ECMWF

European Centre for Medium-Range Weather Forecasts

Newsletter

No. 167 | Spring 2021

Snowfall in Madrid in January

Data assimilation or machine learning?

ECMWF online

A new way to work with IFS experiments

Publication policy

The ECMWF Newsletter is published quarterly. Its purpose is to make users of ECMWF products, collaborators with ECMWF and the wider meteorological community aware of new developments at ECMWF and the use that can be made of ECMWF products. Most articles are prepared by staff at ECMWF, but articles are also welcome from people working elsewhere, especially those from Member States and Co-operating States.

The ECMWF Newsletter is not peer-reviewed.

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

Guidance about submitting an article and the option to subscribe to email alerts for new Newsletters are available at *www.ecmwf.int/en/about/media-centre/media-resources*

Beyond the medium range

ECMWF carries a commitment to 'medium-range weather forecasts' in its name. That means a broad range of forecasts across the globe up to 10 days ahead. The production of such forecasts is enshrined in the Centre's Convention. But it was found long ago that it was best for them to be complemented by research into longer-range forecasts and the release of corresponding products. The starting point was laid in 1986, when the Centre's governing Council decided that "it will be necessary to carry out extended integrations to study systematic model error and as an aid in predictability research". Council also reserved the right to make available "forecasts in the extended range". This has developed into the operational issuing of extended-range forecasts, up to 46 days ahead, and long-range forecasts, up to 13 months ahead. Together with medium-range forecasts, they are produced by the Integrated Forecasting System (IFS). Several articles in this Newsletter illustrate important work that has recently been carried out in the extended-range and long-range fields.

One of them highlights the importance of the Gulf Stream for forecasts. The results of sensitivity experiments show the potential benefits of high-resolution ocean models that can better resolve the position of the Gulf Stream. Another article investigates the possibility that ECMWF's current seasonal forecasting system, SEAS5, will predict marine heatwaves. A third contribution examines the representation of blocking in ECMWF seasonal prediction systems, and a fourth one reviews the interplay of the state of the Indian Ocean and the El Niño – Southern Oscillation in the Pacific. Another case in point that will be covered in more detail in the next Newsletter, but is already on our website, is a change in the calibration of SEAS5 forecasts of the number of tropical cyclones, to reduce their underprediction. These cases show that we have come a long way in our predictions beyond the medium range, but also that a lot remains to be done.

Some of the articles focus on medium-range predictions. They highlight an episode of snow in Spain in January and exceptionally high precipitation in Sardinia in November. The latter article also shows how our forecasts complement more short-range products. It underlines how forecasts in different ranges can complement each other.

The more in-depth articles include a first application of machine learning, in this case in the field of data assimilation. Watch out for more to follow. They also cover a new way to work with IFS experiments, and developments in ECMWF's online presence. The latter article is particularly closely linked to the move of our data centre to Bologna, Italy. The latest news is that the wireless network service is now fully operational, but the start of operations of the new supercomputer in Bologna will have to wait a little longer. We will keep you posted.

Florence Rabier
Director-General

Contents

Editorial
Beyond the medium range 1

News
Unusual snowfall in Madrid in January 2
The impact of Gulf Stream errors on ECMWF forecasts 4
Towards open science 5
Major flash floods in Sardinia in November 2020 6
New flash flood forecast products for Italy 8
Marine heatwaves in the ocean analysis and seasonal predictions 9
Atmospheric blocking representation in ECMWF seasonal prediction systems 10
Pathways of the Indian Ocean to impact ENSO forecasts 11
ECMWF meets its representatives 12
New climate projection datasets in Copernicus 14
Scalable Acquisition and Pre-Processing system used more widely 15
Obituary: Anders Persson 16

Meteorology
Data assimilation or machine learning? 17

Computing
ECMWF online – community, collaboration and empowerment 23
IFShub: a new way to work with IFS experiments 28

General
ECMWF publications 33
ECMWF Calendar 2021 33
Contact information 33

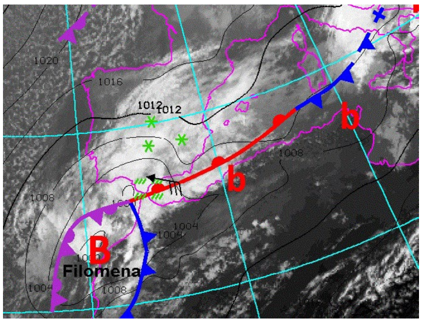
Editor Georg Lentze • **Typesetting & Graphics** Anabel Bowen • **Cover** roberto nova / iStock Editorial / Getty Images Plus

Unusual snowfall in Madrid in January

Estíbaliz Gascón, Linus Magnusson (both ECMWF), Jesús A. Barroso (AEMET, Spain)

Cyclone Filomena affected Spain from 6 January 2021 with strong winds and heavy rain in the Canary Islands and southern Andalusia, and with extreme snowfall in large areas of the interior of the Iberian Peninsula on 8–9 January. The Madrid area saw snow accumulations up to 50 cm in 30 hours, which meant this was the most intense snowstorm since 1971. After Filomena disappeared, an extreme cold wave affected the whole Peninsula. This can also be considered historic due to minimum temperature records in many areas. The snow episode forced Madrid-Barajas Airport to close and led to a cancelling of the entire rail service in the region. Schools were also closed for several days.

The Spanish Meteorological Agency (AEMET) named the storm Filomena on 5 January, when its centre was close to the Azores Islands. On 8 January, Filomena moved in a northeast direction towards the south of the Iberian Peninsula as an extratropical storm structure (cold, warm, and occluded fronts). Meanwhile, a very cold polar airmass with minimums below 0°C was established over practically the entire Iberian Peninsula.

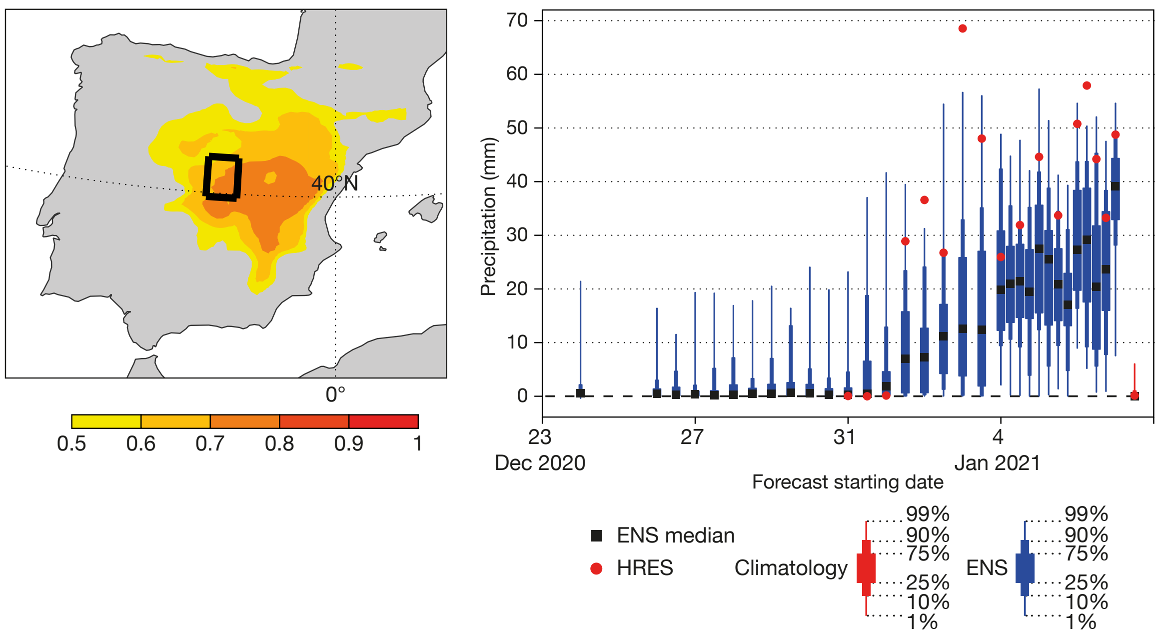


Surface analysis. Excerpt from the AEMET surface analysis for 12 UTC on 8 January 2021.

Therefore, when the storm Filomena reached the Peninsula, the warm and humid air moved over the very cold air below (see the AEMET chart). With the exception of some southern coastal areas, all precipitation occurred as snow during 8 and 9 January, resulting in a snow covering of the entire centre and northeast quadrant of the Peninsula after the storm.
Red warnings were set up in most of those areas due to the high 24-hour precipitation accumulations expected in ECMWF's ensemble forecast (ENS) for several consecutive model runs.

Snowfall accumulations

ECMWF's high-resolution forecast (HRES) and the ENS over Madrid provided some indication of the risk of snowfall from the beginning of January. From 3 January 12 UTC, the ENS median indicated snowfall totals of 20–30 mm for the period 7–9 January, compared to around 10 mm in earlier forecasts. Even in earlier forecasts, a substantial number of ENS members were indicating large amounts of snow. However, a large spread in the medium range was observed due to the uncertainty regarding the exact location of the convergence between the polar and the subtropical air masses.
The Extreme Forecast Index (EFI) was variable according to the changes in the ensemble forecast; however, a clear signal of a rare snowfall event was present six days in advance with EFI values between 0.7–0.8 in Madrid (see the EFI figure and the box-and-whiskers forecast for the area around Madrid). EFI values were close to 1 from four days in advance, with some uncertainty in the exact spatial distribution of the anomalous weather. The averaged 72-hour observed precipitation up to 10 January 00 UTC among 11 stations



Extreme Forecast Index (EFI) and Madrid region precipitation forecast. The EFI with a starting time of 00 UTC on 3 January 2021 is shown for total snowfall from 7 to 9 January 2021 (left). The box-and-whisker plot (right) shows the evolution of forecasts for precipitation in the same period in the box marked around Madrid in the left-hand panel for different starting dates.

inside a 1x1 degree box ranged from 5 to 53 mm, showing huge variability in the snowfall distribution. This compares with more than 60 mm predicted in the south of the box and between 10 and 20 mm in the north by the corresponding HRES forecast starting on 5 January 00 UTC (see the figure on observations and the HRES forecast). The same forecast on 6 January showed generally larger precipitation amounts in the southeast of Spain. However, a considerable decrease in the predicted precipitation amount was observed again at the shortest lead time on the 7 January 00 UTC run, mainly due to uncertainty in the trajectory of the cyclone between consecutive model runs. We should point out that there is also a large uncertainty in the observations, since the SYNOP stations in Spain cannot measure snowfall correctly due to the lack of heating of rain gauges. Therefore, we expect the observed precipitation to be underestimated, which makes this analysis just an estimation.

Very cold conditions after Filomena

When the North Atlantic Oscillation (NAO) enters a negative phase, it usually has a large impact on the Iberian Peninsula. In this phase, the high-level jet stream drifts to lower latitudes, resulting in above-normal rainfall and lower temperatures in southern Europe. From the end of 2020, global models began to forecast a possible negative NAO in early 2021, and this situation may have contributed to create wet and cold conditions in almost the whole Iberian Peninsula during Filomena. After Filomena, many areas in northeast Spain dropped to record low temperatures for several consecutive days. Temperatures from –20 to –26°C were recorded in some areas with anticyclonic conditions. The figure showing the temperature anomaly forecast in Europe up to two weeks ahead indicates that the low temperatures were part of a larger phenomenon. Madrid registered a minimum of –6°C on 12 January.

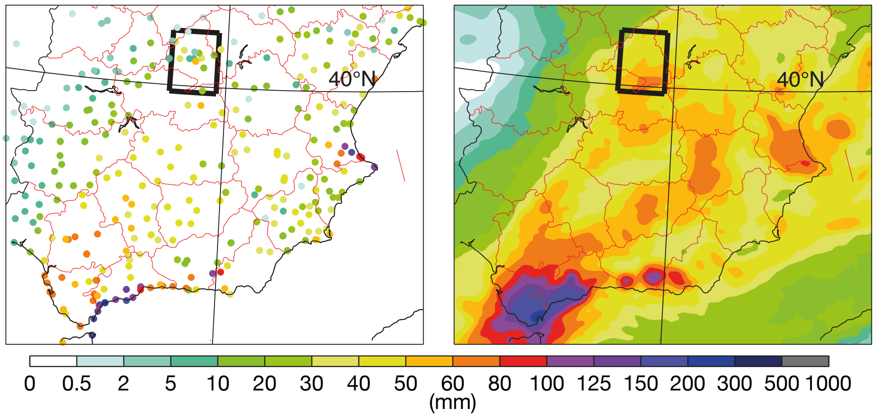
The mean 72-hour temperature forecast up to 10 January 00 UTC inside a 1x1 degree box suggests a tendency of the temperature dropping for those days from the beginning of January.

In forecasts from 9 to 10 January, the ENS median indicated mean temperature values inside that box near –5°C. However, the forecasts issued on 10 January for the same day increased those values to around –2°C and reduced the forecast's uncertainty.

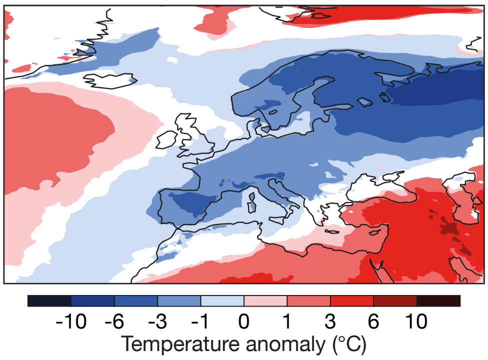
We can also systematically observe higher temperatures in the HRES than in the ENS Control for different lead times of the event (see the temperature forecast for the Madrid region).

An evaluation of the model performance shows that temperatures were generally overestimated in many areas of the interior of the Peninsula. This may have been caused by some disagreements between snow coverage in the forecast and according to observations.

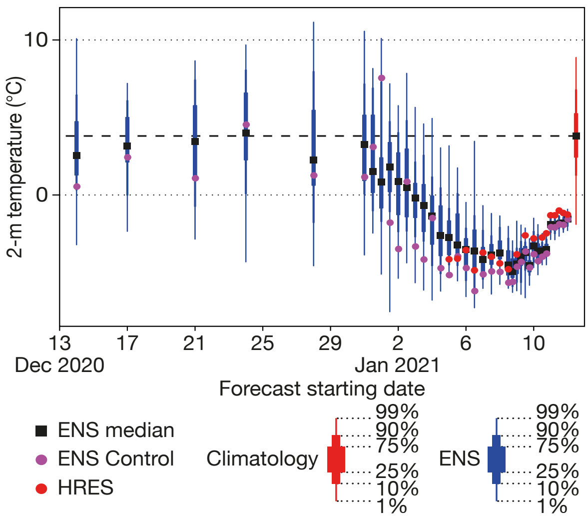
Overall, the IFS successfully predicted this historic snowfall event well in advance. The forecasts included the normal uncertainties associated with the exact location of the polar and subtropical air mass convergence and the total amount and extension of the most affected areas, but they contained enough information to provide the corresponding weather warnings long in advance.



Observed and predicted accumulated precipitation. The charts show accumulated precipitation from 7 January 00 UTC to 10 January 00 UTC in observations (left) and in the HRES forecast from 5 January 00 UTC (right).



Temperature anomaly forecast up to two weeks ahead. The plot shows the 2 m temperature anomaly forecast in Europe initialised at 00 UTC on 4 January 2021 and valid from 11 to 17 January 2021.



Temperature forecast for Madrid region. The box-and-whisker plot shows the evolution of forecasts for 2 m temperature for the 72-hour period from 11 to 13 January 2021 in the box marked in the figures above for different starting dates.

The impact of Gulf Stream errors on ECMWF forecasts

Christopher D. Roberts, Frédéric Vitart, Magdalena Balmaseda

The ECMWF Integrated Forecasting System (IFS) includes a time-evolving representation of the 3D ocean state. This coupling to a dynamic ocean model improves the representation of air–sea interactions and has been shown to improve forecasts at the sub-seasonal to seasonal (S2S) time range (from about 2 weeks to about 2 months). However, the inclusion of additional ocean processes also introduces the potential for systematic errors in sea-surface temperature (SST), which can have a negative impact on forecast quality. In particular, ocean models with a grid spacing of about 25 km, as used in the IFS, struggle to accurately simulate the location and structure of the Gulf Stream and its associated sharp gradients in SST.

Here we describe recent work at ECMWF to evaluate the impact of such SST errors on S2S forecasts using an online SST-bias correction methodology. The reference S2S experiment (CTRL) is a 15-member ensemble initialised on the 1st and 15th of each month of an extended winter period (November–March) from 1989 to 2015 (i.e. 270 start dates). The bias-corrected experiment (BCFC) is the same as CTRL, but the SSTs seen by the atmosphere in the North Atlantic region are adjusted by the model SST bias derived from CTRL, which varies as a function of location, calendar start date, and forecast lead time.

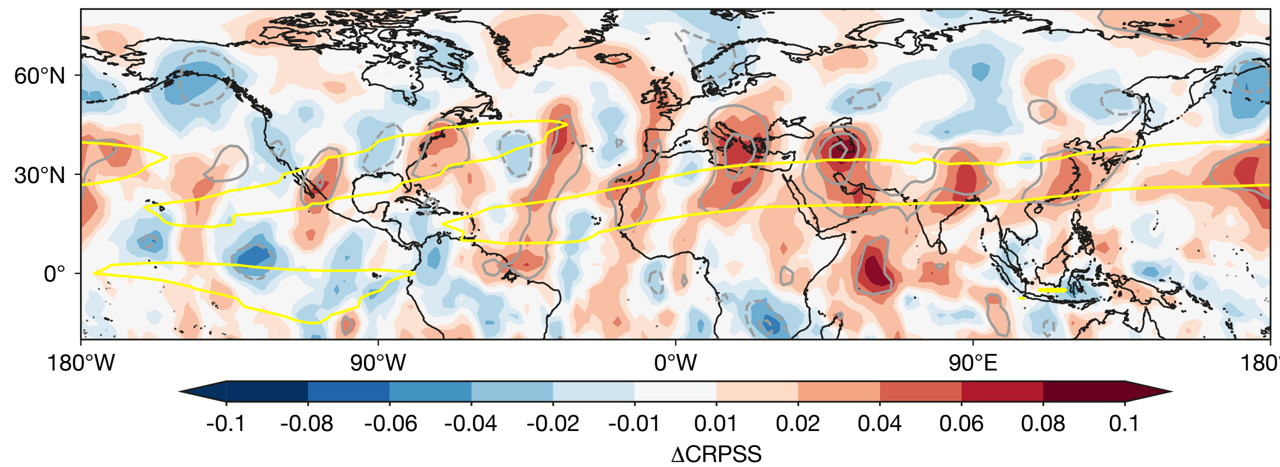
Experiment outcomes

The applied bias correction effectively reduces SST biases in the North Atlantic region. The resulting southward shift of the Gulf Stream drives changes in convective precipitation and vertical motion (not shown), which has consequences for atmospheric predictability that extend beyond the North Atlantic. In particular, we find that reducing North Atlantic SST biases leads to significantly improved S2S forecasts of atmospheric circulation anomalies over Europe. Furthermore, the impacts extend beyond the North Atlantic and Europe and circumnavigate the globe along the northern hemisphere subtropical waveguide (the figure shows an example). This response is typical of the propagation of stationary Rossby wave activity along the jet stream.

