TECHNICAL MEMORANDUM

840

Use and Verification of ECMWF products in Member and Co-operating States (2018)

Tim Hewson

Forecast Department

March 2019

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site www.ecmwf.int.

Contact: library@ecmwf.int

European Centre for Medium Range Weather Forecasts
Shinfield Park, Reading, Berkshire RG2 9AX, England

Abstract

Each summer ECMWF invites Member and Co-operating States (i.e. National Meteorological Services (NMS)) to submit updated reports on the application and verification of ECMWF's forecast products. This report summarises the findings from the 2018 reports, in conjunction with additional feedback gathered in other fora.

Direct use of ECMWF products remains extensive within the Member and Co-Operating States, spanning time ranges from day-1 to seasonal. The peak in direct usage is for lead times of about three to five days, when initial warnings for potential high impact events are commonly considered. At short ranges, up to about day-3, ECMWF model data commonly also provides boundary conditions for deterministic and ensemble limited area model (LAM) runs. Consortia are increasingly being developed across Europe, to perform LAM runs centrally for groups of countries. For example, the Nordic MetCoOp project, which runs the AROME 2.5 km model in ensemble mode, is expected to have expanded from the current three partner countries to eight countries by 2022.

In the last couple of years, the average resolution difference between operational LAMs and everyday output from ECMWF's Integrated Forecast System (IFS) has increased again; 2 km LAM runs are now typical. Probably because of this difference, new verification results show slightly higher skill levels, overall, for LAMs compared to HRES, for surface weather parameters. However, this statement hides what is a very complex picture, with relative performance depending strongly on the LAM itself, on region, on season, on time of day, on weather type and on weather parameter. Two key issues with ECMWF model output continue to be wind strengths over mountains and screen temperatures in very stable conditions. These are being actively worked on at ECMWF. Many NMSs continue to use statistical post-processing to successfully address these and other bias issues.

Most NMSs indicate that they are extremely satisfied with ECMWF operational output. In conjunction, expectations have grown, to the extent that in some fields they can exceed what is physically reasonable. Our inability to pinpoint windstorms, to predict very low visibility in fog, to forecast waves near complex coastlines and to deliver accurate seasonal forecasts have all been noted. ECMWF probably needs to better manage expectations in very challenging areas such as these.

Recent ECMWF product initiatives have been widely praised; e.g. vertical profiles, lightning diagnostics and, most notably, precipitation type meteograms. The ecCharts tool is very widely used, although speed continues to be an issue for some.

ECMWF will continue to deliver efficiency savings for the Member and Co-operating States, by serving as many user requests as possible in the coming months and years, subject to the decisions of ECMWF governing bodies, and to available computational and human resources. At the same time issues with model output will be investigated and addressed as far as possible. ECMWF appreciates and strongly encourage the continued feedback of issues identified by forecasters and others. This mechanism continues to be very effective in initiating an investigative process that eventually improves forecast quality.

1 Introduction

Each summer ECMWF invites Member and Co-operating States to submit updated reports on the application and verification of ECMWF's forecast products. The NMSs (national meteorological services) submitted their reports (22 out of 34), which are available on the ECMWF website. Reports have been provided by Belgium, Croatia, Czech Republic, Denmark, Finland, the Former Yugoslav Republic of Macedonia, France, Greece, Hungary, Iceland, Ireland, Israel, Italy, Latvia, Luxembourg, Montenegro, Norway, Romania, Serbia, Sweden, Switzerland, United Kingdom.

A summary of the reports is presented below. Content has been combined with (i) feedback from the "Using ECMWF Forecasts" (UEF2018) workshop held at ECMWF from 5–8 June 2018 (primarily from breakout groups), and (ii) feedback from official triennial Member State/Co-operating State visits undertaken by ECMWF between July 2017 and June 2018. In chronological order visits were made to: Norway, The Netherlands, Belgium, Luxembourg, Sweden, Romania, Bulgaria, Croatia, Spain, Slovenia, France and Morocco. Please note that this report generally only covers NMS activities and results recorded in the above fora.

For the NMS reports contributions had been invited under the following headings:

a) Summary of major highlights

b) Use and application of products

c) Verification of products (objective and subjective)

d) Feedback on ECMWF "forecast user" initiatives

e) References to relevant publications

For the verification section, this year ECMWF particularly encouraged submission of results pertaining to the following: surface parameter systematic errors; visibility, humidity and clouds; conditional verification; Limited Area Ensemble Prediction Systems (LAM-EPS). This initiative met with some success, although we received only three small contributions relating to LAM-EPS systems.

Note also that the ECMWF IFS is upgraded each year, which naturally affects aspects of performance in-year, so summary information presented here should be read with this in mind. During the past 12 months, a long range (seasonal) forecasting system upgrade (to System 5) was introduced in November 2017, and a new IFS cycle 45r1 for medium and extended range (monthly) forecasts was introduced in early June 2018.

Note that the results of ECMWF's own recent objective verification has been documented separately, in Technical Memorandum 831 (October 2018).

2. Use and application of products

Strategies for using ECMWF model output for operational purposes depend largely on the lead time of the forecasts. Although visits to Member and Co-operating States and NMS reports do not encompass every forecasting activity, there was evidence that at least 22 out of 27 states used IFS data directly in some way to prepare short range forecasts (up to, say, 48-72h ahead), all 27 used IFS data for medium range forecasting, and respectively at least 13 and 16 out of 27 were using the monthly and seasonal forecasts for operational purposes.

In the shorter ranges (typically to 48–72h ahead), ECMWF IFS products are commonly used in conjunction with products from other sources, notably deterministic Limited-Area Model (LAM) systems, but to an increasing extent LAM-EPS too. In the vast majority of cases reported, ECMWF IFS data provides boundary conditions (BCs) for these limited-area runs, commonly four times per day but sometimes more often. LAM-EPS systems vary more in their configuration however. Some use the ECMWF ENS BCs, others use more than one EPS system for BCs, whilst one or two even use a "Two-way-Multi-Model approach". For example, the AEMET-SREFS system utilises four different types of LAM, and five different types of global model - see Figure 1. Conceptually this approach has many attractive aspects, though whether these outweigh the benefits of driving a single model configuration from a single ENS, which by design should have a more even representation of uncertainties, is an interesting question.

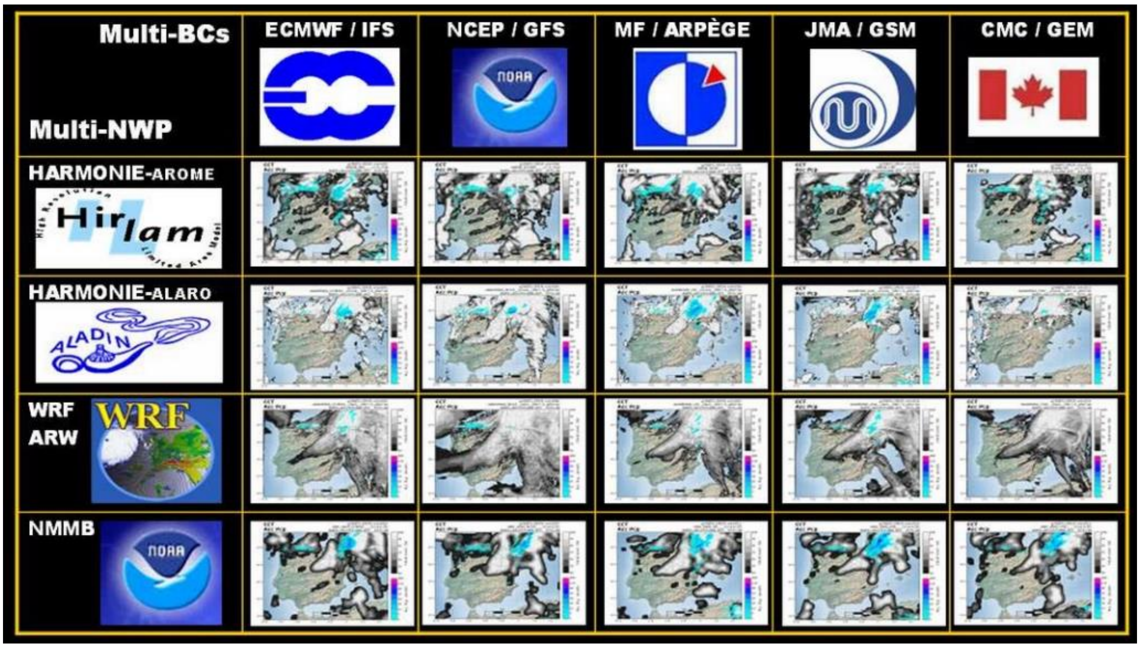


Figure 1: Cloud and precipitation forecasts from the AEMET two-way-multi-model LAM-EPS approach. Columns show global models providing BCs; rows show LAMs run from those BCs.

In the medium, extended and longer ranges, ECMWF products continue to be the main or only output used by NMSs. All NMSs seem to provide forecasts of some sort up to lead times of 10 days. Growth in uptake of longer lead forecasts is commonly being driven by requests from a range of different customers. However, there are suggestions that some customers are frustrated by the lack of skill at longer lead times, particularly for seasonal.

Operational short-range limited area modelling efforts within the Member and Co-operating states seem to have converged somewhat in recent years, such that most now run at about 2.5 km horizontal resolution (whilst in Belgium, France, Greece, Switzerland and UK they go down to about 1 km). Lower resolution runs, close to that of HRES (9 km) have been or are being phased out, partly in recognition, it seems, of the superior skill of HRES at those resolutions for most near-surface forecast parameters.

Whilst the main operational focus for all Member and Co-operating states is on local weather, many also have international commitments, for land and sea areas across the world, for which ECMWF forecasts are very regularly used. For example, The Netherlands provides forecasts for its territories in the Caribbean, France has similar commitments in many different tropical regions, Sweden provides ship-route forecasts globally and predictions for armed forces based overseas, whilst the United Kingdom has a single forecaster rostered 24/7 to handle only forecasting tasks for regions outside the UK.

2.1 Local post-processing of model output

2.1.1 Statistical adaptations

Most countries apply some local statistical ‘recalibration’ procedures to post-process ECMWF forecasts, especially to make forecasts of sensible weather for specific locations. HRES has traditionally provided the main input, but statistical calibration is being applied to the ENS distributions in some countries. Activity in these fields has continued in spite of the fact that raw model output becomes ever more accurate, and, as resolution increases, ever more representative of point locations. Research studies suggest that one can continue to make noteworthy improvements via statistical post-processing even as forecast quality increases. France and the UK are continuing to improve their multi-faceted multi-model post-processing techniques, many other countries report that previously developed systems continue to deliver benefits and continue to run, and no country has reported the full decommissioning of post-processing activity. Within the 22 NMS reports only Denmark, Hungary, Iceland, Ireland and Luxembourg expressly report that they do not use statistical post-processing (and Ireland plans to start).

We do not have enough references this year to describe the full range of statistical techniques applied, although as many are apparently unchanged, reports from previous years probably remain valid. So the main post-processing methods are probably still Model Output Statistics (MOS) and Kalman filtering. Similarly, the following approaches are probably also still used, to varying degrees: height mismatch adjustment, bias removal, multiple linear regression, polynomial regression, logistic regression, linear discriminant analysis, quantile matching and non-homogeneous Gaussian regression. Many countries employ a mixture of approaches; this is most evident in the report from France. Many of the techniques are applied to HRES, some to ENS members one by one, some to ENS distributions. Meanwhile Norway has developed a ‘smart calibration’ methodology for coastal stations, which uses land-sea mask information and observations to establish the best weights to apply to nearby model gridpoints. This approach achieves very positive results for 2 m temperature forecasts. Norway also have a local calibration mechanism for making ENS precipitation forecasts reliable for Norwegian sites, which they say increases spread and is effective.

Most calibration techniques seem to be based on real-time forecast performance, as measured using standard synoptic stations. There is also continued use of analogues by Serbia and the Czech Republic for monthly forecasts.

Statistical combinations of forecasts from ECMWF and other models are quite common, employing weighting according to lead time. The other models are usually limited-area models with higher horizontal resolution than HRES (usually in the range 1–7 km), but in some cases output from other global deterministic or ensemble systems is incorporated - e.g. in the U.K. The UK also stresses, as last year, and with more new evidence, that tropical cyclone forecasting benefits from the use of multi-model ensembles even though some of the constituent ENS systems have less intrinsic skill and less reliability than do others (see Figure 2).

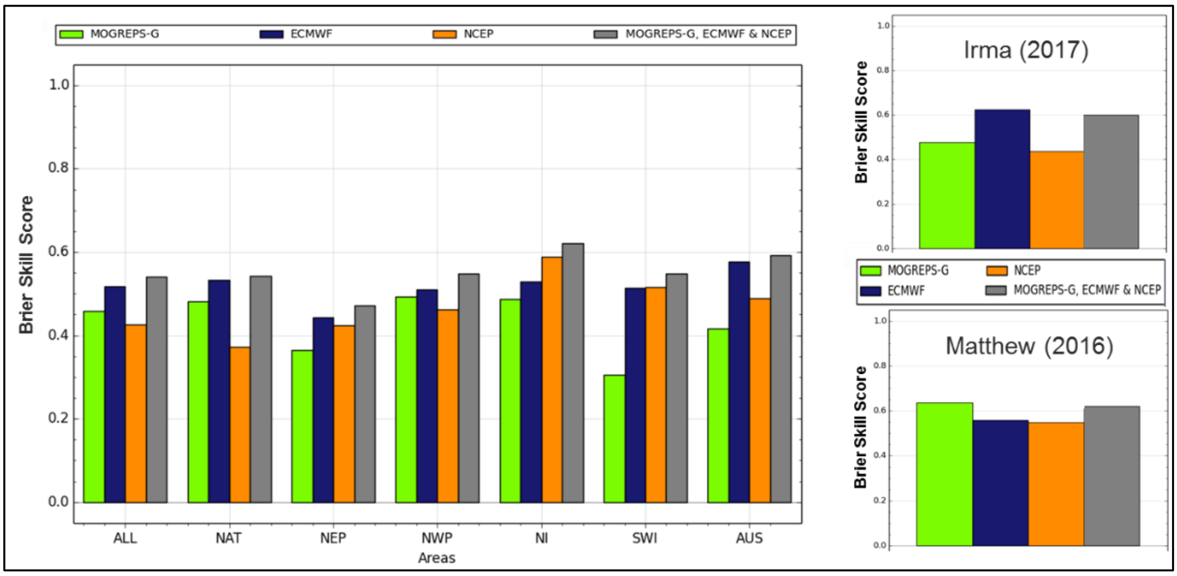


Figure 2: Brier Skill Score of MOGREPS-G, ENS, NCEP GEFS and multi-model EPS forecasts of named storm strike probability, for "a storm passing within 120km within the next seven days": all storms from January to December 2017 and split by tropical cyclone basin (left) and verification for Hurricanes Irma and Matthew (right).

ECMWF reforecasts continue to be a convenient and important dataset for calibration activity, and we continue to see requests at UEF meetings to provide a full year's worth with each new cycle, particularly from commercial customers. It is worth re-iterating that such an initiative would require substantial computer resources that are not available. Evidence of direct NMS use of reforecasts was found for Belgium, Sweden, Norway, Hungary and Romania. Probably other countries use them too. And of course, they are a fundamental component within ECMWF's strategy, for the monthly forecast, and for very-widely-used ECMWF products such as the EFI and the 15-day Meteogram-with-climate.

2.1.2 Physical adaptation

ECMWF output, in one form or another, is used very widely to provide boundary conditions (BCs) for running limited area models, and in some instances initial conditions (ICs) for those models too. Mostly this happens via the Optional ECMWF Programme "Boundary Conditions for Limited Area Modelling", which provides additional HRES and ENS forecasts from 06 and 18Z data times, now at hourly intervals. The limited-area model suites most widely used in this way are ALADIN, AROME, HARMONIE, HIRLAM, COSMO and WRF.

