ECMWF

European Centre for Medium-Range
Weather Forecasts

Newsletter

No. 175 | Spring 2023

980

940

960

----

editorial

Reanalysis

An important activity that goes hand in hand with numerical weather prediction is the establishment of a record of the recent climate. This is done by the EU-funded Copernicus Climate Change Service (C3S) implemented by ECMWF. The method is to apply a modern Earth system data assimilation system, which is also used to initialise numerical weather forecasts, to past observations. This way it becomes possible to obtain as accurate an estimate of the past state of the Earth system as possible. We have already produced several reanalyses of the weather and climate, and the most recent one is called ERA5. It is important to realise that the production of reanalyses is an ongoing process. For example, ERA5 has successively been extended further back, and now it can tell us in some detail how the global atmosphere, land surface and ocean waves have developed from 1940 to the present. The way ERA5 was pushed back from 1958 to 1940 is described in this Newsletter.

But developing reanalysis does not stop there. Every improvement of the forecasting model and the data assimilation system brings new scope for also improving our reanalysis capabilities. And there is a continuing quest to use more observations of the past in our reanalysis efforts. That is why we are already working on the next version of reanalysis, called ERA6. Another article in this Newsletter reports on efforts to make more satellite observations available for ERA6 and other future reanalyses.

Maintaining and developing the data assimilation system is, of course, also important for numerical weather prediction. A feature article in this Newsletter describes how data assimilation tests using the latest version of the linearised physics have shown a systematic and significant improvement for all parameters, levels, and regions. Another article describes the extension of the OpenIFS facility to atmospheric composition. OpenIFS enables external users to carry out research with a portable version of ECMWF's Integrated Forecasting System (IFS). The extension, which was done jointly with Member State scientists in the Netherlands and Finland, will significantly widen the use to which OpenIFS can be put. Two computing articles illustrate the importance of technical developments in weather forecasting: one deals with ECMWF's upgrade of the file format of weather-related parameters, which is to be completed in the next few years; and the other describes the new role of automation in ECMWF processes. The latter was implemented as we moved the data archive to our new high-performance computing facility in Bologna in preparation for the next upgrade of the IFS to Cycle 48r1. The upgrade is expected to happen in June this year, and next year we plan to start the new reanalysis ERA6.

This Newsletter illustrates, on the one hand, how all our activities are interconnected, from computing and numerical weather prediction to atmospheric composition and climate monitoring. On the other hand, it shows the collaborative dimension of our work: ECMWF could not deliver its Strategy without close interactions with meteorological services and other organisations.

Florence Rabier
Director-General

Contents

Forecast performance 2022

Thomas Haiden, Matthieu Chevallier, David Richardson

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 2022 are given in the box.

Cycle 47r3 of ECMWF's Integrated Forecasting System (IFS), implemented on 12 October 2021, included a major upgrade to the moist physics of the model. Together with a number of other changes, this cycle improved upper-air forecasts in the early medium range such that in 2022 new high points in skill relative to forecasts based on the ERA5 reanalysis system were reached (see the figure). This cycle has been ported to the Atos high-performance computing facility in Bologna (Italy), where it has been running operationally since 18 October 2022.

Compared to forecasts from other global centres, ECMWF has been able to maintain the overall lead for upper-air parameters in the medium range. For surface parameters such as 2 m temperature and precipitation, some of the other centres have drawn closer to ECMWF in the medium range and partly taken the lead in the short range. However, the number of large 2 m temperature errors in ensemble forecasts (ENS) has been further reduced according to one of ECMWF's headline scores, defined as the continuous ranked probability score (CRPS) being larger than 5 K for forecast day 5.

The skill of the Extreme Forecast Index (EFI) has increased for 10-metre wind speed and 2-metre temperature, and slightly decreased for precipitation. In the extended range, the skill of predicting terciles of 2-metre temperature anomalies in the extratropics in week 2 now consistently exceeds that of persistence by about 10% and that of climatology by 40%. Corresponding values for skill over weeks 3 and 4 combined are about 5% (with respect to persistence) and 20% (with respect to climatology). Summer

2022 positive temperature anomalies over Europe and northern Russia were captured at weeks 1 and 2 in ECMWF medium-range and early extended-range forecasts, whilst negative temperature and positive precipitation anomalies with signals of severe flooding over Pakistan were indicated up to three weeks ahead in extended-range forecasts.

Position errors for forecasts of tropical cyclones were similar to those in the year before. There is also still a notable slow bias in terms of propagation speed. With regard to forecast skill for ocean wave parameters, there has been a substantial improvement in wave peak period due to IFS Cycle 47r3. ECMWF is now leading compared to forecasts from other global centres both in peak period and significant wave height.

Early in 2022, ECMWF's seasonal forecast predicted a gradual return from La Niña towards more neutral conditions within a few months, but in the observations La Niña conditions persisted throughout the year. Later in 2022, the forecast got closer to observations and shifted the return to

neutral into early 2023. In the extratropics, a strong signal for a warm anomaly in the summer of 2022 in Europe was present in the forecast but its magnitude towards Scandinavia and Siberia was underestimated. The negative temperature anomaly over Pakistan, linked to a strongly positive precipitation anomaly leading to widespread severe flooding in the area, was indicated by the seasonal forecast for the summer of 2022.

The complete set of annual verification results is available in ECMWF Technical Memorandum No. 902 on 'Evaluation of ECMWF forecasts, including the 2021 upgrade', downloadable from http://www.ecmwf.int/en/publications/technical-memoranda.

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

- Verification as part of ECMWF's charts page:
  https://charts.ecmwf.int

- World Meteorological Organization (WMO) inter-comparison of global model forecast skill:
  https://wmolcdnv.ecmwf.int/
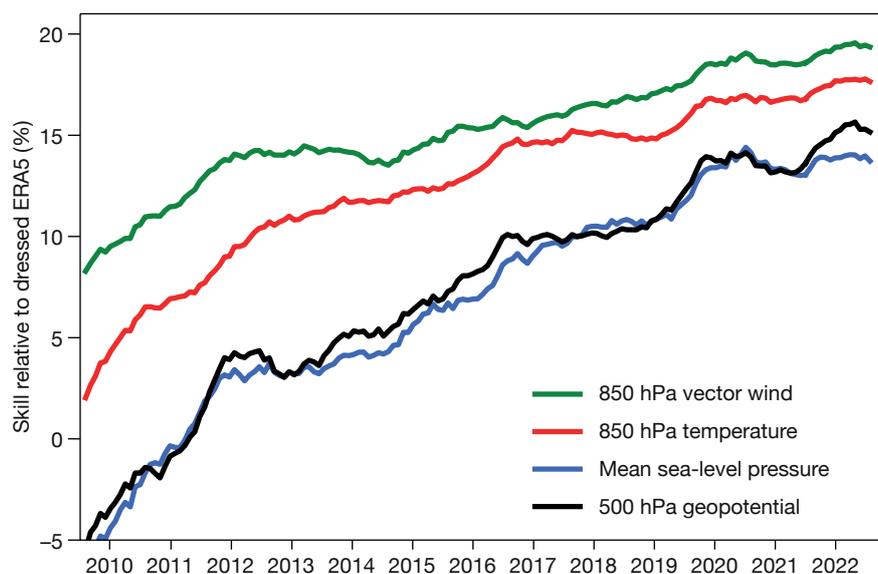
Upper-air skill of the ENS relative to forecasts based on the ERA5 reanalysis system. Skill of the ENS for upper-air parameters at day 5 in the northern extratropics, relative to a Gaussian-dressed ERA5 forecast. Values are running 12-month averages, and verification is performed against own analysis.

- WMO ocean wave model inter-comparison results: https://confluence.ecmwf.int/display/WLW/WMO+Lead+Centre+for+Wave+Forecast+Verification+LC-WFV

- List of 'Known IFS Forecasting Issues': https://confluence.ecmwf.int/display/FCST/Known+IFS+forecasting+issues

- IFS cycle changes since 1985: http://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model

Assessment of ECMWF's Technical Advisory Committee, 6–7 October 2022

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

a) congratulated ECMWF on progress with the new high-performance computing facility in Bologna, noting that a test system for the CAMS global analyses and forecasts is already running, and is looking forward to the successful switch-over of the operational forecast system on 18 October 2022;

b) recognised that ECMWF maintains a lead compared to other centres for a range of upper-air and stratospheric verification scores, medium-range ensemble precipitation, and the medium-range spread-skill relationship;

c) noted, also, that ECMWF now performs less well than some other centres for some surface parameters in the shorter range, for example ensemble 2 m temperature, whilst recognising that proposed improvements in 48r1 and 49r1 will address some of these issues; also, in the southern hemisphere ECMWF's lead in some scores for upper-air parameters has narrowed relative to other centres;

d) congratulated ECMWF on the successful implementation of IFS Cycle 47r3 in October 2021, which included an improved moist physics package and some new products introduced in response to user requests;

e) noted that the introduction of 47r3 had a positive impact on many scores, including precipitation and upper parameters and improvements visible in the early medium-range ensemble scores. However, in verification against SYNOP 47r3 showed limited impact on 2 m temperature and 10 m wind scores and led to a reduction in skill in total cloud cover, which will be addressed in future model cycles;

f) noted that ECMWF has maintained its lead compared to other centres for significant wave height and that the introduction of 47r3 has had a positive impact on peak period skill scores with ECMWF now leading other centres;

g) recognised that EFI ROC skill scores over Europe aimed at high-impact weather showed improved skill for 2 m temperature and 10 m wind speed with new high points reached, whilst scores for 24 h precipitation fell away slightly but still remained high;

h) noted that ERA5 is a useful benchmark for accounting for year-to-year variability and welcomed that the causes of the improvement seen in both HRES and ERA5 precipitation headline scores are to be investigated;

i) noted improvement in tropical cyclone HRES and ENS position errors and speed, whilst central pressure errors remained similar to last year;

j) noted that in the extended range the week 2 forecast continues to show improvement relative to persistence, but no statistically significant trends were apparent in weeks 3 and 4 compared to persistence;

k) recognised that in the extended range the JJA 2022 positive temperature anomalies over southern Europe and northern Russia and negative temperature anomaly/positive precipitation anomaly over Pakistan were captured at weeks 1 and 2 as well as at longer lead times; signals for the Pakistan flooding appeared with 3 weeks' lead time, whilst the positive temperature anomalies over northern Europe were, however, not identified at lead times of 3 weeks or longer;

l) noted that on longer-range seasonal time scales ECMWF, like some other centres, returned ENSO to a neutral state too quickly and so did not capture the ongoing La Niña at lead times of 3 months and longer;

m) noted that seasonal model outputs for DJF 2021–22 and JJA 2022 represented signals related to La Niña as well as capturing some element of the JJA 2022 warm anomaly over Europe and cold anomaly over Pakistan. However, during DJF 2021–22 in more northerly latitudes signals were more poorly captured, for example the warm anomaly over northern Russia and the cold anomaly over eastern North America;

n) acknowledged the continuing high quality of atmospheric composition products and increase in performance of some outputs in CAMS; noted upper-air scores were degraded in 47r3 and recognised that this will be addressed soon;

o) welcomed the introduction of additional datasets in verification, for example TOA net shortwave radiation and downward solar radiation at the surface, and the greater focus on near-surface variables in verification; appreciated the insightful presentation of verification data, for example scale-dependent aspects of precipitation at different lead times; suggested, based on forecaster feedback, that a verification metric reflecting duration of blocking situations, which can be underestimated, be developed;

p) appreciated the continued development of new diagnostics and very good support ECMWF provided to Member and Co-operating States over the last year, with engagement via many mechanisms including online support, the annual UEF, online seminars, site visits and meteorological representatives at Member States; the increase in in-person contact as the COVID-19 pandemic evolves is also noted.

Predicting extreme precipitation over California

Linus Magnusson, David Lavers, Fernando Prates

The winter of 2022/2023 was unusually wet in the southwestern US. The first figure shows time-series of daily area-integrated precipitation for two boxes in California from 25 December 2022 up to 20 January 2023, and a map of mean observed precipitation over the period, based on the Prism dataset (https://prism.oregonstate.edu/). During the period, several extreme days can be seen in northern and southern California. This was caused by extratropical cyclones transporting large amounts of moisture – in the form of atmospheric rivers (ARs) – from the sub-tropical Pacific. The impacts of these storms were widespread and included river flooding, large snow amounts, strong winds and high sea waves. In this article, however, the focus will be on precipitation forecasts.

Predictability

To assess predictability, we consider the most extreme event that occurred during the 24-hour period ending at 12 UTC on 10 January. We will evaluate the prediction for a relatively large region covering parts of southern California (land points within 33°N–39°N, 122.5°W–117°W, corresponding to the lower box in the first figure). The evolution of the ECMWF ensemble (ENS) and high-resolution forecasts (HRES) for this region and period is visualised in the second figure.

In the extended range, a hint of above-average precipitation in the ensemble mean (shown as black diamonds), compared to the climate mean (shown as a black horizontal line), is visible. The strongest signal in this time range is from 22 December. The signal in the extended range could be connected to the presence of a Madden–Julian Oscillation in phase 6–7, which has been shown to be linked with increased AR activity. In the 10–15 day forecast range, the ensemble mean was again somewhat higher than the climate mean, with the most extreme members above the 99th percentile of the climatology, but this was still not a strong signal.

However, between 12 UTC on 1 January and 00 UTC on 2 January the ensemble shifted toward a more extreme forecast, and for all forecasts from 12 UTC on 3 January (a week before the event) and onwards, the ensemble median is above the 99th percentile of the model climatology. From 00 UTC on 6 January, the ensemble median exceeded the model climate maximum (based on 1,200 re-forecast members). In conclusion, there was a very strong signal for a large-scale extreme event in medium-range forecasts.

The third figure shows the spatial pattern of the precipitation from the Prism dataset, from an HRES forecast at 12 UTC on 9 January, and from an HRES forecast at 12 UTC on 6 January. All three had comparable mean precipitation inside the box (54.5 mm, 52.0 mm and 53.3 mm respectively). In the observation dataset the main precipitation area was along the coast and in the Sierra Nevada mountain range to the east. The shortest forecast captures the distribution of precipitation rather well, but the precipitation along the coast was somewhat underestimated while the forecast had too much
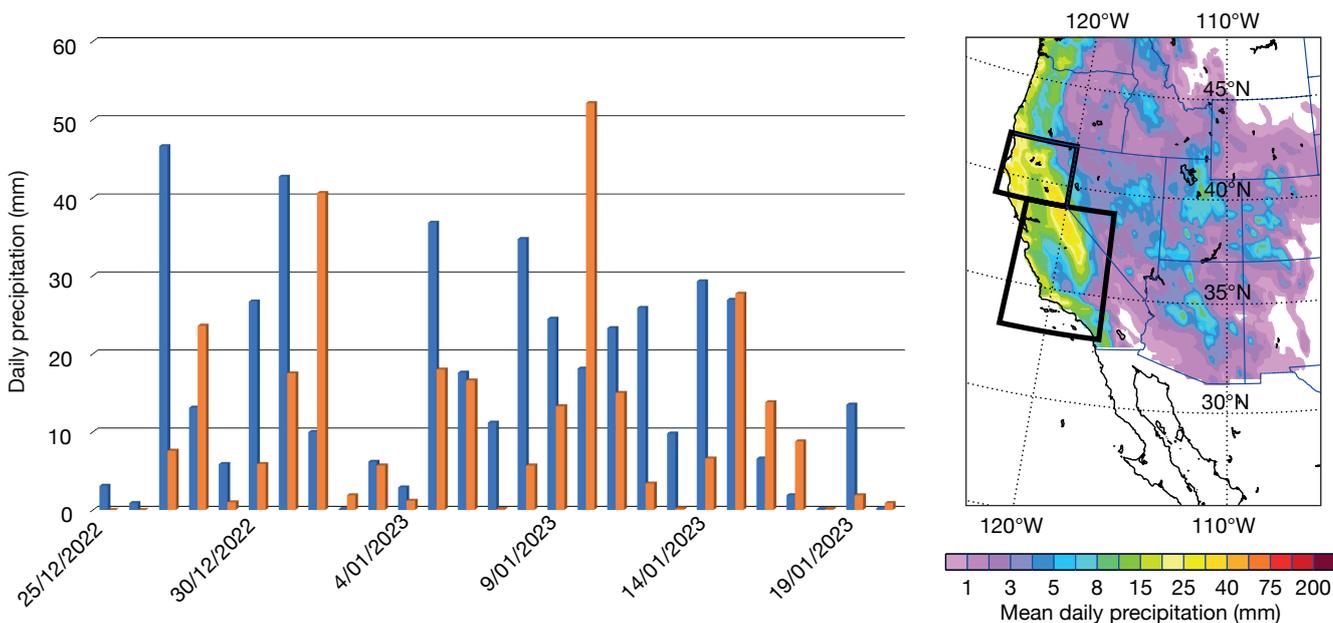
Precipitation according to the Prism dataset. The left-hand diagram shows 24-hour total precipitation ending at 12 UTC on the dates shown, in northern (blue) and southern (orange) California from the Prism dataset, in the boxes shown in the right-hand chart. The right-hand chart shows mean daily precipitation over the period 25 December 2022 to 20 January 2023 in the Prism dataset.

precipitation in the lee of the mountains, which is a sign of too weak an orographic enhancement of the precipitation. For the longer forecasts, the precipitation was shifted more to the north for this case, due to a shift in the synoptic pattern. On average, over the period from 25 December to 20 January, the short-range HRES produced similar precipitation amounts as in the Prism dataset for the southern box, but it had a 5% underestimation for the northern box. A statistical analysis of this type of cases could bring to light whether these errors are due to model biases.

Atmospheric River Reconnaissance: observations and workshop

ECMWF has engaged with the Atmospheric River Reconnaissance (AR Recon) observational campaign for several years. In this campaign, aircraft release dropsondes in dynamically sensitive regions, and additional ocean buoys with pressure sensors are deployed to capture more accurately the mean sea level pressure field across the northern Pacific. At the same time, extra radiosondes were released in California. All these extra data are used in real time in the ECMWF operational data assimilation system.

In June 2023, ECMWF is arranging
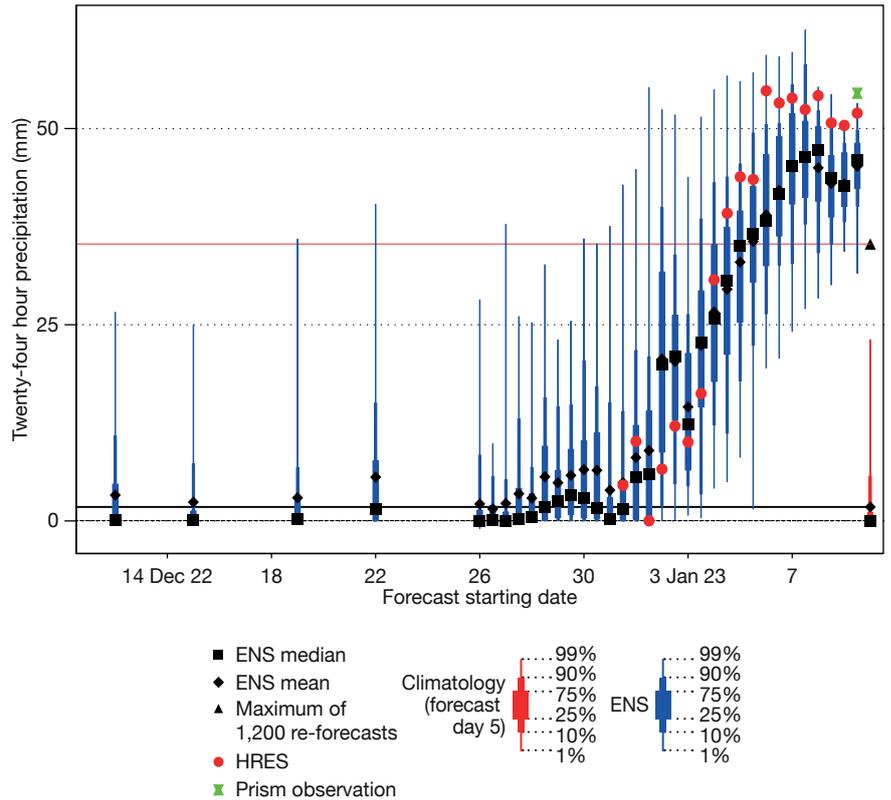
Twenty-four hour total precipitation forecasts. Forecast evolution of 24-hour total precipitation ending at 12 UTC on 10 January in the land area of southern California highlighted in the next figure.

and hosting the AR Recon workshop in Reading, UK. This event, organised jointly with US colleagues, will bring together current AR Recon participants and interested experts to share results and to coordinate and inspire future work on data collection, data assimilation, metric development and impact assessment.

The Research And Operations Partnership approach pioneered by AR Recon will also be discussed.
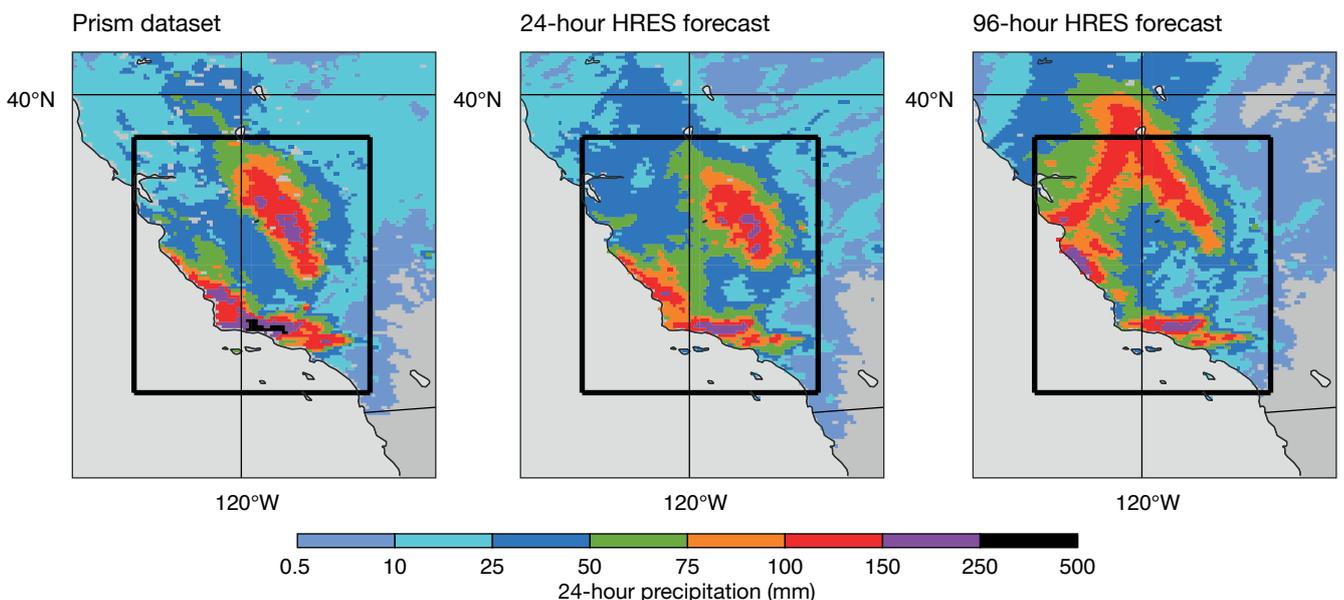
Precipitation observations and forecasts. Twenty-four hour precipitation over land ending at 12 UTC on 10 January in the Prism dataset (left), in the HRES forecast from 12 UTC on 9 January (middle), and in the HRES forecast from 12 UTC on 6 January (right). The black box shows the land integration area used in the figure above.

Supporting the humanitarian effort in Ukraine

Davide Miozzo, Sabrina Meninno, Giorgio Meschi, Fabio Violante, Rocco Masi, Martina Lagasio, Massimo Milelli (all CIMA Research Foundation), Lorenzo Massucchielli (Italian Red Cross), Yoav Levi, Pavel Khain, Alon Shtivelman, Nir Stav (all Israel Meteorological Service), Sari Lappi (WMO), Umberto Modigliani (ECMWF)

Following the escalation of the Russia-Ukraine international armed conflict in 2022, there was a need to support humanitarian operations with the provision of support based on meteorological information. The EU-funded programme for Prevention, Preparedness and Response to natural and man-made disasters in Eastern Partnership countries – phase 3 (PPRD East 3) has adapted the activities originally planned in Ukraine into operational intervention, by supporting the civil protection authorities in structuring and strengthening civil protection and humanitarian operations.

In response to the humanitarian crisis that followed the start of Russia's invasion of Ukraine, planned capacity-building activity in the area was modified, as requested by DG ECHO, addressing new needs arising from the conflict. As a result, the programme worked directly with humanitarian responders in the field to design and establish an Impact-Based Forecast (IBF) bulletin and its consequent Early Warning to Early Action (EWEA) protocols.

Main components of the Impact-Based Forecast bulletin

The IBF is a daily bulletin that provides situational awareness to organisations responsible for managing the crisis caused by the conflict. It offers information on hydro-meteorological events that have a significant impact on humanitarian operations and on the population in Ukraine and Moldova, and in a buffer area where many refugee movements were concentrated especially in the first months of the conflict.

The IBF combines information on weather variables, encompassing low temperatures, rain, wind, snow, and biometric indices, with data on particular vulnerabilities. This information includes the locations of reception centres, important border crossing points, numbers of refugees and internally displaced persons, locations of Red Cross mobile clinics, and other infrastructure, such as roads that are crucial for logistics. Most of the data on exposed assets are shared directly by the Italian Red Cross, which is part of the development team of the bulletin and uses this tool operationally for more accurate short- and medium-term planning of its humanitarian intervention in the region.

Weather forecasts to better plan humanitarian operations

Synoptical overviews of the meteorological situation are provided on a daily basis by the national meteorological and hydrological services (NHMS) of Ukraine and Moldova (Ukrainian Hydrometeorological Center and the State Hydrometeorological Service of the Republic of Moldova), while 24 h to 144 h predictions are developed thanks to the data supplied by ECMWF.

In particular, weather data are retrieved from the numerical weather prediction model ICON run by the Israel Meteorological Service (IMS). This has been kindly shared by ECMWF in the framework of cooperation between PPRD East 3 and the South-East European Multi-Hazard Early Warning Advisory System project (SEE-MHEWS-A – https://www.see-mhews.org/) coordinated by the World Meteorological Organization (WMO).

