ECMWF

European Centre for Medium-Range Weather Forecasts
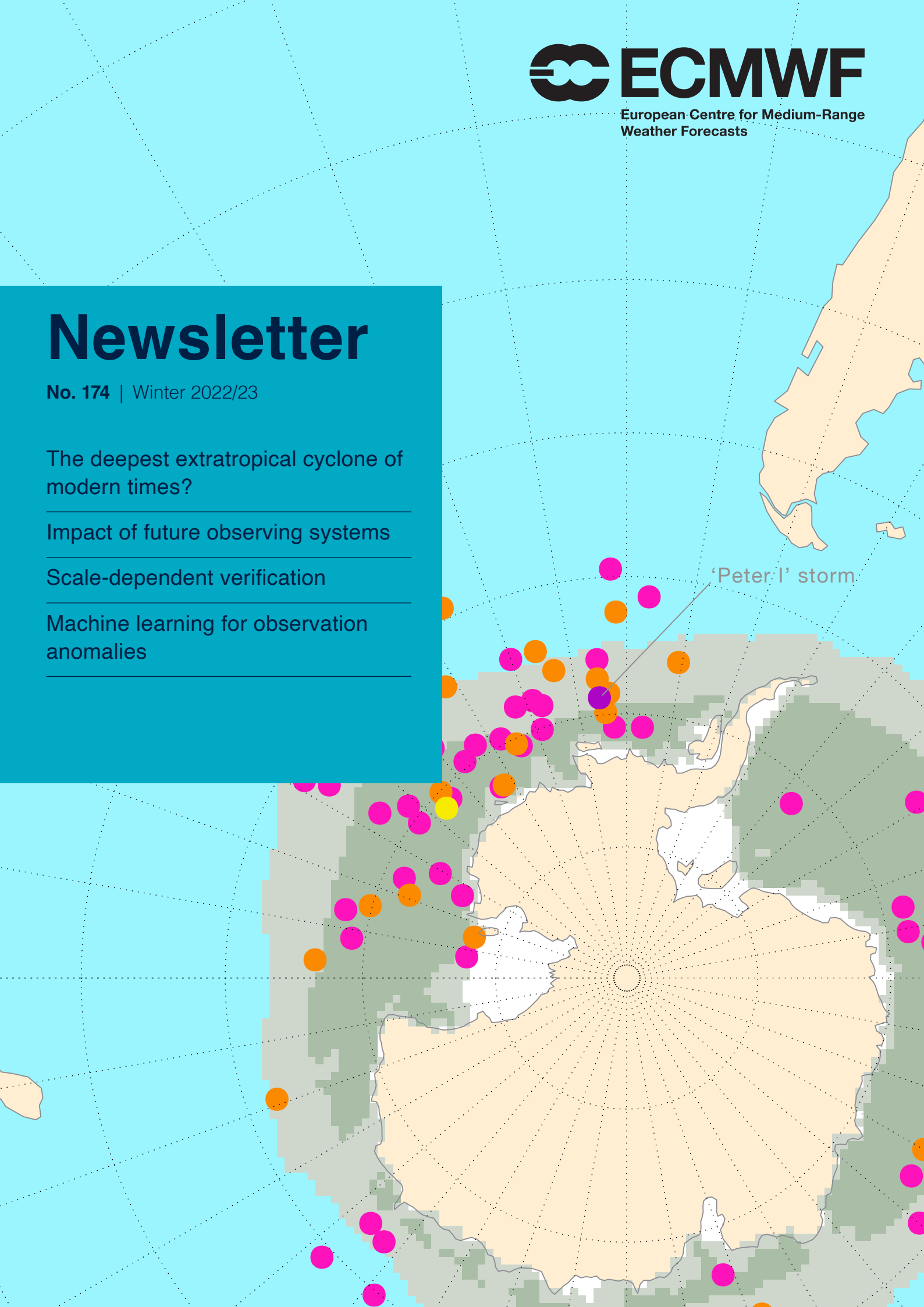
Newsletter

No. 174 | Winter 2022/23

The deepest extratropical cyclone of modern times?

Impact of future observing systems

Scale-dependent verification

Machine learning for observation anomalies

'Peter I' storm

----

editorial

Machine learning

A revolution is beckoning in weather forecasting: machine learning methods are moving into the field, and they could radically change the way weather forecasts are created in the next ten years or so. Machine learning applications are based on sample data to make predictions or take decisions. In numerical weather prediction, the traditional approach is to use the laws of physics to make predictions, although these are combined with statistical approaches for observation processing, data assimilation and post-processing. It is clear that machine learning could be useful along this whole spectrum of activities in weather forecasting. One relevant area is highlighted in this Newsletter: developing a machine learning tool for the detection and classification of observation anomalies. This application, which is still being tested, will improve the classification of events by severity and cause and help to monitor satellite data from a growing number of platforms. This is only one of the many applications of machine learning which are being investigated at ECMWF. For those interested in the potential of machine learning in numerical weather and climate predictions, we have set up a Massive Open Online Course (MOOC) on the subject that will run until April.

Looking ahead to the next 12 months, there are many more developments lined up that will take us, and the EU services in which we participate, forward decisively. One of these is the next upgrade of our Integrated Forecasting System (IFS) to Cycle 48r1, which is planned for June. This is inextricably linked to us moving from our previous Cray high-performance computing facility (HPCF) to a new Atos one in October last year. The new HPCF helps us to increase the frequency of extended-range forecasts and to include twice as many extended-range ensemble members as before. It also gives us the resources to upgrade the resolution of our ensemble forecasts (ENS) from 18 km to 9 km, which is currently the resolution of our single high-resolution forecast (HRES). This materialises the strategic objective, laid out in our ten-year Strategy in 2016, of making the ensemble our primary forecast.

Two articles in this Newsletter describe areas where we have work to do to facilitate further moves to higher resolution: observations, used to help determine the starting conditions of forecasts, should include unconventional ones so that they become as dense as possible; and our facilities to verify forecasts of precipitation and cloudiness have to be adapted to work well at higher resolution. Another article describes how we can predict the forecast impact of future observing systems that are being considered. This capacity is being extensively used in partnership with satellite agencies, such as EUMETSAT and ESA, and can be a key resource for the optimisation of the Global Observing System. Machine learning work is thus important, but it is far from the only issue to keep us busy in the new year.

Florence Rabier
Director-General

Contents

The deepest extratropical cyclone of modern times?

Tim Hewson, Jonathan Day, Hans Hersbach

On 9 October 2022, a small depression developed in the south Pacific, near Tonga, and in subsequent days this moved towards Antarctica. In so doing it developed explosively to become, so far as we can tell, the deepest extratropical cyclone of modern times. The system's central pressure fell by over 100 hPa to reach a minimum around 900 hPa at 06 UTC on the 17th, near the edge of the Bellingshausen Sea, and close to the tiny Norwegian dependency of Peter I island. Accordingly, we name this the 'Peter I storm' (Peter-the-first). Our primary data source for this study is ECMWF's ERA5 reanalysis, as low surface observation density in the areas concerned precludes observation-based analysis.

Evolution

The ERA5-based cyclone track figure shows that following genesis in the tropics, the cyclone spent a day or so (10 October) around 23°S, before progressing slowly south-southeastwards on the 11th and 12th, during which time central pressure changed little. On 13th, the low accelerated and turned towards the east-southeast, near 35°S. In tandem, isobars elongated markedly along this track orientation, with renewed development taking place at the eastern end, causing an apparent track jump between 06 and 12 UTC on the 13th. Such an evolutionary feature is not uncommon in the extratropical North Atlantic, when cyclones enter the core of the main extratropical jet. The forward development is referred to as a warm front wave. In the current case, the warm front wave then started deepening and turning poleward, also in a way that is very reminiscent of North Atlantic cyclones. As it approached the Antarctic sea ice edge on the 16th, deepening rates peaked: the maximum 6 h fall was 19 hPa. Part of the Peter I storm's

pressure fall can be attributed to migration into a broadscale annulus of low pressure 'cyclone debris' that commonly encircles Antarctica (e.g. see track figure). The lowest pressure value (900.7 hPa) occurred over sea ice, with a coincident near-surface model temperature around –10°C. In the next three days, the cyclone filled slowly and moved slowly in a loop, becoming part of the 'debris' and finally losing its identity on the 20th. This 10-day life cycle is much longer than is typically seen over the North Pacific and North Atlantic, though there are cases of (much less deep) Arctic cyclones lasting as long, particularly in summer. Even if the storm ended up being very cold at its core, latent heat input attributable to its tropical origins may have helped the pressure fall so low.

Forecasts

From six days before the event, ECMWF high-resolution forecasts
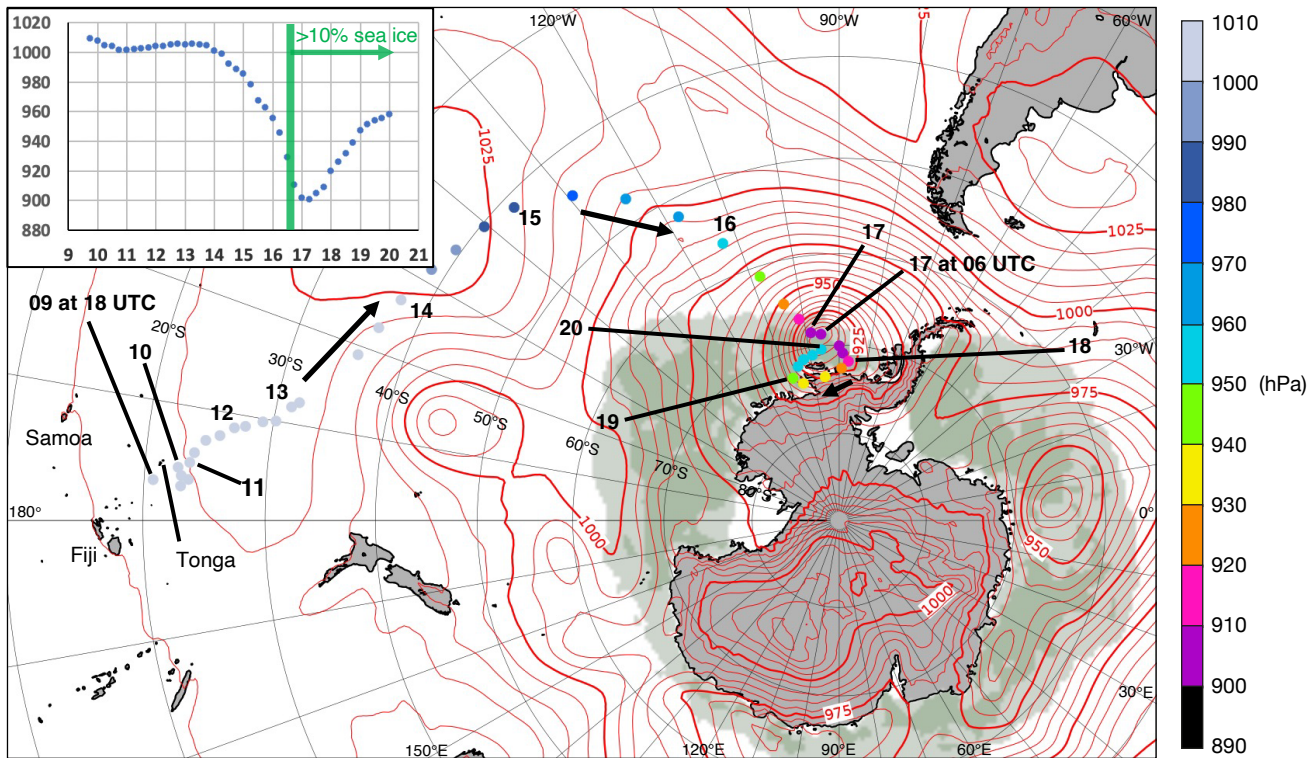
The track of the Peter I storm over its lifetime. Spots show positions from ERA5 at 6 h intervals, numbers indicate the date in October 2022; no UTC time given means 00 UTC. Contours show ERA5 mean sea level pressure at the time of maximum depth (17th, 06 UTC). Green shades denote analysed sea ice cover on 17th (10–90% and >90%). White areas adjacent to Antarctica are permanent ice sheets. The inset shows the storm's central pressure trace in hPa (y-axis) from ERA5 over the lifetime (October date on x-axis).
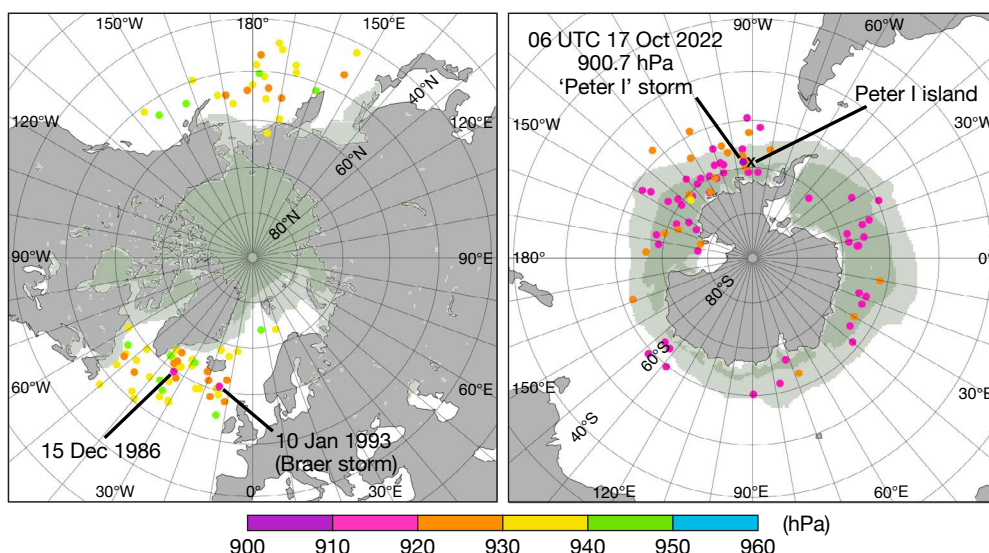
Locations of annual mean sea level pressure minima. The locations and values are shown for each year from 1950 to 2021 (annual extremes), and for 1 Jan – 31 Oct 2022, for latitudes >40°N (left), and >40°S (right), as extracted from full resolution ERA5 reanalysis data at 6 h intervals (00, 06, 12 and 18 UTC). 40° latitude limits aim to exclude tropical cyclones. Green shades denote ERA5 climatological average sea ice cover (10–90% and >90%) for 1990–2020, for 15 January (left-hand panel) and 15 September (right-hand panel).
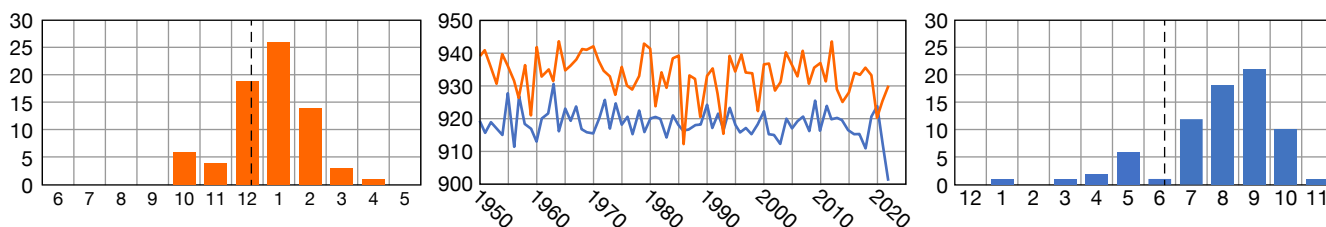
Annual mean sea level pressure minima. The charts refer to annual mean sea level pressure minima in ERA5 in the extratropical northern (>40°N, orange) and southern (>40°S, blue) hemispheres. The central panel shows the value time series (y-axis, hPa) for 1950–2022 (x-axis). The panels on the left and the right show, for the respective hemispheres, frequencies (y-axis) by calendar month (x-axis) of when the minima occurred; month orders were shifted so that the winter equinox (dashed line) lies near the centre in each case.

(HRES) were remarkably consistent in predicting a very deep low in the right general area. Pressure minima were generally in the range 895–910 hPa. This reiterates that modern-day models are perfectly capable of producing realistic forecasts of never-before-seen conditions if the initial conditions are conducive. Note also that one would expect a slightly lower minimum in HRES than in ERA5, due to its higher horizontal resolution (9 km versus 31 km).

Climatological context and physical considerations

Circumpolar plots of extreme annual surface pressure minima, for the extratropics, highlight that the southern hemisphere consistently experiences lower values than the northern hemisphere. This can be for a variety of reasons, with the very different land–sea distributions being one. The southern hemisphere sequence also shows how the 2022 event really stands out, with pressure values about 10 hPa lower than anything else in the 1950–2021 period. This gives us some confidence that

the Peter I storm was the extreme event of modern times, even if a dearth of surface pressure observations reduces our confidence in the absolute value.

