ECMWF
European Centre for Medium-Range Weather Forecasts

Newsletter

No. 159 | Spring 2019

Publication policy

The ECMWF Newsletter is published quarterly. Its purpose is to make users of ECMWF products, collaborators with ECMWF and the wider meteorological community aware of new developments at ECMWF and the use that can be made of ECMWF products. Most articles are prepared by staff at ECMWF, but articles are also welcome from people working elsewhere, especially those from Member States and Co-operating States. The ECMWF Newsletter is not peer-reviewed.

Any queries about the content or distribution of the ECMWF Newsletter should be sent to Georg.Lentze@ecmwf.int

Guidance about submitting an article is available at *www.ecmwf.int/en/about/media-centre/media-resources*

New observations

Progress in numerical weather prediction depends on advances in many areas: in the methods used to solve the equations describing global atmospheric flow; in the physics package representing smaller-scale processes; in the models of different Earth system components and their coupling; in the assimilation of Earth system observations to make the initial conditions of forecasts as accurate as possible; and in computing to ensure the timely production and dissemination of increasingly high-resolution forecasts.

Several articles in this Newsletter illustrate the importance of continuous work in all these areas. The successful forecasting of freezing rain in Romania in January this year would not have been possible without prior work on the correct modelling of melting and refreezing processes. Building on external work, changes in the ocean wave model to be implemented in ECMWF's Integrated Forecasting System (IFS Cycle 46r1) in June this year will improve wave predictions. A new collaboration seeks to address challenging issues in the modelling of orographic drag. The new ECMWF Production Data Store will ensure the continued timely collection of weather data and delivery of forecasts as data volumes grow.

There is another factor fundamental to the success of global numerical weather prediction: the availability of a wide variety of timely, high-quality Earth system observations covering all parts of the globe. ECMWF's new ERA5 reanalysis of the global weather and climate, presented in this Newsletter, illustrates just how important such observations are. ERA5 includes an ensemble of reanalyses, the spread of which gives an indication of the confidence we can have in ERA5 data. If you look at how that spread has changed over the last five decades, you can see how adding new and better-quality observations has successively reduced the spread and thus increased the confidence we can have in the data.

That is why recent additions to the observing system, such as EUMETSAT's Metop-C satellite and the deployment of more buoys carrying pressure sensors in the northeast Pacific, are so important. As described in this Newsletter, they help to refine our estimate of the all-important initial conditions at the start of every forecast. New observations are also benefitting the atmospheric composition forecasts produced by the EU-funded Copernicus Atmosphere Monitoring Service (CAMS) implemented by ECMWF: data from the Sentinel-5P satellite, which is part of the EU's Copernicus Earth observation programme, are now being assimilated by CAMS and are improving CAMS air quality data. Since atmospheric composition interacts with weather processes, we are also working to improve our use of atmospheric composition data for the benefit of numerical weather prediction.

Developing the global observing system requires collaboration between many players. That is why ECMWF collaborates closely with a range of agencies developing and operating weather satellites. In bilateral meetings earlier this year, we reaffirmed our commitment to working with EUMETSAT, ESA and the China Meteorological Administration to ensure that new weather observations will continue to improve forecasts for the benefit of society.

Florence Rabier
Director-General

Contents

Editor Georg Lentze • **Typesetting & Graphics** Anabel Bowen • **Cover image** ERA5 linear temperature trend (see p. 23)

Forecasts of freezing rain in Romania

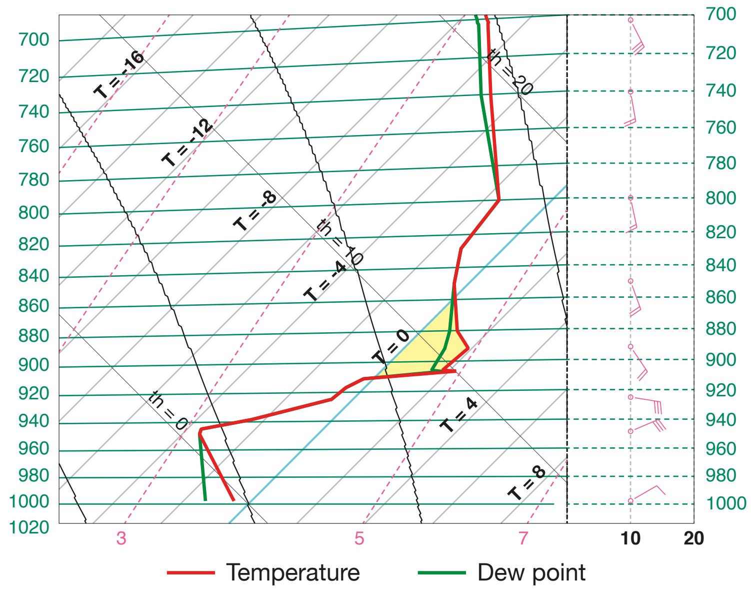
Linus Magnusson, Estibaliz Gascon, Ivan Tsonevsky

On 24 and 25 January 2019, a band of freezing rain affected south-eastern Europe and especially Romania. The precipitation rate was almost 1 mm/h for several hours, leading to a thick layer of ice on the ground. Motorways and many other national roads in the southern and eastern part of the country were closed from the evening of 24 January to the morning of 25 January, and many trains and flights were delayed or cancelled. Ice deposits on electrical cables led to power cuts, and the emergency services reported about 100 people with fractures or sprains due to falling on the ice.

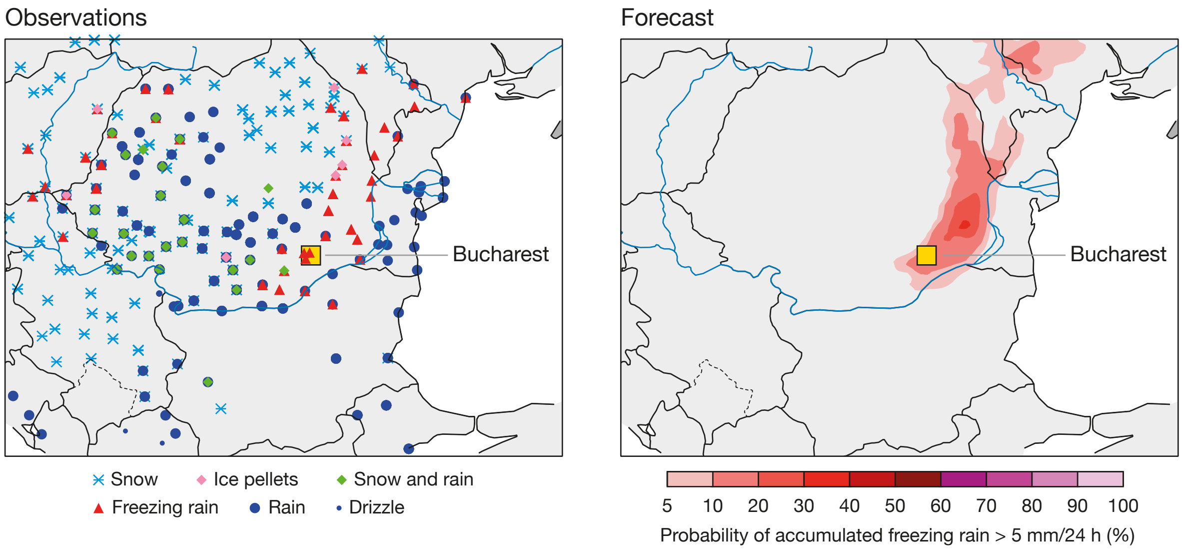
The freezing rain event was associated with a deep Mediterranean cyclone that formed over the Gulf of Genoa on 23 January. Over the following two days, this low pressure system moved initially to the southeast and later eastwards towards Greece, bringing adverse weather conditions to the Balkan Peninsula. A combination of warm advection aloft in the eastern flank of the upper-level low and cold advection from the north and northeast near the surface created favourable conditions for freezing rain that affected southern and eastern Romania on 24 and 25 January.

Predicting freezing rain

In freezing rain conditions, precipitation falls through a layer of air whose temperature is above zero and which is thick enough for the precipitation to have time to melt, before passing through a colder layer closest to the ground producing super-cooled water, which turns to ice when it hits the ground. Such a temperature profile was evident in the radiosonde sounding from 00 UTC on 25 January from Bucharest. In a warm layer between 910 hPa and 860 hPa the temperature was above freezing, while the 2-metre temperature was –0.5°C.



Radiosonde sounding. Sounding from Bucharest for 25 January 00 UTC. The area highlighted in yellow indicates the layer in which the temperature was above 0°C, with colder air below and above.

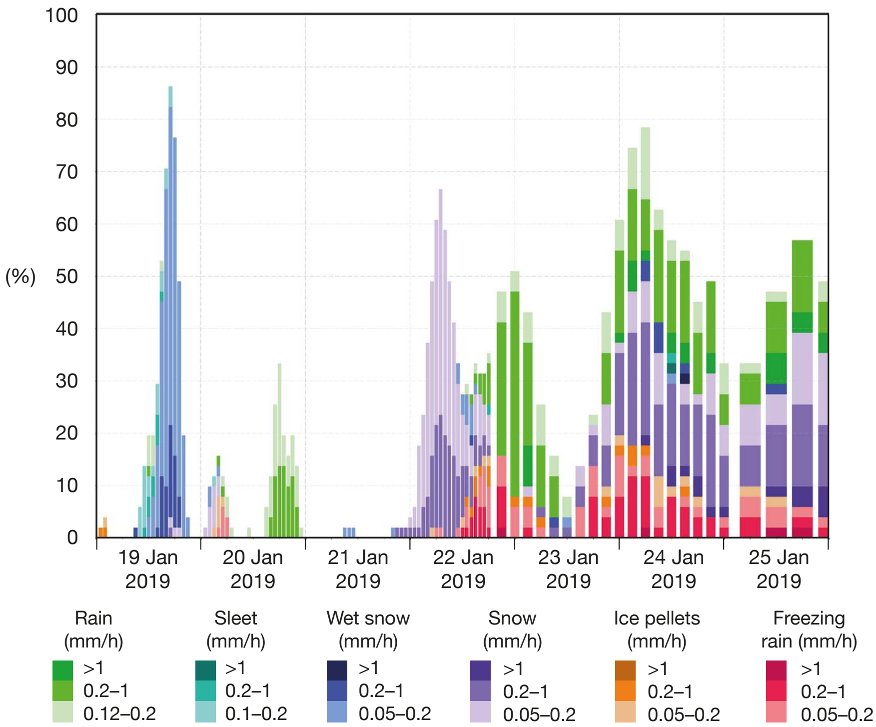


Freezing rain observations and forecast. Observations of precipitation types between 24 January 06 UTC and 25 January 06 UTC (left) and ensemble forecast probability for more than 5 mm of freezing rain in the forecast from 21 January 00 UTC (right) valid for the same period as the observations.

Since 2015, ECMWF has produced freezing rain precipitation type as a model diagnostic. To be able to predict the correct precipitation type at the surface, including freezing rain, the atmospheric model needs to predict the vertical temperature profile and the amount of precipitation sufficiently accurately, but must also include the correct physics for the melting and refreezing processes.The diagnostic is described in ECMWF Newsletter No. 141. Subsequently the precipitation type diagnostic was used for new ensemble-based probabilistic precipitation type products, described in ECMWF Newsletter No. 154.

ECMWF's forecasts of the event

In the forecast from 21 January 00 UTC, the probability for more than 5 mm/24 h of freezing rain (a threshold representing a severe event) in the affected area reached 30% for 24 January 06 UTC to 25 January 06 UTC. This shows that the forecast was able to capture the horizontal structure of the event three to four days ahead of the event. For point forecasts, ECMWF has developed a special ensemble product for precipitation type showing the probability of precipitation. It depicts the temporal evolution of probabilities (in per cent) for a specific location in the form of a bar chart. Probabilities are calculated from the instantaneous precipitation type variable. Different colours represent different precipitation types and different hues provide details on the probability of different instantaneous precipitation rates for different precipitation types. The information on precipitation rates can be key for determining the severity of a potential freezing rain event. Three different rate categories are used: < 0.2 mm/hour, 0.2–1 mm/hour and > 1 mm/hour. As an example, we show the forecast for Bucharest from 19 October, five to six days before the event. It shows an elevated risk for freezing rain on 24 and 25 October. Although the probabilities for freezing rain are low, the forecast gives an early indication of a possible freezing rain event five to six days later.



Precipitation type meteogram product. Forecast for a location near Bucharest, Romania, starting at 00 UTC on 19 January 2019.

Operational assimilation of Metop-C data

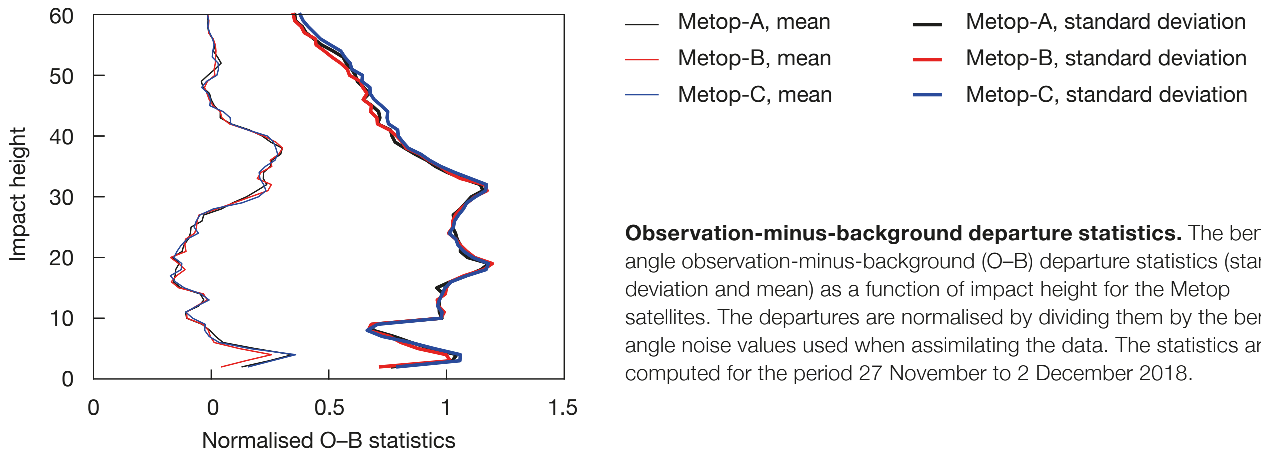
Sean Healy, Niels Bormann, Katie Lean

ECMWF is now assimilating measurements from EUMETSAT's polar-orbiting Metop-C satellite, which was launched on 7 November 2018. The new Metop-C data used at ECMWF includes radio occultation (RO) bending angles from the GRAS instrument, radiances from the AMSU-A and MHS instruments, and atmospheric motion vector (AMV) dual-Metop wind information derived from the AVHRR instrument. These changes followed evaluation and testing of the data in the ECMWF system between December 2018 and February 2019.

The Metop-C satellite joins Metop-A and Metop-B in a 'tristar' configuration in a morning orbit (i.e. passing over the equator at 09:30 local time). These satellites provide key observations for both global and regional numerical weather prediction (NWP).

Metop-C GRAS data

The addition of Metop-C GRAS measurements significantly increases the total number of RO profiles available for operational assimilation by around 30%, bringing EUMETSAT's contribution to typically 1,900 RO profiles per day. The EUMETSAT RO team produced high-quality bending angle profiles by 13 November 2018, within six days of launch. The data was then made available to calibration–validation (CAL–VAL) partners, including the EUMETSAT Radio Occultation Meteorology Satellite Application Facility (ROM SAF), for evaluation. It was quickly established that the Metop-C bending



Observation-minus-background departure statistics. The bending angle observation-minus-background (O–B) departure statistics (standard deviation and mean) as a function of impact height for the Metop satellites. The departures are normalised by dividing them by the bending angle noise values used when assimilating the data. The statistics are computed for the period 27 November to 2 December 2018.

angles were of similar quality to both the Metop-A and B measurements, by comparing them with NWP information mapped to observation space. Assimilation experiments with Metop-C GRAS over a three-month period were then conducted. These showed that the new measurements improved the agreement of ECMWF short-range forecasts with other conventional and satellite measurements, and there was some improvement in stratospheric temperature biases. However, the impact of Metop-C GRAS on medium-range tropospheric forecasts was not found to be statistically significant for this period. Consequently, while assimilation of the data started on 14 March 2019, these assimilation experiments are being extended.

Metop-C AMSU-A and MHS data

Metop-C AMSU-A and MHS data were also made available quickly in November 2018, and the data quality was evaluated by comparing observations against equivalent values calculated from short-range forecasts. The AMSU-A comparisons showed that the overall data quality was consistent with that of other AMSU-A instruments currently in orbit. The Metop-C MHS exhibited poorer noise performance when compared to the instruments on the other Metop satellites for two of the three sounding channels, with some variations of the increased noise over time. Whilst this suggested that it might not have the impact of the other MHS instruments, it did not rule out testing the active assimilation of the data. This is because errors in the short-range forecasts (the ‘background’) used in the data assimilation system for humidity are larger, and error contributions other than instrument noise tend to matter more, when we assimilate the data in all-sky conditions.

Short assimilation trials assimilating both AMSU-A and MHS over three months produced a small but statistically significant positive forecast impact. This is particularly remarkable as the instruments were added as the 9th microwave temperature sounder and the 11th microwave humidity sounder in the assimilation. Furthermore, they are the third set of microwave instruments in the 09:30 orbit, already populated by Metop-A and B. Assimilation of the data started on 14 March 2019.

Dual-Metop AMVs with Metop-C

EUMETSAT provides both polar AMVs from a single Metop satellite and global AMVs derived from imagery using pairs of Metop satellites, including Metop-C, known as dual-Metop AMVs. The first polar AMVs derived from AVHRR data on Metop-C became available in early January 2019, and data characteristics for these are comparable to those from Metop-A or B. Assimilation trials are ongoing

New observations since January 2019

The following new observations have been activated in the operational ECMWF assimilation system since January 2019.

Observations	Main impact	Activation date
32 new drifting buoys in the northeast Pacific Ocean	Mean sea-level pressure	12 and 17 January 2019
Atmospheric Motion Vectors from a combination of Metop-B and -C AVHRR imagery	Tropospheric wind	17 January 2019
Radiances from AMSU-A and MHS on Metop-C	Temperature, humidity, clouds, dynamics	14 March 2019
Bending angles from GRAS on Metop-C	Temperature in upper troposphere/ lower stratosphere	14 March 2019

and activation of the data is expected for later in the year.

The dual-Metop AMVs – including Metop-C – have been used in the operational ECMWF system since 17 January 2019, when the new combination of Metop-C/B was included alongside the pairing of Metop-A/B. At ECMWF, the dual-Metop AMV information is used to fill the gap between the traditional geostationary AMVs and the single-satellite polar AMVs, in the latitude bands between 40 and 60 degrees in both hemispheres. A detailed analysis has shown that the dual-Metop AMVs have characteristics that depend on which satellite pair is used. These differences are primarily in the tropics, rather than where the data is used at ECMWF, and they are currently under further investigation by EUMETSAT.

Other instruments

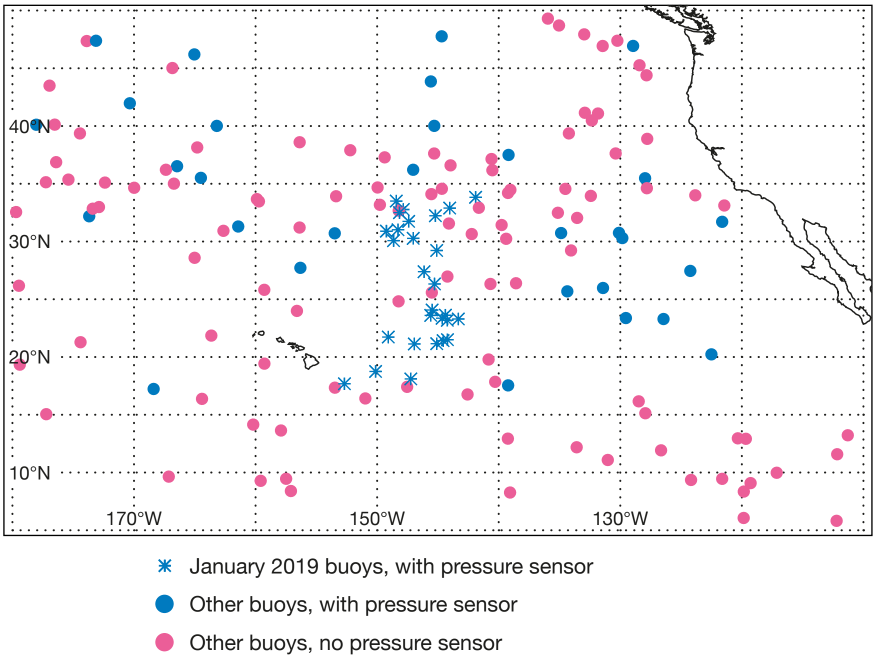
Test data from the Metop-C IASI instrument is expected in April 2019 and will be evaluated in a similar way as the other instruments. Surface wind and soil moisture information retrieved from Metop-C ASCAT measurements is currently being assessed.