Interestingly, this impact of SST biases on forecast skill is modulated by the Madden–Julian Oscillation (MJO). Forecasts with an active MJO in the initial conditions exhibit a stronger impact over Europe and along the northern hemisphere waveguide, whereas forecasts without an active MJO have a stronger response over the Gulf Stream and North Atlantic. This sensitivity is unrelated to changes in MJO forecast skill, which is not sensitive to North Atlantic SST biases. Instead, we speculate that this effect is a consequence of the impact of the MJO on the background state and its associated teleconnections that steer or obstruct planetary wave activity that is initiated in the North Atlantic region.

In conclusion, the results from these sensitivity experiments provide important evidence for the potential benefits to ECMWF forecasts of higher-resolution ocean models (i.e. with a grid spacing of less than 10 km) that can better resolve the position of the Gulf Stream. Such models will be investigated at ECMWF during the next few years as part of the recently funded Horizon 2020 NextGEMS project.

Further information on the impact of SST biases on ECMWF forecasts can be found in an article by the authors entitled 'Hemispheric impact of North Atlantic SSTs in sub-seasonal forecasts' in *Geophysical Research Letters*, https://doi.org/10.1029/2020GL091446.



Impact of SST bias correction on forecast skill of meridional wind at 200 hPa. Changes in the continuous ranked probability skill score (CRPSS; shading) and anomaly correlation (grey contour spacing of 0.1) shown for weekly mean anomalies derived from forecast days 26 to 32. The yellow contours highlight the position of the northern hemisphere waveguide diagnosed from the meridional gradient of absolute vorticity. (See Figure 3 of the article in *Geophysical Research Letters*, https://doi.org/10.1029/2020GL091446.)

Towards open science

Nikolaos Mastrantonas, Milana Vučković

Following a recent publication, this article presents some complementary actions that we took to make the work more accessible to interested users and researchers in the interest of 'open science'.

Sharing science

Scientific publications are crucial for expanding our knowledge. A stand-alone publication nevertheless withholds many aspects of the work from its readers. This is why many researchers and institutes shift towards open science. It means that they make publicly available additional elements that led to the publications, such as the data and the tools developed.

A recent publication about extreme precipitation in the Mediterranean (https://doi.org/10.1002/joc.6985) follows a similar path. The work analysed the spatiotemporal characteristics of extreme precipitation events for regions across the Mediterranean and quantified their connections to large-scale atmospheric flow patterns over the entire domain. Bearing in mind that it could be of interest to many researchers as well as users from our Member and Co-operating States, the idea was to move a step further from the (open-access) peer-reviewed publication. In a move towards open science, this work uses freely accessible datasets and software and provides online all the post-processed data and the scripts developed for the analysis.

Motivation

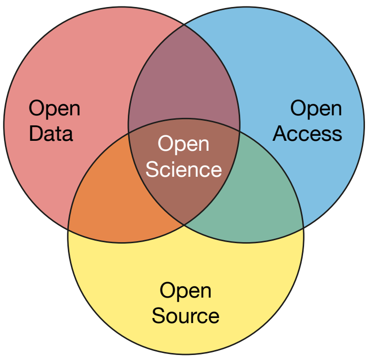
The motivation behind these additional initiatives is threefold. First, sharing tools and data can support research tackling similar topics. We also want to encourage more scientists to implement similar methods and strengthen the already remarkable efforts within the Earth science community for open and accessible science. Finally, by making the work publicly available we are hoping to get valuable feedback from interested users. This feedback can help to further improve the practices used and optimise the analysis. In a nutshell, every contribution towards open science supports the collective efforts of the scientific community and gives the opportunity to improve our individual skills and competences.

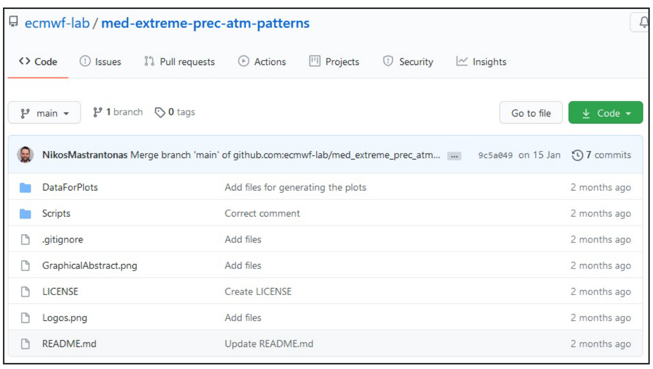
Data & tools used

As part of the Copernicus programme, ECMWF provides ERA5, the latest state-of-the-art reanalysis dataset of very high spatiotemporal resolution. The analysis conducted as part of the research on extreme precipitation in the Mediterranean made use of this dataset, which is freely accessible within the Copernicus Climate Data Store. We also used the Python software, a free and open-source programming language with an extensive community of developers. All the scripts for processing the data and creating the figures were written with Python in the Jupyter notebooks environment, which is easy to use by others and provides an opportunity to accompany the code with figures and explanatory text. All notebooks and post-processed data are available via a repository within the ECMWF account in GitHub (https://github.com/ecmwf-lab/med_extreme_prec_atm_patterns). Anyone can freely access this repository, download its content and reproduce the results locally. More importantly, any users can modify subsections of the workflow and adjust it to their needs and interests.

Open science at ECMWF

This work is only one of many contributions to open science within ECMWF. There are many efforts to make our science more accessible, share workflows and data with the scientific community, and provide tools to the end users. However, just publishing open articles and opening up the data is not enough if only scientists from our domain can use it. The development of open-source software packages that facilitate the use of weather and climate data (such as Metview and ecCodes) and computational routines used on ECMWF's operational products (such as the Integrated Forecasting System) are steps in the right direction. Making the data accessible in a wider open-source software ecosystem by developing the Python interface in Metview, and the cfgrib library for the Python and Julia languages, makes it possible for our data to be used by a much wider audience than meteorological and climatological scientists. All the above are freely accessible by all interested users. The workshop 'Building reproducible workflows for earth sciences', organised by ECMWF in 2019 (https://events.ecmwf.int/event/116/), provides a small summary of many of these initiatives towards open science.





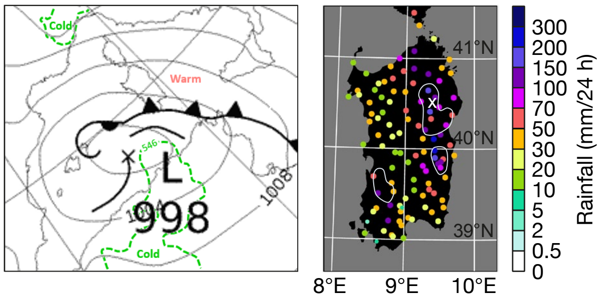
Open science example. On the GitHub page shown on the right, the data and scripts for the research on extreme precipitation in the Mediterranean have been made publicly available.

Major flash floods in Sardinia in November 2020

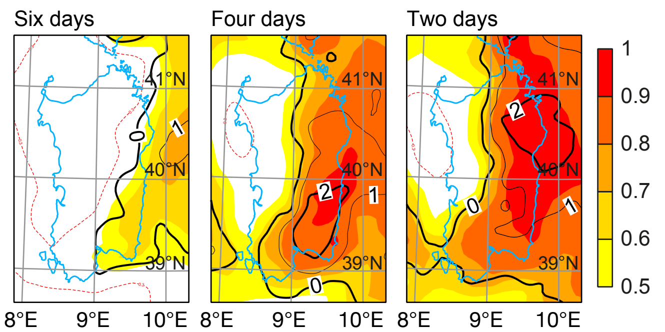
Tim Hewson (ECMWF), Samuele Salis (ARPAS), Enrico Giuseppe Cadau, Dario Secci (both Sardegna Clima Onlus), Fatima Pillosu (ECMWF & University of Reading)

Timely warnings, backed up by tenacious policing of the public's response, helped avert a major humanitarian disaster in Sardinia when devastating flash floods hit on 28 November 2020. This was probably the worst rainfall event in Sardinia since the deadly floods of November 2013, which had led to major policy changes.

Synoptically, the 2013 and 2020 events were strikingly similar. Both times, in the preceding days, a pool of cold air transferred eastwards across northern Africa, with a surface cyclone developing on the northern flank, along with an attendant inverted cold front. The primary rainfall source was then the warm, moist, unstable south-easterly flow ahead of that front, which impinged on Sardinia's complex and often steep topography (see synoptic setting and observation charts).



Synoptic setting and observations. The Met Office surface analysis for 28 November 2020 at 12 UTC, with the added green lines denoting 546 decametres for the 1000–500 hPa thickness (left), and rainfall observations for 24 hours ending at 00 UTC on 29 November 2020 (right). The curved lines in the observations image enclose some regions affected by flash floods according to media reports, and 'X' is Bitti, which suffered major impacts.



Extreme Forecast Index (EFI) and Shift of Tails (SOT) for Sardinia. Six-day, four-day and two-day forecasts of the EFI (shading and red contours) and SOT (black contours) for the 24 h rainfall period ending at 00 UTC 29 November 2020, starting from 00 UTC on 23, 25 and 27 November 2020.

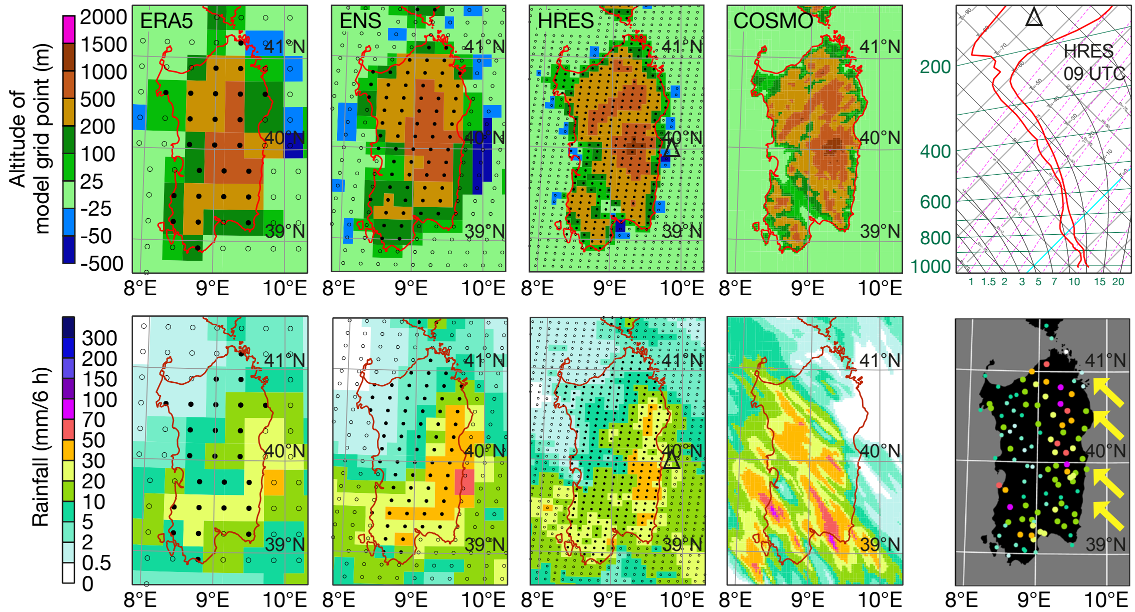
Medium-range forecast guidance

ECMWF's primary warning tools for such events are the EFI (Extreme Forecast Index) and SOT (Shift of Tails) for 24-hour rainfall. These aim to give advance warning of events that may be climatologically extreme at the model grid scale. In this case there were very useful signs of possible extremes as much as six days in advance, with more clarity developing as the event approached. Eastern Sardinia was signified to be at greatest risk (see EFI forecasts).

Whilst broadscale predictions and non-dimensional EFI-type output continue to be fundamental products for ECMWF's user base, nowadays users also expect accurate forecast values. However, models with parametrized convection, such as ECMWF's high-resolution forecast (HRES) and ensemble forecast (ENS), and indeed the ERA5 reanalysis, cannot and should not be expected to represent the very localised and flash flood-related high rainfall totals that can occur during convective outbreaks.

Model output issues

To gauge potential shortfalls, and also investigate and compare systematic resolution-dependent biases, one can assume that, at very short ranges, forecasts from particular systems will have migrated asymptotically towards the predictive skill ceilings that are innate to those systems. Unsurprisingly, the maximum rainfall forecast varies inversely to resolution, with ERA5, ENS Control and HRES here 'underpredicting' observed peaks, at land grid points, by about 75, 65 and 50% respectively (allowing for some location errors – see the forecast and observation figures). Also apparent is an increasing tendency, with larger grid-lengths, for the largest totals to shift eastwards, away from land to sea-based grid points. This might seem surprising given that topographically forced ascent is a strong convective trigger in ECMWF models. However, because topography must be spectrally fitted, near-coast sea grid points can be well above or well below sea level, creating artificial 'marine hills' that can apparently trigger convection. This is most apparent here in the Control run, where peak totals are near the real east coast where the model sea 'rises up' quite abruptly. This is also indicative of another factor contaminating IFS forecasts, namely no convective life cycle: where convection is triggered, rain immediately falls, leading to systematic biases on the grid scale. These inferences are strongly supported by analysis of ensemble mean fields (see: https://confluence.ecmwf.int/display/FCST/202011+-+Rainfall+-+Sardinia). The keen



Topography and rainfall forecasts from different systems, a representative sounding, and observations. The four top charts show the grid-point altitude in the ERA5, ENS, HRES, and COSMO models (respectively 31, 18, 9, 2.2 km resolutions). The open/filled circles are sea/land points (not shown for COSMO). The four bottom charts show 6 h rainfall for 06–12 UTC on 28 November 2020 from the same models (the Control run for ENS, member 17 for COSMO), from 00 UTC on 28 November. The triangles in the HRES images indicate the location of a 9-hour HRES forecast sounding for 09 UTC on the same day, shown in the top right panel. The chart indicates high moisture content and marked potential instability. The bottom right panel shows observations of rainfall corresponding to the forecast times. The arrows show dominant tropospheric airflow direction. Speeds were about 35 kts with minimal vertical shear.

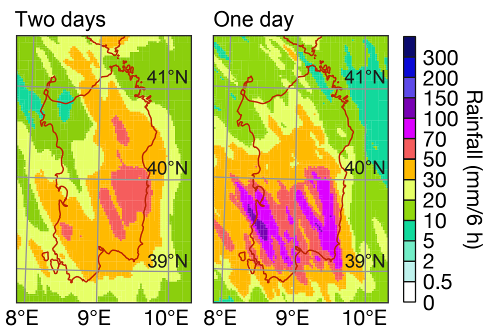
forecaster referencing grid point values (e.g. on meteograms) will benefit greatly from understanding these various modelling issues.

Convection-resolving model and datasets

In reality, the largest 6-hour totals were inland, mostly in elevated regions, whilst eastern near-coast sites saw more modest amounts. The convection-resolving COSMO (Consortium for Small-scale Modelling) model captured this distinction much better, and extremes were of the right order. However, in any one member peaks were not all correctly located, and near-coast rainfall was insufficient, perhaps because of other model issues (see the COSMO panels in the rainfall forecast figure). To reach these instructive conclusions, and indeed for real-time monitoring, high density observations are essential. Whilst the Agenzia Regionale per la Protezione dell'Ambiente della Sardegna (ARPAS) provides one such set, separate initiatives (such as Sardegna Clima: https://www.sardegna-clima.it/) are increasingly providing additional real-time datasets to exploit around the world.

EFI issues

The absence of coastal maxima, which were consistently predicted in ENS members at shorter ranges, suggests that the EFI and SOT fields discussed above may have been contaminated by systematic errors. The explanation would be mixing up of differing weather-type-dependent biases, which is a minor weakness of the EFI approach.



MISTRAL rainfall forecasts. The 98th percentile from MISTRAL probabilistic rainfall forecasts (post-processed and blended COSMO and ENS), for respectively about 36 and 12 hours in advance, valid for 06–12 UTC on 28 November 2020.

Post-processed products

To try to address these various resolution and formulation-related issues, and thus to deliver more accurate forecasts of values for users, one can post-process and blend model output in different ways. This is done in the MISTRAL project (Meteo Italian Supercomputing Portal), which uses ECMWF and COSMO ensembles (also in this Newsletter). Probabilistic output thus derived can provide for forecasters a 'reasonable worst-case scenario' at a user-selected point probability level (see MISTRAL charts). Whilst raw model outputs may not always reach the basic performance level needed for post-processing to add value (as was the case for northeast Sardinia here), overall MISTRAL products did provide reasonable guidance regarding maximum rainfall values for this case.

Outlook

ECMWF's Strategy to 2030 includes the goal of delivering convection-resolving ensemble output at global scales. This article demonstrates some of the benefits and challenges that can lie ahead.

New flash flood forecast products for Italy

Estíbaliz Gascón, Andrea Montani, Tim Hewson (all ECMWF), Gabriella Scipione (CINECA)

The MISTRAL project (Meteo Italian Supercomputing Portal), funded under the EU's Connecting Europe Facility (CEF), has come to an end after 30 months. The project's main goals were to facilitate and foster the re-use of meteorological datasets by various communities, to provide added-value services using high-performance computing (HPC) resources.
The project lead and provider of HPC facilities was CINECA, based in Bologna. Along with ECMWF, the other collaborating partners were Arpae Emilia-Romagna, Arpa Piemonte, Protezione Civile and Dedagroup (all in Italy).

ECMWF participated in MISTRAL to improve probabilistic 6-hour rainfall forecasts to help anticipate flash floods in Italy: the so-called 'flash flood use case'. Localised very heavy rainfall is commonly associated with flash floods, but it is difficult to forecast it accurately in both magnitude and location. Ideally weather forecasts should be provided for points, and not for the large regions represented by the grid boxes of global models, such as those of ECMWF. The mismatch can be addressed by post-processing model output or by using convection-resolving limited-area models (LAMs). In MISTRAL, ECMWF used both approaches in ensemble mode, with blending used to create the final products. We also applied post-processing to the LAM output before blending, to address the limitations of finite ensemble size that play a central role when convection is resolved.