It is also interesting that the vast majority of the southern hemisphere annual minima appear to have been over sea ice. They also coincide with months when sea ice is most extensive (maximum extent tends to be in mid–late September). In the northern hemisphere virtually all the annual minima are, meanwhile, over open sea water, with their peak occurrence tending to be closer to the coldest part of the year. These remarks probably have physical significance. The frictional retardation of lower tropospheric flow is less over sea ice than over stormy seas, since sea ice typically has a lower surface roughness. Indeed, over oceans there are destructive feedbacks, whereby the ocean wave growth that coincides with cyclone deepening and attendant strengthening winds ordinarily increases surface roughness. As friction tends to fill cyclones, if other things are equal greater roughness

would favour lower deepening rates and higher central pressures.

On the other hand, due to lack of experimental evidence there is considerable uncertainty in roughness lengths over sea ice, particularly if the ice has been distorted or broken up by wave action and strong winds, and this coupled with the low observation density could have conspired to make ERA5 cyclones over sea ice deeper (or indeed shallower) than they should be. This is an area where more research is warranted.

Relative to the ERA5 October climatology for 1990–2020, sea ice cover for the current storm (see track plot) was depleted west of the Antarctic Peninsula, out to about 100°W, and then more extensive beyond that; these anomalies may also have physical and/or dynamical relevance for the surface pressure evolution.

Finally, even though the Peter I storm was a standout event, the time sequence graphs do not seem to provide evidence of any long-term trends in these particular extremes.

Updating land and aerosol properties to improve reanalyses and seasonal forecasts

Magdalena A. Balmaseda, Tim Stockdale, Souhail Boussetta, Retish Senan, Gianpaolo Balsamo, Angela Benedetti, Tanya Warnaars

CONFESS is a European Horizon 2020 project initiated in 2020 that aims to provide a consistent representation of boundary forcings in reanalyses and seasonal forecasts (https://confess-h2020.eu/). It is intended to evolve the capabilities of the EU-funded Copernicus Climate Change Service (C3S), run by ECMWF, to monitor and predict extreme events and represent climate trends. CONFESS has three main strategic objectives:

- **introduce** for the first time temporal variations of land cover and vegetation in C3S systems by exploiting state-of-the-art Copernicus observational datasets

- **improve** temporal representation of tropospheric aerosols by harmonising datasets from the Coupled Model Intercomparison Project Phase 6 (CMIP6) and the EU's Copernicus Atmosphere Monitoring Service (CAMS) run by ECMWF

- **increase** predictive skill by inclusion of prognostic vegetation and new capabilities for responding to volcanic and biomass burning emissions.

Together with project partners Météo-France, Italy's Institute of Atmospheric Sciences and Climate (CNR-ISAC), and the Barcelona Supercomputing Center (BSC), CONFESS developments are being tested in a multi-model framework to assess the robustness of the results. This article reports on some of ECMWF's efforts over the past two years targeting the inclusion in the next generation of C3S reanalyses and seasonal forecasts. Beyond the immediate implementation, developments in CONFESS will prepare the ground for the rapid uptake of new Earth observations, thus supporting the continuous evolution of monitoring and forecasting systems used for operational seamless predictions and Copernicus services.

Land cover, use and vegetation

Land cover and vegetation observations are of paramount importance to properly constrain the land surface models that are included in current reanalysis and seasonal-to-decadal prediction systems.

ECMWF has produced a harmonised temporal record spanning the period 1993–2019 of Land Cover (LC), Land Use (LU) and Leaf Area Index (LAI) by
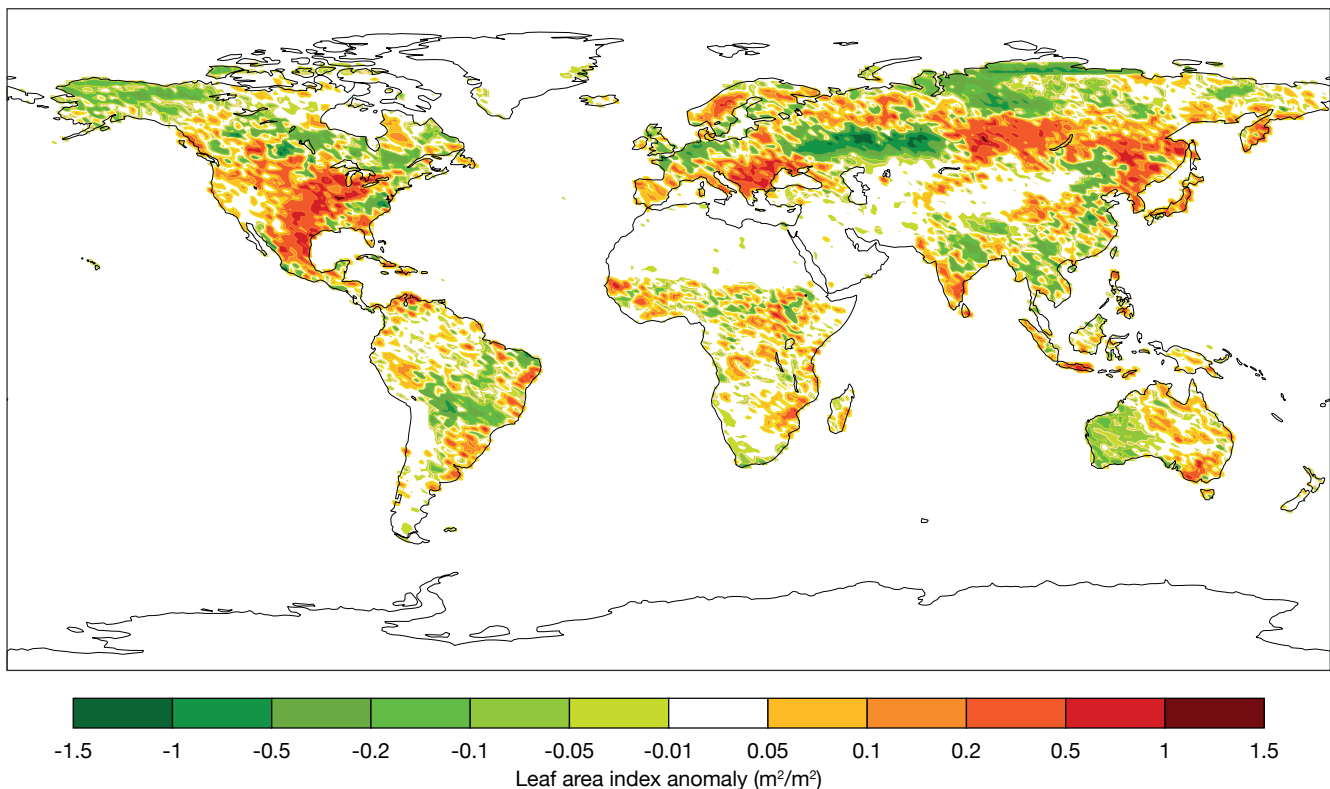
Leaf Area Index anomaly in July 2010. The map shows the leaf area index (LAI) anomaly values in July 2010 with respect to the 2009–2019 climatology. A prominent negative LAI anomaly is visible over the location of the 2010 Russian heatwave.

Black carbon, July 1975

Black carbon, July 2015

Aerosol column burden (mg/m²)
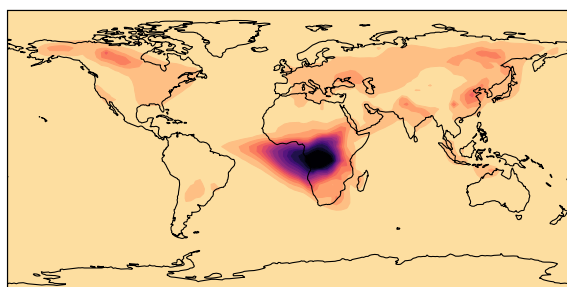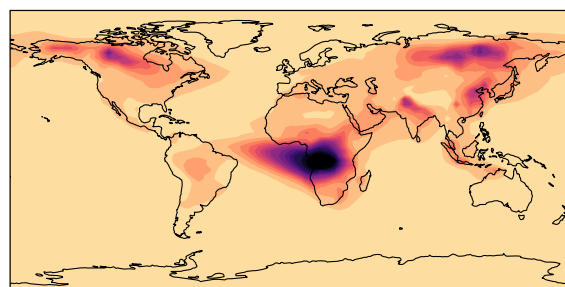
Sulphate, July 1975

Sulphate, July 2015

Aerosol column burden (mg/m²)
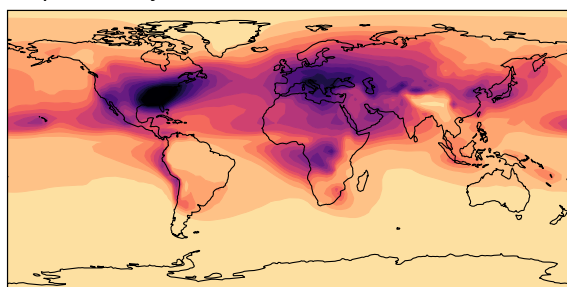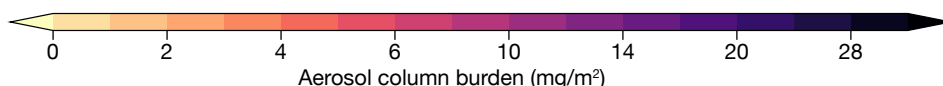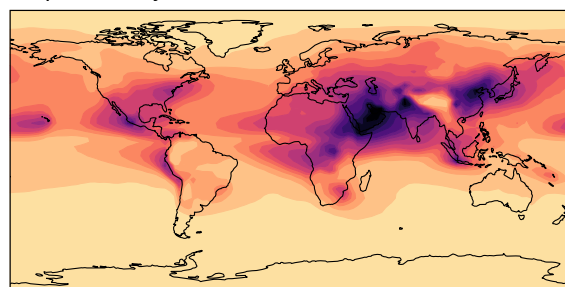
Change in tropospheric aerosol between 1975 and 2015. The image shows the change in vertically integrated black carbon (top) and sulphate (bottom) between July 1975 (left) and July 2015 (right). The effects of increased forest fires at high latitudes, pollution controls in Europe and North America, and the growth of emissions in India and the Middle East are all visible. Sulphate aerosols over China have peaked and are now declining.

merging records from C3S and the Copernicus Global Land Services (CGLS). The first figure shows an example of the anomaly in LAI in July 2010 from the harmonised datasets.

These temporal records have been implemented in different Land Surface Models (LSM), namely ECLand at ECMWF, Surfex at Météo-France and the EC-Earth HTESSEL-LPJGuess. With each LSM, pairs of offline simulations, driven by atmospheric forcing using ECMWF's ERA5 reanalysis, have been conducted with and without time-varying land properties. The sensitivity of water and energy fluxes to the temporal land variations have been evaluated and verified against available observations. The multi-model assessment makes it possible to quantify the uncertainty arising from differences in model configuration and parametrizations. Results show that anomalously low LAI conditions, such as those during the European drought in 2003 and the Russian heatwave in 2010, consistently decrease latent heat flux

and increase sensible heat flux, in agreement with observations. This response is likely to impact the estimation and prediction of extremes.

Aerosol forcing

Decadal variations of tropospheric aerosols (see the black carbon figure as an example) can have a strong impact on the production of reanalyses and seasonal re-forecasts. Currently, ERA5 and ECMWF's seasonal forecasting system (SEAS5) use tropospheric sulphate aerosol forcing from CMIP5, which is outdated and not consistent with the aerosol climatology used in numerical weather prediction (NWP). In preparation for ERA6 and SEAS6, CONFESS has produced a homogenous and consistent multi-decadal record of tropospheric aerosols, exploiting the atmospheric composition capabilities that CAMS has introduced into ECMWF's Integrated Forecasting System (IFS). A time-varying climatology of multiple aerosol species is calculated based on data from a

multi-decadal set of Cycle 47r3 IFS-COMPO forecasts constrained by ERA5 meteorology, and with continuously evolving chemistry and aerosols driven by specified emissions. So far, we have considered the periods 1971–2019 forced by CMIP6-style emission data, and 2003–2020 forced by CAMS emissions. Creating data back to 1940 will be straightforward using the newly created ERA5 reanalysis for that period. The aerosol records are then smoothed with a nine-year running mean to represent decadal variability. The resulting product has the added advantage that the last nine years can be used as a representation of the current climate aerosol values for NWP. We plan to use the resulting forcing fields in the IFS cycle used for the next generation of ECMWF seasonal forecasts SEAS6 and the upcoming C3S reanalysis ERA6. Having an up-to-date aerosol climatology that is consistent with the latest CAMS aerosols also helps us to benchmark the impact of interactive aerosols on NWP.

ECMWF–ESA Workshop on Machine Learning for Earth Observation and Prediction

Massimo Bonavita (ECMWF), Rochelle Schneider (ESA ESRIN Φ-lab)

The third edition of the ECMWF–ESA Workshop on Machine Learning for Earth Observation and Prediction took place from 14 to 17 November 2022 at ECMWF's headquarters in Reading, UK (**https://events.ecmwf.int/ event/304/**). While the first two editions of the workshop were held online due to Covid restrictions, this one ran in a hybrid format, with an in-person component (about 120 people) and a large and active online participation (about 700 registered participants). These attendance numbers, together with a record number of 121 abstract submissions, confirm both the large interest in machine learning (ML) in the Earth system sciences and the fact that the ECMWF–ESA workshops have established themselves as a reference meeting and discussion venue in this area.

One of the aims of the event was to provide an up-to-date snapshot of the state of the art in this rapidly evolving field. The two invited talks by Prof. Stephen Penny (Sofar Ocean Technologies) and Prof. Damien Borth (University of St. Gallen) set the stage, with overview presentations of the state of the art, current challenges, and opportunities for adopting AI/ML solutions in data assimilation and Earth observation. From these talks, it was apparent that increasingly sophisticated ML techniques have further spread into research and operational practice in the Earth sciences and, more importantly, they

are being tailored to this specific domain with compelling results.

Thematic focus

The workshop was structured according to separate themes designed to cover the main application areas of ML in Earth observation, numerical weather prediction (NWP), and climate prediction:

• Machine learning for Earth observations

• Hybrid machine learning in data assimilation

• Machine learning for model emulation and model discovery

• Machine learning for user-oriented Earth science applications

• Machine learning at the network edge and high-performance computing

In this edition, the last theme was added to the first four traditional areas. The aim was to encourage discussion of emerging topical areas of ML applications, such as observation processing on board of satellites (see the talk by Vit Ruzicka and poster presentations by Giacomo Acciarini and Andrea Spichtinger) and futuristic applications of AI/ML in quantum computing. In this latter area, the talks by Lisa Wörner and Bertrand Le Saux provided a fascinating snapshot of current and planned developments for the application of

these techniques to Earth observation.

Another topic at the heart of numerous presentations and discussions is the possibility that in the not-too-distant future ML tools will completely supersede foundational NWP activities, such as data assimilation, model development and model emulation. Unsurprisingly, views were varied. While most participants felt that a more likely development path would involve ML components being introduced in specific parts of the NWP value chain, others advocated a bolder approach. This would involve substituting entire NWP activities with potentially faster and cheaper ML counterparts (see for example the talks by Sid Boukabara and Stephen Rasp). What is not controversial is the fact that things appear destined to move fast as big commercial players, such as NVIDIA and Google, enter the field of ML model emulators.

Working groups

Working group discussions for different thematic areas were organised to help participants explore the main ideas emerging from the in-person and poster presentations. They also served to report on the main current and predicted trends in each of the areas. The working groups were held in a hybrid format, with both in-person and online attendees, which allowed a broad and very diverse participation. Preliminary findings from the working groups are available on the workshop

On-site participants. The event was a hybrid on-site and online event, with about 120 people attending in person.

website (https://events.ecmwf.int/event/304/timetable/), while a more detailed report is in preparation.

Outlook

We are very encouraged by the large and unabated interest in this series of workshops, the excellent level of the presentations and discussions, and the very positive feedback received from participants. This confirms that the field of ML in the Earth sciences is still on an upward trajectory. The ECMWF–ESA workshops have a significant role to play in its development and in building a strong community of developers and practitioners.

Stay tuned for the next edition!

Improvements to the Climate Data Store Virtual Assistant

Kevin Marsh, Michela Giusti, Xiaobo Yang, Anabelle Guillory

The Climate Data Store (CDS) of the EU's Copernicus Climate Change Service (C3S) run by ECMWF now has over 165,000 users and provides a wealth of climate data and information. The aim of the CDS Virtual Assistant (VA) is to help users find the answers they need as simply, quickly and efficiently as possible. The VA, or 'Knowledge Duck', was first released on the CDS on 1 July 2021. It has now been extensively improved to enhance its capabilities.

The Knowledge Duck. Located on the bottom right of CDS web pages, the Knowledge Duck can provide helpful answers to many questions from CDS users.

