TECHNICAL MEMORANDUM

818

Use and verification of ECMWF products in Member and Co-operating States (2017)

D.S. Richardson & T. Hewson

Forecast Department

January 2018

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 under:
http://www.ecmwf.int/en/research/publications

Contact: library@ecmwf.int

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 now available on the ECMWF website.

A summary of the NMS reports is presented below. Content has been combined with (i) feedback from the "Using ECMWF Forecasts" (UEF2017) workshop held at ECMWF from 12–16 June 2017 (primarily from breakout groups), and (ii) feedback from official triennial Member State/Co-operating State visits undertaken by ECMWF between July 2016 and June 2017. In chronological order these were to: Iceland, Austria, Germany, Denmark, Finland, Estonia, Ireland, Montenegro, Serbia, UK, Turkey, and Switzerland.

Reports received are available on the ECMWF website at:

https://www.ecmwf.int/search/elibrary/Green?solrsort=sort_label%20asc&year=2017&secondary_title=Green%20book%202017

Reports came from 22 states: Belgium, Czech Republic, Denmark, Finland, France, Greece, Hungary, Iceland, Italy, Latvia, Lithuania, Norway, Portugal, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey and the United Kingdom.

Contributions were 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

The ECMWF IFS undergoes improvements each year, which naturally affect aspects of performance, so summary information presented here has to be read with this in mind. The results of ECMWF's own objective verification together with details of recent model upgrades are presented in a separate document (Haiden et al, 2017).

2. Use and application of products

In the shorter ranges (typically to 48–72h ahead), ECMWF IFS products are commonly used in conjunction with products from other sources, notably limited-area deterministic systems, but to increasing extent limited area ensemble systems too. In most cases ECMWF IFS data provides boundary conditions (BCs) for these limited-area runs, four times per day.

In the medium, extended and longer ranges, ECMWF products are generally the main or the only output used by Member States/Co-operating States. It seems that all Member States/Co-operating States provide forecasts of some sort up to lead times of 10 days. When looking beyond, to longer and longer lead times, the number of organisations with formal forecast commitments slowly diminishes but it is clear again this year that this is a growth area in many countries. Growth is commonly being driven by requests from customers such as energy suppliers.

The ECMWF resolution upgrade in 2016 allows for a better identification of some weather details that were traditionally the remit of LAMs, and a number of countries report that this has led to changes in their working practices. For example Finland now use the HRES in their aviation forecasts, when previously only HIRLAM was used; while Norway has retired its 8 and 10 km resolution HIRLAM and now ECMWF data for all applications needing NWP output outside of the local (Scandinavian) domain. Switzerland also plan to phase out their COSMO-7 model, replacing with direct use of HRES. These changes allow the Member and Co-Operating States to focus more on modelling at higher resolution, with 1-2km LAM configurations becoming increasingly used.

Whilst the main operational focus for all Member and Co-operating states is on local weather, many also have international commitments, for which ECMWF forecasts are very regularly used. Norway for example has undertaken capacity-building activities in three countries in Southeast Asia, and for this HRES and ENS data are being used.

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. This is borne out by increasing efforts to improve calibration, as reported for example by Finland, France, Iceland and Romania this year. Hungary reported new objective verification results this year, showing that calibration can improve of ENS precipitation forecasts for extreme events.

The main post-processing methods reported this year are Model Output Statistics (MOS) and Kalman filtering, while Perfect Prog (PP) seems to be used less that previously. Statistical techniques include height mismatch adjustment and bias removal. Other more advanced techniques include, multiple linear regression, polynomial regression, logistic regression, linear discriminant analysis, quantile matching and non-homogeneous Gaussian regression. Indeed many countries report using a mixture of approaches. Some of these techniques are applied to individual ENS members as well as to HRES, while additional approaches such as calibration based on rank histograms are applied to the ENS distribution. Sweden report better results from calibration of the ensemble median rather than mean for clouds.

Most calibration techniques are 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 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). The UK also reports on the benefits of multi-model ensembles, using weighted combinations from ENS and the Met Office ensemble to improve tropical cyclones forecast and severe weather.

