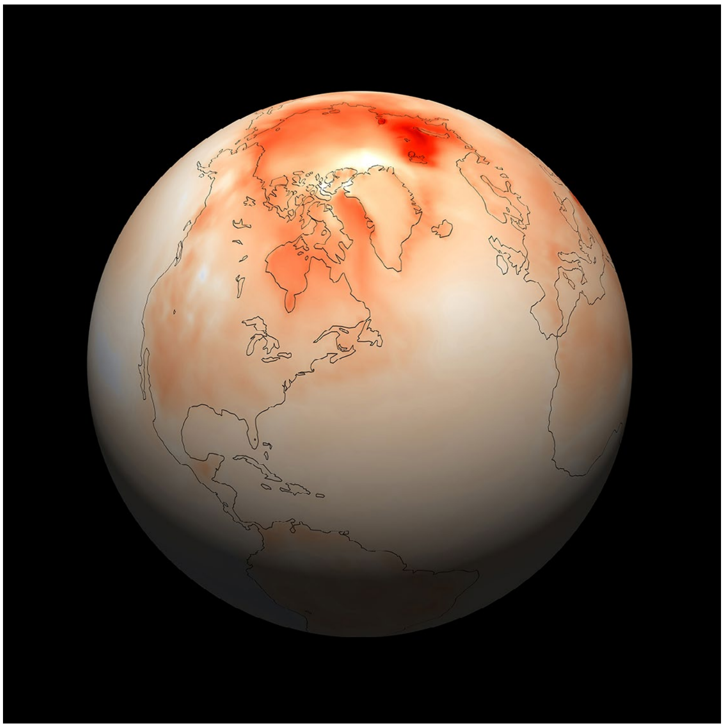
ECMWF Feature article

from Newsletter Number 159 – Spring 2019

METEOROLOGY

Global reanalysis: goodbye ERA-Interim, hello ERA5



Cover image: Point-wise linear trend in surface temperature between 1979 and 2018 from ERA5

www.ecmwf.int/en/about/media-centre/media-resources
doi: 10.21957/vf291hehd7

This article appeared in the Meteorology section of ECMWF Newsletter No. 159 – Spring 2019, pp. 17-24.

Global reanalysis: goodbye ERA-Interim, hello ERA5

Hans Hersbach, Bill Bell, Paul Berrisford, András Horányi, Joaquín Muñoz Sabater, Julien Nicolas, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, Dick Dee

As part of implementing the EU-funded Copernicus Climate Change Service (C3S), ECMWF is producing the ERA5 reanalysis of the global weather and climate. Production is complete for the period 1979 to the present, and by the first quarter of 2020 ERA5 will provide a detailed record of the global atmosphere, land surface and ocean waves from 1950 onwards. This new reanalysis replaces the highly successful ERA-Interim reanalysis that was started in 2006 and spans the period from 1979 to the present. ERA5 is based on 4D-Var data assimilation using Cycle 41r2 of the Integrated Forecasting System (IFS), which was operational at ECMWF in 2016. ERA5 thus benefits from a decade of developments in model physics, core dynamics and data assimilation relative to ERA-Interim. In addition to a significantly enhanced horizontal resolution (31 km grid spacing compared to 79 km for ERA-Interim), ERA5 has a number of innovative features. These include hourly output throughout and an uncertainty estimate.
The uncertainty information is obtained from a 10-member ensemble of data assimilations with 3-hourly output at half the horizontal resolution (63 km grid spacing). Compared to ERA-Interim, ERA5 also provides an enhanced number of output parameters, including for example a 100 m wind product.
The move from ERA-Interim to ERA5 represents a step change in overall quality and level of detail. An overview of the main characteristics of ERA5 and a comparison with ERA-Interim is presented in Table 1. A more detailed description of the ERA5 configuration, how it was produced and how it fits into the framework of other reanalysis activities at ECMWF can be found in Hersbach et al. (2018).

	ERA-Interim	ERA5
Current availability	1979 onwards	1979 onwards
Availability by early 2020	Until August 2019 inclusive	1950 onwards
Availability behind real time	2-3 months	2-3 months (final product) 2-5 days (ERA5T)
IFS model cycle	31r2 (2006)	41r2 (2016)
Atmospheric data assimilation	12-hour 4D-Var	12-hour 4D-Var ensemble
Model input (radiation and surface)	As in operations, inconsistent SST and sea ice	Appropriate for climate, e.g. evolution of greenhouse gases, aerosols, SST and sea ice
Spatial resolution	79 km (TL255) 60 levels to 10 Pa	31 km (TL639) 137 levels to 1 Pa
Ocean waves	1 degree	0.36 degree
Land-surface model	TESSEL	HTESSEL
Uncertainty estimate	none	From the 4D-Var ensemble, 10 members at 63 km (TL319)
Output frequency	6-hourly for analyses 3-hourly for forecasts	Hourly throughout (uncertainty 3-hourly)
Output parameters	Extensive list	Extended list (e.g. 100-metre wind)
Dedicated land product	79 km, HTESSEL	9 km, HTESSEL