The user support journey and the Knowledge Duck

Various support channels are available for C3S users. These include the VA, the Knowledge Base, the User Forum, and the Jira-based service desk. These form the basis of the user support journey which was implemented in early 2022. The aim of this is to promote user self-help and direct users who are looking for information. At the end of the journey, users are still able to contact the Support team if direct assistance is needed. As shown in the user support journey figure, the VA is the first step in this process.

The VA is accessed via the Knowledge Duck icon on the CDS web interface. Here, users can type in a question, and the VA will respond with answers drawn from various information sources. These include the ECMWF parameter database and existing documentation stored on the CDS itself.

During the first six months of operation, the Data Support Team monitored user interactions and improved the responses provided. In particular, cases where conversations led to negative feedback, or the VA was unable to answer/understand ('dead ends'), were identified and addressed as quickly as possible. It was also seen that the questions asked by users rapidly increased in complexity over time, with more free-text questions being asked. These were challenging for the VA to understand and respond to correctly. These issues formed the basis for a short project undertaken by an external contractor and managed by Data Support.

The project ended in July 2022 and delivered several improvements, many of which go beyond what the previous version of the VA provided:

- A significant reduction in the likelihood of conversation 'dead ends'

- The introduction of structured flows ('guided conversations') to direct users through the content of the CDS and C3S products and services. This allows users to 'click through' to find answers to common questions

- The CDS VA interface was redesigned, to improve the look and feel so that information is more accessible to users

- Better variables and improved spatial and temporal searching; pseudonyms, locations and specific years can now be searched for, and spelling mistakes are handled in a much better way. Users are also able to filter the results of a dataset search by a number of facets, such as variable and product type

- The ability for a user to raise a Support Jira ticket directly from the VA interface. In this case, the user is prompted to enter their email address and the details of their question. If the user is not registered with ECMWF, they are prompted to register. Once it has been confirmed that the user has a valid ECMWF account, they can enter the details of their question, and a Jira ticket is automatically created. The user is provided with a link to the query, so they can follow progress

- The ability for users to leave a feedback comment, which is then

User support journey. Four stages are available for users who need support.

11,781 interactions	4,315 sessions
1 August 2022 to 30 November 2022	
Tickets created from CDS VA: 31	Top questions: CDS cache, what is ERA5, CDS variable definitions

CDS Virtual Assistant statistics. The chart summarises VA statistics from August to November 2022 following the update.

sent to the relevant user support team

• Multimedia (e.g. videos) can now be incorporated into VA responses

• The way feedback is managed has been simplified, so that a more agile service can be provided

• Behind the scenes, the VA backend was redesigned so that it now uses Botpress (https://botpress.com), which allows complex conversational flows to be implemented.

Taken together, the recent improvements to the VA mean that it is now a very capable starting point for the user support journey.

Outlook

In its first year, the VA has proved a valuable addition to the user support channels available, and the backend developments have made it more straightforward to manage and extend. In addition, we are considering having a support person behind the service for

specific periods each week, who could either interact with users directly or provide a customised response in a live user conversation. It is anticipated that VAs will also be used for the Atmosphere Data Store (ADS) of the EU's Copernicus Atmosphere Monitoring Service (CAMS) implemented by ECMWF – and ultimately the merged Climate and Atmosphere Data Store – as well as for other ECMWF systems to complement existing user support services.

The VA is an efficient and powerful way to bring information to users. With more users attracted to our data and services, our support function needs to scale up. In this way, we will be able to continue to deliver a high-quality support service to more and more users. We look forward to the continuing development of the VA over the coming years.

You can access the CDS and the Knowledge Duck at: https://cds.climate.copernicus.eu.

Summer of Weather Code becomes Code for Earth

Esperanza Cuartero, Jörn Hoffmann

The ECMWF Summer of Weather Code (ESoWC) has been engaging external coders to work on open-source developments for the last five years. Several of the software solutions developed have advanced our web services, visualisation or other services. It has been an exciting opportunity for mentors from many sections across ECMWF to work with motivated participants from around the world, who receive a stipend of €5,000 per selected project.

ESoWC has become well known in the community and is becoming a key driver of innovation at ECMWF. However, with Earth-science-related services which ECMWF runs for the EU becoming more and more involved in ESoWC, the name no longer fully reflects what ESoWC is about. Also, the community addressed by the event is expanding. We have therefore decided to continue the event in 2023 under a new name: 'Code for Earth'.

ESoWC 2022 teams. Participants in ESoWC 2022 show their certificates during the final day at ECMWF's premises in Bonn.

The programme will continue to drive Earth science innovation while fostering cutting-edge developments

to fulfil specific needs in data analysis, data management, visualisation or services across ECMWF activities.

Six innovations in 2022

ESoWC 2022 ended with a full day of presentations at ECMWF's premises in Bonn, Germany, on 28 September 2022. After two years of virtual events, the six teams selected last year could meet face-to-face with their mentors to present their final outcomes. The talks were also livestreamed. The new open-source projects focused on web development, software development, and applied data science.

Two projects concerned Copernicus activities. They were 'Adjusting climate projections' by Fiona Spuler and Jakob Wessel, who created a flexible and user-friendly toolkit for the bias correction of climate models; and a 'Wildfire emission explorer' by Giovanni Paolini and Ainhoa Murillo Iraola, who developed a graphical user interface (GUI) that simplifies the creation of wildfire emission plots allowing users to select data on demand.

Three projects referred to web development applications to help internal and external users optimise their operational environments. An 'ECMWF user dashboard' by Adarsh Narayan Pandey built on the user dashboard prototype developed initially during ESoWC 2021 through the integration of widgets from individual applications and web services. A 'CliMetLab web application' by Akshaj Verma concerns the implementation of a web-based graphical user interface (GUI) to make the configuration settings of the CliMetLab Python package easier. 'Bringing Magics weather maps to Matplotlib' by Alish Dipani improved the Python interface for Magics using the libraries Matplotlib and Cartopy. Magics is the geospatial visualisation library widely used by ECMWF and its Member and Co-operating States.

Another project, 'CW4Floods', explored the hydrological citizen data CrowdWater and its application in flood forecasting. Mohit Anand, Emiliana Myftari, Beatrice Rinaldi and Enxhi Sulkja developed a Python package to facilitate the use of crowdsourced hydrological measurements for forecast validation.

These include core activities, the EU-funded Copernicus Climate Change and Atmosphere Monitoring Services run by ECMWF, and the EU's Destination Earth initiative, in which ECMWF participates.

Time for new challenges

Now is the time for ECMWF staff to think about challenges for Code for Earth 2023. These are topics relating to ECMWF's core activities, Copernicus and Destination Earth that can be addressed by our participants. They could include solutions for web services, developing innovative visualisation or workflows, managing data and machine learning applications. Challenges are overseen by mentors from ECMWF. In the past, mentors might promote implementations they were looking for, but they also had a chance to pursue innovative approaches in their area of work, test research ideas and learn new perspectives from external talents while enjoying a fun experience.

What's next?

Code for Earth will launch on 27 February with a list of software challenges proposed by ECMWF staff in line with core activity and Copernicus and Destination Earth workflows. The projects will publish their materials on the open-source hosting service GitHub. External data scientists and open-source programmers are invited to submit an innovative and feasible proposal by 7 April.

The announcement of the selected teams on 20 April will lead to the coding phase. During this four-month period, from 2 May to 30 August, ECMWF mentors and developers will team up to work closely on the proposed software challenges.

The Final Presentation Day in September will mark the official closure of the sixth edition of Code for Earth. All developer teams will be invited to showcase their results at an in-person event – for the first time in Bologna, Italy.

Beneficial partnerships

Code for Earth 2023 will be supported by the two Copernicus services operated by ECMWF, Destination Earth, and two cloud services: the European Weather Cloud and the Copernicus DIAS service WEkEO. These partnerships help to reinforce the innovation role of Code for Earth within the meteorological, climate and atmosphere community.

Code for Earth links:

Website: **https://esowc.ecmwf.int**

Twitter: **https://twitter.com/esowc_ecmwf**

GitHub: **https://github.com/esowc**

YouTube: **https://www.youtube.com/channel/UCWLn6evyZ6tTktvUSTE1Xow**

New observations since October 2022

The following new observations have been activated in the operational ECMWF assimilation system in October – December 2022.

Observations	Main impact	Activation date
Atmospheric Motion Vectors from Himawari-9 (replacing Himawari-8)	Tropospheric wind	13 December 2022
Clear-Sky Radiances from Himawari-9 (replacing Himawari-8)	Tropospheric humidity and wind	13 December 2022

Towards using unconventional observations

Ulrike Falk, Simon Smart, Fredrik Wetterhall, Tiago Quintino, Vincent-Henri Peuch

As the complexity and resolution of global Earth system models (ESM) increase, so does the need for high-resolution data in space and time for the initialisation, verification and post-processing of models. As an example, convection-permitting models require very high-resolution initial conditions to provide a gain in forecast skill. A higher data density can be achieved by getting access to more weather stations. However, tapping into so-called unconventional or novel observations can complement conventional meteorological measurements to increase spatial and temporal resolutions. Here we describe ECMWF's involvement in projects that investigate the use of such observations.

Multiple sources

Novel observations can be collected from devices that are designed to collect meteorological observations, such as personal weather stations. Parameters include temperature, air pressure, humidity and geographical information. They can also be time

series of geophysical variables derived from different data sources, e.g. rainfall from telecommunication tower networks. Novel data sources also include the Internet of things (IoT) and crowdsourced information. There is a continuous growth in data from high-frequency, diverse, unmanaged sources. Harnessing these data has the potential to improve weather forecasts. However, it requires substantial development to combine the information from novel data with conventional observations. Challenges of these novel observations include the low reliability of data streams, uncertainties in measurement quality, and the heterogeneity of data types.

Three projects

ECMWF is involved in three European projects that have a focus on using citizen science and collecting novel observational data from private and public sensors, crowdsourced data and the IoT. The projects are I-CHANGE (Individual Change of Habits Needed for Green European

transition – a Horizon 2020 project), TRIGGER (SoluTions foR mItiGatinG climate-induced hEalth thReats – a Horizon Europe project), and AD4GD (All Data for Green Deal). The main objectives of this work are: (1) the development of new infrastructure and the adaptation of existing infrastructure for data acquisition; (2) the adaptation and establishment of data storage infrastructure; and (3) the processing of observations, including a generic quality control framework.

A core task of these projects is developing and maintaining data hubs, which will need to perform quality control, encode observations in standardised forms, and store and index the observations. The data will follow standards including Open Geospatial Consortium (OGC), FAIR data, and European Data Governance standards (see the graph for an overview of the envisioned data infrastructure). The data processing includes data quality monitoring and flags, filtering, and encoding before the data are
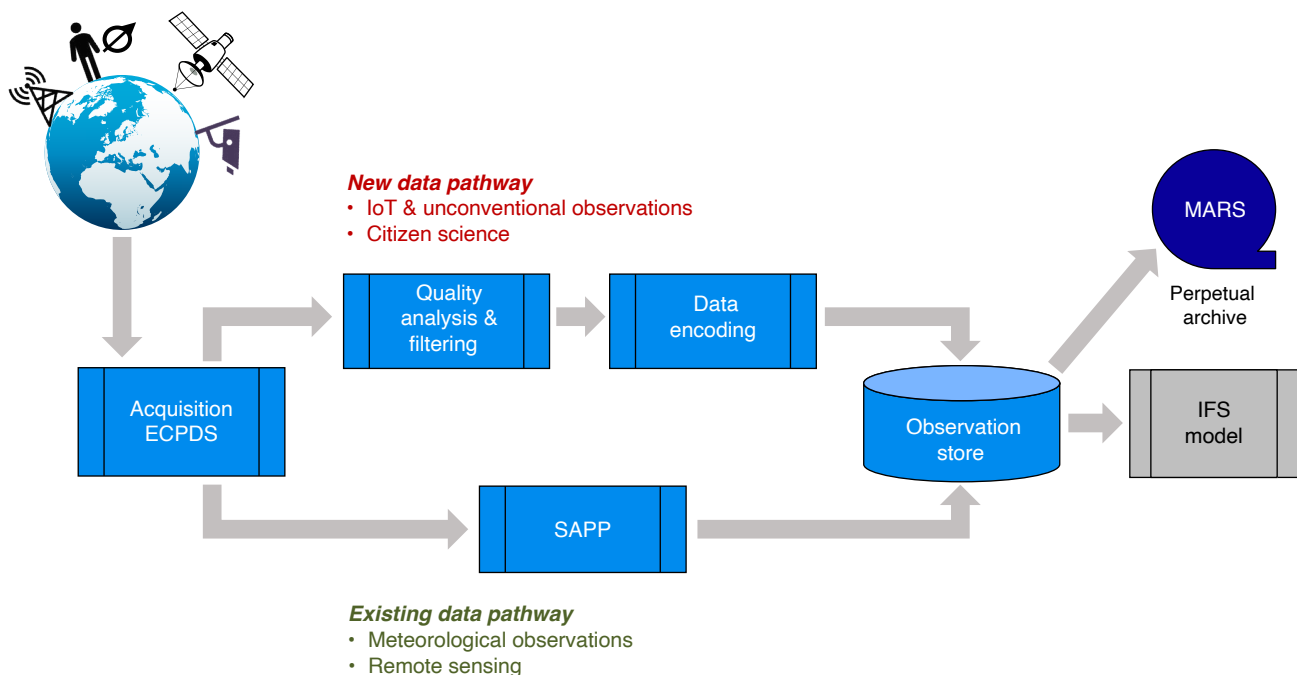
ECMWF's envisioned data infrastructure including IoT and unconventional observations. The infrastructure relies on ECMWF software such as ECPDS (ECMWF Production Data Store), SAPP (Scalable Acquisition and Pre-Processing system), and MARS (Meteorological Archival and Retrieval System) to provide initial conditions for the IFS (ECMWF's Integrated Forecasting System).

made available to users. The TRIGGER and I-CHANGE projects also share the idea of engaging citizens in the active collection of hydrometeorological, health and other data in real-life environments, so-called Living Labs (LLs). A number of these LLs are directly involved in setting up the envisioned data infrastructure to facilitate the assimilation and distribution of datasets to the general public. In TRIGGER, data containing personal health information will be randomised and anonymised already in the labs (i.e. clinics participating in TRIGGER) to protect patient confidentiality. The LLs will also be directly involved in the quality control and uncertainty assessment of the datasets before submitting them to the project's data infrastructure. Project experts are tasked with the development of tools for increasing data usability and interoperability.

These projects will pilot the handling of novel observations with the goal of exploring their potential use in operational models at ECMWF as well as in the EU initiative Destination Earth (DestinE), in which ECMWF participates. The end goal is a system that can merge novel and conventional observations into a high-resolution observational dataset with an estimated uncertainty and quality. There is still a long way to go before this becomes a reality, but these projects are a first step towards using novel observations at ECMWF.

Further information on the three projects can be found at:

https://www.ad4gd.eu
https://trigger-project.eu
https://ichange-project.eu

Register now for the Massive Open Online Course (MOOC) on Machine Learning in Weather & Climate

Discover how cutting-edge techniques impact our characterisation of weather & climate

https://lms.ecmwf.int/

ECMWF
EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS

in partnership with
iFAB
INTERNATIONAL FOUNDATION
BIG DATA AND ARTIFICIAL INTELLIGENCE
FOR HUMAN DEVELOPMENT
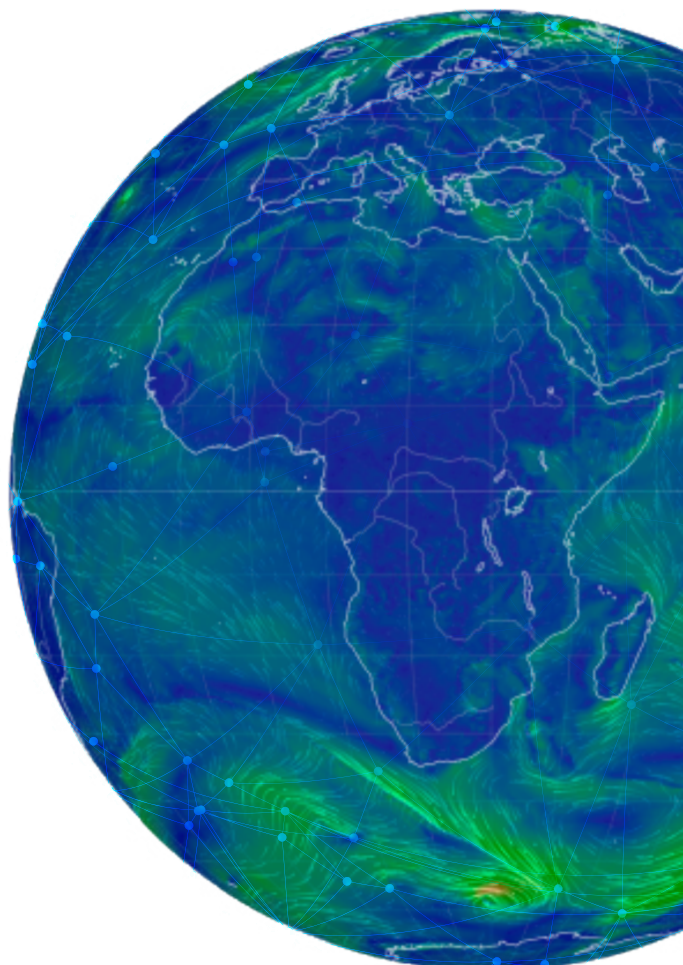
Predicting the forecast impact of potential future observing systems

Niels Bormann, Sean Healy, Katie Lean, Katrin Lonitz

ECMWF's weather predictions rely on global weather observations to help determine the initial conditions of forecasts. A key question for the evolution of the global observing system is how much benefit we expect from new observing capabilities. To answer this question, ECMWF is closely working together with space agencies, and we are increasingly using ensemble methods to estimate expected impacts from potential future satellite missions. A key tool is the Ensemble of Data Assimilations (EDA), which estimates in a statistical sense the expected reduction of uncertainty in the forecast from adding observations to our forecast system. The tool was first used 15 years ago to predict the impact of wind profile observations from the Aeolus satellite mission, which was being planned at the time. It was subsequently applied to gauge the benefit of an increase in available radio occultation data. With both of these observing capabilities becoming a reality in recent years, we are finding that the predictions from many years ago are relevant indicators of the actual impact now obtained. We can also use these new observations to further evaluate the strengths and limitations of the EDA method. Looking again into the future, the EDA method is now being used to assess the value of more passive microwave (MW) sounding observations, demonstrating continued strong benefits from an increase in the number of available MW sounders. The findings help space agencies such as the European Space Agency (ESA) and the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) to design an impactful and cost-effective constellation.