ECMWF reforecasts are increasingly seen as being a very convenient and important dataset for calibration activity, as noted in Norway's report. Indeed there are regular requests at the UEF meetings to provide a full year's worth with each new cycle. However, such an initiative would delay implementations and require substantial computer resources that are not available at the current time.

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 Programme "Boundary Conditions for Limited Area Modelling" (BC Optional Programme), which provides additional forecasts from 06 and 18Z data times. The limited-area model suites most widely used in this way are ALADIN, AROME, HIRLAM, HARMONIE, COSMO and WRF.

As well as HRES BCs, ENS BCs are also available and are steadily being adopted (e.g. Hungary, Italy, Sweden, Switzerland). In November 2016, hourly output fields were introduced for ENS for use in the BC programme.

The improvements associated with the ECMWF horizontal resolution increase (March 2016, cycle 41r2) led to Norway retiring its 8 and 10km resolution HIRLAM in January 2017, using ECMWF data for all applications needing NWP output outside of the local (Scandinavian) domain. Switzerland also plan to phase out their COSMO-7 model, replacing with direct use of HRES; already their high-resolution deterministic COSMO-1 and COSMO-E ensemble are driven directly by ECMWF boundary conditions.

ECMWF's model suite also provides ICs and BCs for wave modelling in several countries. France's internationally competitive wave model is driven directly by IFS winds (and IFS sea-ice since April 2017); the French storm surge model is also driven by IFS as well as by Arpege and Arome. Sweden use ECMWF fields as upper boundary conditions for their 10-day NEMO ocean forecasts.

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 pollutant modelling. ECMWF are also used as back-up boundary conditions for LAM usually driven by other models.

The UK report an additional application of ECMWF initial and boundary conditions to initialise Unified Model LAM configurations in the tropics to investigate the impact of the driving model on performance at the km-scale.

HRES and ENS fields continue to also be used or adapted to drive trajectory and dispersion models and hydrological models in numerous countries. In several countries there are application models of other types (e.g. road state and oil spill 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 also tailored versions of those products that ECMWF has been providing.

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 the ecCharts platform, and by fulfilling specific requests for new products whenever possible. For example, last year Hungary demonstrated a "precipitation type probability diagram" using the new precipitation type diagnostic from the ENS. This was well received by other users, and ECMWF has now implemented a version of the precipitation type probability in ecCharts so that all users have easy access to this type of information. Norway produce their own lightning parameters and looks forward to testing ECMWF new lightning parameter which will be available in the next model cycle upgrade. Hungary has followed up its innovative precipitation type probability diagram with a similar graphical product for visibility, which received positive feedback from forecasters during last winter.

Various derived products are generated across the Member and Co-operating States for use by their forecasters, for the public, and for specific societal or economic applications. Many countries have reported a continuation of products they cited in previous years' reports.

For forecasters, a key challenge is to provide information about severe convective storms and to identify the favourable environment for development of such storms in the coming days. Convection-related indices generated in Member States and Co-operating States for forecasters include Jefferson, Lifted, SWEAT, Showalter, KO, wind shear and helicity. Aviation is another activity for which countries compute specific derived products. For more general forecasting, several countries compute their own clustering of the ENS, and generate ensemble mean, spread, and probabilities for a variety of parameters and event thresholds.

For the public, a typical application is to generate "characteristic weather" symbols for web or for mobile phone "Apps". For example, as part of the re-engineering of the product suite, France has developed a new algorithm to generate weather pictograms for anywhere in the world based on the ECMWF forecasts (Figure 1). For its yr.no web application, Norway stress the importance of the ECMWF reforecasts. These are needed to provide a reliable model climate that is essential to achieve a smooth transition from short-range forecast (based on AROME-Arctic LAM) to the medium-range which uses the ENS.