The ICON model provides data with a horizontal resolution of 2.5 kilometers between grid points and a temporal resolution of 1 hour. They are then aggregated in time and space and classified in colour-coded categories. They thus show alert levels to convey hazard information that is immediately identifiable, recognisable and comprehensible (see the minimum

PPRD East 3

The PPRD East 3 programme, which is funded by the European Commission's Directorate-General for European Civil Protection and Humanitarian Aid Operations (DG ECHO) and coordinated by the Swedish Civil Contingencies Agency, was launched in October 2020. Its aim was to enhance the early warning systems and planning capacities of the civil protection authorities of five Eastern Neighbourhood Partnership Initiative (ENPI) East countries (Armenia, Azerbaijan, Moldova, Ukraine and Georgia). The overarching goal is to work in a multi-stakeholder framework on developing Early Warning to Early Action (EWEA) systems that ensure actions are taken promptly. This is important to reach the high operative standards required for participation in the EU Civil Protection Mechanism.

temperature forecast as an example).

Spatially aggregated values are represented at regional level, in Oblasts for Ukraine and districts for Moldova. They include:

- Biometric indices averaged over 24 hours
- Daily minimum and maximum temperature
- Daily maximum wind
- Rainfall accumulated over 24 hours.

Temperature, cumulative precipitation, and wind forecasts are then classified into four colour-coded categories, whereas biometric indices are classified into five (see the legend table). Snow cover and the probability of T<0 are neither spatially aggregated nor classified to preserve the granularity of the information in specific regions.
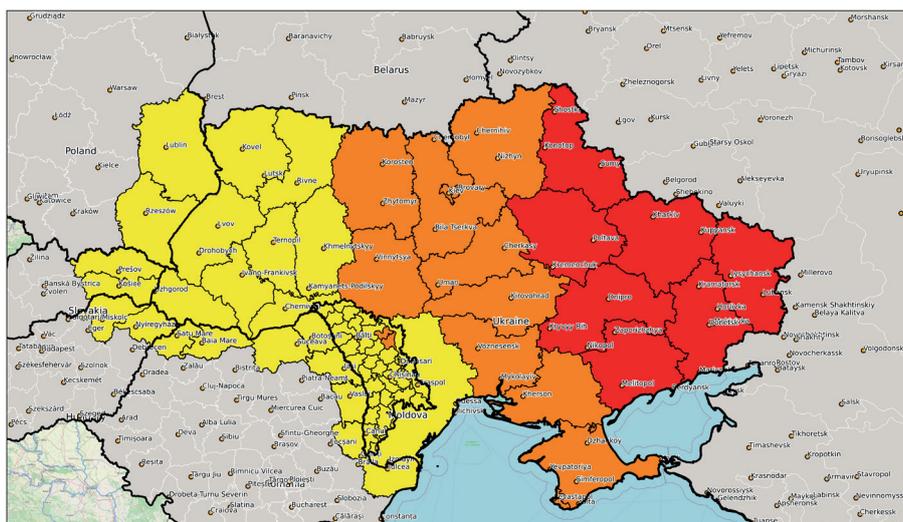
Temp Class	Min Temp (°C)
1	T > 5
2	0 < T < 5
3	-5 < T < 0
4	T < -5

PPRD East 3 minimum temperature forecast. This is an example of a PPRD East 3 forecast of minimum temperature on the following day, 24 February 2023.

Threshold-trigger mechanism for EWEA

Thresholds for each category have been defined in close collaboration with humanitarian responders on the ground (mostly with the Italian Red Cross), based on expert judgment and considering the severity of impacts on specific factors, such as roads, internally displaced persons, and population.

Weather conditions and hydrometeorological hazards, such as heavy snowfall, ice on roads, and flooding, can hinder access to certain areas and aid delivery. Furthermore, they can affect a highly vulnerable population that is often forced to move and may not have access to proper heating. For this reason, the potential impact on exposed persons is also represented through a synoptic visualisation of the population experiencing low temperatures.

To better meet the needs of teams operating on the ground, in areas where most humanitarian operations are taking place, forecasts are provided at a finer spatial scale. This enables a clearer understanding of what responders need to be aware of and how weather conditions may affect their intervention in the coming days. Data are organised at smaller administrative levels (corresponding to Rajons for Ukraine), which enables more precise identifications of alert levels for intense and localised weather events.

Since the bulletin is automated, there is no evaluation of the data by meteorologists, meaning that alert levels depend only on the weather model outputs. Therefore, they might differ from the official warnings issued by the Ukrainian and Moldovan NHMS. To complement the IBF information and to underline the 'one voice principle' in civil protection, daily communications from Ukrainian and Moldovan NHMS are reported in the bulletin. This is also to reaffirm that the IBF does not substitute the official forecasts, but it is meant to be an instrument to establish a proper EWEA threshold-trigger mechanism for humanitarian actors.

A dynamic tool

A close exchange of information with experts actively involved in managing the crisis in Ukraine has enabled ad hoc changes to the bulletin. The IBF is a dynamic tool that has varied over time, evolving according to the seasons (with the inclusion/exclusion of some variables) and following the changes of hazard conditions. Above all, the bulletin was initially based on open-access data. However, through careful coordination work between various initiatives and thanks to support from the WMO, ECMWF, and the IMS, the PPRD East 3 consortium was able to draw on higher-resolution meteorological models that were more suitable to the bulletin's purposes. Indeed, partnerships and collaboration between scientific and civil protection authorities have always been at the core of the IBF development. Such joint efforts of the PPRD East 3 consortium, DG ECHO, the WMO, ECMWF, the IMS, and the NHMS of Ukraine and Moldova have enabled the development of the IBF and its distribution among almost 18 different institutions.

Since the start of the activity, roughly 320 bulletins have been issued, making scientific expertise available to responders in the ongoing emergency. This has led, in turn, to the inclusion of civil protection strategies in humanitarian activities on the ground, enhancing EWEA processes and tools.

More information on PPRD East 3 activities related to Ukraine can be found at: https://www.pprdeast3.eu/news/impact-based-forecast/.

	Temperature		Rain	Wind	Biometric Indices	
	24 h Min	24 h Max	24 h cumulative height (mm)	24 h local maximum velocity (km/h)	Wind chill	Humidex
1	T > 5	T < 27	h < 5	v < 35	–27 < WC < 0	H > 27
2	0 < T < 5	27 < T < 32	5 < h < 30	35 < v < 45	–39 < WC < –27	27 < H < 30
3	–5 < T <0	32 < T < 36	30 < h < 60	45 < v < 62	–47 < WC < –39	30 < H < 40
4	T < –5	T > 36	h > 60	v > 62	–54 < WC < –47	40 < H < 55
5	-		-	-	WC < –55	H > 55

PPRD East 3 forecast legend. The table provides a summary of the colour coding in PPRD East 3 forecasts for Ukraine and Moldova for different parameters.

OpenIFS@home: Using land surface uncertainties and large ensembles for seasonal heatwave prediction

Jamie Towner, Sarah Sparrow, David Wallom, Andrew Bowery (all University of Oxford, UK),
Antje Weisheimer (University of Oxford, UK; ECMWF), Glenn Carver, Florian Pappenberger (both ECMWF),
Hannah Cloke (University of Reading, UK)

OpenIFS@home, a distributed forecasting system developed between the University of Oxford and ECMWF, provides the unique opportunity to run very large ensembles of OpenIFS Cycle 43r3 with specific model configurations from weather to seasonal timescales. Despite running at a relatively coarse resolution of about 110 km horizontally and 60 vertical levels, forecasts using land surface perturbations were comparable to the uncoupled version of ECMWF's seasonal forecasting system (SEAS5) in successfully predicting 2 m air temperature for the extreme June–July–August (JJA) 2021 summer heatwaves over regions of the North Pacific and the Mediterranean. SEAS5 is run in coupled mode operationally, but a version that is uncoupled from the ocean can be run for experimentation purposes.

Forecast set-up

Seasonal forecasts for JJA 2021 were initialised on 1 May 2021, with hindcasts going back each year to 1981 for the same forecast period. With the objective of allowing parameter perturbations to be activated within OpenIFS@home, five land surface variables, considered to be particularly sensitive in the literature, were perturbed: total soil depth excluding the top soil layer, saturated conductivity, Van-Genuchten alpha, soil moisture stress function, and minimum stomatal resistance. This was done through a set of percentages (–80%, –40%, 0%, 40%, 80%), where the perturbation percentage applies to the default parameter for the soil type at a particular grid point. With five parameters and five possible values, a total of 3,125 ensemble members is generated by applying all possible combinations of the five parameters. However, as OpenIFS@home uses a citizen-science approach which utilises volunteers' personal computers, some

model failures are expected. Accounting for these, 2,965 members were used for the 2021 perturbed runs, with an average of 907 members for each hindcast year. A control OpenIFS@home forecast without the perturbations and just 51 members was also run.

An exceptional season

The first figure shows the percentile rank of 2 m temperature for the JJA 2021 season according to ECMWF's ERA5 reanalysis in relation to the model climatology (1981–2021) for the North Pacific land region considered. The light pink shading dominates western North America, indicating that JJA 2021 was the hottest season within the analysed period. Indeed, the

Canadian observed temperature record was broken by 4.6°C, with a new record temperature of 49.6°C set in Lytton, British Colombia.

Forecast performance

The second figure shows a clear shift in probability towards higher temperatures in JJA 2021 for both the control and perturbed OpenIFS@home forecasts over both the North Pacific and Mediterranean land regions. This shift is more pronounced in the control, with a higher probability of forecasts indicating higher temperatures, whilst the distribution for the perturbed runs is more widespread, indicating a larger range of uncertainties and a long tail towards colder temperatures. Observing the final figure, the SEAS5
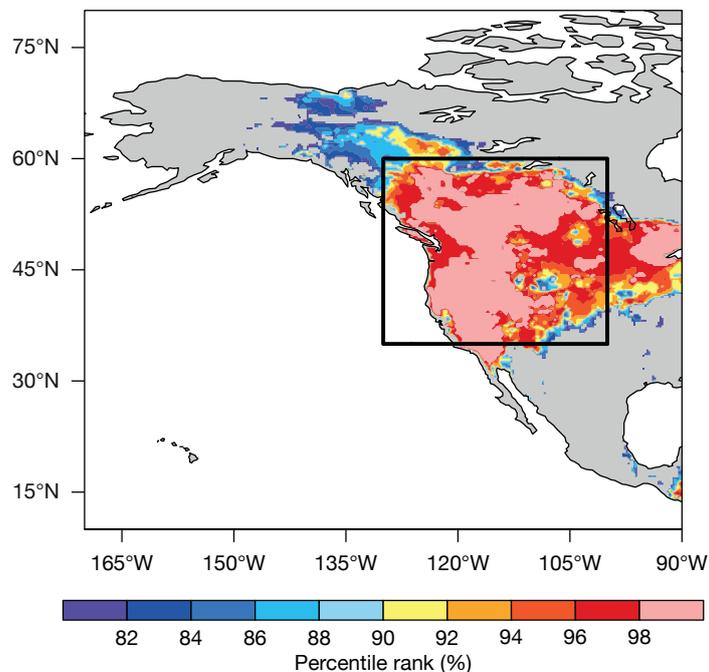
Percentile rank of JJA 2021 season mean. The chart shows the percentile rank of the JJA 2021 season mean of ERA5 in relation to the ERA5 1981–2021 climatology for 2 m air temperature. Light pink shading indicates that the 2021 season was the hottest within the climatological period analysed. The black box represents the region in which temperatures were particularly extreme and where the distribution of the control and perturbed forecast is shown in the next figure.

uncoupled, control and perturbed forecasts are comparable with between 60% and 70% of ensemble members predicting that temperatures for the JJA 2021 season will fall in the highest quintile of the associated model climate. This percentage is slightly higher in some regions of the control and SEAS5 uncoupled forecasts, highlighting the increased spread in the perturbed forecasts.

Next steps and recommendations

These results are the first of multiple upcoming forecasts where OpenIFS@home has been run on a seasonal timescale, and it is the first with land surface perturbations. Further analysis is required using the existing dataset produced from these simulations. The next development phase could attempt to increase the horizontal resolution of the model to about 80 km, with 90 vertical levels. Finally, data produced from OpenIFS@home could be used as input into hydrological and routing models to gain insight into the potential role that large ensembles could play in flood forecasting.
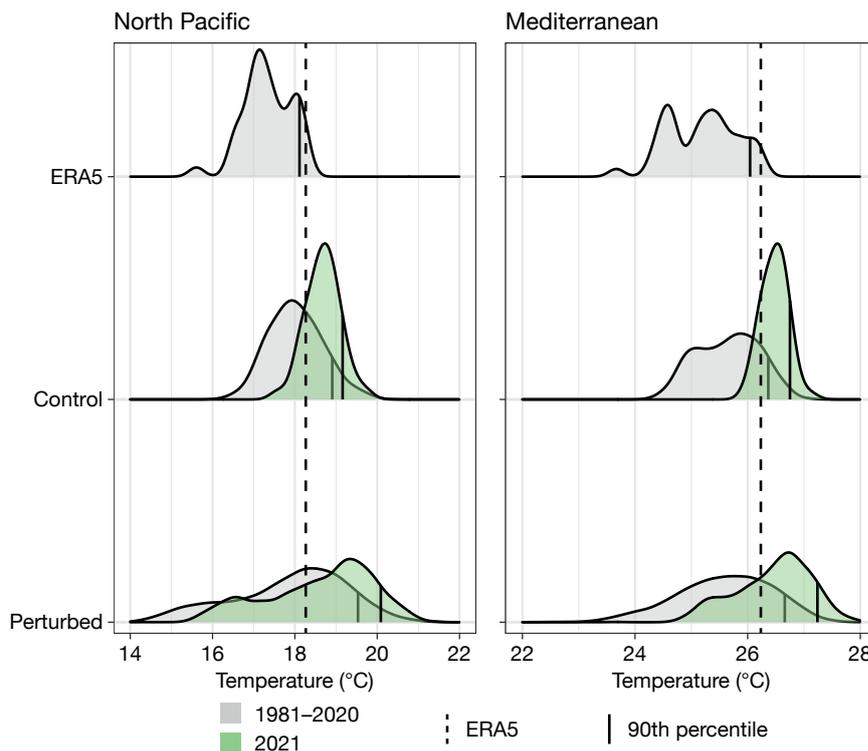
Probability density functions for JJA 2021. The charts show probability density functions for 2 m air temperature over the North Pacific (130°–100°W, 35°–60°N; see black box in previous plots) and the Mediterranean (10°W–40°E, 25°–50°N). The green curve represents the 2021 forecast distribution of all ensemble members whilst the grey curve indicates the climatology (1981–2020) for each dataset. Dashed and solid vertical black lines represent the observed temperature from ERA5 and the 90th percentile for each distribution, respectively.
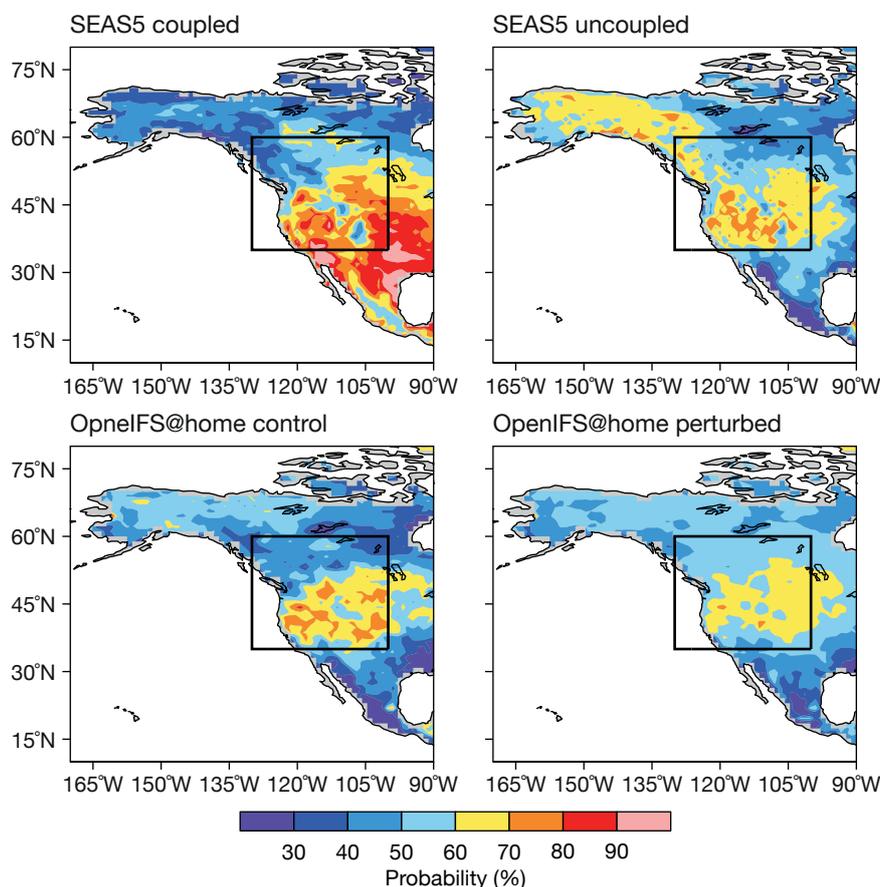
Probability of exceeding the upper quintile of climatology. The charts show the probability of the ensemble for JJA 2021 exceeding the 80th percentile of the model climate in SEAS5 coupled (51 members), SEAS5 uncoupled (51 members), the control forecast (51 members) and the perturbed forecast (2,965 members).

ERA5 reanalysis now available from 1940

Hans Hersbach

On 7 March 2023, ERA5 reanalysis data for 1940 to 1958 were made publicly available from the Climate Data Store (CDS) of the EU's Copernicus Climate Change Service (C3S) implemented by ECMWF. Authorised users can also access those years in ECMWF's Meteorological Archival and Retrieval System (MARS). This release extends the data record to over 83 years of hourly global snapshots for many quantities that describe the global atmosphere, land surface and ocean waves, from 1940 to the present. At a horizontal resolution of about 31 km, ERA5 provides detailed information on historical weather and climate. It has attracted a diverse and rapidly growing user base with over 100,000 registered users.

The additional 19 years of the new release were produced during 2022 in four parallel streams on ECMWF's high-performance computing facility (HPCF) in Reading, UK. They were completed just before the HPCF was decommissioned in October 2022. In addition, the suite that delivers timely updates was ported to ECMWF's new facilities in Bologna, Italy. It continues to add one day per day to the ERA5 record, five days behind real time. The ERA5-Land suite, which provides a downscaled land product at a resolution of 9 km, was also ported. A service of timely updates to ERA5-Land, similar to the updates to ERA5, was introduced on 16 December 2022.

ERA5 final release data were already available from 1959 onwards. The current further extension to 1940 replaces a preliminary extension back to 1950 that was characterised by sub-optimal tropical cyclone properties.

The role of observations

Like the analysis that is computed to initialise ECMWF operational forecasts, the ERA5 reanalysis blends observations with the ECMWF numerical weather prediction model into globally complete fields using the 4D-Var data assimilation method. The main differences are that ERA5 uses a frozen weather forecasting system throughout (as operational in 2016) for reasons of consistency, and it uses a lower resolution than the one used in operational weather forecasting to keep the system affordable. In contrast to longer centennial reanalyses, which ingest only surface observations, ERA5 makes use of all available observations, in-situ and from satellites. In addition,

ERA5 includes observations from the latest instruments as they are used in ECMWF's medium-range weather forecasting system, where possible.

The ability to reconstruct the actual historical weather depends on the availability of observations. In the past, the observing system was sparser than today. In addition, many observations are not yet available in digital form. In ERA5, no satellite data were used before 1970. The in-situ observing system was sparse over the southern hemisphere, and before the mid-1940s no upper-air observations were available. Nevertheless, from 1940 ERA5 should provide a good estimate of the actual synoptic situation for large regions over the northern hemisphere. An example (shown in the figure) is the representation of a severe storm over the Iberian Peninsula 82 years ago, which led to significant damage and disruption over Portugal and northwest Spain.

The spread in ERA5 ensembles provides a guide to the accuracy of such synoptic situations. It can also be used to know when and where ERA5 is dominated by observations, and where it is more controlled by the model. In the latter case, it should still be able to give a good description of low-frequency variability.

On the seasonal timescale, surface temperature for the ERA5 back extension is in good agreement with reconstructions that are solely based on observational sources on sub-daily to monthly aggregations. On the yearly scale, there is less agreement in global mean temperature for the early 1940s, while the lack of upper-air observations in these years also exposes a model cold bias in the lower stratosphere. Agreement is much better from the mid- to late-1940s onwards. Overall, the ERA5 back extension provides an excellent extension of the power of reanalysis to provide 'maps without gaps'.

The full ERA5 reanalysis can be accessed in the CDS: **https://cds. climate.copernicus.eu/**.

15 February 1941, 11 UTC   15 February 1941, 17 UTC   15 February 1941, 23 UTC

Hourly maximum 10 m gusts (m/s)
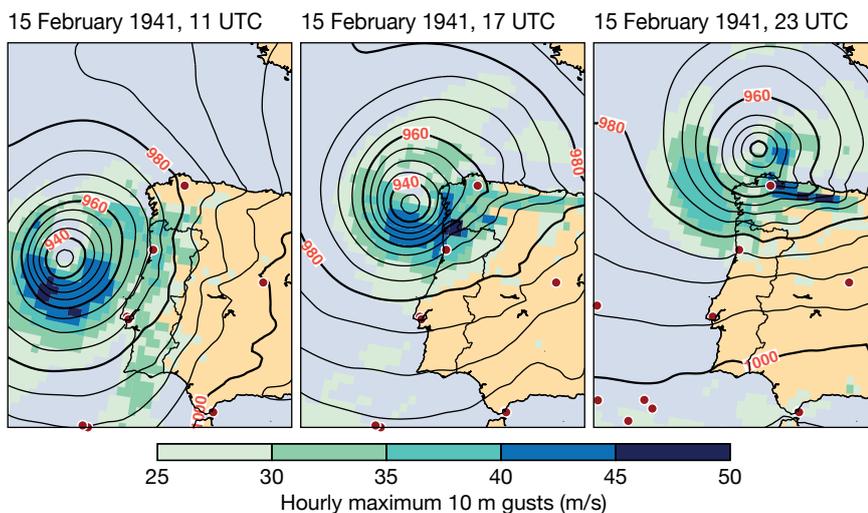
The Iberian Storm of 1941. Three snapshots of the Iberian Storm of 1941, showing ERA5 mean sea-level pressure (contours, in hPa) and hourly maximum 10-metre gusts (colours, in m/s). The locations of assimilated pressure and marine-wind observations from which these ERA5 reanalysis fields were constructed are shown as red dots.

ECMWF collaborates with Swiss partners on GPU porting of FVM dynamical core

Christian Kühnlein (ECMWF), Till Ehrengruber (CSCS), Stefano Ubbiali, Nicolai Krieger, Lukas Papritz, Alexandru Calotoiu, Heini Wernli (all ETH Zurich)

Numerical weather prediction models face an increasingly diverse landscape of supercomputing architectures. The ongoing, overarching trend is the shift from purely CPU-based platforms to heterogeneous systems with GPU accelerators. The resulting boost in computing cores and memory bandwidth offers significant potential for attaining higher numerical resolution and energy efficiency. However, challenges arise because model programming and implementation depends on the specific hardware, and efficient execution requires targeted optimisation. Serving various hardware inevitably involves more complex code that needs to be organised to maintain productivity.

To achieve this for ECMWF's non-hydrostatic FVM dynamical core, which was previously implemented in Fortran, we are rewriting and further developing FVM with the domain-specific library GT4Py of the Swiss National Supercomputing Centre (CSCS) and ETH Zurich. Both CSCS and ETH Zurich are also working closely with MeteoSwiss in a related project that is porting the ICON model to GPUs using GT4Py. ECMWF benefits greatly from the leading software development efforts by these partners and the PASC-(Platform for Advanced Scientific Computing)-funded KILOS project at ETH Zurich. At ECMWF, the GT4Py efforts occur in addition to the ongoing GPU adaptation of the operational Integrated Forecasting System (IFS) model.

GT4Py systematically separates the user code from hardware architecture dependent optimisations, instead of handling all in one source code, which is common in traditional designs. The user code contains the definitions of the computational stencils (e.g. discrete operators of the spatial discretisation, advection schemes, time integration) that are expressed using the embedded GT4Py domain-specific language (DSL). An optimising toolchain in GT4Py then transforms this high-level representation into a finely tuned implementation for the target hardware architecture (see the presentation of results in the first figure).

Productivity can be maintained with the GT4Py approach because the common high-level 'driver' interface is completely agnostic to target hardware, thus allowing support for optimisations and new architectures without changing the application. Furthermore, Python represents a popular and advanced programming language with concise syntax, comprehensive documentation, an extensive set of libraries (e.g. for unit testing, data analysis, machine learning, visualisation), a low barrier of entry for domain scientists and academia, and relatively straightforward prototyping. Python can also integrate seamlessly with lower-level languages such as C++ and Fortran, making it a good fit for DSLs.

Preliminary results

We present results from an intermediate stage of the FVM porting effort, using the first version of GT4Py restricted to structured grids. We consider the FVM dynamical core coupled to the IFS prognostic cloud microphysics scheme, implemented entirely in Python using GT4Py. For the performance measures shown in the first figure, we performed short runs of the moist baroclinic wave benchmark exemplified in the second figure.

