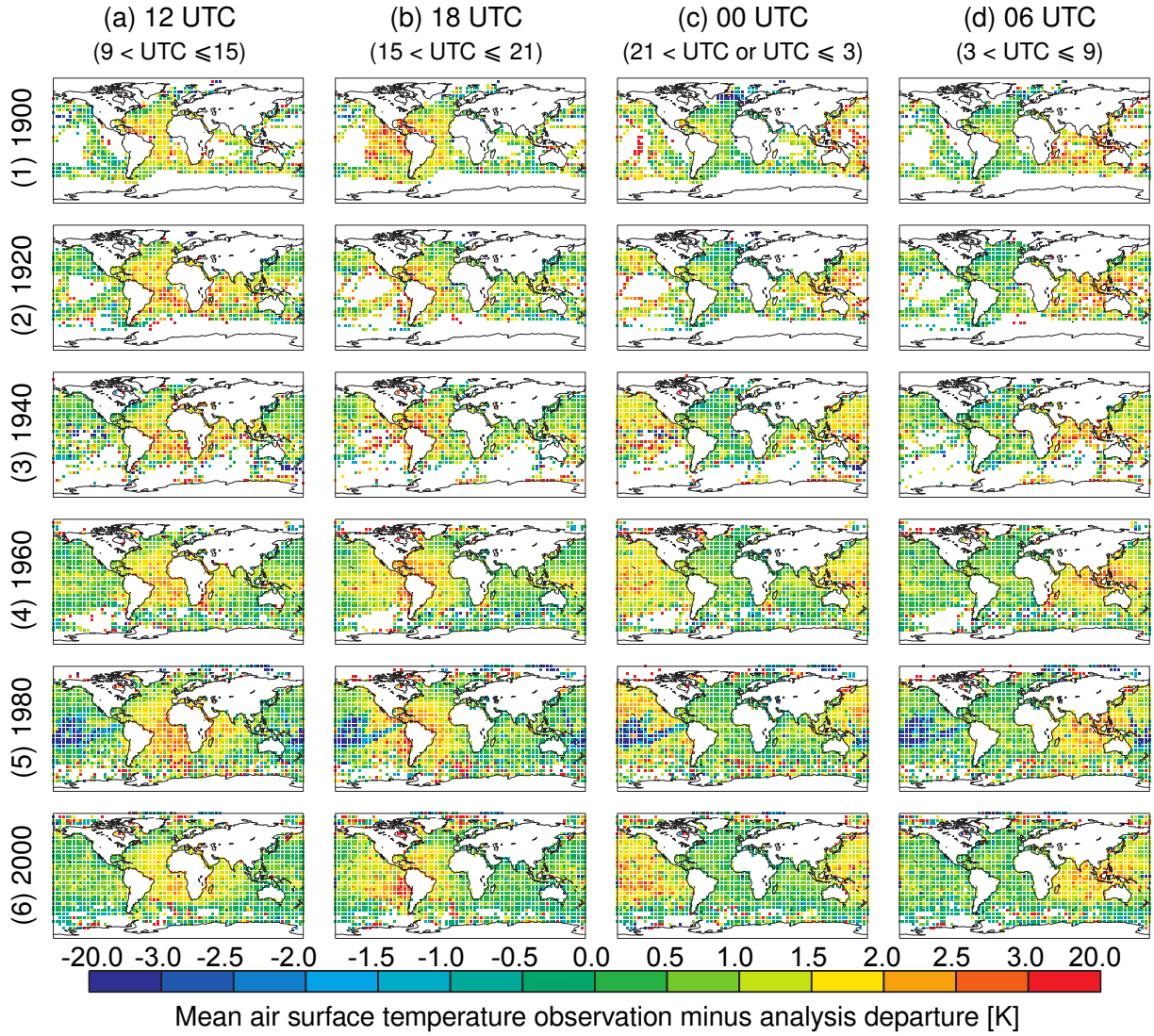

Figure 38: Mean surface air temperature observation minus analysis departures over open ocean, for selected years (rows 1 to 6), and as a function of time of day (columns (a) to (d), ordered by increasing observation time with respect to the beginning of the assimilation window at 9 UTC). Note that these observations are not assimilated

sizes so that, in each map, the dot sizes depend on the relative observation amount. In 1906, we note large areas in the Southern Oceans with negative corrections; this is probably because the same, few, ships reported from these tracks. As the number of reporting platforms increases over the century, the maps appear less structured, with a mix of positive and negative values, of smaller intensity. This is consistent with Figure 36(b) which shows that the bias corrections in the later streams present smaller RMS than those in the earlier streams.

5.5 Fit to withheld or independent observations

The ICOADS 2.5.1 observational input dataset contains more geophysical variables than just surface pressure and wind. It also contains, in particular, air surface temperature. As such observations are submitted for assimilation, they are compared with the background, before being rejected by the blacklist. After the assimilation is complete, these observations are also compared with the analysis state. The resulting background and analy-

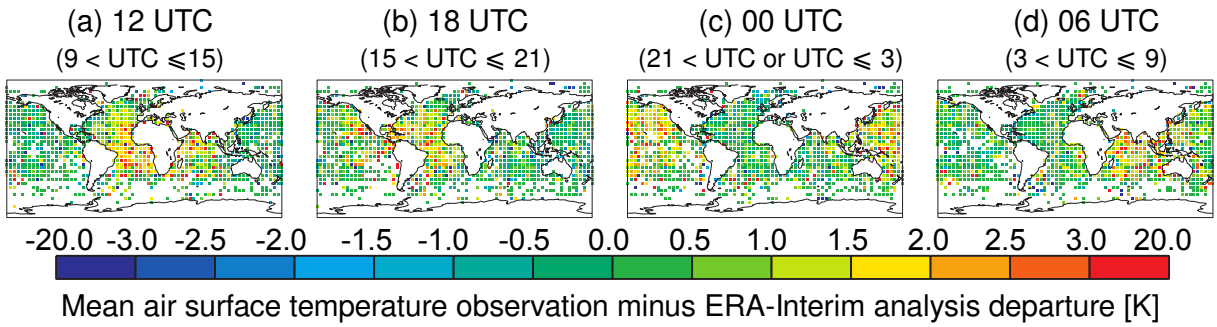

Figure 39: Mean surface air temperature observation minus ERA-Interim analysis departures over open ocean, for the first day of each month in the year 1980 as a function of time of day (columns (a) to (d), ordered by increasing observation time with respect to the beginning of the assimilation window at 9 UTC). Note that these observations were not assimilated

sis departures to observation are saved in the observation feedback archive. Consequently, the ERA-20C OFA contains an air surface temperature verification dataset, using independent observations that are not assimilated. Figure 38 shows average differences between ICOADS 2.5.1 air surface temperature and ERA-20C analyses, for selected years, as a function of time of day, considering only locations where the model land-sea mask indicates open ocean (land fraction less than 1%). This restriction is applied to prevent considering observations that may have been reported at incorrect locations over land, and also to avoid considering locations that are actually over ocean but appear over land at the ERA-20C coarse horizontal resolution. The mean differences are mostly within ± 1 K (green areas). We also note a large pattern of yellow to red departures, mostly visible in the tropical belt. This pattern is centred around 0° longitude at 12 UTC. It travels westward during the day. This indicates that at the local noon air surface observations over ocean are warmer than ERA-20C. This could either reflect an observation bias, as the measuring platforms are heated by the sun, and the thermometers are more influenced by the platform's temperature than solely by the air. It could also be a model bias, with an insufficient representation of the diurnal cycle, underestimating the daily maximum temperature.

For comparison, Figure 39 shows similar quantities for the year 1980, but from the ERA-Interim observation feedback where air surface temperatures above oceans were not assimilated. Note, because the ERA-Interim observation feedback is much more difficult to access than that of ERA-20C, we had to restrict the map to only consider data from the first day of each month in the year 1980. Consequently, the comparison between Figures 38 and 39 is with respect to a different pool of observations, even though a large fraction of ICOADS 2.5.1 observations may already have been among the observations available for ERA-Interim. In any case, we also find, based on Figure 39, that there is a warm temperature bias, when compared to ERA-Interim analyses, for observations over ocean at the local noon. Because the model version is quite different between ERA-Interim and ERA-20C, this result points to a problem of warm bias in temperature observations reported around midday from ships.

Figure 40 shows that the RMS of differences between ICOADS 2.5.1 air surface temperatures and ERA-20C is usually less than 2 K, except for a large set of differences in 1980. This corresponds somewhat with another feature shown in Figure 38, row (5), with a pool of cold biases in the year 1980.