EDA method

The EDA uses a Monte-Carlo approach to estimate the size of short-range forecast error statistics. Multiple data assimilation computations are run in parallel, with perturbations added to the observations and other input parameters (e.g. sea-surface temperature) as well as to the forecast model. If these perturbations are chosen appropriately, it can be shown that the spread of the ensemble is related to the statistical uncertainty in the analysis and short-range forecasts. At ECMWF, the main role of the EDA is to provide flow-dependent uncertainty information for the short-range forecast (background) used in the operational numerical weather prediction (NWP) assimilation system and to help initialise predictions of ensemble forecasts. However, Tan et al. (2007) realised that both real and simulated observations could be included in the EDA at the same time. This means that we can assess how new, simulated observations change the EDA spread, and hence how they reduce the statistical uncertainty in analyses and short-range forecasts. The new observations are simulated from a representation of the actual atmosphere, such as the ECMWF high-resolution analysis. Perturbations are added to the simulated data to mimic expected errors in the measurements, for instance arising from instrument noise.

The EDA method is conceptually quite different from more traditional Observing System Simulation Experiments (OSSEs), which are also used to gauge the expected impact from a future observing system. OSSEs start from a long run of the forecast model, the so-called 'nature run', which simulates the true evolution of the atmosphere. All observations and their error characteristics are simulated from this nature run, and these are subsequently assimilated as if they were real observations. New observations can also be simulated from the nature run, and assimilated alongside the existing observations, and so the impact of new data can be assessed. Ensuring that all existing observations are simulated in a realistic way is a very significant effort. The EDA method has the advantage that real observations can be used alongside simulated ones, and the method directly benefits from any refinement, made for the EDA, used in the background-error modelling in the operational NWP system. While ECMWF provides nature runs to support OSSEs performed at other organisations, we currently use the EDA method for assessing new observations.

Previous impact predictions

The EDA method was first used 15 years ago in an ESA-funded study to predict the benefit expected from line-of-sight wind-profile information from the Aeolus Doppler wind lidar satellite mission. The study suggested that we would see a clear positive impact from the assimilation of Aeolus data, with particular benefits in the tropical troposphere and over oceans. This significant impact was largely confirmed a few years ago when real

Aeolus data became available, at least in a qualitative sense. The underlying observing system and the ECMWF assimilation system of course changed considerably in the meantime, making a strict quantitative one-to-one comparison impossible. But overall the impact of the real data showed many of the characteristics of the earlier predictions. This underlines that the EDA method can indeed provide useful guidance on which observations will make a big impact in the global observing system, and it can indicate where this impact will be strongest.

Another previous application of the EDA method suggested in 2013 that we would see very significant benefits from an increased number of radio occultation (RO) observations. These measurements provide primarily information on temperature and humidity at high vertical resolution, and they play an important role in controlling bias corrections for other satellite observations. At the time the EDA-based predictions were made, RO was a comparatively new observation type with relatively few observations available, and it was unclear how their impact would scale with more available observations. Surprisingly, the study indicated continued positive impact when increasing the number of RO observations per day beyond the 2,500 observations available at the time even out to 128,000 (Figure 1), with the clearest benefit for the first 20,000 profiles (see Healy et al., 2013). This gave strong motivation to invest in increasing the number of RO measurements made. In recent years, the number of real RO observations has indeed increased very significantly, particularly through the introduction of COSMIC-2 data and access to Spire observations. This has resulted in a very significant improvement in the impact of RO observations on forecasts, in line with previous predictions (see Lonitz et al., 2021).

Verifying previous impact predictions for radio occultation data

The new RO data has also allowed us to revisit some specific aspects of earlier predictions to further test the behaviour of the EDA method. Firstly, we looked at how actual short-range forecast error statistics change as the number of RO observations increases, and we compared this with EDA spread reductions. One example is shown in Figure 2, for temperatures at 100 hPa in the tropics. As the number of RO measurements increases, the short-range forecasts fit radiosonde temperature measurements more closely and the EDA spread values are reduced. This illustrates the link between spread-reduction and forecast improvement, as expected. The differences between radiosonde values and short-range forecasts are affected by measurement and representation errors, so a perfect one-to-one relationship is not expected. What is more important here is that the different impact results appear to follow a linear relationship, suggesting a strong link between the two measures. The slope of the linear relationship is larger than one, that is, the forecast errors are
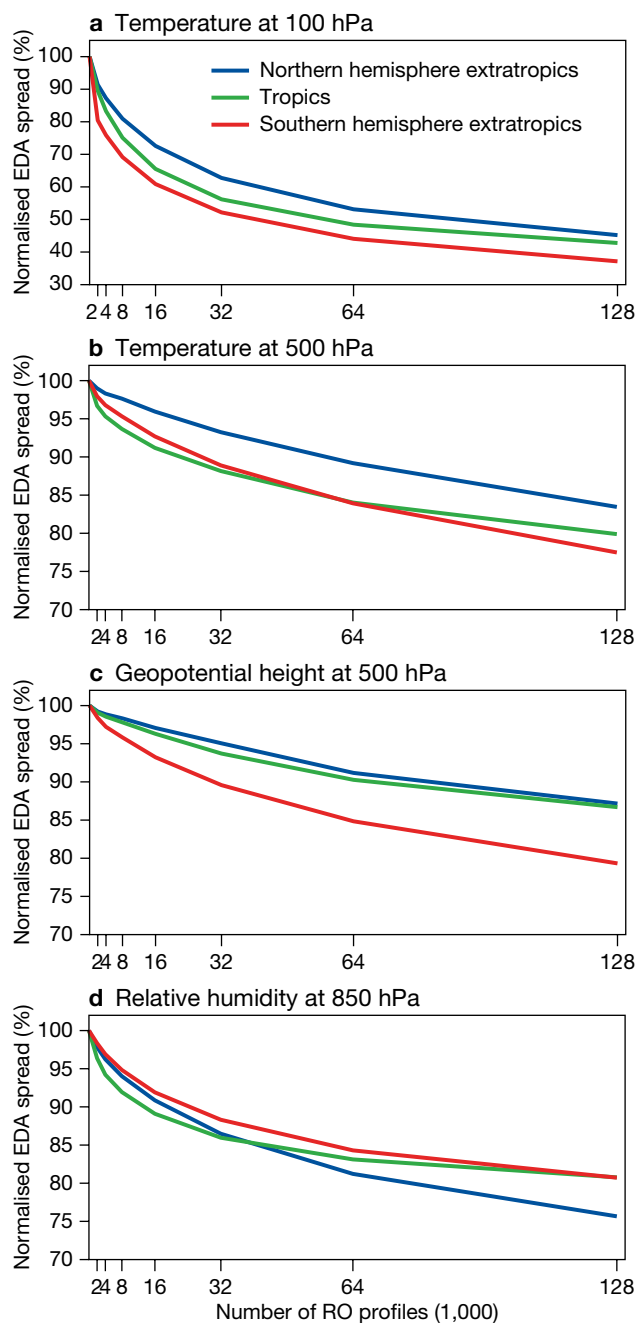
FIGURE 1 Better impact from more RO profiles as predicted in 2013. The figure shows the spread of the EDA for (a) temperature at 100 hPa, (b) temperature at 500 hPa, (c) geopotential height at 500 hPa, and (d) relative humidity at 850 hPa, normalised by the value of an EDA that does not assimilate RO data, as a function of the number of the simulated RO profiles used. The different lines indicate different geographical regions as indicated in the legend. Results are for the period 8 July 2008 to 15 August 2008. (From Healy et al., 2013)

reduced more strongly than the EDA spread values as more RO data is assimilated. This is likely related to the EDA being under-spread, an aspect that is well-established and that is actively being worked on in the context of refining our background error specification for operational NWP. The linear relationship nevertheless suggests that relative impacts appear to be adequately captured by the EDA.

A second aspect we investigated is how EDA spread reductions obtained with real RO measurements compared to those obtained with simulated observations. As real observations we used COSMIC-2 measurements provided by the US University Corporation for Atmospheric Research (UCAR). The simulated measurements have the same times and locations as the real data, but they are simulated from ECMWF analyses in a similar way as in the Healy et al. (2013) study. The agreement in the EDA spread reduction between the real and simulated COSMIC-2 data is quite impressive (Figure 3). The simulated data tends to result in slightly larger spread reductions in the troposphere and slightly smaller reductions in the stratosphere, particularly in the tropics. These small differences are still under investigation but appear to be related to the noise added to the simulated data. However, the key finding is that the EDA calculations with simulated COSMIC-2 RO provide a useful estimate of the spread reductions expected with real data.

New impact predictions: MW sounding

The EDA method is now being used in an ESA-funded study to estimate the expected impact from a potential future constellation of passive MW sounders (Lean et al., 2022). Currently, MW sounders are flown on a few large satellites, such as EUMETSAT's Metop satellites, the US National Oceanic and Atmospheric Administration
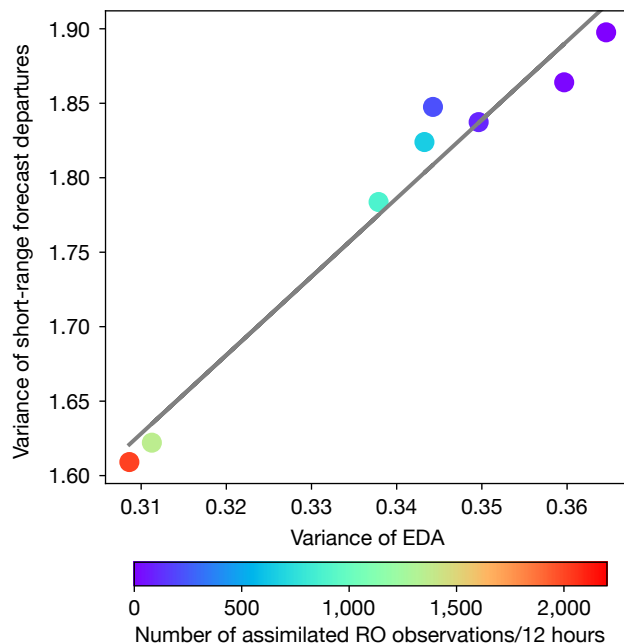
FIGURE 2 Link between EDA spread values and forecast errors as seen by radiosonde data resulting from an increased number of RO measurements assimilated. The figure shows the variance of the short-range forecast departure statistics for 100 hPa temperature measurements from radiosondes as a function of the EDA temperature variance in the tropics. The different data points are the result of varying the number of real RO observations assimilated, as indicated by the colour scale. The experiments were carried out from 10 January to 10 February 2020.
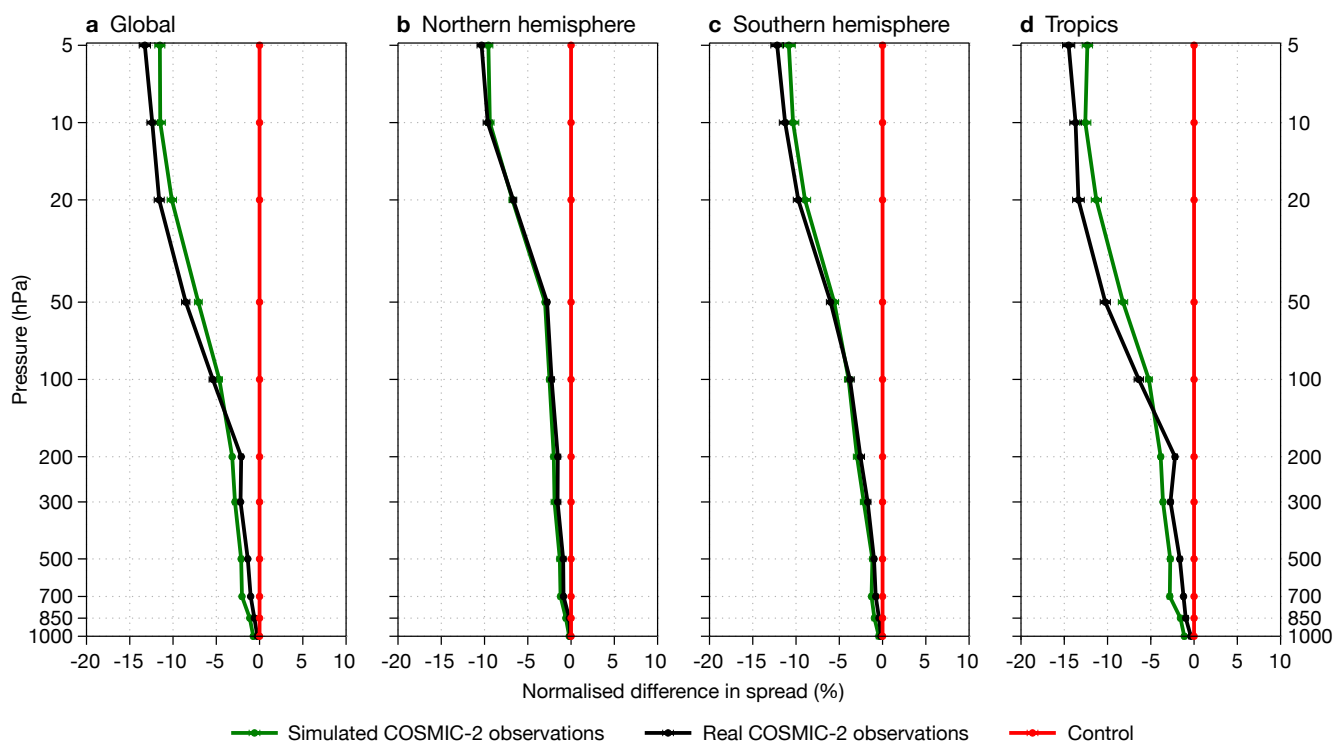
FIGURE 3 Consistency of EDA spread reductions from real and simulated observations. The figure compares EDA temperature spread reductions achieved with real COSMIC-2 measurements (black line) and with simulated COSMIC-2 data (green line). Values are given as a percentage of the spread in the control run without COSMIC-2 data. The charts show (a) global statistics computed for the ± 40° latitude band covered by COSMIC-2, (b) the northern hemisphere covering 20° to 40° latitude, (c) the southern hemisphere covering –20° to –40° latitude, and (d) the tropics covering 20° to –20° latitude.

Constellation name	Type of orbit	Number of orbital planes	Number of satellites
Real data			
Metop/JPSS baseline	Sun-synchronous	2	4 (Metop-A/B; Suomi National Polar-orbiting Partnership (S-NPP), NOAA-20)
Metop/JPSS+	Sun-synchronous	5	8 (Metop-A/B; S-NPP, NOAA-15/18/19/20, F17)
Simulated new data, added to the Metop/JPSS baseline with real data			
Polar	Sun-synchronous	4	8
Polar+	Sun-synchronous	7	14
Polar++	Sun-synchronous	10	20
4x2	Mid-inclination (60°)	4	8
6x2	Mid-inclination (60°)	6	12
Polar & 4x2	Sun-synchronous + mid-inclination (60°)	8	16

TABLE 1 Satellite constellations with MW sounding considered. In the real-data cases, combinations of the Advanced Microwave Sounding Unit-A (AMSU-A)/Microwave Humidity Sounder (MHS) or Advanced Technology Microwave Sounder (ATMS) were normally used, with the exception of the 5th-orbit in the Metop/JPSS+ constellation, for which a combination of the NOAA-15 AMSU-A and the F-17 Special Sensor Microwave Imager/Sounder (SSMIS) was used (both in an early-morning orbit during the study period). For each constellation with MW sounding data, separate EDA experiments were run with humidity-sounding channels assimilated only, and with temperature and humidity-sounding channels assimilated.