New drifting buoys in the northeast Pacific

Bruce Ingleby, David Lavers (both ECMWF), Anna Wilson, Marty Ralph, Luca Centurioni (all Scripps Institution of Oceanography, US)

In January 2019, 32 drifting buoys were added to the global observing network in the northeast Pacific Ocean. Critically, these buoys have pressure sensors which provide important sea-level pressure observations in this data-sparse region. Drifting buoy pressure measurements are very important for improving numerical weather predictions because (1) pressure at mean sea level is an important variable linked to the main mode of extratropical synoptic variability; (2) in this and many other ocean areas there are very few other in situ observations; and (3) satellite data still provide only a small amount of information about pressure at mean sea level. The buoys usually last about two years and are a very cost-effective component of the global observing system. Sadly, up to 50% of drifting buoys are deployed without a pressure sensor.

The deployment of these new buoys, made possible by funds from NOAA's Global Drifter Program, the California Department of Water Resources and the U.S. Army Corps of Engineers, was prompted by a discussion during a visit by Scripps Institution of Oceanography researchers to ECMWF. The meeting, in early September 2018, was organised by Aneesh Subramanian and David Lavers. Preparations for the buoy deployment were co-ordinated by the Center for Western Weather and Water Extremes (CW3E) at Scripps in collaboration with ECMWF, the Lagrangian Drifter Laboratory also at Scripps, and the United States Air Force. These extra



Drifting buoys in the northeast Pacific Ocean. The chart shows all drifting buoys in the northeast Pacific Ocean on 11/12 March 2019. The 32 new ones were dropped in two lines in mid-January and gradually separated as they moved with the ocean currents. Only 28 of them are shown as four have stopped reporting. They also provide valuable observations of sea-surface temperature. In this area, only 36% of the drifting buoys (fewer before the deployment) have pressure sensors. In the North Atlantic, the proportion of drifting buoys that have pressure sensors is slightly higher, at about 42%.

drifting buoy observations formed part of a broader 'atmospheric river reconnaissance' field campaign, a multi-agency observational effort led by CW3E, involving dropsondes and additional soundings on the coast that took place in February and March 2019. The field campaign aims to improve forecasts of storms and extreme precipitation affecting California and adjacent regions. The addition of these buoys should benefit the ECMWF Integrated Forecasting System in the Pacific region and could lead to forecast skill improvements over Europe in the medium range.

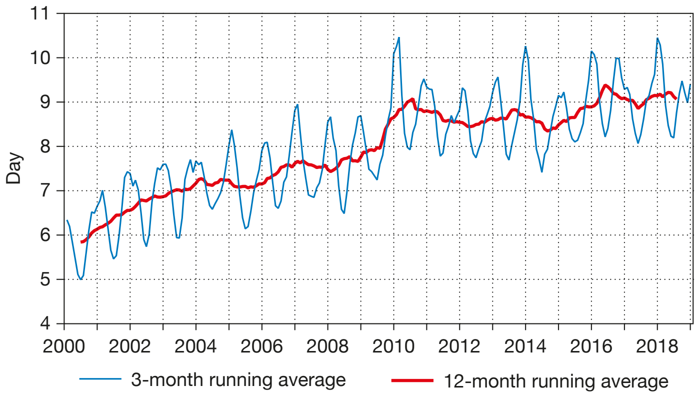
Further information on the impact of drifting buoy pressure data can be found in Ingleby & Isaksen (2018; doi.org/10.1002/asl.822), Centurioni et al. (2017; doi.org/10.1175/BAMS-D-15-00080.1), and Horányi et al. (2017; doi.org/10.1002/qj.2981).

Forecast performance 2018

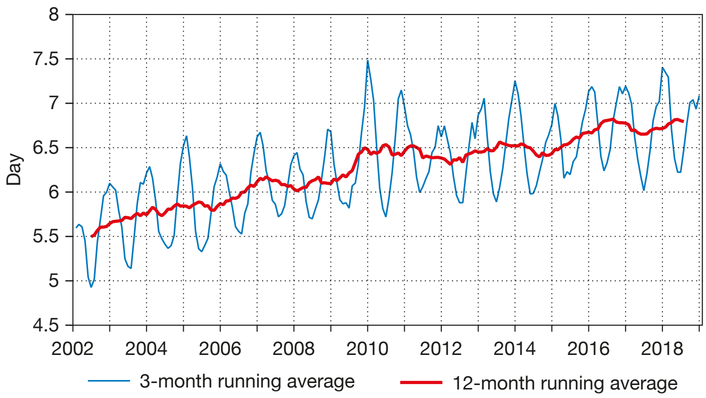
Thomas Haiden, David Richardson, Martin Janousek, Tim Hewson

ECMWF maintains a comprehensive range of verification statistics to evaluate the accuracy of its forecasts. Each year, a summary of verification results is presented to ECMWF's Technical Advisory Committee (TAC). Their views about the performance of the operational forecasting system in 2018 are given in the box.

The overall performance of the operational forecasts is summarised using a set of headline scores endorsed by the TAC, which highlight different aspects of forecast skill. Upper-air performance is monitored through the continuous ranked probability score (CRPS) for temperature at 850 hPa for the ensemble forecast (ENS) and the anomaly correlation of 500 hPa geopotential height of the high-resolution forecast (HRES), both over the northern hemisphere extratropics. After a slight decrease in predictability in 2017 (as inferred from comparison with the new ERA5 reanalysis), both ENS and HRES skill increased again in 2018. The most recent upgrade of the Integrated Forecasting System (IFS Cycle 45r1 on 5 June 2018) has brought improvements in the extratropics for some aspects of the forecast, notably an increase in precipitation forecast skill in the HRES, which is very clear when compared with a baseline of ERA-5-based forecasts, and improvements in the tropics for most parameters. In terms of surface skill, in 2018 there was a slight increase for 2-metre temperature. Forecasts of tropical cyclones have further improved in terms of position, intensity, and speed errors for both the ENS and HRES. With regard to ocean waves, ECMWF has maintained its lead compared to other global wave forecasting systems for forecasts of significant wave height, and its position among the leading systems for peak period. The change from La Niña to El Niño conditions in the first half of 2018 was well captured several months in advance although the positive sea-surface temperature anomalies were weaker than predicted. ECMWF forecasts predicted warm anomalies in Europe in spring and summer 2018 weeks in advance, while the northward extent and intraseasonal variations during the heatwave were reflected in forecasts up to two weeks ahead.



Skill of the ENS as measured by ECMWF's primary ENS headline score. Evolution of 850 hPa temperature ensemble forecast performance in the northern hemisphere extratropics, verified against the corresponding analysis. The chart shows 12-month and 3-month running average values of the forecast range at which the continuous ranked probability skill score (CRPSS) falls below 25%.



Skill of the HRES as measured by ECMWF's primary HRES headline score. Evolution of 500 hPa geopotential height forecast performance in the northern hemisphere extratropics, verified against the corresponding analysis. The chart shows 12-month running average values and 3-month running average values of the forecast range at which the anomaly correlation falls below 80%.

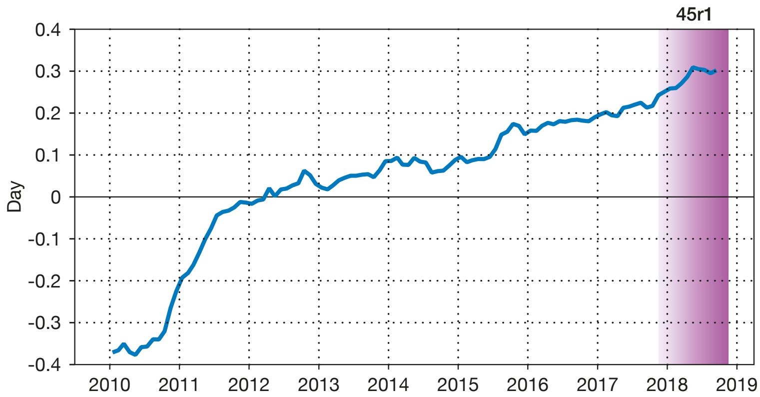
Each summer ECMWF invites Member and Co-operating States to submit updated reports on the application and verification of ECMWF's forecast products. Many reports focus on comparing HRES with forecasts produced by limited-area models (LAMs), and for this reason usually centre on the shorter ranges (up to about 48 h). A common finding was that biases in IFS forecasts have a diurnal cycle. Annual cycles are also often present. Overall, whilst there are large model-to-model and parameter-to-parameter variations, and some large differences between countries, LAM performance tends to be slightly better than HRES performance.

ENS-related verification was only included in a few reports. Comparisons between ENS and LAM-EPS systems were provided by Finland (2-metre temperature), Switzerland (precipitation) and Denmark (10-metre wind), while Hungary (surface parameters), the UK (tropical cyclones and regimes), Israel (ocean waves) and Switzerland (vertically-integrated water transport) provided some evaluation of ENS forecast outputs.

The complete set of annual results is available in two ECMWF Technical Memorandums, No. 831 on 'Evaluation of ECMWF forecasts, including the 2018 upgrade' and No. 840 on 'Use and Verification of ECMWF products in Member and Co-operating States (2018)'. Both are downloadable from http://www.ecmwf.int/en/research/publications.

The following are other sources of information about verification and forecasting system changes.

- Verification pages: http://www.ecmwf.int/en/forecasts/charts
- Inter-comparisons of global model forecast skill: http://apps.ecmwf.int/wmolcdnv/
- A list of 'Known IFS Forecasting Issues': https://confluence.ecmwf.int/display/FCST/Known+IFS+forecasting+issues
- All IFS forecasting system cycle changes since 1985: http://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model



Lead time gain, for HRES over ERA5-based forecasts, in 24-hour precipitation forecast skill. The chart relates to the forecast lead time at which the ECMWF SEEPS headline score for the extratropics drops to 45% (12-month running average values). If for HRES we call that time H, and for ERA5 forecasts we call that time E, then the plot shows H minus E, so above the zero line HRES skill is better. For the years represented, E is about 4 days. The x-axis dates denote the mid-point of each 12-month period. The purple shading shows the period during which the influence of IFS Cycle 45r1 on the score grows linearly from 0 to 1.

Assessment of ECMWF's Technical Advisory Committee, 11–12 October 2018

With regard to its overall view of the performance level of ECMWF's operational forecasting system, the Committee:

a) welcomed the new format of the presentations for the forecast performance and feedback from Member and Co-operating States as a webinar in advance of the TAC; appreciated that many questions raised during the webinars had been addressed by ECMWF in time for the TAC meeting;

b) noted that ECMWF headline scores continue to show high skill, and noted that the recent plateauing is due to natural variability in predictability, as revealed by comparison with ERA5;

c) welcomed improvement seen in the new headline score for large 2 m temperature errors and encouraged ECMWF to investigate a similar measure for 10 m wind;

d) took account of Member and Co-operating States' feedback on precipitation biases and encouraged further investigation;

e) appreciated the very good support ECMWF provided to Member and Co-operating States over the last year, in particular for high-impact weather events such as heat waves, cold spells, windstorms, and tropical cyclones, and through specific products such as the EFI, precipitation type, and improvements in EFAS and fire danger indices;

f) noted that the extended-range forecast was useful for assessing the duration of the summer heatwaves in Europe;

g) welcomed the improvements in snow, but noted the problem of excessive accumulation of wet snow, and slowness in melting accumulated snow; welcomed the changes in the forthcoming cycle to address the first issue and encouraged future work including the multi-level snow model;

h) noted the degradation in spread–error at the longer ranges (T+168 and T+240) from 2013, appreciated the work to understand the changes and encouraged ECMWF to closely monitor this in future;

i) noted that the recent drop in skill for 24-hour precipitation accumulations in Europe is also seen in the ERA5 verification, which indicates that it is related to year-to-year atmospheric variability; encourages Member and Co-operating States to investigate whether this is also seen in high-resolution, convection-resolving LAMs over Europe and report back to ECMWF;

j) welcomed plans to investigate and introduce monitoring of forecast jumpiness and 'flip-flopping';

k) acknowledged that the new products relating to CAPE, CAPE-shear are well appreciated, and welcomed ECMWF plans to review the computation of CAPE, and encouraged exchanges with experts in Member and Co-operating States on this;

l) welcomed the new online Forecast User Guide with the flexibility for more frequent updating;

m) welcomed ECMWF's role in WMO verification, acting as Lead Centre for deterministic NWP verification, and now also as Lead Centre for ocean wave forecast verification;

n) noted the potential impact of data cut-off time and forecast delivery time on the interpretation of forecast scores, and would welcome further work to take this effect into account in forecast comparisons;

o) appreciated the use of high-density observations to improve the verification and the demonstration of their benefits in model evaluation, and encouraged other countries to provide such data;

p) encouraged ECMWF to make use of high-resolution calibrated radar and/or other gridded precipitation datasets in verification;

q) appreciated ECMWF's invitation to Member and Co-operating States to identify additional products and services for use by forecasters, and appreciated ECMWF's management of these requests with the URMS to assess priorities;

r) appreciated the value of ECMWF training courses in increasing the benefit of ECMWF forecasts to Member and Co-operating States' forecasters and end users; welcomed the new blended format combining e-learning with classroom face-to-face interactions; but stressed that such face-to-face interactions remain important and should not be reduced further.

First ERA5-Land dataset to be released this spring

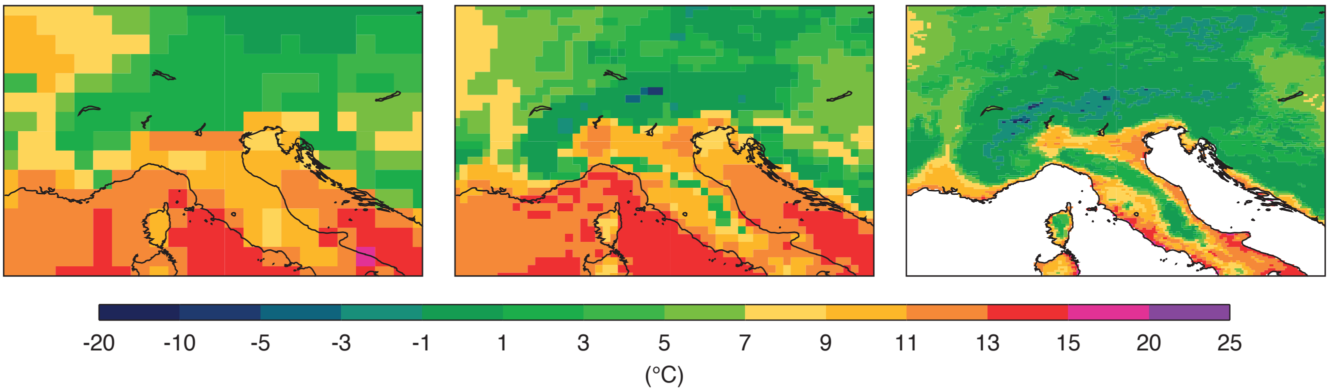
Joaquín Muñoz Sabater

Following the recent release of ECMWF's ERA5 climate reanalysis from 1979 onwards, the release of the first subset of ERA5-Land data covering the period 2001 to 2018 is planned for this spring. The dataset will be updated in a timely manner together with ERA5 updates. Like ERA5, ERA5-Land is being produced by ECMWF as part of implementing the EU-funded Copernicus Climate Change Service (C3S). This is the first time that a global land surface dataset describing the water and energy cycles and spanning nearly two decades will be available at a grid spacing of 9 km and hourly temporal frequency. The main features of this new dataset compared to previous reanalyses and ERA5 are shown in the table.

Meeting user requirements

ERA5-Land provides much finer-scale information than the land surface component of ERA5, which has a 31 km grid spacing. The figure shows the soil temperature on 15 March 2010 according to three ECMWF reanalyses: ERA-Interim, ERA5 and ERA5-Land. The enhanced resolution in ERA5-Land makes it possible to reveal much greater detail in the soil temperature structure and to resolve lower temperatures in the peaks of the Alpine region.

In line with requirements for C3S datasets, ERA5-Land will provide uncertainty estimates for each variable. To begin with, this will be done through uncertainty estimates of ERA5-equivalent variables at a spatial resolution of 62 km, and at a later stage (subject to positive scientific feedback) through a dedicated 10-member ensemble at 31 km with 3-hourly output. ERA5-Land is a single simulation forced by ERA5 low atmospheric meteorological fields, with additional lapse-rate correction. There is no coupling to the atmospheric model, and observations only influence the simulation indirectly through the forcing. This makes running ERA5-Land computationally affordable. Timely reprocessing of some or all of the dataset is particularly important to support seasonal forecasting, which requires



Soil temperatures in ERA-Interim, ERA5 and ERA5-Land. The charts show soil temperature of the top 7 cm of soil at 12 UTC on 15 March 2010 according to ERA-Interim (79 km grid spacing, left), ERA5 (31 km grid spacing, middle), and ERA5-Land (9 km grid spacing, right). The temperature values over the Mediterranean Sea in ERA-Interim and ERA5 are sea-surface temperature values.

	ERA-Interim	Era-Int/Land	ERA5	ERA5-Land
Period	Jan 1979–present	Jan 1979–Dec 2010	Jan 1950–present	Jan 1950–present
Spatial resolution	~79 km grid spacing 60 levels	~79 km grid spacing	~31 km grid spacing 137 levels	~9 km grid spacing
Model version	IFS (+TESSEL)	HTESSEL IFS Cycle 36r4	IFS (+HTESSEL)	HTESSEL IFS Cycle 45r1
Assimilation system	IFS Cycle 31r2	n/a	IFS Cycle 41r2	n/a
Uncertainty estimate	None	None	Based on 10-member 4D-Var ensemble at 63 km	To be based on 10-member atmospheric forcing at 31 km
Output frequency	6-hourly analysis fields	6-hourly analysis fields	Hourly (three-hourly for the ensemble)	Hourly (three-hourly for the ensemble)

consistent land initial conditions. Such reprocessing is envisaged in case of major updates of the land model or the forcing data. In addition, in ERA5-Land the spin-up of surface variables with long memory, such as root-zone soil moisture, is significantly reduced compared to the land surface component of ERA5. This has been achieved by running a single stream for the period 2000 onwards and by using as initial conditions the last year of the ERA5 stream preceding the year 2000.

Preliminary scientific assessments of the first few years of production have been carried out for 2-metre temperature, soil moisture, river runoff and lake fields. They show that ERA5-Land is of very good quality, adding value to ERA5 surface fields and providing users with a more accurate dataset for surface applications. The impact can be particularly important over complex terrain, where accurate orography is very important. ERA5-Land meets the growing requirement from land user communities to gain access to long-term higher-resolution datasets. In the context of the Copernicus programme, this requirement is of special relevance as water resources management, agricultural activities and drought prediction, among others, demand long-term datasets at a finer resolution than what climate reanalysis can currently provide.

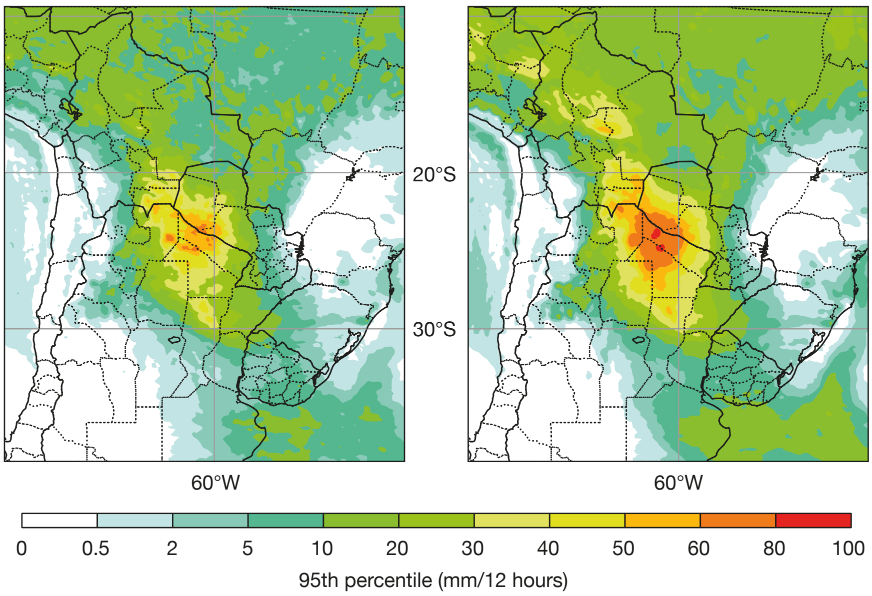
Outlook

Production of the dataset for the period 1979 to 2000 is currently under way, and it is planned to be released in the autumn of 2019, subject to satisfactory quality control checks. ERA5-Land will eventually be extended back to 1950 to match the period covered by ERA5. Production for this last subset is planned to start before the end of the year.