There is clear evidence of increasing levels of co-operation and merging of LAM formulations across Europe, and remote running has also begun (as will be the case for ECMWF once its supercomputer resides in Bologna). For example, a combined HARMONIE-AROME formulation is used jointly by Denmark and Iceland, in deterministic and ensemble modes, and run on a supercomputer in Denmark. There are reports that the HIRLAM and ALADIN consortia have begun to merge, to become one by 2021. Meanwhile Sweden, Norway and Finland are partners in the MetCoOp project, which runs a shared 10-member 2.5 km AROME-based ensemble system for Northwest Europe; currently BC input is just from HRES (in lagged mode), but there are plans to replace this by ENS forcing in future. Also, it is expected that by 2022 five more Nordic countries will have joined this consortium. Based on these overall trends, it may be that contributions to this document in future years become a little less country-specific.

Generally, it is the HRES BCs that are used most widely, for deterministic LAM runs. For LAM-EPS forecasts ENS BCs are often used; such an approach was reported this year by Denmark, Hungary, Switzerland, Italy and Belgium.

The report from France lists as a highlight for this year the "operational implementation" of the use of HRES BCs to drive one of their LAMS (AROME) this year for a Western Europe domain, to complement the runs of that LAM that use Arpege BCs. They report also that the way in which HRES BCs are used for LAMs in other regions of the world has changed, leading to better precipitation forecasts in the first six hours; a similar comment highlighting the advantages of "BC-blending" is made by Sweden.

ECMWF's model suite also provides ICs and BCs for wave and surge modelling in several countries. France, Norway, Greece, Israel, Bulgaria, and Finland run wave models using HRES data, for the world, parts of the world, the Eastern Mediterranean, the Red Sea, the Black Sea and the Baltic respectively. Surge modelling efforts were reported by France and Norway; France only uses HRES, whilst Norway's system, which contributes to Copernicus Marine Services, has, since winter 2017/18, been based on ENS. Sweden uses ECMWF fields as upper boundary conditions for its 10-day NEMO ocean forecasts, whilst Denmark runs ocean / sea ice models over various domains that use HRES fields for upper forcing.

Even NMSs that have the capacity to provide BCs from their own global models still use ECMWF output of certain types for BCs and ICs. For example, IFS and CAMS data are being used in collaborative fashion in France in the evolving MOCAGE suites, for modelling pollutant transport and dispersion in various ways. Recent developments include extending lead times and increasing native resolution; these are facilitated by having, for example, 9 km HRES forecasts to day-6 available as input. In France ECMWF runs also provide back-up BCs for LAMs usually driven by other models.

HRES and ENS fields continue to also be used or adapted in various countries to drive trajectory and dispersion models (Finland, France, Switzerland, Denmark, Iceland, Italy, Norway, and Greece for sea pollution), and local hydrological models (e.g. using HRES: Bulgaria, Czech Republic, Croatia, Finland, Serbia, Slovenia; using ENS: Belgium, Latvia). Some use EFAS flood forecasts in addition. In several countries, there are application models of other types (e.g. road state models) that are or can be driven by ECMWF data.

2.1.3 Derived fields

Many countries perform additional post-processing to provide derived products that historically ECMWF has not provided, and tailored versions of those products that ECMWF has been providing.

Derived products are generated locally in NMSs for several reasons, for use by NMS forecasters, for specific societal or economic applications, and for use by the public. Many countries have reported a continuation of products they cited in previous years' reports.

Convection-related indices are generated in many Member and Co-operating States for forecasters. This evidently reflects the lack of a discrete representation of the hazard at global model resolution, the severe impacts that can arise, and perhaps also a general shift towards impact-based forecasting across the NMSs (e.g. Iceland changed to an impact-based system late in 2017). More specifically, at longer ranges forecasters need to identify when the environment will be conducive to the development of convective hazards, and then, at shorter ranges, to forecast the hazards themselves, ideally giving guidelines on timing and location. The indices generated locally include Lifted, SWEAT, and wind shear and helicity using various level combinations.

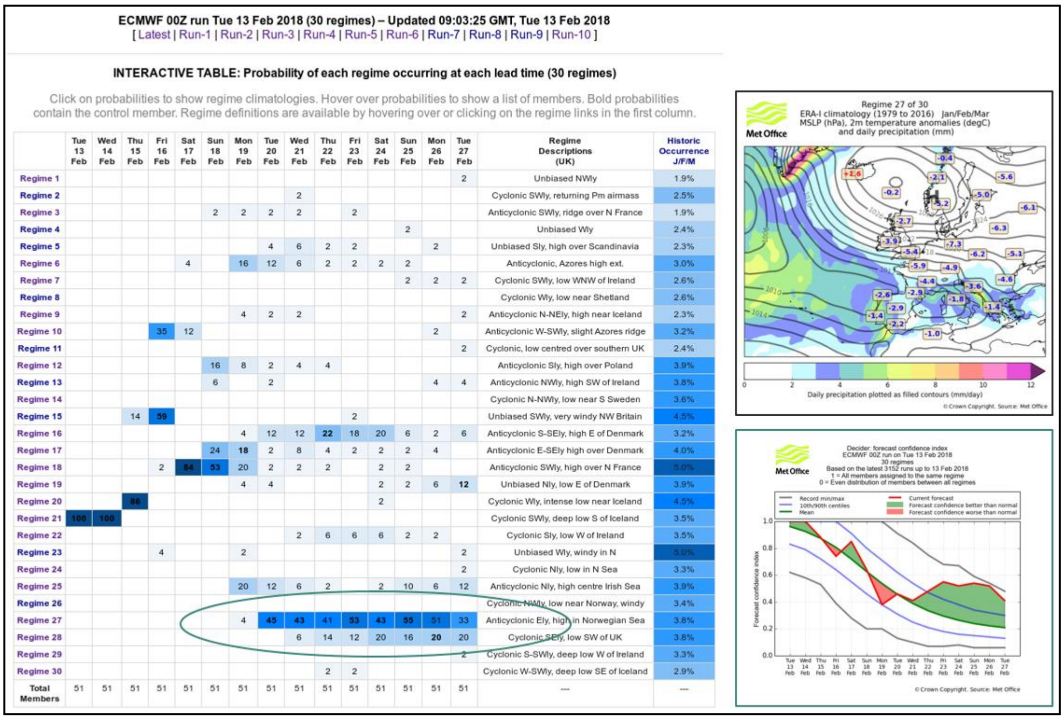
Sweden continues to utilise an interesting approach to convert HRES output (of cloud cover and precipitation) into an ensemble-like product. For each HRES grid point they assume that all values within a 20km radius (approximately 25 HRES points in total) are (equally) relevant, and work out median, 10th and 90th percentiles for that gridpoint accordingly. Where there is no strong topographic or coast-related forcing of rainfall this has the potential to deliver useful results for the forecaster.

Several countries compute their own clustering of the ENS, and generate for example ensemble mean, spread, and probabilities for a variety of parameters and event thresholds. The UK is very active in this general area, and refers extensively in its report to a UK-centric regime-based clustering approach, called "Decider", which is used with output from other global ensemble systems as well as ECMWF's. An example using just ECMWF output is in Figure 3. Here the "spread" in week two, ahead of an exceptional cold outbreak in late February 2018, is unusually low for such lead times; "spread" here refers to how spread out, amongst the different possible regimes, the ENS member solutions actually are. This is a particularly interesting plot because the relatively low spread implied seems to be at odds with spread defined by ECMWF's usual metrics (such as standard deviation of 850hPa temperature), which around this time was exceptionally high. Probably the easterly flow is associated, itself, with a propensity to deliver (at least sometimes) airmasses with many different intrinsic (850hPa) temperature levels, due to an absence of surface-heating-related airmass homogenization by long passage over a relatively warm ocean. It reminds us that ECMWF must remain cognisant of key user needs, and tailor its output accordingly.

Other miscellaneous fields computed for forecasters (many of which ECMWF has been asked to add to its own variables list) include the following: vorticity advection, thermal advection, freezing level, wind chill, river ice formation and breakup and weather phenomena via decision tree.

Aviation is one key societal activity for which countries compute specific derived products (e.g. turbulence-related, icing-related). Drought indices are also used / being developed (Bulgaria and

Romania). Probability of road icing, rime and a ventilation index also appear to be customer-specific derived products.



ECMWF 00Z run Tue 13 Feb 2018 (30 regimes) – Updated 09:03:25 GMT, Tue 13 Feb 2018
[Latest | Run-1 | Run-2 | Run-3 | Run-4 | Run-5 | Run-6 | Run-7 | Run-8 | Run-9 | Run-10]

INTERACTIVE TABLE: Probability of each regime occurring at each lead time (30 regimes)

Click on probabilities to show regime climatologies. Hover over probabilities to show a list of members. Bold probabilities contain the control member. Regime definitions are available by hovering over or clicking on the regime links in the first column.

	Tue 13 Feb	Wed 14 Feb	Thu 15 Feb	Fri 16 Feb	Sat 17 Feb	Sun 18 Feb	Mon 19 Feb	Tue 20 Feb	Wed 21 Feb	Thu 22 Feb	Fri 23 Feb	Sat 24 Feb	Sun 25 Feb	Mon 26 Feb	Tue 27 Feb	Regime Descriptions (UK)	Historic Occurrence J/F/M
Regime 1															2	Unbiased NWly	1.9%
Regime 2									2							Cyclonic SWly, returning Pm airmass	2.5%
Regime 3						2	2	2	2		2					Anticyclonic SWly, ridge over N France	1.9%
Regime 4													2			Unbiased Wly	2.4%
Regime 5								4	6	2	2			2		Unbiased Sly, high over Scandinavia	2.3%
Regime 6					4		16	12	6	2	2	2	2			Anticyclonic, Azores high ext.	3.0%
Regime 7													2	2	2	Cyclonic SWly, low WNW of Ireland	2.6%
Regime 8																Cyclonic Wly, low near Shetland	2.6%
Regime 9							4	2	2						2	Anticyclonic N-NEly, high near Iceland	2.3%
Regime 10				35	12									2		Anticyclonic W-SWly, slight Azores ridge	3.2%
Regime 11															2	Cyclonic, low centred over southern UK	2.4%
Regime 12						16	8	2	4	4						Anticyclonic Sly, high over Poland	3.9%
Regime 13						6		2						4	4	Anticyclonic NWly, high SW of Ireland	3.8%
Regime 14																Cyclonic N-NWly, low near S Sweden	3.6%
Regime 15			14	59							2					Unbiased SWly, very windy NW Britain	4.5%
Regime 16							4	12	12	**22**	18	20	6	2	6	Anticyclonic S-SEly, high E of Denmark	3.2%
Regime 17						24	**18**	2	8	4	2	2	2	4		Anticyclonic E-SEly high over Denmark	4.0%
Regime 18				2	84	53	20	2	2	2		2	2			Anticyclonic SWly, high over N France	5.0%
Regime 19							4	4				2	2	6	**12**	Unbiased Nly, low E of Denmark	3.9%
Regime 20			86									2				Cyclonic Wly, intense low near Iceland	4.5%
Regime 21	100	100														Cyclonic SWly, deep low S of Iceland	3.5%
Regime 22									2	6	6	6	2	2		Cyclonic Sly, low W of Ireland	3.5%
Regime 23				4			2								2	Unbiased Wly, windy in N	5.0%
Regime 24															2	Cyclonic Nly, low in N Sea	3.3%
Regime 25							20	12	6	2		2	10	6	12	Anticyclonic Nly, high centre Irish Sea	3.9%
Regime 26																Cyclonic NWly, low near Norway, windy	3.4%
Regime 27							4	45	43	41	53	43	55	51	33	Anticyclonic Ely, high in Norwegian Sea	3.8%
Regime 28									6	14	12	20	16	**20**	20	Cyclonic SEly, low SW of UK	3.8%
Regime 29															2	Cyclonic S-SWly, deep low W of Ireland	3.3%
Regime 30										2	2					Cyclonic W-SWly, deep low SE of Iceland	2.9%
Total Members	51	51	51	51	51	51	51	51	51	51	51	51	51	51	51	---	---

Figure 3: Probabilistic weather pattern forecast from the 00UTC IFS runs on 13th February 2018, showing a high confidence forecast transition to an anticyclonic easterly weather pattern in week 2 (left), with colder-than-average temperatures (top right). Forecast confidence (=low pattern spread) was unusually high for week 2 (bottom right).

For the public, a typical derived product is “characteristic weather” symbols that an NMS will generate for specific times or periods, for the web or for mobile phone “apps”. France reports again on recently introduced “global pictograms” which based on their example seem to depict the most likely weather type in map format (such as “heavy snow”, “showers with local storms”, “dense fog”). Norway have an interesting way of using ENS in this framework, to deliver a colour-code confidence level for each deterministic forecast component (green, yellow, red) on the yr.no site, although this is only available for Norwegian locations.

There is not much evidence of derived fields related to waves and the ocean, although Italy does generate a sea-state code on the “Douglas scale”, and a risk of icing.

For non-bespoke fields ECMWF continues to provide efficiency-savings to its users, by adding more derived fields to its range of web products and to ecCharts, and by improving the online structure, access tools and usability of those platforms. ECMWF also fulfils specific requests for new products whenever possible. For example, whilst Switzerland reports on the use of (locally computed) vertically-integrated water vapour fluxes for real-time flood prediction around the Alps, related parameters can now be provided directly to users since their introduction in cycle 45r1 in June. In 45r1 ECMWF also provided new lightning parameters and temporally more consistent representations (6-hour maxima) of CAPE and CAPE-shear. These should in time facilitate better prediction of convective hazards, and reduce computational effort in the NMSs, in a key area where, as highlighted above, much work has historically been devoted to creating new diagnostics. Meanwhile Sweden reports that it may phase out its old visibility diagnostic following positive verification results obtained for the ECMWF visibility parameter, introduced into the IFS in May 2015.

2.2 Direct use of ECMWF products, including severe weather prediction

ECMWF products are primarily used from day-3 onwards. At shorter ranges, most services refer to ECMWF products alongside output of their main LAM models (which themselves, as discussed above, usually also incorporate ECMWF input via BCs). For extended (“monthly”) and seasonal ranges use of and demand for accurate ECMWF forecast products remains high. This all means that ECMWF forecasts are vital for a vast range of operational functions in most of the Member and Co-operating states.

Most countries ingest a range of ECMWF products (especially from HRES) into their forecasters’ workstations, where they can be shown alongside other products used by the forecasters (LAM output, observations, satellite data etc.). Certain workstation applications allow some editing by forecasters of the model fields, who may for example apply some subjective weighting to HRES and ENS components.

The main alternative ways to view ECMWF products are as static products on a website internal to the NMS, as static clickable images on the ECMWF website, or within ECMWF’s complementary web-based visualisation tools ecCharts, and, related to this, the Dashboard. All three ECMWF web-based tools continue to undergo upgrades and re-design work to improve usability. Users are also reminded that ecCharts provides a WMS service to facilitate the transfer of ECMWF data into local workstations. Some NMSs also report that they use the ECMWF graphics tool Metview to create plots for forecasters (Croatia, Italy, Montenegro, Romania, Serbia).

In the medium-range ECMWF forecasts are used for general forecasting, but also for more specific forecasts for aviation and marine services. Although the main focus is for Europe, several countries routinely forecast for different areas of the world, for overseas territories and for military, humanitarian or capacity building activities, or for general public forecasts (e.g. yr.no).

Some NMSs report that the introduction of new forecast parameters such as lightning density and integrated vapour transport, in June in cycle 45r1, were very welcome.

ENS and HRES

Whilst improving ENS forecasts lies at the core of ECMWF’s investment program and 10 year strategy, most NMSs use ENS and HRES to varying degrees, and have commented this year on *how* ENS and HRES are used operationally, essentially by forecasters. It is therefore helpful to describe this feedback aspect. The summary below refers mainly to the early medium range (days 3–7):

- “Forecasts are mainly based on HRES” (Iceland, Montenegro, Former Yugoslav Republic of Macedonia)

- “ENS output is used to assess the credibility of HRES” (Czech Republic)
- “ENS is used to assess the possibility of alternatives, and its use is increasing” (Israel)
- “ENS is used, but use is restricted in summer by its later time of arrival” (Italy)
- “Use of ENS is constantly increasing” (Croatia)
- “Use of ENS is widespread” (Spain)
- “Use evolves with lead time from being HRES-based to ENS-based” (Belgium)
- “Both HRES and ENS are widely used” (Denmark, Finland, France, Norway, Romania, Slovenia, The Netherlands, UK)
- “A probabilistic (ENS) approach is very important for decision making and is consistent and effective” (Luxembourg)

So approaches vary, and all have some merits, but for the vast majority the ENS is critical. The comment from Luxembourg seems to be particularly appropriate.