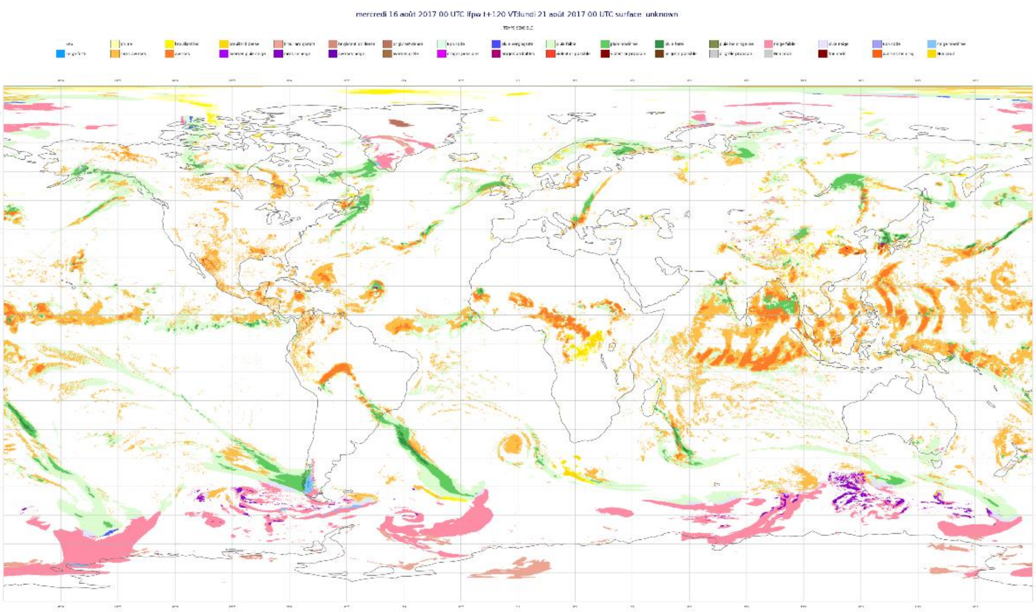


Figure 1: Example from France of weather parameters used for pictogram generation over the globe using ECMWF forecasts.

2.2. Direct use of ECMWF products, including severe weather prediction

ECMWF products are primarily used from day 3, up to around 2 weeks ahead. At shorter ranges, most services also refer to ECMWF products to supplement their main LAM models. As discussed above indirect use for short ranges, via provision of BCs for LAMs, is commonplace. Use of and demand for forecasts for extended and seasonal ranges continues to grow (section 4 below).

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 (including the LAMs, observations, satellite data and nowcasts). Several workstation applications allow some editing by forecasters of the model fields, who may for example apply some weighting to the HRES and ENS components. Others facilitate some degree of processing, for example FMI workstations allow short scripts to be written to generate derived parameters (wind chill, probability of thunder).

It is not always as easy to ingest ENS products into the forecaster workstation systems, and these are sometimes displayed on a local intranet, or used directly on the ECMWF web pages (or increasingly via ecCharts, that also provide a WMS service to facilitate the transfer of ECMWF data into the workstations). Data volumes can be an issue, and use of ENS can be limited by bandwidth or disk space requirements. For example Switzerland only import post-processed ENS products into their NinJo system, and not individual members. Germany note that the use of ensemble products has increased greatly since they have been included in Ninjo. Another application that has been increasing in recent years is to input the ECMWF forecast output data into digital databases (e.g. Belgium ModelBestGrid, UK best data, AEMET Digital Forecast Database) that are used to produce a range of automated products (including text and pictograms) for website or apps. Some of these use ENS data and provide probability information to the automated products.

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 humanitarian or capacity building activities (e.g. Norway is using HRES and ENS in SE Asia), or for general public forecasts (e.g. yr.no).

Severe weather

IFS forecasts are very widely used for official severe weather warnings and for alerting forecasters to potential severe events at longer range. The longest lead time at which a warning will be issued varies by country, e.g. for Italy it is only about 24h, but for other countries can be up to 5 days, and so as with everyday forecasts the extent to which ECMWF output is actively used also varies. Nowcasting tools and LAMs are used more at short leads, although for example Denmark use ECMWF output as a "second opinion" for warnings in the first few days.

Warning systems are becoming more probabilistic in nature, 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 now very well established, and the more recently introduced EFIs for severe convective events have been found useful additions by Member States. There also appears to be a gradual evolution from threshold-based warnings to ones that are more impact-based. This focus on impacts relates closely to the return period philosophy that underpins the ECMWF EFI and SOT products. Norway report how using EFI and SOT together can be useful in giving indications of potential severe events well into the medium range. Finland and the UK have also highlighted how the EFI can be especially useful beyond the local region, when forecasters' knowledge of local climate is lacking.