New point rainfall products in ecCharts

Tim Hewson, Fatima Pillosu, Axel Bonet, Iain Russell, Cihan Sahin, Florian Pappenberger

New experimental 'point rainfall' products with global coverage have been available in ecCharts, a web-based application providing easy access to ECMWF's medium- and extended-range forecasts, since 17 April 2019. They are the result of an ECMWF product idea which reached the final of the Harry Otten Prize for Innovation in Meteorology at EMS 2015. The idea was followed by considerable development work and the creation of a new post-processing suite called ecPoint. The post-processed products add value to existing products by accounting for the different degrees of sub-grid variability and grid-scale bias that one sees in different weather situations.



Example comparing raw ensemble and point rainfall charts. The charts show the 95th percentile for the raw ECMWF ensemble forecast (left) and for the corresponding ecPoint-Rainfall forecast (right) for 06–18 UTC on 2 April 2019 (T+90 to T+102). Although it is occasionally not the case, in this particular example the areas where there is a 5% chance of large accumulations, for example 60 mm or more in 12 hours in the Argentina–Paraguay border region, are more extensive in the point rainfall product than in the raw ensemble.

The new ecCharts products show probabilistic forecasts of 12-hour rainfall for any location (i.e. representing totals as measured by rain gauges). These complement the grid box mean values that the raw ensemble provides. Output for overlapping lead time ranges is available up to day 10 (T+0 hours to T+12 hours, T+6 hours to T+18 hours, ..., T+234 hours to T+246 hours). In the first release, users can display point rainfall in one of two ways. They can either select a percentile (1, 2… 99) and display the corresponding field of rainfall threshold values (mm), or select a rainfall threshold and show a corresponding field of probability (%, of exceedance). New colour schemes are available for both approaches, and these are duplicated for the raw ensemble, allowing users to easily compare the two. Potential applications include issuing warnings – e.g. for flash floods – and more accurate forecasts of dry weather at a given location. We would welcome feedback on the usefulness of the products. User guidelines are now available in ECMWF's online Forecast User Guide (https://confluence.ecmwf.int/display/FUG/Forecast+User+Guide).

Model upgrade improves ocean wave forecasts

Jean-Raymond Bidlot

Changes in the ocean wave model used in ECMWF's Integrated Forecasting System (IFS) are set to improve forecasts of some of the most common ocean wave variables, including significant wave height. The changes are part of IFS Cycle 46r1, which is scheduled to be implemented in June 2019. They include new parametrizations for wind input and deep-water dissipation of waves as previously implemented by Météo-France, based on work by Fabrice Ardhuin (Ifremer, France) and collaborators.

Wave modelling is an important part of numerical weather prediction systems: wind-generated surface waves modulate momentum, heat, and mass fluxes between the atmosphere and the oceans. Ocean waves in turn influence the upper ocean circulation and mixing. As part of ECMWF's Earth system approach, the wave model component of the IFS is actively coupled to both the atmosphere and the ocean modelling subsystems.

The evolution of the wave field over the deep ocean is modelled by the wave energy balance equation. This equation makes use of different parametrizations to represent how waves are generated by wind and how they are dissipated. It also includes a nonlinear interaction term that allows for wave energy redistributions among the different wave components. Once locally generated, waves can propagate across whole ocean basins as swell. Their interaction with surface currents is also represented in the model.

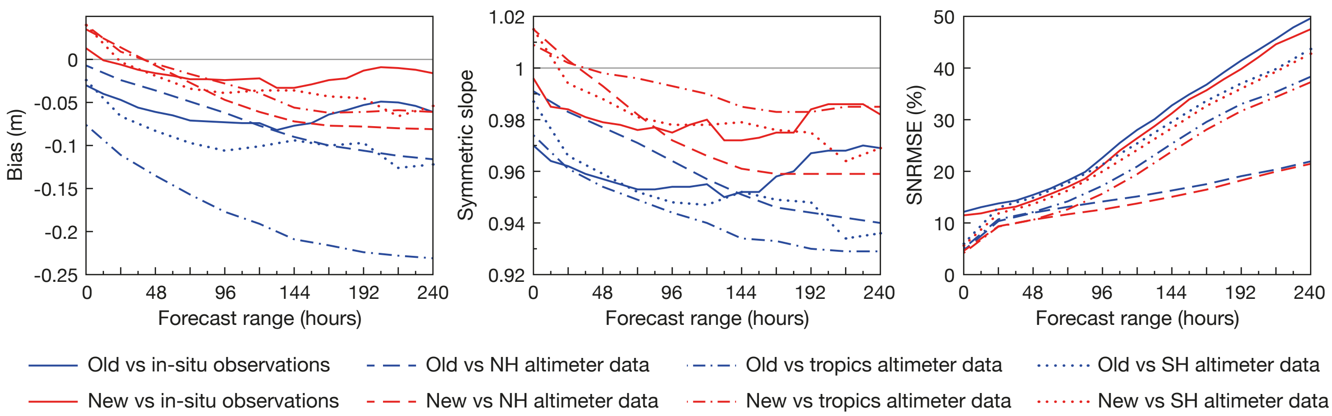
Better forecasts

The Météo-France changes to the wave model code have been adapted and optimised to run efficiently as part of the coupled Earth system models used in the IFS. It was important to ensure that the new parametrizations return a similar level of feedback from the modelled sea state to the atmosphere model. The impact of modelled ocean waves on the modelled ocean circulation when using the new parametrizations was also assessed. Finally, as part of the upgrade, when the ocean and ocean waves are coupled in the forecast model, surface currents have an impact on the wave propagation.

The main impact of these changes is increased accuracy and realism of the most frequently used ocean wave parameters, such as significant wave height (roughly the average height of the highest one third of waves). The new formulation reduces the overprediction of long period swell energy and the small wave height under-estimation in the storm tracks. Forecasts are generally improved up to 10 days ahead. One effect of the new parametrizations is that the activity level of significant wave height forecasts is increased. The figure shows that the new formulation reduces negative wave height forecast biases and increases the activity level, as measured by the standard deviation of the model simulations, in better agreement with the observation standard deviation (symmetric slope closer to 1). There is also a small reduction in the symmetrically normalised root-mean-square error throughout the 10-day forecast range.

Freak wave modelling update

Output wave parameters available from the IFS comprise a set of parameters to describe the mean sea state as well as the different major wind sea and swell wave components. There is also a set of variables to describe the single largest wave in a record, which, when substantially larger than the mean state, is commonly referred to as a freak wave. Based on work by Peter Janssen (ECMWF Technical Memorandum No. 813) subsequently developed by Peter Janssen and Augustus Janssen, the freak wave parameter calculation has been updated. The main impact is a more realistic increased probability of larger waves in shallow water compared to the old version.



Improved scores. The charts show reduced bias (left), increased activity as shown by a symmetric slope closer to 1 (middle) and reduced symmetrically normalised root-mean-square error (SNRMSE – right) in forecasts of significant wave height using the new parametrizations. Bias is determined by comparison against in-situ wave height observations and, for the northern hemisphere (NH), southern hemisphere (SH) and the tropics, wave height estimates from space-borne altimeters. In-situ observations are mostly located in near-coastal areas in the northern western hemisphere. The altimeter data were assimilated by the wave model while the in-situ data are completely independent. The results are based on experiments at TCo399 resolution (25 km grid spacing) for the period June to August 2017 and December 2017 to February 2018.

NWP courses combine face-to-face with eLearning

Sarah Keeley

Over 150 participants have attended this year's numerical weather prediction (NWP) training course modules at ECMWF. We know that participants really value face-to-face learning: it is required when studying advanced materials in the short amount of time that busy schedules permit, and it gives participants the chance to talk to lecturers about course material and their own research issues and to build networks within the research community. Responding to feedback, all modules are now delivered in short one-week face-to-face courses with practical sessions to consolidate learning and supported by pre-course study, mainly in the form of newly created eLearning modules. A 'Data assimilation collaborative course' has been developed and delivered by the University of Reading for those interested in more introductory theory or practical sessions. The course curriculum is focused on the Centre's objective to provide advanced training for the scientific staff of ECMWF's Member and Co-operating States in the field of numerical weather prediction. It draws on the expertise of over 70 research staff, mainly from the Research Department but also from the Forecast and Copernicus Departments, and one or two external experts.

Group photo. This year's course on predictability and ensemble forecasting attracted a large number of participants.

Focus on ensembles and the extended range

Over the last few years, we have received a large number of applications to attend the NWP course on predictability and ensemble forecasting, and this year was no exception. As we enter the era of seamless probabilistic forecasting, with national meteorological services providing probabilistic forecasts from days to months ahead, we have seen forecasters as well as researchers wanting to increase their understanding of the systems that are used to produce the forecasts and the underpinning theory of how we can make predictions beyond the medium range. Attendees this year spent time using the OpenIFS system to look at the use of ensembles within a case study of September 2012, which was challenging for forecasters when Hurricane Nadine interacted with an Atlantic cut-off low. (See the article 'The varied uses of OpenIFS' in this Newsletter)

Advanced numerical methods

In 2014, we introduced a new course in advanced numerical methods, which has grown in popularity. This year it attracted 25 participants to learn about the latest advances to describe the dynamics of the atmosphere considering future computational and resolution requirements. Participants had time for hands-on exploration of elliptic solvers, advection schemes and spectral transform methods, as well as looking at idealised cases with OpenIFS.

Course material is downloadable and eLearning modules can be accessed from the training resource area on the ECMWF website (www.ecmwf.int/en/learning/education-material). Registration for next year's courses will open in the summer (see www.ecmwf.int/en/learning/training for updates).

Feedback on this year's courses

"The lecturers were all very enthusiastic and had great content to their presentations."

"There's no substitute for meeting the lecturers and other attendees in person. Discussing the lectures and ideas during coffee breaks was an important part of the course."

"I find it easier to digest scientific content through e-learning modules rather than traditional lectures, because I can adjust the pace according to my own needs. But to meet lecturers and other course participants in person is ever so important. It is good for networking and makes it possible to have informal on- and off-topic discussions that would never have taken place otherwise."

Numerical weather prediction courses for 2020

- Advanced numerical methods
- Parametrization of the subgrid scale and diabatic processes
- Data assimilation
- Satellite data assimilation
- Predictability and coupled ocean–atmosphere ensemble forecasting

Challenges in forecasting local heavy rainfall in mountainous regions

Artur Gevorgyan (Armenian State Hydrometeorological and Monitoring Service)

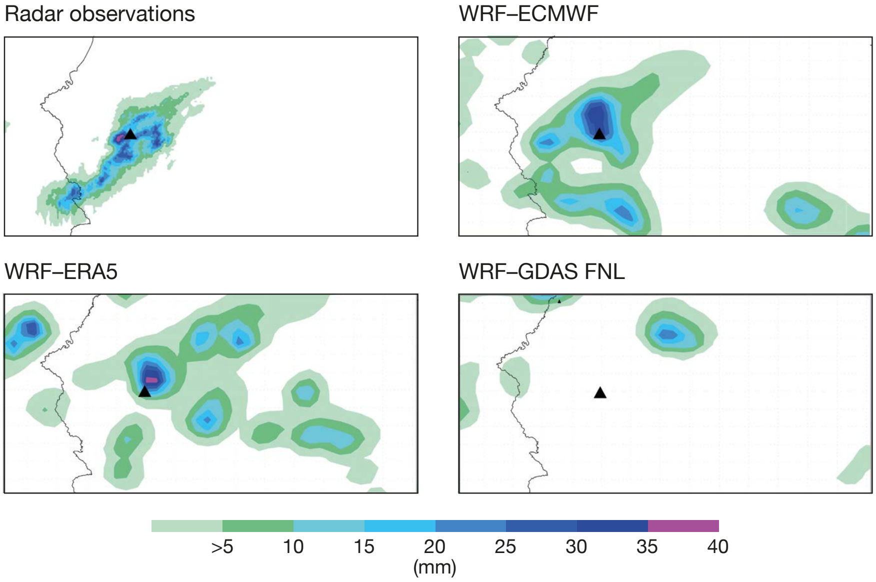
A new study has used initial and lateral boundary conditions produced by ECMWF's Integrated Forecasting System (IFS) to investigate the predictability of a highly localised heavy rain event in mountainous terrain in Armenia using a limited-area model. The study illustrates the importance of high-quality initial and boundary conditions and highlights challenges in accurately forecasting such events.

Details of the event

In the summer, heavy rainfall events in mountainous regions are often accompanied by damaging hailstorms and windstorms, flash floods or landslides, causing the loss of lives and significant economic damage. Accurately forecasting heavy rainfall events is thus important, but it poses a number of challenges. These include small spatial and temporal scales and the inherent nonlinearity of the dynamics and physics of such weather phenomena. This can be especially problematic over mountainous regions characterized by heterogeneous orography. In the afternoon of 2 July 2016, 52.8 mm of rain fell at Talin weather station in western Armenia. This is about 1.5 times higher than the climatological average for the entire month of July. The extreme rainfall lasted only 1 hour and 20 minutes and was accompanied by lightning and wind gusts reaching up to 22 m/s.

Forecasting the event

The US Weather Research and Forecasting (WRF) limited-area model was used in order to perform dynamical downscaling to a 3 km grid. The WRF model requires a minimum set of meteorological and land-surface fields for successful initialisation. The meteorological input data, known as forcing data or initial and lateral boundary conditions, include near-surface and upper-air temperature, humidity, wind, pressure etc. The following forcing data were used in the study:



Observed and predicted precipitation. Accumulated precipitation between 1200 and 1400 UTC on 2 July 2016 according to radar estimates (top left) and WRF model forecasts starting at 0000 UTC on the same day using ECMWF forecast forcing data (top right), ERA5 forcing data (bottom left) and GDAS FNL analysis forcing data (bottom right). Black triangles indicate the location of Talin station.

1. ECMWF's high-resolution operational forecasts starting at 0000 UTC on 2 July 2016 (IFS Cycle 41r2)
2. ECMWF's new ERA5 reanalysis
3. The US National Centers for Environmental Prediction (NCEP) Global Data Assimilation System Final operational analysis (GDAS FNL)

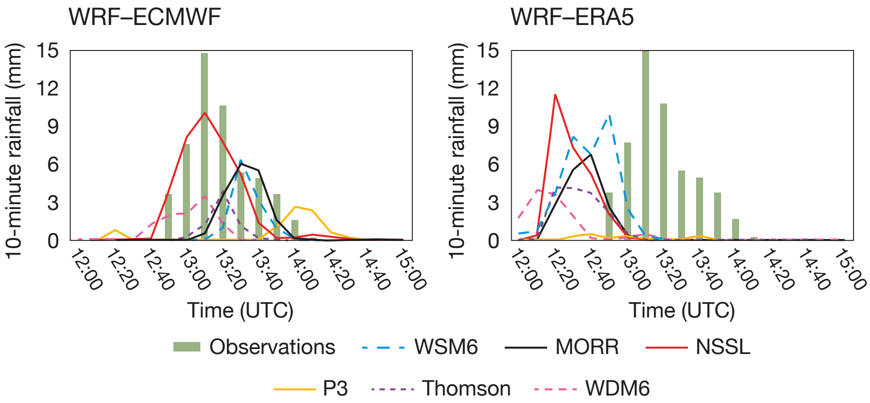
ECMWF's high-resolution global forecast has a grid spacing of about 9 km. The other two global datasets have a coarser grid spacing (31 km and 28 km, respectively) but benefit from vast amounts of observations from satellites, aircraft, weather stations, weather balloons etc. from around the time of the rain event.

The first figure shows the geographic distribution of rainfall according to very high resolution (about 500 m grid spacing) rainfall estimates derived from radar observations and WRF forecasts using different forcing data, all between 1200 and 1400 UTC on 2 July 2016. The northeast to southwest-oriented rainfall band stretched over the southwestern slopes of Mount Aragats. The area of maximum rainfall located around Talin station (the black triangle) saw up to 40 mm of rain. High-frequency radar observations (every five minutes) show that the earliest orographic convective cells were initiated over the Aragats massif at around 0705 UTC. The convective cells then propagated downslope in a southwesterly direction toward the Talin station and beyond. At 4,090 m above sea level, Mount Aragats is Armenia's highest isolated mountain massif. Convection initiation over isolated mountain massifs is known to be the result of thermal and dynamical forcing effects.

The rainfall patterns simulated by the WRF with ECMWF forecast and ERA5 forcing data compare well with the radar observations. By contrast, the WRF with GDAS FNL forcing data simulates much weaker local rainfall (20–25 mm), which is shifted 25–30 km to the northeast from Talin station.

The WRF model configuration

The appropriate configuration of the WRF model is another important issue. The second figure illustrates the sensitivity of the WRF model simulations to various cloud microphysical parametrizations. It can be seen that the NSSL scheme combined with ECMWF operational forecast forcing data performs best. This combination successfully reproduces the onset (at 1240–1250 UTC) and the peak (at 1310 UTC) of the observed convective rainfall in WRF forecasts starting at 0000 UTC. The sensitivity experiments using the ERA5 reanalysis forcing data show earlier peaks in simulated rainfall intensities. In general, the second figure indicates the importance of using 'double-moment' microphysical schemes (e.g. NSSL and MORR).



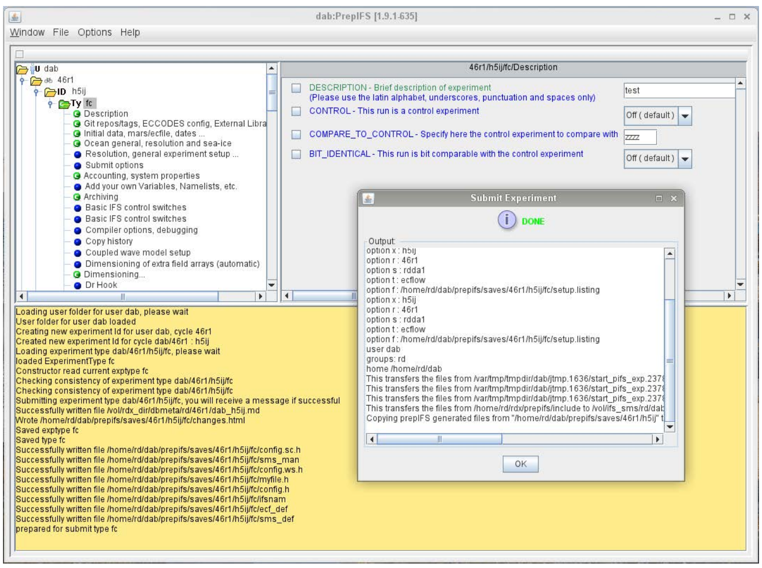
Sensitivity to the microphysical scheme used. Ten-minute observed and simulated rainfall amounts at Talin station derived from WRF sensitivity runs initialised at 0000 UTC on 2 July 2016, using ECMWF forecast forcing data (left) and ERA5 reanalysis forcing data (right). MORR = Morrison double–moment scheme; NSSL = The National Severe Storms Laboratory scheme; P3 = Predicted Particle Property scheme; Thompson = Thompson scheme; WDM6 = WRF double-moment 6–class scheme; WSM6 = WRF single-moment six-class microphysics scheme.

Overall, the results demonstrate that the accurate simulation of extreme rainfall over mountainous terrain is challenging: it requires the use of high-quality forcing data, sophisticated double-moment microphysical schemes and convection-permitting spatial resolution. Further information can be found in an article by Artur Gevorgyan in the *Journal of Geophysical Research: Atmospheres*, doi: 10.1029/2017JD028247.

PrepIFS code modernisation

Stephan Siemen, Paul Burton

For many years, PrepIFS has been the tool which ECMWF and Member and Co-operating State scientists use to configure their model experiments with ECMWF's Integrated Forecasting System (IFS). Recently, with the help of our consultant Paul Davis, who worked for four months with staff from the Research Department and the Development Section, a comprehensive review of the code was carried out and changes were implemented. An important goal was to make PrepIFS work with the latest long-term support version of Java. This work, together with an extensive code clean-up and refactoring, will greatly ease the migration to the systems to be installed in the new data centre in Bologna, and it will ensure that PrepIFS remains easy to maintain into the future.



PrepIFS user interface. The PrepIFS tool has been updated to make sure that it works with the latest long-term support version of Java and remains easy to maintain.

The code has also been moved under git version control, and automatic tests have been set up to make sure no regressions are introduced with future developments. This should ensure that the stability of the application is maintained in future releases. PrepIFS releases are now versioned through the module system. The work was finished by the end of March and users now benefit from better performance, several minor bug fixes and new features which were implemented.

Modelling centres collaborate on orographic drag

Annelize van Niekerk (UK Met Office), Irina Sandu (ECMWF)

A new modelling exercise, launched in autumn 2018 and jointly led by the UK Met Office and ECMWF, focuses on the representation of mountain effects on atmospheric flow. The exercise involves major numerical weather prediction and climate centres and is endorsed by the Global Atmospheric System Studies Panel of the Global Energy and Water Exchanges project and the Working Group for Numerical Experimentation (WGNE) of the World Meteorological Organization.