Blending global and LAM output

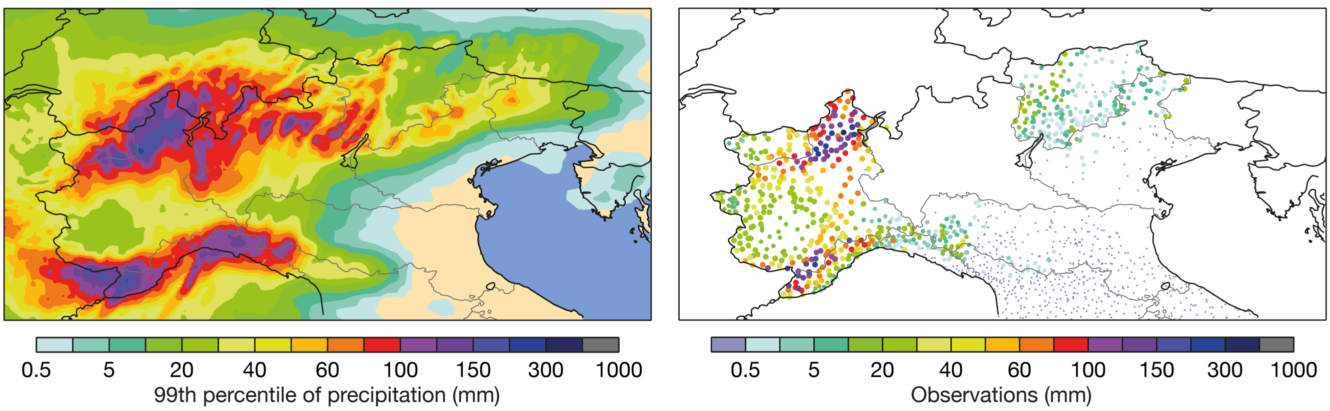
For the global model post-processing, we employed an ECMWF approach based on decision trees (ecPoint-Rainfall), which relies on calibration against rain gauges. Two underlying premises of ecPoint-Rainfall are that the relationship of forecasts versus point-observations depends not on location but on weather types, and that each such weather type has a different sub-grid point rainfall distribution associated. ECMWF already runs ecPoint-Rainfall for 12-hour accumulation periods at ECMWF. In MISTRAL we redesigned the weather types for use with 6-hour totals. Use of 6-hour rainfall is more appropriate here, given the propensity for flash floods to be driven by extreme rainfall over short periods.

For the LAM ensemble, ECMWF used COSMO-2I-EPS, created by the COSMO-LAMI consortium (Consortium for Small-scale Modelling-Limited Area Model Italia). This is a 2.2 km LAM ensemble prediction system run for Italy and surrounding regions. The post-processing applied here was a new state-of-the-art scale-selective neighbourhood technique. This technique aims to selectively and dynamically preserve, at fine scales, the (heavy) rainfall signals that are reliable – assumed to be where ensemble member grid box totals agree – and to smooth out probabilities elsewhere, across neighbourhoods proportionate in size to the level of disagreement.

By blending, we aim to combine the two systems' most skilful aspects. Blended output is available up to 48 hours (when the LAM runs end), with the ecPoint relative weight rising from zero at T+0 to one by T+48. From T+48 to T+240, only ecPoint is used. This strategy minimises the impact of fast error growth with lead time in the LAM and provides product continuity. ecPoint-Rainfall is specifically designed to improve the reliability and discrimination of the forecast.
In verification, it shows particularly good results for large totals.
By combining it with LAM-based output, we can be more specific at short lead times about where extreme totals are most likely. The final products are delivered on the COSMO grid, as percentiles and probabilities of exceeding specific thresholds for 6-hour rainfall. This serves Italian forecasting and feeds into a gateway for Europe. The use of CINECA HPC resources has been crucial to produce the real-time products operationally since the end of 2019.

Open data portal

The new products can be found in the



Italy Flash Flood forecast product. The left-hand plot shows the 99th percentile of precipitation over northern Italy, with verifying observations in the right-hand panel as available from the MISTRAL portal, all for 18–24 UTC on 2 October 2020. The forecast start time is 00 UTC on the same day. Devastating floods were reported in northwest Italy that day.

new National Meteorological Open Data Portal (https://meteohub.mistralportal.it/app/maps/flashflood/) created in the framework of the project. In the portal you can also find other forecasts, satellite data, radar and SYNOP observations. With this project, ECMWF benefited greatly from further development of ecPoint and from a close look at convection-permitting ensemble output and the challenges of post-processing it. Gaining timely access to high-density observations for Italy, for verification and research, was another plus.

Marine heatwaves in the ocean analysis and seasonal predictions

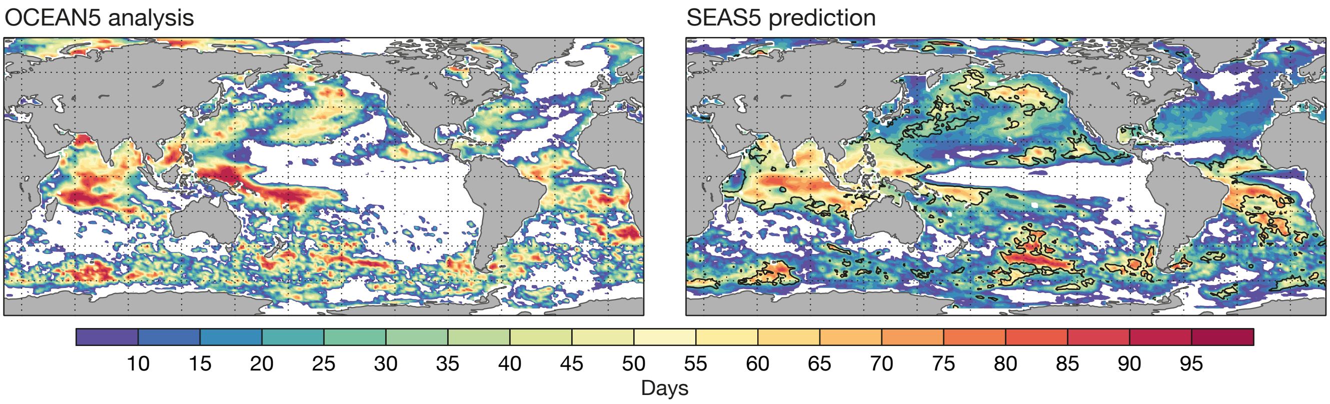
Eric de Boisséson, Magdalena Balmaseda

Marine heatwaves (MHW) are defined as periods of usually five days or more of warmer sea-surface temperatures (SST) than the 90th percentile.
The importance of marine heatwaves was highlighted by 'the Blob', a long-lasting event off the west coast of North America and Alaska over the 2014–16 period that decimated populations of Pacific cod, seabirds, salmon and other species while toxic algae prospered. Another major marine heatwave hit the northeast Pacific in the summer and autumn of 2019 (see ECMWF Newsletter 162). This strong event led the US federal cod fishery in the Gulf of Alaska to close for the 2020 season as a precautionary measure as the number of cod in the area was deemed too low. Advance planning of the fishing season is an example of the value that forecasting marine heatwaves could have for a better management of fish stocks and ecosystems in general. A better understanding of marine heatwaves (origin, characteristics, etc.) and their predictability is the objective of the marine heatwave task force of the Horizon 2020 EuroSea project. Within the project, ECMWF's role is to investigate the representation of such events in ocean real-time monitoring products such as OCEAN5, and to investigate their predictability using the ECMWF seasonal forecasting system 5 (SEAS5).

Analyses and predictions

OCEAN5 is ECMWF's current ocean and sea-ice analysis system.
The OCEAN5 monitoring page has been online since 2019. The high-frequency real-time ocean monitoring capabilities provide an easy way to detect at a glance if any major anomaly is going on in the global ocean with respect to the long term ocean climatology (https://www.ecmwf.int/en/forecasts/charts/oras5_nrt/). Meanwhile, the behind real-time OCEAN5 monitoring offers a record of anomalies on ocean variables from 1975 to the present, with a focus on low frequency variability (https://www.ecmwf.int/en/forecasts/charts/oras5/). Large values of SST anomalies provide only a qualitative indication of heatwaves. Work is ongoing to enhance these monitoring tools for a more quantitative detection of marine heatwaves, e.g. events exceeding the 90th or 95th percentile of the climatological values. In 2020, strong SST anomalies were detected in OCEAN5, and values exceeding the 90% threshold were tagged as marine heatwaves, as shown in the figure.

SEAS5 provides a 50-member ensemble of 7-month-long forecasts every month. They are based on the ECMWF Earth system model that couples the atmosphere, land, waves and the ocean. SEAS5 cannot predict the exact observed weather features but can provide indications on weather statistics on a monthly to seasonal basis. Here we explore if SEAS5 can predict the occurrence of marine heatwaves over the next few months. The forecast starting on 1 May 2020 was chosen to assess the ability of SEAS5 to predict marine heatwave events in the second half of 2020, like those captured in the OCEAN5 analysis. For each member of the forecast, we



OCEAN5 and SEAS5 marine heatwave days in the summer of 2020. The charts show the number of days in June–July–August 2020 when SST anomalies exceeded the 90% threshold of the climatological value in OCEAN5 (left) and on average in the SEAS5 ensemble prediction system (right). Areas where such anomalies persisted for less than five consecutive days have been left blank. The black contour in the right-hand panel indicates areas where the probability of SEAS5 predicting a MHW is over 90%.

compute the number of days per month and season in which the SST anomalies are above the 90th percentile, and we compare the result to OCEAN5 monitoring (see the figure).

The outcome is that, on average, SEAS5 can predict marine heatwaves in the right areas. The average number of extreme days is lower than observed, but the range of the ensemble encompasses the observations. These extreme values are predicted with more than 90% probability over many of the areas of occurrence.

Next step

These first diagnostics show there is potential for ECMWF SEAS5 to predict marine heatwave events in the same areas as seen in OCEAN5 in the summer 2020. The next step is to diagnose several years of seasonal forecasts to get more robust statistics on marine heatwave predictions that would provide indications on the possibility to make these predictions part of the ECMWF catalogue.

This work is supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No 862626, project EuroSea (Improving and Integrating European Ocean Observing and Forecasting Systems for Sustainable use of the Oceans).

Atmospheric blocking representation in ECMWF seasonal prediction systems

Paolo Davini (ISAC-CNR, Italy), Antje Weisheimer, Magdalena Balmaseda

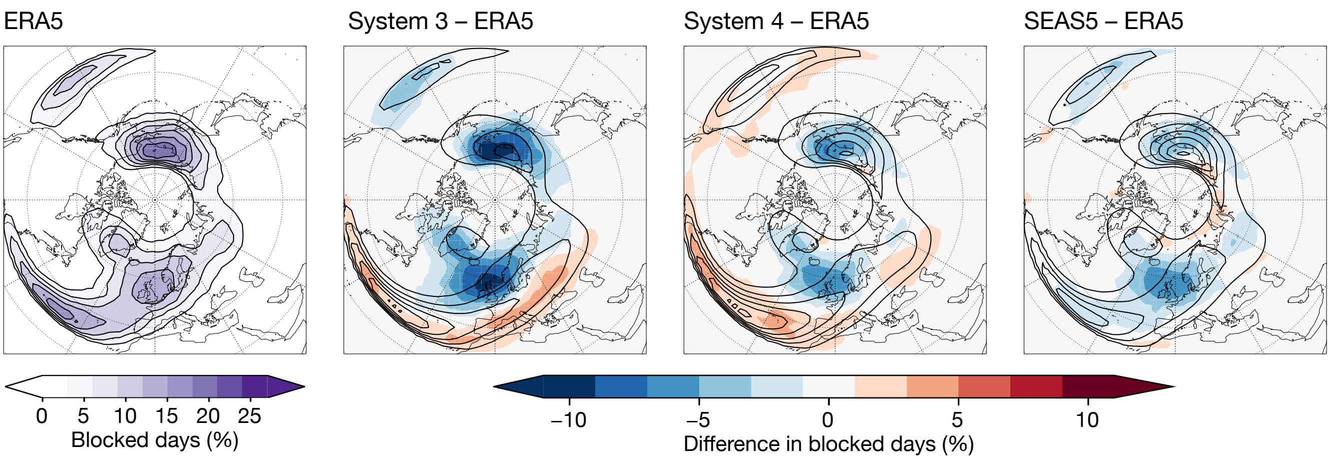
Atmospheric blocking is a recurrent weather pattern characterized by a quasi-stationary persistent large-scale high-pressure system which 'blocks' the westerly flow, typically at the exit of the storm tracks. Blocking has elicited the attention of scientists since the early 1950s, due to its inherent dynamical complexity and its impacts on mid-latitude weather, which may lead to cold spells in winter and to heatwaves in summer.

In addition to being a weather phenomenon only partially understood by the scientific community, blocking has always been a difficult problem in numerical simulations, from both the weather and the climate point of view. However, blocking has been rarely investigated at seasonal timescale. A new study has reviewed the representation of blocking in ECMWF seasonal prediction systems, tested its sensitivity to key model characteristics, explored predictive skill and compared it with climate simulations of the ECMWF model.

Climatology and sensitivities

Blocking representation in ECMWF seasonal prediction systems shows considerable improvements from System 3, to System 4, and to SEAS5, as shown by the blocking frequency climatology in the figure. Reduced biases are observed over the North Pacific and Greenland, even though over Europe – a region where numerical models often struggle the most – a non-negligible underestimation of blocking frequency is still present in SEAS5.

Making use of a series of complementary hindcast experiments, it has been possible to assess that increasing atmospheric and oceanic resolution improves considerably the statistics of blocking. On the other hand, the simulated blocking frequency remains largely insensitive



Climatological frequencies of blocking events. The frequencies of blocking events in December–January–February in the period 1981–2011 are shown here for the ERA5 reanalysis (shading and contours) and seasonal predictions from System 3, System 4 and SEAS5 (contours). The three right-hand charts also highlight the System 3, System 4 and SEAS5 differences from ERA5 (shading). The charts show general improvements from System 3 to SEAS5 compared to ERA5 but also a persistent underestimation of blocking events in western Europe. The figure is adapted from the first figure in Paolo Davini et al., *Quarterly Journal of Meteorological Society*, https://doi.org/10.1002/qj.3974, (CC BY 4.0).

to coupled model sea-surface temperature (SST) errors, suggesting that atmosphere–ocean coupling might not be the key to correctly simulating blocking as long as the mean oceanic state does not show large biases. The implementation of stochastic parametrizations, which are operationally included in SEAS5, tends to slightly displace blocking activity equatorward, possibly pointing to a change in the mean climate.

Variability and prediction

SEAS5 predictive skill of blocking frequency and corresponding signal-to-noise ratios remain globally low, but interesting positive results are found at low latitudes, especially over the Pacific sector. Moreover, positive skill is also obtained over parts of western and central Europe, achieving moderate but significant values (ensemble mean correlation skill of about 0.3) over an extended region. Most importantly, the areas of significant skill are larger and more robust than in previous seasonal prediction systems, pointing not only to an improvement of the mean climate but also of the prediction capabilities.

Interestingly, SEAS5 blocking interannual variability is underestimated too, and its bias is proportional to the climatological frequency. This shows that a negative bias in the blocking frequency implies a negative bias in the interannual variance. As a result, it seems plausible to expect that eventual improvement in the blocking mean state would be reflected by improvement in its interannual variability. It might also influence positively the overall skill of mid-latitude geopotential height which, for SEAS5 over Europe, remains lower than for blocking.

Comparison with climate runs

Seasonal hindcasts have been compared against a set of climate runs of the same model configuration. It has been shown that SST errors have a larger impact on blocking bias in climate runs than in seasonal runs, and that increased ocean model resolution contributes to improved blocking more effectively in climate runs than in seasonal forecasts. Making use of the same methodology, it has been found that the largest contributors to the long-standing underestimation of blocking are persistent errors in the atmospheric model. Seasonal forecasts can thus be considered a suitable seamless test-bed for model development targeting blocking improvement in climate models.

Further information can be found in an article by Paolo Davini and co-authors, *Quarterly Journal of Meteorological Society*, https://doi.org/10.1002/qj.3974.

Pathways of the Indian Ocean to impact ENSO forecasts

Michael Mayer (ECMWF, University of Vienna), Magdalena Balmaseda

After a false alert in the northern winter 2014/15, in 2015/16 one of the strongest El Niño events on record occurred in terms of the standard Niño 3.4 (N3.4) sea surface temperature anomaly index. A study based on reanalyses revealed that the 2015/16 El Niño was also associated with unusual energy transfers (see ECMWF Newsletter No. 155). Most notably, the Indonesian Throughflow (ITF) was exceptionally weak, which led to the retention of warm waters in the Pacific Ocean. The weak ITF appeared related to the relatively high sea level in the Indian Ocean with respect to the Western Pacific.

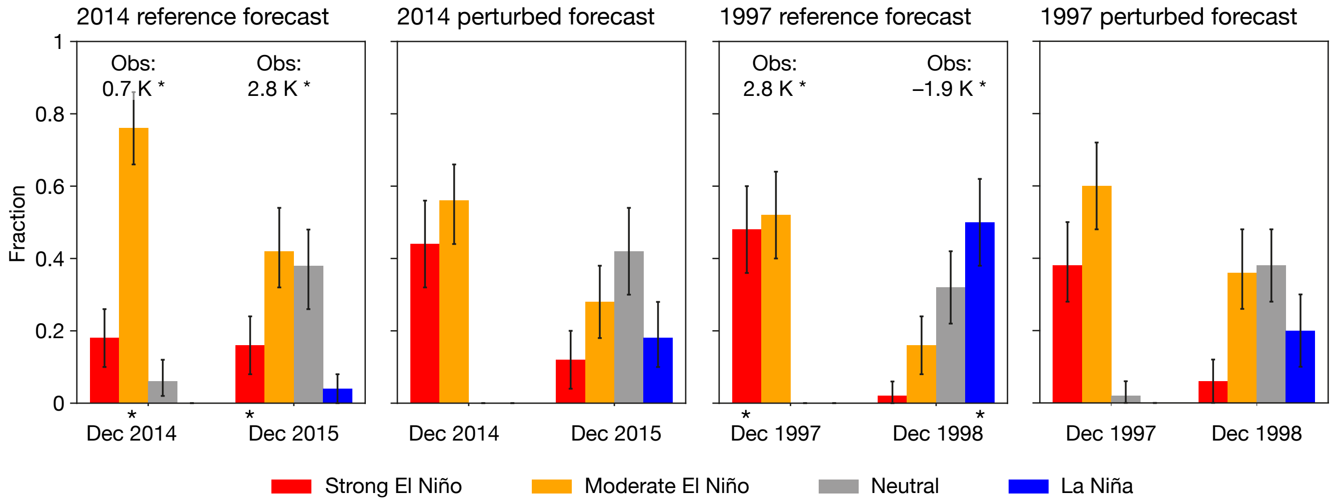
Recent results confirm the earlier hypothesis that the exceptionally warm Indian Ocean state in 2014 was responsible for the unprecedented reduction of observed ITF transports and retention of heat in the Pacific. The Indian Ocean state also decreased the probability of an extreme warm event in the Pacific in 2014/15 and favoured the probability of warm conditions in 2015/16, as later materialised. The results demonstrate the importance of the Indian Ocean low-frequency variability and trends for seasonal forecasts. They also highlight the potential merit of two-year-long ENSO (El Niño – Southern Oscillation) predictions, as forecasters may have interpreted the predictions in 2014 differently with information of lead times beyond year one.