Separating the observational data that form ICOADS 2.5.1 by collection, we display in Table 2 observational statistics from each collection for the tropical latitude band. We expect for this region a fairly low variability in terms of temperature, as compared to global statistics. One collection, 'NODC/OCL World Ocean Database (WOD, and formerly World Ocean Atlas, WOA)', stands out from the others, as presenting the lowest average temperature and the largest variability. Other collections, such as 'US Natl. Centers for Environmental Pred. (NCEP) Ship Data' and 'US Air Force Global Weather Central (GWC)', contain data with larger variability than in the other collections. Removing these three collections from the comparison with ERA-20C, we show

ICOADS 2.5.1 collection		Observation statistics		
Number	Name	Count	Mean	Std. dev.
145	PMEL (Daily) Equatorial Moorings and Island Stations	306	298.3 K	2.4 K
223	UK Met. Office Main Marine Data Bank Selected Ships	41	299.8 K	2.2 K
229	UK Met. Office Main Marine Data Bank British Navy Ships	79	300.9 K	1.4 K
254	UK Met. Office Main Marine Data Bank Int. Maritime Met. (IMM) data (foreign or unknown origin)	5,640	299.8 K	2.3 K
732	Russian Marine Met. Data Set (MORMET) received at NCAR	20,298	300.1 K	3.6 K
735	Russian Research Vessel digitization	11,239	300.4 K	1.8 K
761	Japanese Whaling Ship Data (CDMP/MIT digitization)	11	300.6 K	2.6 K
780	NODC/OCL World Ocean Database (WOD) (and formerly World Ocean Atlas, WOA)	6,675	277.2 K	10.9 K
849	First GARP Global Experiment (FGGE)	87	299.5 K	4.3 K
850	German FGGE	40	300.0 K	2.2 K
888	US Air Force Global Weather Central (GWC)	12,173	299.4 K	5.7 K
889	Autodin (US Dept. of Defense Automated Digital Network)	5,345	300.4 K	4.3 K
892	US Natl. Centers for Environmental Pred. (NCEP) Ship Data	62,301	299.7 K	5.6 K
900	Australian	27	300.2 K	1.8 K
926	International Maritime Meteorological (IMM) Data	128,127	299.9 K	2.4 K
927	International Marine (US- or foreign-keyed ship data)	40,008	300.1 K	2.6 K

Table 2: Summary of ICOADS 2.5.1 air surface temperature data available for year 1980 in the region 20°S–20°N

in Figure 40 the RMS of differences between ICOADS 2.5.1 air surface temperatures and ERA-20C. We find that now the RMS is less than 2 K in 1980. This short discussion illustrates the importance of keeping the traceability to the original record sources within the observation feedback. This will assist its exploitation by users, and enable future improvements to the input dataset so that either the causes of errors are found, or the data are flagged appropriately.

5.6 Case studies: Great Storm of 1987, Europe and Great Blizzard of 1899, U.S.A.

Two case studies are investigated in this report: an extreme low pressure system and an extreme high pressure system. On 16 October 1987, a storm hit Southern England and France, causing great damage, and with little forecast notice regarding the storm location (Morris and Gadd, 1988). This meteorological case is sometimes seen as a tracker for improvement in representation of extreme weather between generations of reanalyses (Dee *et al.*, 2011). Figure 42 shows the short-term ECMWF operational forecasts for 16 October 1987, 00 UTC. Note, 48-hour and 24-hour forecasts are missing, because at the time the ECMWF operations only ran daily at 12 UTC. During the hours preceding the storm, the operational forecasts contain information about a great storm in preparation, possibly over Scotland, but the location is not consistent between the forecasts. Even 12 hours before the event, the storm location is a few hundred kilometres away from its analysed position. The ERA-15 reanalysis (same horizontal resolution as operations in 1987, about 190 km) and the ERA-40 reanalysis (horizontal resolution 125 km) show improved forecasts at 24 hours, but are yet far from the analyses. The ERA-Interim reanalysis (horizontal resolution 80 km) and the ERA-20C reanalysis (horizontal resolution 125 km) 24-hour forecasts appear as the best of the lot shown here. The low in ERA-20C 24-hour forecast is actually the closest to the analysis, both in location and intensity, 1 hPa deeper than ERA-Interim. The 48-hour forecast from ERA-20C also appears more accurate than the other products shown here, even though the low is not deep enough and located too much South. We remind that all the products (other than ERA-20C) assimilated

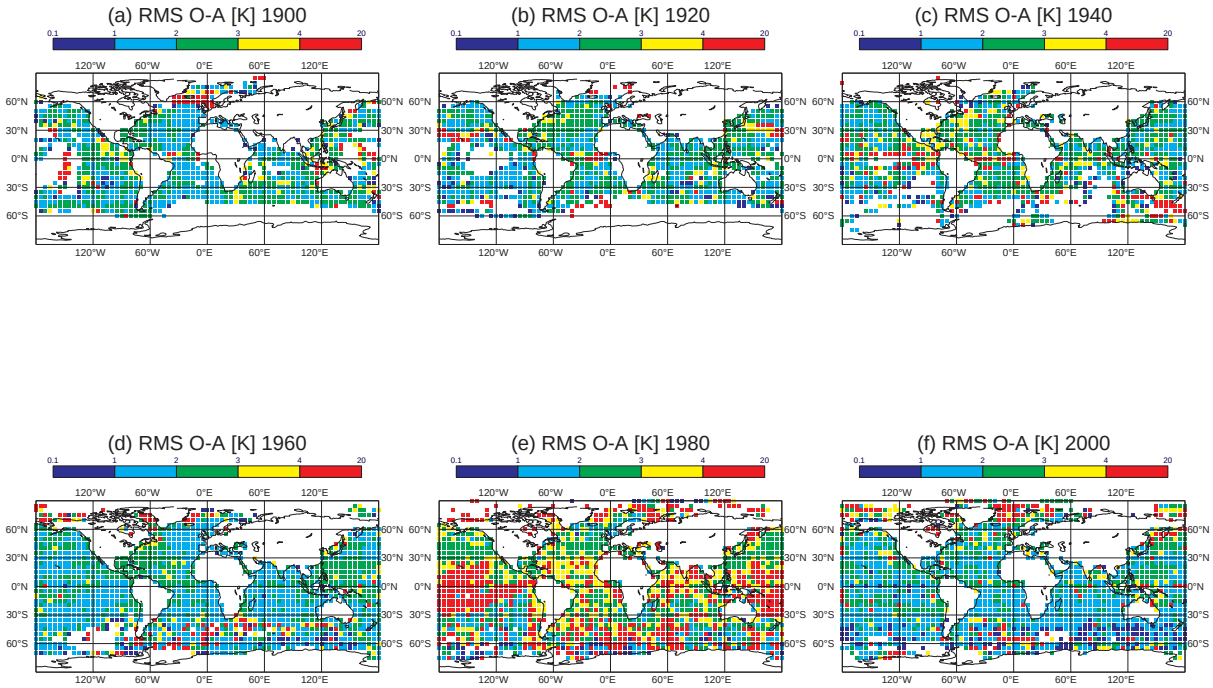

Figure 40: RMS of surface air temperature observation minus analysis departures over open ocean, for selected years (rows 1 to 6). Note that these observations are not assimilated.

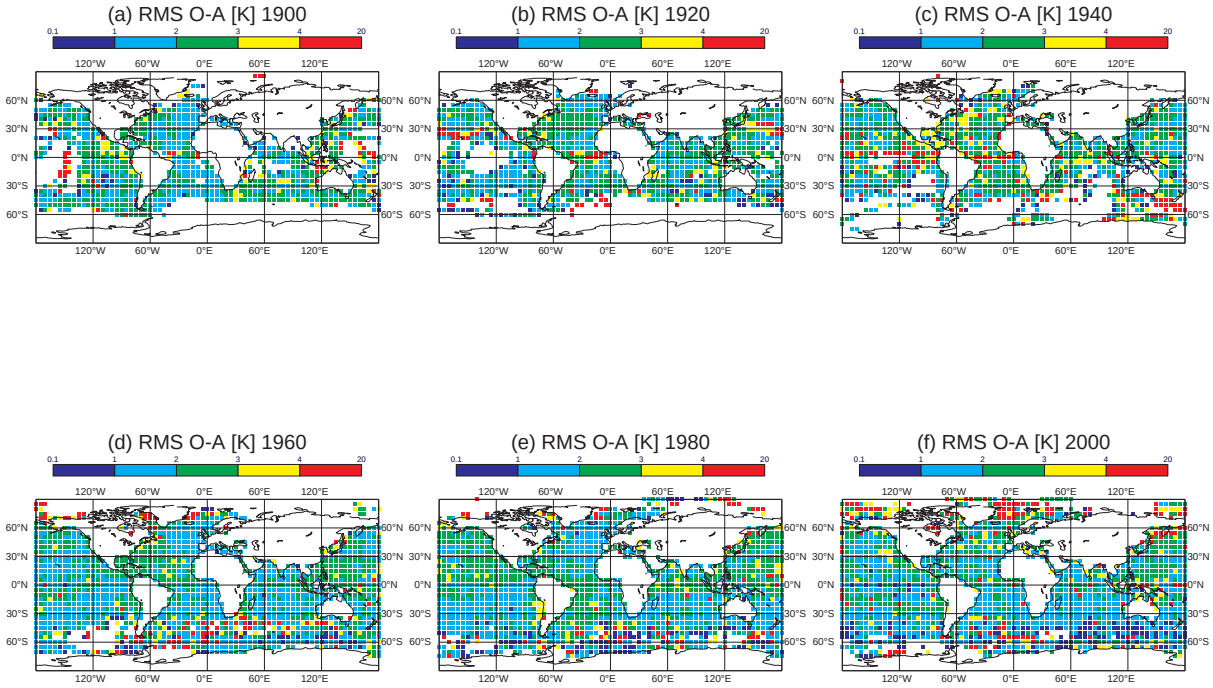

Figure 41: Same as Figure 40 but excluding three collections on the basis of observational statistics, to illustrate the importance of tracing to original sources. See text.