Motivation

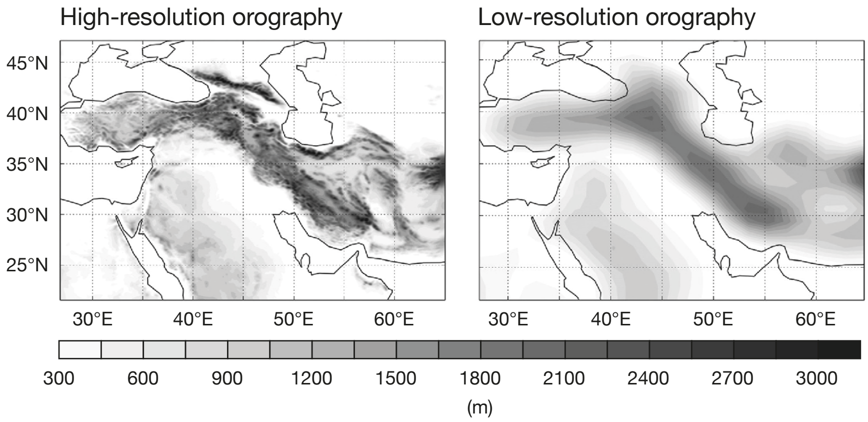
Mountains play a vital role in the predictability of the atmosphere from weather to climate timescales. While large-scale mountain ranges (in the order of 100 km or more) are well resolved in models used for climate projection, seasonal forecasting and numerical weather prediction, smaller-scale mountain ranges (in the order of 5 km to 100 km) are generally unresolved. These small-scale mountains can generate gravity waves that grow in amplitude as they propagate vertically, decelerating the flow in the stratosphere when they break. Small-scale mountains also deflect the flow near the surface, therefore acting as a drag on the atmosphere both locally and remotely. It is now widely acknowledged that these unresolved orographic drag processes significantly impact key aspects of the large-scale atmospheric circulation. As a result, they are approximated in models through parametrizations that have, thus far, relied mostly on idealised modelling, linear theory and empirical relations. Owing to the difficulties in directly measuring gravity wave momentum fluxes and the drag that arises from non-linear interactions with orography near the surface, there are very few constraints on the magnitude and spatial distribution of simulated orographic drag processes. Consequently, the accuracy of orographic drag parametrizations is highly uncertain. This was also highlighted by the recent WGNE Drag project, which showed that, while the total parametrized surface stress was roughly similar across models, the magnitude of the contributing components varied greatly across models at similar resolutions.

Motivated by these findings, the new modelling exercise, called COnstraining Orographic DRag Effects (COORDE), proposes a model comparison that seeks to quantify the impact of resolved orography and parametrized orographic drag on the simulated circulation over complex mountainous regions. The main regions of interest are the Middle Eastern mountainous region and the Himalayas.

Nine modelling centres are participating: ECMWF, Environment Canada, the German national meteorological service (Deutscher Wetterdienst), the Japan Meteorological Agency, the Korea Institute of Atmospheric Prediction Systems, Météo-France, the US National Oceanic and Atmospheric Administration, the US National Center for Atmospheric Research, and the UK Met Office. The first model output was received at the end of January 2019 and results are currently being analysed.

Modelling framework

COORDE uses a novel modelling framework that was proposed in a recent study (Van Niekerk et al., *J. Adv. Model Earth Sy.*, doi:10.1029/2018MS001417). In this study, kilometre-scale simulations over some of the most complex mountain chains on Earth (the Himalayas and Middle East mountains) were used to assess the ability of existing low-level blocking and gravity wave drag parametrizations to reproduce the explicitly resolved impacts on the flow. One question the study aimed to address is: to what extent do the circulation impacts of the parametrized orographic drag mimic those obtained when explicitly resolving the orography in high-resolution (kilometre-scale) simulations over such complex mountain ranges? To answer this question, short-range forecasts were produced with various configurations of the UK Met Office Unified Model and ECMWF's Integrated Forecasting System. The impact of resolved orography on the circulation was deduced by calculating the difference between high-resolution simulations with a high-resolution (4 km to 9 km) and a low-resolution (125 km to 150 km) orography.



High- and low-resolution orography. High-resolution orography (left) and low-resolution orography (right) in the Middle East region, used in the high-resolution (4 km) Met Office Unified Model experiments. (Adapted from Van Niekerk et al., *J. Adv. Model Earth Sy.*, doi:10.1029/2018MS001417, under the CC BY-NC-ND 4.0 licence)

This was then compared with the impact of parametrized orographic drag, deduced from global low-resolution (125 km to 150 km) simulations with and without parametrized orographic drag.

At resolutions ranging from tens to hundreds of kilometres, relevant for predictions from a few weeks ahead to climate timescales, errors in simulated winds were shown to be due not only to the orographic drag parametrizations but also to the way the resolved dynamics interacts with the parametrized drag. This highlights the importance of physics–dynamics interactions. COORDE now aims to:

- expose differences in orographic drag parametrization formulation between models
- gain a better understanding of the impacts of different orographic drag parametrizations on modelled circulation
- use high-resolution simulations to quantify drag from small-scale orography, typically unresolved in models used for integrations from monthly to climate timescales, to evaluate orographic drag parametrizations
- investigate the interaction between the parametrized orographic drag and the resolved dynamics across models.

Further details about COORDE can be found at: https://osf.io/37bsy.

ecCodes and Magics available under Windows

Stephan Siemen

Since the February 2019 synchronised software release, for the first time ECMWF has made its popular ecCodes and Magics software packages available under the Windows operating system. ecCodes enables users to encode and decode BUFR and GRIB messages, while Magics is a meteorological plotting library to visualise data in the BUFR, GRIB and NetCDF formats.

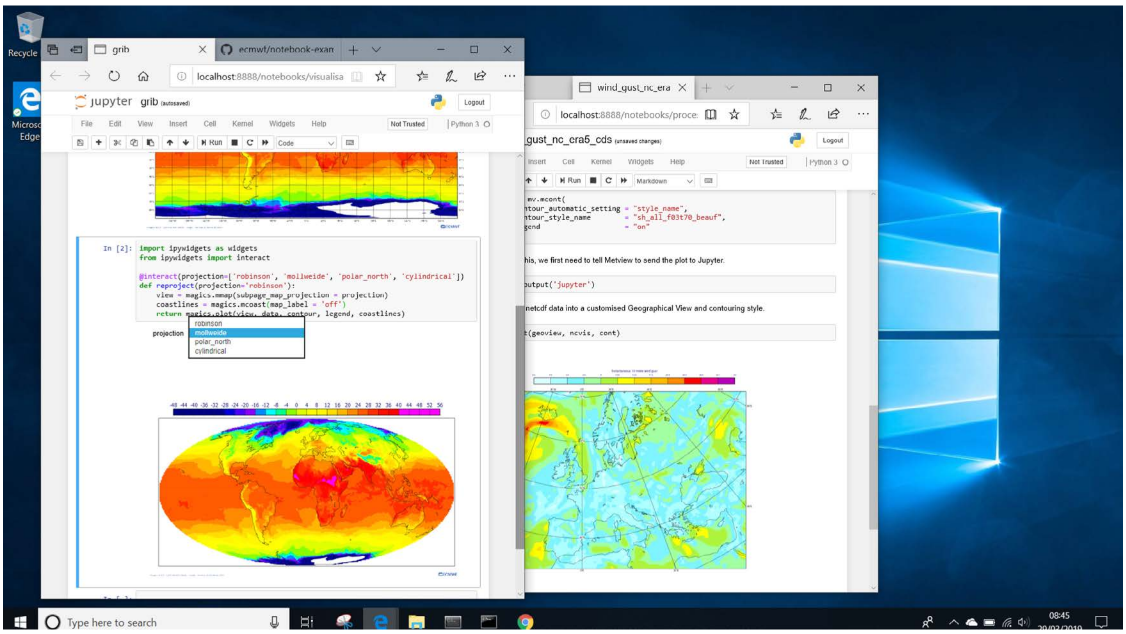
The software company Old Reliable Tech made the necessary changes in both tools and packaged them using the popular ‘pip’ and ‘conda’ package managers used by the Python community. Using conda, users can install both packages by typing

```
conda install -c conda-
forge eccodes Magics
```

This development enables users to decode and visualise BUFR and GRIB data on Windows using ecCodes and Magics. Users of these tools under Linux and MacOS can of course use the same package managers.

Windows is not a native platform for technical work at ECMWF. Therefore, ECMWF will need the help of the external user community to support these versions of ecCodes and Magics in future. ECMWF invites the Windows user community to use the new setup on GitHub (https://github.com/ecmwf) to fork/contribute and issue pull request changes to ecCodes and Magics. Old Reliable Tech has set up automatic continuous integration (CI) tests and builds for Windows on GitHub, which should ensure this can be done easily and safely.

In parallel, the Python APIs to download data from ECMWF’s MARS archive and Copernicus Climate Data Store (CDS) have also been made available with pip and conda thanks to the support of the software company B-Open. This allows Python users to easily install the necessary software to access, process and visualise data from MARS and the CDS.



Using ecCodes and Magics under Windows. This screenshot shows the Jupyter notebook under Windows visualising ECMWF GRIB data through Magics and ecCodes.

Improved user support via CAMS and C3S websites

Xiaobo Yang, Daniel Varela Santoalla

Since January 2019, the two Copernicus Earth observation services implemented by ECMWF on behalf of the European Union have provided improved access to user support facilities on their websites. Visitors to the Copernicus Atmosphere Monitoring Service (CAMS) and Copernicus Climate Change Service (C3S) websites now have immediate access to the Copernicus knowledge base and Service Desk without leaving the page they are visiting.

Previously users of the CAMS and C3S websites could only access the knowledge base and Service Desk through links on the 'Help & Support' page. Those links, which continue to be available, take users to external web pages and thus move them away from the page they are visiting. To facilitate access to user support and to improve the user experience, the CAMS and C3S websites now include a small question mark icon in the bottom right corner of all pages. When users click on it, they can search the Copernicus knowledge base and browse knowledge base articles without leaving the page. After running a search, they can also create a Service Desk ticket, for example to give feedback on a knowledge base article or to raise any other issue. Such a ticket will be logged in the Jira Service Desk system and processed according to workflows set up by the Copernicus User Support team, in the same way as requests received through the 'Help and Support' page. The new feature ensures that any user creating a support request has gone through the knowledge base search first and has been offered access to any relevant articles. This will help users to help themselves. The new feature also makes it a lot easier for users to raise tickets in Jira as the system does not require prior registration and the information users need to provide about themselves is minimal.

User journey. Users can browse the CAMS and C3S knowledge base and can subsequently raise a Service Desk ticket, all without leaving the page.

The Copernicus User Support team keeps looking for new channels to enhance interactions with users. Plans for the future include a forum service to further promote self-support and a chat service for more immediate interactions between the User Support team and users. The team is always interested to hear from our users on how we can improve further. If you have an idea, criticism, or suggestion, you can reach us at copernicus-support@ecmwf.int.

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).

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

	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

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.

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

a Accessing ERA5 data

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.

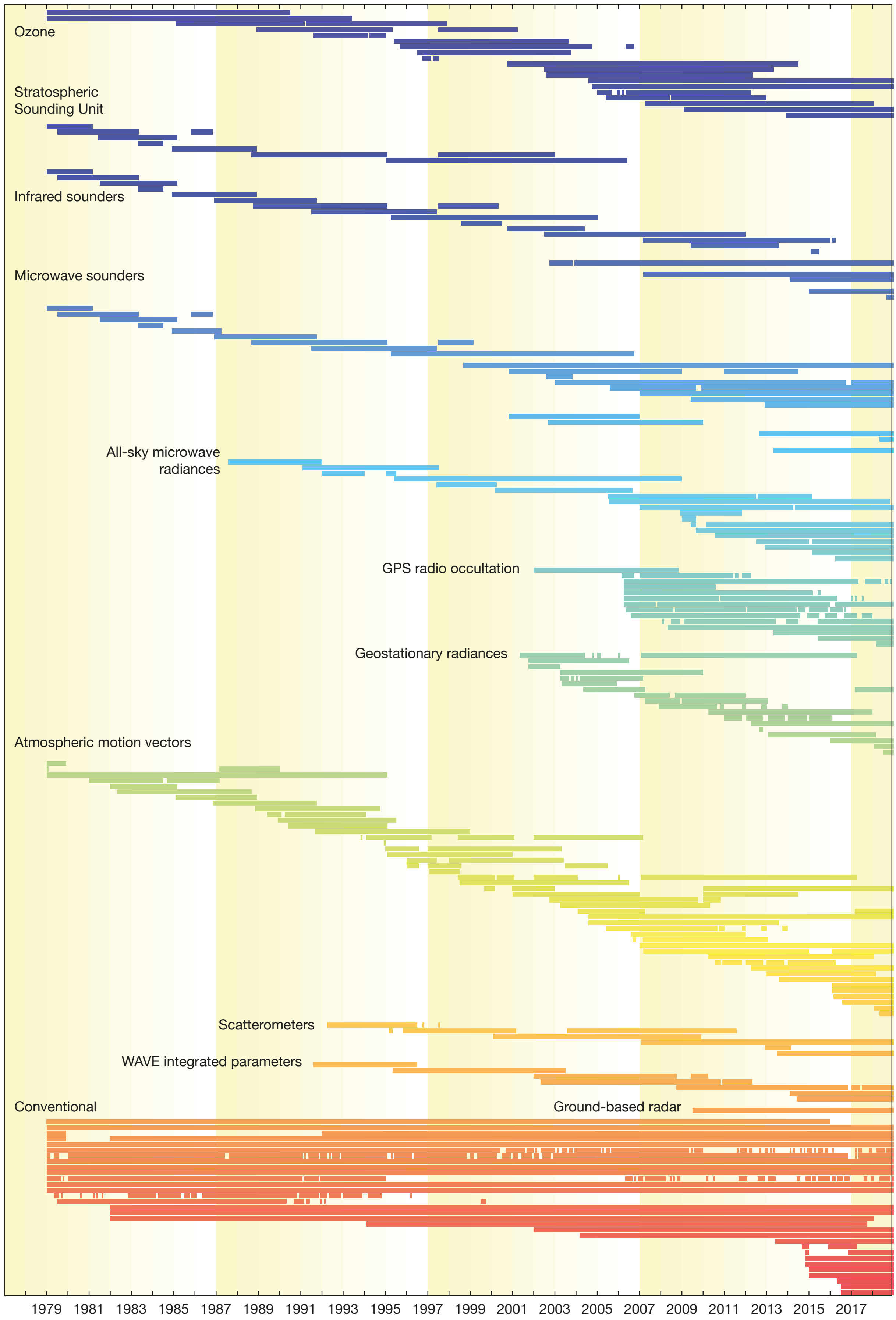


FIGURE 1 Data usage in ERA5 for the segment from 1979. Each horizontal bar represents the use of a particular satellite instrument or ground-based radar or a particular source of conventional data, such as weather stations, aircraft, ships, buoys and radiosondes. (*Image courtesy of Paul Poli*)

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.

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

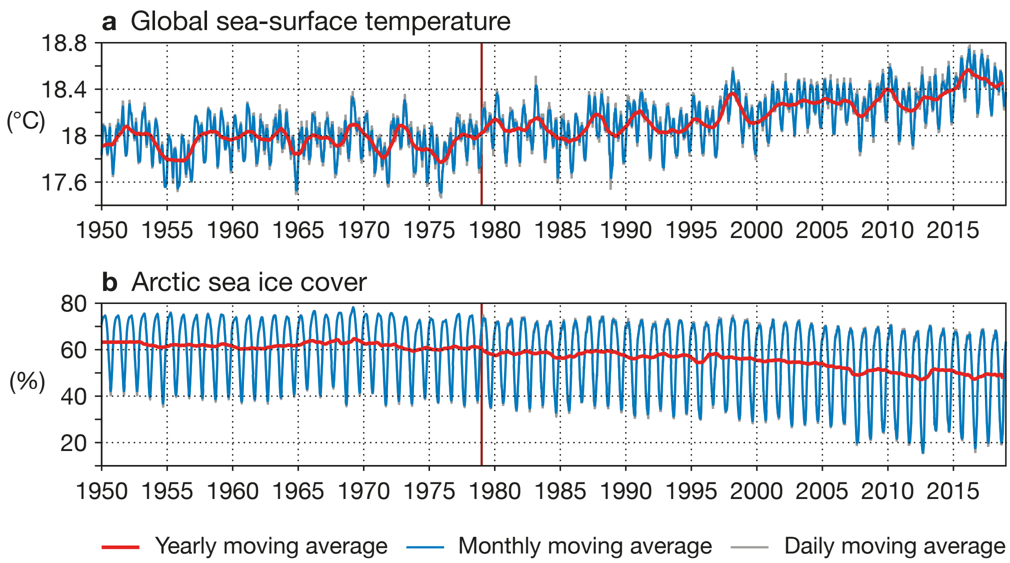


FIGURE 2 Time series of running mean averages of (a) global sea-surface temperature and (b) Arctic sea ice cover (60°N to 90°N) as it is used in ERA5 (right of the vertical line) or will be used in ERA5 (left of the vertical line).

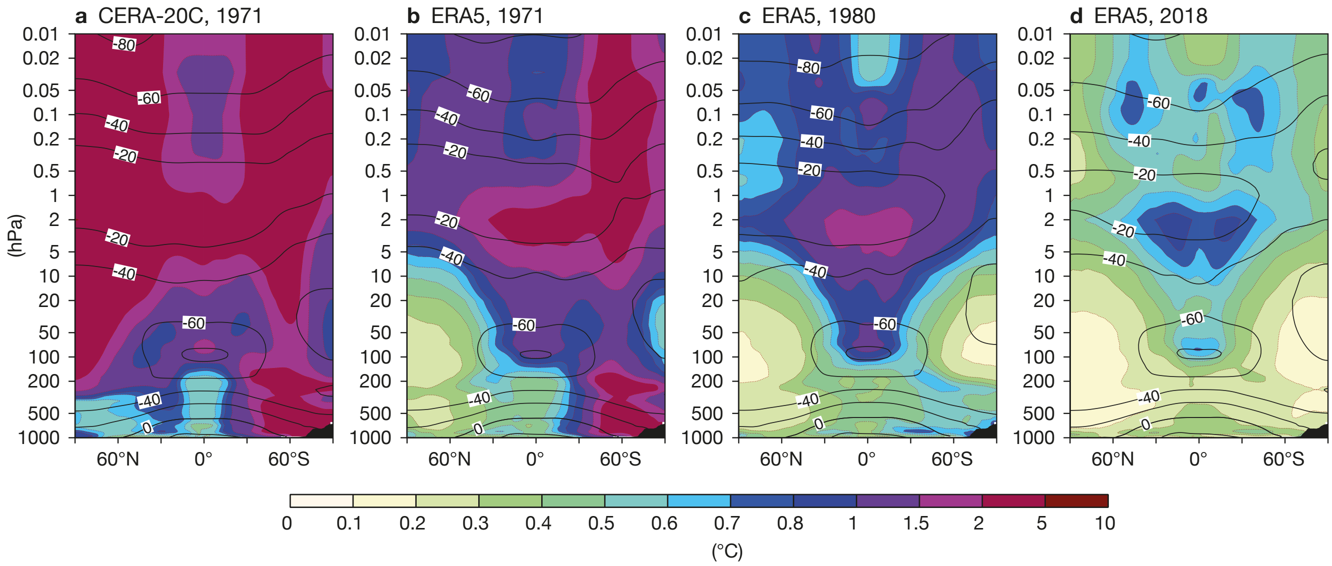


FIGURE 3 Zonal-mean cross section using a logarithmic pressure scale (hPa) of sub-daily ensemble spread of upper-air temperature averaged over March–April–May for (a) CERA-20C in 1971, (b) preliminary ERA5 data in 1971, (c) ERA5 in 1980, and (d) ERA5 in 2018. The contours show isotherms (in °C).

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.

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).

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.