Impact of Indian Ocean on 2-year ENSO predictions

We investigated the impact of the anomalously warm Indian Ocean state in 2014 on the subsequent evolution of ENSO and its energetics with two sets of two-year long twin seasonal forecast experiments. The reference experiments were initialised in February 2014 and 1997, using the SEAS5 forecasting system. These two initial dates were chosen because the state in the Pacific was similar, but the Indian Ocean was much cooler in 1997. In the perturbed experiments, we swapped the Indian Ocean initial conditions between these two years.

Reference forecasts

Consideration of predicted ENSO probabilities shows that the two reference experiments are able to discern the contrasting ENSO evolution during 1997–99 and 2014–16 (see the figure). The forecasts starting in Feb 1997 predict a 100% chance of an El Niño event in year 1 (Dec 1997) with a 48% chance of that event being extreme. In year 2 (Dec 1998), cool La Niña conditions are predicted with a 50% probability. The observed outcome was indeed an extreme warming in 1997/98 followed by strong cold conditions in 1998/99. The forecasts starting in Feb 2014 predict a 98%



Probability of occurrence of ENSO events. The first two charts show the probability of ENSO events in December 2014 and December 2015 according to a reference forecast starting in February 2014 and a perturbed forecast where Indian Ocean initial conditions were swapped with those from February 1997. The third chart shows a reference forecast starting in February 1997, and the fourth one shows a perturbed forecast starting in February 1997 where Indian Ocean initial conditions were swapped with those from February 2014. The probabilities were derived from counting ensemble members. Bars represent probabilities of La Niña (ENSO index N3.4 ≤ –0.5 K), neutral (–0.5 K < N3.4 ≤ 0.5 K), moderate El Niño (0.5 K < N3.4 ≤ 1.8 K), and strong El Niño (N3.4 > 1.8 K) events. Observed values for N3.4 are shown as well. The bars are medians and the whiskers 5% and 95% quantiles of the probabilities. The figure has been adapted from Figure 5 in Michael Mayer & Magdalena Balmaseda in *Climate Dynamics*, https://doi.org/10.1007/s00382-020-05607-6, (CC BY 4.0).

chance of warming in year 1, but only an 18% chance of a strong El Niño, much lower than the 1997 prediction. For year 2 (Dec 2015), they predict high probabilities (58%) for the continuation of warm conditions and only a small chance of a La Niña event (4%).

Perturbed forecasts

The perturbed forecasts exhibit a marked shift in ENSO probabilities compared to their respective reference, especially in year 2. The perturbed 2014–16 forecast shows a more than doubled (44%) probability of a strong El Niño event in year 1 compared to the reference forecast (see the figure). Diagnostics and further experiments showed that the main reason for this change is the increased probability of a positive Indian Ocean Dipole event, which promoted development of a strong El Niño event through atmospheric teleconnections. In year 2, the perturbed forecasts exhibit an enhanced probability of a La Niña event (18%). Analysis of the ocean heat budget (not shown) reveals that the perturbed forecast produces a much stronger Pacific heat loss than the reference forecast during 2014–15, mainly due to much stronger ITF transports in the perturbed experiment, thus preparing the ground for the switch to cooler La Niña conditions. The perturbed 1997–98 experiment confirms that the Indian Ocean initial conditions matter for the 24-month ENSO predictions. Compared to the reference forecast, the likelihood of a La Niña event in year 2 is much reduced (20%). Similar to the situation in 2014–16, this is due to much weaker Pacific Ocean heat loss, which is caused by reduced ITF transports.

For further details, see our article in *Climate Dynamics* published in 2021, https://doi.org/10.1007/s00382-020-05607-6.

ECMWF meets its representatives

Becky Hemingway, Anna Ghelli, Xiaobo Yang, Carsten Maass

On 28, 29 and 30 October 2020, ECMWF virtually met with its Computing Representatives, Meteorological Representatives and Catalogue Contact Points, respectively. The meetings were an opportunity for ECMWF to present updates on a variety of topics and to understand any issues our users in the Member and Co-operating States face. They also provided a forum for the 41 attendees to share experiences and network with other Representatives. The Meteorological User Forum for ECMWF's Meteorological Representatives was the first meeting of its kind.

Representatives are instrumental in ensuring there is a continuous and effective link between ECMWF and Member and Co-operating States. The role includes two-way communication and information flow, including for model updates, events, training and liaison visits, facilitating various administrative transactions between ECMWF and countries that have access to ECMWF's services, and providing feedback on technical solutions. ECMWF supports its Representatives by providing services, listening and acting on feedback received, and organising events and training for Member and Co-operating State users.

All attendees agreed that they found the meetings useful and that they would like them to take place again in the future. ECMWF looks forward to future engagement with its Representatives, including potentially meeting all Representatives together.

IT User Forum 2020

Computing Representatives were invited to the IT User Forum 2020 (formerly the Computing Representatives' Meeting), which took place on 28 October 2020. The meeting, attended by 16 Computing Representatives, focused on providing updates on Bologna Our New Datacentre (BOND). It highlighted the actions that need to be undertaken by Member and Co-operating States to ensure the transition is as smooth and least impactful as possible for users.

Presentations from ECMWF included BOND programme updates, the European Weather Cloud and machine learning. Representatives were invited to provide input on the Test and Early Migration System (TEMS), remote access and working from home. Attendees provided a recorded presentation of updates from their organisations and countries, which highlighted that national meteorological and hydrometeorological services (NMHS) continue to prosper with exciting projects and many achievements to celebrate.

Meteorological User Forum

Meteorological Representatives were invited to the Meteorological User Forum, the first meeting of its kind for this group of people. The meeting was attended by 13 Meteorological Representatives with a variety of experience in the role. The key aims of the meeting were to understand and define their role as a Meteorological Representative and their expectations of ECMWF, get to grips with any issues faced, and facilitate networking within their community. Presentations from ECMWF included new ECMWF forecasting products, the Severe Event Catalogue and updates on model cycles and BOND.

Computing Representatives' meeting. ECMWF organised all three meetings remotely, including the Computing Representatives' meeting shown here.

Prior to the meeting, attendees were asked for their views on the role of a Meteorological Representative, and during the meeting a summary was presented and discussed. The key task raised was to facilitate two-way communication and feedback as well as being the local organiser of liaison visits and contributing to ECMWF activities. ECMWF's role was seen as providing relevant information, training and feedback opportunities and as improving products, services and modelling. Attendees agreed these were useful summaries of the role.

A Meteorological Representatives' Portal has recently been set up in Confluence and includes information on the role of a Meteorological Representative as discussed at the meeting, and other useful information.

Catalogue Contact Points meeting

The Catalogue Contact Point (CCP) meeting was attended by 12 participants from Member and Co-operating States. The aims were to understand issued faced by CCPs and to share and learn from each other's experiences, in particular on the topic of moving towards open data.

The meeting included three insightful presentations from Manon Vignes (Météo-France), Arne Kristensen (MET Norway) and Frank Lantsheer (Royal Netherlands Meteorological Institute), who talked about different elements of supplying ECMWF data to customers from a Member State perspective.

The ECMWF Data Support Team presented new tools which have been developed to help customers manage their data: the shopping cart tool (Product Requirements Catalogue, https://apps.ecmwf.int/shopping-cart/) gives users the ability to choose their required parameters and provides costings, and the Products Requirements Editor manages real-time user requirements, allowing configuration of products from different catalogues.

More information on the groups mentioned here can be found on the ECMWF website: https://www.ecmwf.int/en/about/who-we-are/representatives

New climate projection datasets in Copernicus

András Horányi, Anca Brookshaw, James Varndell

The Copernicus Climate Data Store (CDS) has recently been extended with the introduction of new global climate projection data from CMIP6 and regional simulations from CORDEX. The CDS is the main user interface of the Copernicus Climate Change Service (C3S) implemented by ECMWF on behalf of the EU. The recently published data is a key input to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC).

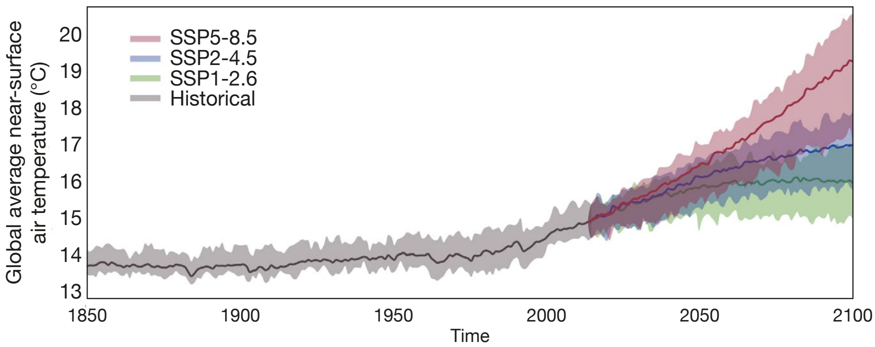
Nature and origin of the data

Climate projections come from global and regional climate models and provide information about the future evolution of the planet's climate system at global and local scales. This quantitative information about climate change is highly relevant for developing adaptation strategies for the future and thus an important component of the catalogue of C3S datasets.

The climate projections in the CDS (https://cds.climate.copernicus.eu) are subsets of data generated under international climate model intercomparison activities, like CMIP (Coupled Model Intercomparison Project) for global projections and CORDEX (Coordinated Regional Climate Downscaling Experiment) for regional projections. The CMIP and CORDEX projects design the intercomparison experiments and collect global/regional climate projection information from climate modelling centres all over the world. These datasets are archived at the Earth System Grid Federation (ESGF), a federated infrastructure maintained by and primarily aimed at research institutions. C3S relies on the ESGF datasets but makes them more appropriate for climate service applications, by creating a robust and resilient system, which ensures continuous availability and includes additional quality control on the data.

Data service

The publication of CMIP6 data in the CDS is an important milestone, providing users of C3S with very new, state-of-the-art global climate projection data, which will be intensively used and referred to in the coming years. The main features of the dataset can be summarised as follows:



Visualisation of a subset of CMIP6 data generated in the CDS Toolbox. The plot shows annual global mean 2 m temperature between 1850 and 2100 for selected models. The grey area shows the range over the historical period, and the coloured shades show pessimistic (red), realistic (blue) and optimistic (green) future greenhouse gas emission scenario simulations.

- Historical simulations (validation runs for the past) and scenario simulations (future projections) are available. Simulations under a number of scenarios are available (see the CDS dataset documentation).
- Around 50 variables from more than 50 different Global Climate Models (GCMs) will be available. These are represented on single levels (typically surface, but also top of the atmosphere) or pressure levels (between 1,000 hPa and 1 hPa).
- The data are available at monthly and daily temporal resolution.
- Computational processes will be on offer in this catalogue entry with subsetting options already included in the interface. This is a new feature of the CDS catalogue, which exploits processing services run at the location of the data, thus making it unnecessary for users to download large datasets. Sub-setting can be done in the vertical direction (choice of pressure levels or surface), horizontally (geographic selection), or in time (selection of the time period required).
- Easy data manipulation and visualisation in the CDS. For example, the figure was created by a simple workflow in the CDS Toolbox.

C3S implicitly supports the IPCC assessment through its contribution to the IPCC interactive climate atlas. This is a tool to visualise and assess different global and regional climate projections and will complement the IPCC AR6 report. C3S has provided support to upload many CORDEX simulations for regions outside Europe to ESGF, and to quality-control European and non-European climate projection data for the regions designated by CORDEX. For Europe, C3S has also significantly enhanced the number of high-resolution simulations, to better sample uncertainties coming from the global and regional models, scenarios and internal climate variability, and published these in the CDS. Data covering other CORDEX domains around the world is now available to users of the CDS, with the Mediterranean, North American, Arctic, African and South American regions recently published, and more domains to come soon.

Outlook

By the end of 2022, the IPCC will release its assessment report based on CMIP6 model simulations. Interest in these climate projections already is and will remain high, in the context of climate services and thus among users of the CDS. We anticipate the number of users accessing CMIP6 data to far exceed the more than 1,000 who have made use of the CMIP5 CDS data in the last two years. In the next phase of the Copernicus programme, C3S will add more global and regional projection data from CMIP6 and its regionally downscaled simulations to the CDS. Better tools and applications are also in the pipeline for the CDS infrastructure to enhance the value of climate projections to users.

Scalable Acquisition and Pre-Processing system used more widely

Saritha Gudur

Some ECMWF Member and Co-operating States have begun to use the Centre's operational acquisition and pre-processing system for observations and other input data. The Scalable Acquisition and Pre-Processing (SAPP) system has been open in an Optional Programme since 2019 to support the provision of SAPP to participating states.

SAPP webinar 2021 participants. Due to COVID-19 restrictions, the SAPP workshop was cancelled and an online seminar via Zoom was conducted on 10 March 2021.

Progress made

Participating states have used SAPP to varying degrees. Some have only very recently joined the programme, whereas a few others have adopted it operationally. Ireland, which was one of the early adopters, is now using the SAPP system operationally and has demonstrated a marked improvement in their capability to perform operational acquisition and pre-processing of observations for numerical weather prediction (NWP). Notably, the implementation of the SAPP system has yielded a substantial increase in the number of observations they are able to use in their data assimilation.

In July 2020, the Turkish State Meteorological Service (TSMS) joined Ireland and adopted SAPP operationally. They are currently mainly using it to provide BUFR observations to the TSMS Numerical Weather Prediction Division in order to run their Local Model and data assimilation. A few countries, such as Denmark and Italy, have demonstrated strong interest in using SAPP operationally soon. Overall, there has been considerable progress towards optimizing the use of SAPP by ECMWF Member and Co-operating States.

SAPP webinar and training

An online seminar in support of the SAPP Optional Programme was successfully held via zoom on 10 March. More than 30 representatives from 16 participating and non-participating States attended the online event. The objectives of the webinar were to enhance the SAPP uptake and to expand the programme. The webinar also aimed to encourage usage of the SAPP system by early users. The attendees were given an overview of the SAPP system, provision of user services to participating States, the latest software updates, and other agenda topics. Ireland and Turkey presented their testimonials about their operational use and SAPP customisation. This webinar acted as catalyst for some countries who were close to making SAPP an operational tool. After the webinar, Iceland expressed its intention of going fully operational. The overall feedback from the webinar was positive, with participants considering it to be relevant, useful and aimed at a diverse level of users. Online training sessions will be integrated with face-to-face events in 2022.

As a part of the training initiatives, earlier this year an online training module was developed and made available in the ECMWF learning portal. This course gives a comprehensive view of the SAPP system through Jupyter Notebook. The module would greatly benefit early SAPP users with little or no experience of SAPP. It includes an introduction to the SAPP Optional Programme to non-participating Member and Co-operating States as part of the expansion drive.

The way forward

In November 2019, ECMWF's Jordan Rice and Mikko Aalto (Secondee from Finland) created a beta version of the SAPP Containerised project. The project's aim was to explore the feasibility of containerising the SAPP system. This project has since been released to a few Member States for testing purposes. Users were able to build the containers from scratch and run them successfully.

The containerised solution has demonstrated that it can handle the same data as the current SAPP system and has other advantages. Work is in progress towards making the containerised solution available to all participating Member and Co-operating States. Further updates on this project will be made available on the SAPP Confluence page (https://confluence.ecmwf.int/display/SAPP/SAPP+Home).

Obituary: Anders Persson

David Richardson

Anders Persson in front of the screens at ECMWF (June 2013).

Anders worked in the Meteorological Operations section during the 1990s, a key decade which saw the introduction of medium-range ensemble forecasts in 1992 and the start of the evolution of probabilistic forecasting to the central role it plays today.

Understanding was important to Anders, and he raised many challenging questions that helped to clarify the ensemble approach. He became a great advocate of the benefits of ensembles and put a lot of effort into communicating these to forecast users at every opportunity. This culminated in the development of the ECMWF Forecast User Guide, which presented clearly and in detail the practical approaches by which users could take advantage of the ensemble and integrate this into their work.

Another major component of Anders' work at ECMWF was the day-to-day operational monitoring of the performance of the forecasting system, and the investigation of cases of poor performance and systematic errors. Anders pioneered the use of several diagnostic approaches, including error tracking to try to identify the sources of forecast error. By always questioning forecast performance and actively following up with staff across the Centre, he helped to ensure that the ECMWF forecasting system continued to improve. Again, much of Anders' insight into these areas was captured in the Forecast User Guide.

After leaving ECMWF in 2001, Anders always maintained a strong link to the Centre, visiting on many occasions and always keeping a keen eye on forecast performance. During a visit to ECMWF in 2010–11 to prepare a new edition of the Forecast User Guide, Anders was interviewed for the ECMWF Newsletter. This conversation provides a fitting testimony in his own words to all his experience over the years at the Centre: https://www.ecmwf.int/sites/default/files/elibrary/2011/17430-increasing-trust-medium-range-weather-forecasts.pdf

Anders has had a lasting influence at ECMWF and will be well remembered also by the many users of ECMWF forecasts who attended training courses and workshops at ECMWF, where he was always passionate about promoting the use of ensemble forecasts. Anders died on 23 January 2021.

I was really saddened and shocked, like all of us, by this news. Anders was still so active and passionate the last time I saw him not that long ago. I have very fond memories of Anders since my first stay at ECMWF in the 1990s. He never lost touch afterwards and had so many intellectually challenging discussions with many of us. He will be sadly missed but not forgotten!
(ECMWF Director-General Florence Rabier)

Many of us will be saddened to hear of the death of Anders Persson. Anders chose a career in meteorology in 1964 in the mistaken belief that meteorology was a simpler science than quantum mechanics or relativity theory. He followed at close range – and with great interest – the spectacular development of numerical weather prediction.
(Austin Woods)

Anders was a wonderful colleague to share ideas with. He was very widely read and would engage pretty much on any topic you wanted to. Why, he would argue, do most of the ensemble members have a higher root-mean-square error than the unperturbed control? Didn't that mean something was wrong? I will miss our intellectual debates.
(Tim Palmer)

I am really very sorry to hear this sad news. I worked a lot with Anders. I liked to call him 'the ensemble conscience', since he was always asking very appropriate and challenging questions about it. Thanks to his questions and engaging discussions, I think we did indeed check more carefully the language we were using.
(Roberto Buizza)

I met Anders for the first time around 1980 when he came to spend a few weeks in Paris to see how French forecasters were working. At ECMWF we often had long discussions about the meteorological situation of the day. After I left ECMWF, he contacted me several times on scientific-historical subjects, and I learnt a lot from the subsequent exchanges with him.
(Jean Pailleux)

I had the difficult task, and the honour, to replace Anders as a synoptic meteorological analyst at ECMWF. I did it with the same enthusiasm and fascination that he transmitted to me during my training. He is still in the Italian forecasting community's memory after the many exciting lectures he used to give at the MeteoTrentino courses.
(Federico Grazzini)

Data assimilation or machine learning?