Once ported to GT4Py, FVM can employ various user-selectable
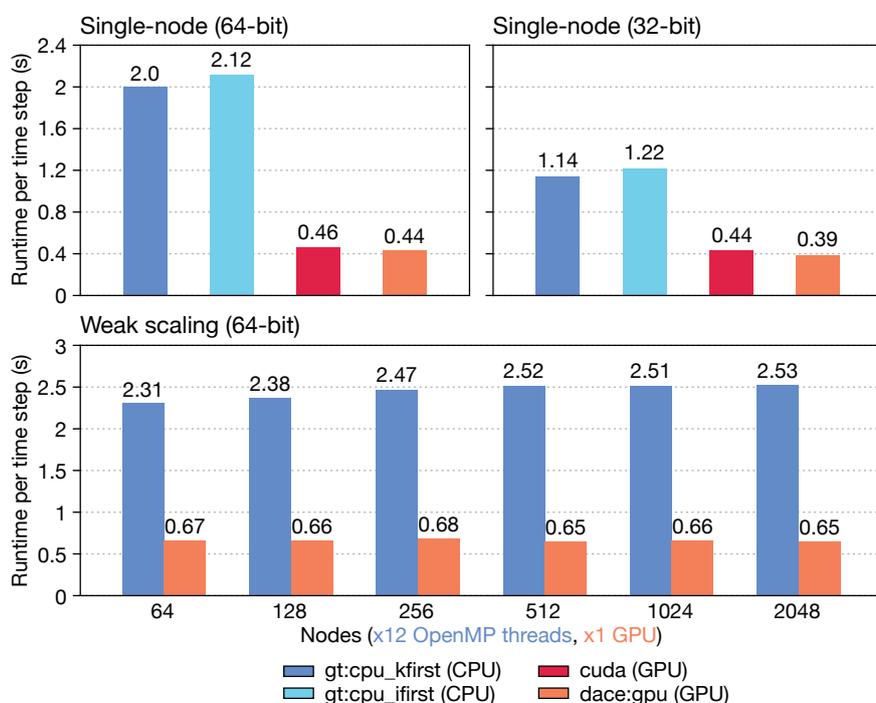
Selected performance measures. The performance measures were obtained on the hybrid partition (CPU: Intel Xeon E5-2690 v3; GPU: NVIDIA Tesla P100) of the Piz Daint supercomputer at CSCS. In the top two plots, the runtime per time step for various GT4Py backends using 64-bit and 32-bit floating point precision is shown for CPUs (gt:cpu_kfirst and gt:cpu_ifirst backends) and GPUs (cuda and dace:gpu backends). The bottom plot shows weak scaling, with the model executing at various model resolutions on either the CPUs or the GPUs, respectively, using the gt:cpu_kfirst or the dace:gpu backends.

backends for CPU and GPU hardware, using either 64-bit or 32-bit floating point precision. Two CPU C++ backends specialised for different array layouts were tested. The two GPU backends tested can execute with either native CUDA C++ or by leveraging the data-centric (DaCe) framework of ETH's Scalable Parallel Computing Laboratory. The first figure's bottom panel shows parallel scaling of the distributed model using the GHEX library (funded by the Partnership for Advanced Computing in Europe, PRACE) across multiple CPU or GPU nodes. Weak scaling is considered from about 14 km down to 1.7 km grid spacing and corresponding increases in compute nodes from 64 to 2,048, respectively (grid spacings are given for the equatorial region, finer spacings apply away from the equator due to the regular longitude-latitude grid that is employed temporarily). Near-optimal scaling (i.e. the same runtime) is achieved in the given range, somewhat better with GPUs. Development and study are continued on other platforms with different hardware.

Outlook

The preliminary results presented
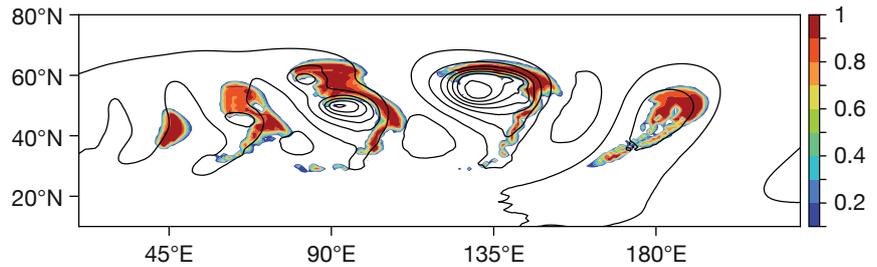
Moist baroclinic wave. Snapshot of a moist baroclinic wave at day ten using the FVM nearly global configuration (a regular grid defined in the latitude range +/–80°). Depicted are cloud fraction at about 2 km above the surface (shading) and surface pressure (contour lines with interval of 10 hPa).

indicate the potential of the GT4Py approach to achieve performance portability and productivity in the context of future numerical weather prediction models. This is also supported by the latest results with the ICON model (a first GT4Py-enabled dynamical core on GPUs already shows equal performance compared to the Fortran & OpenACC version) and an earlier GT4Py porting project for the US FV3 model at the Allen Institute for AI.

The porting and development of FVM is ongoing with the next version of GT4Py, which is equipped with a new declarative interface and supports unstructured meshes required for the global configuration with the quasi-uniform IFS octahedral grid. The declarative GT4Py introduces important new features for more modular and concise programming in the user code and enables GT4Py to optimise more freely and aggressively.

Updates with FVM based on the new declarative GT4Py are expected in 2023 and 2024. From mid-2023, ECMWF will host a computer scientist position for GT4Py, which is funded by the EU ESiWACE3 project.

ECMWF pioneers the use of remote observations in fire forecasting

Francesca Di Giuseppe

Recently ECMWF has been involved in a project funded by the European Space Agency (ESA) to explore the real-time use of remote observations to improve fire danger forecasts. As a result, a new fire danger index has been proposed to overcome one of the most important limitations of current fire danger metrics: the lack of connection with real-time information on fuel available for burning. The proposed fire occurrence probability index (FOPI) combines the most-used model of fire danger, the Canadian fire weather index (FWI), with remote observations of vegetation optical depth (VOD) as a proxy of fuel amount and moisture. The goal is to improve fire danger predictions in all fuel-limited environments where fire is driven by the short-term drying of intermittently available fuel.
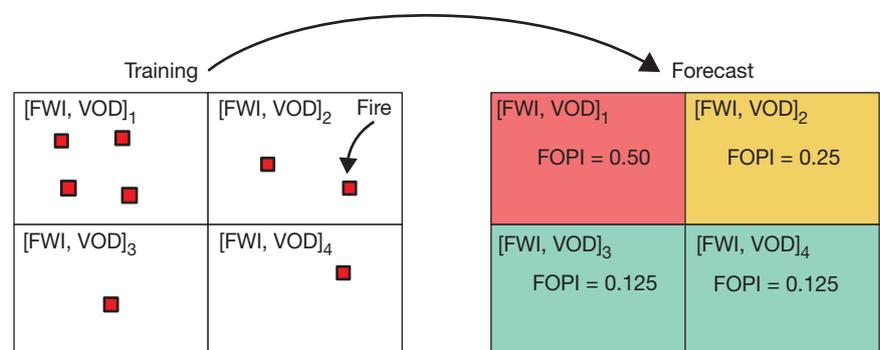
Schematic explaining the derivation of FOPI. FOPI is derived by identifying combinations of FWI and VOD values that have previously resulted in detected fire activity. The higher the fire activity detected, the more severe is the prediction of landscape flammability when those conditions occur. Here this concept is explained assuming a total activity of eight fires unevenly spread across four model grid boxes. The combinations of FWI and VOD of boxes 1, 2, 3 and 4 correspond to FOPI values of 0.5, 0.25, 0.125 and 0.125, respectively. Boxes 3 and 4 might have a different combination of FWI and VOD values but, as they experienced the same fire activity, they have the same flammability.
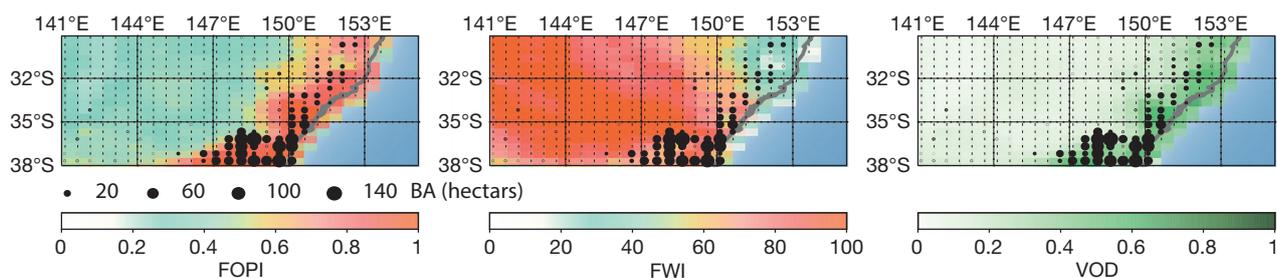
Forecasts for large fire events in 2020 in New South Wales, Australia. The maps show same-day forecasts for 4 January 2020. The leftmost panel shows the FOPI prediction for that day, the middle panel shows the FWI prediction, and the rightmost panel shows observed VOD for the same day. FOPI, FWI and VOD are all dimensionless. Recorded burnt areas for the same day are provided as black circles. The size of the symbols is proportional to the burnt area expressed in hectares.

FOPI considerably outperforms the FWI in all arid biomes and has comparable skill to the FWI where fuel is in abundance.

The fire occurrence probability index

To understand how the flammability of a particular area changes in response to weather conditions, and to assess the potential spread and intensity of a fire, fire danger indices such as the FWI are used. These indices express a measure of landscape flammability. Although they do not provide information on actual fires, as they do not consider ignition, they have been shown to correlate with fire activity expressed in terms of burnt area reasonably well.

Fire weather indices are calculated from daily temperature, relative humidity, wind speed and rainfall to simulate the moisture content of the fuel beds at different depths.

One of the most recognised limitations of fire danger indices is that they exclusively rely on weather inputs and disregard the availability of fuel. This leads to very unrealistic results when applied globally. Over deserts, for example, fire danger is usually very high because of the dry conditions, but fires are impeded by the lack of biomass to burn. Static maps are usually applied to mask out barren areas.

The idea behind the new FOPI index is to improve fire danger forecasts in fuel-limited environments by combining classic fire weather danger with information about fuel and its moisture content. This is done using remote observations of vegetation optical depth. FOPI 'weights' FWI predictions by the available fuel: it decreases fire danger where there is very little fuel to burn or where the landscape is too moist to sustain fires, while it retains high fire danger where fuel is abundant and dry.

FOPI is based on a training dataset collected in 2020, when fire activity in terms of burnt areas was referenced to FWI predictions and VOD observations. Using this dataset, FOPI is trained to predict combinations of fire weather conditions (FWI values) and vegetation status (VOD values) that have in the past resulted in detected fire activity (see the FOPI schematic).

The higher the fire activity detected in the past, the higher the FOPI value is for a given combination of FWI and VOD. By construction, FOPI is limited to values between 0 and 1.

FOPI performance in a real case

To provide an idea of how FOPI could help to localise the potential of critical fires in real-time fire monitoring, I analyse an event in 2020. The 2019/2020 Australian summer, since then named Black Summer, produced hundreds of fires, mainly in the southeast of the country. At its peak, air quality dropped to hazardous levels in all southern and eastern states and the smoke could be detected across the South Pacific Ocean to Chile and Argentina.

The second figure shows the calculated FOPI and FWI on a day when fires raged in the Australian state of New South Wales. The localisation of actual fires is performed through recorded burnt areas. A map of observed VOD is also provided to interpret the differences between FOPI and FWI. FOPI is better correlated with the fire activity than FWI. FOPI's more localised outcome stems from its capability of masking out areas of insufficient fuel load.

Most local authorities are informed of fuel conditions and would be able to exclude areas where fires are not likely to occur despite high FWI. However, the use of

FOPI is certainly a great improvement in global systems that aim to provide an overview of fire danger worldwide as it simultaneously takes into account the interannual variability of fuel and its moisture content. These systems are often used at face value without expert interpretation. FOPI can restrict areas in need of further monitoring.

Benefits and future work

The proposed FOPI overcomes part of the limitations that characterise FWI. There are two innovative aspects in FOPI. The first aspect is that, by combining FWI with remote observations of vegetation characteristics, it provides a framework to account for real-time fuel availability. As an immediate benefit, FOPI limits unrealistically high values registered in desert areas where fire activity is hindered. It also makes it possible to retain a memory of previous burning before vegetation recovery takes place. The second advantageous aspect of FOPI is that it expresses a probability of fire occurrence based on previous observations. By explicitly considering the history of burning for a given landscape, FOPI enables landscape susceptibility to be considered.

FOPI can be shown to outperform FWI in all fuel-limited ecosystems while still retaining comparable skill to FWI where fuel is in abundance. The FOPI index thus advances fire danger monitoring globally and to some extent it creates a real-time link with the available fuel in a fire danger forecast. The use of 10-day running mean VOD products implies that FOPI can be issued as a medium-range forecast (1 to 5 days in advance) by applying the same map. The usability of FOPI in extended-range forecasts will, however, require a predictive model for fuel available for burning to be developed, and this is the next step in this research project.

Four Early Career Fellows start projects at ECMWF

Florentine Weber, Katerina Anesiadou, Luise Schulte, Paolo Andreozzi

Between mid-January and the beginning of April, Paolo Andreozzi, Katerina Anesiadou, Luise Schulte and Florentine Weber started at the ECMWF site in Bonn, Germany. As the first cohort of the German Meteorological Service's Early Career Fellowship Programme, they are initially working at ECMWF for a period of two years. The programme also involves local research institutes, such as the Center for Earth System Observations and Computational Analysis (CESOC), the Universities of Cologne (UoC) and Bonn (UB) and the Forschungszentrum Jülich (FZJ). To support talented research graduates towards careers in international research organisations, ECMWF has carefully selected current research topics, while the German Meteorological Service (DWD) has assisted in pushing the boundaries of science and improving forecasting by coordinating an appropriate programme.

CESOC, with founding directors Susanne Crewell (UoC), Jürgen Kusche (UB) and the late Astrid Kiendler-Scharr (FZJ), is a network for scientific research cooperation and knowledge transfer across disciplines in the area around Bonn and beyond. Building bridges between activities in CESOC and ECMWF will provide ECMWF with links to local research infrastructures and large collaborative research projects, for example on atmospheric and surface processes relevant to climate in the Arctic.

The vacancies for the first cohort of fellows attracted applications from over 100 young scientists from ECMWF Member and Co-operating States. The competitive selection process resulted in four fellowships:

- **Modelling water in Arctic clouds** – Supercooled liquid-containing clouds are crucial for the radiation balance at the surface and thus for the Arctic climate system, but their representation is challenging for numerical weather prediction models, leading to systematic errors in the Arctic and beyond. Luise Schulte will use observational data from the one-year ice drift campaign MOSAiC (2019 to 2020) for verification and

The four Early Career Fellows. Katerina Anesiadou, Luise Schulte, Florentine Weber and Paolo Andreozzi (left to right) met each other for the first time at the DWD headquarters in Offenbach, Germany.

informed model development of Arctic clouds. Her work at ECMWF is supported by Linus Magnusson, Jonathan Day, Richard Forbes (all ECMWF) and Susanne Crewell (UoC). Luise obtained her Master's degree in physics at the University of Frankfurt, where she worked on a case study of Arctic mixed-phase clouds using the ICON model (DWD).

- **Modelling the regional water cycle and temperature in urbanised areas** – By 2050, more than two-thirds of the world's population will live in urban areas, the UN predicts. Due to the increasing importance of cities, their adequate representation in weather forecasts is crucial. Comparing observational data with weather predictions, Florentine Weber will investigate the impact of cities on the hydrological cycle at high resolution. This interdisciplinary research is supported by Gianpaolo Balsamo and Linus Magnusson (both ECMWF) and carried out in collaboration with CESOC. As a physicist, Florentine recently completed her PhD at the University of Sheffield, UK, discovering why the atmosphere over land has become drier.

- **Use of altimeters in a coupled data assimilation system** – Altimeters are currently used in ocean and wave data assimilation, providing sea-surface height and ocean wave height information. Additional information these satellite instruments could provide regarding

the atmospheric state is currently treated as 'corrections' and removed in the pre-processing of the observations. Katerina Anesiadou will investigate the potential of extracting new atmospheric information from altimeter measurements by incorporating them in a coupled data assimilation system, using the ground-based GNSS Zenith Total Delay assimilation methodology. Her research work is supported by Sean Healy and Patricia de Rosnay (both ECMWF) in cooperation with Jürgen Kusche (Institute of Geodesy and Geoinformation, UB). Katerina is a physicist with an MSc in Environmental Physics (University of Bremen, Germany) and has recently completed a fellowship about nowcasting of extreme weather events (University of Padua, Italy).

- **Meteorological effects of atmospheric composition** – Aerosols and their composition play an important role for climate by changing radiation and initiating cloud formation in addition to their impact on air quality. Aiming to improve the operational representation of prognostic aerosols at sensible computational costs, a suitable configuration of ECMWF's Integrated Forecasting System including atmospheric composition will be used to understand the effects of day-to-day variations of aerosol species on relevant forecast fields. Paolo Andreozzi will explore this project under the supervision of Robin Hogan and Richard Forbes (both

ECMWF) in collaboration with Birger Bohn (FZJ) and Ulrich Löhnert (UoC). Such research will benefit from the collaboration with the CESOC research partners on the chemistry and observations of aerosols. Paolo recently graduated at the University of Hamburg with an

MSc thesis on the properties of Kelvin wave errors in ECMWF deterministic forecasts.

At the end of January, the Fellowship programme started with a kick-off meeting at DWD in Offenbach. Further events and workshops as well as

training and mentoring aim to provide an outstanding professional development experience.

For details about the Fellowship programme, visit: https://www.dwd. de/DE/derdwd/arbeitgeber/ einsteigen/fellowship/fellowship.html.

Exploring alternatives to radiance assimilation for hyperspectral infrared sensors

Kirsti Salonen (ECMWF), Thomas August, Tim Hultberg (both EUMETSAT),
Angela Benedetti, Anthony McNally (both ECMWF)

We are soon entering the era of next-generation hyperspectral infrared (IR) instruments, such as the Infrared Atmospheric Sounding Interferometer – New Generation (IASI-NG) and the Meteosat Third Generation Infrared Sounder (MTG-IRS). These instruments will give us more detailed measurements of the state of the atmosphere spatially and temporally, but they will also significantly increase the data amounts to be handled. Radiance assimilation is still the preferred approach to constrain numerical weather prediction (NWP) model analysis, but at the same time there is also increased interest in alternative ways to use satellite observations. This includes direct assimilation of retrieved atmospheric parameters. Here we focus on recent

joint investigations between EUMETSAT and ECMWF to assess the potential of directly assimilating retrieved temperature and humidity from IASI in ECMWF's Integrated Forecasting System (IFS). While ECMWF is not currently planning to assimilate the retrievals in operations, the results are very encouraging, with particular added value for humidity forecasts. They also provide interesting insights for ECMWF's Member and Co-operating States.

All-sky temperature and humidity retrievals

The piecewise linear regression (PWLR) all-sky machine learning retrieval approach has been the cornerstone in EUMETSAT's operational IASI retrieval

processing since 2010 and forms the baseline for future missions. For this study, we have used retrievals exploiting IASI without its microwave sounder companions as a proxy for MTG-IRS. The training set consists of about 68 million IASI spectra from May 2019 to April 2020, paired with collocated temperature and humidity profiles from operational ECMWF analyses and forecasts.

The PWLR retrieval method is based on linear regression. In order to capture the non-linear relationship between the inputs and the outputs better, the input space is divided into several classes. For each of these classes, a separate set of linear regression coefficients is computed from the corresponding
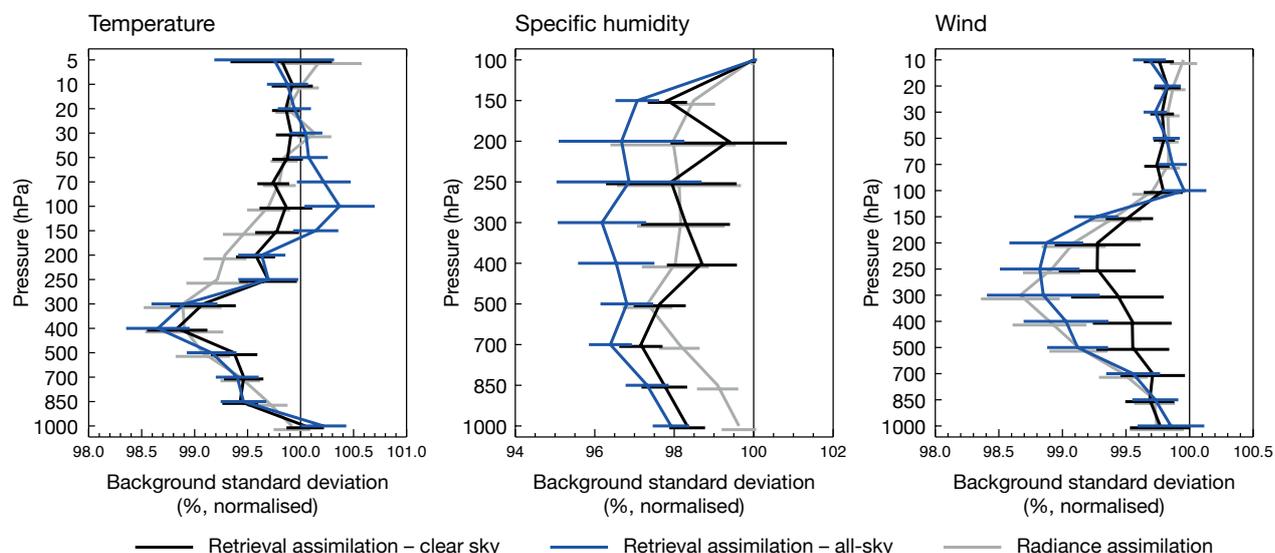
Retrieval assimilation vs radiance assimilation. The charts show normalised observation-minus-background standard deviation for radiosonde temperature, specific humidity and wind if retrieval or radiance assimilation is included in the background forecast. In the case of wind, pilot balloon, aircraft and wind profiler observations are also considered. The period ranges from 1 December 2019 to 28 February 2020.

subset of the training set, such that overall a piecewise linear function from the input space into the output space is obtained.

The retrievals are provided as principal component (PC) scores together with scene-dependent observation operators. These operators guarantee that only the vertical structures which the sounders can resolve are conveyed into the model and that finer model structures are preserved. Use of the scene-dependent observation operators has been the key in achieving a positive impact from the retrieval assimilation.

Impact assessment in the ECMWF system

The impact of temperature and humidity retrievals from Metop-C IASI has been tested in IFS Cycle 48r1. Depleted observing system experiments are a widely used method to emphasise the impact originating from new data. The focus in the assimilation experiments has been on retrievals over sea, as they have high and homogeneous quality.

The baseline experiment uses conventional and AMSU-A observations. In the test experiments, either IASI radiances or retrievals are assimilated on top of the baseline experiment. IASI radiances are currently used from channels diagnosed cloud free. This includes completely clear situations and channels which are diagnosed clear above a cloudy scene. The retrieval assimilation has been tested in completely clear situations and in all-sky scenes, including both clear and cloud-affected areas, using different thresholds for the cloud information provided with the data.

The impact on temperature, humidity and wind forecasts is clearly positive in all depleted system assimilation experiments. The impact on short-range forecasts is shown in the figure in terms of normalised change in background departure standard deviations for radiosonde observations. In clear-sky scenes (black line), the retrieval assimilation brings very similar information into the system to that of the radiance assimilation (grey line). For temperature and wind the impact from retrieval assimilation is somewhat smaller in magnitude than from the radiance assimilation in the upper atmosphere. For humidity, the impact of retrievals is slightly larger in the lowest 3 km (up to about 700 hPa). When the retrievals are assimilated in all-sky scenes (blue line) the magnitude of the impact becomes comparable for temperature and winds. For humidity, the retrieval assimilation even outperforms the radiance assimilation impact. However, for temperature forecasts some degradation is seen

above 150 hPa in the all-sky experiment. This is the first time that such a strong positive impact is seen from a retrieval assimilation.

The experiments have been repeated in the full observing system. As expected, the signal is weaker than in depleted observing system experiments. The results still indicate positive impact on humidity, while the impact on temperature is mainly neutral when the retrievals are used in clear scenes. In all-sky experimentation, the temperature forecasts are currently degraded, which would require further investigations and the fine-tuning of the quality control and observation errors used.

Concluding remarks

The results in the depleted observing system are very positive and robust and the all-sky experimentation is indicating comparable or even better impact than radiance assimilation, especially for humidity. This is probably because of using more humidity-sensitive channels, and overall more observations, in all-sky retrieval assimilation than in the radiance assimilation. In full observing system experimentation, the main benefits are currently coming from the humidity retrievals, and further work to consolidate the results would be needed.

New observations January – March 2023

The following new observations have been activated in the operational ECMWF assimilation system during January – March 2023.

Observations	Main impact	Activation date
Atmospheric Motion Vectors from GOES-18 (replacing GOES-17)	Tropospheric wind	12 January 2023
Clear Sky Radiances from GOES-18 (replacing GOES-17)	Tropospheric humidity and wind	12 January 2023
Additional SYNOP observations from Hungary	Local near-surface fields	19 January 2023
AOD from S-NPP and NOAA-20 VIIRS	Aerosol in COMPO suite	1 February 2023
Nadir O_3 profiles from NOAA-20 OMPS	Ozone in COMPO suite	1 February 2023
Radiances from MWHS-2 on FY-3E	Tropospheric humidity and wind	22 February 2023
Additional SYNOP observations from Greece	Local near-surface fields	6 March 2023
Atmospheric Motion Vectors from Meteosat-10 (replacing Meteosat-11)	Tropospheric wind	23 March 2023
Additional radio occultation bending angles from Spire	Temperature and winds in upper troposphere/lower stratosphere	28 March 2023
All-Sky Radiances from Meteosat-10 (replacing Meteosat-11)	Tropospheric humidity and wind	29 March 2023

Initial assimilation of hyperspectral sounding data from 1970 in preparation for ERA6

Bill Bell, Paul Poli, Dinand Schepers, Hans Hersbach (all ECMWF), Andrzej Klonecki (Spascia)

During the 1960s and 70s, NASA and the US National Oceanic and Atmospheric Administration (NOAA) flew several pioneering satellite missions. They set out to test the technologies which would later become the bedrock of the modern meteorological satellite observing system, delivering key observational data for global numerical weather prediction (NWP) systems. These technologies included infrared and microwave sounding instruments, which provide information on upper air temperature and humidity.

The EU-funded Copernicus Climate Change Service (C3S), implemented by ECMWF, has supported the assessment of many of these early satellite data records for assimilation in the forthcoming ERA6 reanalysis as well as other future reanalyses, and for extending the production of Essential Climate Variables (ECVs) as far back as possible. These activities build on data rescue programmes initiated at NASA and NOAA, which recovered the raw data and collated the original documentation. The C3S activities aim to improve the assimilation readiness of the data, by developing and testing the radiative transfer models required to assimilate the data and assessing the bias characteristics of the data.