Table 1 Overview of the characteristics of ERA5 compared to ERA-Interim.

doi: 10.21957/vf291hehd7

40 years of data ready to use

Production of ERA5 started in early 2016 (Hersbach and Dee, 2016). During 2017 and 2018, segments of ERA5 were made publicly available in stages via the C3S Climate Data Store (CDS) (Raoult et al., 2017; see Box A on how to access the data). Finally, in January 2019, the complete set of hourly ERA5 data from 1979 onwards was published on the CDS. At this time, both the ERA5 and ERA-Interim archives contained exactly 40 years of climate data. This important milestone allows users to compute ERA5 climatologies going back to 1979. The time has now come to phase out ERA-Interim, which is limited in its capacity to use several new important types of observational data, is increasingly difficult to maintain and will not be migrated to ECMWF's future data centre in Bologna. New data covering the period to the end of August 2019 will continue to be made available with a delay of two to three months; after that, the production of ERA-Interim will stop. ERA5 on the other hand will be maintained as an operational product to at least the mid-2020s, when a replacement should be available. There will be early data availability only two to five days behind real time (ERA5T), and a quality-checked final product with a delay of two to three months.

Accessing ERA5 data

A

For practical reasons, ERA5 is produced in a number of parallel streams, with typically one-year overlaps to minimise spin-up effects. Data from these production streams have been consolidated into one continuous dataset in the MARS archive (currently around 5 petabytes). To ensure fast access to ERA5 data, a post-processed product (currently around 1 petabyte), has been made available on the Climate Data Store (CDS) cloud server (*https://cds.climate.copernicus.eu/*). This includes upper-air parameters on 37 pressure levels from 1,000 hPa to 1 hPa, and a large number of near-surface parameters and other two-dimensional fields. The CDS data have been converted from the native reduced-Gaussian grid to a regular lat–lon grid (0.25 degrees for the high-resolution deterministic reanalysis and 0.5 degrees for the ensemble products; 0.5 degrees and 1 degree, respectively, for ocean wave products). Several parameters, such as precipitation, surface fluxes and minimum and maximum temperatures, are provided in the CDS as hourly timeseries that combine hourly analysis fields with short-range forecasts as needed. This simplifies many technical difficulties that users have encountered in the past when retrieving ECMWF reanalysis data. Expert users can also use the CDS to access additional ERA5 data stored in MARS, including model-level data up to 1 Pa. However, data retrieval will likely be much slower in that case.

The CDS offers various ways to interact with ERA5 data. Data can be selected and downloaded using a web form or by means of python scripts using the CDS Application Programming Interface (API). Alternatively, users can perform simple operations on the data (calculations, plotting, etc.) online by using the CDS Toolbox. This avoids the need to download large data volumes and is in line with the principle that big datasets are best handled by bringing tools to the data rather than data to the tools.

The value of reanalysis

The role of reanalyses in climate monitoring applications is now widely recognized. ECMWF reanalysis is the basis for monthly C3S climate bulletins for surface air temperature, sea ice and hydrological variables (*https://climate.copernicus.eu/monthly-maps-and-charts*). In addition, ERA-Interim is used regularly, together with other datasets, as input for the World Meteorological Organization's annual assessment of the State of the Climate presented at the Conference of the Parties of the United Nations Framework Convention on Climate Change (UNFCCC). ERA-Interim is also a resource for the production of Essential Climate Variables (ECVs) and Climate Indicators recommended by the Global Climate Observing System (GCOS). By optimally combining observations and models, reanalyses provide consistent 'maps without gaps' of ECVs and a coherent representation of the main Earth system cycles (e.g. water, energy). ERA5 is beginning to replace ERA-Interim for these purposes.

Reanalysis is also widely used operationally at ECMWF, and by diverse communities for research and commercial purposes. Since 1 January 2015, ERA-Interim data have been downloaded by approximately 40,000 users worldwide. The move from ERA-Interim to ERA5 is now in full swing. At ECMWF, reanalysis data are used as a reference to evaluate trends in the skill of the operational high-resolution forecast (HRES). Reanalysis also provides the initial conditions for re-forecasts. These are in turn used to produce the climatologies needed for probabilistic forecast products, such as the Extreme Forecast Index, and for widely used forecast scores, such as the anomaly correlation coefficient. Re-forecasts are also used to calibrate ensemble forecasts in the extended and seasonal range.

Observations used

Over time the observing system has evolved dramatically. A timeline of data that were used in ERA5 from 1979 is provided in Figure 1. The number of observations assimilated in ERA5 has increased from approximately 0.75 million per day on average in 1979 to around 24 million per day by the end of 2018. Satellite radiances are the dominant and growing type of data throughout the period. Major developments for this class of observations have included the transition from the TOVS to the ATOVS suite of sounding instruments from 1998, the introduction of hyperspectral infrared radiances since 2002, and the increasing availability of data from a growing number of microwave imagers. There has also been a marked increase in the number of other satellite observations assimilated, notably GNSS-Radio Occultation bending angles in large quantities from 2006, scatterometer ocean vector wind and altimeter wave height data (both from 1991), and Level-2 ozone products. ERA5 also assimilates information on precipitation from ground-based radar observations (2009 onwards). The volume of conventional data has increased steadily throughout the period.