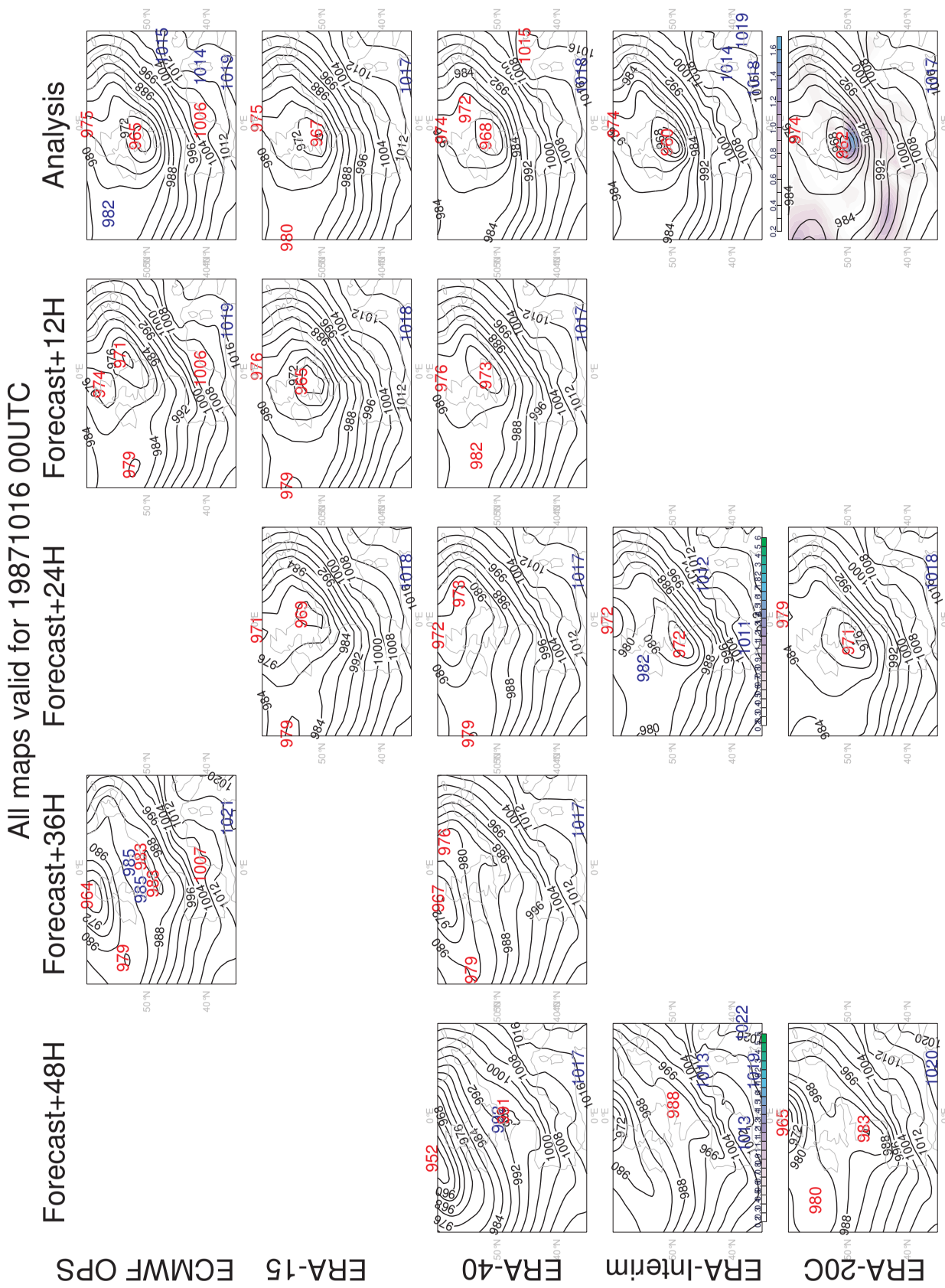

Figure 42: Maps of mean sea level pressure for 16 October 1987, 00 UTC. For ERA-20C analysis, the colour shading shows the ensemble spread.

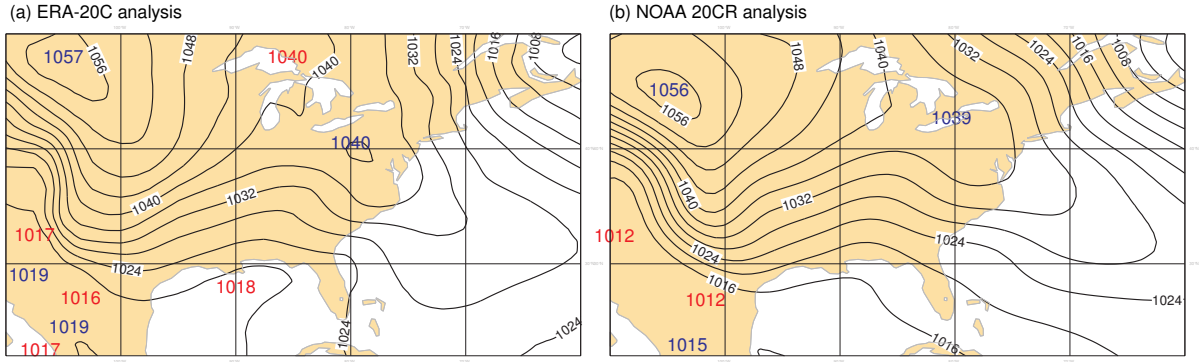

Figure 43: Maps of mean sea level pressure for 11 February 1899, 12 UTC.

upper-air and satellite observations.

The ERA-20C ensemble mean analysis is not as deep as ERA-Interim, at 962 hPa instead of 960 hPa, while the estimate from hand-drawn analyses at the time was around 953 hPa. However, the ensemble spread in ERA-20C (shown with a colour shading in Figure 42) indicates locally larger uncertainties in the vicinity of the storm, whereas the ERA-Interim product comes without any metrics of uncertainty. In hindsight, the performance of ERA-20C for this case echoes Dee *et al.*(2011) who concluded from their evaluation of ERA-Interim for this same meteorological case that *'results could perhaps be further improved by using a better representation of flow-dependent background errors'*.

In the first half of February 1899, the East Coast of the U.S.A. was affected by extremely low temperatures, as far South as Florida (Kocin *et al.*, 1988). This time period is at the very beginning of ERA-20C, before all the components of the ERA-20C system were fully spun-up. Yet this provides an interesting case of an extreme high pressure system. Figure 43 shows the situation as analysed by ERA-20C and NOAA 20CR for 11 February 1899 at 12 UTC. The two reanalyses shown here used approximately the same input surface pressure observations from ISPD, although NOAA 20CR used version 2.2 whereas ERA-20C uses version 3.2.6. In addition, ERA-20C assimilates wind observations, but over ocean only and hence not over the Continental U.S.A. However, as discussed earlier, the two reanalyses used very different approaches in the analysis schemes. In spite of this, both reanalysis products suggest a high pressure system of 1056–1057 hPa located around coordinates 45°N 105°W. The 1 hPa difference in local maximum intensity between the two reanalyses is small, but in ERA-20C the maximum is located approximately 200 km North of the maximum in NOAA 20CR. A close inspection of Figure 2 in the work of Kocin *et al.*(1988) for the map valid at 13 UTC (hence with a one-hour offset) indicates that the two reanalyses are both very close to the hand-drawn analysis, even though the hand-drawn analysis suggests a somehow higher maximum at 1060 hPa. It must be reminded that for ERA-20C these maps are the product of only 6 weeks of reanalysis integration started from inaccurate initial conditions for 1 January 1899 (*i.e.*, ERA-Interim analysis valid for 1 January 1989) and without a yet fully acting surface pressure bias correction.

In their work, Kocin *et al.*(1988) proceed from observations reports from the US Weather Bureau and from marine observations from NCDC. Since the only regular reporting network covers the land domain, the analysis of observations by Kocin *et al.*(1988) is limited to the Continental U.S.A. The fields produced by global reanalyses can be used to construct a picture of the Northern hemisphere circulation during these two weeks. Figure 44 and 45 show that during that time period, according to NOAA 20CR and ERA-20C reanalyses, the Northern Hemisphere saw a nearly stationary wave-2 pattern, with highs over the U.S.A. and Siberia, and lows over the North Atlantic and North Pacific. The highs and lows that develop in that stationary wave are intense. In particular, the Atlantic Ocean sees a low pressure system remain nearly stationary for two weeks; this co-