Much of this work is being carried out in close collaboration with C3S partners EUMETSAT and Spascia. The reprocessed datasets include a range of early infrared imagers carried on the Nimbus series of satellites starting in 1964, as well as several sounding instruments flown as part of NASA, NOAA and US defence programme missions.

The IRIS instrument

During the first phase of the C3S programme, an initial project focused on early infrared sounders, including the Infrared Interferometer Spectrometer (IRIS) hyperspectral instrument operated on Nimbus-4 from April 1970 to January 1971. In common with all pre-1979 satellite missions, IRIS was relatively short-lived and yet provided
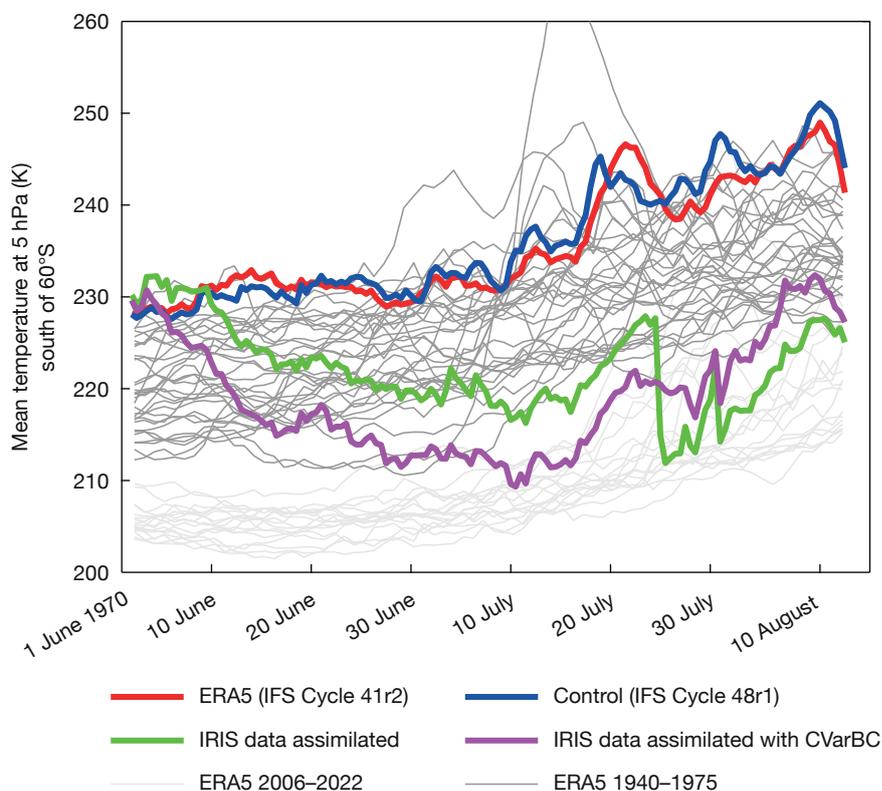
The effect of including IRIS in data assimilation. The light grey lines show the evolution of analysed mean temperatures south of 60°S at 5 hPa from 1 June to 12 August between 2006 and 2022 in ERA5, when GNSS-RO observations provide an accurate observational constraint on stratospheric temperatures. In the early period of ERA5 (1940–1975, dark grey), there were fewer observations of upper-air temperatures and the analyses at this level are biased warm, by 10–20 K. The ERA5 analysis for 1970 is shown in red, and a control experiment based on IFS Cycle 48r1 is shown in blue. The analysis resulting from the assimilation of IRIS data is shown in green, and the result of applying CVarBC is shown in magenta. The assimilation of IRIS data with CVarBC brings the analysis to temperatures more consistent with the climatology of ERA5 during recent years.

valuable and unique measurements in previously unobserved regions, including for example the polar stratosphere. Notably, IRIS pre-dates the first interferometer to be deployed for operational NWP (the Infrared Atmospheric Sounding Interferometer, IASI) by 36 years. The data from IRIS had been the subject of earlier investigations at ECMWF, which showed the data to be of some promise for assimilation. IRIS observations cover the spectral range 400–1,600 cm^{-1} at a spectral resolution of 2.5 cm^{-1}. Recently IRIS data have been introduced into ECMWF's Integrated Forecasting System (IFS) in advance of the start of

the next C3S global reanalysis, ERA6, due to start in mid-2024.

In advance of the assimilation tests, scientists at Spascia confirmed that our current cloud detection methods work well for IRIS, that the data exhibits biases that are manageable using variational bias correction (VarBC), and that the data quality is generally good. Furthermore, through a detailed analysis of independent data from a co-hosted multi-spectral instrument on Nimbus-4 as well as rocket-sonde data from the time, Spascia identified significant biases in the southern winter polar stratosphere in ERA5.

IRIS data was assimilated to assess the impact on analyses and forecasts as well as the effect of the data on the winter polar biases. Despite the relative sparsity of the data, the impacts from these initial tests were significantly positive. For example, southern hemisphere short-range forecasts of surface pressure, upper-air temperatures and winds are improved by 5–7%.

Furthermore, the impact of the IRIS observations on polar stratospheric

biases is dramatic (see the figure). A significant warm bias at 5 hPa in ERA5 during the southern polar winter of 1970 is corrected by the IRIS data over a period of six weeks. The application of constrained VarBC (CVarBC) further improves the impact of the IRIS data by limiting the amount of bias absorbed by VarBC and allowing the analysis state to cool to a more realistic state more rapidly.

For ERA6, model bias correction methods are currently under

development to help address such stratospheric biases. The findings shown here illustrate how the IRIS dataset will be instrumental to assessing the performance of these novel methods during the ERA6 early period. Furthermore, the methodologies developed for this assessment will be applied to similar, high-quality reference data, such as that from the Spektrometer Interferometer (SI-1) flown on Soviet satellites, to be delivered by EUMETSAT in 2023.

An urban scheme for the IFS

Joe McNorton, Gianpaolo Balsamo

The global population reached eight billion in 2022, with about 55% residing in urban areas. The urban environment is not currently considered in ECMWF weather and seasonal predictions or in global reanalyses. A simple, yet effective, representation of sub-grid and resolved urban elements affecting local weather and climate is being implemented in the Integrated Forecasting System (IFS). The scheme will become operational in IFS Cycle 49r1 and future Copernicus reanalyses data products, such as ERA6.

In cities, temperatures are often elevated due to an effect known as the urban heat island (UHI). This further intensifies the impact of extreme heat events on the local population and ecosystems. The atmospheric impact of urbanization is multi-faceted and includes direct contributions to the UHI effect from anthropogenic heating and cooling. Buildings alter the surface morphology, whilst urban materials change the albedo, emissivity, and both thermal and hydrological properties of the surface. Other climatic impacts come from anthropogenic activities, including irrigation, drainage,

snow-clearing and trace gas emissions. These components are known to impact, amongst other things, temperature, wind fields, surface heat and moisture fluxes, the boundary layer height and precipitation.

The urban scheme

Historically, cities have been difficult to resolve in global weather modelling, owing to the coarse spatial resolution of those models and the lack of reliable datasets to specify the spatial distribution of urban areas. With recent and foreseen improvements in the IFS resolution and global satellite-based
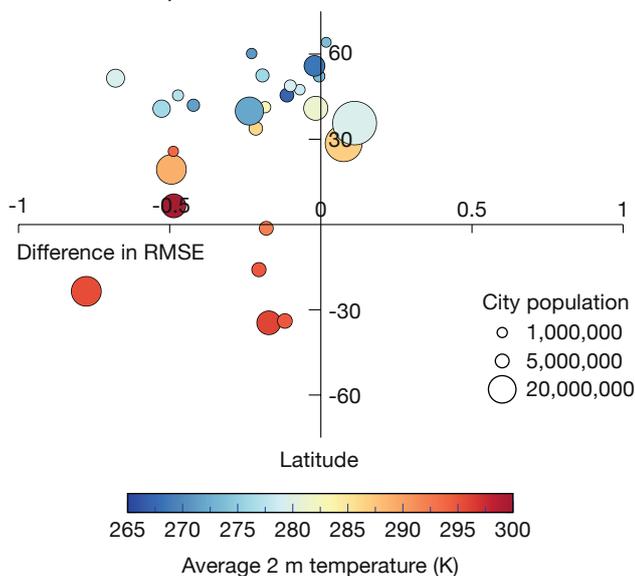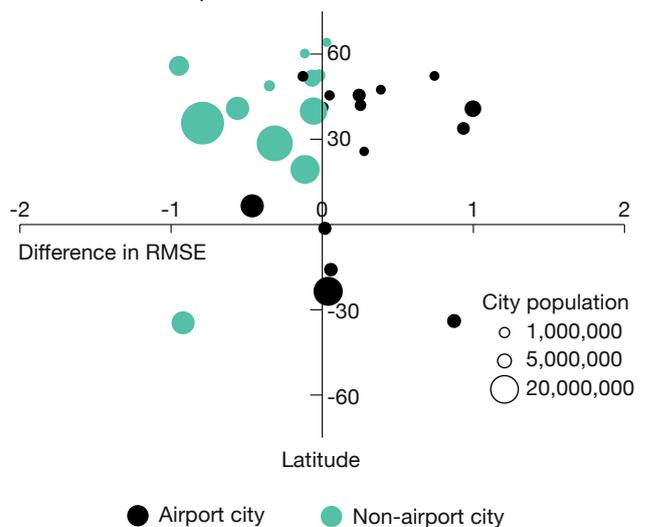
Temperature and wind speed errors. The average difference in root-mean-square error (RMSE), for various forecast times between 12 hours and 10 days, between the urban IFS and the operational IFS when compared with observed 2 m temperature (left) and 10 m wind (right) from 27 urban sites for northern hemisphere winter, 2018. Negative x-axis values indicate model improvement when using the urban scheme.
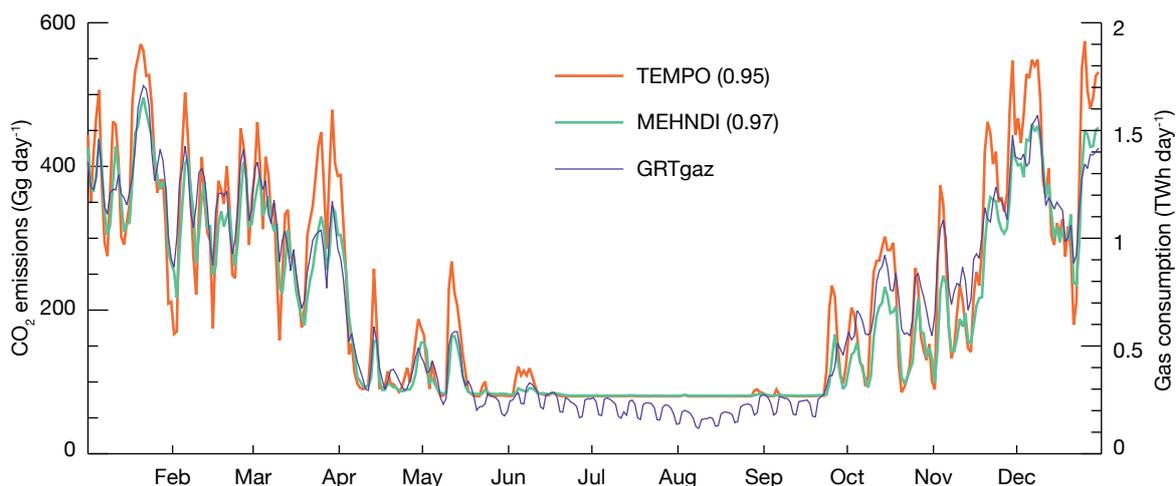
CO$_2$ heating emissions. Daily CO$_2$ heating emissions in 2020 from TEMPO (left axis, orange) and MEHNDI (left axis, green) as well as domestic natural gas consumption data from GRTgaz (right axis, purple) from France. The Pearson correlation coefficient of modelled emissions against gas consumption is provided in brackets in the legend.

datasets, the main weather effects specific to urban areas can now be represented. A simple scheme which considers cities as an interface between the sub-surface soil and the atmosphere has been designed and implemented. This includes an additional urban tile, sub-divided into urban canyons with road bases and roof fractions bordering them on either side. Overall, the scheme aims to improve the energy and moisture exchange between the surface and the atmosphere to better estimate near-surface weather conditions across all forecast times.

Local weather impact

ECMWF typically verifies near-surface model performance using screen-level temperatures and 10 m wind speeds measured by SYNOP weather stations. When averaged over more than 6,000 sites, the urban scheme outperformed the control IFS model for both near-surface temperatures and wind speed at all forecast times. There was a model improvement of about 0.5% for temperature and about 0.8% for wind speed up to ten days ahead averaged across all sites, most of which are in rural environments. When we focus on 27 densely populated urban sites, the improvement further increases to about 10% for both summer and winter months (see the temperature and wind speed error figure). It is well known that the UHI effect is greatest at night-time, and the model reflects this.

As detailed mapping of urban environments is not available at a global scale, we make several broad assumptions in the model. An example of this is an assumed universal average building height of eight metres and a fixed unitary building-to-road ratio. The shortcomings of this approach are evident at sites that include airports, where the model over-estimates building fraction and, as a result, reduces wind speed too much (see the wind speed error graph). These assumptions can be relaxed in future work, when we plan to test more detailed mapping information that recently became available, such as spatially varying building heights.

Environmental impact

The introduction of an urban scheme is of particular interest also for monitoring greenhouse gas emissions. More specifically, ECMWF is helping to build a prototype system for a European Monitoring and Verification Support (MVS) capacity for anthropogenic CO$_2$ emissions as part of the EU-funded CoCO$_2$ project. Such a system relies on accurate high-resolution initial estimates of emissions. Currently the EU-funded Copernicus Atmosphere Monitoring Service (CAMS) implemented by ECMWF provides predicted atmospheric CO$_2$ concentrations based on monthly maps of emissions from different sectors (energy production, manufacturing, settlements, and transport are among the four dominant sources). We utilise the urban map introduced combined with predicted temperature to estimate daily emissions of CO$_2$ from residential heating in urban settings in real time. We validated our approach for Modelling Emissions from Heating in

Near-real-time Driven by the IFS (MEHNDI system) against observed gas consumption from France and compared it with a state-of-the art emissions model, the TEMPO system, currently used in CAMS (see the second figure). We found that MEHNDI provides an accurate temporal and spatial variability of residential heating emissions, which can be used to accurately report emissions as required by the European MVS system.

Atmospheric concentrations of CO$_2$ from IFS simulations driven by both MEHNDI and CAMS emissions were evaluated with total column CO$_2$ observations from the global Total Column Carbon Observing Network. We found that, over or near urban sites, MEHNDI improved model performance and therefore it is scheduled for implementation in CAMS. This work highlights the benefits of an improved surface representation for weather and environmental applications. Future research is planned on other trace gases and aerosols as well as on the simulation of anthropogenic heat sources.

Further details are described in the recent publications McNorton et al., 2021, https://doi.org/10.1029/2020MS002375, and 2023, https://doi.org/10.1029/2022MS003286.

Linearised physics: the heart of ECMWF's 4D-Var

Marta Janisková, Philippe Lopez

In 1997, the incremental four-dimensional variational (4D-Var) data assimilation system became operational at ECMWF. 4D-Var is the system used at ECMWF to combine a short-range forecast with observations to produce the best possible estimate of the current state of the atmosphere. This then provides the initial conditions for a new forecast. The system works by minimising a cost function, which measures the departure of the analysis from observations and the short-range forecast. In this minimisation, linearised versions of the forecast model are used to efficiently determine the optimal 3D representation of the atmospheric state at a given time. From the very beginning, it was recognised that it is crucial to represent physical processes in the linearised model. Parametrization schemes used in linearised models have gradually become more and more complex. Nowadays, comprehensive schemes are included in ECMWF's 4D-Var system. Recent data assimilation tests using the latest version of the linearised physics have again demonstrated a systematic and significant improvement for all parameters, levels, and regions.

Principles

At ECMWF, the 4D-Var analysis system is employed to generate 3D atmospheric states (or analyses) which, together with separately produced land and ocean conditions, can be then used to initialise the Earth system model. This system uses the model dynamics and physics in an optimisation procedure, which aims to improve the overall fit of initial conditions to the observations available inside the assimilation time window, while staying as close as possible to the model first guess (a short-range forecast which forms the starting point in creating the analysis) in poorly observed regions. For more details see Box A.

4D-Var relies on linearised versions of the forecast model, namely its tangent linear and adjoint versions (see Box B), to describe the time evolution of small perturbations around the reference state, typically in terms of temperature, wind, humidity and surface pressure.

In the 1980s, when experimentation and developments towards 4D-Var data assimilation were under way, only the adiabatic version (i.e. dynamics only) of the linearised

models was used. However, the importance of representing physical processes was soon recognised. It started to become obvious that the mismatch between the model analysis and observations can remain large during the 4D-Var minimisation when only the adiabatic linearised model is used to evolve the model state from the beginning of the assimilation window to the time of each observation. Besides, without the inclusion of physics in variational data assimilation, the assimilation of many satellite observations, such as all-sky radiances, precipitation and cloud-affected measurements, would simply be impossible.

ECMWF linearised physics

Although already started in the mid-1990s, the development of the linearised physics (LP) really gathered pace at the turn of the century, with the operational implementation of 4D-Var.

At ECMWF, parametrization schemes for linearised models started from very simple ones. They aimed to remove very large increments, especially those produced close to the surface by the adiabatic model (Buizza, 1994). More complex but still incomplete schemes were then developed by Mahfouf (1999). Over the following years, more comprehensive schemes were gradually implemented. This led to a set of schemes able to describe most physical processes and their interactions with almost as much detail as in the non-linear model, but with just slight simplifications and/or regularisations (Janisková et al., 2002; Tompkins & Janisková, 2004; Lopez & Moreau, 2005).

The current set of physical parametrizations used in ECMWF's linearised model (called simplified or linearised parametrizations hereafter) comprises six different schemes: radiation, vertical diffusion, orographic gravity wave drag, moist convection, large-scale condensation/precipitation and non-orographic gravity wave activity. The schemes are sequentially called in this order (Janisková & Lopez, 2013). This LP package is quite sophisticated and is believed to be the most comprehensive one currently in use for operational global data assimilation.

Developing and testing the linearised model

The construction of the linearised model starts with the design of a set of suitable non-linear physical

a   4D-Var methodology
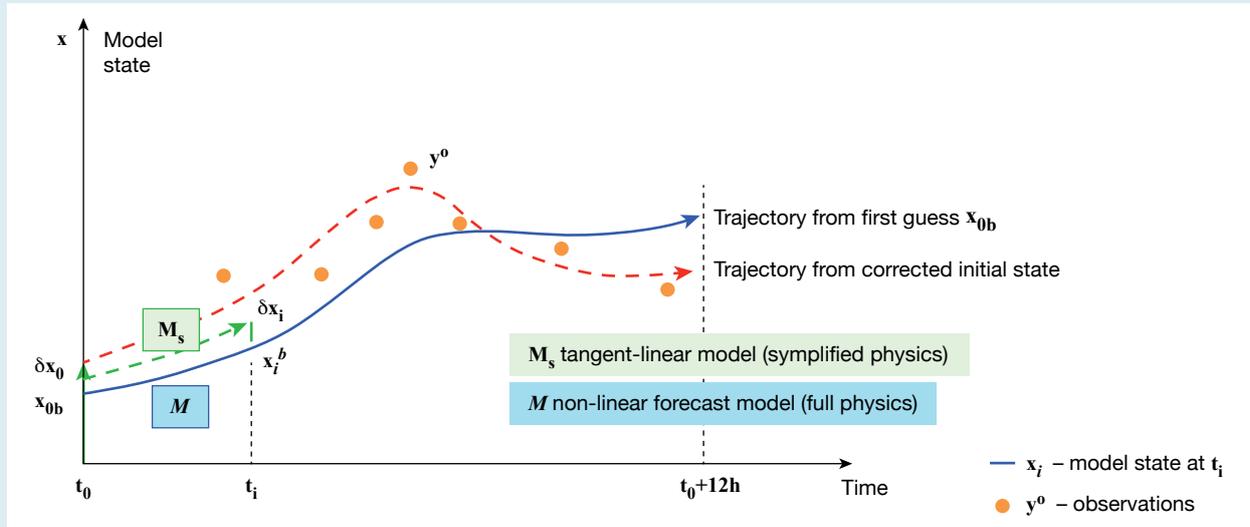
Schematic description of 4D-Var.

The goal of 4D-Var is to define the initial atmospheric state such that its distance to the original model trajectory on the one hand and to the observations on the other is as small as possible over the assimilation time window (see the figure in this box).

At ECMWF, an incremental approach is used as a cost-effective formulation of 4D-Var (Courtier et al., 1994). Using the incremental formulation, the model state at any time is defined in terms of an increment, i.e. a small perturbation around the background state (typically, a previous short-range forecast). To the first order, the 4D-Var problem can be solved by finding the analysis increment δx_0 at initial time t_0 which minimises the following cost function:

$$J(\delta \mathbf{x}_0) = \frac{1}{2} \delta \mathbf{x}_0^T \mathbf{B}^{-1} \delta \mathbf{x}_0 + \frac{1}{2} \sum_{i=0}^{n} (\mathbf{H}_i(\delta \mathbf{x}_i) - \mathbf{d}_i)^T \mathbf{R}_i^{-1} (\mathbf{H}_i(\delta \mathbf{x}_i) - \mathbf{d}_i) \tag{1}$$

At any time t_i, $\delta \mathbf{x}_i = \mathbf{x}_i - \mathbf{x}_i^b$ is the analysis increment representing the departure of the model state ($\mathbf{x}$) with respect to the background ($\mathbf{x}^b$), which consists of temperature, humidity, vorticity, divergence and surface pressure in the current 4D-Var system. $\mathbf{d}_i = \mathbf{y}_i^o - H_i(\mathbf{x}_i^b)$ is the so-called innovation vector, which provides the departure of the model background equivalent ($H_i(\mathbf{x}_i^b)$) from the observation ($\mathbf{y}_i^o$). $\mathbf{H}_i$ is the linearised observation operator (also describing the spatial interpolations to the observation locations and the forecast model integration propagating the initial state $\mathbf{x}_0$ to the time of observation). $\mathbf{R}_i$ is the observation-error covariance matrix (including measurement and representativeness errors) and $\mathbf{B}$ is the background-error covariance matrix of the state $\mathbf{x}^b$.

The minimisation requires an estimate of the gradient of the cost function:

$$\nabla_{\delta \mathbf{x}_0} J = \mathbf{B}^{-1} \delta \mathbf{x}_0 + \frac{1}{2} \sum_{i=0}^{n} \mathbf{M}^T (t_i, t_0) \mathbf{H}_i^T \mathbf{R}_i^{-1} (\mathbf{H}_i(\delta \mathbf{x}_i) - \mathbf{d}_i) = 0 \tag{2}$$

where $\mathbf{M}^T$ is the adjoint of the forecast model M and $\mathbf{H}^T$ is the adjoint of the observation operator H, as explained in Box B.

The incremental approach reduces the computational cost of 4D-Var since the perturbations $\delta \mathbf{x}_i$ and the gradient of the cost function are computed at a lower resolution using the simplified model during successive minimisations (currently four at ECMWF). After each minimisation, the model trajectory is recomputed at high resolution.

In operational practice, innovation vectors are computed with the nonlinear model at high resolution including full physics. Increments are computed with the tangent-linear model at low resolution including simplified physics. The gradient of the cost function is computed with the low-resolution adjoint model using simplified physics.

b   Tangent-linear (TL) and adjoint (AD) models

If the model M describes the time evolution of the model state $\mathbf{x}$ from time t_i to time t_{i+1} as:

$$\mathbf{x}(t_{i+1}) = M[\mathbf{x}(t_i)] \qquad (3)$$

the time evolution of a small perturbation $\delta\mathbf{x}$ can be estimated to the first order approximation by the tangent-linear version $\mathbf{M}$ of the non-linear model M:

$$\delta\mathbf{x}(t_{i+1}) = \mathbf{M}[\mathbf{x}(t_i)]\, \delta\mathbf{x}(t_i)$$

$$\delta\mathbf{x}(t_{i+1}) = \frac{\partial M[\mathbf{x}(t_i)]}{\partial\mathbf{x}}\, \delta\mathbf{x}(t_i) \qquad (4)$$

The adjoint of the linear operator M is the linear operator $\mathbf{M}^*$ such that:

$$\forall \mathbf{x}, \forall \mathbf{y} \qquad \langle \mathbf{Mx}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{M}^*\mathbf{y} \rangle \qquad (5)$$

where $\langle\,,\,\rangle$ denotes the inner product and $\mathbf{x}, \mathbf{y}$ are randomly chosen input vectors. The adjoint operator, for the simplest canonical scalar product $\langle\,,\,\rangle$ in Eq. (5), is actually the transpose of the tangent linear operator, $\mathbf{M}^T$.

Variational assimilation is based on the adjoint technique, since the adjoint model $\mathbf{M}^T$ can provide the gradient of any objective (cost) function, J, with respect to $\mathbf{x}(t_i)$ from the gradient of the objective (cost) function with respect to $\mathbf{x}(t_{i+1})$:

$$\frac{\partial J}{\partial\mathbf{x}(t_i)} = \mathbf{M}^T\left(\frac{\partial J}{\partial\mathbf{x}(t_{i+1})}\right) \qquad (6)$$

The integration of the AD forecast model works backward in time and all adjoint equations in the code are in reverse order compared to the tangent-linear equations.

---

parametrizations, from which the tangent-linear and adjoint versions can be derived. Thus, the whole development is rather demanding since it requires a good knowledge of both physical parametrizations and the adjoint. The validation must be very thorough, and it must be performed for the non-linear, tangent-linear and adjoint versions of the physical parametrization schemes.

Building the linearised model: a tricky compromise

In the real world, physical processes can be highly non-linear (NL) and often discontinuous. Since 4D-Var relies on the linearity assumption, it is necessary to either discard processes which could lead to instabilities, or to apply some regularisation to smooth the parametrizations' discontinuities and render the schemes as differentiable as possible. Another constraint when developing parametrization schemes for data assimilation is related to the minimisation of the 4D-Var cost function, which is solved with an iterative and therefore computationally demanding algorithm. Because of the constraints related to linearity and computational cost, the full NL model needs to be simplified before its successful inclusion in 4D-Var computations. However, at the same time, parametrization schemes used in the linearised model must maintain a realistic representation of physical processes and remain as close as possible to the reference full NL model.