Severe weather

IFS forecasts are very widely used for official severe weather warnings and for alerting forecasters to potential severe events. The longest lead time at which a warning will be issued varies by country; for Italy it is only about 24 hours, for many other countries it is now five days and for the UK it is seven days (increased from five in 2017). Because of these variations the extent to which ECMWF output is actively used also varies, with the output of Nowcasting tools and LAMs given a lot of weight at short leads.

Warning systems are becoming more probabilistic in nature, except perhaps at short leads, and overall ENS usage is increasing. The widespread use of the EFI and SOT for alerting forecasters to severe weather events (rain, wind, heat, cold, snow) is very well established - numerous references can be found in the NMS reports - and the more recently introduced EFIs for severe convective events have been found to be useful additions by Member and Co-operating States. Probably the forecaster training that ECMWF provides has helped considerably in the uptake of these various products. The trend to go from threshold-based warnings to ones that are more impact-based continues. This focus on impacts relates closely to the return period philosophy that underpins the ECMWF EFI and SOT products. The UK again highlights how the EFI can be especially useful beyond the local region, when forecasters’ knowledge of local climate is lacking.

Other ENS products, such as CDFs (Cumulative Distribution Functions) and probabilities, are widely used in conjunction with the EFI to provide additional uncertainty estimates. These can be particularly valuable in countries where warnings continue to be threshold-based, and there are reports (e.g. Denmark) that the ENS is tailored accordingly, via ecCharts or other mechanisms. Denmark, Finland, Montenegro, Norway and the UK noted how the extra-tropical cyclone tracking products assist in their warning process, by highlighting uncertainties in the behaviour of cyclones that are “responsible” for severe weather. Meanwhile Italy has praised the long-standing ECMWF “colour-cube” approach to showing HRES cloud cover at different levels on one map.

ECMWF is also very happy to report that precipitation type products, notably the precipitation type meteograms introduced for winter 2017/18, are already being regularly used, and that feedback has been widespread and unanimously positive. Croatia (specifically CroControl, its air traffic command centre) describe this as ECMWF’s “star product”! An example from the UK report is shown in Figure 4.

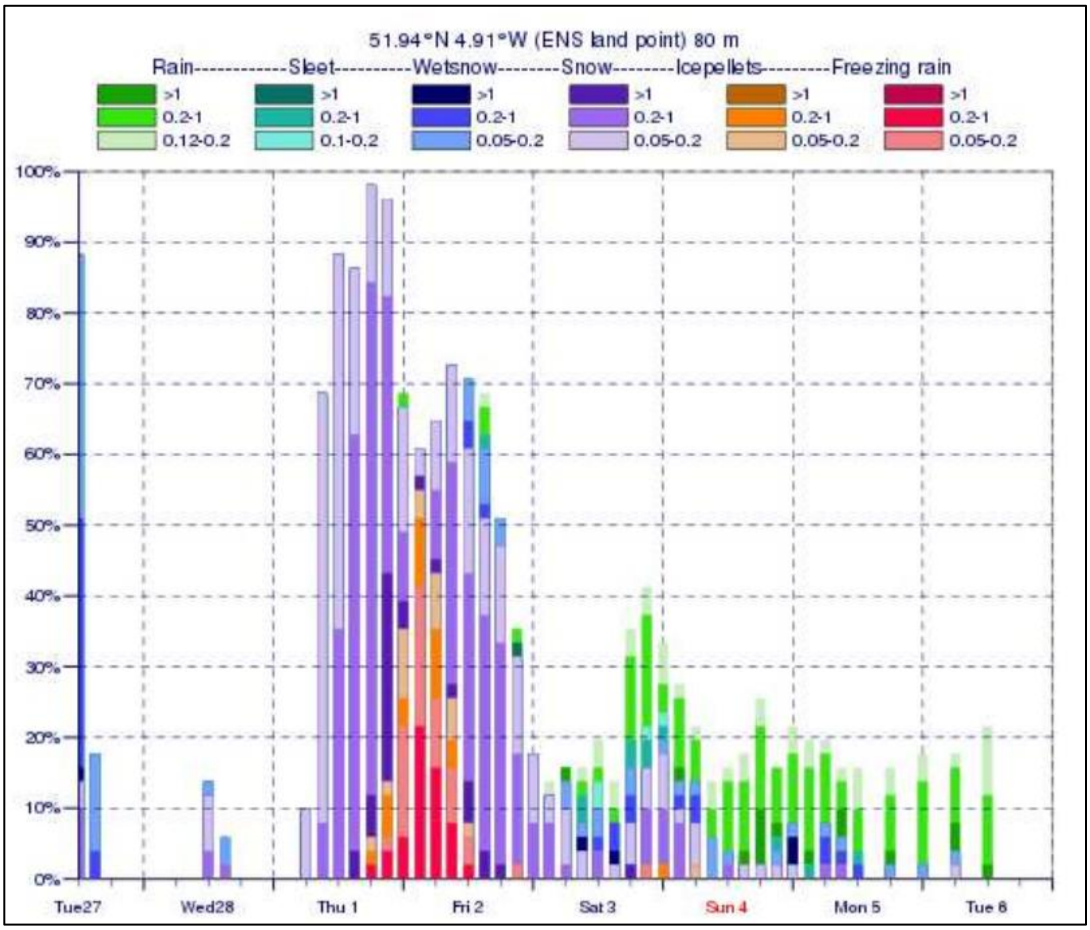


Figure 4: 12UTC ENS runs on 27th February 2018: Precipitation type probabilities for Fishguard in southwest Wales, highlighting a risk of freezing rain or ice pellets on Thursday night.

Tropical cyclone-tracking products are used by various NMSs, due to them having certain responsibilities for tropical regions (notably France, for La Réunion, which hosts one of the WMO’s six Regional Specialised Meteorological Centres (RSMCs), but also The Netherlands, Sweden and the UK).

Whilst many warnings focus on standard meteorological parameters such as wind and rain that are well served by ECMWF, warnable phenomena in some countries are more diverse, including for example avalanches and landslides (Iceland). For these more diverse categories IFS data is also used, but only indirectly.

ecCharts and other web products

ecCharts is actively used in many countries, probably more than in previous years due to introduction of new products like the precipitation type meteograms, and users report that they greatly appreciate its functionalities. In some countries, however (e.g. Hungary) ecCharts is not used directly because it is not needed, as ECMWF data is ingested fully into advanced local forecaster workstations, which provide the required tools as well as the additional facility to overlay other models and observational data (note again that ecCharts' WMS capabilities permit product transfer onto such a platform).

In response to user requests, ECMWF continues to regularly add parameters, display schemes and extra functionalities. Key update times are early summer and early winter; details are here: https://confluence.ecmwf.int/display/ECCHARTS/Updates. One major change in summer 2018 was the addition of a vertical profile display that shows both ENS and HRES data, at 6h intervals, up to T+120. This was very well received during special training sessions provided at the UEF meeting, and has even been mentioned, already, as a positive development in a few of the NMS reports (Finland calls this "an excellent product").

Of particular concern are the continuing complaints about the slow speed of ecCharts (this issue was raised again at the June UEF meeting, and also appears in reports from Belgium, Iceland, Luxembourg, The Netherlands and the UK). In part, this attests to the frequent deadlines faced by forecasters, and the complexity of their work. We continue to try to address these concerns, for example by code optimisation, by extensive speed tests when adding new and complex products, such as the vertical profiles, and by upgrading ECMWF hardware. Training can also be important to help users circumvent some speed issues via better exploitation of the many ecCharts functionalities, and by regular use of the Dashboard tool. Note also that ECMWF has provided special training sessions during some NMSs visits, and will remain receptive to requests for similar in the future.

The standard ECMWF web charts are also still widely used, especially for the ENS, with meteograms probably being the most frequently used products.

2.3 User Requests

Products

As usual there have been many requests this year for new output from ECMWF, right across the product range. These are generally dealt with through the URMS (user-request management system).

Note also that some users have requested items that have already been made available - for example Italy requested regional boundaries to be added to ecCharts; this facility was added late in 2017 (users should add a "Boundaries" layer, then edit that to include "Administrative boundaries" for Europe, see Figure 5). So there may be a requirement here for ECMWF to better publicise its updates.

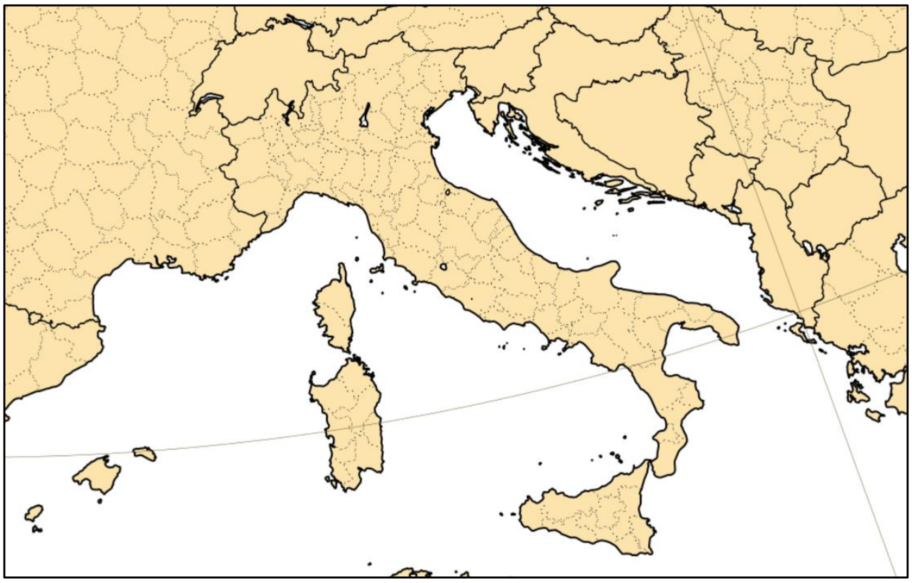


Figure 5: Snapshot from ecCharts, with administrative boundaries set to "Europe" (instead of the default, which is "OFF").

An indication of the main output requests is given below but for more details please refer to the Member State/Co-operating State reports and to the UEF2018 meeting presentations and posters.

- More turbine-relevant levels for winds for renewable energy industry needs
- Other variables targeting the renewables sector (e.g. radiation-related)
- Various aviation-related products (e.g. turbulence / icing)
- Drought index, heat stress index, icing index for shipping, fire index*
- Thermal advection, vorticity advection
- Smoother mean sea level pressure fields
- More EFI/SOT parameters (e.g. for integrated water vapour transport)
- Surge ensemble output
- Cloud layer output (Low, Medium, High) consistent with WMO height definitions
- Additional convection/stability indices (long lists!)
- Additional products for tropical waves (MJO, Kelvin waves)
- More ocean wave products in ecCharts

*Note that real-time forecasts of the FWI (Canadian) fire index based on ECMWF output are available on the EFFIS website here: http://effis.jrc.ec.europa.eu/static/effis_current_situation/public/index.html

Other

At UEF meetings in 2018 and previous years strident requests have been made, primarily by commercial customers, for HRES and ENS data from Boundary Conditions runs (from 06 and 18UTC) to be released. Approval was given by the ECMWF Council in June 2018 to make this generally available as a (chargeable) ECMWF product. Delivery began in autumn 2018. Users should note that the accuracy of these runs is much the same as the accuracy of the 00 and 12UTC. Release of this data has the potential to make ECMWF output, in general terms, even more competitive internationally. The general release of hourly data from HRES and (where available) ENS runs was also approved by Council in June. This has particular applications in the renewables field.

Other miscellaneous requests include providing an ecCharts tool for trajectory calculations, akin to the US web-based "HYSPLIT" tool (Israel). Meanwhile Italy reports that they use the ECMWF Metview model FLEXTRA for trajectory calculations, albeit not in an ecCharts framework.

Several NMS have again asked for cross-sections in ecCharts, in three formats: spatial, temporal, and spatio-temporal for ship routeing for example. Users would also like to see a cities layer in ecCharts. And as happens every year, specific requests have been made for certain contour/shading settings for certain variables; ECMWF can usually add these options fairly easily, provided the logic is clear.

Following advertisements at UEF 2018 and in other fora, Slovenia, Hungary and the UK have expressed interest in the ecPoint point rainfall products that will soon be added to ecCharts. The UK has also requested more model-based climatological information ("novel methods to get better statistics of extremes"). To this end ECMWF could in principle apply the ecPoint concept to ERA-5 output, to derive a calibrated model-based global climatology of point rainfall distributions. This would be a powerful tool in itself, providing gridded data representing point rainfall everywhere. And if blended directly with observational data one could greatly improve even the best global analyses of observed daily rainfall for past years.

The ECMWF method of CAPE computation continues to cause consternation amongst some users (e.g. Belgium, Croatia, Finland, France), due primarily to incompatibilities between this (which does not use virtual temperature) and methods used to generate CAPE values from other models used by the NMSs (which do). ECMWF has in the past cited reasons for adopting the method it does, such as computational efficiency, but agrees now to undergo an internal review of this and formally report back. It should also be recorded that some NMSs have reported regular use of CAPE fields without highlighting any issues to us (Romania, the Former Yugoslav Republic of Macedonia, Montenegro).

Finally, the UK has requested an "interactive meteogram facility", which would show for example how weather in one scenario on one day related to weather in the same scenario on the next day. This would require non-trivial development work for ECMWF, but we should acknowledge that some advanced and impressive tools that do this and more with other ENS systems are already freely available on the web. For example, the German "Topkarten" site delivers interactive plumes, where one can also click on a line and see a synoptic chart for a given member at a given time, e.g.:

http://www.wetterzentrale.de/de/show_diagrams.php?model=gfs&lid=ENS&var=2&bw=&geoid=48401

3. Verification of products

3.1 Objective verification

Most countries have reported results from the verification of ECMWF forecasts, generally by comparison with observations in the local area of interest. Of relevance to interpretation are the dates of the most recent upgrades to the IFS:

Cycle 43r1 became operational 22 Nov 2016

Cycle 43r3 became operational 11 July 2017

Cycle 45r1 became operational 5 June 2018

This means that in this year's reports verification corresponds mainly to cycle 43r3, although in some the verification period used is 2017 which will encompass cycles 43r1 and 43r3 in about equal measure.

As always, *year-on-year* changes in IFS performance depend also on the prevalence of different synoptic patterns, that can have different associated error characteristics, so apparent changes in performance relative to "last year" need to be treated with caution. Internally, to assess the long-term skill evolution, ECMWF subtracts from statistics for the operational forecast the equivalent statistics derived from a fixed model version run over the same period, which can help eradicate impacts of this type.

And when considering a *fixed period*, there are likewise several reasons why one would not expect consistency, a priori, in the verification results (e.g. bias, RMSE, etc.) reported by different countries. Firstly, different weather patterns will have very probably prevailed in different regions. Secondly, the impact that a certain weather type has on skill and biases will manifest itself differently in countries with different (fixed) geographical characteristics. For example, issues handling orographic rainfall, which we know exist, will clearly have little or no impact on a flat country, but can have a substantial impact in mountainous regions. And thirdly, a range of "interpolation" and "site-selection" techniques are being used. Full resolution IFS output is not always being exploited, and in some reports received the method(s) of extraction and interpolation are not entirely clear.

A conditional verification approach (which is now being increasingly used at ECMWF) can go some way towards resolving some of the issues listed above, and some results deriving from that are presented below.

3.1.1 Direct ECMWF model output, and comparison with high-resolution models

Many reports focus on comparing HRES with LAMs, and for this reason usually centre on the shorter ranges (up to about 48h). A common finding, seen in virtually every verification result, for almost every sensible weather parameter, was that biases in IFS forecasts have a diurnal cycle. Annual cycles are also often present.