Other ENS products, such as 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 several countries report that the ENS is tailored accordingly, via ecCharts or other mechanisms. Denmark, Norway and the UK noted how the extra-tropical cyclone tracking product can assist in the warning process, by highlighting uncertainties in the behaviour of cyclones that are responsible for severe weather.

Tropical cyclone-tracking products are used by some Member and Co-operating States, due to certain NMSs having specific responsibilities in tropical regions (notably France, from La Réunion).

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 and Norway), and also flooding related to ice break-up in spring (Latvia) which is not addressed by the EFAS setup at present. For these more diverse categories IFS data is also used, but only indirectly.

ecCharts and other web products

ecCharts is actively used in in several countries, and these users greatly appreciated its functionalities. In other countries the use of ecCharts is limited or nil. Reasons for this varied. In some countries (e.g. Hungary) ECMWF data is ingested into advanced local forecaster workstations, which provide all the tools needed, as well as the additional facility to overlay other models and observational data (note also that ecCharts' WMS capabilities permit product transfer into such a platform). More concerning was the fact that other countries had issues with ecCharts. Complaints about slow speed continue. We have continued to work hard to address these concerns, for example by proposing ways to tackle local IT/firewall issues, which can sometimes explain the lack of speed in a particular country. Recent optimisation efforts have improved the speed of some products such as the ENS meteograms and “spaghetti” maps. Training can also be important to benefit fully from the ecCharts capabilities and ECMWF provided special training sessions during some of their visits to the national meteorological services.

In response to user requests, ECMWF continues to add parameters, display schemes and extra functionality, on a regular basis (key updates being in early summer and early winter).

The standard ECMWF web charts are also still widely used, especially for the ENS products, with EFI, SOT and the meteograms being the most frequently used products. To support these users, the web chart browser has been improved, and more interactions were added to most of the medium range maps allowing users to get access to the ECMWF points based products such as ENSgrams, and CDF/EFI graphs with a single click on the map.

2.3. User Requests

Products

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

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 UEF2017 meeting presentations and posters at: http://www.ecmwf.int/en/learning/workshops/using-ecmwfs-forecasts-uef2017.

- More model output fields in archive/dissemination/catalogue (e.g. stratospheric levels from ENS);
- more reforecast-based M-climate output products, e.g. relevant quantiles (10,25,50,75,90) for EFI/SOT
- Higher temporal resolution: e.g. for EFI/SOT and M-climate, 3 hourly timesteps in meteograms, probabilities for 6h precip
- More parameters and additional time periods for the monthly and seasonal forecasts, including EFI and related products
- Additional convection indices (long list!), other indicators for severe convective storms, including lightning products
- forecast vertical profiles (tephigrams or skew-T plots)
- More aviation parameters (e.g. icing/turbulence indices, TAF-related parameters)
- ENS wind level between 10m and 100m for offshore wind power

- Additional products for tropical waves (MJO, Kelvin waves), and for tropical cyclone structure (TC phase space products)

Other

As regards more general requests relating to ECMWF operating policy, the recurring themes this year were similar to last year. These include hourly outputs for many applications (e.g. hydrology, renewable energy), access to 06 and 18 UTC forecasts for all users, (even) earlier delivery of data. Some users want a full year of reforecasts to be made available with each new cycle to facilitate their internal forecast calibration processes. Mostly, responses to these requests will depend both on the recommendations of the ECMWF Advisory Committees, and on computational resources.

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 41r2, with its resolution upgrade, became operational 8 Mar 2016

Cycle 43r1 became operational 22 Nov 2016

Cycle 43r3 became operational 11 July 2017

This means that in reports submitted this year (2017) verification correspond mainly to periods since the introduction of the higher resolution HRES and ENS in March 2016, and to cycles 41r2 and 43r1. It is reasonable therefore to expect to see some improvements, overall, in performance relative to last year (when only the first few months of higher resolution forecasts were available). However note that, 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).

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

Overview

Biases and errors in sensible weather parameters can depend strongly on the prevailing weather types, which differ each year. When considered alongside the (fixed) geographical differences between regions, and the different impacts that those can have in a given synoptic pattern, it is no surprise to sometimes see different findings from different countries.