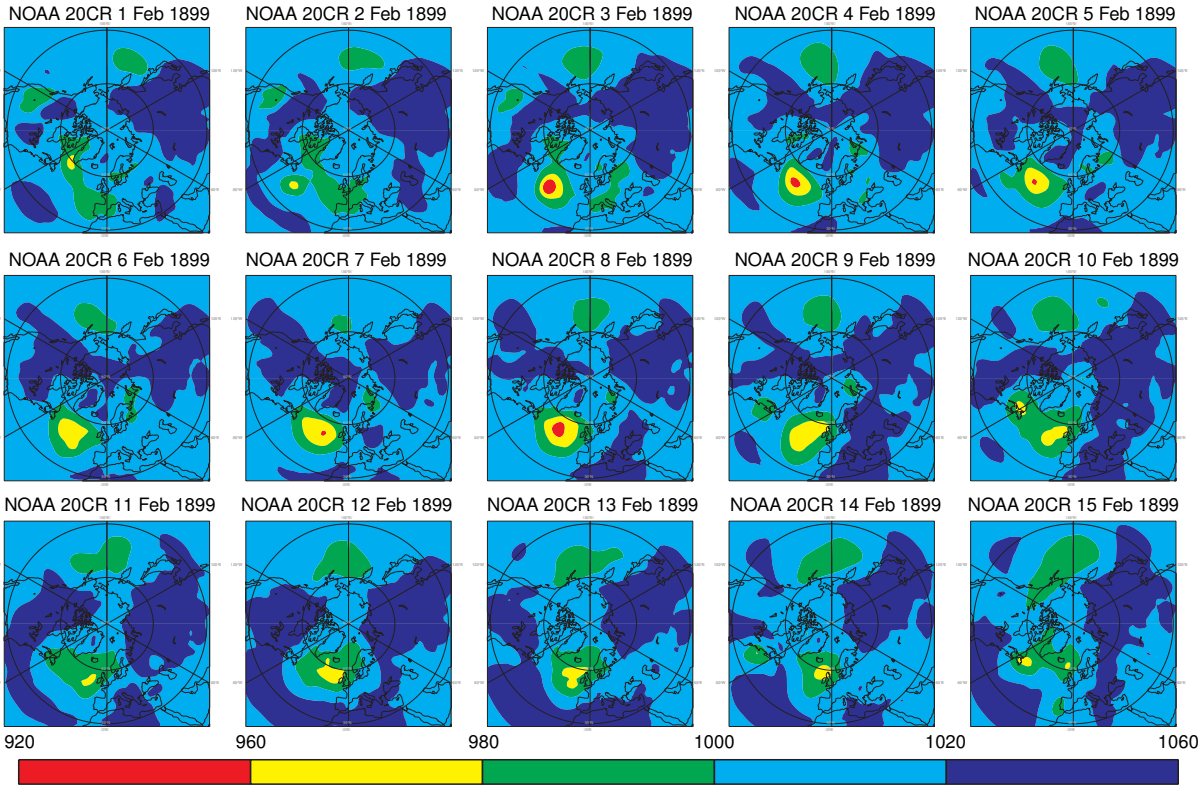

Figure 44: Maps of mean sea level pressure from NOAA 20CR ensemble mean, all valid at 00 UTC

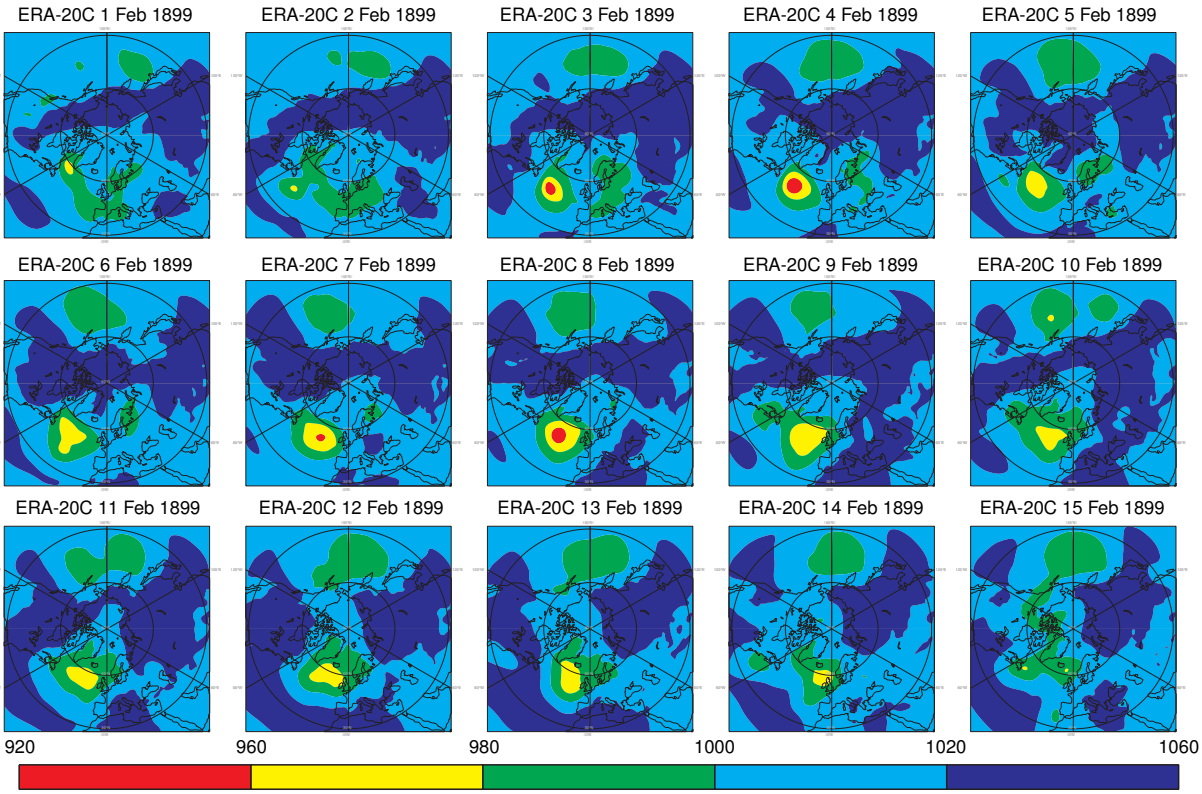

Figure 45: Same as Figure 44, but from ERA-20C ensemble mean

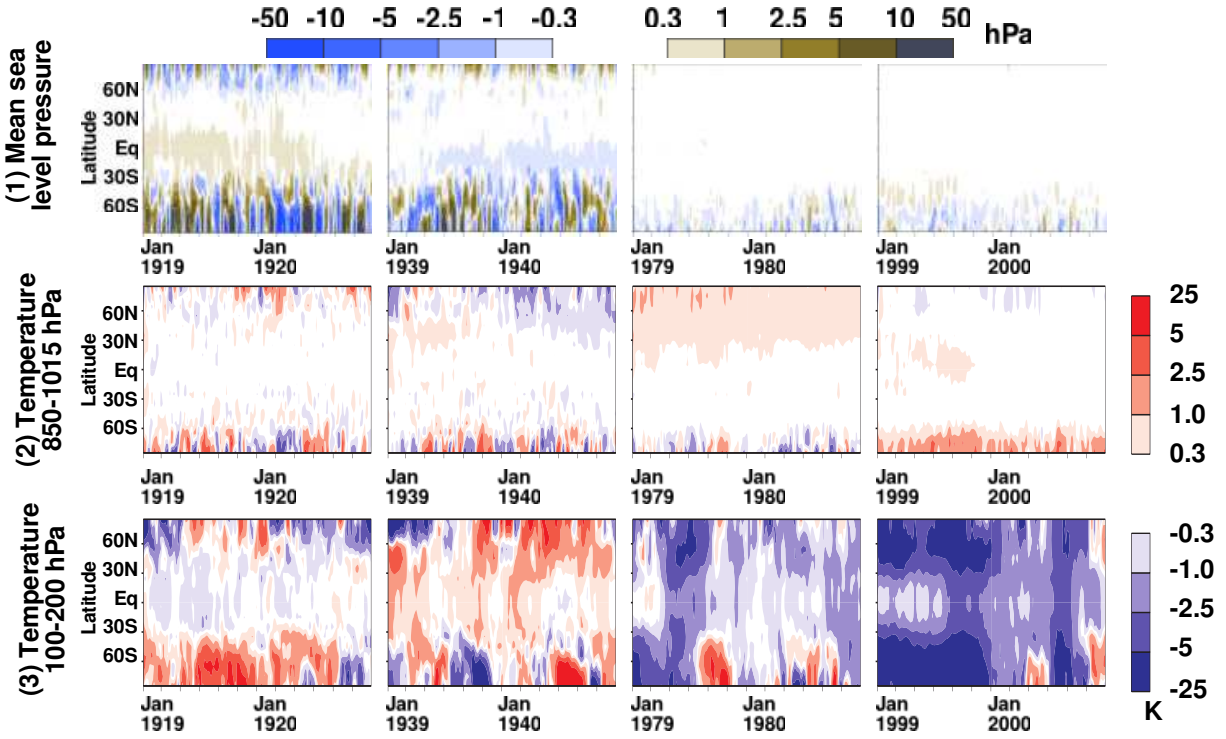

Figure 46: Mean zonal differences between consecutive ERA-20C production streams, as a function of time within the overlap time periods. Row (1) shows mean differences between analyses of mean sea level pressure, row (2) shows mean differences between analyses of temperature for the model levels 78–91 or approximately 850–1015 hPa, and row (3) shows mean differences for the model levels 38–49 or approximately 100–200 hPa

incides with an article published on 16 February 1899, reporting serious difficulties from ships arriving in the New-York harbour (New York Times, 1899). This meteorological case of an extreme high pressure system, although briefly discussed here, illustrates the value of placing historical extreme events in the context of global circulation to help understand them.

5.7 Match between consecutive production streams and comparison with ERA-Interim

At the time of writing, ERA-20C production streams 1, 2, 3, and 5 are complete. Stream 6 is stopped in mid-September 2009, waiting for an extension of HadISST2.1.0.0 for dates thereafter. Stream 4 is completing the year 1960. As can be inferred from the production layout shown in Figure 6, this means that 4 of the 5 intended 2-year overlaps are complete, namely for the years 1919-1920, 1939-1940, 1979-1980, and 1999-2000.