(NOAA) Joint Polar Satellite System (JPSS), and China's FY-3 satellites, providing global coverage at selected overpass times. MW sounding from core polar orbits is part of the backbone observing system identified by the Coordination Group for Meteorological Satellites (CGMS). The motivation for complementing this backbone with further MW sounders is two-fold: 1) we continue to see significant forecast benefits from adding further existing MW sounding observations to our NWP data assimilation system, prompting the question of how much further forecast benefit could be obtained from even better temporal sampling from an optimised constellation of these instruments; and 2) developments in satellite and sensor technology make it feasible to launch MW sounding instruments on small satellites or even cube-sats, thus making constellations of these instruments a possibility. But how many more satellites would be cost-effective, in what constellation should they be flown, and what sensing capabilities should they have?

To answer these questions, we have considered a number of potential constellations, designed in collaboration with ESA and other partners (Table 1). In these constellations, we explore different numbers of satellites (ranging from 8 to 20, in 4 to 10 orbital planes) and different types of orbit (sun-synchronous polar orbits as well as mid-inclination orbits, which cover latitude bands between 75°S and 75°N only). For each of these, we also considered whether the satellite instrument only includes channels for humidity sounding (around 183 GHz, complemented by channels at 89 and 165 GHz) or additionally includes temperature-sounding capabilities (in the 50 GHz band). This is an important question as it strongly affects the design of the instrument; an instrument with humidity-sounding only can be accommodated on an even smaller and more economical satellite. For our investigations, a

hypothetical MW sounding instrument was considered, consistent with being deployable on a small satellite, with characteristics broadly in line with those of the instrument considered for ESA's Arctic Weather Satellite (*www.esa.int/aws*).

Results from the EDA experimentation show a clear continued benefit from adding further MW sounders. This can be seen in Figure 4, which shows the spread reduction from MW sounder assimilation as a function of the number of sounding locations assimilated. Results are shown relative to a system in which no MW sounders are assimilated, but otherwise the full observing system is used. The different data points represent spread reductions for the different constellations considered. The data point furthest to the left represents a baseline system in which real MW sounding data from two Metop and two JPSS satellites is assimilated. The next two data points depict the spread reduction resulting from adding three existing MW sounders to this baseline, either with humidity-sounding capabilities only (red) or with temperature and humidity sounding (black). The remaining data points show spread reductions from adding instead simulated data from the potential constellations considered here, again either with humidity-sounding (red) or temperature- and humidity-sounding capabilities (black). The trend of the data points from adding simulated data extends smoothly from those using only real observations, adding further confidence in the simulation results.

The results show that even the smallest constellations considered ('Polar' or '4x2', i.e. eight satellites in four orbital planes) bring sizeable benefits, more than doubling the impact of MW sounders compared to the Metop/JPSS baseline. This aspect is further highlighted in Figure 5, which compares the spread

reduction of the two 'Polar' constellations with that achieved by four existing MW sounders. From observing systems with real data, we know that the impact of these existing MW sounders translates to gains in forecast skill of several hours, suggesting that this would be a very significant improvement.

There is a very clear benefit from having the temperature sounding channels available for the additional MW sounders. While the scenarios with just humidity-sounding show significant benefit from more observations, the benefit is even larger when temperature and humidity sounding capabilities are combined (Figure 4 and Figure 5). This result was not necessarily expected, since atmospheric humidity exhibits smaller spatial and temporal scales than temperature. Higher temporal sampling could thus be expected to be even more beneficial for humidity. It is possible in this context that our findings are underestimating the benefit of the humidity-only constellations as the observation simulations may not capture the full variability, or as the spatial resolution employed in the EDA may not be sufficient to make full use of the spatial variability (model resolution: TCo399, i.e. 25 km; final incremental analysis resolution: TL255, i.e. 80 km). In the case of the added temperature channels, it appears that the reduction in the effective noise from having multiple observations plays an important role, most likely a result of the already rather small size of our typical short-range forecast errors for these channels.

Our findings also indicate a clear benefit for wind forecasts from the assimilation of MW sounding data. This is a well-established result from Observing System Experiments with real observations, and the study shows that this impact increases further when even more data are available. The wind impact is commonly attributed to the ability of 4D-Var data assimilation to infer wind information from balance relationships as well as through tracing the evolution of humidity or cloud structures during the 12-hour assimilation window. Temperature and humidity channel sets are contributing to this wind impact in different ways, and this appears to be reflected in the geographical variations of the additional benefit that results from including the temperature-sounding channels (Figure 5): the benefit of the temperature-sounding channels for wind is largest in the extratropics and at high latitudes, enabled by geostrophic balance relationships. In contrast, the benefit of adding these channels is smaller in the tropics, probably as the humidity/cloud-tracing effect dominates. The EDA captures these different mechanisms in a physically plausible way.

One rather practical aspect encountered during the study was the role of thinning applied prior to the assimilation of MW sounding data, an aspect that becomes particularly relevant for large constellations. In line with current practice for real data, observations for the
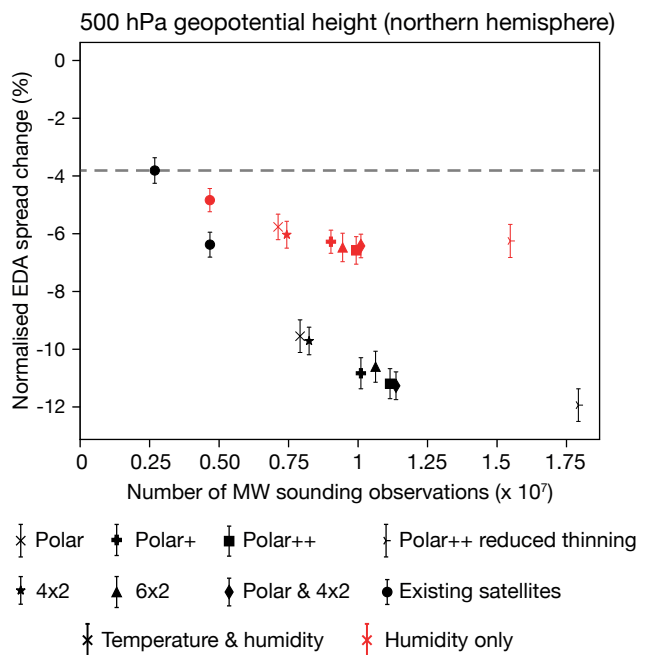

FIGURE 4 Continued benefit from adding more MW sounding data. The graphic shows the normalised EDA spread reduction relative to a MW-sounder denial EDA experiment as a function of the number of sounding locations added. Statistics are shown for 500 hPa geopotential height over the northern hemisphere extratropics for the period 8–28 June 2018. Black and red symbols indicate, respectively, the addition of temperature and humidity sounding or humidity channels only to the baseline of four existing MW sounders with temperature and humidity sounding capability (point furthest to the left). Different symbols denote the different simulated data scenarios: 'Polar', 'Polar+', and 'Polar++' use 8, 14, and 20 satellites in 4, 7, and 10 sun-synchronous orbital planes, respectively; '4x2' (eight satellites in four orbital planes) and '6x2' (12 satellites in six orbital planes) use 60° inclination orbits. The effect of adding three existing satellites to the baseline is also shown.

simulated data were thinned spatially, selecting only one observation within a 110 km distance per half-hour time-slot. This is to limit the effect of spatially correlated observation errors, which we can currently not account for during the assimilation. Data from all satellites of the hypothetical future constellation were thinned together, resulting in many observations being thinned out at higher latitudes, where there is most overlap in coverage. This is the reason why the increase in the number of data points shown in Figure 4 for the different constellations is not as large as might be expected, given the increase in the number of satellites. To investigate the role of thinning the constellation together, we also ran a pair of EDA sensitivity experiments in which we assimilated the largest constellation, Polar++, with the satellites spatially thinned separately rather than all together. This most extreme point in the number of observations in Figure 4 suggests that for using temperature and humidity or humidity-only channels, it is possible to approach a point where the benefit from additional measurement slows considerably. These findings highlight the sensitivity of the EDA results to practical assimilation choices and
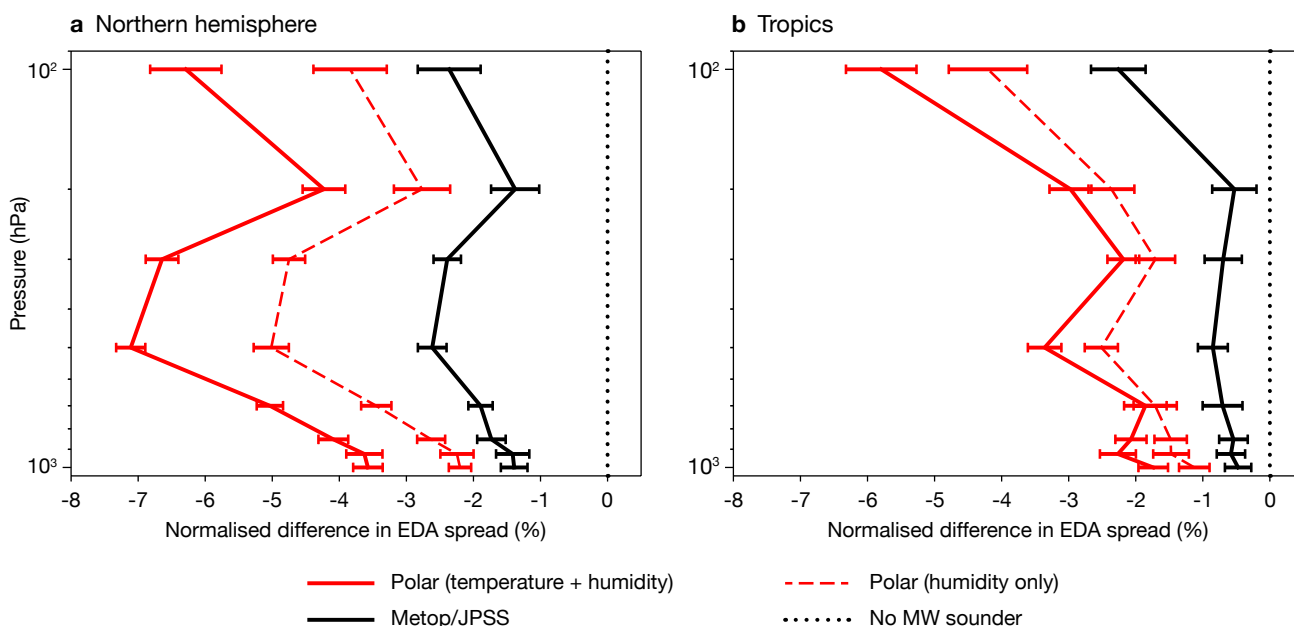
a Northern hemisphere

b Tropics

FIGURE 5 Impact on short-range wind forecasts from the 'Polar' constellation (red), with humidity sounding channels only (dashed) and both temperature and humidity sounding channels (solid line), and the 'Metop/JPSS' baseline (black), relative to a 'no MW sounder' EDA experiment. Data are over (a) the northern hemisphere extratropics and (b) the tropics, for the period 8–28 June 2018.

assumptions on error characteristics, an important element to keep in mind when interpreting the results.

Outlook

In partnership with EUMETSAT, we have now started applying the EDA method to two specific future satellite missions currently under consideration at EUMETSAT: the Sterna constellation of small satellites carrying MW sounding instruments, and a future Doppler wind lidar as a follow-on to the Aeolus satellite mission. These two mission concepts will be evaluated in a consistent way, over the same study period, and the comparability of the benefits and the synergies of the two missions will be a subject of the study. We also intend to collaborate with NOAA on the evaluation of their planned future MW sounding mission.

Our experience with using the EDA method has shown that it is a powerful tool to provide useful estimates of expected impact of future observations. We continue to actively investigate the performance of the EDA method and its strengths and limitations. However, as with any tools aimed at predicting future impacts, the results are subject to the assumptions made and they require careful analysis and interpretation, to make sound decisions about the future evolution of the global observing system. Importantly, any prediction of future impact can only reflect our current use of observations. For instance, the results for MW sounders shown here benefit from the all-sky assimilation of observations. By that we mean the use of satellite data in clear, cloudy and rainy conditions, a relatively recent innovation at several NWP centres. Results would probably have looked quite different

15 years ago, when MW sounding data were assimilated in clear-sky situations only. These aspects are important to bear in mind for future instrument decisions, particularly with the current drive to a more complete exploitation of all Earth system information in a coupled data assimilation framework. This development is expected to see significant changes in the amount and type of information that we extract from observations. Evaluations using the EDA method should hence be accompanied by further scientific considerations that take into account the untapped potential of future observations that is not yet accessible through our present use of observations.

Further reading

Healy, **S.**, **F. Harnisch**, **P. Baue**r & **S. English**, 2013: Scaling of GNSS radio occultation impact with observation number using an ensemble of data assimilations, *ECMWF Newsletter* **No. 135**, 20–24.

Lean, **K.**, **N. Borman**n, **S. Healy** & **S. English**, 2022: Final Report: Study to assess earth observation with small satellites and their prospects for future global numerical weather prediction. *ESA Contract Report 4000130590/20/NL/IA*, *doi:10.21957/kp7z1sn1n.*

Lonitz, **K.**, **C. Marquardt**, **N. Bowler** & **S. Healy**, 2021: Final Technical Note of 'Impact assessment of commercial GNSS-RO data', *ESA Contract Report 4000131086/20/NL/FF/a*, *doi:10.21957/wrh6voyyi.*

Tan, **D.G.H.**, **E. Andersson**, **M. Fishe**r & **L. Isaksen**, 2007: Observing-system impact assessment using a data assimilation ensemble technique: application to the ADM–Aeolus wind profiling mission. *Q. J. R. Meteorol. Soc.*, **133**, 381–390.

Scale-dependent verification of precipitation and cloudiness at ECMWF

Llorenç Lledó, Thomas Haiden, Josef Schröttle, Richard Forbes

As part of ECMWF's continuous efforts to improve the representation of physical processes of the Earth system, an increasing number of km-scale simulations are performed. These activities are supported by the EU's Destination Earth initiative (DestinE), in which ECMWF aims to create a high-resolution digital twin of our planet. Although higher-resolution simulations can improve the prediction of smaller-scale features and increase model fidelity, often the features' exact location cannot be precisely determined. This results in traditional metrics such as root-mean-square error (RMSE) being degraded. Spatial verification techniques can provide a better indication of the value of such forecasts than traditional verification metrics by evaluating if the forecasts have the right statistics over a certain neighbourhood. Such techniques have been used in the limited-area modelling community for some time. In this article we describe how the Fractions Skill Score is being tested at ECMWF as a first step towards a more comprehensive evaluation of performance improvements at high resolution.

Verifying high-resolution forecasts

ECMWF's current operational high-resolution global forecast (HRES) has a grid spacing of 9 km. Experimental forecasts with 4.5 km grid spacing are being generated as part of the first phase of DestinE, and even higher resolution runs have been performed in test mode. In order to evaluate the skill of surface fields at increasingly high resolution, methods beyond the standard point-wise matching of forecasts and observations need to be adopted. This is because of the so-called 'double-penalty' issue. A forecast predicting a feature with sharp gradients such as precipitation from a convective cell will be doubly penalised if it predicts the feature at a wrong space or time: once for missing the feature in the correct spot/at the right time, and once for the false alarm in the wrong place/at the wrong time (Figure 1). Hence the error will be twice as large as for a 'flat' forecast that does not predict the feature at all. Given the potential value of the forecast indicating the event, even if somewhat shifted, human judgement would consider the flat forecast as worse than the

wrong-location/wrong-time forecast, especially if the displacement is small. The interpretation of gridded forecasts to issue warnings or assist decision-making typically relies on neighbourhood analyses rather than a literal interpretation of grid-point values. Double penalty issues are especially relevant for high-resolution fields with strong gradients and sharp features, such as precipitation or cloudiness.

For a given level of imperfect association between forecasts and observations (a correlation coefficient smaller than 1), forecasts that minimise the point-wise RMSE have less variability than the observations. Therefore, a model with low variability (i.e. a lack of fidelity) will score better than a model with the right amount of variability. Based on the RMSE or related measures, it might be tempting to tune models towards lower variability. Hence, if physical realism of a model is sought, the RMSE may not be the proper metric to optimise.