Overall, whilst there are large model-to-model and parameter-to-parameter variations, and some large differences between countries, the impression one gets this year is that LAM performance is overall slightly better than HRES performance, meaning also that LAM performance has, in relative terms, improved somewhat (in the last two years the two were subjectively ranked as similar in this document). One of the reasons for this change may be that LAMs that had a resolution similar to or slightly better than that of HRES (9 km) have been retired in favour of using runs at higher resolution (which is now typically about 2.5 km). It took a while, following the HRES resolution change in March 2015, for the benefits of upgrading LAM resolutions, and retiring lower resolution versions, to be recognised and implemented.

10 m wind continues to be the parameter where resolution really matters and several countries show that LAMs provide significantly better results than HRES. One expects this to be particularly apparent in countries that have complex topography and societally important gap flows, such as Croatia. Indeed, Croatia runs large eddy simulation models in re-forecast mode to tackle this issue (for design purposes), although at the same time experts there have reported significant improvements over the years in HRES handling of Bora events. Precipitation biases in LAMs, for both small and large totals (versus point observations), are mostly smaller than for HRES. It is also clear that handling of surface weather parameters by different models can vary greatly according to synoptic situation, geographical region and parameter in question.

Ultimately all models have their strong and weak points, and the impression one gets from the wealth of statistics provided in this year's reports is that in the short ranges at least (where the bulk of the comparisons were performed) a multi-model approach to forecasting has considerable merit. This would be particularly true if one could vary weightings according to known synoptically-varying performance characteristics. There is little evidence that such a strategy is being applied automatically at the moment. However, during the Member State visit to Belgium one exciting development was revealed - the introduction of "situation-dependant weighting". In the meantime this concept is undoubtedly being used in subjective fashion by forecasters across Europe.

Details, by parameter, are given below. Some of the IFS issues raised here are known, and most of these are also listed in the ECMWF's publicly accessible 'Known IFS Forecasting Issues' web page at which has been updated a number of times in the last year.

ECMWF made a specific request for ENS-related verification statistics this year. In spite of this, limited verification of this type has been received, as in previous years. However, the report from Greece states that new software to perform ENS verification has been installed, and Ireland is also planning work in this area, so perhaps we can be optimistic for the future. In the short and medium ranges reports on ENS-related verification were as follows: surface parameters (Hungary, IFS ENS mean), tropical cyclones and regimes (UK, Global ensembles), ocean waves (Israel, IFS ENS), a form of vertically integrated water vapour transport (Switzerland, IFS ENS), 2 m temperature (Finland, IFS ENS and its LAM-EPS systems), 12h precipitation (Switzerland, IFS ENS and its LAM-EPS system) and 10 m winds (Denmark, IFS ENS and its LAM-EPS system, as last year). Some monthly and seasonal forecast verification is also provided, but this was mostly performed by treating the ensemble mean or median as if they were deterministic forecasts. Discussion of these various ENS-related contributions is incorporated into the sub-sections below.

2 m temperatures

Combining the results from all reports one can conclude that for 2 m temperature LAMs perform as well as if not slightly better than HRES, especially if the resolution difference is large. By better we mean smaller MAE and RMS errors, and these are usually accompanied by smaller biases. Better representation of complex terrain (mountains, lakes, small islands) undoubtedly contributes to the relatively high skill of high resolution LAMS in some regions.

Norway, Sweden and France all indicate that their AROME model versions, at 1.3 and 2.5 km resolution, have notably lower MAE/RMS errors than HRES. Denmark and Serbia indicate that their LAMs used to be better than HRES, but that with HRES improving that gap has diminished or disappeared. Meanwhile in Greece, Hungary, Croatia and Romania LAM skill is reported to be somewhat inferior to that of HRES. In Croatia the MAE difference at day 2 in their 4 km ALADIN model is double that seen for HRES. In Romania the superiority of HRES will be partly because the LAMs they use are, relative to many other LAMs, “low resolution” (6–7km).

In terms of bias, it has been frequently reported that HRES forecasts have a cold bias (Greece, Iceland, Ireland, Italy, Luxembourg Norway, Serbia), typically of order 0.5-1ºC. Similarly, Belgium, Luxembourg, Norway and the UK discuss the under-prediction of maxima on hot/summer days, by 2ºC or more. ECMWF is aware of this issue, which is under investigation within an internal task team. It should also be noted however that Croatia provides evidence of good HRES forecasts of maxima during a hot spell. Meanwhile Hungary depicts the diurnal cycle of errors in 2017 and shows, for days 1 and 2, as well as a cold bias in late afternoon, a warm bias for most other times (both ≤ 1ºC magnitude), which is structurally the same as the picture one gets of errors on clear days (Figure 13, top panel), albeit with a lower amplitude.

HRES performance in Finland seems markedly worse than it is in most other countries, at least in winter, due mainly to difficulties handling strong low-level inversions which are commonplace due to frequent snow cover, light winds and minimal insolation. Lower minima, below about -10C, tend to not be cold enough in HRES, and for observed minima around -30C a typical error is ~10C - see Figure 6. Another error important for Finland occurs in spring, which is that evening temperatures there are systematically too low in IFS forecasts. This issue has been reported many times, but has proved very difficult to correct in raw model output. Regarding EPS performance Finland also shows, for autumn 2017, that its MetCoOp EPS system (based on Harmonie) achieved slightly smaller biases than ENS, although its RMS errors were still very slightly worse.

Finally, we note that France investigated the impact, on AROME 1.3 and 2.5 km forecasts, of providing BCs from HRES instead of ARPEGE. Overall it made no difference, not just for 2 m temperature, but also 10 m wind speed and 2 m relative humidity.

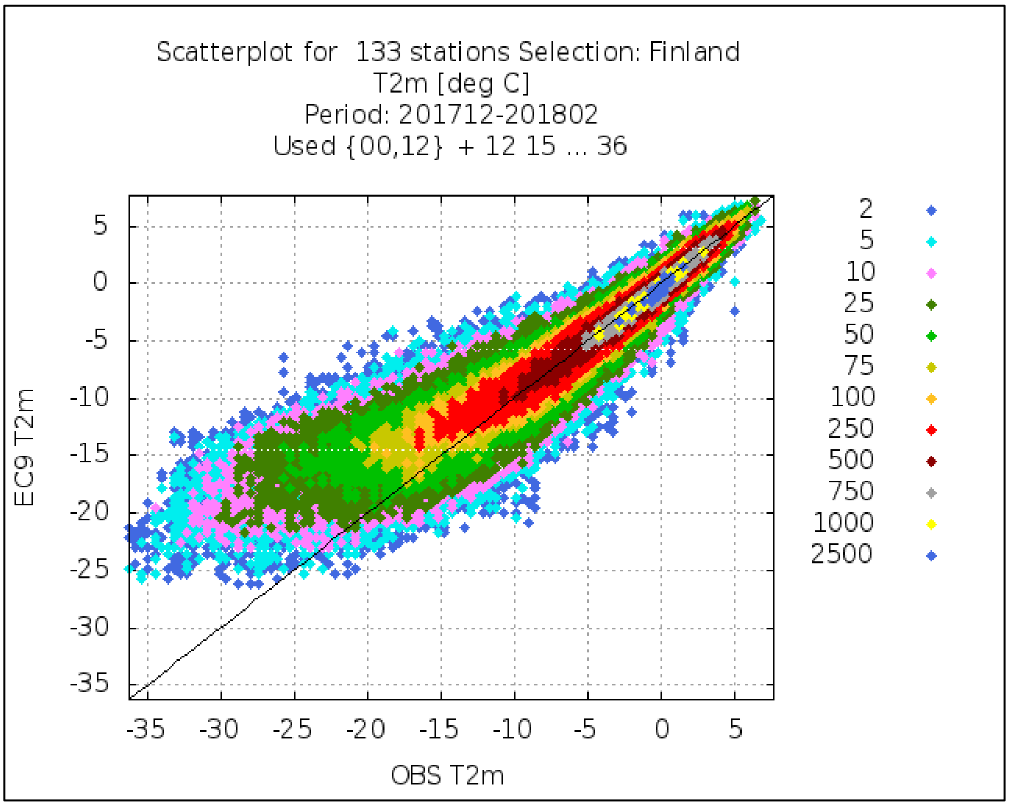


Figure 6: HRES 2m temperature forecast verification for Finland, days 1 to 2, DJF 2017-18

10 m wind

Combining results from different reports, as for temperature, it seems now that HRES is providing mean wind speed forecasts that are somewhat inferior to those of LAMs (Denmark, France, Hungary, Iceland, Norway, Sweden, all of whom use AROME or HARMONIE-AROME at 2.5 km, and Italy that uses COSMO at 7 km). In relative terms this represents a deterioration for HRES since last year, although a key factor here may be the trend to higher LAM resolution in reports this year. Physically, one also expects a much bigger impact of topography on local wind speed than on some other variables, and therefore resolution can be critical (as noted by Croatia). The same general picture of LAM advantage applies to gust forecasts. A curious result apparent in data from Latvia is an increase, with lead time, in a positive bias in HRES gust forecasts (e.g. from 1.8m/s at D1 to 3m/s at D6). Not all countries include gust verification, although encouragingly this seems to be a growth area. ECMWF particularly welcomes this development, due to the difficulties of procuring and using gust reports when reporting practices differ between countries, as noted in ECMWF Technical Memorandum 834 - Use of in situ surface observations at ECMWF.

Iceland shows the gap between HRES and HARMONIE-AROME forecasts to have widened, in favour of the latter (Figure 7), which it attributes to roughness length changes. The first of these related to the sub-grid orography (September 2013), the second, more subtly, to a change in leaf area index (September 2015); they manifest themselves as step changes in bias on Figure 7 (lower purple line). There may be merit in ECMWF examining the influence of LAI on roughness.

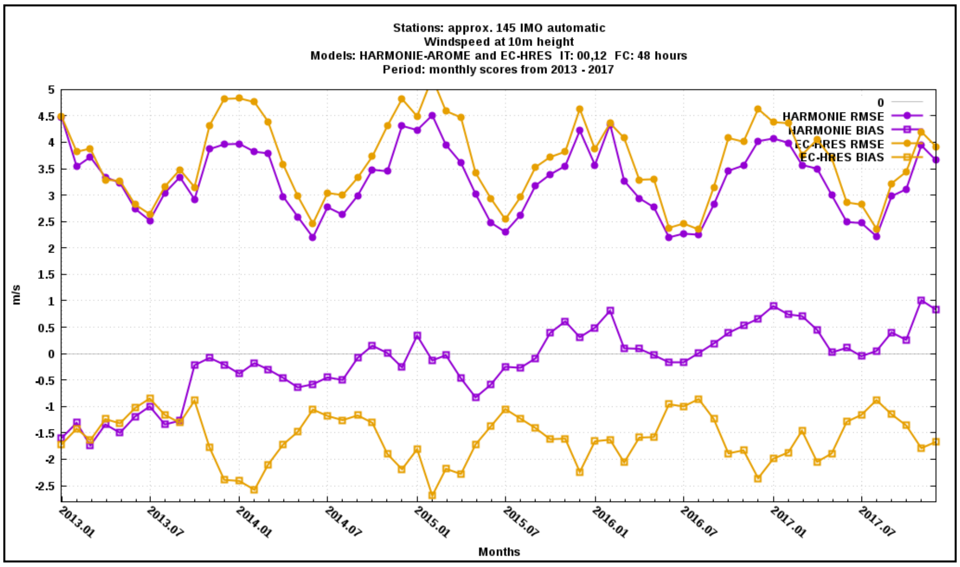


Figure 7: Monthly values of RMSE and bias for 48h forecasts of 10-metre wind speed (m/s) for 2013-2017. HARMONIE-AROME (purple) and HRES (yellow), initialized at 00 and 12 UTC

With regard to biases, there are strong indications of a general slight over-prediction of speeds in HRES, up to ~1m/s, with larger biases generally being seen at night (Belgium, France, Hungary, Latvia, Serbia, Sweden). This broadly concurs with ECMWF's own verification results. However as reported before the bias signal is opposite in sign, and larger in magnitude, over Iceland, causing forecasters there to use 100 m wind speeds as a guide to 10 m winds in generally windy situations. Similar issues have again been noted over mountains in Norway. Curiously there is also slight underprediction (up to ~0.5m/s) over Ireland, Italy and Greece on average, in all seasons, with the largest errors typically found by day. Sweden notes that strong *gusts* appear in shorter range forecasts about twice as often as they should, although at coastal sites there is no such bias. The mountain winds problem is recorded on the unrestricted "Known IFS forecasting issues" web page.

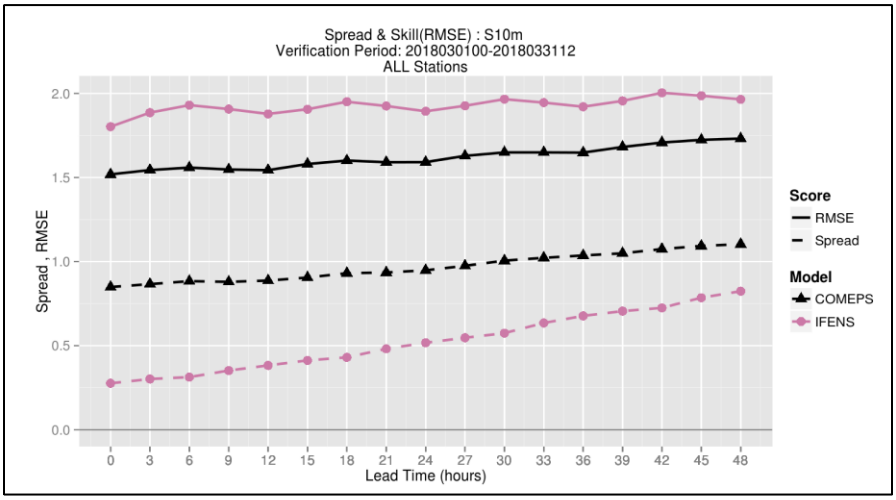


Figure 8: Average Spread (dashed) and RMSE (Solid) versus lead time for forecast 10m wind, March 2018, comparing ENS (magenta) and DMI-COMEPS (black) with synoptic reports over Scandinavia.

Denmark again shows encouraging results for their COMEPS EPS system, which has better skill than ENS and a much better spread-skill relationship (ENS spread is clearly insufficient). This is shown on Figure 8, which is for all of Scandinavia but only for March 2018. They also suggest COMEPS has a much better representation of coastal windstorms near Greenland. On the other hand, over-prediction of speeds in stable conditions is often a problem for both EPS systems.

Precipitation

We begin this sub-section with “overall impressions”, which are backed up in many but not all cases with data.

Italy suggests HRES performance is similar to previous years, whilst Serbia and Slovenia have identified improvements. For Slovenia this is in marked contrast to five years ago, when they indicated, during an official visit by ECMWF, that precipitation forecasts were not used or even examined in summer, as they were so unreliable. At the same time, a common viewpoint in NMSs is that summer convective precipitation is handled better by LAMs (stated by Denmark, France, Luxembourg and The Netherlands).

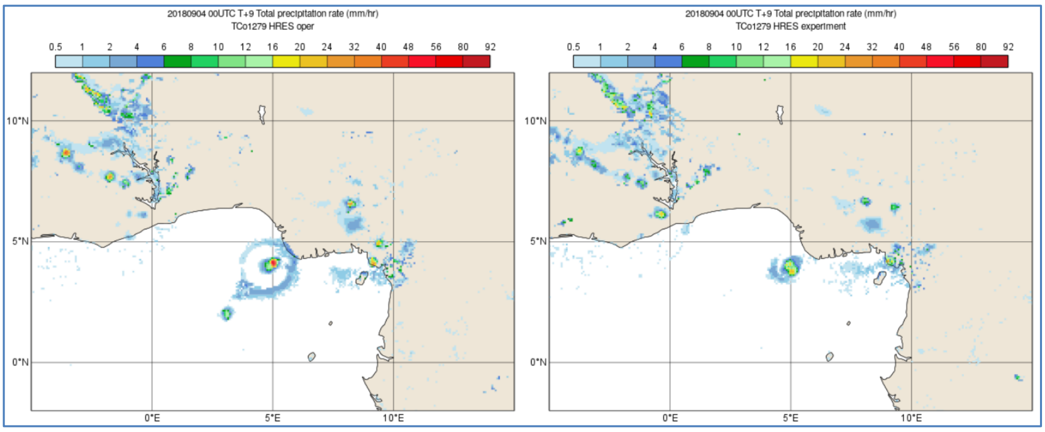


Figure 9: Precipitation rates (mm/h) from short range (9h) HRES forecasts compared, over the Gulf of Guinea region - operational (left) and experiment with coding error corrected (right).