The first evaluation of these overlap time periods concerns the boundaries on January 1st. Row (1) in Figure 46 shows the mean zonal differences between mean sea level pressure analyses from two consecutive streams, computed in 10° latitude bins as a function of time (showing 7-day averages), for the 4 overlap time periods completed so far. The differences on January 1st 1920 (1940, 1980, 2000, respectively) are a measure of the discontinuities between two streams. In the first two overlaps we see that for high latitudes (above 60° in either hemisphere) the differences typically exceed 2.5 hPa. In these regions, alternating patterns of negative and positive differences suggest that consecutive streams may place at different times the low and high pressure systems in these regions. This is consistent with the weak observational constraints in the early part of the 20th century for those regions and points to regions where recovery of additional observations would be of high value to improve the quality of the reanalysis product. For regions located at latitudes 30°S–60°N, smaller

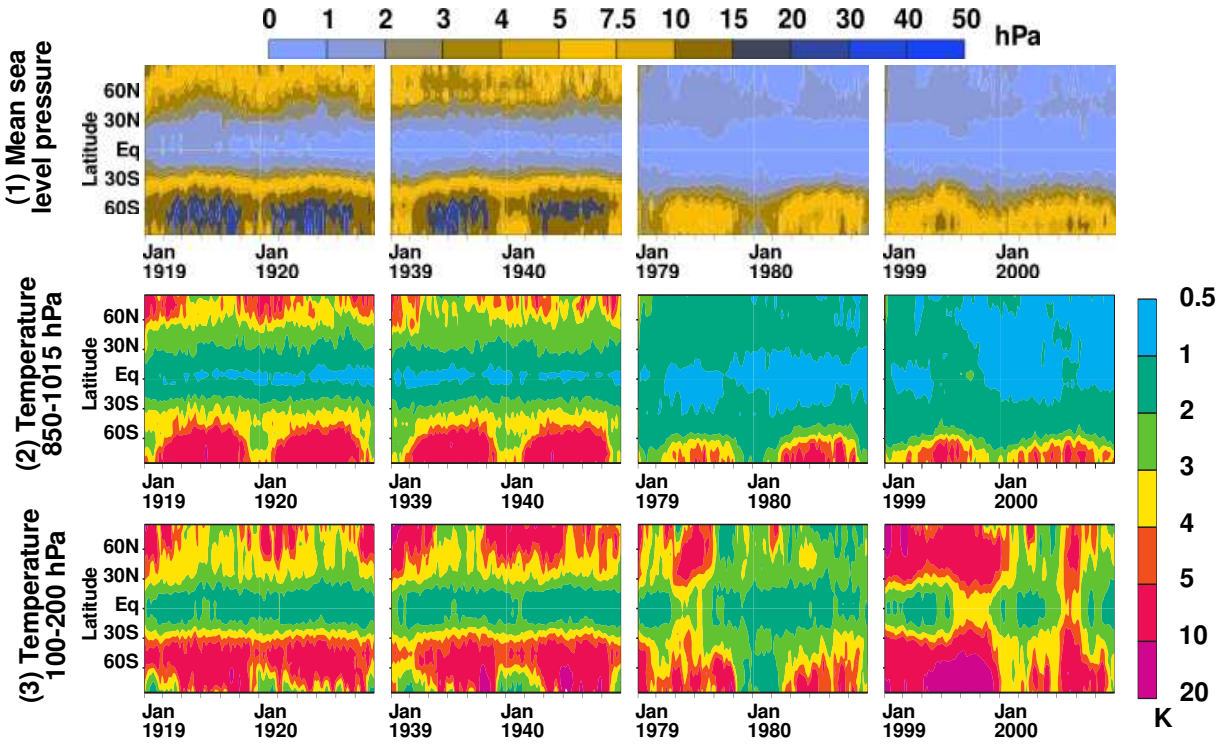

Figure 47: Same as Figure 46, but showing RMS instead of mean differences

mean differences are found, sometimes within ±0.3 hPa but always within ±1 hPa. In the last two overlaps we find mean differences that are always within ±0.3 hPa in the whole Northern Hemisphere and Tropics, and usually within ±1 hPa in the Southern Hemisphere high latitudes.

Row (2) shows comparisons between consecutive streams for analysed temperatures at the model levels closest to the surface (model level numbers 78–91, or approximately 850–1015 hPa). The discontinuities are much smaller in recent years, within ±0.3 K mean differences in the year 2000 in large parts of the Northern Hemisphere. In the first part of the century the discontinuities can be larger, occasionally as much as ±2.5 K, but otherwise within ±1 K.

Row (3) shows that for the levels near the tropopause (model level numbers 38–49, or approximately 100–200 hPa) the streams do not agree well in the mean, with differences typically around ±2.5–5 K.

In an attempt to consider not only the instantaneous discontinuities (mean differences) but also discontinuities in the representation of weather patterns, row (1) in Figure 47 shows the RMS of differences between mean sea level pressure analyses from two consecutive streams, in the same fashion as Figure 46. In the first two overlaps, we see that small differences (within 2 hPa) are only found in the Tropics. Elsewhere, the RMS differences exceed 3 hPa, up to 20–30 hPa in the Southern high latitudes. In the last two overlaps we find RMS differences within 2 hPa in the whole Northern Hemisphere and Tropics, and usually within 7.5 hPa in the Southern Hemisphere.

Row (2) shows RMS differences between analysed temperatures near the surface (model levels number 78–91, or approximately 850–1015 hPa). There again the quality of the overlap is much greater in recent years (within 1 K RMS in the year 2000 in most of the Northern Hemisphere) than in the first part of the century (usually around 2–3 K RMS at mid-latitudes, and 3–5 K RMS at high latitudes). Although these differences are no direct measures of accuracy, such qualitative conclusions are in line with expectations; as more observations

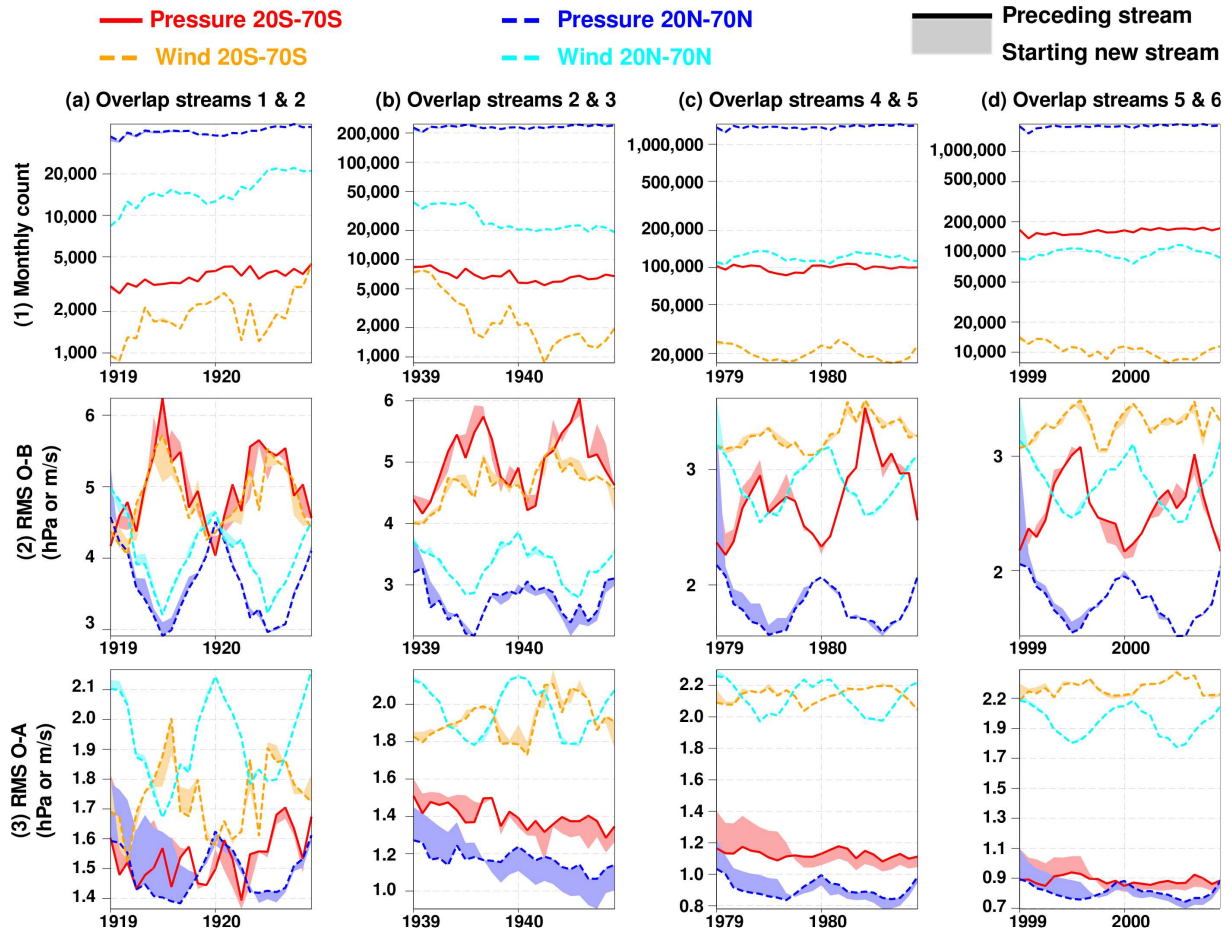

Figure 48: Monthly statistics for observations assimilated in ERA-20C production streams, focusing on the common years analysed by consecutive streams. Shading shows the area between two consecutive streams (the smaller the area, the better the quality of the overlap). Row (1) shows the observation count, row (2) shows the RMS of observation minus background departures, and row (3) shows the RMS of observation minus analysis departures

are assimilated, the analyses are more constrained in the timing of synoptic events, and have fewer chances of adopting different solutions (note, similar analysis solutions in two streams does not imply that these solutions are necessarily correct, as they could both be similarly wrong and in particular biased). In areas which are notoriously unobserved, such as the Southern high latitudes pre-World War II, the differences between analyses in consecutive streams are the largest, which is consistent with greater uncertainties there.