Compared to ERA-Interim, ERA5 benefits from many improvements in the observation operators, which convert model values to observation equivalents, and in the handling of observations in the IFS. It uses RTTOV-11 as the observation operator for radiance data instead of the RTTOV-7 operator used in ERA-Interim. It also assimilates a number of humidity-sensitive satellite channels using the all-sky approach instead of the clear-sky approach used in ERA-Interim. Besides providing new information in cloudy and precipitating areas, this also rectifies a problem with an earlier assimilation technique of radiances in rainy conditions that led to anomalous precipitation in ERA-Interim over the global ocean in the 1990s.

Improvements in the characterisation, inter-calibration and processing of conventional and satellite data enable providers to progressively refine the quality of historical observations and extend their geographic and temporal coverage. ERA5 has made use of several reprocessed satellite datasets, which were acquired from space agencies and institutes based in Europe, the US and Japan. These include atmospheric motion vector winds; ozone, radio occultation and altimetry data; soil moisture and wind data from scatterometers; and the SSMI record of satellite data sensitive to humidity over the ocean.

Overall, ERA5 has used many more observations than ERA-Interim, which is unable to ingest data from the latest satellite instruments, such as hyperspectral data from IASI and CrIS, or ground-based radar data. ERA-Interim thus suffers from a gradual decline in the number of observations as instruments and channels fail. ERA5 can also handle the BUFR format increasingly adopted for surface pressure, upper air wind and temperature data since 2013. ERA-Interim cannot handle this format and, as a result, it has experienced a sharp decline in the number of observations assimilated. At the end of 2018, ERA5 used about 24 million observations per day, about five times more than ERA-Interim.

Dedicated model input for ERA5

In addition to assimilated observations, the IFS relies on information about radiative forcing and boundary conditions. ERA5 makes use of the findings of the EU-funded ERA-CLIM collaborative research project (2011 to 2013) on the selection of forcing data that are suitable for climate reanalysis. For radiation, ERA5 includes forcings for total solar irradiance, ozone, greenhouse gases and some aerosols developed for the World Climate Research Programme (WCRP) initiative CMIP5, including stratospheric sulphate aerosols. This represents a major improvement on ERA-Interim, which, for example, does not account for stratospheric sulphate aerosols due to major volcanic eruptions.