Massimo Bonavita, Alan Geer, Patrick Laloyaux, Sébastien Massart, Marcin Chrust

Machine learning and deep learning (ML/DL) techniques have made remarkable advances in recent years in a large and ever-growing number of disparate application areas, such as natural language processing, computer vision, autonomous vehicles, healthcare, finance and many others. These advances have been driven by the huge increase in available data, the corresponding increase in computing power and the emergence of more effective and efficient algorithms. In Earth sciences in general, and numerical weather prediction (NWP) and climate prediction in particular, ML/DL uptake has been slow at first, but interest is rapidly growing. Today innovative applications of standard ML/DL tools and ideas are becoming increasingly common, and ECMWF has recently set out its approach in a 'roadmap for the next 10 years' (Düben et al., 2021).

The interest of Earth scientists in ML/DL techniques stems from different perspectives. On the observation side, for example, the current and future availability of satellite-based Earth system measurements at high temporal and spatial resolutions and the emergence of entirely new observing systems made possible by ubiquitous internet connectivity (the so-called 'Internet of Things') pose new challenges to established processing techniques and ultimately to our ability to make effective use of these new sources of information. ML/DL tools can potentially be useful to overcome some of these problems, for example in the areas of observation quality control, observation bias correction and the development of efficient observation operators.

ML/DL approaches are also interesting from the point of view of data assimilation, in other words the combination of the latest observations with a short-range forecast to obtain the best possible estimate of the current state of the Earth system. This is because such approaches can be typically framed as Bayesian estimation problems which make use of a similar methodological toolbox to the one used in variational data assimilation. This fact has been recognised in the new ECMWF Strategy 2021–2030, which acknowledges that "4D-Var data assimilation is uniquely positioned to benefit from integrating machine learning technologies because the two fields share a common theoretical foundation and use similar computational tools". It is apparent that some of the techniques already widely adopted in the data assimilation community (e.g. model error estimation, model parameter estimation, variational bias correction) are effectively a type of ML algorithm. The question is then, what lessons can the NWP/climate prediction community learn from the methodologies and practices of the ML/DL community? How can we seamlessly develop the ML/DL techniques already used in the NWP/climate prediction workflow and integrate new ideas into current data assimilation practices?

In this article we briefly discuss the underlying theoretical equivalence between today's data assimilation and modern ML/DL. We then provide some concrete examples of how the adoption of ML/DL ideas in the NWP workflow has the potential to extend the capabilities of current data assimilation methodologies and ultimately lead to better analyses and forecasts.

Machine learning: a form of data assimilation

The aims of data assimilation and machine learning are similar: to learn about the world using observations. In traditional weather forecasting we assume we have a reasonably accurate physical model of the Earth system, and the biggest unknown is the initial conditions from which to start the forecast. In machine learning the physics is unknown, and the aim is to learn an empirical model directly from the observations. In both fields, Bayes' theorem provides an underlying description of how to incorporate information from observations. This gives a way of mathematically comparing the two approaches.

Rather than show the equations, Figure 1 gives the Bayesian network diagram for both approaches (Geer, 2021). The black text represents data assimilation, such as the 4D-Var system used in ECMWF's Integrated Forecasting System (IFS). The arrows together represent the action of the forecast model and observation operator, which take the model state of the Earth system (x) and the parameters of the physical model (w) and provide estimates of the observations (y). By comparing these estimates to real observations, data assimilation finds a better estimate of the initial conditions (x). In the IFS, we do not yet try to improve the parameters of the model at the same time. Such a process is known as parameter estimation and it is routinely done in fields like chemical source inversion or groundwater modelling. However, we already estimate some parameters in 4D-Var: this is done in the weak-

doi: 10.21957/ut51mb7c39

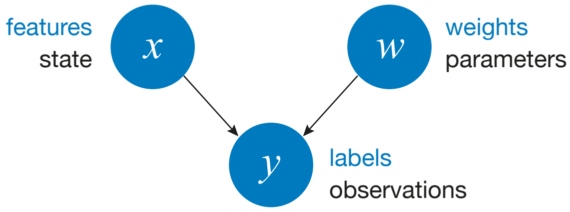


FIGURE 1 Machine learning (blue text) and data assimilation (black text) represented as a Bayesian network. Circles are probabilistic variables; arrows are dependencies representing, in data assimilation, a physical model of the Earth system and the observation forward model, or in machine learning, a neural network. Typical data assimilation is a Bayesian 'inverse problem' that holds the parameters of the model, *w*, constant to estimate the geophysical state of the atmosphere, *x*, from observations *y*. Training a neural network is a Bayesian inverse problem that uses a large dataset of features *x* and labels *y* to estimate the weights *w* of a neural network. See Geer (2021).

constraint term, which corrects model bias in the stratosphere, and in variational bias correction, which corrects observational bias. Research on estimating model parameters in the IFS is only just starting and is an exciting prospect (see later).

In a typical form of machine learning like a neural network, the training is done between data pairs, known as features (*x*) and labels (*y*). These are shown in blue text in Figure 1. A feature might be a picture, the label might be 'cat'. A neural network is trained on a huge set of labelled features, with the aim to estimate weights (*w*) for a neural network. The trained network can then be used, in this example to identify animals in pictures. But the technique is very flexible, and applications range from playing games with higher skill than the best human players to learning to translate from one language to another.

The Bayesian network in Figure 1 is a mathematical abstraction, but by making a few assumptions we can get to the usual forms of variational data assimilation and machine learning, both of which minimise a cost function (known as a loss function in machine learning) to find the best *x* or *w*, or both (see Box A). The minimisation of the cost function in machine learning is usually done through a process called 'back-propagation'. The same process takes place in data assimilation when we calculate the gradient of the cost function using the adjoint of the observation operator and the forecast model. As shown in the box, there is a mathematical equivalence between the processes of machine learning and those of variational data assimilation.

One big advantage of data assimilation is that it can incorporate constraints from prior knowledge, such as the physics of the atmosphere, which helps provide a more accurate result than can be learnt solely from the observations. There is now a rapid development of 'physically constrained' machine learning approaches, but we might wonder if the end result will be to re-invent data assimilation. Machine learning's strength is to operate in areas where there is scarce or no existing physical knowledge to constrain the solution – for example, there is no rule-based algorithm available for translating from one language to another. Within data assimilation for weather forecasting, the obvious place to use machine learning is also where we lack physical

a

The aim of variational data assimilation or the training phase in machine learning is to reduce the cost function $J(x,w)$ as much as possible by varying x and w. Here x and w are as defined in Figure 1, in other words state and parameters in data assimilation, or features and weights in machine learning:

$$J(x,w) = \underbrace{\frac{(y-h(x,w))^2}{(\sigma^y)^2}}_{J^y} + \underbrace{\frac{(x^b-x)^2}{(\sigma^x)^2}}_{J^x} + \underbrace{\frac{(w^b-w)^2}{(\sigma^w)^2}}_{J^w}$$

The minimum of $J(x,w)$ gives the maximum likelihood estimate of x and w. To find this, we must reduce the misfit between the model-simulated observations (or the neural-network simulated labels) and the real observations, which is measured by the observation term in the cost function, J_y, while not increasing the other terms in the cost function too much. Here y is the observation value and $h(x,w)$ represents either the neural network or the physical model (including observation operator) in 4D-Var. For simplicity all the variables are scalars. Data assimilation weighs the observations according to their accuracy, here shown by the observation error standard deviation σ^y. Machine learning does not explicitly represent the observation error, but the process of data pre-processing (quality control, batch normalisation) can represent some (but not all) sources of observational uncertainties. The J_x term is what we call the background term in 4D-Var, with x^b being the background forecast and σ^x the background error. This has no equivalent in machine learning, which assumes the features (e.g. the cat pictures) are free from error. The final term, J_w, constrains the estimated weights. In data assimilation, there are background estimates for the parameters, w^b, which have an error σ^w. In machine learning, the equivalent is ridge regression, which regularises the weights, but the background weight estimate is 0 and its error standard deviation is 1.

knowledge. This leads to some of the examples in this article, which explore learning the patterns of bias in the forecast model or in the observations, where this is hard to understand by purely physical reasoning. The highly accessible software tools provided by the machine learning community have made it easy to get started. A second scenario is where we might be reasonably confident in the equations of the physical system, but the parameters, such as roughness length in the boundary layer turbulence scheme, are not well known and may vary geographically on fine scales. This is probably better done using parameter estimation within the existing data assimilation framework. Another big application area for machine learning, which is not covered here, is not directed at improving the models but instead to produce faster statistical models from existing, slower and more computationally demanding physical models (i.e. model emulation).

Model error estimation and correction

ECMWF continues to make good progress in driving down model error, which together with initial uncertainties is one of the main obstacles to improved accuracy and reliability of weather and climate predictions. However, realistically there will always be some residual error. Recent advances in the context of weak-constraint 4D-Var showed that it is possible to estimate and correct for a large fraction of systematic model error which develops in the stratosphere over short forecast ranges (Laloyaux & Bonavita, 2020). In each data assimilation cycle, weak-constraint 4D-Var (WC-4DVar) optimises the fit of the short-range forecast trajectory to observations in two ways: by suitably changing the initial conditions (as in strong-constraint 4D-Var), and by suitably adjusting a forcing term in the model equations to correct for the model error over the assimilation window.

An Artificial Neural Networks (ANNs) approach has been recently investigated as a way to predict the analysis increments by means of ML/DL. The systematic component of those increments can be viewed as another estimate of the cumulated model error over the assimilation window (Bonavita & Laloyaux, 2020). A set of climatological predictors is selected (latitude, longitude, time of day, month) to capture the part of model error which is related to geographical location, the diurnal cycle, and the seasonal cycle. State predictors are also used to represent the part of model

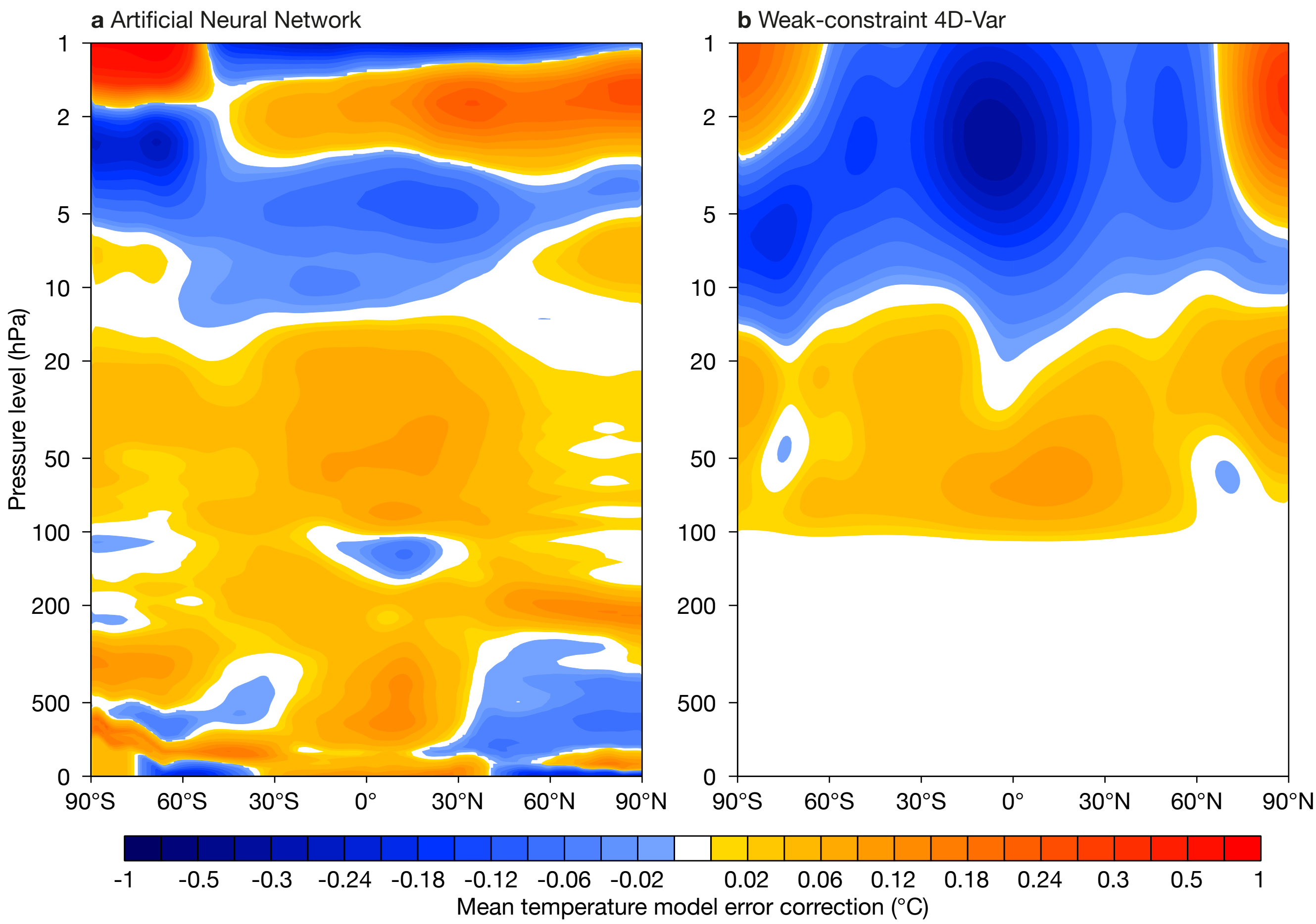


FIGURE 2 Zonal cross section of the mean temperature model error correction estimated by (a) an Artificial Neural Network and (b) weak-constraint 4D-Var, averaged between 15 and 24 August 2019.

error linked to the large-scale state of the flow, for example in oceanic stratocumulus areas, the Intertropical Convergence Zone, and extratropical cyclonic areas. To produce the input and the target of the neural network dataset, these predictors were extracted from the operational first guess and analysis increment for mass, wind, and humidity over the whole year 2018. Fully connected ANNs with multiple hidden nonlinear layers are trained and then used to produce model error tendencies over August 2019. These tendencies are applied to the model equations similarly to what is done in weak-constraint 4D-Var. The plot in Figure 2 compares the temperature model error estimates from ANN and from weak-constraint 4D-Var (applied only in the stratosphere at the moment).

WC-4DVar corrects the warm bias in the upper stratosphere and the cold bias in the mid/lower stratosphere. It correctly captures the transition layer (20 to 10 hPa) where the model bias changes from cold to warm. The transition level between the cold and the warm bias layers is estimated at the same pressure level by the ANN. The main difference is in the upper stratosphere, where the neural network produces a positive correction of around 2 hPa. This is because this signal is present in the analysis increments used to train the neural network. On the other hand, WC-4DVar extrapolates the cold bias to the top of the model, due to the deep vertical structure of the prescribed model error covariance matrix.

WC-4DVar is an online example of a statistical estimation technique as the algorithm updates the model error correction after each assimilation window. The current ANN produces a static correction dataset computed prior to the assimilation of the observations, but work is ongoing to apply the ANN online in a 4D-Var system. In this context, the neural network will be applied in every assimilation window to the actual first-guess trajectory, instead of being applied offline to the pre-computed first-guess trajectories of the operational suite.

Model parameter estimation

Parameter estimation is widely used to make a model better fit a target. From this perspective, parameter estimation techniques can be considered as machine learning techniques. Estimating model parameters as part of a data assimilation process has been little studied so far. In such a framework, the parameters are re-estimated at each assimilation cycle using present information and past accumulated knowledge on the parameter.

The ability to estimate a parameter within the IFS 4D-Var was demonstrated a few years ago with the estimation of the solar constant, which is a model parameter constrained indirectly by satellite observations. We are currently testing this idea on the sub-grid scale orography parametrization, which is a component of the physical parametrizations with a significant impact on the surface pressure field. The standard deviation of

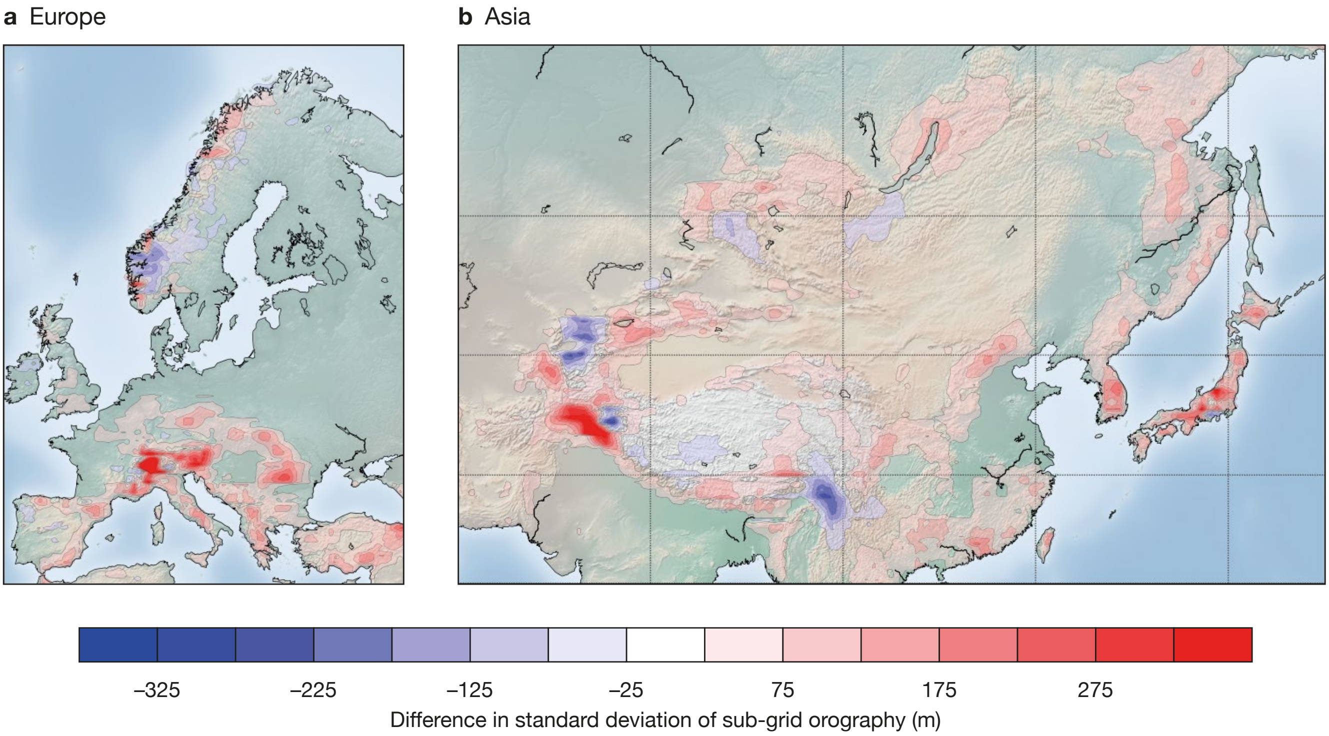


FIGURE 3 Example of where the 4D-Var analysis wants to increase (red) or decrease (blue) the drag over (a) Europe and (b) Asia (note the largest corrections over areas with complex orography).