Related also to the summer convection question, Belgium states that MCSs pose significant problems for the IFS, which is a complex model issue well known to ECMWF. Another well-known and well-documented issue is the problem of inland penetration of SST-based convection, highlighted this year by Israel. Others have again referenced the problems with excessive (warm rain) precipitation over lakes and coasts. Here we have positive news: substantial improvements related to this were made in the 45r1 model cycle introduced in June 2018. However, Norway have suggested, conversely, that in the previous cycle there was evidence of over-prediction of orographic rain in warm sectors. In light of this we would encourage close monitoring of the performance of 45r1 in similar situations. We also have positive news regarding the oceanic “precipitation bullseye” problem referenced by France on several occasions. A long-standing subtle coding error, related to the handling of supersaturation, has just been identified, and testing is now underway for a related model change to hopefully be implemented in the next-but-one cycle. Figure 9 shows an example of the impact of correcting the coding error - extreme rates on the left panel are diminished on the right panel (notably over the ocean near 5ºE), whilst the unrealistic (expanding) ring of convection over the Gulf of Guinea, "triggered" by the bullseye at its centre, has almost disappeared.

For precipitation forecasts the verification metrics used have generally been different to those used for temperature or wind. This aspect, the sometimes-disaggregated nature of precipitation fields, and the different interpolation methods used together make result inter-comparison rather more difficult than it is for wind or temperature.

When examining results, it should also be noted that almost all countries compare forecast totals, which innately apply to a gridbox, with point observations. Reported results will therefore be influenced by the representativeness issue caused by the inherent difference between grid box (area) totals and point values. Sweden's results are something of an exception; they are dealt with separately, at the end.

Many countries quote the frequency bias index, or FBI, for point measurements. For HRES results are again fairly consistent between countries; small total frequencies tend to be overestimated, and large total frequencies underestimated, which is a virtually inevitable consequence of the point versus gridbox approach, so this is not necessarily a model problem. For LAMs FBI profiles versus threshold are almost always more horizontal, and usually close to 1, which is also what one expects, for smaller gridboxes (although ALADIN was reported by Hungary to over-predict in every category). For HRES the "crossover" FBI=1 is around 10-20 mm/24h, although this can also vary markedly. Italy's FBI results are revealing; in winter the crossover is ~16mm/12h, for different lead times, but in summer it varies a lot - e.g. 3.5mm/12h for T+12-24 (12-24UTC), and 15mm/12h for T+24-36 (00-12UTC). Summer results for Greece have a very similar character. The main reason for the differences is probably the diurnal-cycle-related biases in convection handling which manifest themselves in summer only. It should be re-iterated here that ECMWF has developed a way of converting forecast gridbox totals into point total pdfs (probability density functions); a procedure which can define likely sub-grid variability and bias in different meteorological/geographical situations. It can deliver FBI profiles that are almost horizontal and close to 1. The initiative is called "ecPoint-rainfall". Experimental products related to this work are now undergoing real-time testing in different parts of the world, and are due to be introduced into ecCharts in spring 2019 (see here: https://www.ecmwf.int/en/about/media-centre/focus/new-collaborative-agreement-between-ecmwf-and-ecuador-related-extreme-rainfall-and-flood-risk). The diurnal cycle errors are not addressed directly; however, that is planned for a future version.

Switzerland reports briefly on a new initiative to verify *hourly* totals forecast for 560 Swiss sites. And for the model data they now use a nearest gridpoint approach, rather than interpolating to site locations. This revised approach is better and is what ECMWF would recommend for rainfall. Their results highlight an overprediction bias (FBI=1.4) for modest rates of 0.5mm/hr. Again, sub-grid variability is probably the main reason.

Other than bias metrics, which innately favour the LAMs, not that many other measures were used to compare HRES and LAM performance. Using the Equitable Threat Score, Serbia shows that ECMWF performs better than their 4km LAMS for thresholds of 4 and 20mm/24h, whilst Italy shows the COSMO 5km model performance to be similar to HRES across a wide threshold range, albeit with a slight advantage for very small totals. Using SEDI Hungary shows HRES to perform better than LAMS (2.5km and 8km) up to thresholds of 10mm/24h, and worse for higher values. In this regard note also that ecPoint-rainfall substantially improves upon ECMWF's raw ENS for all totals up to and beyond 50mm/12h (using ROC area and reliability metrics). It would be interesting in due course to see how this improved performance compares with LAM and LAM-EPS systems. Indeed, in a collaborative EU project with Italian partners (called "MISTRAL"), ECMWF is working to optimally blend together point rainfall output with post-processed LAM-EPS output, to achieve the best possible rainfall forecasts for a central Mediterranean region. Knowledge gained during MISTRAL will be valuable in helping ECMWF assess future ENS configuration options, as resolution approaches the "grey zone".

Sweden verifies and compares HRES and (2.5 km) AROME on different length scales, using the Fractions Skill Score (FSS), requiring that there be at least three observations in each box for inclusion. Whilst they report that there may be some result contamination because of HRES interpolation, their overall conclusions were (i) AROME is better, (ii) AROME adds most value for small totals, but is better also for heavy rainfall events, (iii) HRES is better than AROME at *localizing* heavy rainfall events (i.e. for the smallest boxes, ~30 km, in the FSS calculations). Separately Sweden also examines the diurnal distribution of rainfall for two summer months (365 sites), which has similar cycles in HRES, AROME and observations, although the HRES peak comes too soon (Figure 9) as expected from parametrisation limitations. Most significantly, their plot implies that HRES over-predicts rainfall in some net sense, by about 25% (whilst AROME bias is close to zero). This over-prediction concurs qualitatively with ecPoint-rainfall bias correction, and with more standard rainfall verification results from ECMWF, although 25% still seems rather high. ECMWF is now actively investigating the issue of net biases in its rainfall forecasts.

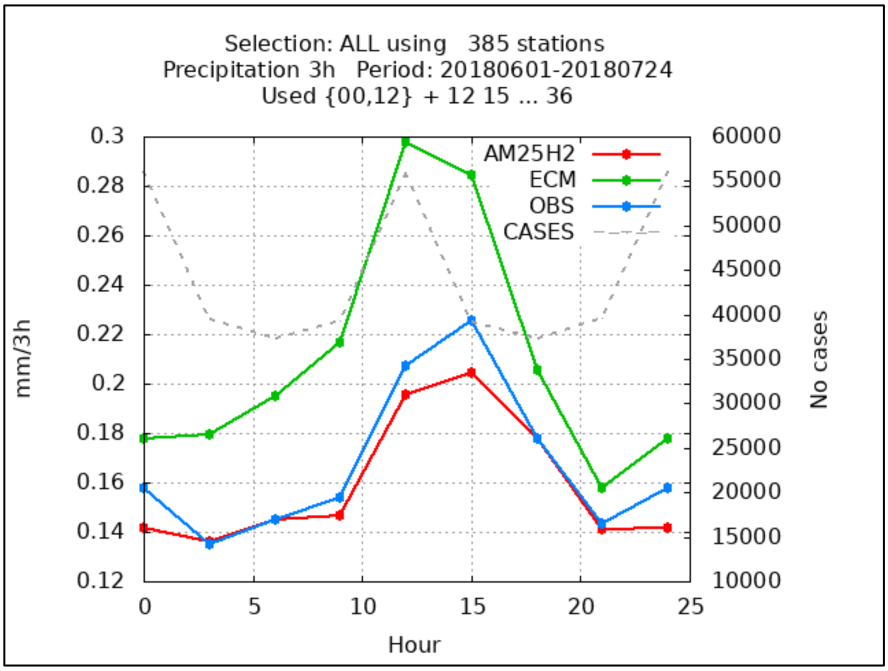


Figure 10: Diurnal cycle for 3h precipitation over Scandinavia 1 June to 24 July 2018. Red is AROME, green is HRES and blue is observed precipitation.

Switzerland shows some interesting EPS reliability diagrams for 1mm/12h totals. For such totals, the COSMO-E LAM-EPS is more reliable than ENS in both summer and winter over Switzerland, and indeed almost perfectly reliable in summer. Again, sub-grid variability should be *much* less of an issue at COSMO-E resolution (2.2 km) than it is for ENS (18 km). Data for higher thresholds would also be interesting to see.

Screen-level humidity / cloud cover

ECMWF would like to thank the Member and Co-operating States for the increased level of contributions regarding humidity made this year. Low level humidity may not be a priority component in forecasts for customers, but it can be a very important factor for CAPE and convection.

The biases and RMSEs reported in screen-level humidity (commonly represented as dewpoint) were generally better in HRES than they were in LAMs. This may be testament to an effective land surface scheme in the IFS, though of course other aspects play a big role too. Greece and Hungary show HRES performance to be better than their LAM systems, for Sweden the two are comparable, whilst for France the AROME LAM is marginally better (over one French domain, but also two tropical island domains which would be innately more challenging for HRES). A common general theme is that HRES has a slight dry bias in Europe. For example, Greece shows dewpoint biases ~ -0.5ºC generally for HRES. For their LAMs it ranged from -1 to -3ºC. However, in Switzerland the equivalent HRES value is -2ºC, all year round, which may be symptomatic of the extra challenge, for moisture handling, afforded by topographic complexity. This may also have a detrimental impact on forecasts of summer convection there, something which ECMWF should probably examine more closely.

Finland again reports somewhat unusual results, specifically that absolute humidity in HRES tends to be too low in winter, in spite of there being a large positive 2 m temperature bias at the same time. These biases, which imply an even larger negative *relative* humidity bias, are consistent with a HRES bias to concurrently predict visibilities that are too good, though at the same time seem inconsistent with a propensity to forecast stratus sheets at very low temperatures when skies stay clear (forecaster comment). The complex interplay between these various aspects at low temperatures, and modelling thereof, is a topic being studied during YOPP (the Year of Polar Prediction), which ECMWF is involved in through the APPLICATE project. We would like to re-iterate that in this context regular radiosonde ascents in cold climates, and indeed data from dedicated multi-sensor surface measurement sites, such as the one at Södankyla in Finland, can be extremely valuable to ECMWF.

Another useful result, highlighted by Serbia and Belgium, is a slow drift to larger magnitude (negative) biases at longer leads (day 4+) in HRES (Figure 11 shows data for just one Belgian site but this trend is present for all 13 shown). So there has been net drying out of the lower troposphere during HRES integrations. In showing single site verification statistics Belgium is also highlighting that there can be large local variations in bias, attributable, for example, to industrial moisture sources (as on Figure 11, hence the dry bias), and/or topographical aspects, and/or land-sea issues. Seasonal variations in dewpoint accuracy/bias are also shown to be large in some NMS reports.

With regard to cloud, Sweden makes some useful points about the limitations of ground based observations of different types - automatic and manual - and systematic differences one should expect between them, which suggest that we should always be careful to not read too much into small reported biases.

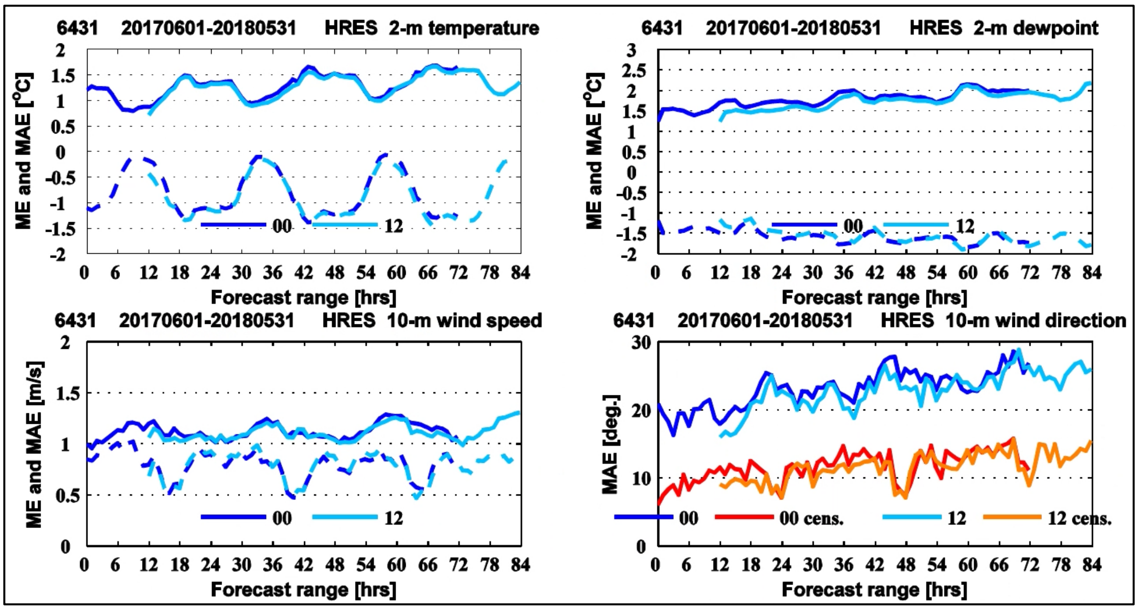


Figure 11: Verification for site "Ghent/Industry Zone" (station 06431) for 4 surface variables. Mean errors (ME, dashed) and mean absolute errors (MAE, solid) of HRES forecasts based on 00UTC (dark blue/red) and 12UTC analyses (light blue/orange). For 10-metre wind direction, only MAE scores are shown, before (dark blue/light blue) and after (red/orange) removing light winds <4m/s from the dataset. Period is June 2017 to May 2018.

Regarding cloud cover biases in general, Belgium, Switzerland and Sweden all suggest that for HRES they are small. The fact these are also geographically diverse countries seems encouraging. On the other hand, Hungary show a marked negative bias in HRES of 20 to 30%, whereas their 8 km ALADIN version has a bias of less than 10%, and an RMSE that is, strikingly, almost 50% smaller (Figure 12).

It may be that we can learn from the ALADIN formulations, though physically the picture is complicated because ALADIN also suffers from marked over-prediction of rainfall, as discussed above. If one examines cloud at different levels, more issues can become apparent. Sweden, which shows minimal HRES bias in summertime cloud cover when considering all base heights up to 7500m, also illustrates a concurrent under-prediction of *low* cloud, with AROME at 2.5 km resolution performing much better.

Greece shows similar skill levels for their LAMs relative to HRES. With only the results from Greece, Hungary and Sweden to reference we cannot say anything definitive about how HRES and LAM performance compare with regard to cloud handling in general.

Two countries comment on handling of cloud at polar latitudes. Norway states that in the Arctic there is too much. Conversely the UK states that the IFS is "highly deficient in low cloud and fog around Antarctica in Austral summer" (which is important for operations there). This came from forecasters seconded to the British Antarctic Survey, working in the region. They postulate that moisture exchanges related to the sea ice are not being correctly modelled. Observations made during YOPP should help.