 doi: 10.21957/vf291hehd7

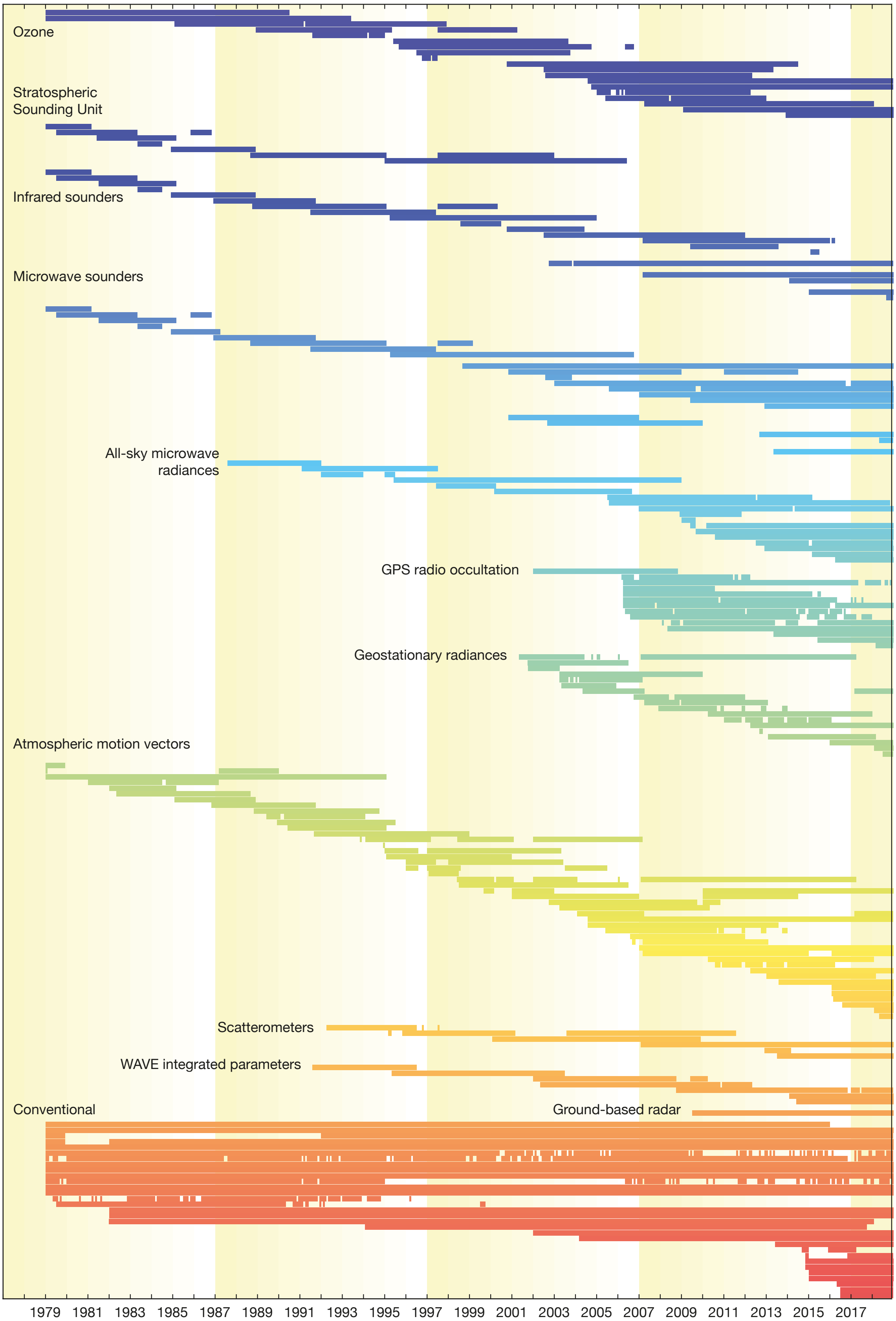


Figure 1 Data usage in ERA5 for the segment from 1979. Each horizontal bar represents the use of a particular satellite instrument or ground-based radar or a particular source of conventional data, such as weather stations, aircraft, ships, buoys and radiosondes. (*Image courtesy of Paul Poli*)

The evolution of sea-surface temperature (SST) and sea ice cover is based on a number of products for different periods of time: the UK Met Office Hadley Centre HadISST2 product for SST and sea ice, the EUMETSAT OSI-SAF reprocessed product for sea ice, and the UK Met Office OSTIA product for SST and sea ice that is also used in ECMWF's operational forecasting system. Details can be found in Hirahara et al. (2016). The aim was to produce a merged dataset that i) is as accurate as possible at each moment in time, ii) has no significant discontinuities, and iii) can be reliably updated close to real time. All these input datasets vary daily. The long-term evolution of SST and sea ice cover according to the merged dataset used in ERA5 is shown in Figure 2. The global-mean SST shows the impact of global warming from the mid-1970s as well as the influence of El Niño events (e.g. in 1997/98 and 2015/16) and major volcanic eruptions. Arctic sea ice shows a general decline over time, especially during summer time.

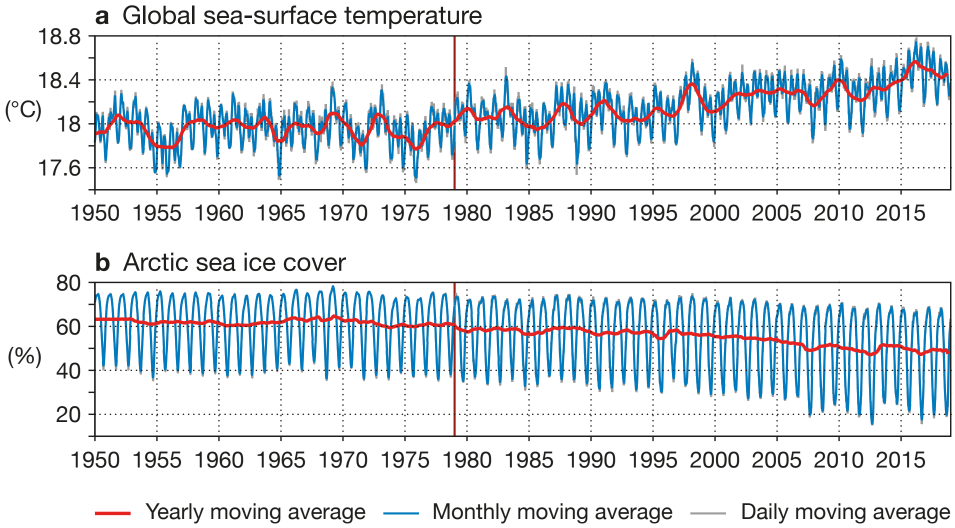


Figure 2 Time series of running mean averages of (a) global sea-surface temperature and (b) Arctic sea ice cover (60°N to 90°N) as it is used in ERA5 (right of the vertical line) or will be used in ERA5 (left of the vertical line).