Another issue, affecting inter-comparability between different countries' results for HRES, and indeed comparisons between HRES and LAM forecasts, is the range of "interpolation" and "site-selection" techniques that can be used. Sometimes, for example, full resolution IFS output is not being exploited. Moreover in some reports received the method(s) of extraction and interpolation used were not entirely clear.

Many reports focussed 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.

Overall, HRES seems to perform competitively with respect to most LAMs. Verification results in this year's reports correspond mainly to periods since the resolution increase in March 2016, and several countries note recent improvements in the verification results for HRES compared to the LAM systems. For example Denmark report recent improvement of HRES for surface parameters where LAM have historically had the advantage, due to the improved resolution (and physics) of HRES. Belgium performed a detailed evaluation of key surface parameters comparing HRES with their 7km resolution Alaro LAM, which provides the basis for the ModelBestGrid automated forecast up to 48 hours, specifically to review whether the Alaro is still the optimal choice for this. They conclude that HRES is the more accurate model for all Belgian stations, except for 10m wind speed where the best model varied with station. They also show that there was little impact from either the resolution of the output from HRES (retrieved at the new 9km resolution or the previous 16 km resolution), or the resolution of the Alaro forecast (run at 7km or 4km resolution).

The results from Belgium on 10m wind are reflected more widely in the results from other countries. This seems to be the parameter where resolution is still most noticeable and several countries show that LAMs, especially at the higher 1-2km resolution, provides significantly better results than HRES. 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, and it seems that in mountainous regions LAMs do, in relative terms, tend to perform a bit better than HRES (again this is most significant for the LAMS running at the 1-2km scale).

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, though it is undoubtedly used in subjective fashion by forecasters.

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 'Forecasting Issues' web page at: https://software.ecmwf.int/wiki/display/FCST/Known+IFS+forecasting+issues which has been updated a number of times in the last year.

There is as usual less verification reported for ENS. However, Denmark report the verification of ENS forecasts is used as a reference (benchmark) for the development of their COMEPS limited area ensemble system. Switzerland compare probabilistic scores for ENS and the COSMO-E ensemble.

2 m temperatures

Combining the results from all reports one can conclude that for 2 m temperature HRES performs slightly better than LAMs. By better we mean smaller MAE and RMS errors, and these are usually accompanied by smaller biases. However, better representation of complex terrain (mountains, lakes, small islands) undoubtedly contributes to the relatively high skill of high resolution LAMS in some regions.

Where reference is made to equivalent statistics from previous years, several countries noted improved scores for temperature since last year, with for example Greece, Iceland (Figure 2) and Norway each reporting improvement in winter temperature forecasts compared to previous years, consistent with the expected improvements from IFS cycle 41r2 (resolution increase) and 43r1 (e.g. modified surface coupling for 2m temperature). Finland report a continuing problem of missing inversions in stable radiation conditions in winter, with consequent large positive biases in HRES, but note that overall winter temperature is well forecast.