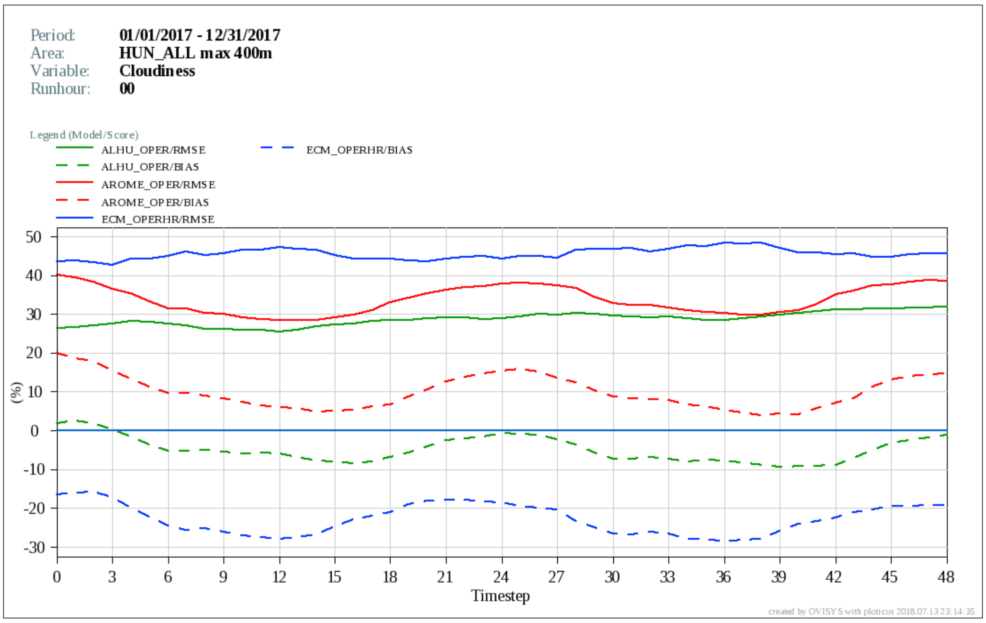


Figure 12: 2017 comparison of RMSE (solid) and bias (dashed) for total cloudiness forecasts of 00UTC runs of HRES (blue), ALADIN/HU (green) and AROME/HU (red) over Hungary

Finally, in the subjective feedback category, Israel notes that stratus (and marine inversions) are too low near coastlines in HRES, whilst Belgium suggests that overall HRES has too much stratus and too little cumulus.

Visibility

Statistics relating to visibility were provided by Finland, France, Iceland and Latvia, whilst many other countries provide comments. Happiness with ECMWF's visibility product varies. Luxembourg is very content, and use it during fog episodes, whilst Bulgaria, at the other end of the range, considers the product in general not good enough to use. Sweden gives positive feedback whilst Croatia uses ENS visibility frequently, but in just a qualitative fashion. Finland states that visibility has some utility, but is not trustworthy, and is too high when snow is falling. Morocco expressed interest in visibility verification.

Although the visibility output parameter was revised in July 2017 (cycle 43r3) to incorporate the new seasonally varying aerosol climatology (based on CAMS aerosol re-analysis), no NMS reported any step change in quality, although one can clearly see step changes in values, more in some regions than others. The revision affected higher visibility values, which may be why it has been somewhat "transparent"; interest in poor visibility and fog is much higher.

Iceland's verification covered a short period, with only a few stations, but suggested that performance of its HARMONIE-AROME LAM was markedly superior to that of HRES. France meanwhile shows just HRES verification for a somewhat "foggy" autumn 2017, using mainly observations from airport automatic sensors. Skill in different visibility categories exhibits diurnal variations, with very low visibility forecast too often, and better visibilities, e.g. >5 km thresholds, predicted about half as often as they should be. Although ECMWF was aware of some over-prediction (as recorded on the model issues web page), the result that visibility<150 m is predicted 6x too often at 06UTC is rather concerning (Figure 13). If used directly this could have a major impact on warning credibility.

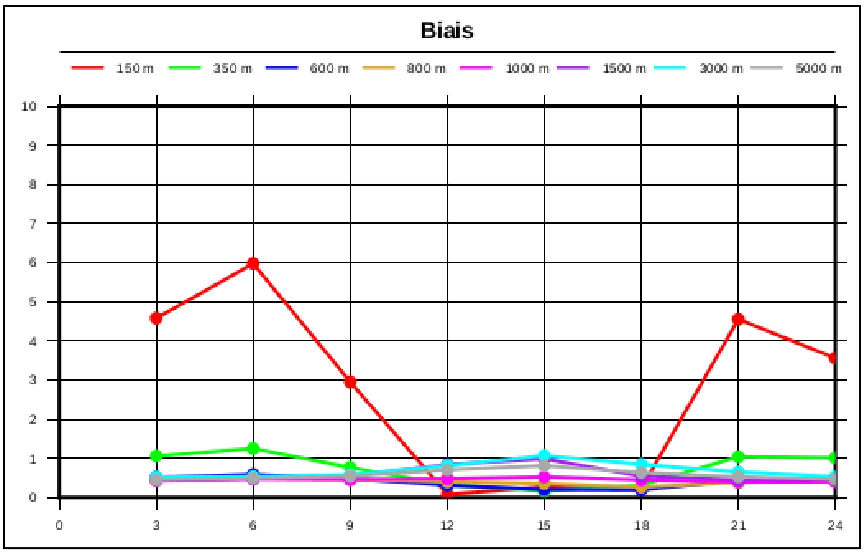


Figure 13: Frequency bias in HRES visibility forecasts, autumn 2017, for lead times 3-24h, for ICAO thresholds in metres: 150 (red), 350 (green), 600 (navy), 800 (yellow), 1000 (magenta), 1500 (purple), 3000 (grey), 5000 (cyan).

In future, ECMWF may need to focus more on the physics of fog production, and visibility changes. Generally, fog that is very dense in a horizontal sense tends to be fog that is vertically thin. This is because horizontal visibility in fog that thickens vertically tends to improve at ground level as the ground heat flux begins to warm (and dry) the air there faster than net outgoing long wave radiation can at the same time cool (and moisten) it. Any snow cover would however be a complication. Correct representation of the physics would require many vertical levels; to this end the current lowest HRES levels at 10, 31 and 54 m are too far apart, and separation in ENS is worse. Another issue, raised at the UEF meeting, is aerosol effects, which provide condensation nuclei on which fog droplets form. Incorporation of this aspect may in time be possible through links with CAMS. In the meantime, expectations probably need to be managed.

A viable alternative in the shorter term to expensive model development is probably post-processing; a related Austrian poster at the UEF showed that visibility forecast improvements can potentially be delivered, relative to HRES and ENS, using for example model dewpoint depression and latent heat flux as predictors.

SST and Sea Ice

Two very similar comments about ECMWF SST fields were raised by Spain and Croatia. They relate to the “long-standing” issue of near-coastline SSTs in interpolated datasets, which tend to be too low because of erroneous use, in the interpolation, of land-based SST values configured to be 0ºC in the archived native grid fieldsets. This can be particularly problematic when ingesting such data into a LAM. Use of ECMWF's new interpolation package MIR will unfortunately not resolve this issue automatically; ECMWF recommends interpolating from the MARS dataset using the "nearest neighbour" method instead of the default setting.

There has also been adverse feedback about water temperatures for lakes Geneva and Constance, which can be unrealistic. This issue probably relates to the lake mode FLake, and its sub-optimal handling of temperature initialisation for deep lakes (as included on the Know IFS forecasting issues web page). Although we need to retain a relatively simple lake model for computational reasons, improvements should be possible and are being investigated.

Miscellaneous feedback over some years now has repeatedly referred to sea ice issues near to coastlines, which ECMWF has tried to tackle in various ways, not always successfully. The introduction of coupled ocean and sea ice models into HRES in June in cycle 45r1 will certainly help alleviate any remaining problems, e.g. close to Iceland and in the Baltic.

Waves

France shows that wave forecasts provided by the French wave model are very accurate compared to competitors. Whilst this is in part due to use of HRES winds, it relates also to their parametrisation of physical effects; ECMWF is now incorporating some of these physical effects, such as dissipation by open ocean wave-breaking, into its own wave model, for release in the next cycle, 46r1, in 2019.

Israel discusses at length wave model verification for several buoys close to its Mediterranean coastline. They raise various issues with wave height forecasts, such as systematic under- or over-estimation of extremes, and far too little spread in the wave ENS. Whilst it is naturally true that the wave model and wave ENS can be further improved, the location of the buoys used, which is close to the coastline, introduces so many complicating factors that it is difficult to make any concrete conclusions. Careful grid point selection/interpolation, that they apply, can only mitigate these effects to a certain extent. The complicating factors include details in ocean bathymetry, currents and current-wave interaction, inadequate representation of coastal winds in general and inadequate representation of the influence on winds, waves and currents of complex topography such as that found around the Haifa promontory (which is about 6 km long, up to 400 m high). One of the buoys is 2 km from this promontory, whilst the resolution of the wave ENS is 28 km. To deliver better wave forecasts near to coastlines would require introduction of an unstructured grid in a global model (which is not currently in ECMWF plans), or running of a local, high-resolution coupled wave-atmosphere model. It is worth noting that to provide good near-coast predictions for Olympic sailing events in 2012 the UK ran a model with 300m resolution. Israel does not comment on whether there is wind data from their buoys; if available this could help somewhat to place the wave forecast "errors" into context.

ECMWF would also like to encourage all Member States to push wave observations to the GTS wherever possible, because they can help ECMWF enormously to provide feedback, to perform verification and to improve the model.

Other

Switzerland uses, in an operational medium-range flood alert system, the causal relationship between floods within mountainous regions and (extreme) integrated water vapour transport (IVT). A key parameter here is flow orientation relative to the orographic barrier (i.e. the orientation of its anisotropy), and to this end they have devised a direction-weighted IVT parameter for prediction. They have also created an extreme-event-weighted continuous ranked probability skill score (CRPSS) for this metric to assess its utility. As is often the case with extremes verification, sample size has been an issue, but for regions with sufficient sample they show utility for flood prediction in the ENS, with this method, up to ~ day 5. For all events utility extends much further, to ~ day 10. In June ECMWF introduced IVT into its real-time product suite; conceivably this could be extended to account for topographic orientation (globally) as in the Swiss approach.

There have again been references to snow depth issues (e.g. by UK), which ECMWF is aware of and is working on, partly with its upcoming multi-layer snow scheme. The main problems are spurious accumulation of snow on the ground when a mix of rain and snow (sleet) is falling, and snow on the ground not melting quickly enough in general. During the year ECMWF did identify and fix one bug relating to the former problem, following a report from Bulgaria. This helped a little.

Forecaster Impact

Some countries verify or at least comment upon forecaster performance compared to models. Whilst performance gaps are small, Hungary shows that forecasters continue to add a little value, on average. This added value in a composite, multi-parameter index equates to a lead time gain of one day over the best of the other model scores shown, which is for the ENS mean. This applies up to day-3. At longer leads “ENS mean” and “forecaster” are about equally good, and remain the best options on average. The forecaster adds most value to the maximum and minimum temperature parameters, but can also improve cloud cover forecasts significantly at shorter lead times. In the Croatia report, there are hints of a positive forecaster contribution to temperature forecasts during one short summer period, whilst Finland reports that forecasters can add value even to MOS forecasts during inversion situations (note the poor quality HRES forecasts on Figure 6). Conversely Denmark illustrates how forecasters’ predictions over a one year period, again for 2 m temperature, have been marginally worse than raw HRES output, for lead times up to day-5.

Conditional Verification

“Conditional verification” is the concept of verifying one forecast parameter for a *subset* of all cases, in which the subset (=the conditions) is defined by value range(s) of one or more parameters, in model and/or reality. It underpins many insightful subjective comments provided by forecasters over the years - e.g. “in this situation the model tends to get this wrong” (see parameter-specific sub-sections of 3.1.1 above for examples). Objective conditional verification is a growth area that holds a key position in ECMWF’s plan of work for verification, and that can also add value in post-processing if model-based subsets are used, as in ecPoint discussed above. Used carefully conditional verification can give physical meaning to certain types of bias/RMSE behaviour, which can help with improving the model formulation, particularly the physics. To this end ECMWF specifically requested conditional verification results this year. All responses used reality-based subsets; Belgium and Hungary provided direct input, whilst some other NMSs included self-referential conditional verification (e.g. showing very strong winds to be underpredicted when they occur; results like this require a cautious interpretation, particularly if they are for longer leads).

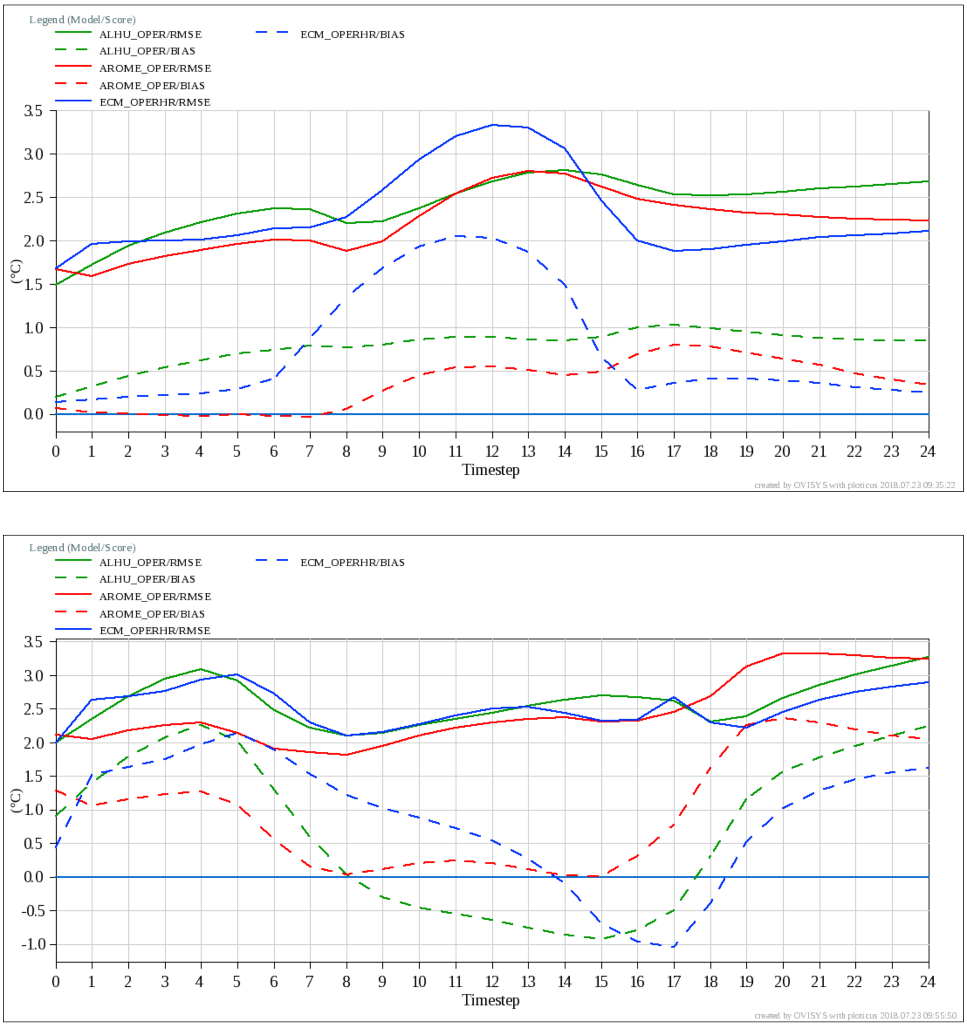


Figure 14: Comparison of RMSE (solid) and BIAS (dashed) values of 2 m temperature forecasts, 00 UTC runs, of HRES (blue), ALADIN/HU (green) and AROME/HU (red) over Hungary when the sky was clear (top) and covered by low-level clouds in most of the country (bottom).

Belgium shows how errors in wind direction forecasts are strongly dependent on the magnitude of the wind speed observed; lighter winds have much larger direction errors associated (e.g. see Figure 11) at all their sites. Whilst this is conceptually reasonable it does have important practical implications, e.g. for pollutant dispersal (as highlighted by Norway).

Hungary stratified day 1 diurnal cycles in 2 m temperature forecast errors in 2017, for HRES and two LAMs, according to whether observations over Hungary showed mainly clear or mainly cloudy conditions (Figure 14). The two classes exhibit very different behaviour; for example, the clear sky cases show larger RMS errors in general, the diurnal cycles in bias are very different, and the mean bias on cloudy days is positive (forecasts too warm) for virtually every time step and every model. Even more illuminating for model development purposes might be "dual conditional verification", in which one examines days when forecasts as well as observations were clear or cloudy, which would eliminate cloud forecast errors themselves as the main contributors. Hungary does indeed touch on this, showing also that HRES has an underprediction bias for cloud cover on 'cloudy days'.

ECMWF has are strongly encouraged NMSs to explore and report upon (dual) conditional verification results in future years.