The ERA5 uncertainty estimate

Reanalysis provides data that cover the whole globe at any time as a result of a data assimilation system that blends model information with observations. Reanalysis data are more accurate overall today than they were in the pre-satellite era, when observations were relatively sparse. The 10-member ensemble 4D-Var system used in ERA5 provides estimates of uncertainties in the data. These estimates depend on flow-dependent uncertainties in the short-range forecasts used in data assimilation. They also crucially depend on observation coverage. This is illustrated in Figure 3, which shows the evolution of sub-daily ensemble spread for upper-air temperature averaged over the months March to May for particular years. The CERA-20C reanalysis, produced as part of the EU-funded ERA-CLIM2 project, was the first ECMWF coupled centennial reanalysis to include an ensemble component (Laloyaux et al., 2018). Its atmosphere component assimilated only surface pressure and marine wind data. As a result, as shown in Figure 3a for the year 1971, areas where its ensemble spread is relatively small (i.e. where confidence in the data is relatively high) are limited to the lower part of the troposphere in the tropics and the extratropical northern hemisphere, where the bulk of data coverage is. Elsewhere the spread is larger (i.e. confidence is lower). For the same year, the ERA5 ensemble has much smaller spread over the entire northern hemisphere troposphere and lower stratosphere, thanks to the additional usage of upper-air data (Figure 3b). For the early satellite era, comprising mainly TOVS instruments, the situation has much improved over the southern hemisphere (Figure 3c), while in 2018, with the full current observing system, higher confidence is obtained almost everywhere (Figure 3d).

 doi: 10.21957/vf291hehd7

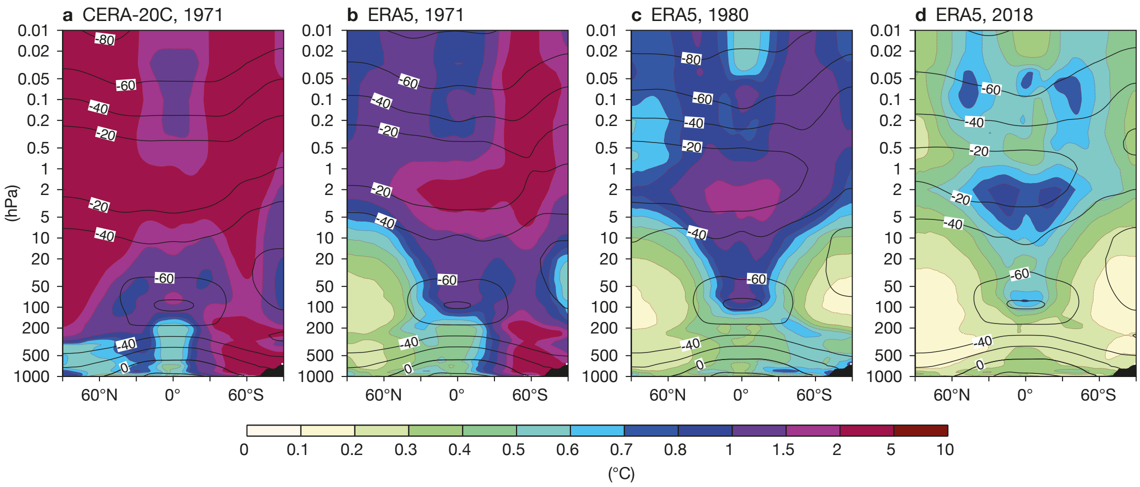


Figure 3 Zonal-mean cross section using a logarithmic pressure scale (hPa) of sub-daily ensemble spread of upper-air temperature averaged over March–April–May for (a) CERA-20C in 1971, (b) preliminary ERA5 data in 1971, (c) ERA5 in 1980, and (d) ERA5 in 2018. The contours show isotherms (in °C).

Improved weather and climate data

Re-forecasts starting from the ERA5 reanalysis show a gain of up to about a day in skilful range with respect to ERA-Interim, as illustrated in Figure 4. In general, tropical cyclones are much better resolved with lower central pressure and more realistic amounts of precipitation. The hourly time resolution enables a much-refined view of the evolution of day-to-day weather systems. An example comparing ERA5 and ERA-Interim output is shown in Figure 5, which shows the monthly-mean precipitation in the North Atlantic for September 2017, when rainfall is dominated by the contribution from tropical cyclones. ERA5 shows much more detail than ERA-Interim and is much closer to the precipitation in the first 12 hours of ECMWF's high-resolution forecast (HRES) operational at the time.