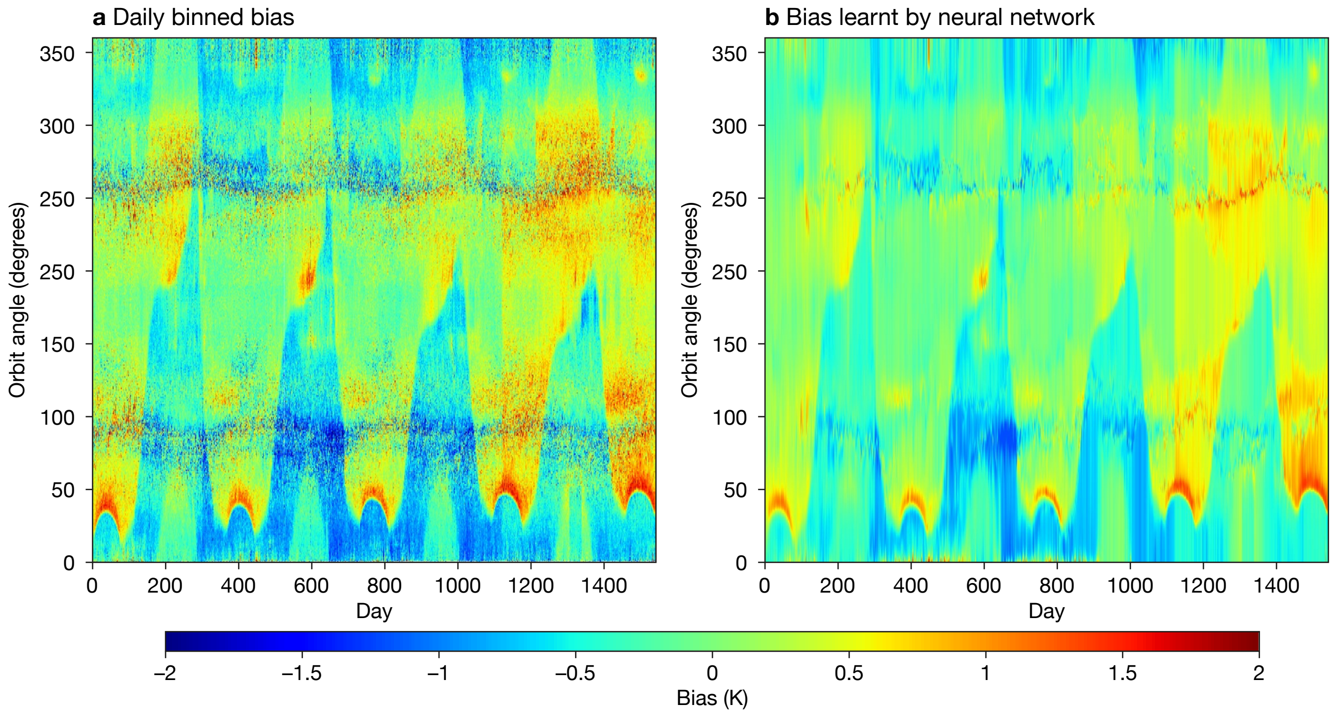


FIGURE 4 Bias diagrams showing (a) daily binned bias in Kelvin between SSMIS F-17 channel 11 observations and the simulations from the IFS, stratified by the angle around the orbit, and (b) bias learnt by an evolving neural network that is then used as a bias prediction model for the next day.

sub-grid orography is well known as a key variable of the sub-grid scale orography parametrization. As a demonstrator, we consider it as a parameter to be optimised. As the surface pressure is relatively well observed, at least over landmasses, 4D-Var produces a correction to the standard deviation of sub-grid orography, which helps to assess where the drag should be modified according to the surface pressure measurements (Figure 3). This demonstrates the potential of obtaining information on the sub-grid scale orography parametrization.

The next step is to try to optimise directly the parameters of the sub-grid scale orography parametrization instead of the well-known standard deviation of sub-grid orography. This parameter estimation step can be performed either using machine learning techniques offline, or directly inside the 4D-Var minimisation.

Bias correction of observations

Observational bias correction might be a good application for machine learning. If we knew the physical explanation for a bias, we would correct it at source in the forecast model or observation operator. But it is often difficult to find physical explanations for biases between the simulated observations and their real equivalents, so biases are often corrected using empirical models. In the IFS, variational bias correction (VarBC) is one example. Currently it only allows linear bias models as a function of predictors like layer thickness, skin temperature or satellite viewing angle. Nonlinear models can be added in VarBC by using linearising transforms, such as terms from the Fourier transform or a polynomial expansion. The latter is already used with the satellite scan angle. However, these transforms are only capable of fitting some nonlinear behaviours. For a completely free-form nonlinear empirical fit, a neural network could be a good alternative. It might even be possible to use a neural network as a bias model within VarBC, although that has not yet been tested. On the other hand, a neural network with a single layer and a linear activation function is just an offline version of VarBC, in other words they both simplify to multiple linear regression.

We show an example using an offline neural network bias correction applied to the solar-dependent bias of the Special Sensor Microwave Imager/Sounder (SSMIS). This varies through the orbit and through the year in a complex pattern (Figure 4a). The red arches near the bottom of the plot are the most difficult part: at certain times of the year, the instrument calibration changes rapidly as the satellite emerges from the dark into the sunny part of the orbit. VarBC absorbs some of this bias but cannot deal with these sharp features (not shown). However, VarBC has the advantage of being able to evolve through time, following the changing bias as the satellite's orbit drifts. The difficulty of applying a neural network is, first, that there is a lack of training data: one year's bias cycle is just a single training point; second, it would need to be continually re-trained to keep up with the evolving bias; in machine learning it is more typical to train once and then use a fixed network for prediction.

Figure 4b shows a simple adaptation, where the neural network is only expected to predict the bias one day ahead, and it is re-trained every day on the new day's data, so that it slowly adapts. The neural network produces a smooth fit to the strongly nonlinear variations in bias through the orbit. With the cycled training, it can also evolve through time. This is an example of 'transfer learning', where a pre-trained neural network is re-trained on new data. However, standard machine learning is only just starting to evolve the tools to control the learning rate of such a network, in other words to control how fast it evolves. Further, it is hard to specify how much of the information gets fitted; the big worry with VarBC has always been that, with too many degrees of freedom, useful observational information could end up being aliased into the bias correction rather than being used, as intended, to update the initial conditions.

To move from multiple linear regression towards a free-form nonlinear empirical model is just a possible evolution for VarBC. There is no binary choice between 'data assimilation' or 'ML'. The real issues are more technical: whether the solver in 4D-Var is capable of fitting a neural network embedded in 4D-Var (or whether quick gains could be made with an offline neural network of the type explored here); how to constrain the shape of the empirical fit so that useful observational information goes into the atmospheric state analysis and not the bias model; and how to constrain the evolution of the bias model through time. These issues are equally relevant for any data assimilation or ML approach.

Outlook

From a data assimilation perspective, ML/DL does not introduce completely new or revolutionary ideas. As we have shown here, ML/DL techniques have much in common with the standard workflow of variational data assimilation, though these similarities tend to be partly obfuscated by the different nomenclature used in the two fields. These similarities make it conceptually easy to integrate standard ML/DL techniques and ideas into the 4D-Var workflow. However, the practical and technical challenges, e.g. arising from the use of different programming languages or the difficulties associated with developing and deploying online ML/DL algorithms, should not be underestimated. We have shown some concrete initial examples of how this can be done in the IFS, focussing on improving and correcting the models used in the data assimilation cycle. There are obviously many other possibilities and application areas in the whole NWP workflow (as discussed in Düben et al., 2021). As the number of possible developments at the intersection of data assimilation and machine learning is vast, collaborations with the broader research community from both Member and Co-operating States and academia is crucial to accelerate the pace and quality of developments. An example of this type of collaboration is the joint work with ECMWF Fellow Marc Bocquet and his group at the École des Ponts ParisTech, which has already resulted in significant insights on how to more closely embed machine learning techniques in the data assimilation system for the estimation and correction of model errors (Farchi et al., 2021).

More generally, the most recent ML/DL wave of interest has brought into renewed focus the opportunities offered by data-driven techniques to improve and/or correct our knowledge-based models, thanks to the huge and ever-increasing amount of accurate Earth system observations available for NWP and climate prediction initialisation. However, for reasons discussed in e.g. Bonavita & Laloyaux (2020), while it is clear that these tools can be useful complements and additions to our physics-based models, it would be naïve to believe that they can replace them entirely.

Further reading

Bonavita, **M.** & **P. Laloyaux**, 2020: Machine learning for model error inference and correction. *Journal of Advances in Modeling Earth Systems*, **12**, e2020MS002232. *https://doi.org/10.1029/2020MS002232*.

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. *ECMWF Technical Memorandum* **No. 878**. *https://doi.org/10.21957/ge7ckgm*.

Farchi, **A.**, **P. Laloyaux**, **M. Bonavita** & **M. Bocquet**, 2021: Using machine learning to correct model error in data assimilation and forecast applications, *Quarterly Journal of the Royal Meteorological Society* (under review).

Geer, **A.J.**, 2021: Learning earth system models from observations: Machine learning or data assimilation?, *Phil. Trans. R. Soc. A*, **379**: 20200089. *https://doi.org/10.1098/rsta.2020.0089*.

Laloyaux, **P.** & **M. Bonavita**, 2020: Improving the handling of model bias in data assimilation. *ECMWF Newsletter* **No. 163**, 18–22.

ECMWF online – community, collaboration and empowerment

Helen Setchell, Sylvie Lamy-Thépaut, Andrew Brady

The web enables global collaborative efforts to advance numerical weather prediction and gives us the ability to disseminate our products. We are committed to continuously developing and improving the ECMWF web, our online interface with the world. The aim is to support ECMWF's Strategy and respond to changes in technology and the behaviours and needs of our users. By building an ECMWF web which is useful and usable, we provide a place where an online community of empowered users can collaborate and share knowledge. Through user-friendly and intuitive online interfaces, they help themselves to make effective use of ECMWF products and services.

Since our last update on web developments at ECMWF (Setchell, 2018), we have significantly updated those areas of our web which bring most value and are most used. We have made our real-time charts freely available with an open licence (CC-BY-4.0) through a new and improved chart browser, and we have improved the performance of our interactive online charts, 'ecCharts'. Users can also expect to see on-demand plotting for observation statistics in the future. For those wanting to experiment with data, we have created new online 'hubs'. We have also begun work on a new data browser to make our data easier to discover, select, and use. We are growing our online communities of stakeholders and users by providing more opportunities for online engagement and knowledge sharing through news, blogs, social media, publications, staff profiles, online events, eLearning courses, a support portal, documentation and forums. All these developments are reliant on technology. We have been keen to use the forthcoming migration of the ECMWF data centre to Bologna to make significant improvements in the performance of the ECMWF web, and to develop a user account and profile service to help manage the relationship between users and ECMWF. More details about these advances can be found below.

Charts

Charts continue to be one of the most highly accessed and used parts of the ECMWF web. They enable users to visualise real-time forecast data at the click of a button.

Open charts

As part of the road map to Open Data, ECMWF opened its web forecast charts to the public in October 2020; all charts became freely available under the terms of the CC BY 4.0 licence (with some additional terms of use). This was the occasion to launch a brand new chart browser application. This new application makes it easier for users to browse and share charts on social media. We also plan to release a widget to enable anyone to embed live ECMWF charts in their website.

More recently we added meteogram tooltips to our charts to help users discover one of the rich features of this application (see Figure 1). We immediately saw an increase in usage, advancing users' understanding of uncertainty in the forecast. The application was launched on a modern Kubernetes infrastructure (see Brady, 2019), ensuring we were prepared for the increase in traffic. This is the first forecast web application deployed in such a way. The others, including ecCharts and the data services applications, will follow shortly, making them highly performant and ready to be migrated to Bologna.

ecCharts: interactive online charts

The development of ecCharts is still ongoing, with the introduction last year of well-established open-source software for the map viewer, generalising the use of tiles for better performance of the cache and making use of vector tiles for static layers. The Web Map Service (WMS) offers access to all ecCharts layers through the OGC WMS standard. The service is becoming popular through the Member and Co-operating States, with new layers constantly requested and, when possible, added to the system.

Migration to a Kubernetes environment and continuous integration and deployment methodology will improve performance and the speed at which we can release new features for users. The possibility of offering access to archive data in ecCharts will also be investigated over the next three years.

On-demand plotting

Improved access to observation monitoring statistics for internal and external users will be enabled by an on-demand-plotting web interface instead of the current

doi: 10.21957/md371hsp83

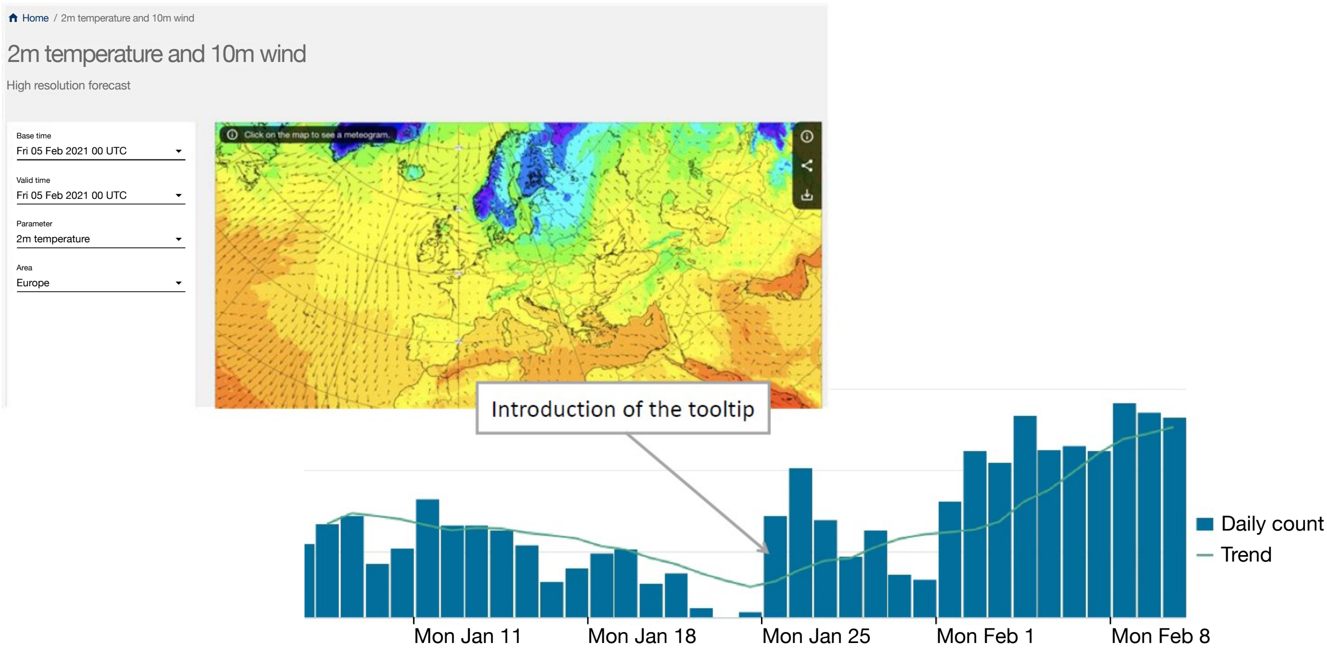


FIGURE 1 Some of the open charts are clickable, giving access to the ENSgram (a type of ensemble forecast meteogram) and other point-based products, such as the vertical profile and the Extreme Forecast Index/cumulative distribution function (EFI/CDF) at the location. A tooltip was introduced at the end of January to highlight this feature, resulting in an immediate increase in its use.

static monitoring pages updated daily or weekly. The new tool will be using the concept and technical elements of ecCharts and will reduce the need to store a large volume of static files.

Data and research

We continue to make our data more discoverable, accessible, and shareable. As requirements have evolved over the years, we have built several user interfaces to our data, each with its own set of features. Recently this has included an interface for customer licensing enquiries and an interface for amending product dissemination requests with the aim of creating a set of self-serve tools. We are now starting a project to rationalise these many interfaces and explore bringing them together into a unified ECMWF data portal. We expect this to significantly help users to quickly discover, select, and access the best data for their purposes.

To make our datasets easier to reference and share, and to improve how they are indexed by search engines, we have begun assigning Digital Object Identifiers (DOIs) and standardised metadata to our data records, and we have made our research experiment data open access when linked to publications. Fellow researchers are now able to follow journal citations to a simple interface which allows them to download the data and reproduce results (Figure 2).

The Integrated Forecasting System hub (IFShub), which is presented in greater detail in a separate article in this Newsletter, will provide a set of higher-level web-based interfaces to set up IFS experiments, monitor progress and obtain a first graphical and statistical look at the outcomes. IFShub will be designed with IFS, OpenIFS, research and operational needs in mind and will allow processes to be controlled from remote web browsers and results to be retrieved from the data archive using the Meteorological Archival and Retrieval System (MARS).

Users will be able to create, share and execute Jupyter notebooks on a JupyterHub hosted at ECMWF, a web-based interactive environment widely used in the research community. This will allow users to combine explanatory text with processing and visualisation in notebooks, which is also very useful for training. Users will then be able to execute these notebooks on ECMWF computational resources from their browser on their laptop.