Once the simplified code is developed, it is necessary to tune and validate it. This is done by evaluating short forecasts (typically 12 hours; the length of the 4D-Var time window) to ensure that the new simplified scheme is stable and does not depart too much from its full NL counterpart over the targeted integration period.

Tangent-linear approximation and regularisation

The tangent-linear (TL) model is used in 4D-Var to evolve increments of the model state in time, from the beginning of the assimilation window to the time of the observations.

To build the tangent-linear model, a linearisation is performed with respect to the local tangent of the model trajectory. For the validation of the TL approximation, the accuracy of the linearisation is evaluated using pairs of two non-linear 12-hour forecasts. First the difference between two non-linear integrations (one starting from a background field and the other one starting from the corresponding analysis) is computed with the full NL model taken as a reference. Then a corresponding TL integration calculates the time evolution of the analysis increments (the difference between the analysis and the background field), with the trajectory being taken from the background field. The TL integration can then be compared with the NL difference to assess how well the TL model approximates its full NL counterpart. For a quantitative evaluation of the impact of linearised schemes, mean absolute errors between TL and NL integrations are computed. The mean absolute error for the TL model without physics (i.e. the adiabatic TL model) is usually taken as a reference in these comparisons.

In some situations in which strong nonlinearities are present, especially when physical processes are included, special care must be taken to avoid the
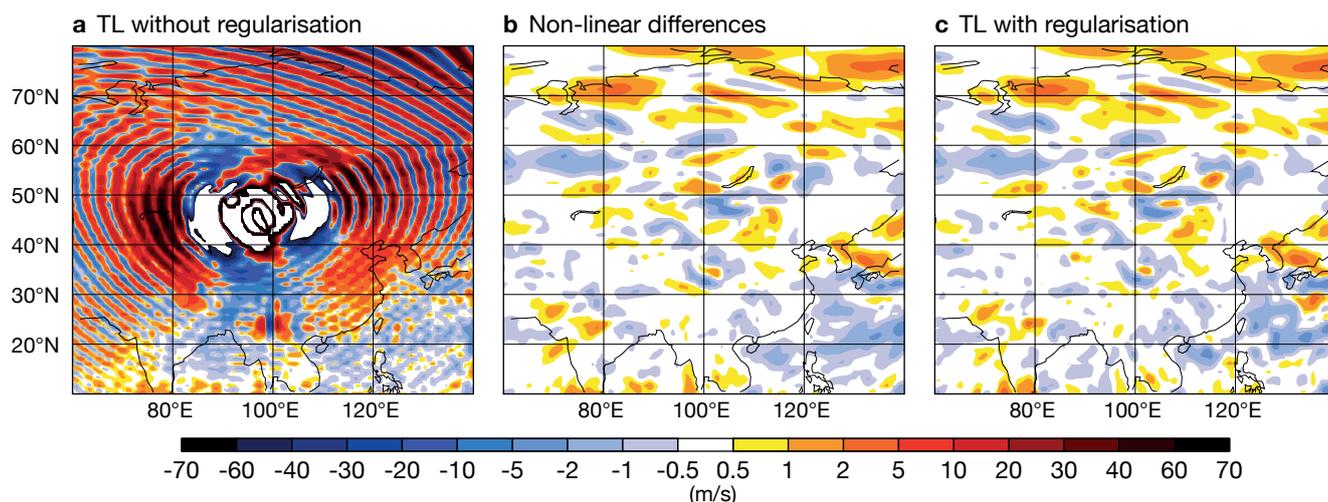
FIGURE 1 Zonal wind increments around 850 hPa after 12-hour evolution: (a) the tangent-linear (TL) model without regularisation in the orographic gravity wave drag scheme, (b) non-linear differences, and (c) the TL model with applied regularisation in the orographic gravity wave drag.

spurious development of perturbations in the linearised model. Without a proper treatment of the most significant nonlinearities and thresholds, the TL model can quickly become too inaccurate to be useful. Figure 1 illustrates some erroneous behaviour of the TL model due to too large local derivatives in the linearised model. When using the TL model without any regularisation in the orographic gravity wave drag scheme, strong spurious noise develops in the TL model (Figure 1a) compared to the differences between two NL integrations (Figure 1b). When regularisation is applied (consisting here in retaining the low-level blocking, while significantly reducing contributions from all other sub-grid scale orographic effects), TL increments (Figure 1c) agree well with the NL differences.

The impact of the LP on the TL approximation is illustrated in Figure 2. It shows the zonal mean cross-section of the change in TL error for temperature when the full LP is included in the TL model relative to the adiabatic TL model. The TL error is significantly reduced almost everywhere when physics is included in the linearised model. The largest improvement is observed in the lower troposphere, where many physical processes are active. Similarly, when looking at mean vertical profiles of the change in the TL error (Figure 3), there is a clear improvement for all variables (temperature, zonal and meridional wind, and humidity) and throughout the vertical.

As illustrated, the TL model describing the evolution of perturbations, with simplified physical parametrizations included, generally fits the corresponding NL differences much better than the adiabatic TL model, provided discontinuities in the different physical parametrizations are properly smoothed. However, a regular effort is needed to ensure that the linearised model remains stable

and noise-free in all meteorological situations, and after any modifications to the full NL model or to the TL model itself, or any change in the computing environment.

Adjoint test

In variational data assimilation, the adjoint version of the TL model (which includes physical parametrizations at ECMWF) is used to estimate the gradient of the cost function to be minimised. The correctness of the adjoint must always be thoroughly checked by testing the identity described in Equation 5 (Box B). In this test, the TL and AD codes must agree to the level of machine
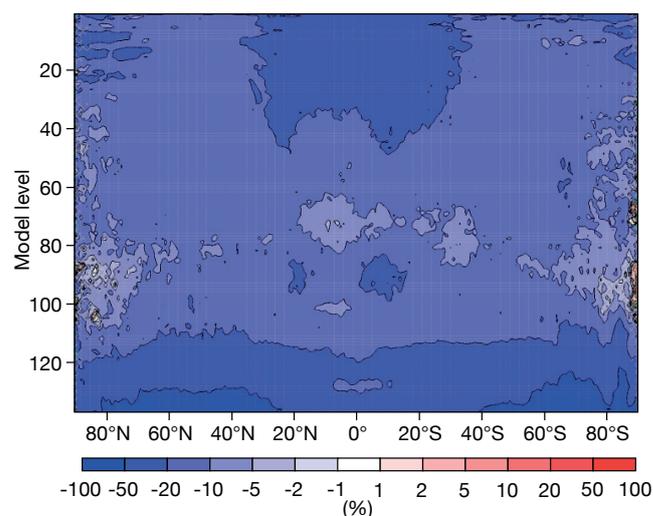
FIGURE 2 Zonal (latitudinal) mean impact of ECMWF's operational linearised physics on the TL approximation error for temperature. Blue colours show a relative improvement of the TL approximation due to the inclusion of the full linearised physics in the linearised model, compared to the purely adiabatic TL model. Red colours show a degradation. The evaluation is based on a set of 20 runs at 50 km resolution and is valid after 12 hours of integration.
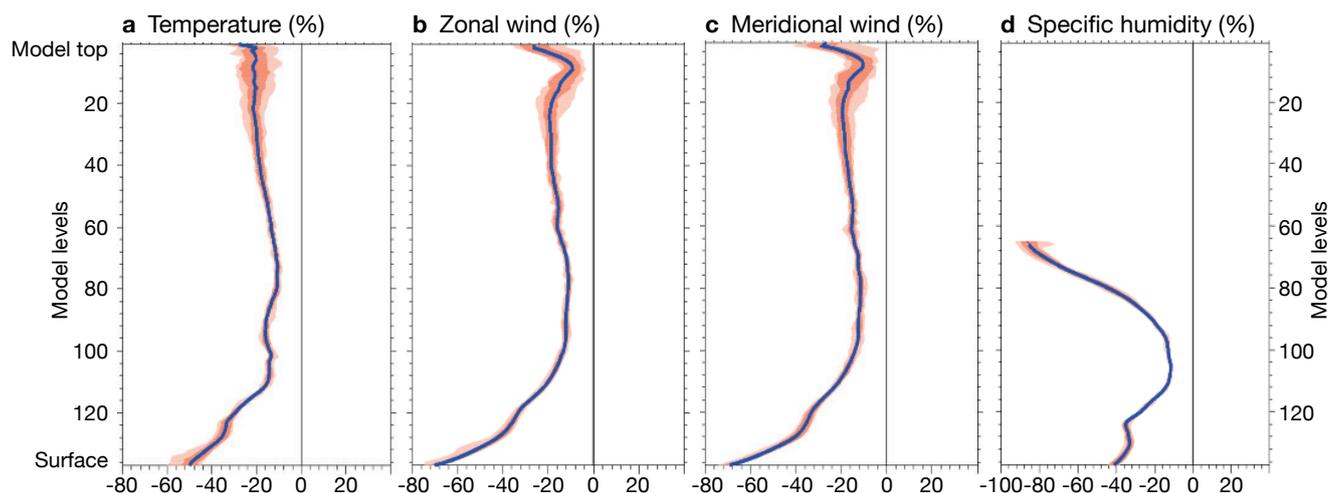
FIGURE 3 Mean vertical profiles of the change in global TL error (in percent) when full linearised physics is included in the TL model. Negative values indicate a relative error reduction for the TL model with physical parametrizations included compared to the adiabatic TL model. The variables shown are (a) temperature, (b) zonal (latitudinal) wind, (c) meridional (longitudinal) wind and (d) specific humidity. The evaluation is based on an ensemble of 20 runs at 50 km resolution and is valid after 12 hours of integration. The spread among the 20 runs is shaded in red.

precision, as illustrated in Figure 4. This test must be satisfied for global 3D atmospheric states and for integrations up to 12 hours (i.e. the maximum length of the 4D-Var window at ECMWF). Achieving a correct adjoint test can be really time-consuming, as even the tiniest error in the code will make the test fail. It should be emphasised that the correctness of the adjoint test does not guarantee the correctness of the tangent-linear code. It only indicates that the adjoint is correct with

respect to the TL code from which it was derived.

Impact of the linearised physics on 4D-Var analysis and forecast

The incremental impact of any change of the model's physics on both TL approximation and on the performance of ECMWF's Integrated Forecasting System (IFS) is regularly evaluated for every new model cycle. However, it has been a decade since the last time the impact on forecast skill from including the full linearised physics package in the 4D-Var minimisation was assessed (Janisková & Lopez, 2013). Since then, several updates have been made to the linearised model to follow the changes of the full NL model. For instance, the cloud and convective schemes were significantly modified, and a simplified linearised parametrization of surface processes was implemented. Other adjustments were also required to deal with additional problems due to non-linearities and discontinuities in the parametrization schemes, which resulted from the increase in the horizontal and vertical resolutions of the model.

To evaluate the impact of the latest whole linearised physics package, three experiments have been performed for the period of July–September 2021, using IFS Cycle 47r3 at resolution TCo639 L137,

Number of matching digits

	1	2	3	4	5	6	7	8	9	10
20	12	12	12	12	13	13	13	13	13	13
19	14	14	13	13	14	13	13	14	15	13
18	11	11	13	13	13	13	14	14	12	14
17	11	11	13	14	13	13	14	13	13	13
16	13	13	13	14	12	12	13	15	14	13
15	13	13	14	14	14	13	14	12	14	13
14	12	12	14	14	14	13	14	13	13	12
13	13	13	14	13	13	14	13	13	13	14
12	14	14	12	13	15	14	13	14	13	13
11	13	13	14	13	13	14	13	13	14	13
10	13	13	13	11	14	14	14	13	14	13
9	14	14	14	13	13	14	14	13	14	14
8	12	12	13	12	13	13	13	14	13	13
7	12	12	13	11	13	14	13	14	13	12
6	13	13	12	12	13	15	14	13	14	12
5	13	13	13	12	12	13	12	13	13	13
4	13	13	14	12	13	13	13	14	13	12
3	13	13	14	13	14	13	13	13	13	14
2	13	13	13	13	13	13	13	13	13	13
1	12	12	13	14	14	13	14	12	13	13

Physics configurations

FIGURE 4 Evaluation of the correctness of the adjoint code for ten different physics configurations (i.e. activating or not activating the different physical parametrization schemes). Here, the number of matching digits is computed from an ensemble of 20 members run at TL399 and L137 resolution, and at the end of a 12-hour integration, using the next model cycle, 48r1. Ideally, the number of matching digits should be at least ten (green and blue boxes). The two leftmost columns correspond to configurations with full linearised physics on.
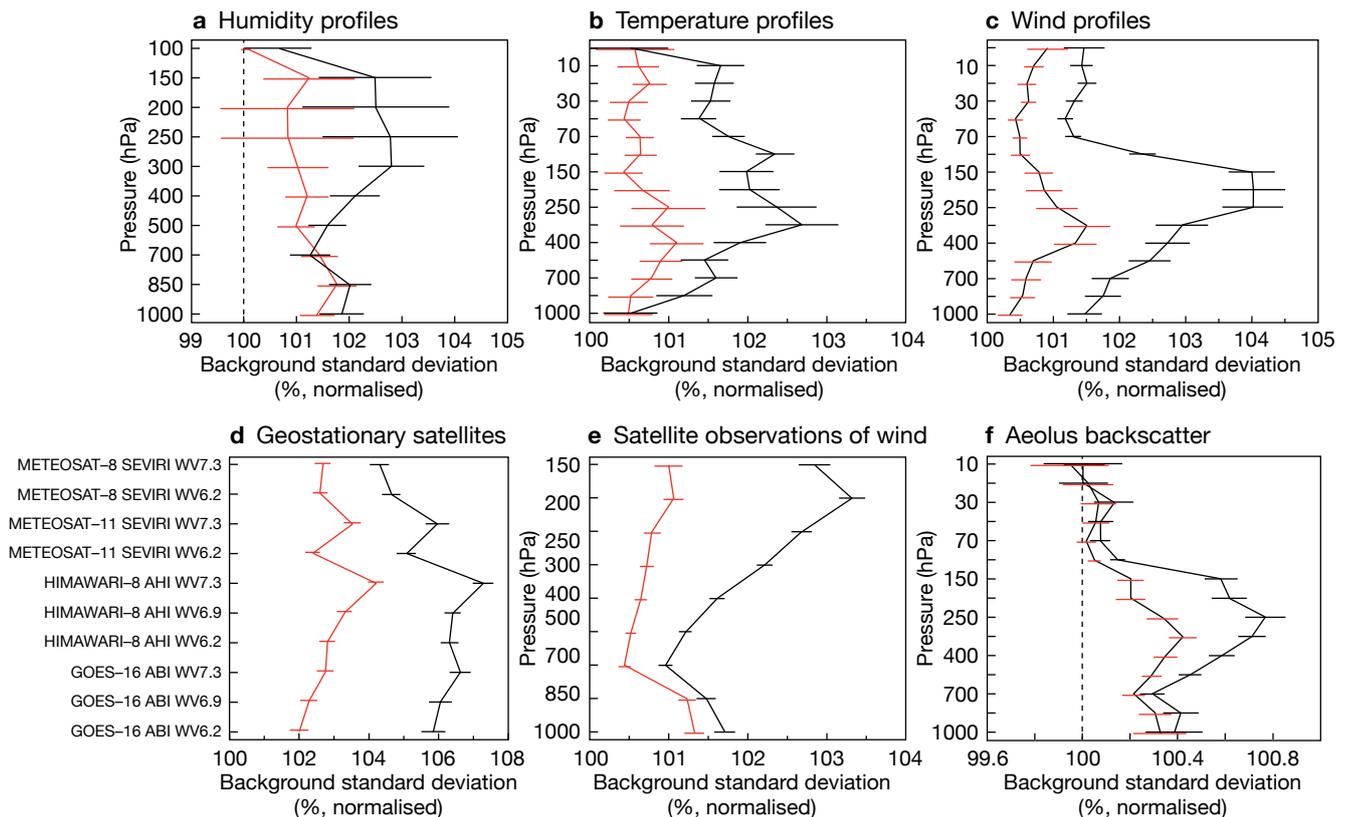
FIGURE 5 Difference between experimental runs and the reference run in terms of the standard deviation normalised by the standard deviation of the reference, for the first guess fits to different observations. The top three charts are for conventional observations: (a) humidity profiles, (b) temperature profiles from radiosondes, and (c) wind profiles from radiosondes, aircraft and wind profilers. The bottom three charts are for satellite observations: (d) geostationary satellite water vapour channel brightness temperatures at different wavelengths (WV in μm) from METEOSAT-8 and 11, HIMAWARI-8 and GOES-16, using the SEVIRI, AHI and ABI instruments, (e) satellite observations (SATOB) of wind and (f) Aeolus backscatter. The reference run is represented by the 100% vertical line and positive values correspond to a worse fit to assimilated observations in the runs without linearised physics. Horizontal bars indicate statistical significance at the 95% level. The results are shown for experiments excluding all observations requiring input from the linearised physics (in red) and additionally excluding the whole linearised physics (in black). Statistics are for the whole globe and over the period of July to September 2021.

corresponding to a horizontal resolution of approximately 18 km and 137 model levels in the vertical. 4D-Var had inner loops at resolution T159, T191 and T255, corresponding approximately to 130 km, 100 km and 80 km respectively, and L137. The first experiment included the linearised physics and all observations as used in operations (reference); the second one excluded all observations (mostly precipitation and cloud-affected) which could not be used without the linearised physics in 4D-Var; the third one also excluded the whole linearised physics.

The impact of including physical processes in the linearised model on 4D-Var analyses was assessed by comparing the first guess fit to assimilated observations. The results are shown in Figure 5 both for conventional observations (humidity and temperature profiles as well as wind observations) and for selected satellite observations (satellite observations of wind and Aeolus satellite backscatter). They clearly indicate that the inclusion of LP in data assimilation provides significantly better 4D-Var analyses.

Furthermore, the impact on forecast scores coming from including the linearised physics in the 4D-Var system is illustrated in Figure 6. The systematic and significant improvement for all parameters, levels and regions is clear. As expected, the impact of the LP in 4D-Var is largest close to analysis time. The positive impact is highly remarkable, especially in the tropics. Overall, these results demonstrate the vital role of the LP in the 4D-Var system, not only through its own huge positive impact, but also because it allows the assimilation of a whole range of observations which strongly rely on information provided by physical parametrizations, such as cloud, precipitations and surface parameters.

Conclusion

Including physical parametrization schemes in the linearised model clearly brings a large positive and significant impact in the data assimilation system at ECMWF. The inclusion of LP in 4D-Var is essential not only to improve the realism of 4D-Var increments, but also to assimilate observations that are highly sensitive to physical processes (e.g. cloud,

a 500 hPa geopotential

b 850 hPa vector wind

c 1,000 hPa temperature
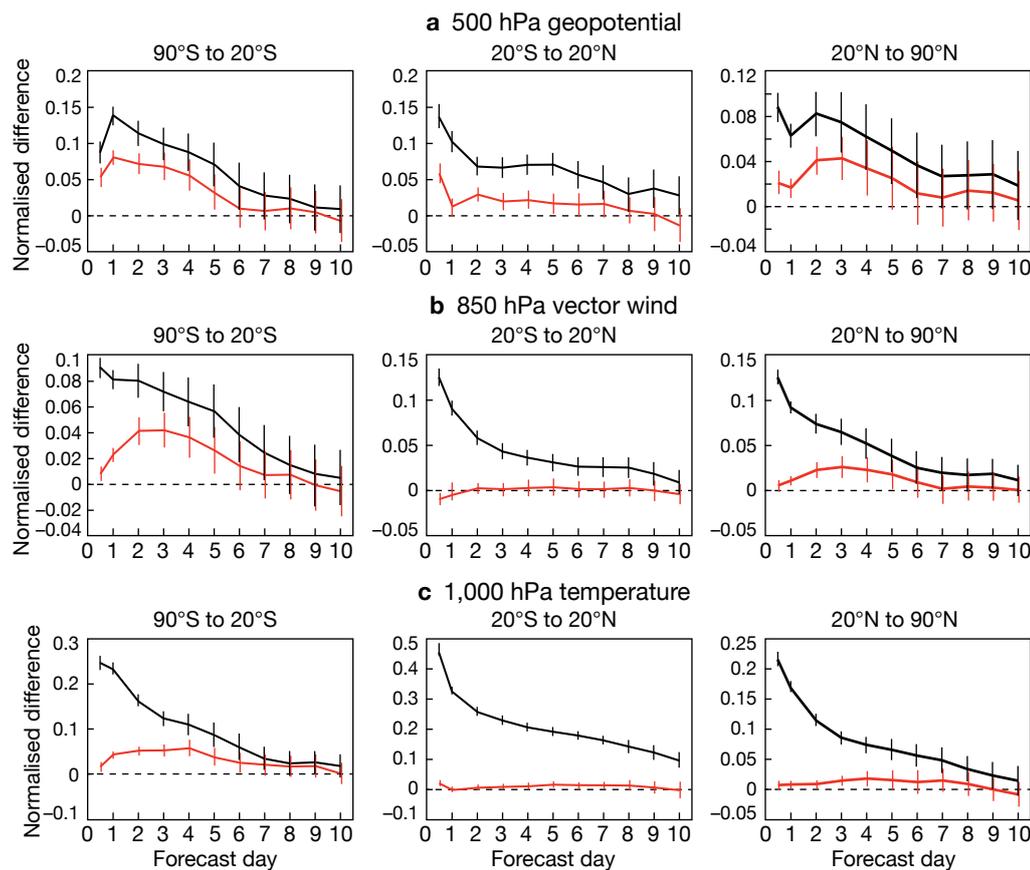
FIGURE 6 Relative impact from the inclusion of the linearised physics in ECMWF's 4D-Var system on forecast scores against own analysis, shown in terms of root-mean-square errors normalized by the Cycle 47r3 reference. Positive values correspond to an improvement when including the linearised physics in 4D-Var, and bars indicate significance at the 95% confidence level. The impact from using observations related to the physics is in red, while the combined impact of including the whole linearised physics and observations requiring physics information is in black. Results are shown for (a) 500 hPa geopotential, (b) 850 hPa vector wind, and (c) 1,000 hPa temperature, and for different regions as indicated. The scores are computed for the period July to September 2021.

precipitation, surface temperature and moisture). Therefore, proper maintenance and development of the LP will be an absolute prerequisite to the longevity of 4D-Var, unless a competitive alternative can be found. For instance, ECMWF is looking at the potential role of machine learning emulators in the EU-funded MAELSTROM project.

Developing well-behaved and efficient linearised physics can be a complex and demanding task. For any new or major revision to a parametrization, a substantial amount of work can be required to simplify and regularise the code so that discontinuities and nonlinearities are eliminated or smoothed out. It is also essential to verify that the TL approximation is not degraded whenever a physical parametrization is modified, or the model horizontal or vertical resolution is changed, and to swiftly address any deterioration through appropriate updates to the TL and AD codes.

To maintain the best analysis and forecast performance possible, the best compromise between the constraints of linearity and computational cost on the one hand and realism on the other must be achieved. Increasingly complex parametrizations in the nonlinear model and increasing model and data assimilation resolution make maintaining the LP package challenging. However, these results demonstrate the considerable benefit to forecast skill achieved by continuing to meet these challenges.

Further reading

Buizza, **R.**, 1994: Impact of simple vertical diffusion and of the optimisation time on optimal unstable structures. *ECMWF Technical Memorandum*, **No. 192**.

Courtier, **P.**, **J.-N. Thépaut** & **A. Hollingsworth**, 1994: A strategy for operational implementation of 4DVar using an incremental approach. *Q. J. R. Meteorol. Soc.*, **120**, 1367–1387.

Janisková, **M.**, **J.-F. Mahfouf**, **J.-J. Morcrette** & **F. Chevallier**, 2002: Linearised radiation and cloud schemes in the ECMWF model: Development and evaluation. *Q. J. R. Meteorol. Soc.*, **128**, 1505–1527.

Janisková, **M.** & **P. Lopez**, 2013: Linearised physics for data assimilation at ECMWF. In: S.K. Park and L. Xu (eds.), *Data assimilation for Atmospheric, Oceanic and Hydrological Applications* (**Vol. II**), *https://doi.org/10.1007/978-3-642-35088-7*, Springer-Verlag Berlin Heidelberg, 251–286.

Lopez, **P.** & **E. Moreau**, 2005: A convection scheme for data assimilation: description and initial tests. *Q. J. R. Meteorol. Soc.*, **131**, 409–436.

Mahfouf, **J.-F.**, 1999: Influence of physical processes on the tangent-linear approximation. *Tellus*, **51**, 147–166.

Tompkins, **A.M.** & **M. Janisková**, 2004: A cloud scheme for data assimilation: description and initial tests. *Q. J. R. Meteorol. Soc.*, **130**, 2495–2517.

Enhancing OpenIFS by adding atmospheric composition capabilities

Marcus O. Köhler, Adrian A. Hill (both ECMWF), Vincent Huijnen, Philippe Le Sager (both KNMI)

OpenIFS is a portable version of ECMWF's global weather forecasting model. Over the past decade, its user community has grown continually, and the model is used for research in many areas of atmospheric science. This has been further expanded with the release of the OpenIFS/AC model in 2022. This extension of OpenIFS permits the simulation of atmospheric composition, including trace gases and aerosols and their interaction with other atmospheric processes.

The road to OpenIFS/AC

For more than a decade, the OpenIFS activity has provided supported, portable configurations of the operational ECMWF Integrated Forecasting System (IFS) (Carver, 2022). The broad aims of OpenIFS are to facilitate easier ways for scientists in ECMWF Member and Co-operating States to access the IFS, to enable its external use for research and teaching on numerical weather prediction (NWP), to promote knowledge exchange, and to enhance collaborations between national hydrological & meteorological services, universities and research institutes.

The OpenIFS model is a 3D global forecasting model which consists of the IFS atmospheric model with all the subcomponents for forecast-only simulations, including the fully coupled land surface and ocean surface wave model. Since OpenIFS is designed to be light-weight and for forecast-only applications, the observational data processing system and data assimilation are excluded.

OpenIFS is distributed under a free licence agreement which permits its use for teaching and research. Its user community has grown from year to year, and it now consists of over one hundred licensed institutes, which are mostly located in and around Europe. OpenIFS users engage in a wide spectrum of scientific and technical modelling activities, which cover simulation timescales from short forecasts to climate simulations. The research carried out by OpenIFS users extends over an impressive range of topics that is increasingly reflected in the published literature. OpenIFS is also a useful tool for professional training and for teaching the next generation of atmospheric scientists, both internally at ECMWF and at other institutions.