3.1.2 Post-processed products and end products delivered to users

Several countries have reported on the evaluation of post-processed products that they provide to users, or use internally for their forecasters. Post-processing can provide substantial improvements compared to the direct model output and can account for some of the known biases in the model data. For example, Romania, using a long-standing MOS-system, which reduces the RMSE in HRES forecasts of 2 m temperature maxima and minima by up to ~50%, and achieves a lead-time gain of ~five days (Figure 15). Meanwhile France has upgraded their multi-model MOS-MIX system this year, also for 2 m temperature, achieving, *relative to the old system*, an RMSE reduction of up to ~15%, and a lead time gain of ~6h (although day 1 improvements were much more substantial). Finland meanwhile is successfully correcting errors in HRES absolute humidity forecasts using a MOS system, which at the same time manages to retain relative humidity values within sensible bounds (≤100%) without incorporating copula techniques.

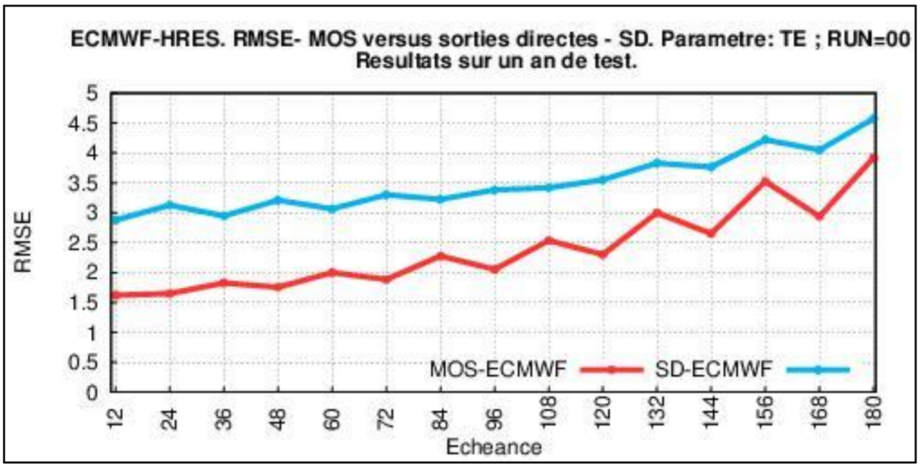


Figure 15: RMSE comparison between HRES forecasts (blue) and HRES-based MOS forecasts (red) for 2m temperature extrema for 1 year of data from 163 stations (Romania); x-axis shows lead time in h.

Israel illustrates how simple bias correction (based on performance in the last 1–3 weeks) can reduce HRES 2 m temperature errors by up to 70%. For their LAM, the improvement is about half what it is for HRES. Interestingly the actual RMSEs after post-processing are only ~0.45ºC, compared to ~1.3 and ~ 1.7ºC in France and Romania respectively (day 2). This probably attests to a more static (and sunny) climate for much of the year in Israel; it seems probable that simple bias correction based on the recent past would not work nearly so well where weather is more changeable, and where cloud prediction is more challenging.

With regard to ENS, Finland shows that it is possible to markedly improve under-dispersion in 2 m temperature forecasts, up to day-7, using calibration based on performance in the last 30 days - e.g. at day 3, after calibration, the average spread and RMSE values of 1 and 3ºC, become, respectively 2 and 2.5ºC.

The UK applies its own tropical cyclone tracker to the ECMWF, Met Office and NCEP ensembles. Verification for 2017 shows again that the ENS is the most skilful single-model ensemble, but that combining the three systems into a multi-model ensemble brings additional benefits (Figure 2).

The UK also provides several useful illustrations this year of verification of their regime-classification system applied to ENS (the regime list is given on Figure 3), and a limited comparison between this and equivalent results from their own GLOSEA extended range prediction system. Brier skill scores suggest that the predictive skill for different regimes varies quite a lot. A forecaster can potentially use this information, and it may also have application in model development. The best predicted regime seems to be "Anticyclonic, Azores high extension" (regime 6), whilst the worst was "Straight westerly" (regime 4). Perhaps the higher "predictability" of the former relates to innately strong persistence characteristics, whereas the lower predictability of the latter may be because it is vulnerable to distortion (into another regime) by cyclogenesis events. Additional data provided suggests, encouragingly, that biases in ENS for some regimes have lessened over the last eight years - for example whilst the frequency of cyclonic south-westerlies used to diminish with lead time, nowadays there is much less evidence of that (Figure 3, regime 7). ECMWF also invests in regime-related work, but in so doing has to cater for all the Member and Co-operating States, which limits the options regarding regime choices. The UK's choices (see Figure 3) would not be useful in Greece, for example.

3.2 Subjective verification

Synoptic studies

Several centres report on specific severe weather events that have affected them in the last year or so, one case by each of France (cyclonic windstorm), Greece (thunderstorms), Ireland (windstorm "Ophelia"), Israel (flash floods) and the UK (extreme cold spell), two cases by Montenegro (heatwave, broadscale floods) and Norway (windstorm, orographic rainstorm), and 11 cases by Italy (heavy rain: six, snowstorm: two, medicane: one, extreme cold: one, freezing rain: one).

Very brief summaries of some of these are given below but for more details please refer to the Member State/Co-operating State reports and to the UEF2018 meeting presentations and posters.

Greece shows how the IFS sometimes under-predicts cloud and convective outbreaks over adjacent seas, using one case in particular (15-16 Aug 2018), but indicating that there were others in summer 2018.

Israel discusses a tragic flash flood event in which ENS rainfall forecasts fell well short of observed (point) values. In following this up it has been shown that ecPoint output, which should be well-suited to this type of event, performed better (Figure 16, shared with Israel). Similarly, the fatal Livorno flash flood event of 10 September 2017, recorded by Italy, was discussed in the autumn 2017 ECMWF Newsletter from a point rainfall perspective.

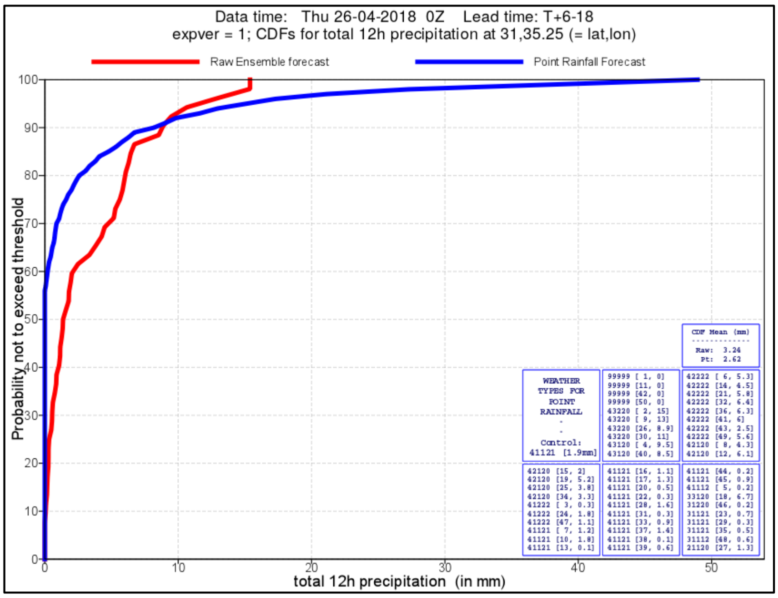


Figure 16: Test ecPoint point rainfall cumulative distribution product for 12h rainfall, for the site of a flash flood (6-18h forecast). The point rainfall (blue) shows a risk, albeit low, of 40mm or so being measured at a point. Raw ENS (red) shows a much lower upper limit.

The broadscale floods in Montenegro were well forecast by HRES, albeit with some under-prediction, apparently because orographic enhancement was under-done (a known model limitation). Comparing the LAM and HRES plots in their report reminds us that the rain shadow effect is also under-done by HRES; over Lake Skadar, downwind of a narrow orographic barrier, ~40mm was forecast, when reality appears to have been ~4mm. The barrier is ~350 m high in HRES, but ~1200 m high in reality.

Norway illustrate nicely how HRES and ENS handled a major orographic rainstorm in the more populated south of the country, in October 2017. The EFI/SOT signal was reassuringly consistent in the lead up. At shorter ranges HRES forecasts were good but underplayed somewhat rainfall totals near windward coasts, and over-predicted at most sites well inland. Again, as in the Montenegro case, a complex interplay between under-represented enhancement and under-represented rain shadow may be at work here.

Referring to the severe wintry outbreak centred on the end of February 2018 (see also Section 2.1.3 above), the UK indicates that they were generally happy with ECMWF guidance over one week in advance regarding the onset of this event, and also with the shorter-range indications of heavy frontal snowfall, and a major freezing rain event (which is extremely rare in lowland UK). Snow events in Rome and Naples cited by Italy related to the same broadscale cold outbreak. These were not handled so well, due to the known model parametrisation issue, of non-advecting marine convection.

Cyclonic windstorm ‘Ana’ on 11 December 2017, cited by France, nicely illustrates the exacting requirements of forecasting operations today, and expectation levels for ECMWF output accuracy.

Damaging strong winds were slightly misplaced in a HRES forecast at 18-24h lead times. However, if one compares the 12h mean sea level pressure field from that “poor” HRES forecast, with the EDA (Ensemble of Data Assimilation) realisations of the *analysis* for the same valid time, it falls within their range. This suggests that the “requirements” for HRES accuracy exceed the level of confidence we can have in even the model analysis details over the ocean. And for cyclonic windstorms, which due to dynamical instabilities are intrinsically a major challenge for a forecasting system, it is not that unusual anyway for short range forecasts to exhibit large errors. A case in point is the forecast for windstorm Johanne (10 August 2018), highlighted by email by The Netherlands. Mean sea level pressure comparisons for both cases are shown in

Figure 17. ECMWF can in time expect to improve upon developmental timing errors such as we saw for Johanne (right panel, Denmark did experience a major windstorm after this time), but without improvements in observational coverage, errors that are more directly related to surface analysis uncertainty, such as we saw for Ana (left panel), seem likely to continue.

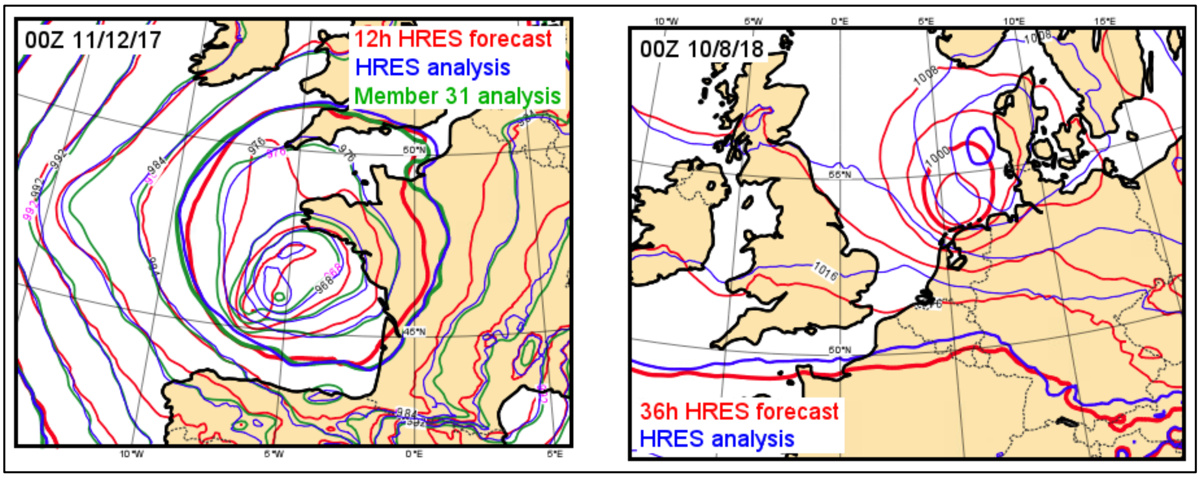


Figure 17: Two cyclonic windstorms, as represented in short range IFS forecasts and analyses, “Ana” (left) and “Johanne” (right) (also known as “Yves” and “Oriana” respectively). For Ana note how the 12h HRES forecast and member 31 analysis are similar, suggesting very large uncertainty in the analysis.

4. Longer Range Forecasts

Many Member and Co-operating States continue to make operational use of monthly (=extended range) and/or seasonal (=long range) forecasts whilst Finland has been trialling major new initiatives in these areas. There is evidence that expectations, regarding accuracy and an absence of jumpiness, are too high.

4.1 Extended Range (Monthly)

The UK shows UK-regime-based verification metrics for ENS, which, encouragingly, show minimal drift in frequencies of almost all the different weather types, over the years 2015 to 2018 (Figure 18). The primary exception is cyclonic south-easterly types, for which there is an upward trend with lead time such that in weeks 3 and 4 these occur ~twice as often as on day 0. The characteristics of GLOSEA, the Met Office’s in house monthly (and seasonal) forecast system, are shown in the UK report for comparison. Its regime frequencies are also quite stable; the cyclonic south-easterly seems to be better modelled than in ENS, but on the other hand the frequency of cyclonic south-westerly types, the second most populated class (after anticyclonic west-south-westerly), is underdone by about 50% in weeks 3 and 4.

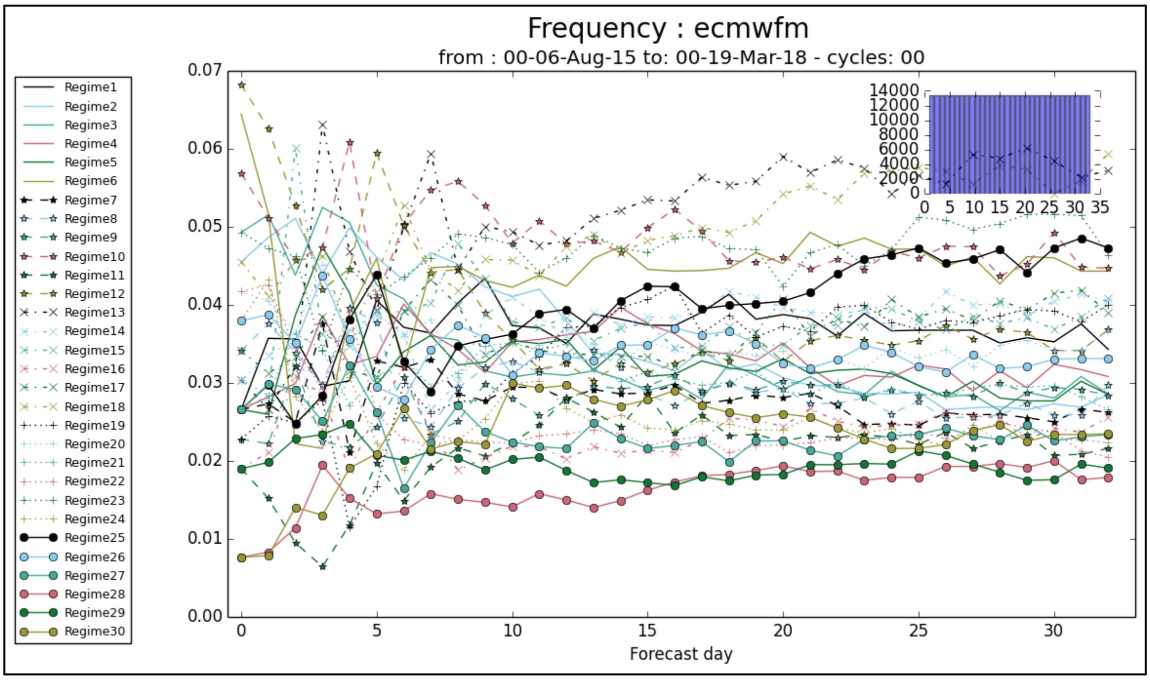


Figure 18: ENS weather pattern forecast frequencies for all 00UTC ECMWF monthly forecast runs between 6th August 2016 and 19th March 2018. Day 0 frequencies are observed values for each pattern over the verification period. Regimes are described on Figure 3.