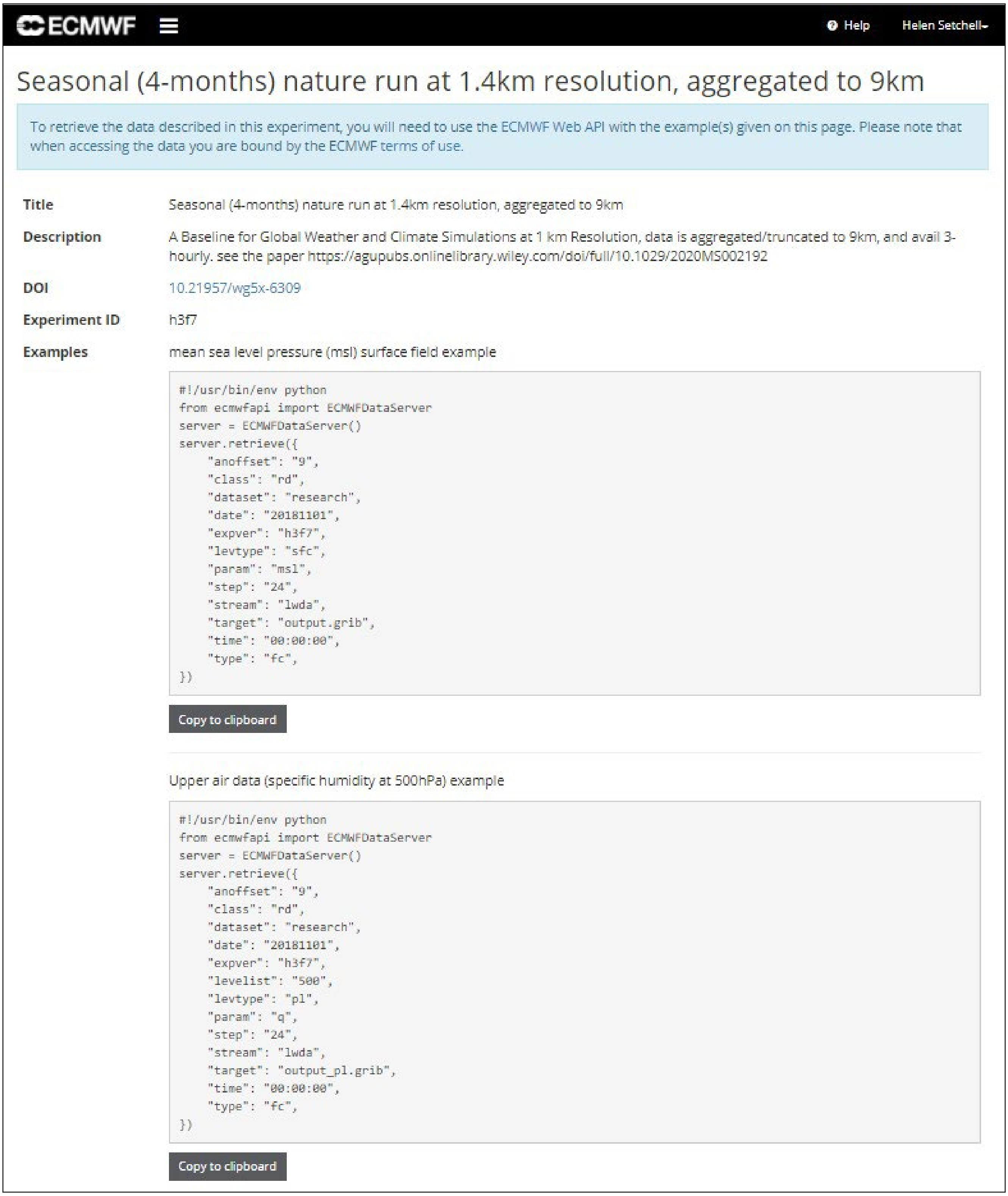
The new System Billing Unit (SBU) application will give Member and Co-operating State users better visibility of their use of the high-performance computing facility (HPCF). It will also enable administrators to reallocate resources between projects, and to account for and manage computing resources.

Community, knowledge sharing, and support

As a leading organisation in the meteorological community, ECMWF uses web technology to build and strengthen our community and advance knowledge.

We provide staff profiles (*https://www.ecmwf.int/en/about/who-we-are/staff-profiles*) to help others find

ECMWF collaborators, and a refreshed online publications repository (*https://www.ecmwf.int/en/publications*), which makes use of Digital Object Identifiers (DOI) and Open Researcher and Contributors IDs (ORCID) to connect ECMWF staff and their publications with the wider community. We also provide blogs and forums which facilitate conversation, and an active social media presence. We aim to continue to grow activity in these areas to build communities and provide more opportunity for collaboration. In the coming years we will look particularly at building smoother and effective links to websites providing third-party activities.


ECMWF ≡ Help Helen Setchell▾

Seasonal (4-months) nature run at 1.4km resolution, aggregated to 9km

To retrieve the data described in this experiment, you will need to use the ECMWF Web API with the example(s) given on this page. Please note that when accessing the data you are bound by the ECMWF terms of use.

Title	Seasonal (4-months) nature run at 1.4km resolution, aggregated to 9km
Description	A Baseline for Global Weather and Climate Simulations at 1 km Resolution, data is aggregated/truncated to 9km, and avail 3-hourly. see the paper https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2020MS002192
DOI	10.21957/wg5x-6309
Experiment ID	h3f7
Examples	mean sea level pressure (msl) surface field example

```
#!/usr/bin/env python
from ecmwfapi import ECMWFDataServer
server = ECMWFDataServer()
server.retrieve({
    "anoffset": "9",
    "class": "rd",
    "dataset": "research",
    "date": "20181101",
    "expver": "h3f7",
    "levtype": "sfc",
    "param": "msl",
    "step": "24",
    "stream": "lwda",
    "target": "output.grib",
    "time": "00:00:00",
    "type": "fc",
})
```

Copy to clipboard

Upper air data (specific humidity at 500hPa) example

```
#!/usr/bin/env python
from ecmwfapi import ECMWFDataServer
server = ECMWFDataServer()
server.retrieve({
    "anoffset": "9",
    "class": "rd",
    "dataset": "research",
    "date": "20181101",
    "expver": "h3f7",
    "levelist": "500",
    "levtype": "pl",
    "param": "q",
    "step": "24",
    "stream": "lwda",
    "target": "output_pl.grib",
    "time": "00:00:00",
    "type": "fc",
})
```

Copy to clipboard

FIGURE 2 Example of a recently published research experiment, including examples of the scripts needed to download the data via the ECMWF Web API.

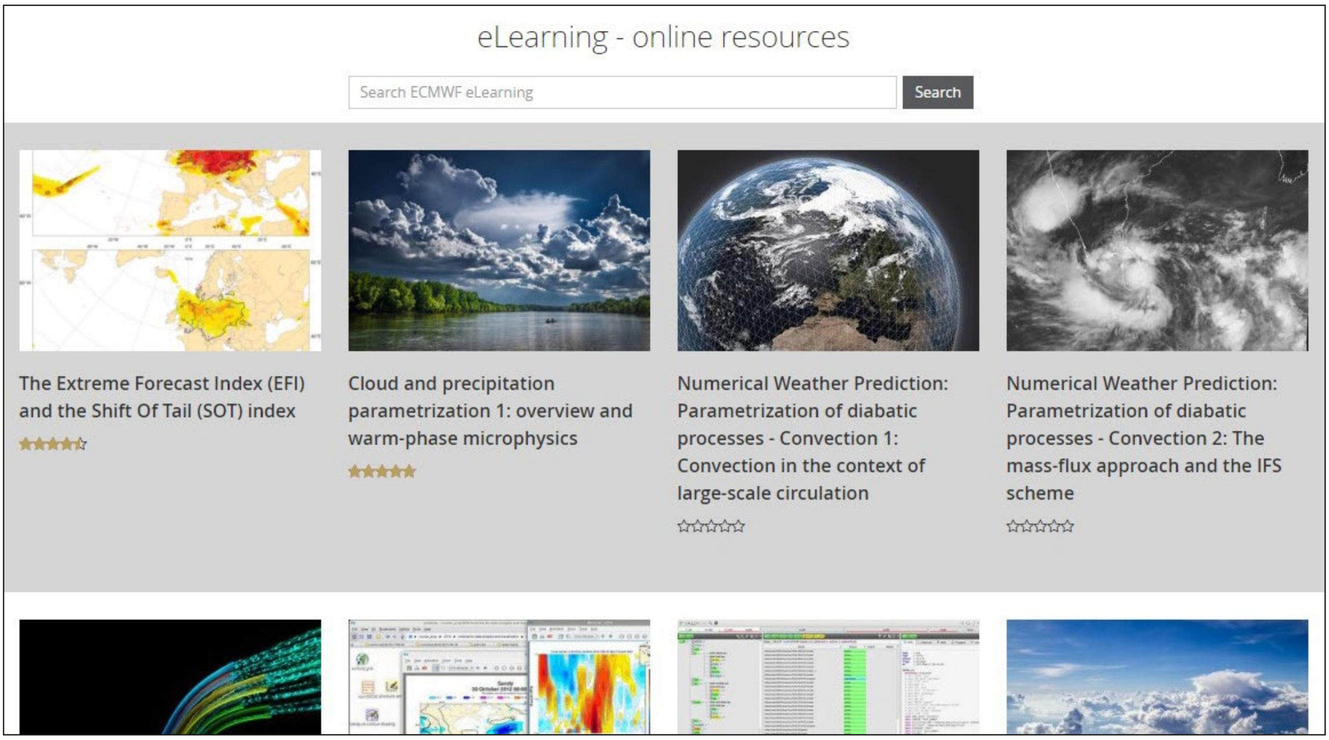


FIGURE 3 A user view of the ECMWF eLearning courses, showing all the available lessons with ratings and your progress, on *https://www.ecmwf.int/en/learning/education-material/elearning-online-resources*.

For those considering joining the ECMWF staff community, in 2021 we will be implementing a new recruitment portal. This will make our recruitment and selection process more efficient, as well as enhancing the onboarding experience for those who seek to join ECMWF so they can 'hit the ground running'.

Knowledge sharing

To ensure users have the knowledge needed to make the best use of ECMWF products and services, we have consolidated our learning content into our eLearning portal. This provides an excellent choice of online training covering a variety of topics (Figure 3). It is open to all, and to date nearly a thousand users have registered for one or more of the 31 lessons available. Users can take lessons in their own time and manage their own learning through the portal. We have also run two fully virtual courses using the platform.

Workshops and training events are now managed through a dedicated events management platform, integrated with our website for a seamless user experience. This has made it possible for users to self-register and for speakers to upload and manage their own materials (including last-minute presentation updates), empowering users and bringing efficiency gains to ECMWF as event hosts. Our use of Vimeo allows us to easily record events and share recordings to ensure that the knowledge imparted at those events is available any time to anyone, anywhere in the world.

In the future, and using the lessons learned during this pandemic, we expect to increase our provision of virtual workshops, seminars and training courses.

Support portal, chatbot, and documentation

As ECMWF becomes more open in its approach, we will see an increase in the number of users. This makes it imperative that we modernise our support service and ensure that users can help themselves to the answers contained in our documentation.

At the beginning of 2021 we launched our new online support portal to help direct users to the documentation they need. Should a user need to raise a query, the portal provides an interface for users to track their open queries, and it improves our efficiency in dealing with these queries (Figure 4).

As part of the ECMWF Summer of Weather Code 2020, we were able to prototype a chatbot or 'virtual assistant' for the ECMWF web. In 2021 we will be rolling out a pilot on the Copernicus Climate Data Store (CDS) with the intention of implementing it on the ECMWF web if it proves successful in helping users. The idea is to provide immediate answers to straightforward common questions without the user needing to leave the web page they are on. In the future we will take advantage of AI techniques and usage data to see if the chatbot can begin to answer more complex queries.

Underpinning effective self-service systems, such as the unified support portal and chatbot, is a major drive to review, rationalise, and standardise our user documentation. We are analysing the kinds of questions users ask and how our documentation is being used in order to eliminate the need for queries (by improving our web interfaces) or to improve the answers provided by our documentation.

Your relationship with us

To better understand and manage our relationships and interactions with our communities of users, and thus provide better services, we are unifying user identities across systems and between organisations and overhauling the Centre's Identity and Access Management (IAM) systems.

In January 2020 we took a first major step to rationalise our web login systems and implement Single Sign On (SSO). We will continue to roll this out across the whole of the ECMWF web so that users only need to log in or out once, even if they visit multiple areas of the ECMWF web. In the future we will be investigating the use of federated login; this would potentially allow users to log in using third party credentials.

Throughout 2020, in the background, we have been preparing to update our user management system. Later in 2021 this will be delivered into production, positioning us to make further improvements in usability and how we identify and authorise access for users of ECMWF services. Features we will have in the pipeline include multi-factor authentication and improved delegated admin and self-service user management. The aim is to improve security without decreasing usability.

We are looking at building an interface for users to be able to view a summary of their relationship with us: their own private profile web page. The intention is to securely display any personal details we have on file as well as provide a gateway to access favourite graphical products, the MARS and Web API request queue, open support tickets, favourite user documentation pages, mailing list memberships, satellite alerts, MARS activity, HPCF activity and usage, and so forth. Some of these already exist, such as the chart dashboard and the Web API queue, while some are currently in development.

Performance of our websites

Maximising our use of web technology is key to sharing our knowledge and providing access to ECMWF products and services worldwide. One of the ongoing challenges we face is keeping up with advances and disruptions in technology. Our goal is to always select the right technology to match requirements and provide our users with online interactions which are quick, easy, reliable, secure – and modern.

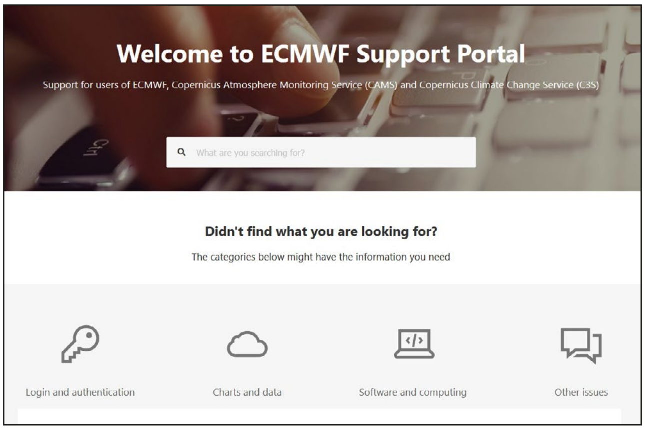


FIGURE 4 The new support portal, allowing users to easily search all ECMWF documentation, or to follow a guided path to the topic of interest and raise a request if needed (*https://confluence.ecmwf.int/site/support*).

Driven partially by the BOND programme (the migration of the HPCF to Bologna) and by the market, ECMWF is taking the opportunity to modernise and reshape how we deliver critical web-based services, ensuring that they provide sustainable good value for both ECMWF and our users. In some cases, this means migrating from traditional in-house systems to high-availability and well-supported cloud-hosted systems. In others, it means seeking partners with the necessary focus and expertise to help us deliver such services more effectively and efficiently.

Just as there is one ECMWF in the real world, there will always be one ECMWF web, even when it is delivered across multiple, ever-changing technologies and sites.

Call for feedback

ECMWF will continue to develop its use of web technologies to deliver its strategy and meet its users' behaviours and needs, focusing on those areas which deliver highest value and ensuring our web is high-performing: charts, data, documentation and online support. If you have any feedback on the ECMWF 'web of websites', we are always interested to hear from you. You can get in touch by going to our online help pages, which can be accessed from anywhere across the ECMWF web.

Further reading

Brady, A., 2019: ECMWF adopts new application platform, *ECMWF Newsletter* **No. 161**, 17–18.

Setchell, H., 2018: ECMWF improves web user experience, *ECMWF Newsletter* **No. 156**, 16.

IFShub: a new way to work with IFS experiments

Paul Burton, Manuel Martins, Stephan Siemen, Michael Sleigh

Roughly 500 experiments on ECMWF's Integrated Forecasting System (IFS) run at any given time, comprising the research work of users from the Centre and our Member and Co-operating States. A host of different bespoke tools are needed to enable researchers to manage this experiment workload. Different applications are available to prepare and submit experiments, to monitor the status of running work, to manage the resulting data archived in our meteorological archive (MARS) throughout its lifetime, and to scientifically analyse and evaluate the results. Most of these tools are desktop applications, some with graphical user interfaces, and all are developed independently of each other by different teams. Each is unique in its language, interface, design, dependencies, and architecture.

In particular, PrepIFS is the tool that researchers use to create, check, and store experiment configurations, and to submit those experiments to the high-performance computing facility (HPCF). PrepIFS has been in use for several decades and has stood the test of time, but in recent years it has developed some maintainability problems. The application is almost unique now at ECMWF in being written in Java, in which there is very little expertise at the Centre anymore. This is because nearly all development has moved to Python and C++ in recent years, while the IFS itself is still written mainly in Fortran. Additionally, there is limited expertise in creating user interfaces amongst the scientists developing the IFS, and this task is better performed by application developers who have considerable experience in this area. These kinds of maintainability problems lead naturally to usability problems, as non-essential improvements and bug fixes become too time-consuming to implement and development necessarily reduces to the bare minimum. PrepIFS also uses a bespoke language to define logical rules that govern valid configurations, making it difficult to implement new or improved rules.

There has therefore been an aspiration for some years to replace PrepIFS with a new tool which solves many of these problems, and which is actively developed, tested and maintained. From this starting point, however, it is only a small step to thinking about the whole end-to-end workflow – from starting to prepare an experiment to having it fully evaluated and archiving its results – and how this might also be improved. To meet this challenge, the vision of IFShub was developed. This article will set out how this was done and explain where we stand in developing the new system.

FIGURE 1 Impressions from the requirements gathering workshops held by Wavestone.

Aims of the project

Based on the shortcomings of existing systems, it was clear that a more coherent organisation of the overall model development workflow could help researchers in their day-to-day work. Especially for staff new to running the IFS, this would mean a faster learning process.

The IFS is increasingly run on different systems to support training and research activities, such as within the OpenIFS project and as part of a digital twin in the EU's Destination Earth initiative. A higher-level interface to the workflow will make it possible to hide some of the technical challenges of running experiments on different infrastructures while still keeping easy access to the outcomes. These developments should ease migrations to different systems and could be shared with other initiatives with similar challenges in the community.

Requirements gathering

It was clear from the beginning that this would be a big multi-year project and involve many stakeholders across ECMWF. To ensure the success of such a project, it is vital to agree on and communicate a vision and its scope. It was decided to use external help, and the company Wavestone was chosen. In the first three months of 2020, two consultants spent time with staff at ECMWF to interview them and organise small workshops to gather requirements and develop a vision.

doi: 10.21957/bu599oxq27

IFShub vision

It became clear from user feedback that there is a multitude of tools currently in use with IFS experiments, and that it can sometimes be difficult for users to easily access all the available tools and learn how to use each one of them. Additionally, some tools require a user to be logged on to a specific platform and may require them to copy data across machines.