An overview of some of the uses of OpenIFS in training and teaching is shown in Szépszó et al. (2019).

An expanding area of research is the linkage between air quality, meteorology and associated feedbacks of air quality on weather forecasting. This includes the short-term variability of trace gas and aerosol processes in the atmosphere at a range of spatial scales. When linked to weather forecasts, predicted air quality can contribute to mitigating the public health impacts caused by air pollution. It can also help in the identification of air pollution sources and rapid changes in emissions of primary pollutants or their chemical precursors. Growth in supercomputer performance and improved accuracy in operational forecasting, combined with an improved understanding of atmospheric chemistry and aerosol processes, make it possible to consider full two-way interaction and feedback between chemistry and meteorology in global models.

The standard OpenIFS distribution does not include the scientific or technical capability to simulate composition processes (trace gases and aerosol) alongside meteorology. The OpenIFS – Atmospheric Composition project (OpenIFS/AC) aimed to address this by adding interactive composition functionality. OpenIFS/AC is a collaborative project between ECMWF and Member State scientists in the Netherlands and in Finland, under the leadership of the Royal Netherlands Meteorological Institute (KNMI). The main aim of OpenIFS/AC is to incorporate atmospheric composition components of the IFS into OpenIFS. These components are used by the EU's Copernicus Atmosphere Monitoring Service (CAMS) implemented by ECMWF. Such a development extends the capability of the OpenIFS model to include additional code that allows the simulation of many processes that affect atmospheric composition coupled to meteorology. This includes both gas-phase chemistry, in the troposphere and in the stratosphere, and aerosol-related processes in the troposphere. This new capability widens the scope of research that can be undertaken by the OpenIFS community, by providing a tool for studying the interactions between atmospheric composition and either NWP or the climate system.

On the role of atmospheric composition

Understanding and modelling atmospheric composition is fundamental for the prediction of local air quality, which

is an important aspect of public health and depends on the ambient concentrations of trace gases and aerosol in the atmosphere. Furthermore, gases and aerosols can interact with Earth's radiative balance, resulting in local heating or cooling, which can influence atmospheric processes in the short term, relevant for NWP, and on longer timescales, relevant for the climate system. On NWP timescales, experimental and modelling studies have shown that many meteorological processes and phenomena, such as cloud and fog formation and lifetime, precipitation, and radiative balance, are influenced by aerosol processes and the chemical composition of the atmosphere (e.g. Miltenberger et al., 2018). So, in future, greater complexity may be justified for global NWP. On longer timescales, atmospheric composition is relevant for studies of the present as well as for past and future climates, reflecting natural variability in the abundance of chemical compounds and their historical change due to human activities. In the context of Earth system modelling, the representation of atmospheric composition and its associated radiative impacts on the climate system have featured prominently in many assessments of the Intergovernmental Panel on Climate Change. In the context of identifying policy measures aimed at improving air quality or mitigating near-term climate change, it is essential to gain an understanding of the processes that govern the presence and the amounts of these constituents and of the factors that affect their variability and temporal evolution.

Many of these constituents are released into the atmosphere through direct natural and anthropogenic emissions. They often form chemical precursor components for complex chemical reaction chains. The distribution of trace gases is partly determined by small-scale processes, such as turbulent mixing, and by large-scale transport processes, such as convection. Physical removal also plays a significant role, either by dry deposition on the Earth's surface or through wet scavenging, involving uptake in cloud droplets and precipitation processes. The largest part of the reactive constituents is, however, controlled through chemical processes, especially photochemistry. This can either occur in the gas phase between trace gases, as seen for instance in summertime ozone smog, or it can involve liquid or solid aerosol particles, for which the processes leading to stratospheric ozone depletion are one of the best-known examples.

When considering NWP spatial and temporal resolutions, these processes have historically been simulated with considerable detail through Chemical Transport Models (CTMs). CTMs are driven by external meteorological fields, such as operational forecasts or global reanalyses. Importantly, they cannot give direct feedback on the meteorological fields. This dependence on meteorological inputs without any feedback represents a limitation in the sense that the full loop of interactions between constituents

and physical processes is not closed. An alternative to CTMs are weather forecasting models in which the explicit treatment of atmospheric composition is coupled to the meteorology. An overview of such models is shown in Baklanov et al. (2014). One of these is the IFS in CAMS configuration, which is an extended version of the IFS used to simulate trace gases and aerosol in the atmosphere in an operational context and to provide reanalyses that contain assimilated observations of atmospheric composition. Such interactive coupling between composition and meteorology generates additional computational costs when they both use the same model grid. The increase in the performance of high-performance computing facilities in recent years enables the inclusion of explicit and detailed chemistry in global weather forecasting models, such as OpenIFS. OpenIFS/AC is a research tool that offers some of the capabilities of the operational IFS as used in CAMS, by permitting the full interaction between meteorology, its impacts on atmospheric composition, and related radiative feedbacks.

Creating OpenIFS/AC

Version 1 of OpenIFS/AC, which was released in 2022, includes the implementation of a chemistry and aerosol scheme in the model. This has been achieved by porting modules developed for the IFS as part of CAMS activities. The tropospheric chemistry module originates from the TM5 CTM, as developed and updated at KNMI (Flemming et al., 2015), and the reactions are based on a modified version of the Carbon Bond 2005 (CB05) chemical mechanism, originally developed for the US Environmental Protection Agency. Additionally, a stratospheric chemistry module can be enabled, which is based on the Belgian Assimilation System for Chemical ObsErvations (BASCOE) (Huijnen et al., 2016). The AER bulk bin aerosol scheme to represent tropospheric aerosol (Rémy et al., 2019) has also been included. The OpenIFS/AC model can therefore be used in various configurations, either with tropospheric chemistry (CB05), additional stratospheric chemistry (CB05+BASCOE) and optionally with the AER tropospheric aerosol code enabled. In the current release of OpenIFS/AC, the tropospheric chemistry and aerosol schemes are closely aligned with those used for the CAMS reanalysis. Note that in the stratosphere the tropospheric chemistry version uses a linearised ozone parametrization. A more complete technical description of the OpenIFS/AC model is available in Huijnen et al. (2022). Examples for results from atmospheric composition modelling are given in Figures 1 to 3, showcasing the ability of OpenIFS/AC to simulate stratospheric ozone, enhanced aerosol due to anthropogenic and fire emissions, and tropospheric NO_2, which is a marker of various pollution sources and plays an important role in tropospheric ozone production.

The chemistry schemes have been designed to represent essential chemical processes on a global scale, suitable for a global forecast model in an NWP framework. It is,
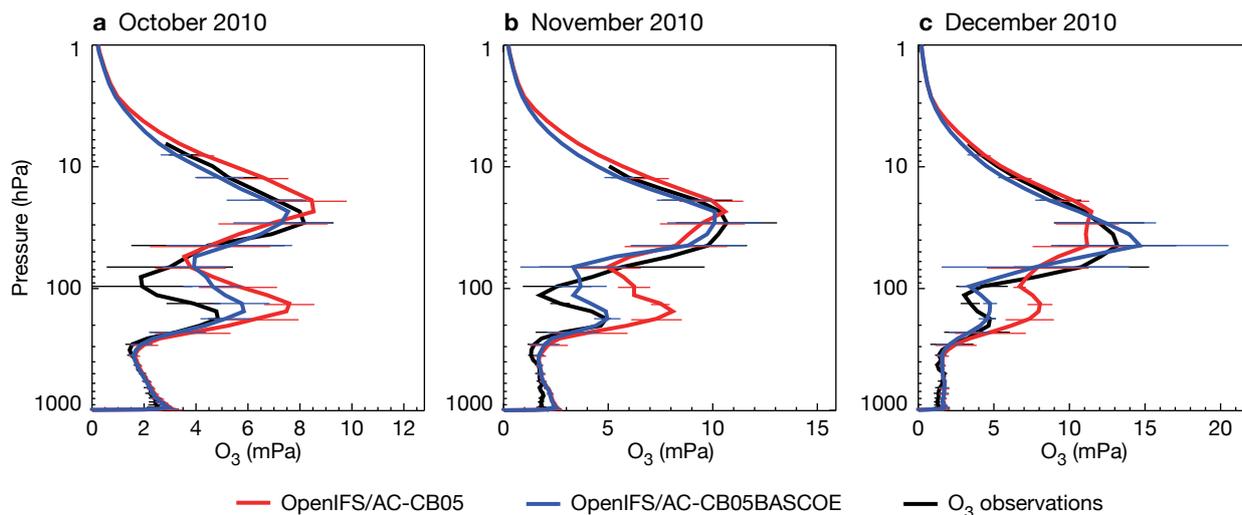
FIGURE 1 Evaluation of OpenIFS/AC stratospheric ozone against sondes (black lines) during (a) October 2010, (b) November 2010, and (c) December 2010 at Neumayer station (Antarctica) showing the performance of OpenIFS/AC-CB05 (red lines) and the OpenIFS/AC-CB05BASCOE (blue lines) configurations to simulate stratospheric ozone depletion.

however, possible to adapt these schemes to different purposes. Complexity can be added or removed for specific applications or to provide various constraints to the model to simulate certain time periods. This can be achieved by providing relevant input parameters for the initial conditions and composition surface fluxes at the start of the run. Different ways to constrain meteorology and surface fluxes throughout the run can also be used. One feasible way to achieve this is by running the model in 'nudged mode', relaxing various meteorological variables either globally or regionally towards provided input fields. The nudging method is particularly suitable to constrain the model towards specific atmospheric conditions that occurred during a given period in the past.

The computational requirements for the simulation of complex atmospheric chemistry are high. The reason is, firstly, the introduction of additional model fields for atmospheric composition data, including calculations

related to their transport and associated diagnostics. Secondly, additional calculations required for chemical reactions and physical loss processes also make a big contribution. Huijnen et al. (2022) estimate that activating the tropospheric chemistry scheme increases computational costs between a factor of three to four compared to an OpenIFS experiment without chemistry. As a result, a compromise between model resolution and the complexity of chemistry needs to be found to reduce the computational overhead. Because atmospheric composition uses the same grid as the other model processes, global model experiments with atmospheric composition often use grid resolutions that are significantly coarser compared to those used in NWP. Also, for longer simulation time periods, alternative chemical mechanisms with reduced complexity are used. To make simulating atmospheric composition computationally affordable, a range of chemistry and aerosol schemes with varying complexity need to be employed depending on the science goals.
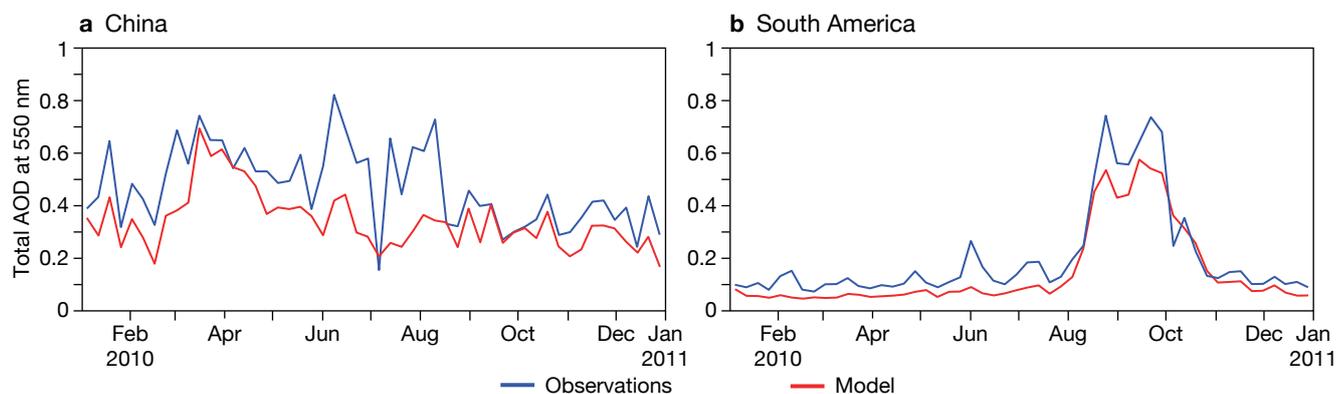
FIGURE 2 Evaluation of total aerosol optical depth (AOD) at 550 nm from OpenIFS/AC (red) against observations from the AERONET network of ground-based sun photometers (blue) between January 2010 and January 2011 for stations in (a) China, and (b) South America, representative of pollution due to anthropogenic and biomass burning sources.

a Model tropospheric NO_2 columns
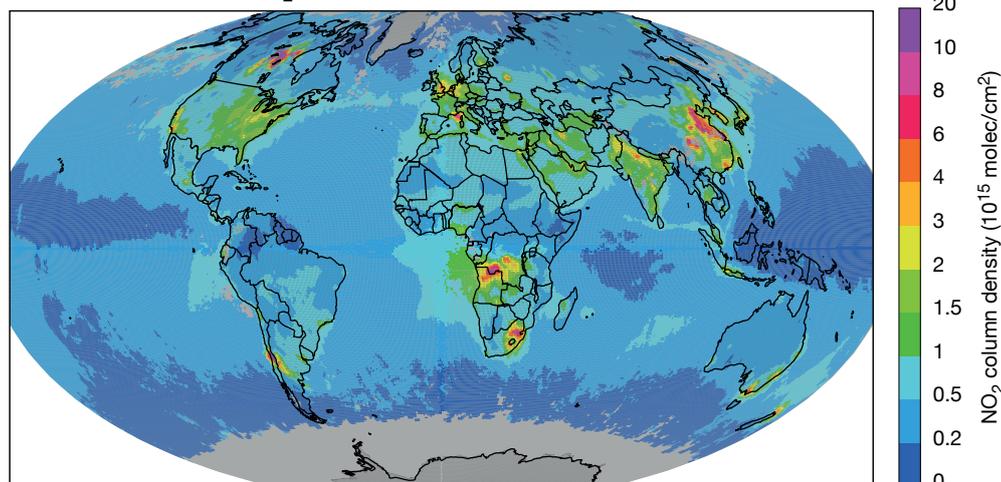
b Model bias with respect to observations
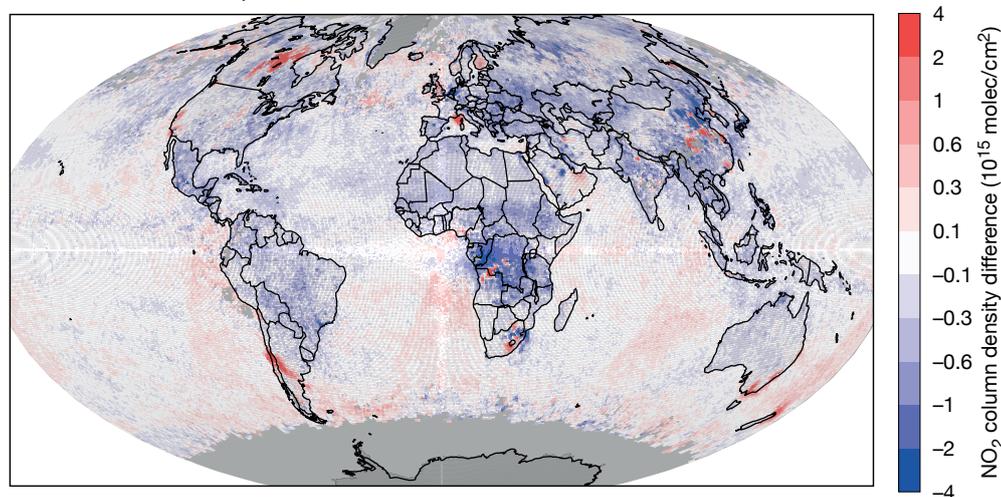
FIGURE 3 Evaluation of OpenIFS/AC-CB05 tropospheric NO_2 columns against Ozone Monitoring Instrument (OMI) satellite observations, averaged for July 2010, showing (a) model tropospheric NO_2 columns, and (b) model bias with respect to the observations. The hotspots in the biases are mainly driven by emissions.

While OpenIFS can be used with a wide range of horizontal model grids and a varying number of vertical levels, equivalent to those available for the IFS, using this flexibility for OpenIFS/AC would require providing input data at all grid resolutions. Currently the input data for the model chemistry is only made available for one model grid configuration, a linear reduced Gaussian grid of TL255, corresponding to approximately 80 km grid point spacing, and for 91 vertical model levels. This grid resolution was selected as a compromise between computational costs and expected usefulness for a range of model applications. Input data for other model grid resolutions will become available in the future.

Next steps for OpenIFS/AC

OpenIFS/AC is an ongoing activity. The aim is to continue providing and further expanding the capability of atmospheric composition modelling in future OpenIFS releases. OpenIFS/AC benefits from recent model developments undertaken in CAMS. In addition, new developments and improvements will be included that are currently not available in CAMS. The next release of OpenIFS/AC, Version 2, will be the last to be based on the current OpenIFS 43r3, with subsequent releases using the next available cycle of the OpenIFS model. Version 2 of OpenIFS/AC will include an alternative scheme to represent tropospheric aerosol in the model, together with a parameterized sulphate production scheme, to make it particularly suitable for climate applications.

The main motivation for this effort is that OpenIFS has been adopted as the atmospheric modelling component of the EC-Earth consortium (*https://ec-earth.org*), who develop an Earth system model for climate modelling by its European user community. Earth system climate modelling requires the representation of processes relevant to longer timescales and of compounds relevant to the climate system (such as primarily greenhouse gases and aerosol), many of which are also controlled to a certain degree by atmospheric chemistry. To answer various questions on climate sensitivities, different levels of composition modelling are required. The consortium's latest model, version EC-Earth4, includes the current OpenIFS 43r3 together with its extension for atmospheric composition.

Simulations over longer timescales often benefit from the

reduction in computational costs when chemistry schemes with reduced complexity are employed. In Version 2, this will become available through a reduced parametrization focusing on sulphur compounds. The scheme employs monthly climatologies of oxidant fields (e.g. oxides of hydrogen and ozone). For consistency, these oxidant fields have been pre-computed with the full-complexity chemistry scheme, and they provide the necessary inputs for a reduced-complexity sulphate aerosol chemical mechanism. In this case, the explicit simulation of photochemical processes for the short-lived oxidants has been removed. This is an acceptable limitation for studies which extend over long timescales or which have a predominant focus on aerosol processes.

This reduced chemistry can be coupled to a more detailed aerosol code, M7, which will also become available in Version 2. The M7 scheme is a double-moment modal aerosol microphysics scheme, in which aerosol is represented in size modes, with a prognostic mass mixing ratio and number concentration (the number of particles per unit volume of air) for each mode. It accounts for internal mixing of different chemical components within aerosol particles, and it is designed to provide detailed aerosol particle size information. As such, M7 enhances the complexity of aerosol representation compared to the existing AER scheme. AER is a single-moment scheme, wherein aerosol is only represented by mass mixing ratio while the number concentration is assumed. In climate models, double-moment schemes have been shown to improve the description of aerosol optical properties and the representation of important and uncertain aerosol-cloud forcings (e.g. Bellouin et al., 2013), which impact the life cycle of the aerosol particles and their direct and indirect radiative effects. The need for double-moment aerosol schemes in NWP and air quality modelling is an open research question. The inclusion of M7 in OpenIFS/AC creates a flexible system that can tackle this research topic.

Finally, simulating atmospheric composition requires a range of additional model inputs to drive the chemical modules. Initial conditions in the form of global fields for all chemical tracers need to be provided at the start of the experiment, as well as boundary conditions for the duration of the run. These include daily fluxes of emitted chemical species, and information on removal rates by deposition subject to surface type. In the current release of OpenIFS/AC, the emission fluxes from natural biogenic sources, from wildfire emissions, and from a range of anthropogenic sectors are lumped together into a single flux per emitted species. These emission fluxes match those used in the Coupled Model Intercomparison Project (CMIP6). To simplify the input data generation process, we aim to expand the capabilities of the OpenIFS Data Hub in a future release to include initial experiment data for OpenIFS/AC experiments.

Conclusion

The release of OpenIFS/AC in 2022 has provided a first model version of OpenIFS that allows the interactive simulation between atmospheric composition and meteorology, based on code extensions that are used in the CAMS operational model. In forthcoming model releases, we plan to expand this new capability by including alternative codes for composition modelling and by supporting more grid resolutions. We also intend to make the generation of input data more convenient, thus making this new research capability attractive to a wider model user community.

Further reading

Baklanov, **A.**, **K. Schlünzen**, **P. Suppan**, **J. Baldasano**, **D. Brunner**, **S. Aksoyoglu et al.**, 2014: Online coupled regional meteorology chemistry models in Europe: current status and prospects, *Atmos. Chem. Phys.*, **14**, 317–398, *https://doi.org/10.5194/acp-14-317-2014*.

Bellouin, **N.**, **G.W. Mann**, **M.T. Woodhouse**, **C. Johnson**, **K.S. Carslaw** & **M. Dalvi**, 2013: Impact of the modal aerosol scheme GLOMAP-mode on aerosol forcing in the Hadley Centre Global Environmental Model, *Atmos. Chem. Phys.*, **13**, 3027–3044, *https://doi.org/10.5194/acp-13-3027-2013*.

Carver, **G.**, 2022: Ten years of OpenIFS at ECMWF, *ECMWF Newsletter* **No. 170**, 6–7.

Flemming, **J.**, **V. Huijnen**, **J. Arteta**, **P. Bechtold**, **A. Beljaars**, **A.-M. Blechschmidt et al.**, 2015: Tropospheric chemistry in the Integrated Forecasting System of ECMWF, *Geosci. Model Dev.*, **8**, 975–1003, *https://doi.org/10.5194/gmd-8-975-2015*.

Huijnen, **V.**, **J. Flemming**, **S. Chabrillat**, **Q. Errera**, **Y. Christophe**, **A.-M. Blechschmidt et al.**, 2016 : C-IFS-CB05-BASCOE : stratospheric chemistry in the Integrated Forecasting System of ECMWF, *Geosci. Model Dev.*, **9**, 3071–3091, *https://doi.org/10.5194/gmd-9-3071-2016*.

Huijnen, **V.**, **P. Le Sager**, **M.O. Köhler**, **G. Carver**, **S. Rémy**, **J. Flemming et al.**, 2022: OpenIFS/AC: atmospheric chemistry and aerosol in OpenIFS 43r3, *Geosci. Model Dev.*, **15**, 6221–6241, *https://doi.org/10.5194/gmd-15-6221-2022*.

Miltenberger, **A.K.**, **P.R. Field**, **A.A. Hill**, **P. Rosenberg**, **B.J. Shipway**, **J.M. Wilkinson et al.**, 2018: Aerosol–cloud interactions in mixed-phase convective clouds – Part 1: Aerosol perturbations, *Atmos. Chem. Phys.*, **18**, 3119–3145, *https://doi.org/10.5194/acp-18-3119-2018*.

Rémy, **S.**, **Z. Kipling**, **J. Flemming**, **O. Boucher**, **P. Nabat**, **M. Michou et al.**, 2019: Description and evaluation of the tropospheric aerosol scheme in the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS-AER, cycle 45R1), *Geosci. Model Dev.*, **12**, 4627–4659, *https://doi.org/10.5194/gmd-12-4627-2019*.

Szépszó, **G.**, **V. Sinclair** & **G. Carver**, 2019: Using the ECMWF OpenIFS model and state-of-the-art training techniques in meteorological education. *Adv. Sci. Res.*, **16**, 39–47, *https://doi.org/10.5194/asr-16-39-2019*.

Migration from GRIB1 to GRIB2: preparing ECMWF model output for the future

Robert Osinski, Matthew Griffith, Sébastien Villaume

I n 2022, ECMWF started a multi-year effort to migrate its daily operations data output from the file format GRIB edition 1 (GRIB1) to GRIB edition 2 (GRIB2). The project is partly a response to the call for global numerical weather prediction (NWP) at convection-permitting resolutions set out in ECMWF's ten-year Strategy 2021–2030. Such resolutions require GRIB2 rather than GRIB1 data because of the limitations of GRIB1 grid definitions. ECMWF has already produced Integrated Forecasting System (IFS) output on vertical model levels using the GRIB2 format for several years. Here we present an overview of where we are in the transition from GRIB1 to GRIB2 for all of our output.

Setting the scene

GRIB1 was created in 1985 and was commonplace by the early 1990s. It was not designed to accommodate horizontal grid resolutions needed to resolve convective-scale phenomena (1–4 km), which are called for in the ten-year Strategy and which will be used in the digital twins for the EU's Destination Earth initiative, in which ECMWF participates.

GRIB1 also has other limitations and disadvantages. The most important one is that GRIB1 was deprecated by the World Meteorological Organization (WMO) more than a decade ago in favour of GRIB2, and that it has not been referenced in the WMO Manual on Codes since 2016 (World Meteorological Organization, 2022). Other limitations of GRIB1 include:

- A maximum vertical resolution of 127 levels; this limitation was hit in 2011 and is the reason why the data on vertical model levels were migrated to GRIB2 during the implementation of Cycle 37r2 (137 vertical model levels).

- GRIB1 allows for the definition of only 128 different parameters and does not have the necessary metadata to describe modern NWP outputs, such as ensembles and probabilities.

- GRIB1 does not have an official built-in mechanism to extend its metadata. Over the years, ECMWF has extended the limited GRIB1 metadata through the permitted local section found in section 1 of the header. This is how ensemble members and ensemble size were introduced in GRIB1. The drawback of these extensions is that they are not endorsed by the WMO and thus not part of the official data format.

GRIB2 resolves these limitations and brings critical new features:

- **Support for vertical resolutions with more than 127 levels:** as above, this was a critical feature when the model was upgraded to use 137 vertical model levels.

- **Support for horizontal resolutions at sub-kilometre scales:** planned resolution increases in ECMWF's ten-year Strategy are well within this scope.

- **Support for millions of different parameters:** NWP parameters can now be encoded with much more freedom.

- **Support for ensemble, re-forecast and post-processed products:** this addresses the most common limitations of metadata to describe products with context.

- **Support for a wide range of compression methods:** this is critical with increasing data volumes at higher model resolutions.

- **Support for rich metadata:** this enables more prescriptive parameter descriptions and improves discoverability and indexing.