Hungary has just begun verifying their monthly forecasts, based on ENS. Using a day-by-day approach, they show that in 2017 skill for 2 m temperature forecasts asymptotes to zero around day 13. Vestiges of skill after this date would probably show up more clearly in a week-by-week approach. Skill for minima is less than half as good as it is for maxima, which is the type of difference one would expect from energy considerations.

France reports on 12 years of categorised subjective verification of 2 m temperature forecasts for France. Week 1 forecasts have clearly improved, but for the other weeks forecast quality seems to have stayed the same or deteriorated slightly.

Finland has embarked on an initiative to issue six-week “Climate impact outlooks” year round, for the public, which involves a lot of user-oriented tailoring, such as providing beach forecasts in summer. Figure 19 shows preliminary ENS-based verification results, based on 20 years of reforecasts, for weekly mean temperature. Surprising here is the reduced skill over the Baltic, compared to land; one would expect the opposite. This may relate to issues with using ERA-Interim as truth. Skill disappears over most land areas during week 3. This broadly concurs with ECMWF’s in-house verification (albeit using different metrics). At the UEF meeting there were requests to include forecasts for weeks 5 and 6 as standard web products. Extensive white areas on Figure 19 suggest that that may not be very helpful, although users should note that these forecasts are now available in ecCharts.

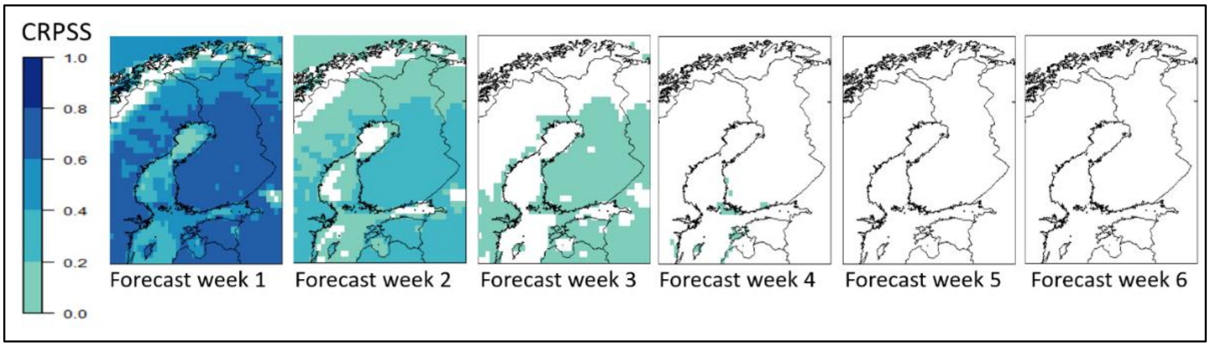


Figure 19: Preliminary results for verification of weekly mean 2m temperature for each grid point, showing CRPSS of the reforecasts for years 1997-2016 using ERA-Interim climatology of 1996-2016 as the reference.

Croatia expresses some dissatisfaction with jumpiness in the monthly forecast, suggesting also that too much “momentum” (i.e. signal retention) can be carried forward between consecutive forecasts (until, presumably, there is a jump). Meanwhile Italy highlights consecutive week two forecasts in late August 2017 that showed strong positive precipitation anomalies in northern Italy (and countries further north) that did not verify.

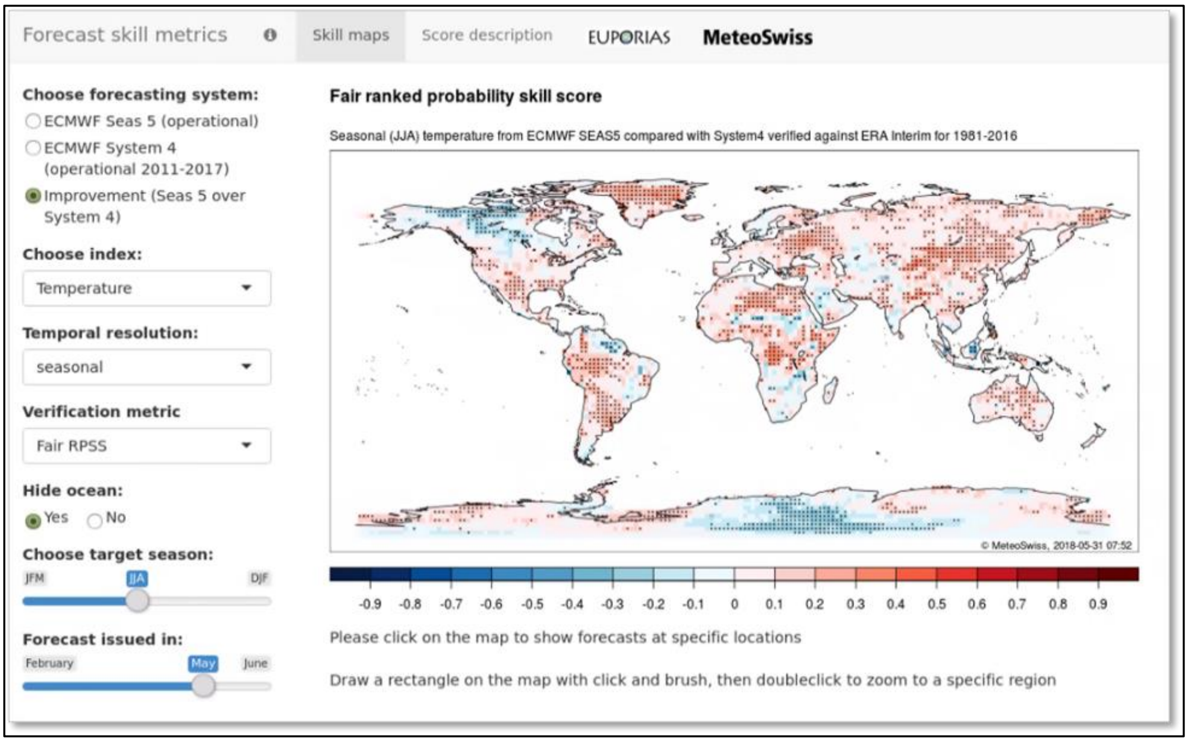
4.2 Long Range (Seasonal)

ECMWF formally introduced a new seasonal forecast system, System 5 (or SEAS5), in November 2017, to replace System 4. The new system has many new features, such as much increased resolution, and use of a sea ice model. Many NMS reports reference the new system.

Switzerland advertises their own publicly-accessible verification website, which facilitates examination and comparison of skill metrics, for the world, for Systems 4 and 5. Feedback on the site was described as “overwhelmingly positive”. The example in Figure 20 illustrates higher skill over most of the world in SEAS5 for JJA forecasts.

Feedback from other NMSs seems to express or imply some disappointment with the seasonal forecasts. Perhaps this is because demand and need are high, which maybe fuels unrealistic expectations. Statistics presented in NMS reports sometimes relate to four forecasts or less; although those particular forecasts are naturally important for many users this is not a sample from which much can be concluded.

Spain has expressed disappointment with winter-time forecasts for the North Atlantic Oscillation. Meanwhile the Director of the Moroccan met service highlighted to ECMWF representatives issues with seasonal rainfall forecasts in a wet year, and Israel shows that SEAS5 DJF rainfall forecasts were worse than those from System 4 for their region. Both Israel and Croatia view seasonal rainfall forecasts as little better than random chance. And in a month-by-month breakdown for 2017 Hungary shows basically no precipitation forecast skill for any month, including month 1.



Choose forecasting system:
○ ECMWF Seas 5 (operational)
○ ECMWF System 4 (operational 2011-2017)
● Improvement (Seas 5 over System 4)

Choose index: Temperature

Temporal resolution: seasonal

Verification metric: Fair RPSS

Hide ocean: ● Yes ○ No

Choose target season: JFM JJA DJF

Forecast issued in: February May June

Fair ranked probability skill score

Seasonal (JJA) temperature from ECMWF SEAS5 compared with System4 verified against ERA Interim for 1981-2016

© MeteoSwiss, 2018-05-31 07:52

Please click on the map to show forecasts at specific locations

Draw a rectangle on the map with click and brush, then doubleclick to zoom to a specific region

Figure 20: Snapshot from the Swiss seasonal forecast website showing improvements in forecast quality, using the fair ranked probability skill score, for ECMWF SEAS5 versus ECMWF System 4, for summer (JJA) temperature forecasts issued in May. Red shows where SEAS5 outperforms System4; stippling denotes significance at the 5% level.

Israel also complains about the reference period for temperature forecasts, and illustrates with the plot on Figure 21, stating that “due to climate change, a forecast relative to the 1981–2010 reference period is useless”. Although one should also not assume that cooler summers in the future are impossible ECMWF is aware of the substantial upward trend in temperatures in this part of the world, and encourages NMSs to conduct any local “climate-change-related” adjustments deemed necessary, to seasonal forecasts, by using seasonal re-forecast fields downloaded from MARS. Tercile probabilities, or other metrics, can then be re-computed from fields adjusted in this way. ECMWF does not have resources to provide a service of this type, although we are happy to offer advice. In a similar vein Croatia states that a rudimentary “always-warmer-than-average” forecast for their region scores better than the actual forecasts. But if one ignores this aspect the temperature forecasts themselves do at least exhibit some skill. Hungary meanwhile shows that for temperature forecasts for month 1 there is clearly skill, whilst in months 2–6 skill is only slightly positive and much the same. Finland has just begun using seasonal forecasts, and in four seasons verified two forecasts were good, two bad.

Users are reminded that most of the impressions above concur with SEAS5 skill metrics readily available on the ECMWF website; for Europe skill levels tend to be zero or only slightly positive. One should also recognise that seasonal forecasts tend to be over-confident. At the UEF meeting jumps in successive seasonal forecast were reported, notably between months 2 and 1. Again, whilst this is unfortunate from a user perspective, it is nonetheless what one should expect to happen quite often within a system, such as ours, that delivers little skill in month 2, but some in month 1.

Users are also encouraged to reference guidelines for seasonal forecast product interpretation within the new Forecast User Guide.

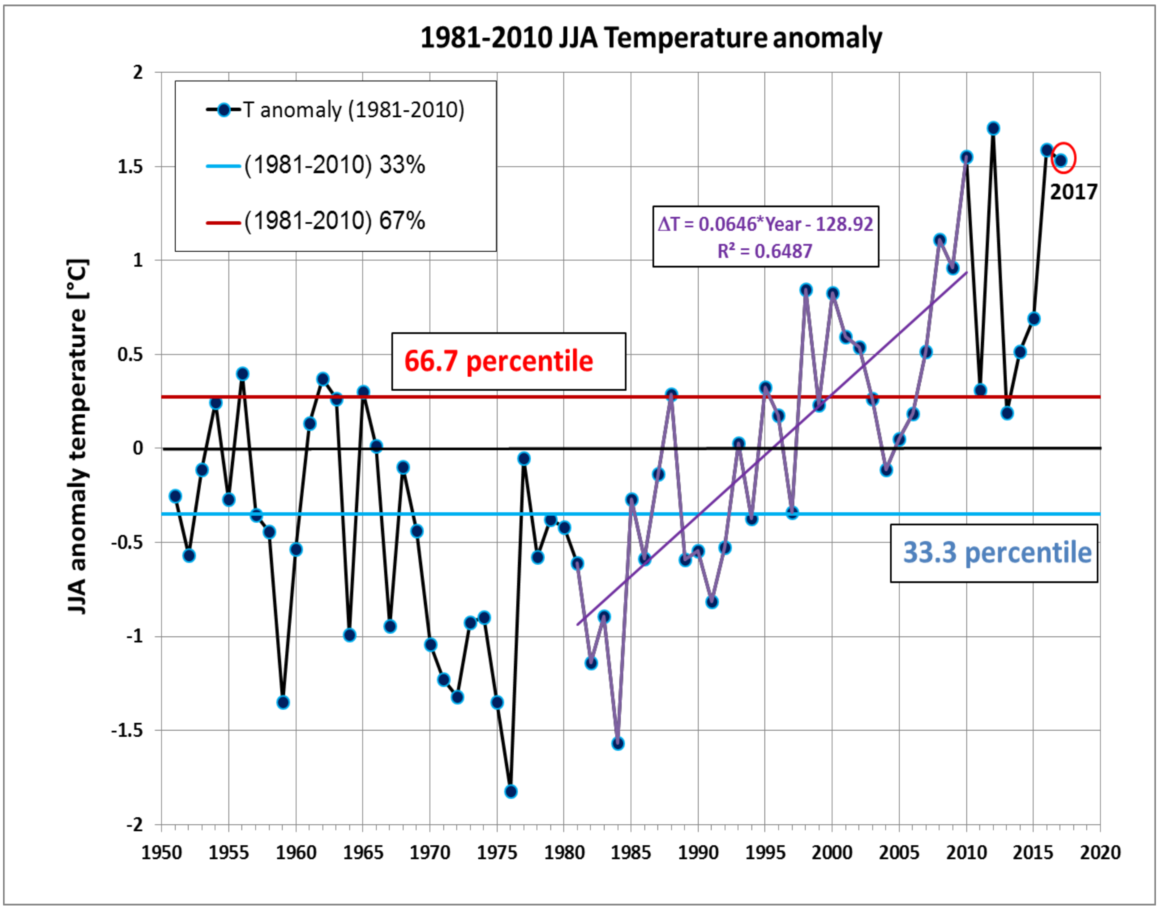


Figure 21: JJA average temperature anomalies for Israel since 1951. The horizontal lines represent the upper and lower tercile thresholds for the 1981-2010 reference periods.

Latvia report that they are using the 12-month seasonal forecast output of SEAS5 (available 4 times a year) to drive a local hydrological model.

5. Interactive feedback facility for forecast users

In June 2014 ECMWF introduced a new online “Forecast User” portal to provide assistance to its forecast users, and a mechanism for interacting with them. Historically the portal contained two main sections, one about known forecasting issues and one about severe weather events. In May 2018, a new item was created - a comprehensive and extensively updated web-based version of the Forecast User Guide. All three sections are now being regularly updated through the year. As in previous years, we have asked for feedback on whether countries were aware of these various facilities, and how useful they are. The overall response rate of ~50% was better than last year.

Table 1: Member and co-operating state responses to questions about ECMWF web-based facilities for forecasters, with last year's numbers in parentheses

Facility	No response	"Didn't know about this"	"Not much used"	"Useful"	"Very useful"
"Known IFS forecast issues"	(13) 11	(0) 2	(2) 1	(2) 4	(4) 4
"Severe event catalogue"	(13) 11	(0) 2	(5) 3	(0) 4	(3) 2
Forecast User Guide	13	1	2	4	2

Users generally made positive comments about these various initiatives, even if they were not used much. One would instinctively expect the "known issues" to be used a bit more often than the "severe event catalogue" which is more specialised, as the reports reveal, albeit with a small sample. Italy commented that the pages were hard to find from the main website, which we will look into. Given the relatively recent release date for the Forecast User Guide this facility probably needs more time to gain traction within the forecaster community, although Norway have already described two sections in it as 'excellent'. We will also investigate Norway's suggestion to include in this a section for 'recently added products'. The UK suggested that a paper version of the Forecast User Guide could be useful; this should be possible but could involve a non-trivial amount of work as some page items do not convert easily.

References

NMS Reports that contributed to this technical memorandum:
https://www.ecmwf.int/en/publications/search/?secondary_title=%22Green%20Book%202018%22

Presentations and Posters from the 2018 UEF meeting:
https://www.ecmwf.int/en/learning/workshops/using-ecmwfs-forecasts-uef2018

ECMWF Technical Memoranda:
https://www.ecmwf.int/en/publications/search?solrsort=ts_biblio_year%20desc&f%5B0%5D=sm_biblio_type%3ATechnical%20memorandum

ECMWF Newsletters:
https://www.ecmwf.int/en/publications/search?solrsort=ds_biblio_date%20desc&f%5B0%5D=sm_biblio_type%3ANewsletter

ECMWF's online Forecast User Guide:
https://confluence.ecmwf.int/display/FUG/Forecast+User+Guide

The ECMWF "Forecast User" portal:
https://confluence.ecmwf.int/display/FCST/Forecast+User+Home

Known IFS forecast issues:
https://confluence.ecmwf.int/display/FCST/Known+IFS+forecasting+issues