One way of dealing with the double penalty issue is to compute scores for differently sized areas instead of just point-wise. A number of different spatial verification techniques have been developed (Brown et al., 2011). One of these, the Fractions Skill Score (FSS) introduced by Roberts and Lean (2008), has
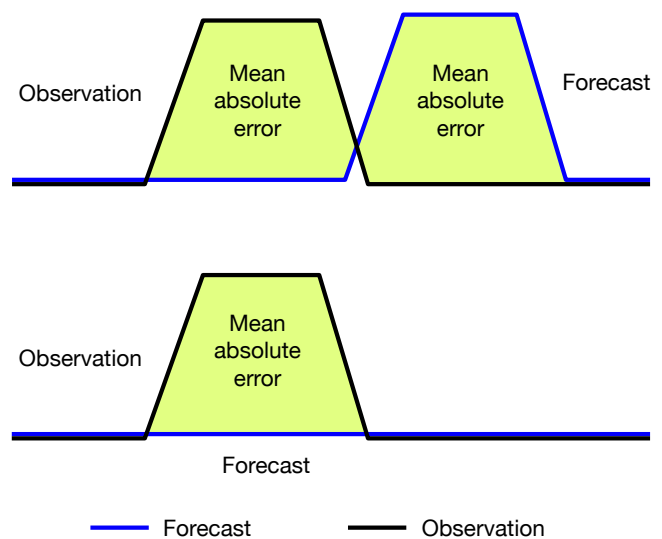

FIGURE 1 Illustration of the double penalty effect: a forecast that is able to predict an observed feature but not its exact location or time (top panel) has a higher mean absolute error (MAE) than a forecast with no feature (bottom panel).
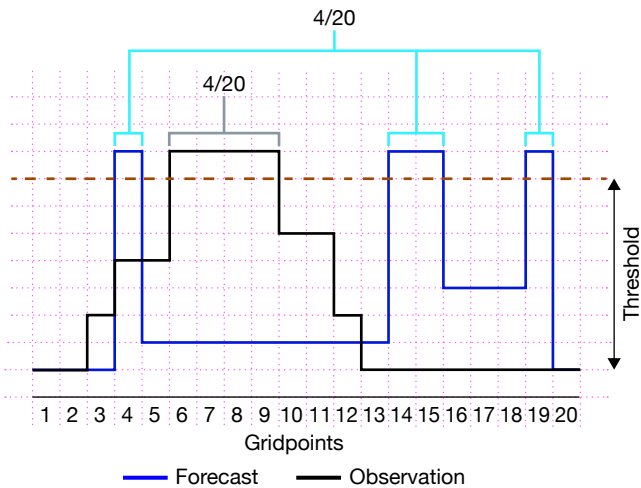
FIGURE 2 The Fractions Skill Score measures the squared difference of the number of exceedances above a certain threshold over a neighbourhood. In this example the model and observations agree perfectly on a scale of 20 grid points.

been widely used for the verification of limited-area model precipitation forecasts and is being tested at ECMWF as part of the move towards higher resolution. It gives a more complete picture of forecast performance for fields such as precipitation or cloudiness, since they may exhibit little skill at the grid scale but significant skill at larger scales. It also provides a natural framework for comparing forecasts at different resolutions, which is needed for km-scale model development and evaluation.

The Fractions Skill Score

The FSS is a spatial verification technique that does not automatically penalise location errors such as the one depicted in Figure 1. It answers the question: did the observed feature occur in a nearby location in the forecast? To do so, it counts the fraction of grid points in a neighbourhood at which a given threshold (e.g. of precipitation amount) is exceeded, both in forecasts and in observations (Figure 2). Then it compares the two fractions by computing their squared difference. The neighbourhood analysis is done separately at each grid point and then averaged over the region of interest. The Fractions Skill Score is then obtained by normalising by the worst score that could be obtained from rearranging the forecast fractions field. Analysing multiple thresholds and neighbourhood scales makes it possible to get an idea of the spatial scales at which the statistics of the number of exceedances in the forecasts match the statistics of the observations. By constraining the statistics to a neighbourhood, the FSS requires some degree of association at larger scales, while allowing for smaller-scale location and shape errors.

Properties of the FSS

Like any verification metric, the FSS has some specific

properties to be aware of. If the number of threshold exceedances in the forecast and observations are overall different due to imperfect calibration of a forecast, this will be penalised. It can be avoided by using quantile thresholds instead of absolute thresholds. Another property to note is that the FSS is not symmetric with respect to the definition of feature and non-feature. It will give different results depending on whether the grid points exceeding a threshold are counted or the ones below it. Finally, the FSS is not a traditional skill score that measures improvement over a fixed reference forecast, and although it ranges between 0 and 1, positive values do not automatically mean that the forecast is useful. According to Mittermaier & Roberts (2010) and Skok & Roberts (2016), a threshold of FSS = 0.5 may in most cases be a useful lower limit, although FSS values below 0.5 may still be regarded as useful for some applications if the forecasts are not perfectly calibrated.

Verification of HRES precipitation

Figure 3 shows the skill of daily-accumulated precipitation forecasts from the operational HRES over Europe (12.5°W–42.5°E and 35°N–60°N) for different thresholds in a winter and a summer month of 2019 (top and bottom rows, respectively). The observational dataset used for this evaluation is GPM-IMERG, which is a satellite-based, gauge-corrected gridded precipitation estimation. GPM-IMERG covers the latitudinal band between 60°S and 60°N and provides data every 30 minutes with a spatial resolution of 0.1 degrees. As a satellite-derived precipitation product, it has the advantage of providing high-resolution information across a large portion of the globe, while it shares the general weaknesses of satellite-derived estimates. Due to the indirect nature of the relationship between satellite measurements and precipitation amounts, both random and systematic errors are introduced. Gridded precipitation analyses based on rain gauges are more robust in this regard. However, they are typically available only at coarser resolution (0.25 to 0.5 degrees at most) and quite uneven in coverage. On the other hand, radar products provide information at very high resolution, but coverage is limited, and calibration procedures can be inhomogeneous across different countries' networks.

The 9 km HRES forecasts have been re-gridded to match the observations with a conservative interpolation method. Each panel in Figure 3 shows the FSS as a function of forecast lead time and neighbourhood scale from the grid-scale up to about 200 km. As expected, skill increases with scale at all lead times. Values above 0.5 (in purple) indicate that the forecast is useful at that scale. At a threshold of 1 mm in winter (leftmost panel of Figure 3a), the HRES is skilful for nearly all lead times and scales shown. In summer (leftmost panel of Figure 3b), forecast skill is lost after about 8 to 9 days,
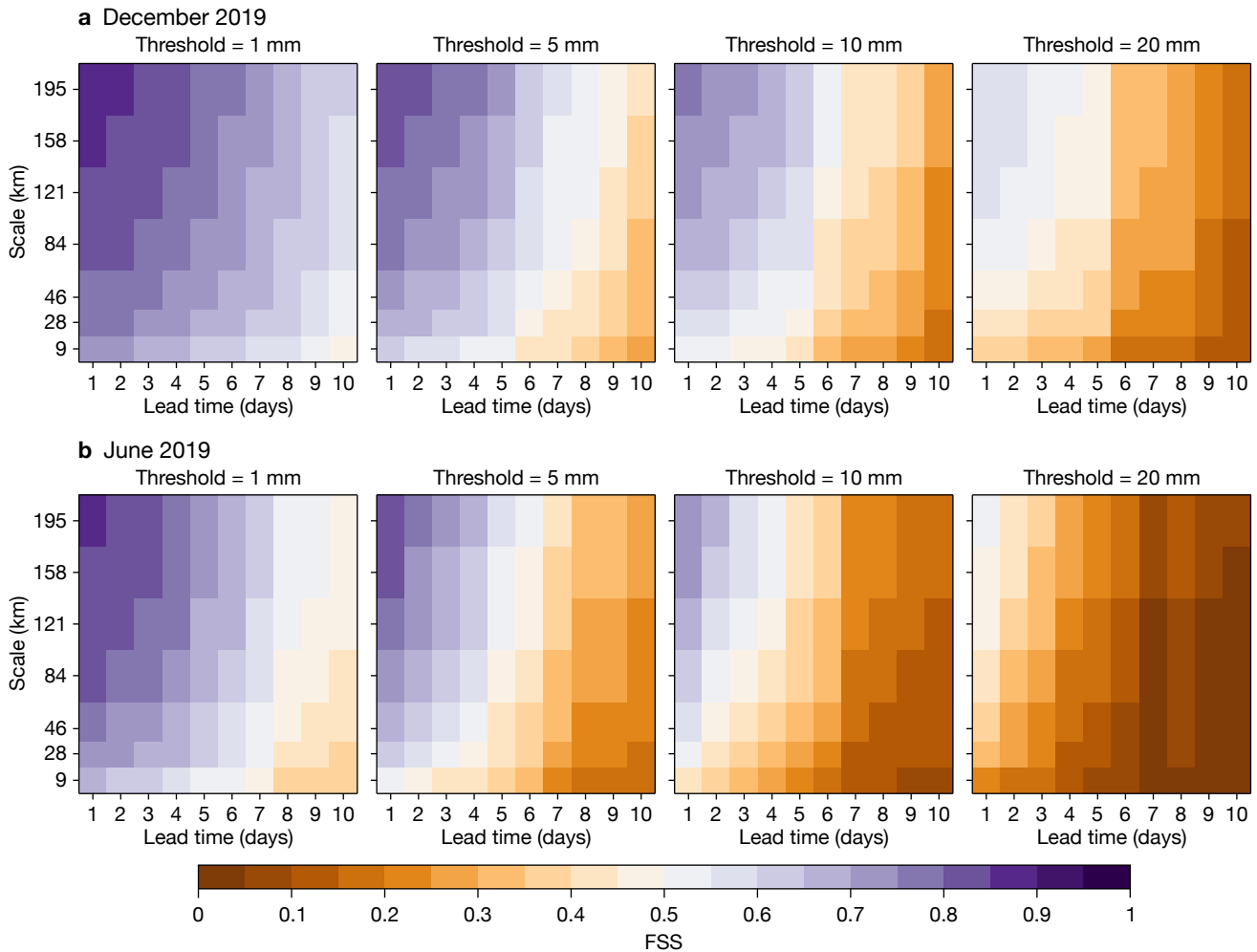
FIGURE 3 Fractions Skill Score of HRES precipitation forecasts over Europe for four thresholds, presented at multiple spatial scales and for lead times of up to 10 days ahead, for (a) December 2019 and (b) June 2019. Purple colours indicate a useful spatial scale at particular lead times.

even at larger scales of 100 to 200 km. Moving to higher thresholds, the FSS generally decreases, so that at 20 mm some skill is left only in the short range and at large scales in winter, and little skill in summer. Generally, there is a substantial gain in skill when moving from the grid-scale to the next-bigger scale (boxes with 3x3 grid points).

Using a threshold of FSS = 0.5 to determine the usefulness of a forecast is strictly applicable only if the forecast and observation datasets are unbiased. As this is not the case, even FSS values below 0.5 may be useful and provide better-than-random guidance on the spatial precipitation distribution.

Verifying radiation with the FSS

A useful proxy for the verification of cloud macro-properties, such as total cloud cover and cloud optical depth, is the downward shortwave radiation flux at the surface. Like precipitation, it is a flux quantity that can exhibit strong gradients and high variability in space and time, and it can be estimated using satellite observations. Figure 4 shows how the operational HRES forecast skill for this quantity varies with spatial scale and lead time in the summer season in the domain 60°W–60°E and 60°S–60°N. The verification dataset used is EUMETSAT's Climate Monitoring Satellite Application Facility (CM SAF) 24-hour average of downward shortwave radiation at the surface. A threshold of 200 W/m^2 has been found suitable for the domain during summer. The FSS is not sensitive to the exact value of this threshold as long as it separates predominantly clear and cloudy areas.

In terms of absolute FSS, we can see similar values (0.85 to 0.9) at large scales and short lead times as for summer precipitation with a low threshold of 1 mm (shown in Figure 3b). As for precipitation, skill becomes small beyond forecast days 7 to 8, and this limiting lead time is not very sensitive to scale.

Evaluating brightness temperature for case studies

The Fractions Skill Score is also useful for evaluating

individual cases, such as the ones currently being investigated in the Extremes Digital Twin of DestinE. The FSS makes it possible to compare the performance of different model configurations during specific events that were relevant to society due to their impacts. This type of individual case study analysis does not replace a statistical verification over longer periods but can uncover weaknesses which are difficult to detect in statistical performance summaries.

The left-hand panel in Figure 5 is a satellite image showing cloudiness during storm Ciara that affected western Europe between 7 and 9 February 2020. A large-scale synoptic trough over the North Atlantic has a clearly developed frontal cloud. The grey scale indicates brightness temperature from the infrared window band of 10.8 µm. Light colours can be used as a proxy for cloud top height, while darker colours indicate surface temperatures under clear-sky conditions. The central and right panels show simulated satellite images derived from experimental ECMWF forecasts at 9 and 4.5 km with a lead time of 48 hours. In the blue domain, a region of cold air advection, the 4.5 km experiment produces smaller-scale clouds than the 9 km run, which are prone to double-penalty issues. In the orange domain, the frontal system is wider in the 4.5 km experiment than in the 9 km run. Both experiments underpredict a narrow band of lower clouds to the southeast of the front.

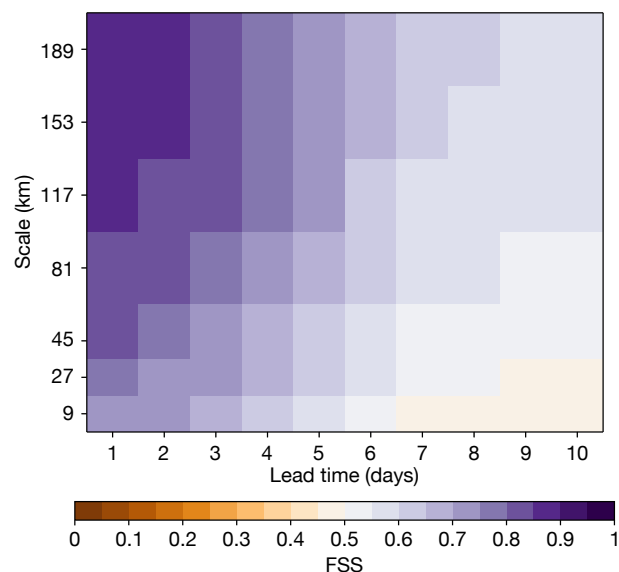Figure 6 compares the FSS of the 9 and 4.5 km experiments for the two selected regions. As these

FIGURE 4 Fractions Skill Score for HRES forecasts of daily averages of downward solar radiation in June–July–August 2022 in the domain 60°W–60°E and 60°S–60°N, verified against CM SAF data.

simulated satellite images are known to have a warm bias, the FSS has been computed for quantiles instead of absolute thresholds. These quantiles are derived from the distribution of brightness temperatures within the selected domains. For the cold air region, we can see that the FSS is below 0.5 at grid scale, and there are no clear differences between the two runs. However, when looking at larger scales, the 4.5 km run produces better
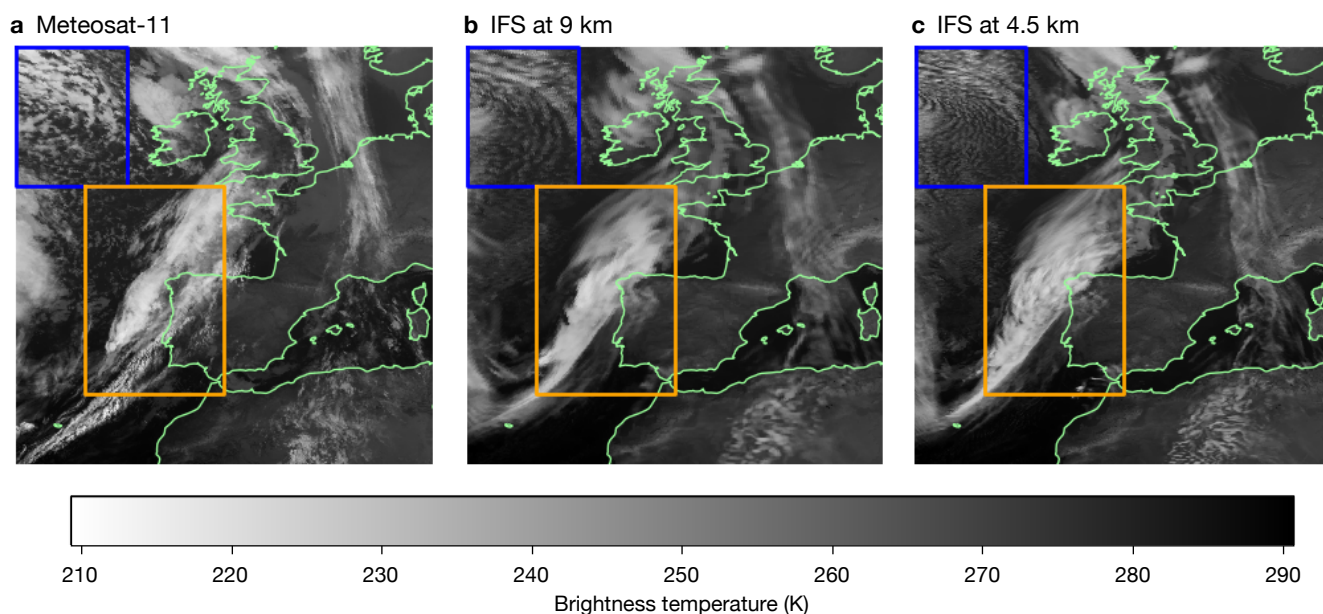
FIGURE 5 Brightness temperature images (infrared 10.8 µm band) over western Europe for the Ciara storm on 8 February 2020 at 00 UTC, (a) as observed by Meteosat-11, (b) as simulated by ECMWF's Integrated Forecasting System (IFS) at a grid spacing of 9 km, and (c) as simulated by the IFS at a grid spacing of 4.5 km, in both cases with a lead time of 48 hours. The blue box (~500 x 1,100 km) encloses a region of cold air with smaller-scale cloud structures, and the orange box (~800 x 1,700 km) contains the bulk of a frontal system.