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

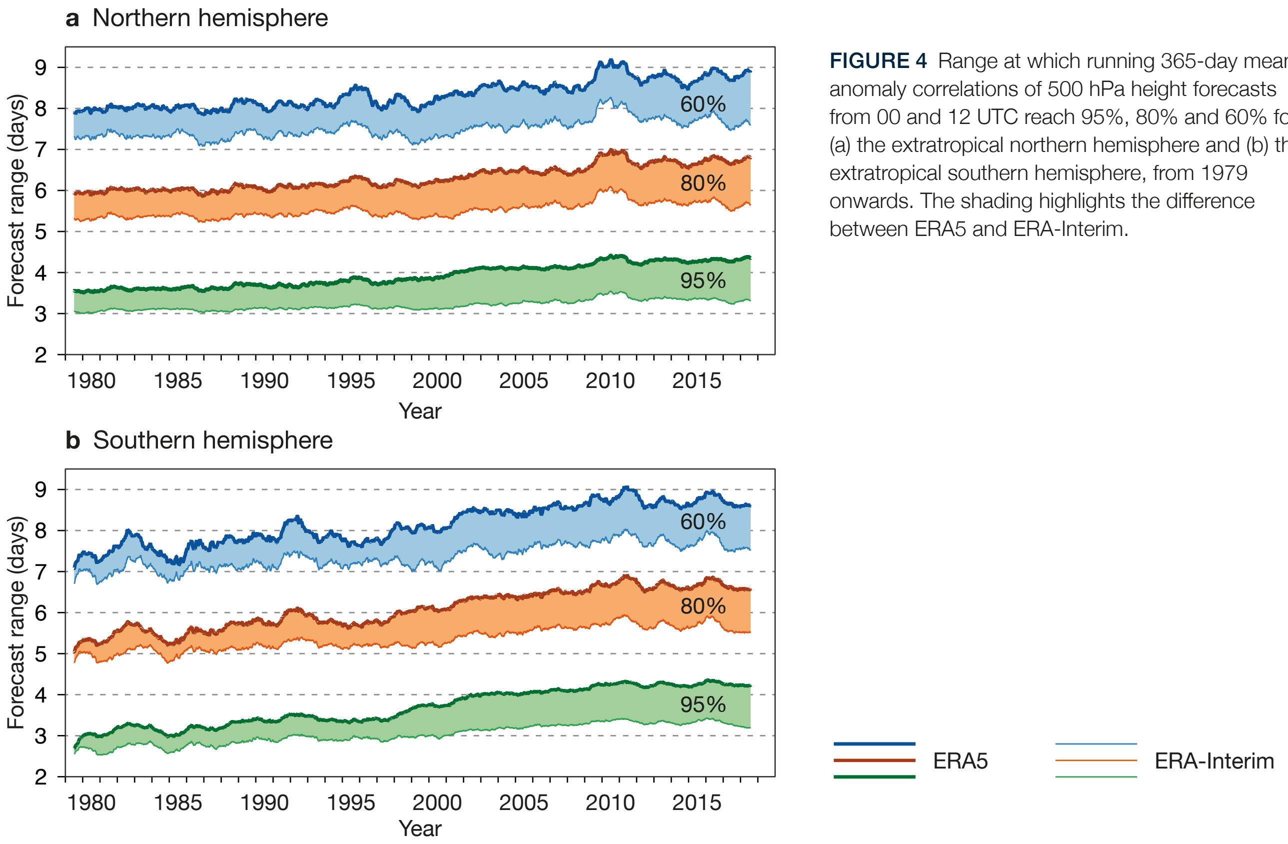


FIGURE 4 Range at which running 365-day mean anomaly correlations of 500 hPa height forecasts from 00 and 12 UTC reach 95%, 80% and 60% for (a) the extratropical northern hemisphere and (b) the extratropical southern hemisphere, from 1979 onwards. The shading highlights the difference between ERA5 and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

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

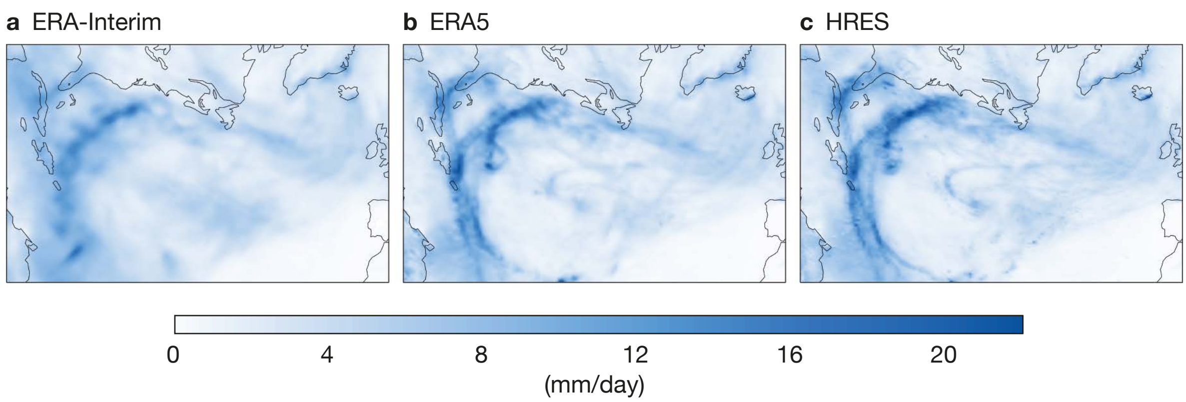


FIGURE 5 Mean precipitation rate for September 2017 over the North Atlantic from (a) ERA-Interim, (b) ERA5 and (c) ECMWF's operational high-resolution forecast (HRES).

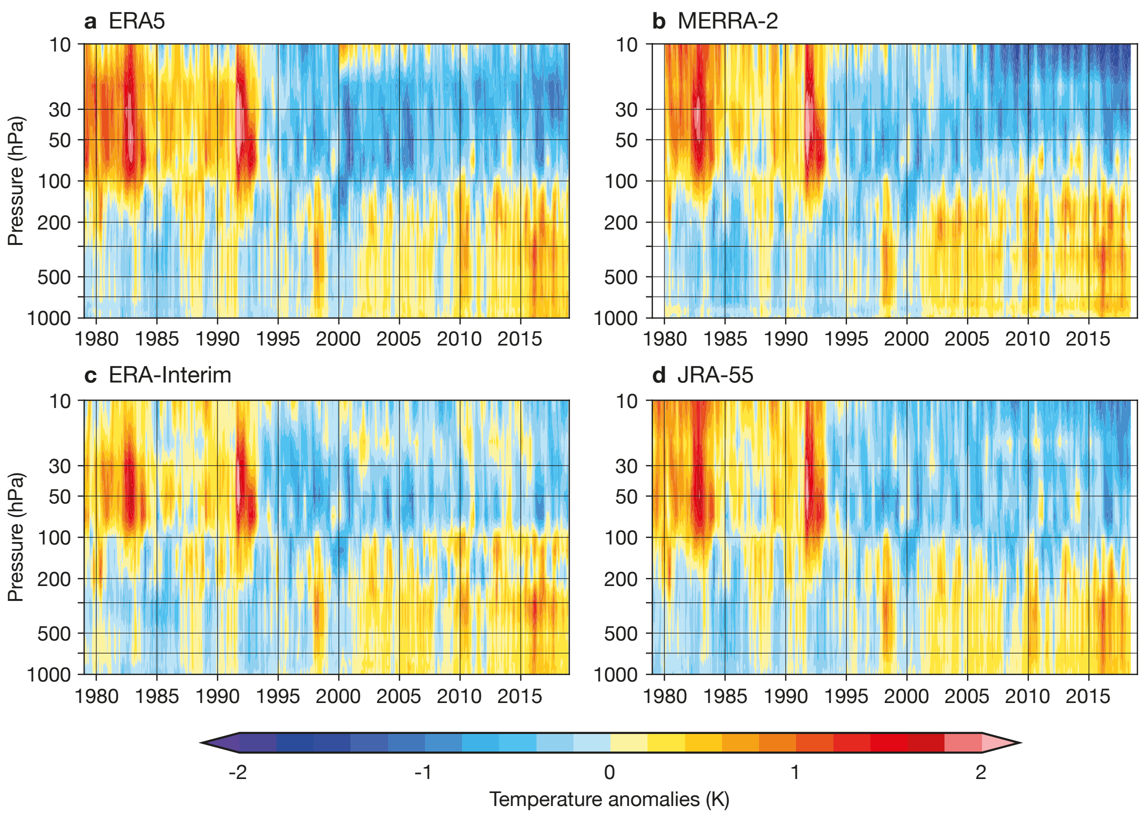


FIGURE 6 Vertical profiles of monthly and globally averaged temperature anomalies as a function of time from 1979 to 2018 from (a) ERA5, (b) the MERRA-2 reanalysis, (c) ERA-Interim, and (d) the JRA-55 reanalysis. Each monthly anomaly was computed by removing the 1981–2010 mean for the corresponding month.

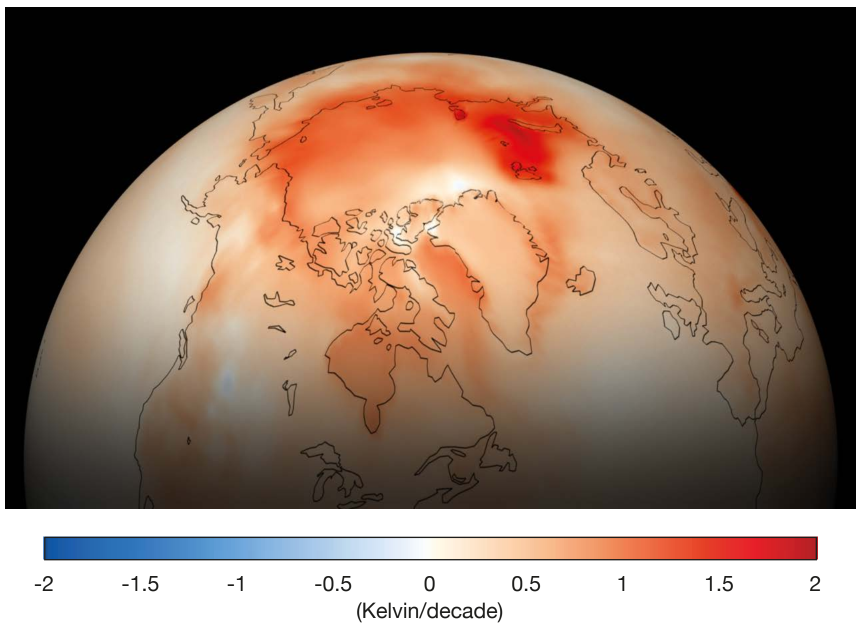


FIGURE 7 Point-wise linear trend (Kelvin/decade) in surface temperature between 1979 and 2018 from ERA5 over the Arctic region. The global-mean linear trend is 0.18 Kelvin/decade.

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.

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.

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.

The use of Sentinel-5P air quality data by CAMS

Antje Inness, Roberto Ribas, Richard Engelen

Air pollution is a global public health risk. The World Health Organization (WHO) estimates that 4.2 million people worldwide, including hundreds of thousands of people in Europe, die prematurely each year because of poor ambient air quality and that 91% of the world's population live in areas where air quality fails to meet WHO guidelines. It is therefore important to provide air quality forecasts on global, regional and local scales to support public health authorities and policy-makers and to enable vulnerable people to take preventative action during pollution episodes. The Copernicus Atmosphere Monitoring Service (CAMS), implemented by ECMWF on behalf of the European Union, provides services related to atmospheric pollution on the global and European scale, combining information from models and observations to enable a continuous assessment of the air we breathe. For its global forecasting system, CAMS uses a wide range of atmospheric composition observations from satellites to improve the initial conditions for its daily 5-day forecasts. The latest satellite to provide such data for CAMS is Sentinel-5 Precursor (S5P). S5P was launched by the European Space Agency (ESA) in October 2017 and is the first satellite mission in Europe's Copernicus Earth observation programme to be dedicated to monitoring atmospheric composition. Tests carried out at ECMWF show that the use of S5P data on ozone and carbon monoxide improves the CAMS analysis of atmospheric composition, which is produced using ECMWF's Integrated Forecasting System (IFS). S5P ozone data are now assimilated operationally by CAMS, and the assimilation of other S5P species is expected to follow later in 2019.

Trace gases observed by S5P/TROPOMI

S5P carries the Dutch-developed TROPOspheric Monitoring Instrument (TROPOMI), which provides high-resolution measurements in the ultraviolet (UV), visible (VIS), near infrared (NIR) and shortwave-infrared (SWIR) part of the spectrum. This wide spectral range enables several atmospheric trace gases to be retrieved, e.g. ozone (O_3), nitrogen dioxide (NO_2), sulphur dioxide (SO_2) and formaldehyde (HCHO) from the UV-VIS, and carbon monoxide (CO) and methane (CH_4) from the SWIR part of the spectrum. These species are all included in the CAMS system, making TROPOMI the perfect instrument to provide observations for CAMS at the unprecedented horizontal resolution of about 3.5 km x 7 km for the UV-VIS and 7 km x 7 km for the SWIR products. See Box A for details on the species that are relevant to monitoring and forecasting air quality.

Figure 1 shows examples of TROPOMI data products. It shows elevated CO and HCHO columns over Australia on 2 December 2018 in areas strongly affected by bushfires at the time. A map of NO_2 values over Europe on 27 June 2018 illustrates that TROPOMI's resolution is good enough to detect plumes of pollution coming from individual cities (e.g. London, Paris, Hamburg). The map also shows high NO_2 values in the Netherlands, the Ruhr region in Germany and the Po valley in northern Italy as well as high NO_2 values from the Saddleworth Moor fires in northeast England. Figure 1e shows low total column O_3 values in the southern hemisphere ozone hole area on 1 October 2018, and Figure 1f shows elevated values of volcanic SO_2 over the eastern Mediterranean from the eruption of the Mount Etna volcano in Sicily on 26 December 2018.

Monitoring and assimilating S5P O_3 data

During the first nine months after launch, the S5P satellite was in its so-called commissioning phase, which mainly served for functional testing, in-flight calibration and testing of processing chains. Nevertheless, ECMWF and ESA agreed to start monitoring these early data in the IFS. The model fields are interpolated in time and space to the location of the observations, and model equivalents of the observations are calculated (i.e. what TROPOMI would see if the state of the atmosphere corresponded to the interpolated model fields). This enables us to calculate temporal and spatial statistics of the differences between the observations and collocated model fields and gives us information on how the data compare to the daily CAMS analyses, which are constrained by various other satellite instruments. At the monitoring stage, the data are not actively used in the assimilation and therefore do not influence the analysis and subsequent forecast. Once the quality of the data has been established, active assimilation of the new data is tested in research

 doi: 10.21957/ko66ais8yp

a Some species relevant to air quality

For the purposes of air quality monitoring and forecasting, it is essential to have access to observational data on levels of ozone, nitrogen dioxide, carbon monoxide, formaldehyde and sulphur dioxide in the atmosphere.

Ozone: While stratospheric ozone is important and protects us from some of the sun's harmful UV radiation, tropospheric ozone at high concentrations can be harmful to humans, animals and vegetation. Tropospheric ozone is formed when nitrogen oxides (denoted NO_x, which means NO_2 and NO), CO, and volatile organic compounds react in the presence of sunlight. In urban areas in the northern hemisphere, high ozone levels in the lower troposphere usually occur during spring and summer. Over the Antarctic, stratospheric ozone destruction during austral spring leads to strong and rapid depletion of the ozone layer, the so-called 'ozone hole'.

Nitrogen dioxide: NO_2 is another key atmospheric pollutant, which is produced by road traffic, power plants and fossil fuel or biomass burning processes. Higher up in the troposphere, lightning and aviation also produce NO_x. NO_2 has a short lifetime so that concentrations are larger over land than in the cleaner air over the oceans. The largest concentrations are found over industrial and urban regions of the eastern US, California, Europe, China and Japan.

Carbon monoxide: CO has natural and anthropogenic sources. It is emitted from soils, plants and the ocean, but its main sources are incomplete fossil fuel and biomass burning. It is also produced by oxidation of CH_4 and other hydrocarbons. The highest CO concentrations are found over the industrial regions of Europe, Asia and North America. CO has a lifetime of several weeks and can serve as a tracer for regional and inter-continental transport of polluted air. The main loss process is the reaction with the hydroxyl radical.

Formaldehyde: HCHO is produced by the oxidation of methane and volatile organic compounds and also by incomplete combustions of hydrocarbons. It has sources from industrial processes, wildfires and biogenic emissions from vegetation. HCHO has a short lifetime of a few hours, making it a good indicator of hydrocarbon emission areas.

Sulphur dioxide: SO_2 also has natural and anthropogenic sources. It is emitted from volcanoes and produced by coal-fired power stations, industrial processes or other fossil fuel burning activities (e.g. cars or ships). The lifetime of SO_2 in the troposphere is a few days, in the stratosphere it can be several weeks. Volcanic eruptions that put ash and SO_2 high into the atmosphere can be a major risk to aviation.

experiments and the impact of their assimilation on the CAMS analysis is assessed with the help of independent validation data.

TROPOMI total column O_3 (TCO3) retrievals going back to 26 November 2017 were monitored with the CAMS system in this way. Time series of TROPOMI TCO3 observations and differences between the observations and the CAMS analysis (analysis departures) are shown in Figure 2a,b in the form of Hovmöller diagrams, which show TCO3 observations and departures by latitude and time of year, averaged over the respective latitude circle. The highest O_3 columns are found in the northern hemisphere spring and the lowest values over Antarctica during the ozone hole season. Even at this early stage, the TROPOMI TCO3 data generally agreed well with the CAMS analysis over large parts of the globe. They also agreed well with TCO3 retrievals from the Ozone Monitoring Instrument (OMI) and the Global Ozone Monitoring Experiment-2 (GOME-2), which are routinely assimilated by CAMS (not shown). However, the TCO3 near-real-time data from TROPOMI showed some retrieval anomalies at high latitudes and over snow/ice (e.g. Antarctica), where the differences with the CAMS analysis and the other datasets are larger (Figure 2c). These differences come mainly from the surface albedo climatology that is used in the near-real-time TROPOMI TCO3 retrieval. This climatology has a coarser horizontal resolution than the TROPOMI TCO3 data, which leads to problems in areas where there are large changes in reflectivity from pixel to pixel, e.g. pixels that are/are not covered by snow/ice. More details about this can be found in Inness et al. (2019a).

The impact of assimilating TROPOMI TCO3 observations was tested in research experiments, again going back to November 2017. Assimilating the observations was found to have a small positive impact on the ozone analysis compared to TCO3 data from ground-based spectrometers (not shown) and ozonesonde observations in the tropical troposphere (Figure 3a) as well as in the troposphere and parts of the stratosphere over Antarctica (Figure 3b) during September to October 2018. There was also an improved fit to In-service Aircraft for a Global Observing System (IAGOS) data at West African

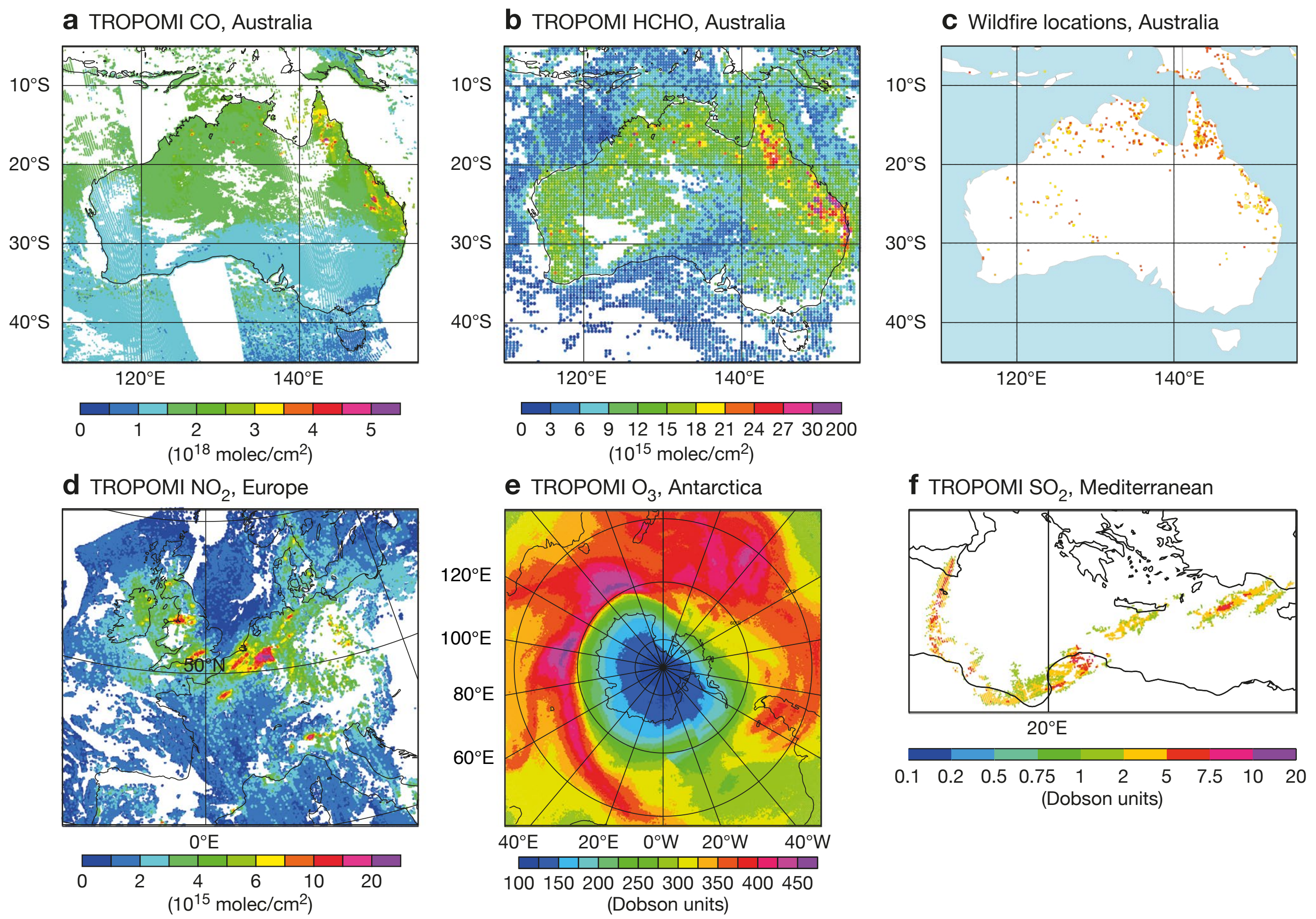


FIGURE 1 Examples of TROPOMI data for a range of chemical species. The charts show (a) total column CO and (b) total column HCHO from TROPOMI over Australia on 2 December 2018; (c) the location of fires from the Global Fire Assimilation System (GFAS) on the same day; (d) tropospheric column NO_2 from TROPOMI over Europe on 27 June 2018; (e) total column O_3, showing the Antarctic ozone hole as seen by TROPOMI on 1 October 2018; and (f) total column SO_2 as seen by TROPOMI on 26 December 2018, resulting from the eruption of Mount Etna in Sicily. (Images contain modified Copernicus Sentinel data)

airports (Figure 3c). In other areas, the impact was small. The overall impact of the TROPOMI data is relatively small because the CAMS analysis is already well constrained by several other ozone retrievals that are routinely assimilated (GOME-2, OMI, MLS, SBUV/2 and OMPS). Averaged over the periods February–April and September–October 2018, the differences between experiments with and without assimilation of TROPOMI data are within 2% for TCO3 and within 3% in the vertical for seasonal mean zonal mean O_3 mixing ratios, with the largest relative differences found in the troposphere. It should be noted that the current CAMS system cannot fully exploit the high resolution of the TROPOMI data at present because its horizontal grid spacing of 40 km is coarser than that of the TROPOMI data.