- **Introduction of templating:** this allows the continuous integration of new templates when additional metadata is required.

These headlines, some of which are shown in Figure 1, describe some of the fundamental design changes which have been put in place within the GRIB2 format. We discuss these in more detail below.

GRIB2 design philosophy

The new features and improved design of GRIB2 allow for a much more self-descriptive data format with an improved user experience. Expanding on the above points, GRIB2 encompasses the following:
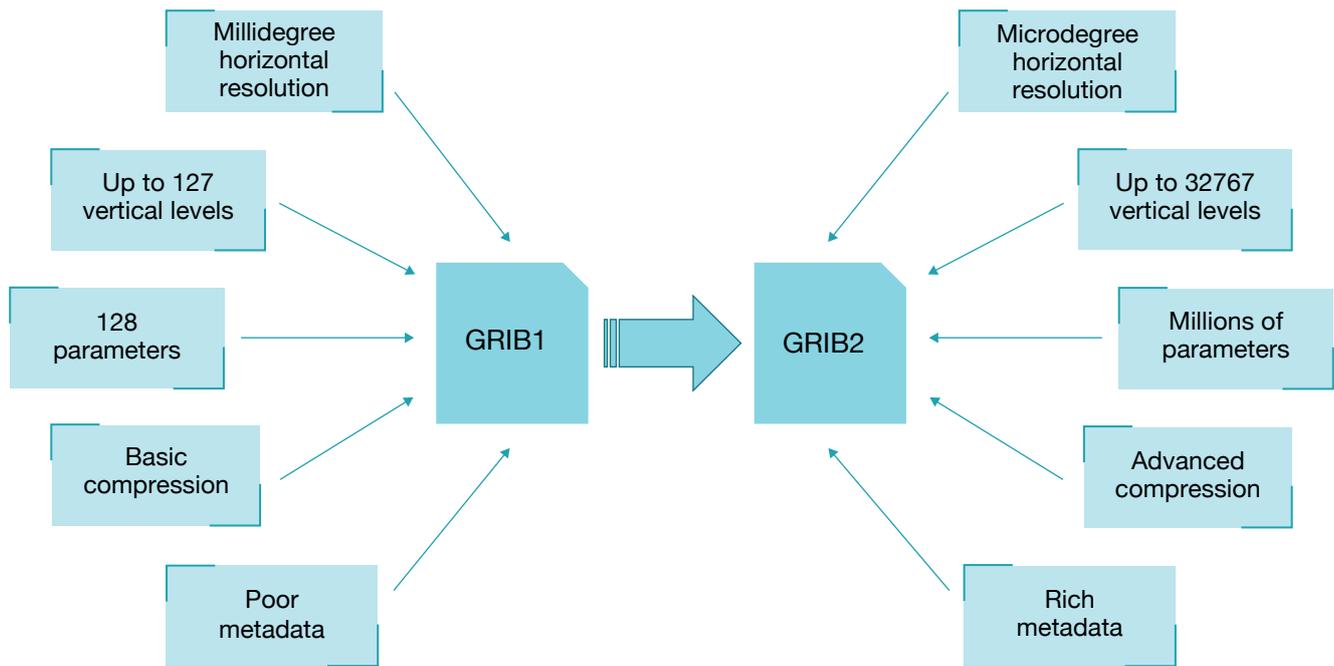
FIGURE 1 Some of the differences between the GRIB1 and GRIB2 file formats.

- **Horizontal resolution:** previously limited to millidegree precision, this can be encoded up to the precision of a microdegree, allowing resolutions below the kilometre scale. This is sufficient to encode data according to the resolution increases planned at ECMWF for the next decade.

- **Parameters:** the total number of parameters that can be defined is in practice unlimited. In GRIB2, a parameter is no longer simply represented by a single entry in a code table but by a combination of entries in various code tables. Thus, the minimum number of metadata keys is now a triplet: discipline, parameter category and parameter number. The top level of the hierarchy gives the discipline within which the parameter is defined, such as meteorology, hydrology or oceanography. Then, within each discipline, the parameters are organised into categories. For instance, in the discipline 'meteorology', one can find the categories 'momentum', 'temperatures' or 'short-wave radiation'. Finally, one selects a parameter within the category. For complex parameters, additional keys are required. The GRIB2 section 2, called the local section, is reserved for encoding centre-specific, local metadata for a parameter. This is used at ECMWF for Meteorological Archival and Retrieval System (MARS) keys, e.g. class and stream. It has the advantage that the message itself must conform to the standards of the GRIB2 data format.

- **Compression:** the data representation section offers a wide range of compression methods. This helps reduce the size of archived data, which will increase

significantly with higher model resolutions. Recently, a fast and lossless compression algorithm with a high compression ratio, developed by the Consultative Committee for Space Data Systems (CCSDS), has been implemented for GRIB2 in ECMWF's IFS, and this feature will be activated in the implementation of IFS Cycle 48r1 later in 2023 (Betke et al., 2022).

- **Rich metadata:** The metadata set describing a given parameter is rich, enabling better parameter descriptions and improved discoverability and indexing in accordance with FAIR data principles. In the context of the migration-to-GRIB2 project, we have developed and will develop new templates to extend the metadata to enable the encoding of all of our products.

- **Templates:** The grid section, product section and data representation section of a GRIB2 message can now be templated, allowing the continuous integration of new templates when additional metadata is needed. Template extensions can be requested from the WMO twice a year through an amendment procedure of the manual on codes called the 'Fast Track procedure'. The request is then processed by the WMO over a period of around six months, after which the amendments are published by the WMO and are ready to be used operationally.

To aid the understanding of the approach behind GRIB2, it is good to look at an example. We shall compare the metadata between a GRIB1 and GRIB2 parameter encoding for "Maximum temperature at 2 metres in the last 24 hours". This is presented in Table 1.

GRIB2 metadata follows a strategy that can be summarised by the 'what, where, when, and how' approach. This is illustrated in Table 1, using a colour-coded key. It can be thought of as follows:

- '**What** is being encoded?' This is the base or core of the parameter and is always defined by the keys discipline, parameterCategory and parameterNumber (in red). For this example, we have discipline 0 (meteorology), parameter category 0 (temperatures) and parameter number 0 (temperature). If the parameter requires no more metadata, we could stop here. However, in the vast majority of cases we then use additional keys to extend the scope of the parameter:

- '**Where** is the parameter defined?' This refers to the vertical spatial range or spatial position for which my parameter is valid (in purple). Here, it is at a specific fixed level – at 2 metres above the surface.

- '**When** is the parameter defined?' This refers to the time range or time point for which my parameter is valid or is processed (in green). Here, we can see that we perform processing over a 24-hour time period. This combines with the 'how' key to indicate the kind of processing we perform.

- '**How** is the parameter processed in time?' This key tells us how we want to statistically process the

parameter in time (in orange). Here, we can see it is 'maximum'. This combines with the 'when' key to give us a maximum in the last 24 hours.

This key-value type design is very powerful and flexible and allows for a direct and intuitive mapping to keywords used in MARS.

Challenges

The migration to GRIB2 poses several challenges. The GRIB format has been tightly coupled to the Centre's dataflow and many of our tools are designed to take advantage of the GRIB data format. While most tools use GRIB data in a transient manner and will only require migrating once, the MARS archive must continue to handle GRIB1 data properly for decades to come. At the time of writing, MARS has more than 200 PB of data in GRIB1 stored on tapes. Converting this data to GRIB2 to completely deprecate GRIB1 at ECMWF is not realistic as it would require a significant amount of time and resources. Instead, we are planning to keep serving this data as is but will offer a tool to convert on-the-fly to GRIB2.

Another challenging aspect of the migration will be the implementation of the migration in operations. This will require preparation upstream and should be implemented in the form of a technical cycle (although this has not been decided yet). Test data in GRIB2 will be released more than six months ahead of implementation to enable our Member and Co-operating States and other users to adapt their workflows accordingly.

The biggest challenge of this migration will be to handle the legacy data formatting standards accumulated over the life of GRIB1. The MARS language and GRIB1 have been around since the early 1990s. Throughout the years, both have been extended to accommodate new types of data that could not have been foreseen and planned for during their design phases. These are types of data that are commonplace for ECMWF now, such as ensembles, seasonal forecasts, hindcasts, probabilities, waves, oceanography, hydrology, and land surface modelling. Understandably, this has created technical debt which has accumulated over the years, making certain aspects of the migration very tricky.

In some cases, a direct migration will be impossible and will require some redesign. For instance, GRIB2 prescribes the units of the parameters to specific SI units and does not allow for alternative, equivalent units. A good example is the precipitation parameters produced by the IFS in units "metres of water", while GRIB2 expects the parameters to be expressed in kg m^{-2}. If we were to switch to producing our precipitation data in WMO standard units, this would create a discontinuity in the archive: a user trying to retrieve data spanning over the transition would receive part of the data in the old units and part in the new units. To solve this issue, we

Maximum temperature at 2 metres in the last 24 hours	
GRIB1	GRIB2
Table2Version = 128 (bears no specific meaning) IndicatorOfParameter = 51 (Maximum temperature at 2 m in the last 24 hours) IndicatorOfTypeOfLevel = 1 (surface)	discipline = 0 (meteorology) parameterCategory = 0 (temperatures) parameterNumber = 0 (temperature) typeOfFirstFixedSurface = 103 (height above ground in metres) scaleFactorOfFirstFixedSurface = 0 scaledValueOfFirstFixedSurface = 2 typeOfStatisticalProcessing = 2 (maximum) lengthOfTimeRange = 24 indicatorOfUnitForTimeRange = 1 (hour)

TABLE 1 Comparison of metadata describing a meteorological parameter in GRIB1 and GRIB2. The names of the keys correspond to those used in ecCodes, an ECMWF package for decoding and encoding messages in WMO formats. When the value taken by a key is followed by an explanation in brackets, it is because it references an entry in a table. For example, entry 103 in the table representing "Fixed surface types and units" corresponds to "height above ground in metres".

must define two sets of precipitation parameters, one set using standard WMO units and a second set with legacy units. The downside is that this second set is defined locally and not endorsed by the WMO. We could then either produce and archive both sets for convenience or we could only archive our local parameters and offer a conversion on-the-fly to the WMO parameters.

Timeline

A roadmap for the migration has been drafted (see Figure 2). The amount of work and the scale of the changes will not allow everything to be migrated at once. Several factors have been considered to set priorities and derive a workplan:

- Any new dataset with a new type of data (not existing in GRIB1) shall be produced entirely in GRIB2. This will be the case for the ocean reanalysis ORAS6 and the real-time OCEAN6. Between 2018 and 2022, the European Flood Awareness System (EFAS), the Global Flood Awareness System (GloFAS) and the Fire Copernicus Emergency Management System (CEMS-Fire) were all released as GRIB2 only datasets following this principle.

- Any new dataset replacing an existing dataset (produced in GRIB1) shall also be produced entirely in GRIB2. By this, we mean any datasets with a well-defined beginning and end. The atmospheric composition reanalysis EAC5 (replacing EAC4), the next global reanalysis ERA6 (replacing ERA5), and the new seasonal forecasting system SEAS6 (replacing SEAS5) fall into this category.

- Any new parameter shall be defined only in GRIB2. This has already been common practice for the past five years. It is the main reason why some surface parameters or non-model-level parameters, in addition to those on vertical model levels, are encoded in GRIB2. This is acceptable because the parameters are new, and therefore they do not introduce a change of behaviour or discontinuity in workflows or in the MARS archive. An example of recently added parameters are new thermal comfort indices, such as the UTCI (Universal Thermal Climate Index).

- Our existing IFS GRIB1 parameters, produced by the operational suite, will be the last to migrate to GRIB2 with an extended period of testing prior to implementation.

Ongoing migration progress

The next dataset in the scope of this work to be released in GRIB2 is the ocean reanalysis ORAS6. The work for this dataset and for OCEAN6 started several years ago independently of this migration project. This is due to the use of unstructured ocean grids (ORCA grids) by the model, which cannot be represented in GRIB1. This makes this dataset a natural candidate for the GRIB2 data format. The ocean grids have now been implemented in GRIB2 and are fully supported by ecCodes 2.20.0 and higher and ECMWF's Meteorological Interpolation and Regridding (MIR) software package. ORAS6 and OCEAN6 are scheduled for production during the second half of 2023.

The next major milestone is concerned with the datasets which will be based on IFS Cycle 49r1, namely ERA6,
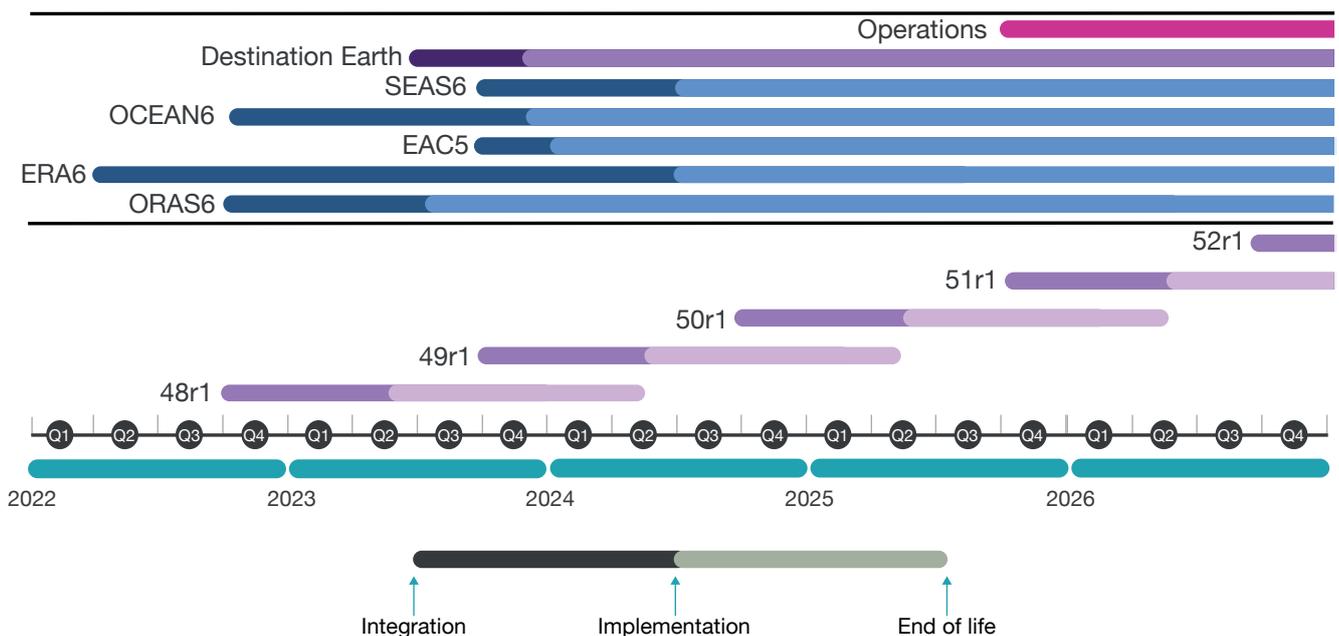
FIGURE 2 This migration roadmap indicates when GRIB2 is to be introduced in all of ECMWF's weather forecasting operations (from IFS Cycle 51r1) and when various other services that will use GRIB2 will become operational.

SEAS6 and EAC5. Early last year, we conducted an exhaustive inventory of all the parameters and concepts needed in GRIB2 for these projects. For EAC5, several hundreds of new parameters would be required, due to two main factors:

- The introduction of many new chemical species and aerosols. For each species, we would need a complete set of physical observables: wet deposition of <species>, dry deposition of <species>, mass mixing ratio of <species>, etc.

- The emissions are now to be resolved by emission sector, leading to yet more parameters: emission of <species> from <sector>. Typical sectors include agriculture, industry, road and volcanoes.

Fortunately, we can use the rich metadata in GRIB2 and the flexibility to extend this metadata through new templates. It is now possible to specify the chemical species or aerosol and the source of emissions through separate metadata keys. The implementation in IFS-COMPO (IFS composition) of this new scheme is well under way.

ERA6 will also require several new implementations to support its release in the GRIB2 format. It will be the first dataset to have wave parameters in GRIB2 including 2D wave spectra (directions and frequencies). These spectra cannot be represented in GRIB2 with existing templates. For the wave spectra and wave parameters, we submitted six new templates to the WMO in November 2022. Additionally, ERA6 will also offer many new parameters, such as new water and energy budget parameters. These parameters, together with the parameters already produced in ERA5, were reviewed, mapped and requested through the WMO approval process when required. The templates and the parameters for ERA6 have just been accepted and will be published in the WMO Manual on Codes in May 2023.

We are also working on other aspects of metadata modelling for the migration. Recent developments, which will also be used in the Destination Earth initiative, include the new snow, soil and sea-ice multilayer schemes. The multilayer snow scheme and corresponding multilayer GRIB2 output will already be introduced in IFS Cycle 48r1.We also looked at how to encode metadata for the Extreme Forecast Index (EFI), the Shift of Tails (SOT), and anomalies based on climate distributions. Four new templates were created to encode these. Finally, we worked on a way to encode the metadata for optical parameters which are wavelength dependent, and we proposed four new templates to achieve this. These templates will be useful for parameters related to simulated satellite images and the radiation parameters used to produce the popular 'space view' images for IFS output.

Future developments

We are currently working on the design of templates to enable the encoding of tile-based parameters. Recent developments in land-surface modelling make use of the partitioning of the grid box into 'tiles' or 'patches' with their own properties and modelled physical processes. Typical tile classes include high vegetation, low vegetation, oceans, lakes, urban land and bare land. However, tile schemes can also be much more detailed in the granularity of the chosen tiles, including more than 20 different types of tiles accounting for different vegetation types across the globe. This kind of partitioning is particularly useful for parameters like 2-metre temperature, as this has large variations depending on the surface over which it is measured. For example, the effect of urban areas, lakes, oceans, and a forest could all be taken into account by encoding the temperature on the tile it is associated with. We are actively working with modelling teams from various European meteorological services to draft templates that could be used by all major European land surface models.

Finally, a template to encode a new type of horizontal grid, the HEALPix (Hierarchical Equal Area isoLatitude Pixelization) grid, is also in the pipeline. This grid, originally designed for cosmological applications, has recently gained much attention due to its attractive and versatile properties. This template, together with the tile templates, will be submitted to the WMO for validation during the next WMO Fast Track procedure.

Expected user impact

The migration from GRIB1 to GRIB2 is comparable to the migration from the Python 2 to the Python 3 programming language: it will require changes in workflows, from the modification of scripts to changes in existing practices. Certain features or parameters will need to be deprecated, too. However, importantly, this will not require a complete rewrite of applications and tools. As in the case of Python 2 and Python 3, we are expecting both ecosystems of dataflow to co-exist for several years. IFS Cycle 49r1, which is due to be implemented next year, is probably the most relevant example of this. This IFS cycle should be able to run in legacy mode in operations, but it should also be able to run in GRIB2 only mode, for ERA6 and Destination Earth. We are actively working on a technical solution allowing this switch rather than maintaining parallel releases of ecCodes.

The user interaction with GRIB2 messages via ecCodes will be the same as in GRIB1, i.e. mostly by setting/ accessing edition-independent keys: dataDate, dataTime, paramId, typeOfLevel, etc. However, for certain parameters there will be changes in the representation used and in the method of access in

GRIB2. The following are the most common examples of such changes:

- Some parameters will obtain a different paramId in GRIB2. The 'soil temperatures level 1/2/3/4' are a good example of this. In GRIB1, these are represented by four separate paramIds all on a unique level called 'surface'. In GRIB2, they will all be represented by a unique paramId on four discrete soil levels/layers.

- A paramId in GRIB2 may need to be complemented by additional keys, for example a wavelength for optical parameters or a chemId to specify a chemical species/aerosol.

- A pre-existing GRIB2 representation of a parameter may become deprecated. This can happen when the representation was erroneous, incomplete or for other technical reasons. In this case, it will still be possible to read and decode such a parameter, but ecCodes will use the new representation when writing the parameter to a file.

To date, development has focused on changes with limited impact for users. However, we are now entering a phase of the project where we are tackling the more visible changes with higher user impacts. These will be clearly announced in ecCodes release notes and through other appropriate channels of communications (see below). It is therefore recommended to use the latest version of ecCodes at all times to avoid issues in the migration of workflows and to benefit from the latest features.

MARS requests for operational data will also be affected by the migration to GRIB2. This is because the migration will inevitably create a discontinuity at the time of implementation. A user who wants to retrieve data overlapping the transition may need to use separate requests for the retrieval depending on the parameters of interest.

We are still working on the best approach to handle this transition. The options include (a) a minimal impact on the user side at the cost of adding more technical debt on the application side; (b) a balanced approach consisting of compromises on both sides; (c) a disruptive approach, clearing as much technical debt as possible and preparing MARS for the long term. A couple of concrete examples of MARS requests are presented in Table 2.

Stay informed

Users are encouraged to follow the progress of the migration to GRIB2. For this purpose, a mailing list has been set up: mtg2@lists.ecmwf.int. This list is intended to be used to inform users about progress and changes in the migration. Users are also invited to continue to check ecCodes release notes for a general understanding of changes and bug fixes, as many may

	Before migration	After migration
Soil temperatures in soil layers 1 to 4	param = stl1/stl2/stl3/stl4, levtype = sfc	param = st, levtype = sol, levelist = 1/2/3/4
Mass mixing ratio, total column mass density and emission mass flux for ozone and carbon dioxide	param = o3/tco3/e_o3/co2/tcco2/e_co2	param = mmr/tcmd/emm chem = o3/co2

TABLE 2 A few examples of MARS requests before and after the migration to GRIB2.

not be related to the migration. To subscribe, send an email to sympa@lists.ecmwf.int with the subject 'SUBSCRIBE mtg2@lists.ecmwf.int'. To report a problem related to the migration to GRIB2 or if you have a question about it, please follow the normal procedure and contact user support via the Service Desk.

The information which will be distributed via mailing list, as well as more details, data and code examples, can also be found on the MTG2 Confluence web page (*https://confluence.ecmwf.int/display/MTG2US/Migration+to+Grib+2+-+User+Space+Home*). You can get informed of any updates and changes by clicking the watch button on that page.

Conclusion

The migration to GRIB2 is an essential step in reaching the goals set out in ECMWF's ten-year Strategy. Naturally the migration will require adaptation, both for users and internally in various ECMWF workflows. However, it is imperative we make this change to support the future data requirements of ECMWF. In addition, the migration will bring numerous advantages, such as more detailed metadata, a more efficient compression of the data, and a more consistent encoding of parameters. Users are invited to stay informed on the migration to GRIB2 via the emailing list and web page mentioned in the previous section.

Further reading

Betke, **E.**, **T. Quintino**, **S. Smart** & **T. Wilhelmsson**, 2022: Impact of GRIB compression on weather forecast data and data-handling applications. *ECMWF Technical Memorandum*, **No. 900**. *https://www.ecmwf.int/en/elibrary/81320-impact-grib-compression-weather-forecast-data-and-data-handling-applications*.

World Meteorological Organization (WMO), 2022: Manual on Codes – International Codes, Volume I.2, Annex II to the WMO Technical Regulations: Part B – Binary Codes, Part C – Common Features to Binary and Alphanumeric Codes, **WMO No. 306**. *https://library.wmo.int/?lvl=notice_display&id=10684#*.

Automated weather data services deployed in Bologna

Sébastien Denvil, Simon Smart, Manuel Fuentes, Baudouin Raoult

I n 2022, ECMWF migrated its 450+ PiB data archive (700 PiB when counting backup copies) from Reading, UK, to its new data centre in Bologna, Italy. It did so without interrupting or delaying its normal operations. We briefly describe ECMWF's Meteorological Archival and Retrieval System (MARS), the Fields DataBase (FDB) and ECMWF's File Storage system (ECFS), and we present the transition periods of this migration. We then discuss the challenges and opportunities the migration offered for the evolution and consolidation of our operational practices.

To do so, we will need to define what software deployment automation, Infrastructure-as-Code and DevOps are, and to explain why these concepts are important for modern software development and operations. Doing this, we will clarify concepts that are often confused with each other. The article will finally reaffirm what the keys elements are that enable a rapid and reliable transition from service development to operational release.

FIGURE 1 MARS architecture, from requests, to references, files, and data. The MARS Server does not handle files but data references. When a user request is processed, the MARS Server translates it into a list of data references, which are passed to the Data Server. The Data Server translates data references into actual files and returns the data.

MARS, FDB and ECFS services

MARS and FDB are domain-specific object stores developed in-house for storing, indexing, and retrieving weather-related data in the GRIB format. Each GRIB message is stored as a field and indexed according to semantically and scientifically meaningful metadata (including physical variables, such as temperature, pressure, ...). A set of fields can be retrieved specifying a request using a specific language developed for accessing the MARS archive or FDB content. ECFS on the other hand, is ECMWF's unstructured archive, providing users with a logical view of a seemingly very large file system. It is used for data not suitable for storing in MARS. UNIX-like commands enable users to copy whole files to and from any of ECMWF's computing platforms. ECFS uses the storage hierarchy of disks and tapes within the high-performance storage system (HPSS) to store the files and a dedicated database for their associated metadata (file ownership, directory structure, etc.).

MARS is a scalable system as it decouples the physical organisation of the data from its logical organisation. The system is split in two parts. The first part, the MARS Server, contains the semantic knowledge of the data. It knows what a meteorological field is and what a forecast is. The second part, the Data Server, has a physical knowledge of the data. It knows if a piece of data is on tape, on disk or cached. Figure 1 shows the architecture of MARS and Figure 2 is an example of an object archived into MARS.