a Cold air domain **b** Frontal system region

FIGURE 6 Fractions Skill Score of brightness temperatures for the 4.5 km and the 9 km experiments in (a) the cold air domain indicated by the blue box in Figure 5, and (b) in the frontal system region indicated by the orange box in Figure 5.

agreement with the observations for all quantiles, becoming useful already at a neighbourhood size of 5x5 grid points for quantiles 0.2 and 0.25. For the frontal region the situation is reversed: the 9 km experiment performs better than the 4.5 km one at all scales and for all chosen quantiles. In the 4.5 km run, the front is too wide when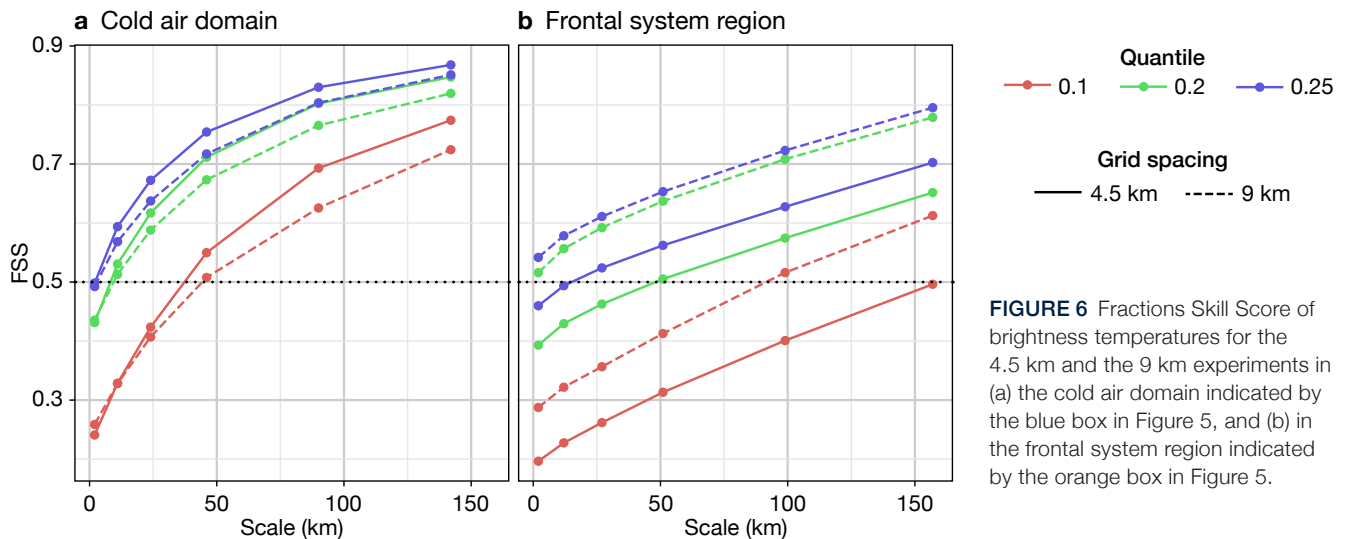 compared to observations, and this large-scale error is only compensated in the FSS when evaluating rather large neighbourhoods. This single case serves to illustrate how the relative skill in predicting cloud systems between runs at different grid spacings can vary strongly with the type of system. A more systematic evaluation of the FSS for different cloud regimes will help identify possible causes of such differences.

Outlook

The method described here to evaluate forecasts with a scale-dependent metric as an additional verification measure will be important in the future as model resolution increases. Spatial verification techniques have been developed and used for some time within the limited-area modelling community, such as in ECMWF's Member and Co-operating States, and ECMWF is benefiting from these efforts. Although those techniques are well established, new developments that simplify and facilitate the interpretation of verification results are still taking place (e.g. Skok, 2022). Another essential aspect to be improved is the availability of high-resolution, high-quality observational datasets with wide coverage for verification. There is a need to understand the associated uncertainties and potential biases of existing observational datasets used in verification. For precipitation in particular, it will be important to compare different gridded verification datasets to assess their respective strengths and weaknesses. Scale-dependent metrics are relevant for other highly variable fields, such as clouds, and

infrared and visible satellite images can be used to verify cloudiness as a function of scale. Computation of the FSS or similar scores will therefore be part of future evaluations of operational changes for the HRES as well as ensemble forecasts (ENS), and it will help to provide a more complete picture of the performance of new model cycles.

Acknowledgement

Further reading

Brown, **B.G.**, **E. Gilleland**, & **E.E. Ebert**, 2011: Forecasts of Spatial Fields. In Jollife, I.T., D.B. Stephenson (editors), *Forecast Verification*, Wiley, 95–117. *Doi:10.1002/9781119960003.ch6*

Mittermaier, **M.** & **N. Roberts**, 2010: Intercomparison of Spatial Forecast Verification Methods: Identifying Skillful Spatial Scales Using the Fractions Skill Score, *Weather and Forecasting*, **25**(1), 343–354. *Doi:10.1175/2009waf2222260.1*

Roberts, **N.M.** & **H.W. Lean**, 2008: Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. *Monthly Weather Review*, **136**(1), 78–97. *Doi:10.1175/2007mwr2123.1*

Skok, **G.** & **N. Roberts**, 2016: Analysis of Fractions Skill Score properties for random precipitation fields and ECMWF forecasts. *Quarterly Journal of the Royal Meteorological Society*, **142**(700), 2599–2610. *Doi:10.1002/qj.2849*

Skok, **G.**, 2022: A New Spatial Distance Metric for Verification of Precipitation. *Applied Sciences*, **12**(8), 4048. *Doi:10.3390/app12084048*

Use of machine learning for the detection and classification of observation anomalies

Mohamed Dahoui

For the last few years, an automatic data checking system has been used at ECMWF to monitor the quality and availability of observations processed by ECMWF's data assimilation system (Dahoui et al., 2020). The tool is playing an important role in flagging up observation issues and enabling the timely triggering of mitigating actions. The system is performing well and has a good detection efficiency. However, its behaviour is less optimal when assigning a severity level to detected events. The statistical procedure used to assign the severity requires tuning, and the behaviour is different from one kind of observation quantity to another. As a result, occasionally less significant events can be communicated as considerable or severe. When the day-to-day variability is small, moderate changes can be interpreted as severe from a statistical point of view. Given that not every threshold violation is a problem, there is a need for an improved way of inferring severity.

Another weakness of the current system is its inability to consider warnings affecting individual data types in the context of what is happening with the rest of the observing system and the type of weather activity dominating in the affected areas. Most anomaly detection tests are based on first-guess departures, i.e. the differences between a short-range forecast and observations. In these, uncertainties from observations and the short-range forecast are combined, which means that the generated warnings are not necessarily caused by observation problems. Factors causing the statistics to deviate are diverse. They require novel methods to attribute the cause and decide on the relevance of the detected event.

Machine learning techniques offer the possibility to improve the anomaly detection via a better detection of patterns, and to improve the classification of events by severity and cause. They do not need a periodic adjustment of threshold limits, either, which makes them useful for the monitoring of satellite data from a growing number of satellite platforms. As part of a wider movement at ECMWF to use machine learning operationally (see Düben et al., 2021), a new version of the automatic data checking system has been designed. It is based on an unsupervised recurrent neural network algorithm for the detection of abnormal statistics, and on a supervised learning algorithm (random forest) to classify the detected events. The automatic checking of observations is mainly used internally at ECMWF, but severe notifications are shared with selected users from EUMETSAT and the Numerical Weather Prediction Satellite Application Facility (NWP SAF) consortium. Improving the severity assignment will ensure delivered warnings are reliable. The new automatic detection framework is planned to be implemented operationally in 2023 after further testing.

In this article, we describe the design and technical implementation of the new system and how it aims to address the limitations of the current operational framework. Avenues for evolving the system are also presented.

Design of the machine learning observational data checking system

The machine learning version of the observational data checking system (Figure 1) has inherited many aspects of the current operational framework, in particular the statistics pre-processing, a set of static plausibility checks, the ignore facility, and the delivery of warnings to subscribed users. The anomaly detection module has been completely modified to rely on an unsupervised neural network algorithm to detect large deviations of statistics. This module aims to flag up sudden changes and slow drifts of statistics. The anomaly detection is performed separately for all observation types. The combined results are analysed by a supervised machine learning classifier (random forest) to adjust the severity (including a dismissal of the event), indicate the likely cause, and suggest whether action is needed. The classification results are then processed for each individual data type in order to generate relevant plots and archive warnings in an event database.

Unsupervised detection of observation anomalies

Two neural network models are applied to each individual data group to learn from the short-term behaviour (past three months) and the long-term evolution (past 12 months when available). The neural networks are autoencoders with long short-term memory (LSTM) cells. The choice of LSTM is mainly
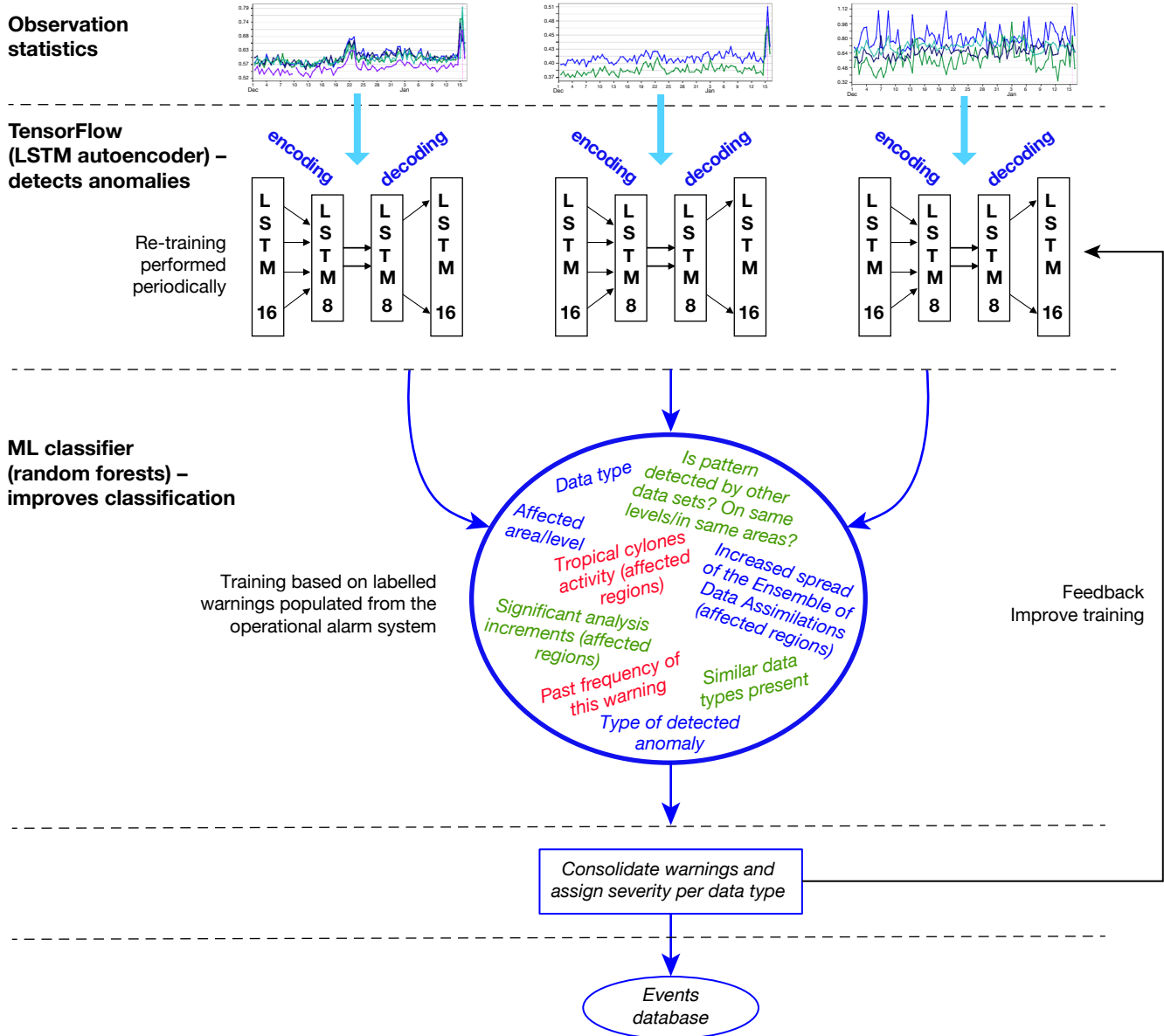
FIGURE 1 Schematic of the data checking system. The autoencoder LSTM has five layers. The first two encoding layers (with 16 and 8 units respectively) are designed to create a compressed representation of the input data. The third layer processes the compressed vector to provide input for the subsequent decoding layers, and the last two decoding layers (with 8 and 16 units respectively) aim to reconstruct the input data from the compressed representation.

intended to enable multi-feature analysis, which is useful to scale up the system in order to support large amounts of data.

The short-term model is trained every data assimilation cycle using recent statistics and excluding the last two days. The training dataset contains only statistics that are considered to be 'normal'. Previously detected events and outliers are excluded. As part of the training, we determine the resulting reconstruction error, which is conservatively chosen as the upper tail of the calculated loss in the training set. The trained model is then applied to the latest data sample (spanning the last few days) to reconstruct/predict the current statistics.
The comparison of the neural network model and actual

statistics will be larger than the reconstruction error when abnormal statistics are encountered (Figure 2). Statistics that are provided as input to the short-term model must be scaled typically between 0 and 1 based on minimum/maximum values. The scaling is necessary to ensure a better convergence of the neural network training. Some observation quantities need to be adjusted to remove periodic signals in order to avoid interpreting ups and downs as abnormal signals. For short-term models, a periodicity removal is necessary for satellite data counts due to periodic dips in counts caused by orbital movements or routines operations. For long-term models, it is important to remove seasonal periodic signals affecting random errors of departures and bias correction.
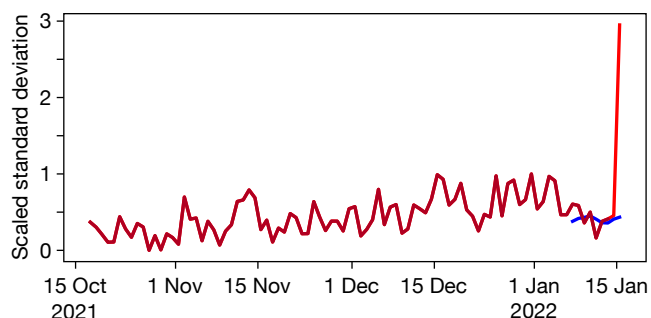
FIGURE 2 Time series of scaled standard deviation of observations minus first-guess forecasts (background departures) for AMSU-A channel 11 on Metop-B satellite over the tropics. Actual statistics are shown in red, predicted statistics are shown in blue.

The number of training epochs is set high to ensure convergence. However, an early stopping mechanism is used to avoid overfitting and reduce the training time.

The aim of the long-term trained model is to detect a slow drift of statistics. The model is trained once every quarter using the past 12 months of statistics (if available). To speed up the training and smoothen day-to-day variability, the data are sampled over periods of ten days. As part of the training, we determine the resulting reconstruction error, which is chosen as the upper tail of the calculated loss in the training set. The trained model is then applied to the latest data sample (spanning the last few weeks sampled every 10 days) to reconstruct the current statistics. Large differences between reconstructed statistics and observed ones indicate a significant change compared to long-term behaviour. Such a change can take the form of a step change (due to a model upgrade) or a slow drift of the observation quantity being monitored. The main interest is to detect a slow drift of statistics. This is achieved thanks to a monotonic slope detection algorithm applied to cases flagged up by the neural network. If the slope is not monotonic, the event is discarded.

The distribution of both neural network reconstruction errors is used to define an initial estimation of the event severity, by deriving a Z score. The classification module will adjust or consolidate these attributes.