Because of the small positive impact on the CAMS analysis, it was deemed to be beneficial to include the data actively in the operational CAMS system. The assimilation of the near-real-time TROPOMI TCO3 data in the operational CAMS system was started in December 2018. Beyond the immediate impact, the addition of TROPOMI data makes the CAMS system significantly more resilient in case some of the older instruments whose retrievals are currently assimilated by CAMS stop working.

Monitoring and assimilating TROPOMI CO data

The use of TROPOMI total column CO (TCCO) data has been explored by CAMS in research experiments. CAMS routinely assimilates thermal infrared (TIR) TCCO retrievals from the Measurement of Pollution in the Troposphere (MOPITT) instrument and the Atmospheric Sounding Interferometer (IASI). TROPOMI data successfully capture the global TCCO distribution (Figure 4a) with relatively high TCCO values in the northern hemisphere, lower values in the southern hemisphere, and high TCCO values over the biomass burning areas in the tropics and the areas of high anthropogenic pollution over India and southeast Asia. However, there are some systematic differences between the TROPOMI data and the CAMS TCCO analysis (Figure 4b and 4c), particularly in the northern

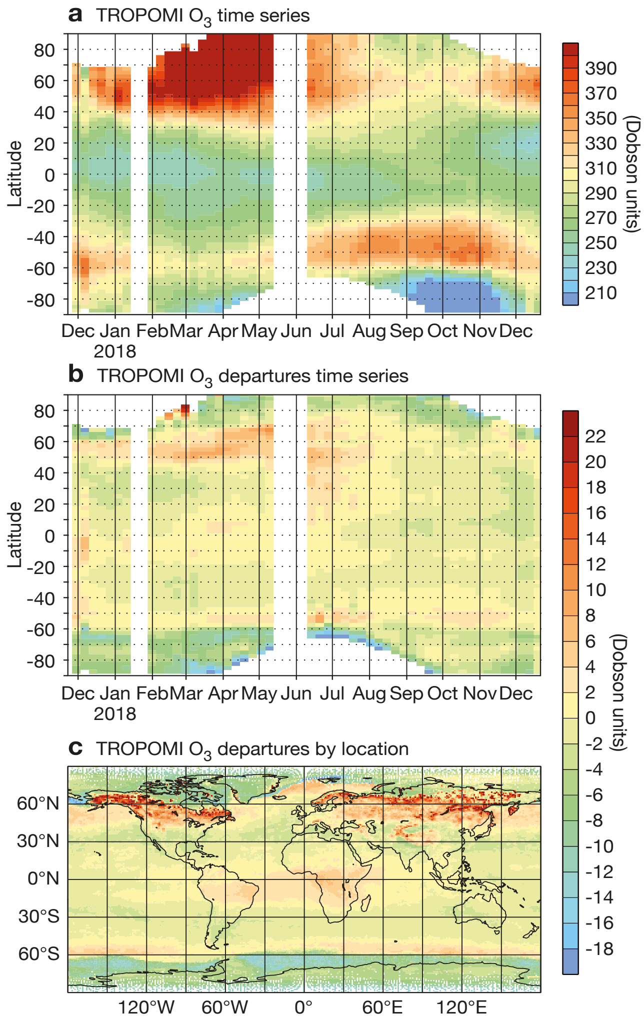


FIGURE 2 The charts show (a) a time series of zonal mean TROPOMI total column ozone from the end of November 2017 to mid-December 2018, (b) a time series of zonal mean TROPOMI analysis departures for the same period and (c) a map of TROPOMI analysis departures averaged over the period 26 November 2017 to 3 May 2018. (Images contain modified Copernicus Sentinel data)

hemisphere at low solar elevations north of 40°N, where TROPOMI has higher TCCO values than CAMS. Since the CAMS system is known to underestimate CO in the northern hemisphere extratropics, particularly during winter/spring and in the lower troposphere, some of these differences are likely to be the result of a CAMS model bias.

When TROPOMI data are assimilated in the CAMS system, they lead to increased CO values in the extratropics and lower values in the tropics, with TCCO changes of up to 30% at high northern latitudes (Figure 5a). Even though the TROPOMI data are total columns, their assimilation has a large impact on the vertical structure of the CAMS CO analysis because of different sensitivities of the TROPOMI SWIR and the MOPITT and IASI TIR retrievals to CO in the atmosphere. While the TIR retrievals are most sensitive to CO in the mid-troposphere, TROPOMI SWIR measurements are sensitive to the integrated amount of CO along the light path, including the contribution of the planetary boundary layer (the atmospheric layer that interacts with the surface). The largest absolute and relative changes due to the TROPOMI TCCO assimilation are found in the lower and upper troposphere (Figure 5b), i.e. the part of the atmosphere that is not well constrained by the assimilated MOPITT and IASI TIR retrievals.

The assimilation of TROPOMI TCCO leads to an improved fit to European Global Atmosphere Watch (GAW) surface CO observations over Europe (not shown) and IAGOS aircraft CO profiles in several areas (Figure 6). Particularly noteworthy is that it reduces the long standing low bias of the CAMS system over Europe. Relative to IAGOS, there are also

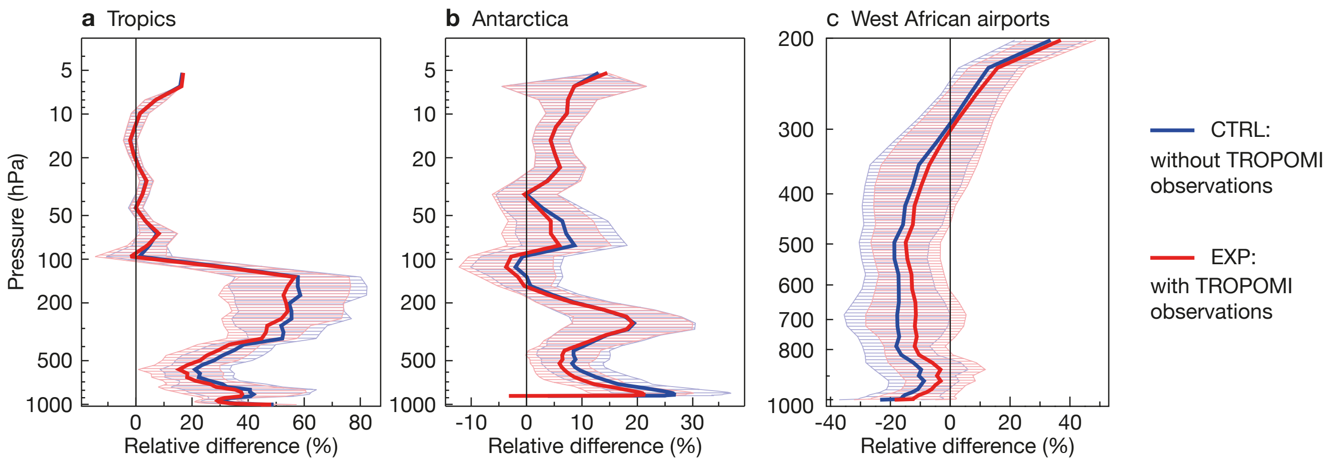


FIGURE 3 Profiles of mean relative differences between the CAMS analysis and ozonesonde observations from a TROPOMI TCO3 assimilation experiment (EXP) and a control experiment without TROPOMI assimilation (CTRL) for the period September to November 2018 for (a) the tropics and (b) Antarctica; and (c) profiles of mean relative differences between the CAMS analysis and IAGOS aircraft ozone observations at West African airports from a TROPOMI TCO3 assimilation experiment (EXP) and a control experiment without TROPOMI assimilation (CTRL). Horizontal hatching indicates one standard deviation in the distribution of relative differences.

reduced biases over Australasia, Southeast Asia and in parts of the free troposphere over India and West African airports. However, in some areas the fit to IAGOS observations is degraded, and the assimilation leads to increased negative biases over India and West Africa in the lower and upper troposphere. There is also an overestimation of surface CO relative to GAW observations at Cape Verde and a degraded fit to CO data from the Total Carbon Column Observing Network (TCCON) in the northern hemisphere, while the bias at the Lauder TCCON site in the southern hemisphere is reduced (not shown; see Inness et al., 2019b). More work is required, including work on the bias correction of the TROPOMI TCCO data, before the data can be assimilated in the operational CAMS system.

Use of S5P in the operational CAMS system

Near-real-time O_3 and NO_2 from TROPOMI were included in monitoring mode in the operational CAMS system in July 2018. This was followed by CO in November 2018 and HCHO and SO2 in December 2018. CH_4 retrievals are not available in near-real time but will be monitored in the CAMS greenhouse gas analysis, which runs a few days behind real time. CAMS started to assimilate TROPOMI TCO3 data in its daily operational air quality forecasts on 4 December 2018 and assimilation tests with near-real-time NO_2, CO and volcanic SO_2 from TROPOMI are under way. It is hoped that the assimilation of all these species will be activated in the operational CAMS system during 2019.

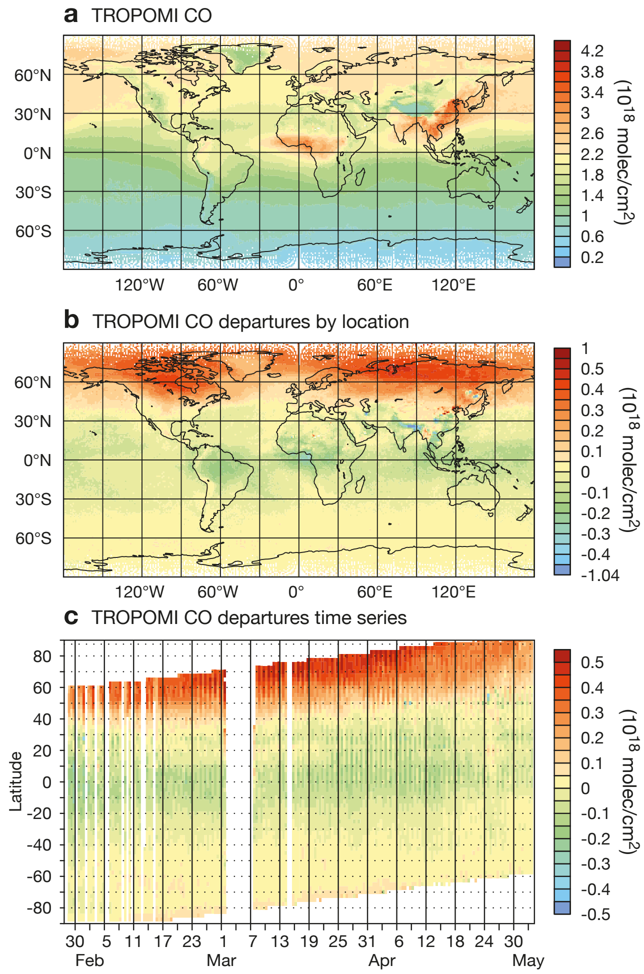


FIGURE 4 The charts show (a) mean TROPOMI total column carbon monoxide, (b) TROPOMI analysis departures averaged over the period 28 January to 3 May 2018, and (c) a Hovmöller diagram (zonal mean time series) of TROPOMI analysis departures for the same period. (Images contain modified Copernicus Sentinel data)

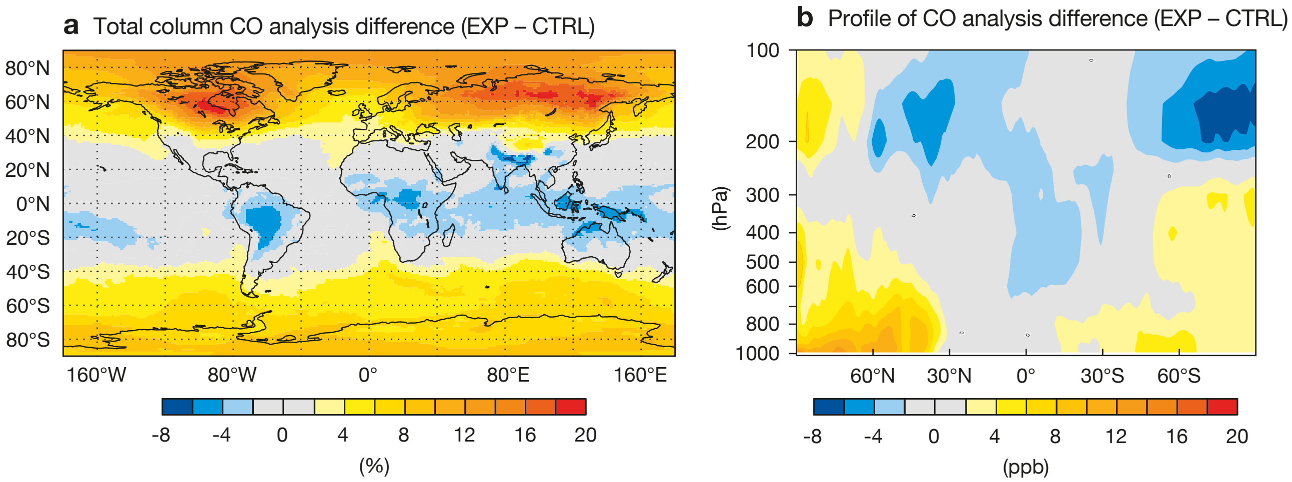


FIGURE 5 The charts show (a) the mean relative difference between the CAMS total column carbon monoxide analysis with TROPOMI assimilation (EXP) minus the analysis without TROPOMI assimilation (CTRL) and (b) a cross section of zonal mean differences between EXP and CTRL averaged over the period 28 January to 3 May 2018.

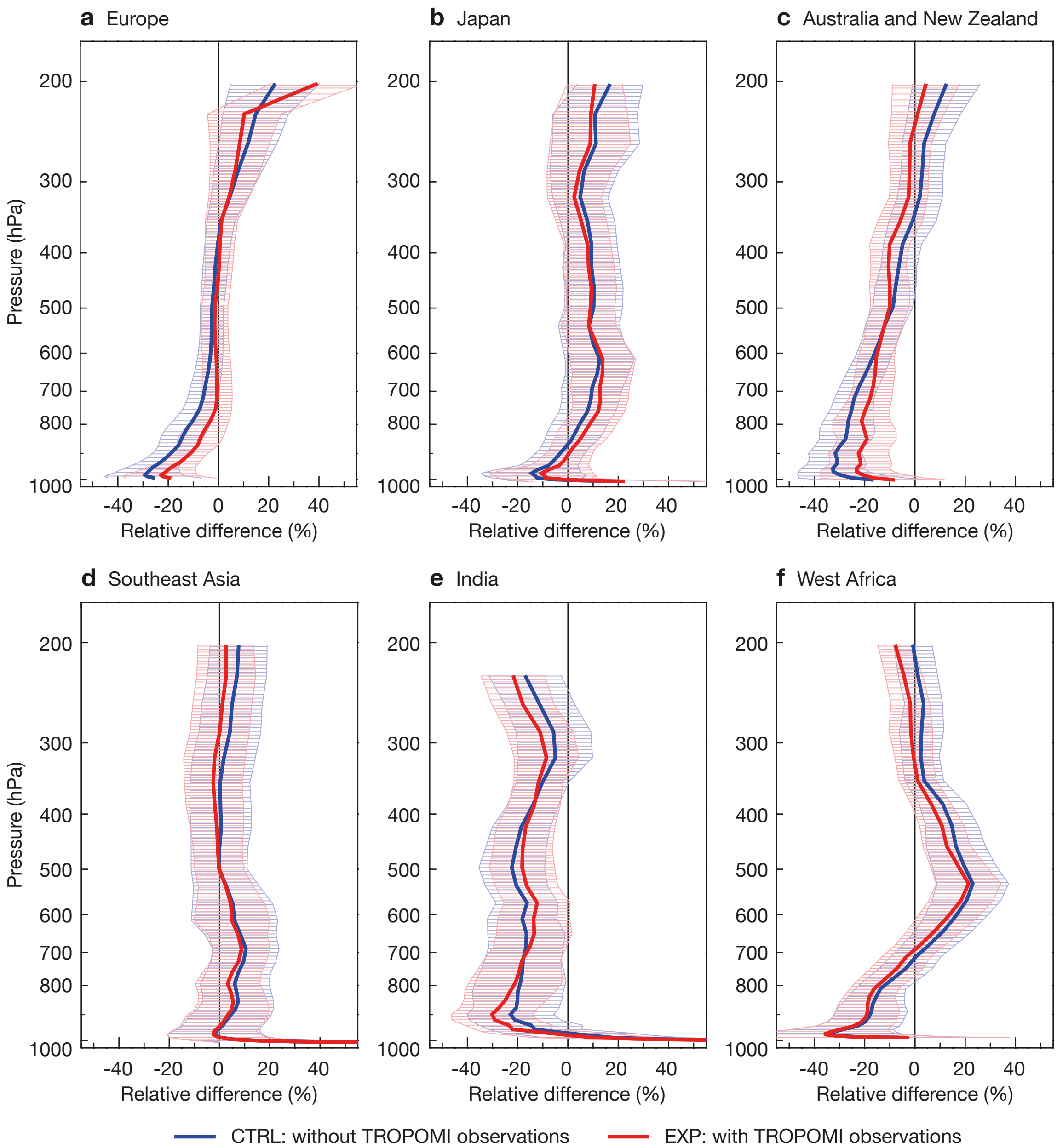


FIGURE 6 Mean relative difference between the CAMS analysis of carbon monoxide and IAGOS aircraft data for an experiment with TROPOMI TCCO assimilation (EXP) and a control experiment without TROPOMI TCCO assimilation (CTRL) for February to April 2018 in (a) Europe (3 sites, 99 profiles), (b) Japan (4 sites, 76 profiles), (c) Australia and New Zealand (4 sites, 54 profiles), (d) Southeast Asia (14 sites, 447 profiles), (e) India (3 sites, 26 profiles) and (f) West Africa (16 sites, 74 profiles). Horizontal hatching indicates one standard deviation in the distribution of relative differences.

Further reading

Inness, A., **J. Flemming**, **K.-P. Heue**, **C. Lerot**, **D. Loyola**, **R. Ribas**, **P. Valks**, **M. van Roozendael**, **J. Xu** & **W. Zimmer**, 2019a: Monitoring and assimilation tests with TROPOMI data in the CAMS system. near-real time total column ozone, *Atmos. Chem. Phys. Discuss.*, **19**, 3939–3962, doi:10.5194/acp-19-3939–2019.

Inness, A., **I. Aben**, **A. Agusti-Panareda**, **T. Borsdorff**, **J. Flemming**, **J. Landgraf** & **R. Ribas**, 2019b: Monitoring and assimilation tests of TROPOMI total column carbon monoxide data in the CAMS system. *ECMWF Technical Memorandum* **No. 838**.

The varied uses of OpenIFS

Glenn Carver (ECMWF), Kerstin Hartung (University of Munich), Pirkka Ollinaho (Finnish Meteorological Institute), Victoria Sinclair (University of Helsinki), Hamish Struthers (NSC at Linköping University, Sweden), Gunilla Svensson (Stockholm University), Gabriella Szepszo (ECMWF and Hungarian Meteorological Service)

The OpenIFS activity at ECMWF provides portable versions of ECMWF's Integrated Forecasting System (IFS) and single-column models for use by universities, the meteorological services of Member and Co-operating States, and research institutes. Some of these have used OpenIFS models since the activity was created in 2011 and now have well-established research and teaching activities using them. In this article, we highlight several such activities to illustrate the varied uses of the models provided. For more information on OpenIFS and how to obtain the models, please see the OpenIFS website at: *https://confluence.ecmwf.int/oifs/*.