The MARS Server does not handle files but rather data references. When a user request is processed, the MARS Server translates it into a list of data references that are passed to the Data Server. The Data Server translates data references into concrete locations and returns the corresponding data. By using this design, we have a system that is independent from the underlying hardware (disks, tape libraries, etc.) and from the underlying software (such as HPSS). The data can be physically reorganised without any impact on the system. Data files are split or joined and moved from disk to tape without involvement of the MARS Server. We have learnt from experience that the fewer files we have,
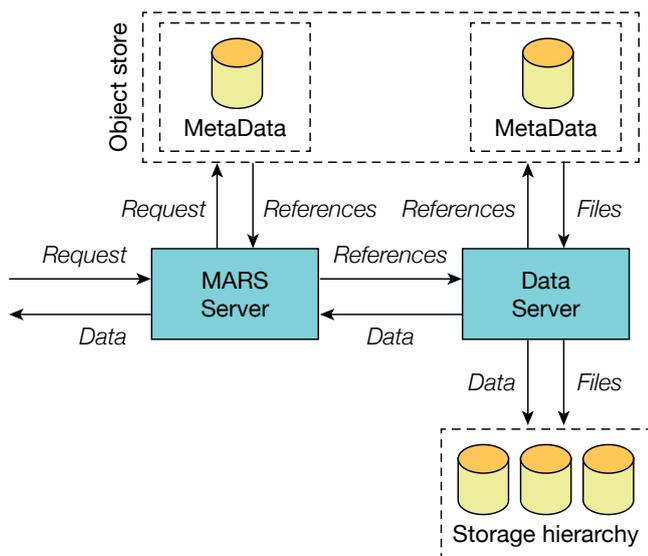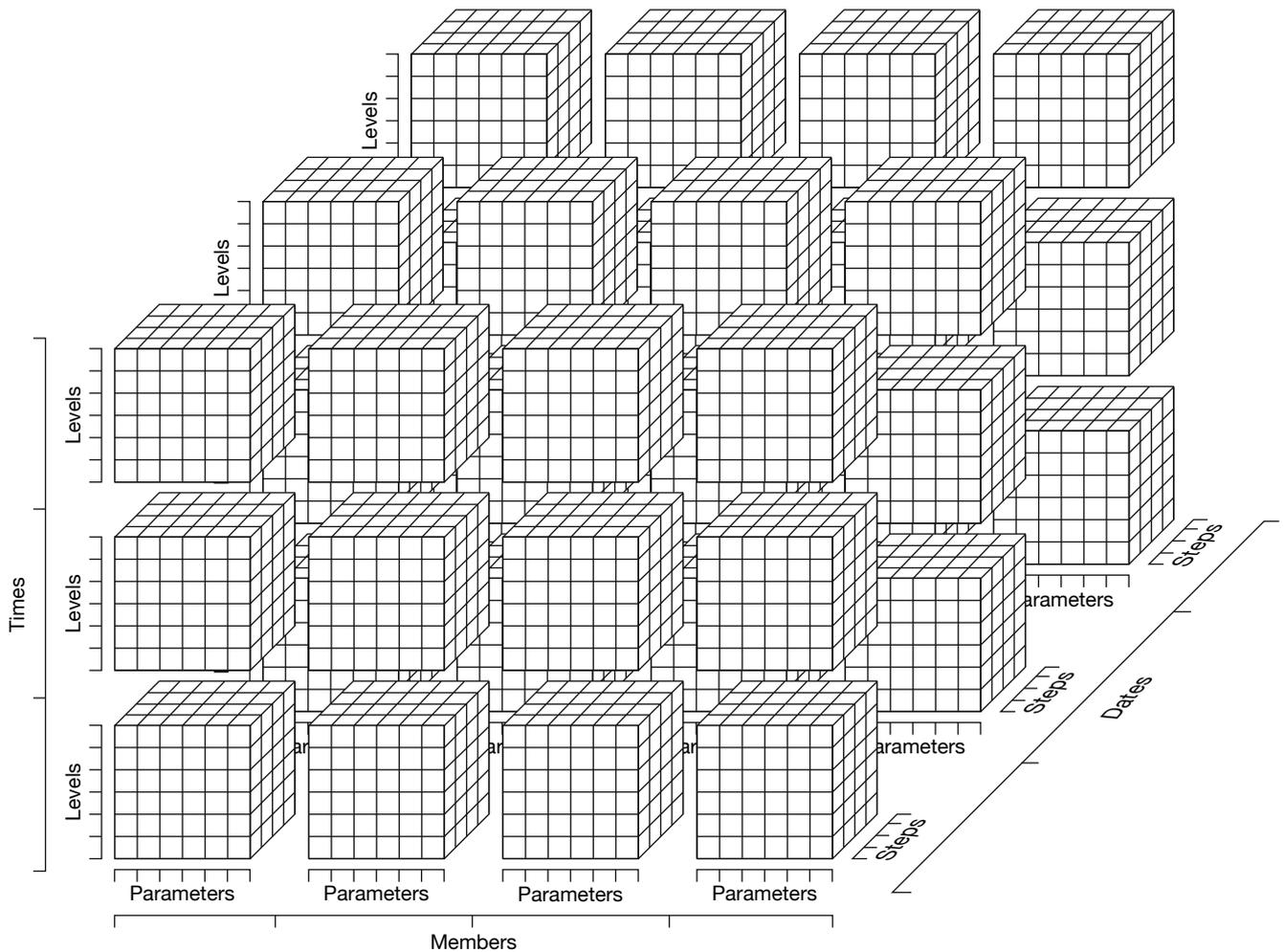
FIGURE 2 Example of an archived object in MARS, stored as hypercubes. In this example, each individual cube corresponds to a particular time, forecast date and ensemble member, and it is subdivided according to parameters, levels, and time steps.

the more manageable the system is. With this architecture, we can reduce the number of files by merging fields into larger files.

On an average day, the system handles requests for more than 13,000 tape mounts, the archive grows by about 287 TB, and 215 TB is retrieved. MARS data represents about 75% of the volume of data stored, but only about 4% of the number of files. ECFS data represents almost all the remaining 25% of the data, corresponding to 96% of the files.

Between 8 September and 11 November 2022, the complete data archive was moved from Reading to Bologna (see the Box for further information). At the beginning of the BOND (Bologna Our New Datacentre) programme, ECMWF's Technical Design Authority (TDA) was set up to manage the overarching technical governance for ECMWF in the context of BOND. An automation work stream was defined that helped shape the work that has been done and that is presented in this article. The corresponding automation journey for FDB/MARS/ECFS, together with the

challenges inherent in software-defined infrastructure and services, is discussed below.

The role of automation

Regardless of the complexity of the environment, an operations automation strategy will help improve existing processes. Automation can save time, increase quality, and reduce costs. Centre-wide approaches can help to realise the full value of automation for modern, digital operations. Centre-wide automation enables an organisation to manage complex IT environments more readily and integrate new technology and processes more effectively. ECMWF, being both a research institute and a 24/7 operational service, must position itself very carefully in this context to ensure both reliability of service and a fast movement from service development to production cycle. The move to Bologna has been used to make a significant step change in the way we automate the deployment of central services.

Automation uses repeatable instructions to replace manual work in the field of IT. Imperative and declarative forms of automation come up frequently. The distinction

a   Moving the data archive to Bologna

Through the migration of the data archive, ECMWF had to preserve its forecast, product generation and archival services, maintaining appropriate levels of service for its users. The planning and testing of a continuity plan of operations for operational forecasts, time-critical suites belonging to Member and Co-operating States, the services provided by the EU's Copernicus services implemented by ECMWF, and research activities was established. As it was both practically and financially unviable to maintain access to the full 450 PiB archive through the migration, the plan relied on exhaustive identification and traceability of the data flows involved in time-critical operations.

The figure shows the transition plan from the Data Handling System (DHS) being operated from Reading to being operated from Bologna. The plan was successfully trialled on three occasions (labelled

dry-runs in the figure): on 5 April 2022 for ten hours, from 26 to 28 of April 2022 for 48 hours, and from 28 June to 5 July 2022 for seven days. Together with key requirements being met in the Bologna data centre, the success of these trials gave the green light to the real move, which started on 8 September 2022 at 9:30 UTC.

At that point, two temporary systems in Bologna took over from the production systems in Reading. On 15 September, the final systems in Bologna, which were mostly empty at the time, were reconfigured to take over from the temporary systems. The dismantled tape libraries, disks, and servers from Reading were gradually transferred, and reassembled in Bologna. Half of the archive was accessible by 28 October, 75% was accessible by 3 November, and the complete archive was accessible by 11 November.
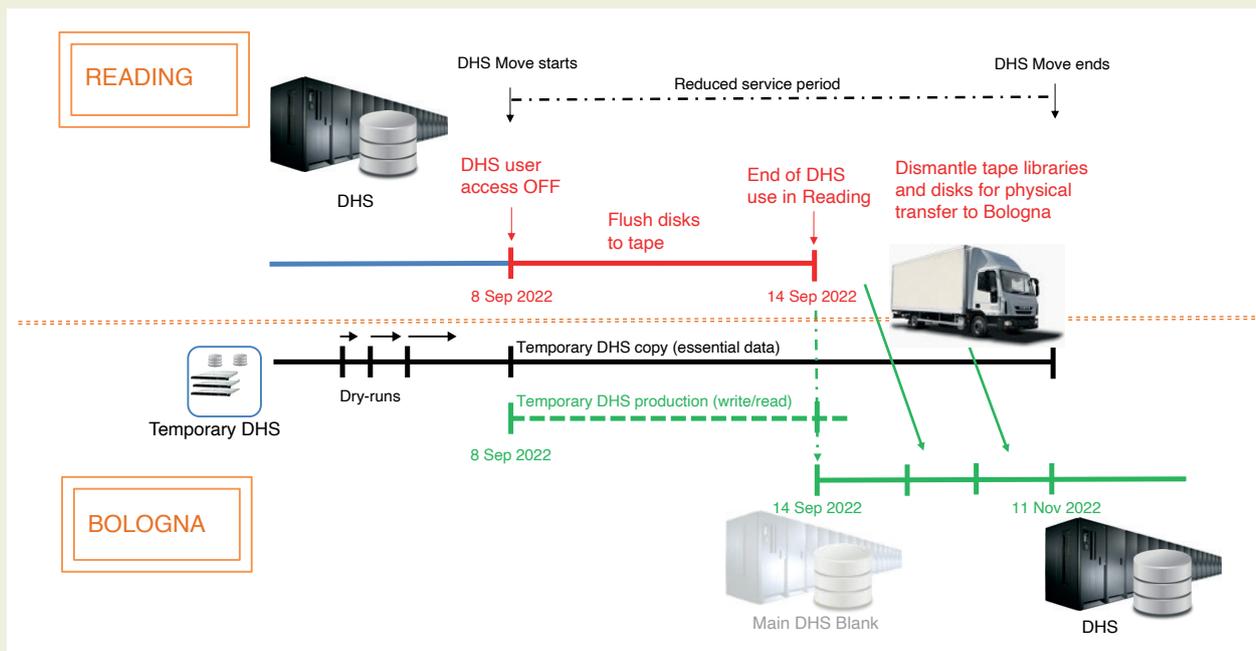
The MARS and ECFS service transition between Reading, UK, and Bologna, Italy, took place between 8 September and 11 November 2022.

between those two forms of automation is important. Both terms refer to how to provide direction to the automation software. With an imperative tool, you define the steps to execute to reach the desired solution. With a declarative tool, you define the desired state of your system, and the automation software determines how to achieve that state. In a software engineering context, declarative programming means writing code to describe what the program should do as opposed to how it should do it. An example of a declarative form

would be for you to state that a service should be in a state of running. The imperative alternative is to simply start the service.

Both have their benefits and drawbacks. Imperative languages are more focused on giving specific instructions to a computer to solve a problem. They can handle more complicated tasks and tend to be faster. Declarative languages, on the other hand, are focused on describing the end goal without worrying about the

specific steps to get there. This makes it easier to ensure that the system is always in the desired state, even if the process is repeated.

Idempotency is a key concept in automation that refers to the property of a process or task of being able to be executed multiple times without producing different results. In other words, when a process is idempotent, running it once has the same effect as running it multiple times. In the context of automation, idempotency is important because it helps to ensure that the desired state of a system is always achieved, regardless of how many times the automation tool is run. For example, if a task is designed to ensure that a particular configuration is applied to a set of servers, running it multiple times should not produce any unexpected changes if the desired configuration is already in place.

By ensuring that a process is idempotent, automation tools can reduce the risk of errors and unintended consequences, while also increasing the efficiency of the automation process. It also makes it easier to manage large-scale environments with many interconnected systems, as the same task can be run repeatedly without worrying about unexpected changes. With a declarative language having the idempotency property, you always end up in the same place, no matter where you start. In contrast, imperative language defines a series of steps that must be followed to complete a task, which can lead to different results depending on the starting point.

Ansible and Puppet are popular automation engines that have a strong track record as declarative systems, amongst many other players. Both were evaluated and used during the preparation of the move to the Bologna data centre. They are usually classified as best serving different use cases. Puppet is more of a configuration management tool, whereas Ansible is more of a provisioning, configuration, and deployment tool, which

is closer to what we are after across the variety of services ECMWF is managing.

Ansible is the automation engine now in use at ECMWF. It has been used before, during and since the completion of the move to Bologna. Ansible is an open-source, command-line IT automation software application written in Python.

Ansible has the concept of a control node and a managed node. The control node is where Ansible is executed from. Managed nodes are the devices being automated, for example a MARS mover server (see Figure 3).

Ansible's basic concepts are the following:

- **Inventory** lists and groups **hosts**

- **Playbook** contains **plays** and can run against hosts or groups of hosts

- **Plays** contain **tasks**

- **Tasks** call **modules** (Files modules, System modules, Storage modules ...)

- **Tasks** run sequentially on a given host

- **Tasks** can run on multiple hosts in parallel

- **Handlers** are triggered by **tasks** and run once, at the end of **plays (stop, start, restart)**

- **Roles** automatically load **tasks** and **handlers**. Grouping content by roles also allows **easy sharing** of roles with **other users**.

MARS, FDB and ECFS share many design and deployment concepts, and hence large fractions of their codebases that are core reusable building blocks. This is reflected in the deployment and configuration structure, with the ability of Ansible to define role dependencies and with the Ansible variable inheritance
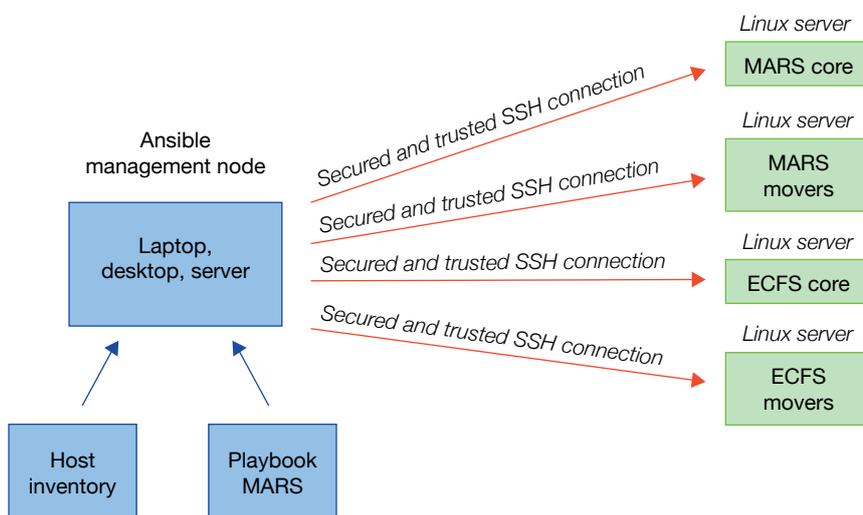
FIGURE 3 A view of the Ansible control node and managed nodes. The management node in blue is where Ansible is executed from. Managed nodes in green are the devices being automated, which are listed in the inventory, for example a MARS mover server.
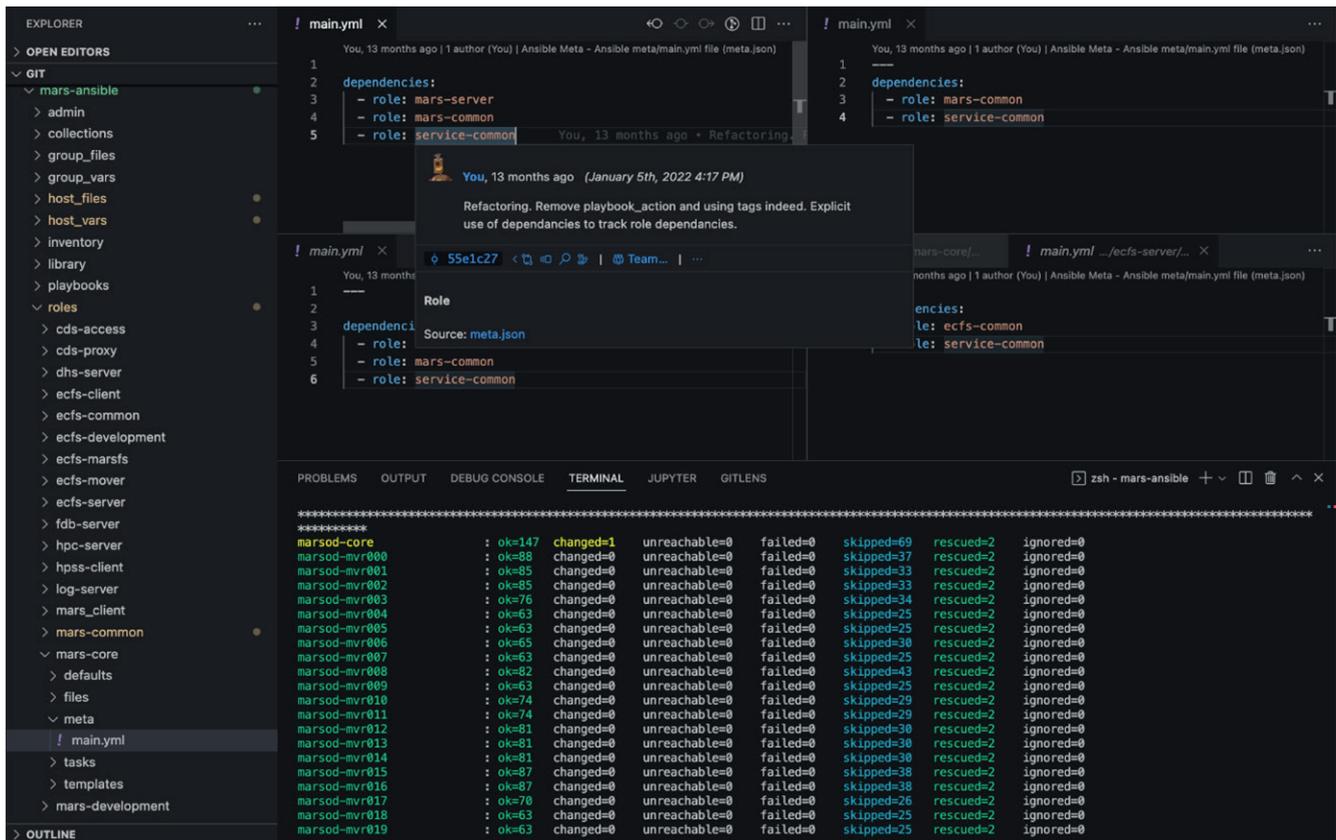
FIGURE 4 This computer screenshot shows on the left-hand side the available roles composing MARS/FDB/ECFS services and on the right-hand side some examples of the structure of MARS/FDB/ECFS automation and role dependencies.

and scope mechanism. Role dependencies and variable inheritance are essential to avoid redundancy of information.

Ansible uses variables to control how roles and tasks behave. If multiple variables with the same name are defined in different places in the code hierarchy, they override each other in a specific order. Default values can be set for included or dependent roles, and these will have the lowest priority of any variables available. They can be easily overridden by any other variable, including those supplied in the inventory or on the command line.

Figure 4 shows Ansible roles forming the building blocks of MARS/FDB/ECFS automation and their dependencies. For example, 'service-common', 'mars-common' and 'ecfs-common' roles are the foundational roles on top of which we specialise deployments and configurations. This is paramount to avoid defining variable or file content at multiple places and to ensure consistency across our deployed services.

Automation deployment structure and the underlying software guiding design principles and architecture turn out to be two sides of the same coin. The close collaboration between software developers and the

operation team has been an essential component in this endeavour during the move to Bologna.

Going beyond automation

Automation alone is not sufficient to significantly accelerate the development to production release cycle. DevOps is a methodology that emphasises collaboration and communication between software developers and operations teams to streamline the software development process and improve the delivery of software like MARS, FDB and ECFS. It is a combination of the words 'development' and 'operations'.

Automation and Infrastructure-as-Code (IaC) are both pillars of DevOps. DevOps teams use automation tools to simplify and streamline the software development process, reducing the time and effort required to deploy, test, and release software. IaC is the practice of managing infrastructure using code, rather than manual processes. It involves defining infrastructure elements, such as servers, networks, or storage, as code and then using automation tools to provision and manage those elements. IaC allows DevOps teams to manage infrastructure in a repeatable and scalable way, reducing the risk of configuration errors and increasing the consistency of infrastructure across

different environments.

Together, automation and IaC enable DevOps teams to manage software development and infrastructure in a streamlined and efficient way. DevOps seeks to break down silos between development and operations teams and encourages them to work together to ensure that software products are delivered efficiently and effectively. The goal of DevOps is to increase the speed of software development, improve the quality of software, and reduce the risk of errors and downtime.

DevOps also emphasises the use of continuous integration and continuous delivery (CI/CD) practices, which rely heavily on automation. CI/CD pipelines automate the process of building, testing, and deploying software, enabling DevOps teams to release software quickly and reliably.

In advance of the MARS, FDB and ECFS migration to Bologna, ECMWF has been intensifying its DevOps approach, culture of collaboration, automation, and continuous improvement, so that those services could be developed and delivered more quickly, reliably, and securely. Many new developments were necessary to complete the Bologna migration, and a whole new infrastructure was available for us to use. We brought together the traditionally separate functions of software development and operations into a single, integrated approach, working daily together.

Environmental drift, which refers to the configuration of a software environment changing over time, thus deviating from the desired or documented state, can be a major challenge that we wanted to avoid imperatively. This can happen due to various reasons, including manual changes made to the environment, differences in software versions or dependencies, or configuration errors in any slice of the underlying infrastructure or software dependency. In a context where numerous software code changes were made, and where a new data centre was coming to life, avoiding any environmental drift has been paramount to success. By having the environmental drift under control, we protected ourselves against inconsistent testing results, which could lead to wrong conclusions. We typically deployed MARS and ECFS software to different environments (such as development, testing, pre-production, and production environments). Each environment serves a different purpose in the software development lifecycle and has its own configuration and dependencies. Knowing exactly what changes have been made that can explain the behaviour you are observing was key to success. The DevOps approach was essential to achieving that.

The way forward

In preparation of the ECMWF move to the Bologna data centre, significant progress had been made about automation and IaC practices for the development and deployments of MARS/FDB/ECFS services. This modernisation played an important role in ECMWF being able to migrate its 450+ PiB primary data archive without interrupting or delaying its normal operations. This endeavour needs to be sustained and expanded.

We have seen that automation alone is not sufficient to provide a fast development to production cycle capability. Automation must be done using declarative language, in an idempotent (having the same effect being run once or multiple times) and reproducible way. This way it can be a foundation for further practices in which infrastructure is managed and provisioned through version-controlled code, rather than manual processes. Automation alone being the silver bullet of modernised IT practices is a myth.

Future work includes focusing on a seamless approach to MARS, FDB and ECFS deployment by further integrating and orchestrating the different slices of the data infrastructure and their dependencies: provisioning, network, servers' configuration, storage, HPSS. This will consist of a better integration of various classes of existing stress test suites, and of the integration and orchestration in an automation platform (like Ansible Tower) of the steps of our deployment workflow, to get the most out of our CI/CD pipeline.

ECMWF publications

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

Technical Memoranda

905 **Duncan**, **D.**, **N. Bormann**, **A. Geer** & **P. Weston**: Superobbing and Finer Thinning for All-sky Humidity Sounder Assimilation. *March 2023*

904 **Quintino**, **T.**, **U. Modigliani**, **F. Pappenberger**, **S. Lamy-Thepaut**, **S. Smart**, **J. Hawkes et al.**: Software Strategy and Roadmap 2023–2027. *January 2023*

ESA Contract Reports

Weston, **P.** & **P. de Rosnay**: Quarter 2 2022: Operations Service Report (Jul–22). *February 2023*

Weston, **P.** & **P. de Rosnay**: Quarter 3 2022: Operations Service Report (Oct–22). *February 2023*

Weston, **P.** & **P. de Rosnay**: Quarter 4 2022: Operations Service Report (Jan–23). *February 2023*

Weston, **P.** & **P. de Rosnay**: Annual SMOS brightness temperature monitoring report 2021/22 (Nov–22). *February 2023*

Weston, **P.** & **P. de Rosnay**: Multi-year SMOS brightness temperature monitoring (Jul–22). *February 2023*

EUMETSAT/ECMWF Fellowship Programme Research Reports

61 **Scanlon**, **T.**, **A. Geer** & **N. Bormann**: Microwave Imagers in the ECMWF–IFS: Adding further observations and improving convective anvils in the observation operator. *April 2023*

ECMWF Calendar 2023

Apr 26	Policy Advisory Committee (virtual)	Oct 4–6	Scientific Advisory Committee
Apr 27	Finance Committee (virtual)	Oct 9–12	Training course: Use and interpretation of ECMWF products
May 2–5	Training course: High Performance Computing – Atos	Oct 9–13	20th workshop on high-performance computing in meteorology
May 15–19	Training course: EUMETSAT/ECMWF NWP-SAF satellite data assimilation	Oct 19–20	Technical Advisory Committee (virtual)
May 22-26	Training course: Data assimilation	Oct 24–25	Finance Committee
May 22–26	6th OpenIFS User Meeting, Barcelona Supercomputing Center	Oct 25	Policy Advisory Committee
Jun 5–8	Using ECMWF's Forecasts	Oct 31	Advisory Committee of Co-operating States (virtual)
Jun 21–22	Council (virtual)	Nov 7	MAELSTROM Dissemination Workshop
Jun 27–30	Atmospheric River Reconnaissance Workshop	Nov 8–10	MAELSTROM Boot Camp
Sep 4–8	Annual Seminar	Nov 13–17	Training course: A hands-on introduction to Numerical Weather Prediction Models: Understanding and Experimenting
Sep 20	Final Code for Earth Day		
Sep 26	European Weather Cloud user workshop (virtual)	Nov 20–24	Training course: Parametrization of subgrid physical processes
Sep 27–28	Machine learning for numerical weather predictions and climate services - a workshop for ECMWF's Member States	Nov 27–1 Dec	Training course: Predictability and ensemble forecast systems
		Dec 7–8	Council

Contact information

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

Telephone National 0118 949 9000

Telephone International +44 118 949 9000

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.

ECMWF

Newsletter | **No. 175** | Spring 2023

European Centre for Medium-Range Weather Forecasts

www.ecmwf.int