The data checking is applied separately to individual data groups. The grouping represents the desired granularity of observation quantities to be checked. For satellite data, a group represents an observation quantity (such as the standard deviation of observations minus first-guess forecast) from a specific channel (or a pressure layer) over a specific geographical area. For some satellite data, additional dimensions are considered, such as the orbital mode (ascending/descending orbits), phase (for GNSS Radio Occultation measurements), or wind type (for atmospheric motion vectors). For in-situ data, the data groups are similar to satellite data for area-based statistics. In addition to area-based statistics, the data checking is monitoring observation quantities for each individual station, leading to a large number of items to check. The training of neural networks is fast for each data group, but when done sequentially for all items (for each observation type), the training can be very time-consuming. The LSTM ability to use multi-features enables a more efficient way of performing the training. Each data group (such as the data count from channel 16 from the IASI instrument on EUMETSAT's Metop-B satellite over the tropics) is considered to be a feature. A multi-feature vector is constructed from a large number of data groups. Such a structure enables the training to be done efficiently at once for a multitude of groups. Although features are considered together, the neural network can learn the behaviour of each data group. This enables the possibility to detect anomalies affecting one data group and not any other. The multi-feature vector is nevertheless constructed from data groups that are likely to be correlated (e.g. because they originated from the same satellite or from surface stations in the same country), which enables the detection of events affecting the whole group.

The use of neural networks to detect anomalies without the need for periodic adjustment of threshold limits is important to efficiently monitor the evolving number of satellite data. Data providers are planning to fly constellations of small satellites to provide weather data. This is expected to significantly increase the number of satellite platforms to monitor. In-situ data are also expected to increase in number and diversify due to the emergence of crowdsourced data and the inclusion of national second-tier observations.

Supervised classification of detected anomalies

Once the anomaly detection has been performed separately for all data types, all detected events are grouped together in a warnings basket. Each event is then augmented by a list of additional features reflecting common events from other data types, significant weather conditions, and the number of past occurrences of the event. A machine learning classifier (random forest) is then applied to define attributes of the detected warnings. These include false alarm (yes/no), slight event (yes/no), considerable event (yes/no), severe event (yes/no), cause (data/other) and action required (yes/no). The machine learning classifier has been trained using a population of previously generated warnings from the current operational system. The training set has been labelled to define the target attributes. Through the training process, the system is expected to learn the rules that lead to labelling decisions based on event attributes (see Figure 3).
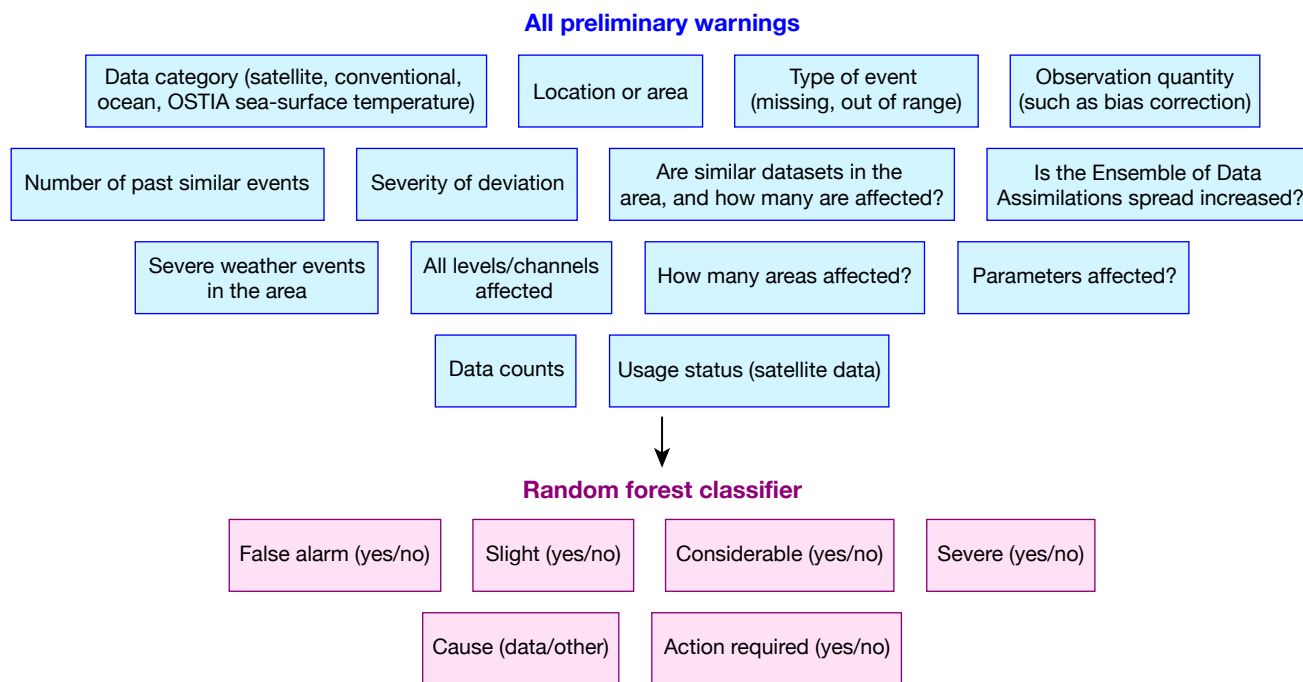
All preliminary warnings

Data category (satellite, conventional, ocean, OSTIA sea-surface temperature)	Location or area	Type of event (missing, out of range)	Observation quantity (such as bias correction)
Number of past similar events	Severity of deviation	Are similar datasets in the area, and how many are affected?	Is the Ensemble of Data Assimilations spread increased?
Severe weather events in the area	All levels/channels affected	How many areas affected?	Parameters affected?
	Data counts	Usage status (satellite data)	

Random forest classifier

False alarm (yes/no)	Slight (yes/no)	Considerable (yes/no)	Severe (yes/no)
	Cause (data/other)	Action required (yes/no)	

FIGURE 3 Features used in the machine learning classifier.

These rules (in the form of decision trees) are then applied to warnings to label them. The training dataset needs to be pre-processed for each target attribute to enable the balancing of the population of possible outcomes. The balancing simply involves the duplication of items for the less populated categories.

The labelling process is time-consuming and requires domain knowledge. In this first implementation, the labelling process was mostly done semi-automatically involving some rules gained from the experience of using the data checking system. The performance of the classification depends largely on the quality of the labelling of the training dataset and more importantly on the data features selected to characterise an event.

Figure 3 shows the important features used by the classifier to determine the cause of events. Once the system is operationally implemented, we plan to repeat the training procedure based on a more refined manual labelling and to allow ad-hoc labelling of generated warnings when relevant (in case of unusual events, for instance). Improving the labelling is very important to improve the reliability of the data checking system.

Once the classification of detected events has been achieved, a consolidation step is performed for each data type. This involves merging common events to reduce the number of warnings communicated to users; generating time series of warnings; and archiving warnings in the event database. An example of consolidated events are
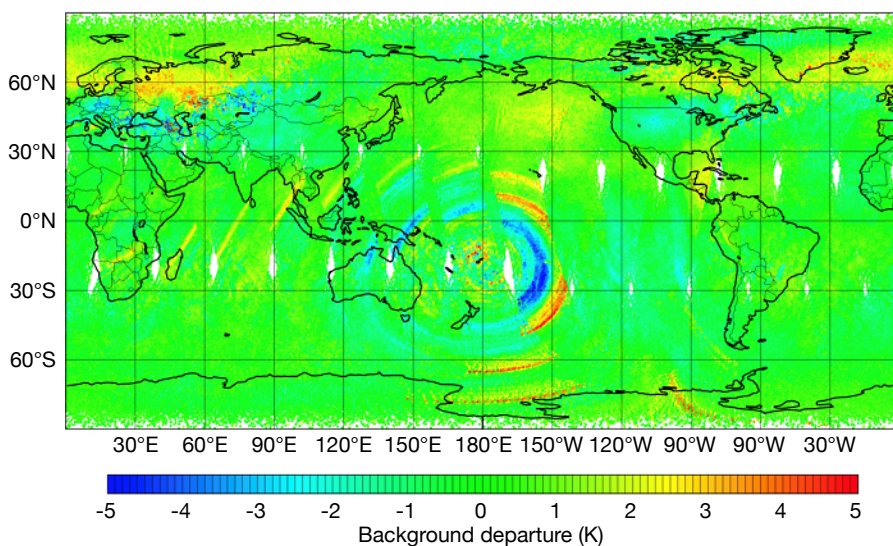

FIGURE 4 Background departures from IASI channel 92 from Metop-B and Metop-C satellites on 15 January 2022, showing the effect of the Hunga Tonga–Hunga Ha'apai eruption.
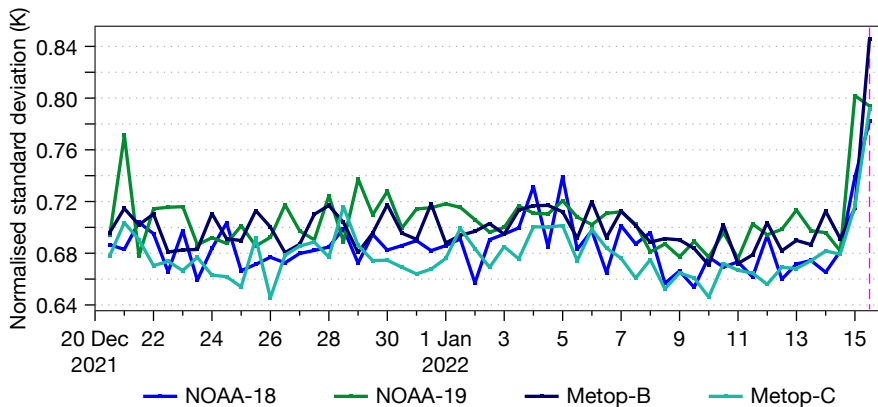
FIGURE 5 Time series of normalised standard deviation of background departures for AMSU-A channel 11 from four different satellites. The statistics are computed over the southern hemisphere extratropics. On 15 January, the standard deviations became much bigger because of the Hunga Tonga–Hunga Ha'apai eruption.

warnings triggered for many satellite data as a result of the shockwave associated with the Hunga Tonga–Hunga Ha'apai eruption in mid-January 2022 (Figures 4 and 5). The cause of the warnings was attributed to other causes (volcano eruption in this case) by the automatic data checking system. The classification was mainly driven by the number of independent data sources affected by the same event.

Evolution of the machine learning data checking system

This first implementation of the machine learning data checking system aims to incorporate novel techniques in the detection of observation anomalies. The new system tends to have fewer false alarms than the current operational framework, and it is able to detect all relevant anomalies and to assign appropriate severity levels. The random forest classifier manages to consider each warning in the context of what is happening with the other components of the observing system. This leads to a better distinction between observation anomalies and issues caused by other factors, such as data assimilation limitations and unusual atmospheric activities. However, the current classifier is mostly reproducing rules used during the semi-automatic labelling of the training dataset. Improved labelling will greatly benefit classification and severity assignment. Future upgrades will offer the possibility to continuously evolve the training dataset by enabling ad-hoc labelling of interesting events. The training of classifiers will also

benefit from simulated scenarios for hacking and data tampering that might affect parts of the observing system. This will enable the system to issue warnings if such scenarios materialise in the future.

Automatic data checking is currently performed after data assimilation takes place, which means that corrective actions are applied at a later stage. ECMWF is planning to implement automatic data checking of incoming data before the start of data assimilation. The detection results will potentially contribute to automatic data selection. A machine-learning-based system is well placed to perform pre-assimilation checks thanks to its reduced reliance on manual tuning and the possibility of improvement due to improved labelling. Parallel efforts are being pursued at ECMWF to use machine learning within the forecasting system to improve data selection and quality control, for example relating to machine-learning-based cloud detection for infrared satellite data.

Further reading

Dahoui, **M.**, **N. Bormann**, **L. Isaksen** & **T. McNally**, 2020: Recent developments in the automatic checking of Earth system observations, *ECMWF Newsletter* **No. 162**, 27–31.

Düben, **P.**, **U. Modigliani**, **A. Geer**, **S. Siemen**, **F. Pappenberger**, **P. Bauer et al.**: 2021, Machine learning at ECMWF: A roadmap for the next 10 years, *Technical Memorandum* **No. 878**.

ECMWF Council and its committees

The following provides some information about the responsibilities of the ECMWF Council and its committees. More details can be found at:
http://www.ecmwf.int/en/about/who-we-are/governance

Council

The Council adopts measures to implement the ECMWF Convention; the responsibilities include admission of new members, authorising the Director-General to negotiate and conclude co-operation agreements, and adopting the annual budget, the scale of financial contributions of the Member States, the Financial Regulations and the Staff Regulations, the long-term strategy and the programme of activities of the Centre.

President Dr Daniel Gellens *(Belgium)*

Vice President Prof. Penny Endersby *(UK)*

Policy Advisory Committee (PAC)

The PAC provides the Council with opinions and recommendations on any matters concerning ECMWF policy submitted to it by the Council, especially those arising out of the four-year programme of activities and the long-term strategy.

Chair Mr Eoin Moran *(Ireland)*

Vice Chair Ms Virginie Schwarz *(France)*

Finance Committee (FC)

The FC provides the Council with opinions and recommendations on all administrative and financial matters submitted to the Council and exercises the financial powers delegated to it by the Council.

Chair Dr Gisela Seuffert *(Germany)*

Vice Chair Mr Lukas Schumacher *(Switzerland)*

Scientific Advisory Committee (SAC)

The SAC provides the Council with opinions and recommendations on the draft programme of activities of the Centre drawn up by the Director-General and on any other matters submitted to it by the Council. The 12 members of the SAC are appointed in their personal capacity and are selected from among the scientists of the Member States.

Chair Prof. Thomas Jung *(Germany)*

Vice Chair Dr Susanna Corti *(Italy)*

Technical Advisory Committee (TAC)

The TAC provides the Council with advice on the technical and operational aspects of the Centre including the communications network, computer system, operational activities directly affecting Member States, and technical aspects of the four-year programme of activities.

Chair Dr Sarah O'Reilly *(Ireland)*

Vice Chair Ms Anne-Cecilie Riiser *(Norway)*

Advisory Committee for Data Policy (ACDP)

The ACDP provides the Council with opinions and recommendations on matters concerning ECMWF Data Policy and its implementation.

Chair Mr Paolo Capizzi *(Italy)*

Vice Chair Ms Monika Köhler *(Austria)*

Advisory Committee of Co-operating States (ACCS)

The ACCS provides the Council with opinions and recommendations on the programme of activities of the Centre, and on any matter submitted to it by the Council.

Chair Mr Nir Stav *(Israel)*

Vice Chair Dr Elena Mateescu *(Romania)*

ECMWF publications

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

ESA Contract Reports

Lean, **K.**, **N. Bormann** & **S. Healy**: WP-1000 Review of EDA approach and recommendations for small satellite configurations. *February 2022*

Lean, **K.**, **N. Bormann** & **S. Healy**: WP-2000 Calibration of EDA spread and adaptation of the observation error model. *February 2022*

Lean, **K.**, **N. Bormann** & **S. Healy**: WP-3000 Developing a flexible system to simulate and assimilate small satellite data. *February 2022*

Weston, **P.** & **P. de Rosnay**: Annual SMOS brightness temperature monitoring report 2020/21. *May 2022*

Weston, **P.** & **P. de Rosnay**: Q3 2021: Operations Service Report. *May 2022*

Weston, **P.** & **P. de Rosnay**: Q4 2021: Operations Service Report. *May 2022*

Weston, **P.** & **P. de Rosnay**: Q1 2022: Operations Service Report. *May 2022*

Lean, **K.**, **N. Bormann**, **S. Healy** & **S. English**: Final Report: Study to assess earth observation with small satellites and their prospects for future global numerical weather prediction. *November 2022*

EUMETSAT/ECMWF Fellowship Programme Research Reports

60 **Warrick**, **F.** & **N. Bormann**: Prospects for improving AMV spatial coverage between geostationary and polar AMVs: LeoGeo and Dual-Sentinel. *November 2022*

ECMWF Calendar 2023

Feb 8–9	CEMS 2nd global Flood Forecasting and Monitoring Meeting
Mar 14	Joint European Weather Cloud Advisory Group
Apr 21	Advisory Committee for Data Policy (virtual)
Apr 26	Policy Advisory Committee (virtual)
Apr 27	Finance Committee (virtual)
May 2–5	Training course: High Performance Computing – Atos
May 15–19	Training course: Data assimilation
May 22–26	Training course: EUMETSAT/ECMWF NWP-SAF satellite data assimilation
May 22–26	6th OpenIFS User Meeting, Barcelona Supercomputing Center
Jun 5–9	Using ECMWF's Forecasts
Jun 21–22	Council
Jun 27–30	Atmospheric River Reconnaissance Workshop

Sep 4–8	Annual Seminar
Sep 25–29	20th workshop on high-performance computing in meteorology
Oct 4–6	Scientific Advisory Committee
Oct 9–12	Training course: Use and interpretation of ECMWF products
Oct 19–20	Technical Advisory Committee (virtual)
Oct 24–25	Finance Committee
Oct 25	Policy Advisory Committee
Nov 13–17	Training course: A hands-on introduction to Numerical Weather Prediction Models: Understanding and Experimenting
Nov 20–24	Training course: Parametrization of subgrid physical processes
Nov 27–Dec 1	Training course: Predictability and ensemble forecast systems
Dec 5–6	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.