Last, row (3) shows that for the levels near the tropopause the quality of the overlaps is poor at all times with RMS differences generally above 5 K, except in the Tropics owing to lower natural variability. The differences are particularly large in the high latitudes polar winters, probably because the position of the polar vortex then is not known well and different streams give different solutions. This ties well with the initial discussion in introduction where we give initial indications that a reanalysis assimilating only surface observations may not generate realistic products in the stratosphere in the absence of additional observational or model constraints.

Assessing the agreement between consecutive streams can also be carried out in observation space. Figure 48 shows monthly statistics for assimilated observations in each production stream, separated in large geographical regions (Southern and Northern extratropics) and by geophysical variable (pressure and wind). In each plot, the thick line shows statistics for the stream started 20 years before the dates shown. The shading shows the area between similar statistics for consecutive streams. Columns (a) to (d) in the figure show the overlap time

periods completed so far.

Row (1) in Figure 48 shows that consecutive streams use nearly exactly the same number of assimilated observations. Row (2) shows that the RMS of observation minus background departures for surface pressure are larger for the new streams in the first year of production and converge within a year to statistics found for the preceding streams. For wind, the situation is somehow different with a near-instantaneous convergence, except maybe for the first overlap year in 1919. This suggests that the production streams have a longer memory for pressure than wind, probably because of the variational bias correction applied to surface pressure observations, which has a slow update time as indicated earlier. Generally, consecutive streams are nearly indistinguishable in the second years of the overlaps in terms of observation minus background statistics.

Row (3) in Figure 48 shows that the RMS of observation minus analysis departures is very similar between two consecutive streams for wind, but isn't as close for pressure. In fact, we observe, in all overlaps, that preceding streams have smaller pressure residuals in the beginning year than the new streams (because the initial conditions of the new stream are not as good, this is expected). However, after about 9 months, the new streams present systematically smaller pressure residuals than the preceding streams. This indicates that the new streams are able to fit the surface pressure observations better than the preceding streams after a given point. This is consistent with the design of the variational bias correction for pressure observations: the adjustments become gradually stiffer over time.

Figure 49 shows temperature anomalies in ERA-20C and in ERA-Interim between January 1979 and August 2009, computed from 6-hourly analyses and averaged to 14-day resolution for plotting. The anomalies are computed by subtracting from each product its own monthly climatology, calculated over the whole time period considered here. The row (1) shows that global anomalies are more different at 850 hPa than they are at 1000 hPa. In both cases however, the temperature trends in ERA-20C differ from those of ERA-Interim (ERA-20C showing a greater global warming then ERA-Interim). From the zonal plots shown in rows (2)–(6), one notes that the temperature anomalies are most similar over the Northern extratropics, but differ significantly in areas where the observational constraint in ERA-20C is weak, such as the Tropics and the Southern latitudes. This figure suggests that trends that can be derived from ERA-20C will probably be close to ERA-Interim (*i.e.*, realistic) only in areas which are consistently observed: near the surface and in the Northern extratropics latitudes only. Everywhere else, one should not expect to find correct temperature trends in ERA-20C unless further investigation demonstrates otherwise.

6 Conclusions and issues

The ECMWF pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C) attempts to reconstruct the weather history between years 1900 and 2010, using a 10-member ensemble of 24-hour 4D-Var analyses. The horizontal resolution is approximately 125 km (same as ERA-40, with analysis increments at horizontal resolution 210 km). A variational bias correction of surface pressure observations is applied. A novel method updates the background error covariances, so as to represent changes in background quality that result from an increase in observation coverage over time.

During the century, the numbers of surface pressure and wind observations from ISPD 3.2.6 and ICOADS 2.5.1 increase approximately by a factor 50 in both hemispheres. Over time, this results in a reduced ensemble spread, an improved fit of the background to observations, and reduced analysis increments. The magnitudes of the bias corrections decrease over time, which could result from improved observation quality (reduced observation biases). Between 1900 and 2010, the ensemble spread for 3-hour forecasts decreases from around 1 hPa in the Tropics and Europe and 15 hPa in the Southern oceans and high-latitudes to under 1 hPa for most of the globe, except the Southern Oceans, where it remains above 3 hPa (even in present times, the largest spread is observed

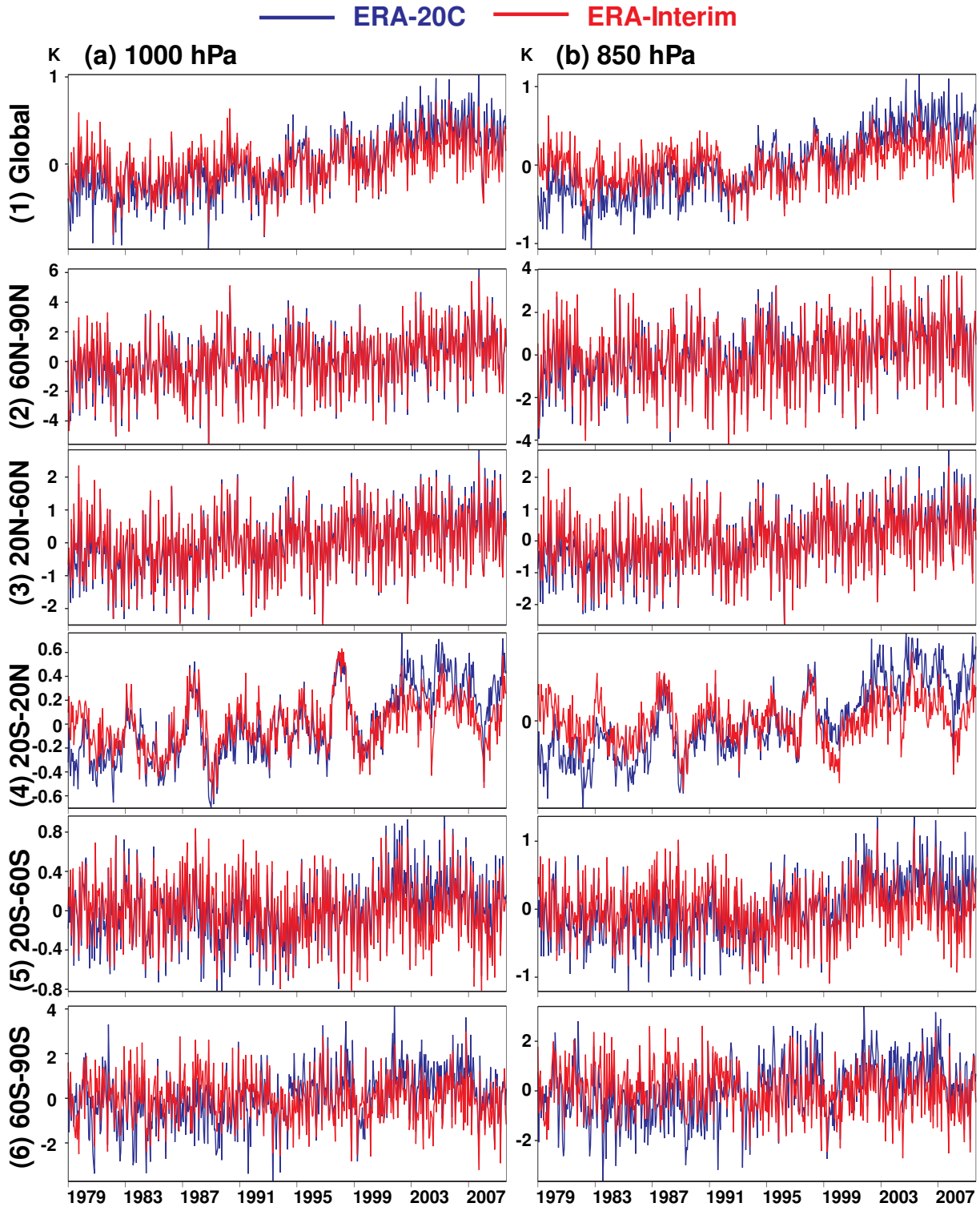

Figure 49: Global (row (1)) and zonal (rows (2)–(6)) anomalies in temperature at 1000 hPa (column (a)) and 850 hPa (column (b))

in the Southern Pacific). In parallel, the RMS of observation minus background departures for surface pressure drops from 4–5 hPa to around 2 hPa. The global RMS of analysis minus background increments decreases from 3 to 2 hPa. The global RMS of bias corrections drops from 1.5 to 0.8 hPa. From several metrics, it seems that the improvement in product quality in not as smooth as the increase in observation numbers, but sees a step around the 1950s. The quality of the Southern hemisphere reconstruction in the 1960s matches that of the Northern hemisphere in 1900. Because of an unequal geographical observational coverage for surface pressure and wind, the Southern hemisphere quality lags behind the Northern hemisphere quality at all times. However, in the recent years, even though the observation numbers continue to increase in both hemispheres, the quality of the Southern (Northern) hemisphere reconstruction keeps (stops) improving. This levelling-off effect in the Northern hemisphere probably suggests that the ERA-20C analysis system is unable to extract additional information from additional surface pressure and wind observations past a certain point, in the current data assimilation configuration.