Atmosphere–ocean single column model

Single-column models (SCMs) are tools that have been used for process studies and parametrization development for a long time. The benefit is that they nicely isolate vertical exchange processes, which are relevant for almost all parametrizations, from the advection tendencies. They also take very little time to run. There have been a number of model intercomparison projects for SCMs in which the focus has been on specific processes that rely heavily on parametrizations, such as turbulence in stably stratified conditions, the representation of the diurnal cycle, or the transition from stratocumulus to cumulus. Upper ocean turbulence has also been studied using SCMs, although to a lesser extent.

A fully coupled SCM was recently developed as an initiative within the Swedish e-Science Research Centre (*e-science.se*), partly funded through the APPLICATE Horizon 2020 project (*www.applicate.eu*). The Atmosphere–Ocean Single-Column Model (AOSCM) is based on OpenIFS for the atmosphere and NEMO/LIM for the ocean/sea ice. The two model components are coupled using the OASIS software (Figure 1). The first version of the AOSCM (AOSCM.v1_EC-Earth3) is described in detail in Hartung et al. (2018) and the code is available through the EC-Earth development portal. The motivation for developing the AOSCM is to be able to improve parametrized processes interacting closely with the surface and also to study the vertical coupling itself. Using the SCM framework enables us to test and generalise findings from idealised large eddy simulations. The AOSCM makes it possible to assess how processes, or the sensitivity to perturbations, depend on whether the surface conditions are prescribed or whether coupling with the surface is allowed.

One example application is to study in detail the coupling between the atmosphere and sea ice when several ice categories are used. The choice of variables to pass between the ocean and the atmosphere increases substantially in the presence of sea ice, and the resulting interaction also depends on the frequency of coupling. It is thus advantageous to be able to perform a multitude of sensitivity tests in a controlled large-scale environment with prescribed advection, at relatively small computational cost.

A case of warm air advection over sea ice in the Arctic, as observed by the Swedish icebreaker Oden in August 2014, is one of the first cases simulated by the

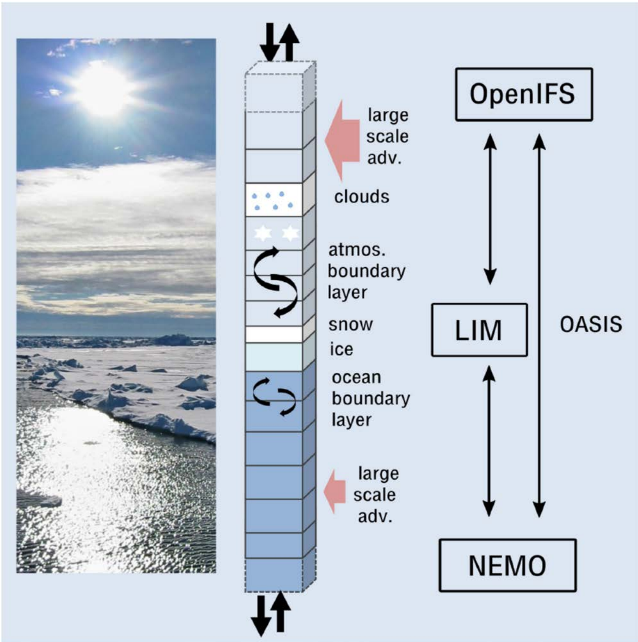


FIGURE 1 A fully coupled single-column model using OpenIFS has been developed at the Swedish e-Science Research Centre. It includes an atmospheric model (OpenIFS single-column model) and an ocean model (NEMO), which are coupled using the OASIS software. It also includes a sea ice model (LIM). (*Photo: Michael Tjernström*)

doi: 10.21957/b5wf2gclr6

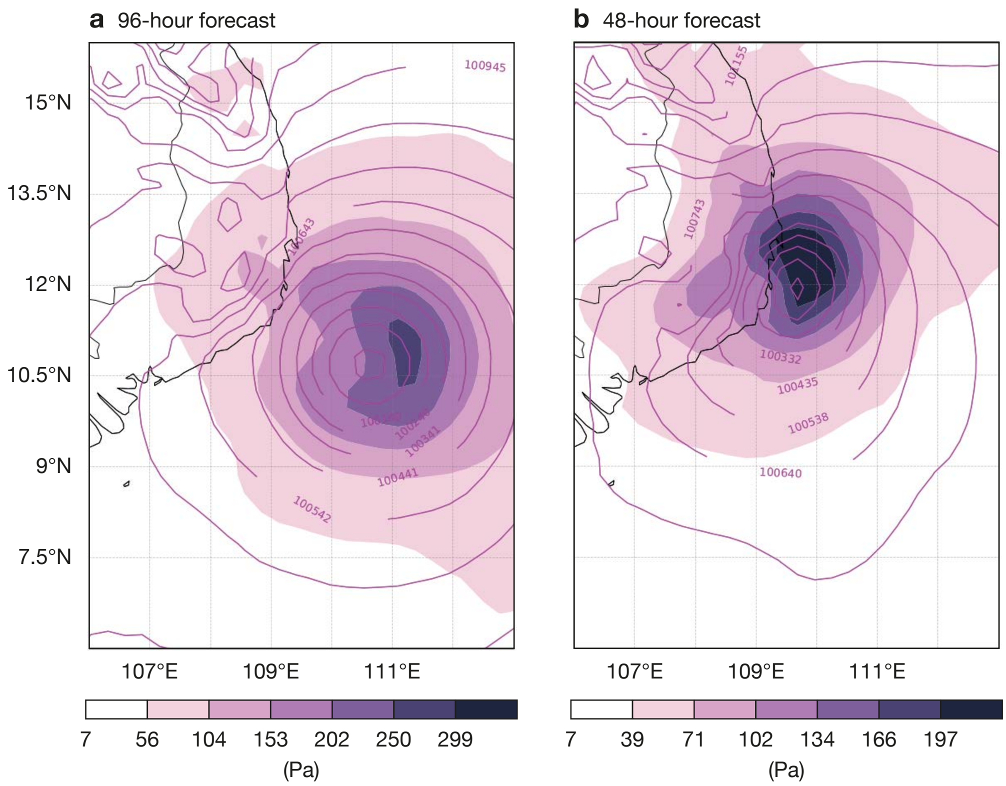


FIGURE 2 OpenIFS 96-hour (left) and 48-hour (right) ensemble forecasts at TL639 resolution with 19+1 members, starting from two of the ensemble initial states, valid at 00 UTC 4 November 2017 (landfall of Typhoon Damrey in Vietnam), showing the ensemble mean (contours, in Pa) and the ensemble spread (shading) for mean sea-level pressure.

AOSCM. In this situation, the advection of moisture is very important for the surface energy budget. However, it turns out that the difference in net surface radiation between coupled and uncoupled AOSCM simulations is as large as omitting moisture advection in the uncoupled large eddy simulation. (*Kerstin Hartung, Hamish Struthers and Gunilla Svensson*)

Ensemble prediction

The responsibility for advancing ensemble forecasting rests largely on the shoulders of major operational forecasting centres, like ECMWF, or big limited-area modelling consortia. The reason is that ensemble forecasting is computationally very demanding and requires means to generate ensemble spread through initial state and model perturbations. For the latter, OpenIFS includes the Stochastically Perturbed Parametrization Tendencies (SPPT) scheme to generate forecast ensembles. However, until recently no solution outside of ECMWF was available for the former.

Recently researchers at the Finnish Meteorological Institute recreated all 50+1 IFS operational ensemble initial states at a range of resolutions, so that both Ensemble of Data Assimilations (EDA) and singular vector (SV) perturbations can be used independently. Initial states covering one year (December 2016 to November 2017) twice a day were created with IFS Cycle 43r3 for TL639, TL399 and TL159 resolutions. The contributions of initial state perturbations and SPPT to OpenIFS ensemble skill are currently being assessed. Once this has been done, the dataset will be made available through an ftp-server under a Creative Commons licence.

Predicting Typhoon Damrey

To illustrate the potential of running your own ensemble experiments with OpenIFS, a case study of Typhoon Damrey, which severely affected Vietnam, is shown in Figure 2. Both ensembles consist of 19+1 OpenIFS forecasts produced at TL639 resolution (corresponding to a grid spacing of about 32 km). Ensemble spread here is the result of starting the forecasts from different atmospheric states (EDA + SV perturbations). The ensemble mean in the 96-hour forecast indicates that most ensemble members have propagated the typhoon too slowly and on a track that is too far south compared to what was observed (not shown). However, the ensemble spread indicates that several ensemble members place the typhoon further north and closer to the coast. The ensemble mean in the 48-hour ensemble forecast places the typhoon closer to the observed landfall location. The ensemble spread still indicates a range of possible landfall locations, but the most likely area starts to be relatively compact. (*Pirkka Ollinaho*)

Modern meteorological education

Several universities provide training in meteorological modelling and computing for Masters and PhD students. The courses at the École Nationale de la Météorologie of Météo-France, University of Helsinki and Eötvös Loránd University in Budapest, for example, aim to develop and improve their students' work-relevant skills, bridging the gap between their studies and real research using a state-of-the-art numerical weather prediction model with a variety of approaches. The courses strengthen the students' ability to work in teams and discuss questions in English, skills which are highly desirable in meteorology, where the best results are often achieved in international collaborations.

École Nationale de la Météorologie

The OpenIFS team returned to Météo-France for the third year in 2018 to run a training course for students at the school. The students followed a tutorial using ECMWF ensemble forecast data of Hurricane Nadine during the HyMeX (Hydrological cycle in Mediterranean experiment) observational campaign in 2012. The focus was on Nadine's life cycle when it interacted with the mid-latitude flow. The high forecast uncertainty of this situation (Figure 3) is ideal for teaching, as it demonstrates the need to produce ensemble forecasts. The exercises simulated a real forecast situation and put the students in the position of a duty forecaster supporting the HyMeX scientific director in decision-making for flight planning based on probabilistic forecast information. For more details on this event, see Carver et al. (2016).

University of Helsinki

The NumLab laboratory course has been taught at the University of Helsinki since the 1970s. OpenIFS has been used in NumLab since 2015, motivated by the fact that the EC-Earth model and the limited-area model HARMONIE, applied in climate research and operational short-term forecasts in Finland respectively, have the same or a similar atmospheric dynamical core as the IFS. In the first part of the course, the students are given an overview of OpenIFS and learn how to compile and run the model, post-process and analyse model outputs and improve their

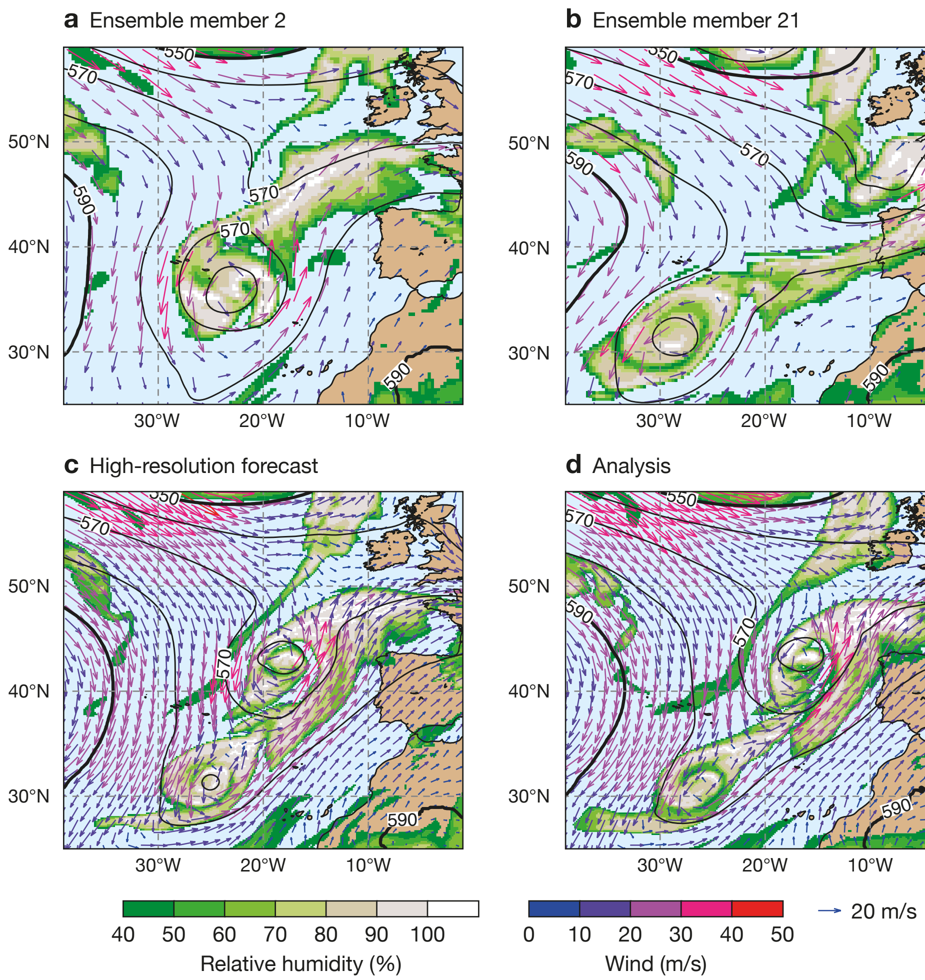


FIGURE 3 The charts show 60-hour forecasts of relative humidity at 700 hPa (shading), geopotential height (contours, in decametres) and wind (arrows) at 500 hPa at 12 UTC on 22 September 2012 according to (a) member 2 of ECMWF's operational ensemble forecast, (b) member 21 of the operational ensemble forecast, (c) the high-resolution forecast, and (d) the operational analysis. The ensemble members predict two substantially different scenarios for the interaction between Nadine and an Atlantic cut-off low. Member 2 represents a scenario where the two systems rotate around each other tending to merge, while member 21 shows that Nadine moves westwards while the cut-off low shifts eastwards. The high-resolution forecast indicated both moving eastward. The ECMWF operational analysis showed divergence between the two systems.

Linux and high-performance computing skills. In the second part, they work in small groups of three to four students on a small research project. A different scientific topic or specific historical weather event is selected each year. The students design their research questions and conduct numerical experiments. For example, in the year when the focus was on deep convection, a group explored the impact of entrainment on the cloudiness and convective precipitation over central Africa. To date more than 50 students have learnt how to use OpenIFS, and the university plans to open up the course to remote participation in the future.

Eötvös Loránd University

A two-semester Master's course is dedicated to numerical weather prediction and climate modelling at Eötvös Loránd University in Budapest. In the first term, meteorology students learn about the basics of meteorological modelling from modelling practitioners at the Hungarian Meteorological Service. In the following semester, applied mathematics students join them and form small teams with the meteorology students. The teams work on different modelling topics and conduct simulations with numerical models available for educational purposes, such as OpenIFS. In 2018, a group tested the evaluation tool developed at ECMWF for OpenIFS 40r1v2 before its release (Szépszó & Carver, 2018). The programme closes with the teams presenting the background of their chosen topic and the experiments they have carried out and discussing the conclusions that can be drawn from their results. (*Gabriella Szépszó, Victoria Sinclair and Glenn Carver*)

Further reading

Carver, **G.**, **S. Kertész**, **F. Váňa**, **F. Ferry** & **E. Chabot**, 2016: Météo-France hosts OpenIFS workshop, *ECMWF Newsletter* **No. 149**, 2–3.

Hartung, **K.**, **G. Svensson**, **H. Struthers**, **A.-L. Deppenmeier** & **W. Hazeleger**, 2018: An EC-Earth coupled atmosphere–ocean single-column model (AOSCM) for studying coupled marine and polar processes, *Geosci. Model Dev.*, **11**, 4117–4137.

Szépszó, **G.** & **G. Carver**, 2018: New forecast evaluation tool for OpenIFS, *ECMWF Newsletter* **No. 156**, 14–15.

The ECMWF Production Data Store

Laurent Gougeon

To ensure the timely delivery of ECMWF's forecasts to Member and Co-operating States and other users, observations from around the globe have to be collected fast and forecasts have to be delivered to users reliably and to a tight schedule. These processes of collection and dissemination used to be taken care of by separate software applications. However, the two activities have similar requirements for data services such as storage, transmission, scheduling, security and monitoring. After careful consideration, in 2016 ECMWF decided to combine the two applications into a single one, the ECMWF Production Data Store (ECPDS).

Since then, the ECPDS software has been developed in-house to support the goals of ECMWF's Strategy with the following objectives in mind:

- secure the evolving needs of the forecasting system for larger volumes (higher resolution) and a greater variety of observations
- support the increasing number of parameters for our forecast users
- speed up the delivery of our forecast products
- enable the ECMWF Data Services function to deliver products to national meteorological and hydrological services around the world within the framework of the World Meteorological Organization (WMO) and to a steadily growing number of commercial customers
- make the transition to cloud computing infrastructure for increased scalability and reliability.

This article presents the solutions that enable ECPDS to meet these objectives and explains its main features.

Basic features

ECPDS has been designed as a multi-purpose repository, hereafter referred to as the Data Store, delivering three strategic data-related services (Figure 1):

- Data Acquisition: the automatic discovery and retrieval of observational data from data providers
- Data Dissemination: the automatic distribution of meteorological products to our Member and Co-operating States and other forecast users
- Data Portal: the pulling of meteorological products and pushing of observational data initiated by remote sites.

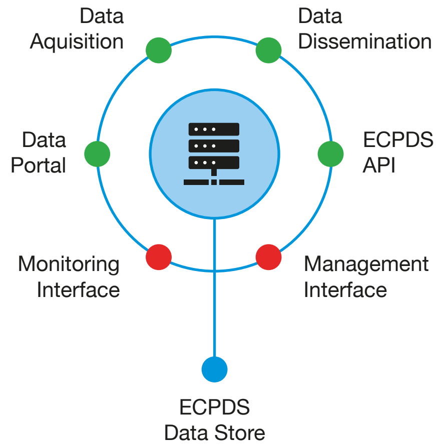


FIGURE 1 Schematic representation of the main components of ECPDS.

Data Acquisition and Data Dissemination are active services which are initiated by ECPDS, whereas the Data Portal is a passive service triggered by incoming requests from remote sites. In other words, the Data Portal service provides interactive access to the Data Dissemination and Data Acquisition services.

Unlike a conventional data store, ECPDS does not necessarily physically store the data in its persistent repository but rather works like a search engine, by crawling and indexing metadata from data providers. However, ECPDS can cache data in its Data Store. This is useful to ensure data availability in ECPDS without relying on instant access to the data providers. Feeding data into the Data Store can be achieved via different mechanisms:

- the Data Acquisition service discovering and fetching data from data providers
- data providers actively pushing data through the Data Portal
- data providers using the ECPDS API to register metadata: the ECPDS retrieval scheduler will asynchronously fetch data from the data providers or the data will be retrieved on the fly when needed.

doi: 10.21957/83deq5lgc0

Data products can be searched for online in ECPDS by name or metadata and their content can be either pushed by the Data Dissemination service or pulled from the Data Portal by users to their place of choice. Whenever data is required, ECPDS will either stream it on the fly from the data provider or send it straight from its Data Store, providing it was previously fetched by the ECPDS retrieval scheduler.

One key aspect of ECPDS is that it can easily interact with all sorts of environments without putting the burden on data providers or users to support one protocol or another:

- outgoing connections initiated by the Data Acquisition and Dissemination services use the most common protocols, such as ftp, sftp, ftps, GlobusFtp, http/s and AmazonS3
- incoming connections to the Data Portal service can be made using various protocols, such as ftp, sftp, scp and https.

Subsequent options vary depending on the selected protocol and authentication method (e.g. password-based vs key-based authentication, plaintext vs cyphertext connection, connection pooling vs straight connection, parallel connection vs serial connection, and more).

The ECPDS software is modular and every protocol is deployed as an extension to the system. All extensions make use of a common API which hides the complexity from the user and gives a unified view of the Data Store. On a regular basis, new protocols will be added when they become available or are required by some of our users. With the emergence of cloud computing, the AmazonS3 protocol was added recently and support for Microsoft Azure is currently in the pipeline. These new protocols are being introduced in ECPDS to make it possible to disseminate meteorological products directly to cloud object storage.

Another key aspect of ECPDS is its ability to reach a wide range of data providers (since forecast skill depends on the availability of observations from a range of sources around the world), users in Member and Co-operating States and others by operating through multiple networks:

- the Internet with a high-speed connection
- the Regional Meteorological Data Communication Network (RMDCN)
- the ECaccess Network with many ECaccess gateways deployed worldwide
- superfast bandwidth networks with dedicated leased lines.