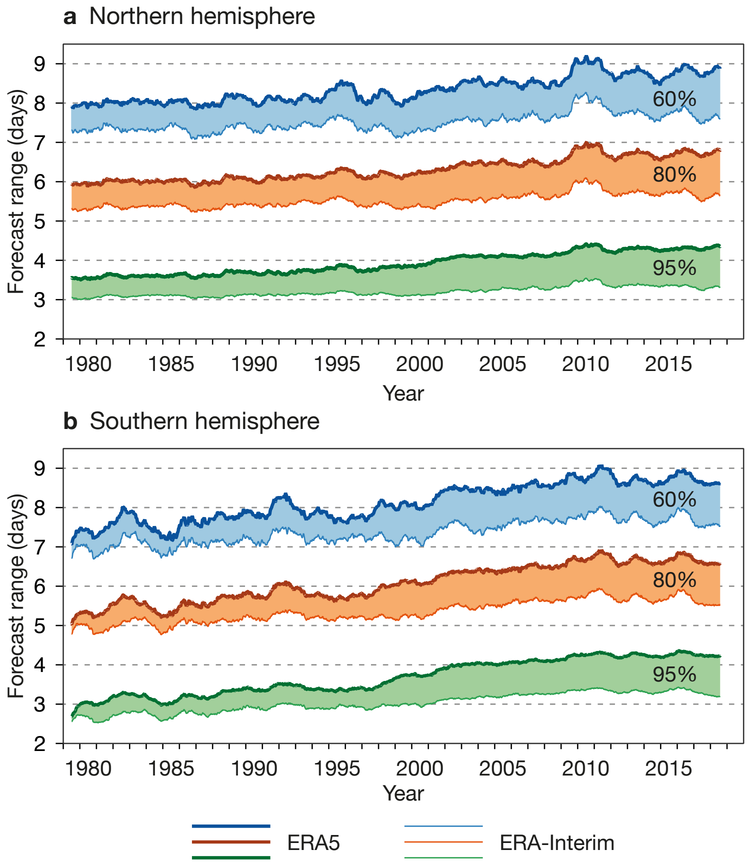


Figure 4 Range at which running 365-day mean anomaly correlations of 500 hPa height forecasts from 00 and 12 UTC reach 95%, 80% and 60% for (a) the extratropical northern hemisphere and (b) the extratropical southern hemisphere, from 1979 onwards. The shading highlights the difference between ERA5 and ERA-Interim.

doi: 10.21957/vf291hehd7

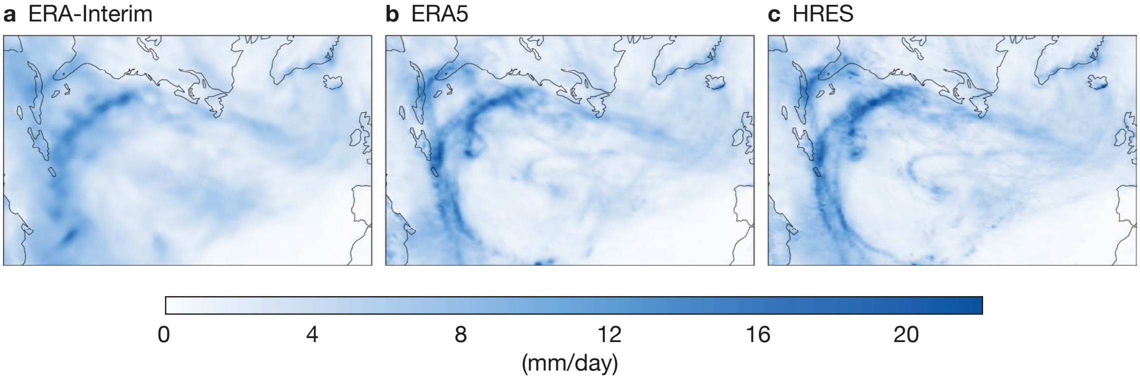


Figure 5 Mean precipitation rate for September 2017 over the North Atlantic from (a) ERA-Interim, (b) ERA5 and (c) ECMWF's operational high-resolution forecast (HRES).

One way of assessing the validity, strengths and weaknesses of a global climate reanalysis is by inter-comparison. Global reanalyses are produced at several institutes worldwide. Examples are NASA's MERRA-2 reanalysis (1980 to the present) and the Japanese JRA-55 reanalysis (1958 to the present). Figure 6 shows the results of these two reanalyses and ERA5 and ERA-Interim for global-mean temperature anomalies over the troposphere and the lower stratosphere (up to 10 hPa) for the period from 1979. In general, the differences are reassuringly small. All four reanalyses show a warming of the troposphere and a cooling of the stratosphere. The enhanced positive temperature anomalies in the lower stratosphere in 1982 and 1991 are the response to the eruptions of El Chichón and Mount Pinatubo, respectively. During these events, large amounts of sulphate aerosols were produced and stayed in the lower stratosphere for many months. Absorption of longwave radiation at these heights led to an increase in stratospheric temperature, while increased scattering of solar radiation to space resulted in a small cooling of the global average surface temperature. This is better reflected in ERA5 than in ERA-Interim, as confirmed by a comparison with radiosonde observations (not shown). The reason is that the ERA5 model 'knows' about these events through the applied CMIP5 forcing for volcanic sulphate. Although these events are also reflected in assimilated observations, the response in ERA-Interim is slightly tempered by the fact that there is no corresponding forcing in the model.