The 10 different realisations produced by ERA-20C represent the first attempt in ECMWF reanalyses to provide users with a useful estimate of uncertainty. The true level of uncertainty cannot be assumed as equal to the ensemble spread, but relative variations between ensemble spread (*e.g.*, over time, or in space) appear qualitatively meaningful as illustrated in this report. Even though the ensemble size is rather small at this point, it is expected that user feedback and research on downstream applications based on ERA-20C will help improve the value of probabilistic reanalysis products. It should be recognised, however, that the underlying observational uncertainties of the early 20th century are fundamentally unknown and cannot be accurately estimated by any data assimilation methodology.

The production speed of ERA-20C, averaging between 20 and 40 reanalysed days per elapsed day of production stream, was found to be mostly affected by non-related upgrades needed by the operational NWP computing environment. Should such production be repeated in a sustained mode, it would benefit to run in a somehow separate non-changing computing environment.

In terms of meteorology, the quality of the ERA-20C 3- to 4-day forecasts of geopotential height at 500 hPa are comparable to that of ERA-Interim at day 7 in 2004. Preliminary investigations indicate that the ERA-20C dataset represents with some fidelity a well-known case of an extreme low pressure system. This case is the great storm of October 1987, regularly assessed *a posteriori* in new reanalysis projects. The 24-hour forecast of ERA-20C for that storm is much improved compared to ERA-40 and operations at the time, and comparable if not better than ERA-Interim, although ERA-20C uses only surface observations. Another extreme case, for an unusually large high pressure system associated with the U.S. East Coast Blizzard of February 1899 is also briefly studied. Although formally before the beginning of the target time period of ERA-20C, in the first 4 to 6 weeks of the production for the year 1899, the ERA-20C reconstruction agrees fairly well with NOAA 20CR reanalysis and hand-drawn analysis of Kocin *et al.*(1988). Furthermore, this case shows that there is possibly a connection between this extreme high pressure system and the unusually bad weather over the Atlantic in the first 2 weeks of February. A planetary wave-2 patter is nearly stationary for 2 weeks. This second case study suggests some ability of ERA-20C to also represent high extremes in observed areas.

Although a similar concept is part of the design of the NOAA 20CR ensemble reanalysis (Compo *et al.*, 2011), the ERA-20C reanalysis uses for the first time in ECMWF reanalyses and unlike ECMWF NWP operations (as of 2012) a set of self-updating background error covariances. The importance of adjusting these covariances is found to be significant on forecast scores (this result is from an evaluation conducted for the year 2004). Over time, the background errors estimated by ERA-20C become smaller in magnitude and in their horizontal extent. In other terms, the ERA-20C system employs observations in the early years of the century to correct large-scale errors in the background, while at the end it employs many more observations which allow to correct for smaller-scale errors, in many more locations.

In a planned rerun of ERA-20C, the following points need to be addressed. The first-guess check limits of ERA-20C, at 18 times the expected error, would need to be reduced to probably 10 times the expected error. This would reduce occurrences of assimilation of surface pressure observations affected by gross errors. The variational bias correction is found to evolve very slowly in the first years, and to stabilise only between years 2 and 4 of each stream. The surface pressure observation errors specified for land stations (ships) would need to evolve in time, as suggested by several statistical metrics in this report, from 1.6 hPa (2.0 hPa) at the beginning of the century to 0.8 hPa (1.2 hPa) in the years 2000s, instead of the constant assumed error of 1.1 hPa (1.5 hPa) that was employed at all times. The CTD/XCTD/MBT/XBT reports seem to contain larger errors in terms of surface pressure than other report types and this would need explaining before repeating assimilation of these data in a future reanalysis. Regarding near-surface wind, the ship observation errors are larger than assumed (1.5 m/s), at least around 2.0 m/s, or probably even more as the representativeness error seems to be significant at the low horizontal resolution considered for the increments. The background error estimation currently uses a past sample of 90 days. This causes the background error covariance estimates to lag behind seasonally and also to probably miss the seasonal extrema. Using a larger ensemble size would allow to reduce the sample size for background estimation and track better the errors in the reanalysis background over time. This, combined with modified specifications of observation errors as suggested here would very likely reduce the differences between predicted and observed RMS of observation departures.

Because of the long experimental setup and the computing costs of the ERA-20C productions, choices had to be made regarding model settings, sacrificing sometimes quality to cost. The model time-step was set at 1 hour, although it would have been more accurate to use 30 minutes. This halving of the production cost in real terms has the bad consequence of introducing a lag in the atmospheric tides, with a wave-2 pattern of ± 1 hPa error, visible on maps of analysis increments and fits to surface pressure observations before assimilation.

The model biases alias into the ERA-20C upper-air temperature products as the variational analysis spreads the analysis increments vertically. Combined with a changing observation coverage, these temperature increments are displaced vertically over time, with an increasing amplitude so as to produce similar net corrections in terms of surface pressure model bias. Consequently, the ERA-20C upper-air products are affected by spurious trends. Overall, the ERA-20C trends in the stratosphere are likely of no value, while those in the troposphere are only probably realistic in well-constrained regions, namely near the surface in the Northern Hemisphere extratropics. Everywhere else, the trends from ERA-20C should not be considered as realistic unless shown otherwise, and any slow evolution noted in the products should be interpreted with due caution before drawing firm conclusions.

Overall, we find indications that the online update of background error covariances is qualitatively working as intended. The lessons learnt from the automatic update of background error covariances give confidence that the scheme employed for this estimation is stable. This is expected to benefit ECMWF NWP operations in the short term. For a climate reanalysis application, one aspect that would need dedicated attention is how to restrict the propagation of increments in unobserved areas. For example, the vertical localisation employed in ensemble Kalman filter systems is not readily available in a variational system, but a similar concept could be explored.

In terms of observations, the ERA-20C reanalysis uses for the first time the observation feedback archive (OFA) developed for ERA-CLIM. This report illustrates the value of maintaining a traceability to original observations in the OFA, showing an example of air temperatures in the year 1980. Also, the importance of data recovery is illustrated by pointing out to data collections that bring in the first observations of the high-latitude Southern Oceans. After the point when these observations are assimilated, the quality of the Southern hemisphere reconstructions are much improved, according to the fit between background and observations.

As of writing, the ERA-20C production is about complete, except for a supplementary year of production in

1960, intended to evaluate overlap quality between two streams. The present report shows that the overlaps achieved by completed consecutive streams are as expected in the production plan by allowing for a one-year spin-up time. We find indistinguishable statistics for the observation minus background departures in the second year of each seam. We also observe a greater memory of the system for the fit to surface pressure observations, owing to the design of the variational bias correction.

The agreement between consecutive streams is better overall in the second part of the century. In 1980 and 2000, the differences in the Northern Hemisphere between analyses produced by consecutive streams for mean sea level pressure are within ± 0.3 hPa in the mean and 2 hPa RMS (for lower tropospheric temperatures: 1 K mean and 2 K RMS). By contrast, these numbers in 1920 and 1940 are about ± 1 hPa in the mean and 7.5 hPa RMS (for lower tropospheric temperatures: 2.5 K mean and 5 K RMS). In the Southern Hemisphere, the corresponding numbers are higher but there also we observe reductions over time as the observation coverage improves. In the upper tropospheric/lower stratospheric levels (100–200 hPa), the large differences between consecutive streams confirm that these levels are not constrained by surface observations. The analyses at such levels represent thus more the model variability and should be used with caution.

For future reference, Table 3 lists the issues to be addressed in a rerun of the ERA-20C control.

Acknowledgement

The European reanalysis of global climate observations project (ERA-CLIM) is funded by EU research FP7 Environment under grant agreement no. 265229. In addition, ERA-CLIM benefits from significant contribution from ECMWF, for high-performance computing, archive, data services, and supplementary staff support. We thank NOAA and NCAR for providing the ICOADS v2.5.1 and ISPD v3.2.6 observational datasets. We are grateful to Rob Allan and his collaborators for their efforts in setting up and coordinating the Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative, which facilitates and undertakes data rescue activities essential for the continuous improvement of ISPD and ICOADS. The HadISST2.1.0.0 dataset was provided by the Met Office Hadley Centre with partial funding from ERA-CLIM. We thank the ERA-CLIM advisory board for fruitful comments and suggestions during the preparation of ERA-20C.