Object storage

ECPDS stores data as 'objects', in other words as data with associated metadata and a globally unique identifier. The object storage system has been built on top of a file-system-based solution, with an efficient replication mechanism that allows continuous data availability across ECMWF and cloud platforms. To this end, ECPDS duplicates data in geographically separated storage locations, such as the US or Canada. This also brings ECMWF products closer to forecast users, enhancing data transfer performance to provide users with more immediate access to forecast products at scheduled release times. The ECPDS data persistence layer, along with its underlying raw storage, is also capable of interfacing with other object storage systems: in the future, ECPDS could expand its data storage capacity across object stores in the cloud, which would be useful in terms of scalability and reliability.

Like any object storage system, the one used in ECPDS is hierarchy-free and has no nested tree structure to store its data. However, ECPDS is able to emulate a directory structure when required. For instance, data providers can create directories when they transmit data to the Data Portal, and ECPDS will store this information as metadata along with the data. Later on, when the data is requested, ECPDS will be able to reuse this information to build a virtual file system with directory trees. Using the metadata enables ECPDS to present different views of the same data, depending on the configuration. For example, one user might see the data organised by date and another might see the data organised by type, depending on the users' preferences, managed via their ECPDS profile.

Other specificities of the ECPDS Data Store are:

- Data compression: this enables the Data Dissemination service to compress data. Compression can be performed either on the fly while

Useful links

Access to the ECPDS monitoring and managing interface:

- *https://ecpds-monitor.ecmwf.int*

Access to the Data Portal through the Internet:

- *https://dissemination.ecmwf.int*
- *https://acquisition.ecmwf.int*

Access to the Data Portal through the RMDCN:

- *https://acquisition-rmdcn.ecmwf.int*

transferring the data to the remote site, or in advance while the data is sitting in the queue before transmission. Most commonly used compression algorithms are supported, including zip, gzip, bzip or lzma. This is important because compression reduces dissemination time and therefore enables faster access to forecast products.

- Data checksumming: this makes it possible to generate, in advance or on the fly, MD5, CRC32, SHA-1 or SHA-256 one-way hashes to preserve the integrity of data against corruption. The Data Dissemination and Data Portal services can provide the cryptographic hashes along with the data, for verification. This feature is used routinely to identify corrupted files at remote sites.
- Garbage collection: every piece of data recorded in ECPDS is given an expiry date. The ‘garbage collector’ makes it possible to automatically remove data that no longer need to be stored. When required, the garbage collector can also take care of cleaning up on the data provider side. It is worth mentioning that there is no limit on the expiry date, which means that, if required, data can stay in the Data Store for ever.
- Data backup: this makes it possible to map entire sets of data in ECPDS to a wide variety of archiving systems (e.g. ECFS).

Monitoring and management

ECPDS offers an uninterrupted 24/7 service and support for the acquisition and dissemination activities. Its monitoring interface is essential for the smooth operation of ECPDS. The software is interfaced with a Nagios server so that ECMWF operators can monitor its internal operations, but the system also has its own monitoring interface with dedicated tools to trace and debug issues specific to the ECPDS services.

An administration interface is also available to manage the various components of ECPDS:

- data storage: managing metadata, data content, transfer groups and Data Movers (activating or deactivating servers)
- transmission: managing destinations, transfer hosts (remote site settings) and transfer requests
- access control: managing users who perform monitoring tasks and users who have access to the Data Portal
- monitoring: a monitoring display for the Dissemination and Acquisition services.

Theses functionalities are available either through a website or a REST API for easy integration with other systems.

ECPDS concepts for users

Understanding some key ECPDS concepts will help users to benefit fully from the tool’s capabilities:

- data files
- data transfers
- destinations
- dissemination and acquisition hosts.

A user connecting to the ECPDS web interface will come across each of these entities, which are related to each other through the different services.

A **data file** is a record of a product stored in the ECPDS Data Store with a one-to-one mapping between the data file and the product. The data file contains information on the physical specifications of the product, such as its size, type, compression and entity tag (ETag) in the Data Store, as well as the metadata associated with it by the data provider (e.g. meteorological parameters, name or comments concerning the product).

A **data transfer** is linked to a unique data file and represents a transfer request for its content, together with any related information (e.g. schedule, priority, progress, status, rate, errors, history). A single data file can be linked to several data transfers as many remote sites might be interested in obtaining the same products from the Data Store.

A **destination** should be understood as a place where data transfers are queued and processed in order to deliver data to a unique remote place, hence the name ‘destination’. It specifies the information the Data Dissemination service needs to disseminate the content of a data file to a particular remote site (Figure 2).

Each destination implements a transfer scheduler with its own configuration parameters, which can be fine-tuned to meet the remote site’s needs. These settings make it possible to control various things, such as how to organise the data transmission by using data transfer priorities and parallel transmissions, or how to deal with transmission errors with a fully customisable retry mechanism.

In addition, a destination can be associated with a list of **dissemination hosts**, with a primary host indicating the main target system where to deliver the data, and a list of fall-back hosts to switch to if for some reason the primary host is unavailable.

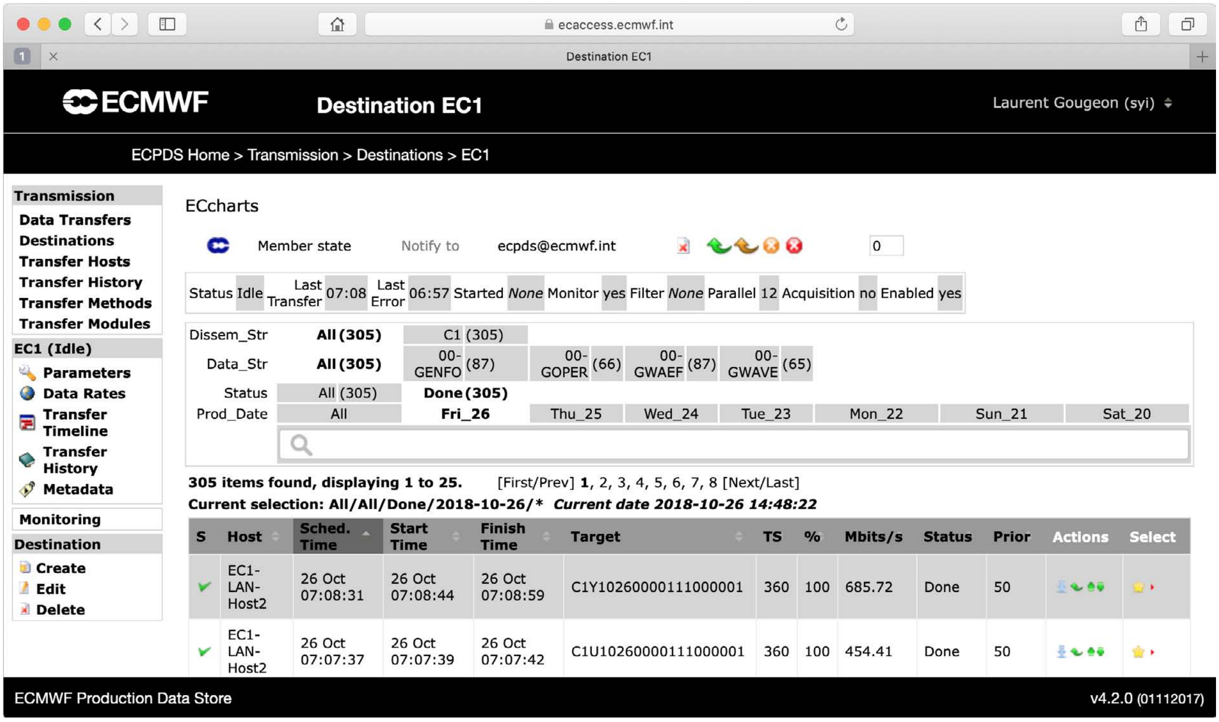


FIGURE 2 The ECPDS interface for internal and external users. A breadcrumb trail at the top shows where a user currently is in the tool. In this case, a user has created a destination called EC1. Users with the right credentials can see the status of this destination and can review the progress of data transmission. They can manage the destination by, for example, requesting data transfers, changing priorities and stopping or starting data transmissions.

A dissemination host is used to connect and transmit the content of a data file to a target system. It enables users to configure various aspects of the data transmission, including which network and transfer protocol to use, in which target directory to place the data, which passwords, keys or certificates to use to connect to the remote system, and more.

If the data transfers within a destination are retrieved by remote users through the Data Portal service, then there will be no dissemination hosts attached to the destination. In this particular case, the destination can be seen as a 'bucket' (in Amazon S3 terms) or a 'blob container' (in Microsoft Azure terms). The transfer scheduler will be deactivated, and the data transfers will stay idle in the queue, waiting to be picked up through the Data Portal.

A destination can also be associated with a list of **acquisition hosts**, indicating the source systems where to discover and retrieve files from remote sites. Like their dissemination counterparts, the acquisition hosts contain all the information required to connect to the remote site, including which network, transfer protocol, source directory and credentials to use for the connection. In addition, the acquisition host also contains the information required to select the files at the source. Complex rules can be defined for each source directory, type, name, timestamp and protocol, to name just a few options.

A destination can be a dissemination destination, an acquisition destination or both. It will be a dissemination destination as long as at least one dissemination host is defined, and it will be an acquisition destination as long as at least one acquisition host is defined. When both are defined, then the destination can be used to automatically discover and retrieve data from one place and transmit it to another, with or without storing the data in the Data Store, depending on the destination configuration. This is a popular way of using ECPDS. For example, this mechanism is used for the delivery of some regional near-real-time ensemble air quality forecasts produced at ECMWF for the EU-funded Copernicus Atmospheric Monitoring Service implemented by the Centre.

There is also the concept of destination aliases, which makes it possible to link two or more destinations together, so that whatever data transfer is queued to one destination is also queued to the others. This mechanism enables processing the same set of data transfers to different sites with different schedules and/or transfer mechanisms defined on a destination basis. Conditional aliasing is also possible in order to alias only a subset of data transfers.

Developing all these features has made it possible for ECMWF to meet the requirements of our main user groups and stakeholders:

- Member and Co-operating States
- the WMO community
- commercial customers.

It also means that ECMWF can flexibly interface with all the observation providers around the world the Centre depends on.

Physical architecture

ECPDS is a distributed application with three main software components (Figure 3):

- the ECPDS Master implements the business logic of the application
- the MySQL Database enables persistent storage for configurations, metadata and history
- Data Movers make it possible to store objects and to perform incoming and outgoing data transfers.

The ECPDS Master and MySQL Database both run in a highly resilient environment. The Internet, RMDCN and LAN Data Movers currently run on physical servers deployed in their own secure environment. There are currently 37 Data Movers available for a total capacity of 1.2 petabytes.

In order to prevent downtime and data loss, which is a critical requirement of ECMWF's Member and Co-operating States and is needed for the real-time running of the forecasting system at ECMWF, the following mechanisms have been implemented:

- The ECPDS Master and MySQL Database are each replicated on three servers in an active/passive mode: one active instance handles requests and two passive instances are on standby. When the active instance fails or requires maintenance, one of the passive instances takes over and the service resumes as normal.
- Each Data Mover group includes multiple servers. In order to guarantee the availability of the files in a group, a replication process is run to copy the files across the Data Movers (each file is replicated three times). If relevant, replication is also performed between ECMWF Data Movers and Data Movers located in the cloud on other continents.

A load balancer is configured to distribute incoming requests to the ECPDS Data Portal amongst the Data Movers. The use of multiple Data Movers and load balancing increases availability through redundancy.

ECMWF upgrades its forecasting system on a regular basis, and such upgrades are usually accompanied by a large increase in the volume of data to distribute. The way to scale up ECPDS in this situation is by adding more Data Movers. Each Data Mover adds CPU and I/O capability and disk space resources. More Data Movers can handle a bigger workload and more data.

Future challenges

In order to support their delivery of forecast services to society, ECMWF's growing number of forecast users require quick and reliable access to ECMWF forecasts. As a result, the average data transmission volume handled by ECPDS is approaching 1 petabyte per month with an exponential increase in Internet traffic. ECMWF forecast products are disseminated to 547 places in 78 countries, and observational data are retrieved from 557 places in 34 countries. Figure 4 shows a map of locations from where data are retrieved in Europe. The huge rise in traffic observed in recent years has successfully demonstrated the reliability,

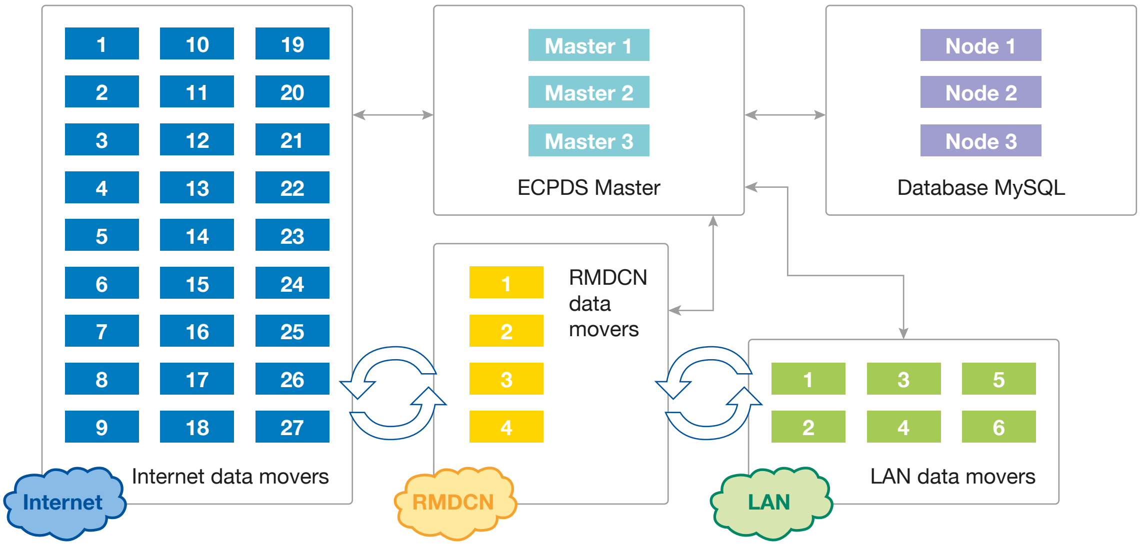


FIGURE 3 ECPDS physical architecture. The key components are the ECPDS Master, the MySQL database and the Data Movers.

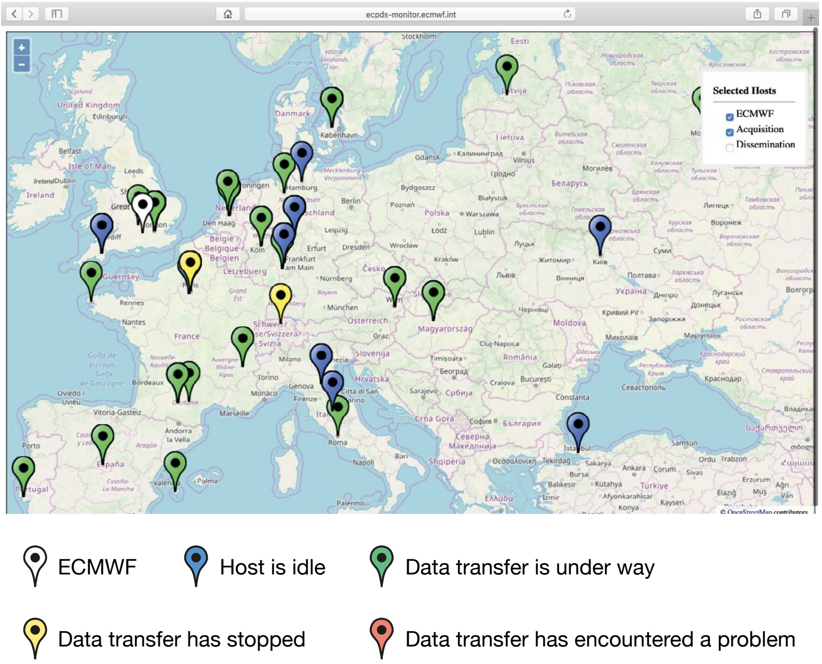


FIGURE 4 The interactive map in the ECPDS user interface shows the locations of hosts across the world, in this case over Europe for acquisition hosts only. A similar map is available for dissemination hosts. Different colours indicate the status of data transfer requests. A concentration of red hosts might indicate a network issue in the region. (Copyright OpenStreetMap *www.openstreetmap.org/copyright*)

availability and scalability of ECPDS. However, there are other challenges that will need to be addressed in the near future:

- ECMWF's new data centre in Bologna: there will be two independent data halls, a new network topology and a new high-performance computing facility. ECPDS will have to adapt simultaneously to a changing infrastructure and changing technologies on top of increasing traffic.
- ECPDS is turning to cloud computing to expand its potential: ECPDS is already running Data Movers in the cloud, but this is just a beginning and our engagement with the wider community as part of the European Weather Cloud and the HiDALGO and LEXIS EU-funded projects should help us to consolidate ECPDS's position in this field.

Overall, ECPDS is now a mature solution which has helped to significantly improve efficiency and productivity of our data services by using proven and innovative technologies. With ECPDS, ECMWF delivers a portable and adaptable application which can fit diverse environments as well as providing a user-friendly tool to run data-related services. Stay tuned!

ECMWF publications

(see *www.ecmwf.int/en/research/publications*)

Technical Memoranda

844 **Balan-Sarojini**, **B.**, **S. Tietsche**, **M. Mayer**, **M. Alonso-Balmaseda** & **H. Zuo**: Towards Improved Sea Ice Initialization and Forecasting with the IFS. *March 2019*

843 **Munoz-Sabater**, **J.**, **H. Lawrence**, **C. Albergel**, **P. de Rosnay**, **L. Isaksen**, **S. Mecklenburg**, **Y. Kerr** & **M. Drusch**: Assimilation of SMOS brightness temperatures in the ECMWF IFS. *February 2019*

841 **Vitart**, **F.**, **G. Balsamo**, **J.R. Bidlot**, **S. Lang**, **I. Tsonevsky**, **D. Richardson** & **M. Alonso-Balmaseda**: Use of ERA5 to Initialize Ensemble Re-forecasts. *February 2019*

840 **Hewson**, **T.**: Use and Verification of ECMWF Products in Member and Co-operating States (2018). *March 2019*

839 **Bormann**, **N.**, **H. Lawrence** & **J. Farnan**: Global observing system experiments in the ECMWF assimilation system. *January 2019*

838 **Inness**, **A.**, **I. Alben**, **A. Agusti-Panareda**, **T. Borsdoff**, **J. Flemming**, **J. Landgraf** & **R. Ribas**: Monitoring and assimilation of early TROPOMI total column carbon monoxide data in the CAMS system. *January 2019*

836 **Browne**, **P.**, **P. de Rosnay**, **H. Zuo**, **A. Bennett** & **A. Dawson**: Weakly coupled ocean-atmosphere data assimilation in the ECMWF NWP system. *January 2019*

EUMETSAT/ECMWF Fellowship Programme Research Reports

49 **Lean**, **K.** & **N. Bormann**: Moving to GOES-16: a new generation of GOES AMVs. *January 2019*

ECMWF Calendar 2019

May 2–3	Finance Committee
May 3	Policy Advisory Committee
May 13–17	Online training week: Software and computing at ECMWF
Jun 3–6	Using ECMWF's Forecasts (UEF)
Jun 10–13	Workshop on observational campaigns for better weather forecasts
Jun 17–21	OpenIFS user workshop (University of Reading)
Jun 27–28	Council
Sep 2–5	Annual Seminar
Sep 24	Extraordinary Council
Sep 25–27	Workshop on next generation HPC I/O
Oct 7–9	Scientific Advisory Committee
Oct 7–10	Training course: Use and interpretation of ECMWF products
Oct 10–11	Technical Advisory Committee
Oct 14–16	Workshop on robust scientific developments with reproducible workflows
Oct 28–29	Finance Committee
Oct 29	Policy Advisory Committee
Nov 18–21	Workshop on stratospheric modelling, predictability and influence on the troposphere
Nov 25–28	Satellite inspired hydrology for an uncertain future: an H SAF and HEPEX workshop
Dec 10–11	Council

Contact information

ECMWF, Shinfield Park, Reading, RG2 9AX, UK

Telephone National	0118 949 9000
Telephone International	+44 118 949 9000
Fax	+44 118 986 9450

ECMWF's public website *www.ecmwf.int/*

E-mail: The e-mail address of an individual at the Centre is firstinitial.lastname@ecmwf.int. For double-barrelled names use a hyphen (e.g. j-n.name-name@ecmwf.int).

For any query, issue or feedback, please contact ECMWF's Service Desk at *servicedesk@ecmwf.int*.

Please specify whether your query is related to forecast products, computing and archiving services, the installation of a software package, access to ECMWF data, or any other issue. The more precise you are, the more quickly we will be able to deal with your query.

Newsletter | **No. 159** | Spring 2019
European Centre for Medium-Range Weather Forecasts
www.ecmwf.int