Several subtle differences between the four reanalyses can be seen in the lower stratosphere. ERA-Interim shows a rather patchy behaviour in the 2000s, while the cooling in the stratosphere in MERRA-2 is enhanced by the assimilation of Microwave Limb Sounder temperature data from August 2004. In ERA5, the abrupt transition to warmer temperatures in 2000 from 20 hPa upwards is artificial and is the result of a change in the data assimilation configuration, specifically the choice of background error covariances. Details are described in Hersbach et al. (2018), which also includes a description of other issues encountered during the production of ERA5. Several of these could be rectified prior to publication of the dataset in the CDS. In the top part of the stratosphere and in the mesosphere, the agreement between different reanalysis products is much poorer (not shown). For example, ERA5 suffers from a spurious mesospheric equatorial jet that occurred in IFS Cycle 41r2. In later IFS cycles, this has been rectified.

 doi: 10.21957/vf291hehd7

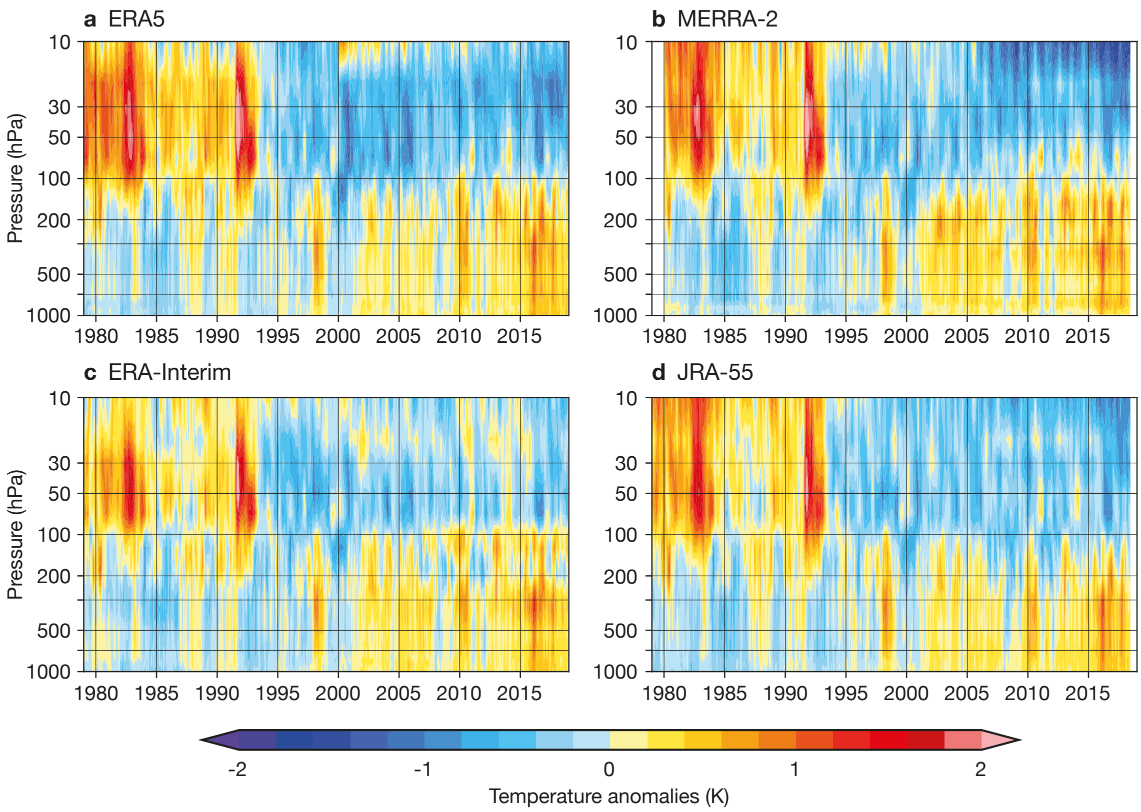


Figure 6 Vertical profiles of monthly and globally averaged temperature anomalies as a function of time from 1979 to 2018 from (a) ERA5, (b) the MERRA-2 reanalysis, (c) ERA-Interim, and (d) the JRA-55 reanalysis. Each monthly anomaly was computed by removing the 1981–2010 mean for the corresponding month.

All these reanalysis products show a consistent pattern of global warming. Figure 7 shows the point-wise linear trend in surface temperature between 1979 and 2018 from ERA5 in the Arctic region. The Arctic amplification is clearly visible. Especially regions where sea ice has receded over time show an increase in temperature that exceeds the global average trend of 0.18 K per decade many times over.