References

Andersson E, Järvinen H. 1999. Variational quality control. Q. J. R. Meteorol. Soc., 125, 697–722

Bonavita M, Isaksen L, Holm E. 2012. On the use of EDA background error variances in the ECMWF 4D-Var. TM664. January 2012

Bonavita M, Raynaud L, Isaksen L. 2010. Estimating background-error variances with the ECMWF Ensemble of Data Assimilations system: the effect of ensemble size and day-to-day variability. TM 632. July 2010

Bonavita M. 2011. Changes in EDA variances from CY36R4 to CY37R2. RD Memorandum R43.2/MB/1134

Compo GP, Whitaker JS, Sardeshmukh PD. 2006. Feasibility of a 100-Year Reanalysis Using Only Surface Pressure Data. Bull. Amer. Meteor. Soc., 87, 175–190. DOI: http://dx.doi.o rg/10.1175/BAMS-87-2-175

Compo GP, Whitaker JS, Sardeshmukh PD, Matsui N, Allan RJ, Yin X, Gleason BE, Vose RS, Rutledge G, Bessemoulin P, Brönnimann S, Brunet M, Crouthamel RI, Grant AN, Groisman PY, Jones PD, Kruk MC, Kruger AC, Marshall GJ, Maugeri M, Mok HY, Nordli O, Ross TF, Trigo RM, Wang XL, Woodruff SD, Worley SJ. 2011. The Twentieth Century Reanalysis Project. Q.J.R. Meteorol. Soc., 137: 1–28. DOI: 10.1002/qj.776

Issue number	Nature	Diagnostic	Possible solution
1	Analysis increments spread vertically introducing spurious trends	Vertical mean analysis increments show slow, but large time variations	Ideally: Remove model bias More likely: Modify background errors
2	First-guess check too loose (18 times the expected error) and some bad data not rejected	Instances of large obs minus background departures assimilated	Revise f.g. check settings and edit blacklist to remove gross errors by inspecting ERA-20C feedback
3	Variational bias correction of surface pressure slower than expected	Global RMS of bias correction converges after 2–4 years	Check procedure
4	Model bias in representation of atmospheric tides	Geographical patterns in analysis increments of surface pressure	Reduce model time-step
5	Suboptimal observation errors	Desroziers diagnostics and comparison between observed and expected RMS of innovations	Use estimates as documented in this report
6	Large error estimates for pressure obs. from CTD/XCTD/MBT/XBT	About 2 times that of other observations	Understand the cause, or blacklist
7	Background error lags seasonally	Maximum of error in observed RMS innovation lags behind the predicted values	Use past and future states from ERA-20C ensemble in future estimation

Table 3: Summary of the issues found in the present report, to be addressed in a rerun of the ERA-20C control

Dee DP. 2004. Variational bias correction of radiance data in the ECMWF system. Proc. ECMWF workshop on assimilation of high spectral resolution sounders in NWP. Reading, United Kingdom, 28 June - 1 July 2004

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette JJ, Park BK, Peubey C, de Rosnay P, Tavolato C, Thépaut JN, Vitart F. 2011a. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q.J.R. Meteorol. Soc., 137: 553–597. DOI: 10.1002/qj.828

Dee DP, Källén E, Simmons AJ, Haimberger L. 2011b. Comments on Reanalyses suitable for characterizing long-term trends. Bull. Amer. Meteor. Soc. 72: 65–70.

Dee DP. 2013. ERA-20C production has started. ECMWF Newletter 134: 6–7. Reading, United Kingdom. http://www.ecmwf.int/publications/newsletters/pdf/134.pdf

Derber JC, Wu WS. 1998. The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. Mon. Wea. Rev. 126: 2287–2299

Desroziers G, Berre L, Chapnik B, Poli P. 2005. Diagnosis of observation, background and analysis-error statistics in observation space. Q.J.R. Meteorol. Soc., 131: 3385–3396. DOI: 10.1256/qj.05.108

ECMWF. 2013. IFS documentation CY38R1, available from http://www.ecmwf.int/research/ifsdocs/CY38r1

Fisher M, Courtier P. 1995. Estimating the covariance matrices of analysis and forecast error in variational data assimilation. TM220. August 1995

Fisher M. 2003. Background error covariance modelling. Seminar on Recent developments in data assimilation for atmosphere and ocean, 8-12 September 2003

Fisher M. 2004. Generalized frames on the sphere, with application to the background error covariance modelling. Seminar on Recent developments in numerical methods for atmospheric and ocean modelling, 6-10 September 2004

Isaksen L, Bonavita M, Buizza R, Fisher M, Haseler J, Leutbecher M, Raynaud L. 2010. Ensemble of data assimilations at ECMWF. TM 636. December 2010

Knapp KR, Kruk MC, Levinson DH, Diamond HJ, Neumann CJ. 2010. The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone best track data. Bull. Amer. Meteor. Soc. 91: 363–376. DOI: 10.1175/2009BAMS2755.1

Kocin PJ, Weiss AD, Wagner JJ. 1988. The Great Arctic Outbreak and East Coast Blizzard of February 1899. Wea. Forecasting. 3: 305–318. DOI: http://dx.doi.org/10.1175/1520-0434(1988)003¡0305:TGAOAE¿2.0.CO;2

Kuchta. P. 2009. Observation feedback archiving in MARS. 12th Workshop on Meteorological Operational Systems, 2-6 November 2009. ECMWF, Reading, United Kingdom. http://www.ecmwf.int/newsevents/

meetings/workshops/2009/MOS_12/Presentations/Kuchta.pdf

Morris RM, Gadd AJ. 1988. Forecasting the Great Storm. Weather 43: 70–89

New York Times. 1899. Terrific storms at sea; Steamships from All Quarters Report Extremely Rough Voyages. All more or less Battered Vessels Sighted in Distress and Abandoned, Blinding Snow and Waves Like Mountains. New-York, New-York. 16 February 1899 Available from http://spiderbites.nytimes.com/free_1899/-articles_1899_02_00001.html

Palmer TN, Buizza R, Doblas-Reyes F, Jung T, Leutbecher M, Shutts GJ, Steinheimer M, Weisheimer A (2009): Stochastic parametrization and model uncertainty. ECMWF Research Department Technical Memorandum, 598, ECMWF, Shinfield Park, Reading RG2-9AX, UK, pp. 42

Poli P. 2011. Data assimilation for atmospheric reanalysis. Seminar Proc. ECMWF data assimilation for atmosphere and ocean. Reading, United Kingdom, 6 - 9 September 2011: 231–247

Poli P, Tan D. 2012. Activation of digital filtering of temperature and vorticity to suppress large increments in poorly observed areas. Technical Report RD49. November 2012. Available from ECMWF, Shinfield Park, Reading RG2 9AX, United Kingdom

Raynaud L, Berre L, Desroziers G. 2009. Objective filtering of ensemble-based background-error variances. Q.J.R. Meteorol. Soc. 135: 1177–1199. DOI: 10.1002/qj.438

Saarinen, S. 2004. ODB User Guide. 7 May 2004, ECMWF, Reading, United Kingdom. http://www.ecmwf.int/-research/ifsdocs/CY28r1/pdf_files/odb.pdf

Tavolato C, Isaksen L. 2010. Huber norm quality control in the IFS. ECMWF Newsletter No. 122, pp. 27–31

Thorne PW, Vose RS. 2010. Reanalyses suitable for characterizing long-term trends: Are they really achievable? Bull. Amer. Meteor. Soc. 91: 353–361

Whitaker JS, Compo GP, Thépaut JN. 2009. A Comparison of Variational and Ensemble-Based Data Assimilation Systems for Reanalysis of Sparse Observations. Mon. Wea. Rev., 137, 1991–1999. DOI: http://dx.doi.org/-10.1175/2008MWR2781.1

World Climate Research Programme (WCRP). 2011. Coupled Model Intercomparison Project - Phase 5. Special Issue of the CLIVAR Exchanges Newsletter, No. 56, Vol. 15, No. 2

World Meteorological Organization (WMO) Commission for Instruments and Methods of Observation (CIMO) Expert Team on Standardization. 2012. Pressure Reduction Formula. Geneva, Switzerland, 26-29 November 2012, CIMO/ET-Stand-1/Doc. 10 (20.XI.2012). http://www.wmo.int/pages/prog/www/IMOP/meetings/SI/ET-Stand-1/Doc-10_Pressure-red.pdf

World Meteorological Organization (WMO). 1954. Reduction of atmospheric pressure, preliminary report on problems involved. Geneva, Switzerland, Tech. Note 7. WMO-No.36.TP.12. http://library.wmo.int/-pmb_ged/wmo_36.pdf

World Meteorological Organization (WMO). 1964. Note on the standardization of pressure reduction methods in the international network of stations. Geneva, Switzerland, Tech. Note 61. WMO-No.154.TP.74. http://library.wmo.int/pmb_ged/wmo_154.pdf

World Meteorological Organization (WMO). 1968. Methods in use for the reduction of atmospheric pressure. Geneva, Switzerland, Tech. Note 91. WMO-No.226.TP.120. http://library.wmo.int/pmb_ged/wmo_226.pdf

Woodruff SD, Worley SJ, Lubker SJ, Ji Z, Freeman JE, Berry DI, Brohan P, Kent EC, Reynolds RW, Smith SR, Wilkinson C. 2011. ICOADS Release 2.5: Extensions and enhancements to the surface marine meteorological archive. Int. J. Climatol., 31, 951-967. DOI: 10.1002/joc.2103

Yin X, Gleason BE, Compo GP, Matsui N, Vose RS. 2008. The International Surface Pressure Databank (ISPD) land component version 2.2. National Climatic Data Center, Asheville, North Carolina, United States of America. 12 pp.