We concluded that users would benefit from a single access point to all tools which relate to IFS experiments and the data produced by them. These tools should be accessible via a consistent user interface on a web browser, which can be accessed easily from anywhere, across a range of devices.

The technical complexity of a tool's implementation can be hidden behind the web interface. Users will no longer have to concern themselves about which system they need to log in to or where their data must be copied from and to. If hardware platforms or data locations change in the future, users should be little impacted by such changes, as the web interface should remain largely the same. The adoption of the changing underlying platforms will therefore be hidden to the user.

Our vision for IFShub is to provide a web-based interface, accessible from anywhere, allowing ECMWF and external users to control the full workflow of running an IFS experiment, through a consistent and intuitive user interface. IFShub should facilitate and encourage strong cooperation and interaction between users, by allowing them to easily share configurations, data, analysis and associated documentation whenever they wish to.

Wireframe mock-ups

To make this vision more concrete, the next step was to produce some mock-ups of how the user interface could look. Together with our consultants, we identified the key areas of functionality we wanted to reproduce, and new functionality we could introduce. We then considered how this could be presented to users within a web user interface.

This was purely to provide a physical manifestation of the ideas produced from the requirements capture process – it was not a working prototype of a system. Despite this, it was still a very useful process to help us ensure we had captured all the major areas of functionality required in a way that was intuitive for users to interact with. It provides a very useful reference point for feedback and will be an invaluable aid to our developers as they start to develop a working user interface.

We present here two wireframes which illustrate some of the core concepts and functionality we plan to provide with IFShub. Of course, we are still very early in the development of IFShub, so they are still just ideas. They are very useful for generating comments and feedback as we work towards the final implementation, but they

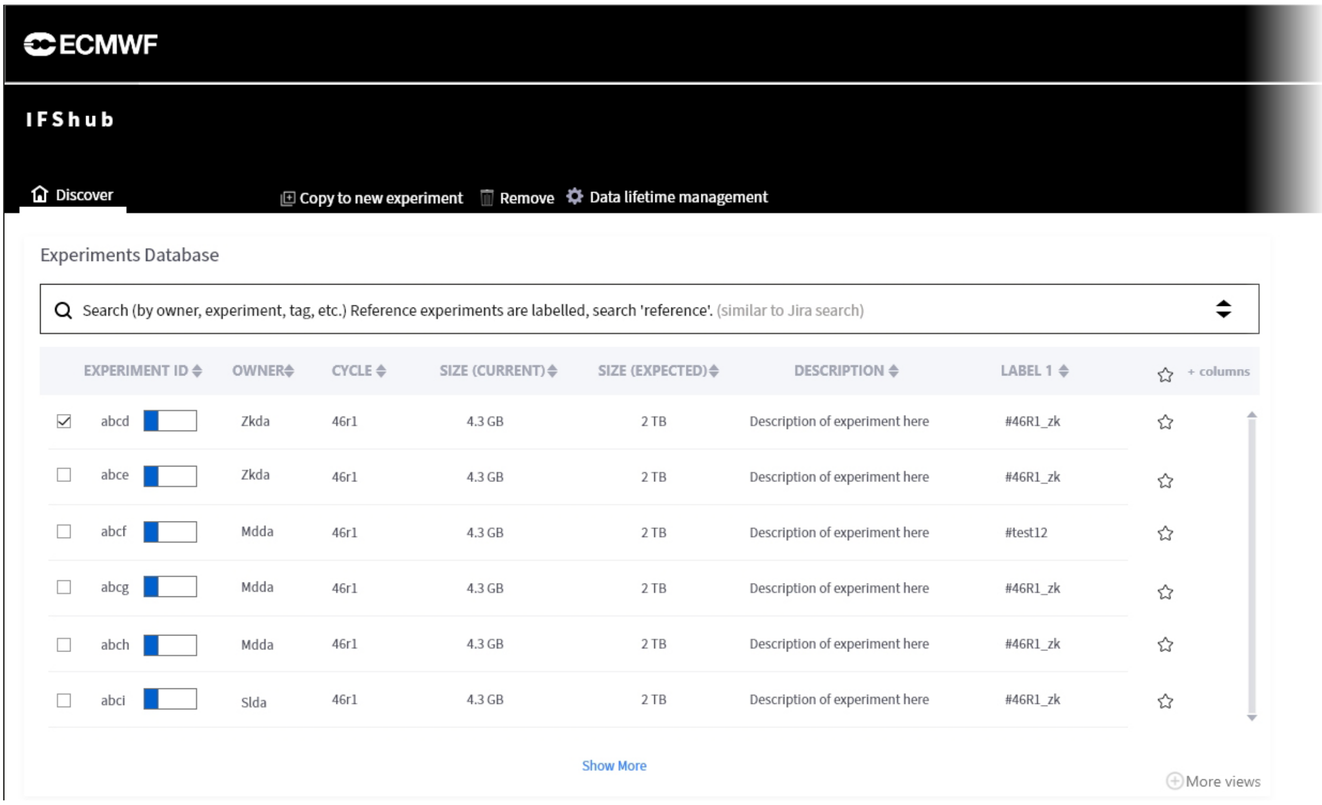


FIGURE 2 The experiment discovery page will give access to the full database of IFS experiments.

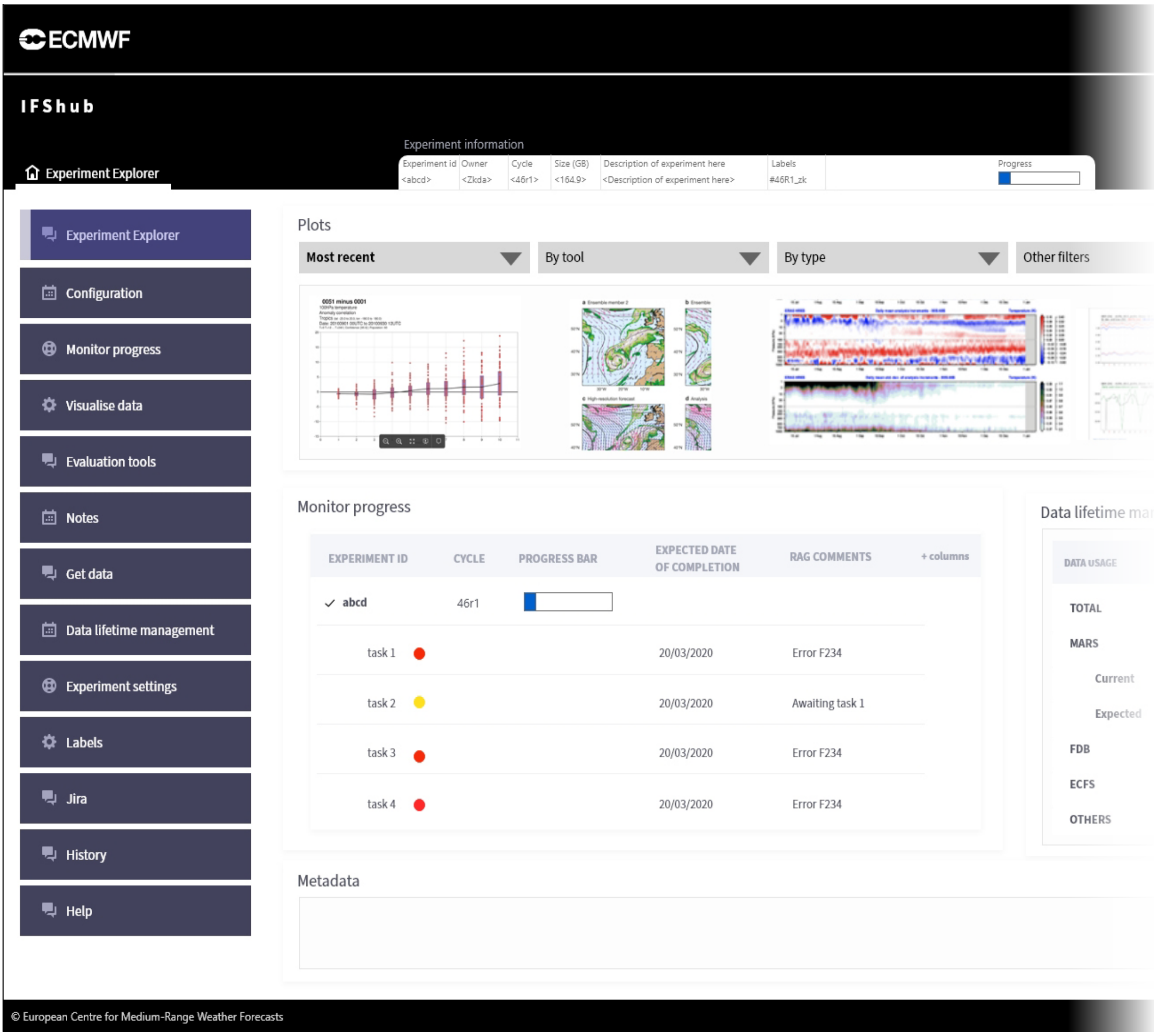


FIGURE 3 Overall summary of an IFS experiment that can be opened in a new tab. This 'experiment explorer' page will give access to all aspects of an experiment in the left-hand menu.

are not cast in stone and will no doubt be subject to revision as we develop and test the system and get feedback from our users.

Experiment discovery

The 'experiment discovery' page is central to IFShub: this is where a user can browse and manage the full database of IFS experiments (both their own and those of other users), using a range of search tools (Figure 2). We aim to provide as much useful information about experiments as possible on this page, to enable a user to quickly and easily identify the experiment of interest to them. Once an experiment has been identified, the most common action will be to browse the experiment in more detail in the 'experiment explorer', but other useful actions from this page will include the ability to copy the experiment's configuration to a new experiment, manage a user's data assets, and remove old experiments which are no longer required.

Additionally, this page will also include a summary of any useful information about ECMWF computer systems which may impact a user's experiments – for example any current problems or planned system sessions.

Experiment explorer

Once a user has identified an experiment of interest to them, they can open it in a new browser tab in the experiment explorer. From this page, the user can manage all aspects of the experiment. The main body of the page provides an overall summary of the experiment (Figure 3). This includes, for example, a progress monitor to quickly alert the user to any problems that need resolving; a summary of any plots produced from the experiment's data; and an overview of the data assets of the experiment.

On the left side of the page is a menu to access further pages which allow the user to interact with the experiment, the main items being:

- **Configuration –** will open WebPrepIFS, a replacement for the existing PrepIFS application, presented within a web browser. WebPrepIFS will allow a user to comprehensively configure their experiment, check all the settings are sensible and consistent, and then submit the experiment to run on the HPCF.
- **Monitor progress –** will allow users to monitor the ecFlow suite relating to their experiment. ecFlow enables users to run a large number of programs in a controlled environment and is used at ECMWF to run all our operational suites across a range of platforms. Users can use this page to quickly identify any problems which require further attention. They will be able to view log files from failed tasks and restart failed tasks. We do not propose to replace ecflow_ui with this tool but provide the most important functionality to allow users to monitor and manage running experiments.
- **Visualise data –** will provide a web interface to existing ECMWF tools, such as Metview and ecCharts, and provide the ability to view and manage the plots generated by these tools.
- **Evaluation tools –** provides a range of web-based tools (many of which will be new web interfaces to existing tools) used by researchers to evaluate and compare data from their experiments. Users are given the ability to manage the associated data and plots that these tools generate.
- **Notes –** will integrate with Confluence (the wiki software that hosts ECMWF's intranet) to enable comprehensive notes and documentation to be kept for each experiment, with automatically added links to access other relevant experiment-related information.
- **Get data –** allowing the user to easily browse the available datasets generated by their experiment and download selections to their current machine.
- **Data lifetime management –** a tool allowing the user to easily view how much space their experiment is using on MARS or ECMWF's File Storage (ECFS) system and how that data is being accessed. This enables redundant data to be easily identified and marked for removal.

Implementing the vision

As can be seen from the overall design of IFShub (Figure 4), most components already exist. The two major new developments are a loose web-based framework, which presents a unified interface to the different services, and a new application, WebPrepIFS, which allows users to define and manage their experiments. The latter will be developed in two steps.

The initial WebPrepIFS will be very similar in behaviour to the existing standalone PrepIFS application, so users should be able to easily migrate to this new

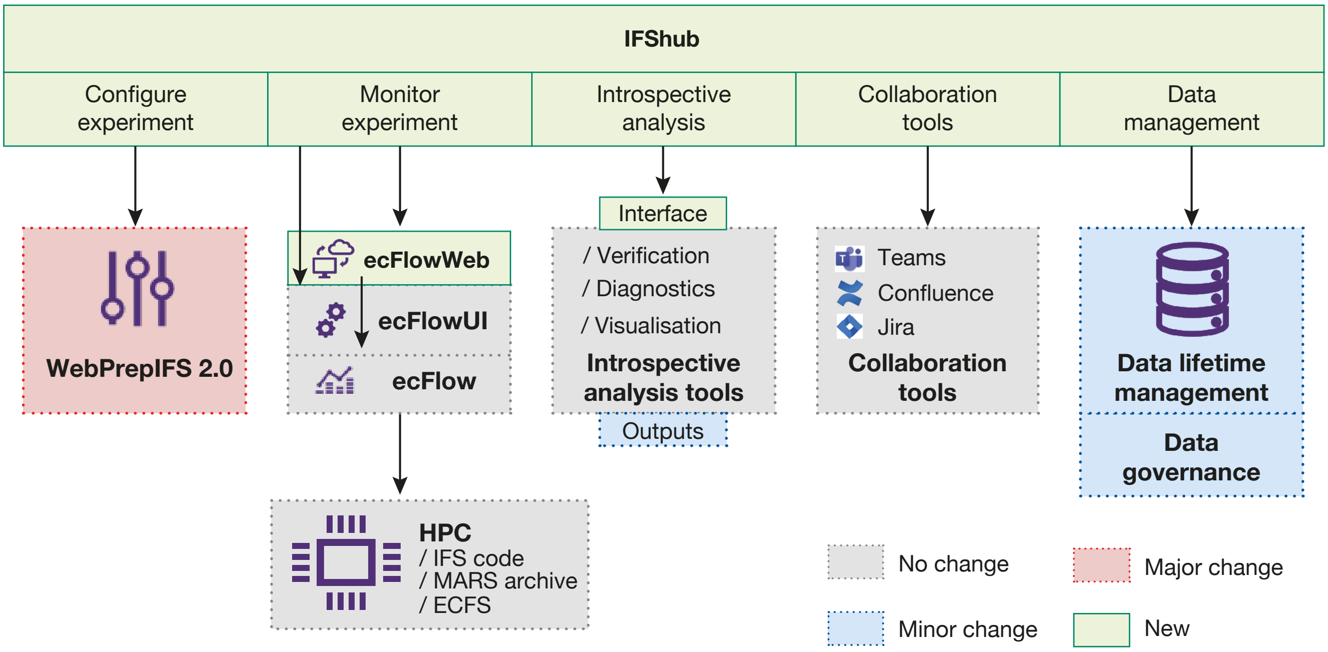


FIGURE 4 The different components which make up IFShub. Many parts already exist and might need only minor changes and integration.

web-based user interface as soon as it is available. We hope to be able to move all of a user's existing experiments into the new user interface. This will allow users to have the advantages of the powerful PrepIFS system in an easily accessible web application within the IFShub framework.

Following on from this, we plan to start work on the successor, WebPrepIFS 2.0. Plans for this are at a very early stage, but the intention is to improve the user interface, making it easier to use, and to improve flexibility and maintenance aspects. One of the most valuable objectives will be to allow the creation of combinations of existing experiments. This will allow production analysts to use the same tool to set up their configurations, and research scientists to replicate an operational-like configuration, which can then be easily shared between research and operations, thus significantly simplifying and streamlining the research-to-operations handover.

How will we get there?

With many key project participants busy with the migration of the computer centre to Bologna, it is unlikely that 2021 will see any major developments. Instead, it was decided to use the time to explore various aspects of the future system. For example, trials are underway on how information on experiments can be best stored and shared. As mentioned, the IFShub will link many existing web services. Work has started to explore how this can be achieved in a flexible way to allow services to evolve independently. In 2021, the management team of the IFShub project will also develop a long-term development road map for the next few years.

ECMWF publications

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

Technical Memoranda

882 **Janssen**, **P.A.E.M.** & **J.R. Bidlot**: On the consequences of nonlinearity and gravity-capillary waves on wind-wave interaction. *February 2021*

881 **Simmons**, **A.**, **H. Hersbach**, **J. Munoz-Sabater**, **J. Nicolas**, **F. Vamborg**, **P. Berrisford**, **P. de Rosnay**, **K. Willett** & **J. Woollen**: Low frequency variability and trends in surface air temperature and humidity from ERA5 and other datasets. *February 2021*

879 **Pucik**, **T.**, **P. Groenemeijer** & **I. Tsonevsky**: Vertical wind shear and convective storms. *January 2021*

876 **Zsótér**, **E.**, **C. Baugh**, **E. Hansford**, **K. O'Regan**, **C. Vitolo**, **F. Wetterhall** & **C. Prudhomme**: Floods, droughts, fire and beyond ... are existing forecasts enough? *January 2021*

ECMWF Calendar 2021

May 4–7	Training course: EUMETSAT/ECMWF NWP-SAF satellite data assimilation
May 10–14	Training course: Data assimilation
May 17–20	Joint ECMWF/OceanPredict workshop on advances in ocean data assimilation
May 17–21	Online computing training week
May 24–27	Training course: A hands-on introduction to numerical weather prediction models: understanding and experimenting
Jun 1–4	Using ECMWF's Forecasts (UEF2021)
Jun 3	Extraordinary session of Council
Jun 29–30	Council
Jun 29–Jul 1	Joint workshop on connecting global to local hydrological modelling and forecasting
Sep 13–18	Annual Seminar 2021
Sep 20–24	19th Workshop on high-performance computing in meteorology
Oct 4–6	Scientific Advisory Committee
Oct 4–7	Training course: Use and interpretation of ECMWF products
Oct 6	Advisory Committee of Co-operating States
Oct 7–8	Technical Advisory Committee
Oct 27–28	Finance Committee
Oct 28	Policy Advisory Committee
Nov 15–19	ESA-ECMWF workshop on machine learning for Earth system observation and prediction
Dec 2–3	Council

Contact information

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

Telephone National 0118 949 9000

Telephone International +44 118 949 9000

Fax +44 118 986 9450

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

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

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

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

ECMWF

Newsletter | **No. 167** | Spring 2021

European Centre for Medium-Range Weather Forecasts

www.ecmwf.int