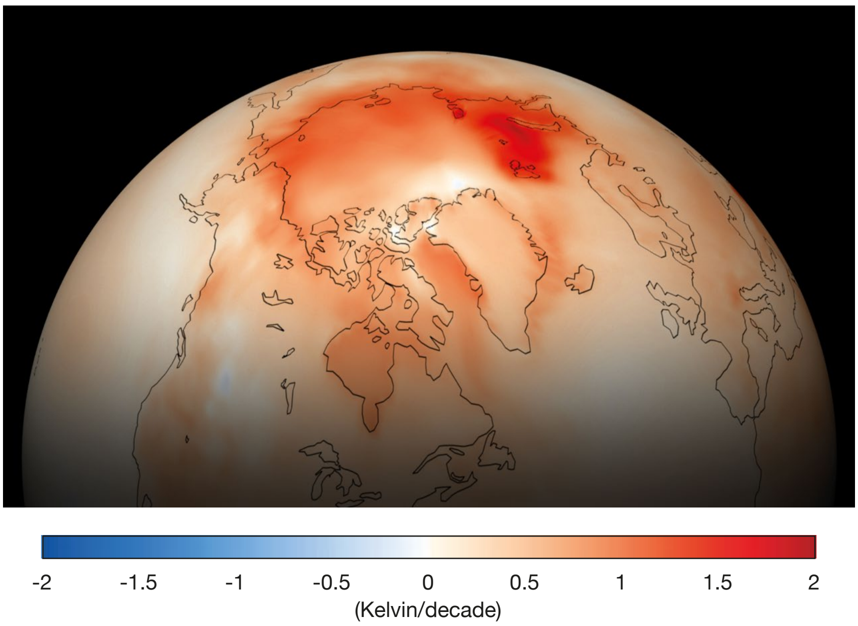


Figure 7 Point-wise linear trend (Kelvin/decade) in surface temperature between 1979 and 2018 from ERA5 over the Arctic region. The global-mean linear trend is 0.18 Kelvin/decade.

Still to come…

The reanalysis of the period 1950 to 1978 is currently in production. It includes VTPR satellite instrument data from the 1970s that were assimilated in ERA-40 and JRA-55, and BUV ozone data. It also assimilates historical in situ and upper-air data that were used in ERA-40, augmented with historical holdings that were used in CERA-20C (surface pressure) and an ERA-CLIM pilot reanalysis using upper-air data, as well as recently acquired surface data from the National Centers for Environmental Prediction (NCEP) that precede the ERA-40 BUFR archive.

ERA5 will be complemented with an enhanced land-surface dataset (ERA5-Land) at 9 km global resolution, which is produced by running the IFS land-surface model using meteorological input from ERA5. ERA5-Land data from 1979 onwards will be made available in the CDS during 2019 (see separate article in this Newsletter). Later it will be extended to 1950 as well. Also in 2019, the CDS will provide access to observations assimilated in ERA5, together with detailed information about data use, statistics on data fit, and estimates of data biases used in the 4D-Var data assimilation scheme.

The ongoing production of ERA5 is undertaken at ECMWF within the Copernicus C3S framework. Many other reanalysis-related tasks are being carried out by C3S using providers outside ECMWF. For example, two high-resolution regional reanalyses, for Europe and the Arctic, are under way and will deliver results by 2019/20. ECMWF has also awarded several contracts for the preparation of input observations for climate reanalysis. These address the need for satellite data reprocessing, data rescue (both satellite and conventional), and the collection of in situ surface and upper-air data into well-maintained archives.

All these datasets will feed into the next generation of global and regional reanalyses. ECMWF's vision for C3S post-2020 continues to give high priority to reanalysis, in line with the Centre's Strategy to 2026. A centennial global reanalysis back to possibly 1850 is proposed to start by 2021. The production of the next full-observing-system reanalysis, ERA6, is planned to start by 2023. This vision is currently being discussed with the European Commission in the context of defining the evolution of Copernicus services beyond 2021.

Further reading

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

Hersbach, **H.**, **P. de Rosnay**, **B. Bell**, **D. Schepers**, **A. Simmons**, **C. Soci**, **S. Abdalla**, **M. Alonso Balmaseda**, **G. Balsamo**, **P. Bechtold**, **P. Berrisford**, **J. Bidlot**, **E. de Boisséson**, **M. Bonavita**, **P. Browne**, **R. Buizza**, **P. Dahlgren**, **D. Dee**, **R. Dragani**, **M. Diamantakis**, **J. Flemming**, **R. Forbes**, **A. Geer**, **T. Haiden**, **E. Hólm**, **L. Haimberger**, **R. Hogan**, **A. Horányi**, **M. Janisková**, **P. Laloyaux**, **P. Lopez**, **J. Muñoz-Sabater**, **C. Peubey**, **R. Radu**, **D. Richardson**, **J.-N. Thépaut**, **F. Vitart**, **X. Yang**, **E. Zsótér** & **H. Zuo**, 2018: Operational global reanalysis: progress, future directions and synergies with NWP, *ECMWF ERA Report Series* **27**.

Hirahara, **S.**, **M. Alonso Balmaseda**, **E. de Boisseson**, & **H. Hersbach**, 2016: Sea Surface Temperature and Sea Ice Concentration for ERA5, *ECMWF ERA Report Series* **26**.

Laloyaux, **P.**, **E. de Boisseson**, **M. Balmaseda**, **J. Bidlot**, **S. Broennimann**, **R. Buizza**, **P. Dalhgren**, **D. Dee**, **L. Haimberger**, **H. Hersbach**, **Y. Kosaka**, **M. Martin**, **P. Poli**, **N. Rayner**, **E. Rustemeier** & **D. Schepers**, 2018: CERA-20C: A Coupled Reanalysis of the Twentieth Century, *Journal of Advances in Modeling Earth Systems* **10**, 1172–1195 doi:10.1029/2018MS001273.

Raoult, **B.**, **C. Bergeron**, **A. López Alós**, **J.-N. Thépaut** & **D. Dee**, 2017: Climate service develops user-friendly data store, *ECMWF Newsletter* **No. 151**, 22–27.

 doi: 10.21957/vf291hehd7