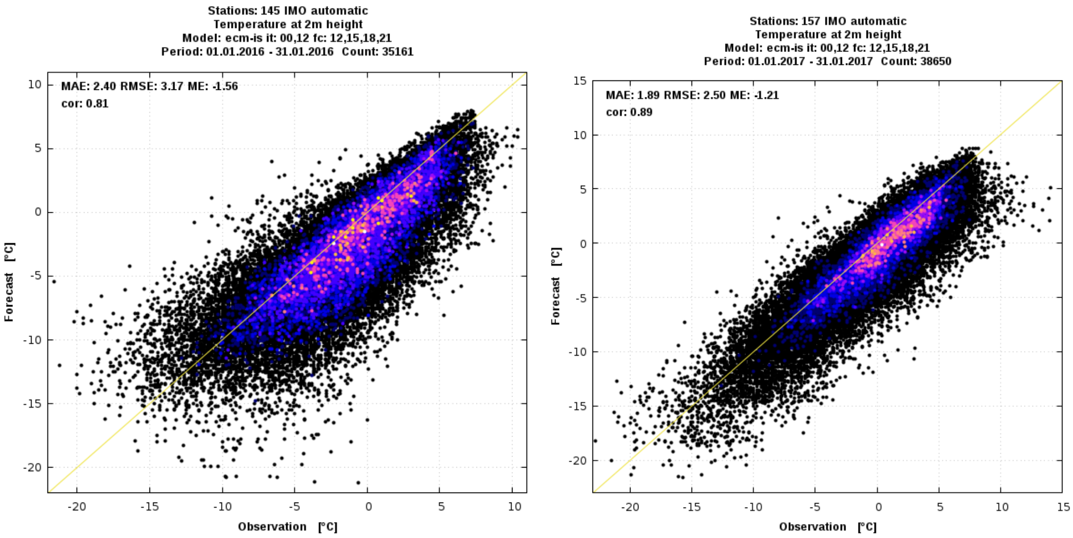


Figure 2: Scatter plot of 2-metre temperature (°C) forecasts and observations for forecasts length 12, 15, 18 and 21 hours, initialized at 00 UTC and 12 UTC for January 2016 (left) and 2017 (right). ECMWF HRES.

Finland report some improvement in the spring temperature bias (forecast temperatures being much too low in the evenings), although this is not completely eliminated. This can still cause problems for hydrological applications during the spring snow melt, but Finland note that their MOS calibration of ECMWF data performs much better in these situations, demonstrating the benefit of appropriate calibration for end users. ECMWF continues to work towards improved performance.

10 m wind

Combining results from different reports, as for temperature, it seems that on balance HRES is currently providing mean wind speed forecasts that are as good as those from LAMs. However, model resolution seems to be a significant factor, with both Iceland and France showing that the highest resolution LAMs (2.5km) perform better than HRES, while HRES compares well to the lower resolution LAMs (at around 5km resolution).

Denmark, Iceland and Norway continue to highlight a marked under-prediction bias in HRES winds in complex mountainous terrain, especially in winter. Their 2.5 km HARMONIE model continues to perform much better than HRES, as reported last year. Iceland and Norway also comment and illustrate that performance is much more comparable at coastal sites, and overall across Norway. Denmark also comment how their COMEPS similarly outperforms the ENS because of its much higher horizontal resolution (2.5km HARMONIE-AROME grid), as shown by CRPS results.

Precipitation

For precipitation forecasts many different skill metrics have been quoted, including, this year, SEDI (Symmetric Extremal Dependence Index), FSS (the Fractions Skill Score) and SEEPS (Stable Equitable Error in Probability Space). This wide range, the sometimes disaggregated nature of precipitation fields, and the different interpolation methods used all together make result intercomparison rather more difficult than it is for wind or temperature.

Firstly it should be noted that all countries compare forecast totals, which innately apply to a gridbox, with point observations. Reported results will therefore be affected to some extent by the representativeness issue caused by the inherent difference between grid box (area) totals and point values. For example, many countries quote the frequency bias index, or FBI, for point measurements. For HRES plots are fairly consistent between countries. Small totals tend to be overestimated, and large totals underestimated, which is a virtually inevitable consequence of a point versus gridbox approach, so this is not necessarily a model problem. For LAMs FBI profiles are almost always more horizontal, and relatively close to 1, which is also what one expects, for smaller gridboxes. For HRES the "crossover" FBI=1 is quoted as around 10 mm/24h. It should also be re-iterated here that ECMWF is currently investigating ways of converting forecast gridbox totals into point total pdfs (probability density functions); a procedure which can take account of both sub-grid variability and different types of bias, and that in the future new products may result from this work.

Norway report that HRES compares well to AROME-MetCoOp for moderate precipitation, and that the underestimation of high precipitation amounts is reduced since the previous year. This is consistent with the expected improvements from the IFS resolution upgrade, but results for just one season may also be sensitive to atmospheric variability so it will be interesting to see if this change is confirmed over more cases. Portugal also report similar performance between AROME and HRES for 24-hour accumulations, though the higher resolution AROME performed better for shorter 3-hour accumulations (Sweden also noted 2.5 km AROME was better for 3-hour accumulations). However, other countries also report on the benefits seen in the higher resolution LAMs for precipitation especially in mountain areas (Czech Republic, Switzerland).

Switzerland show a comparison of reliability diagrams for the ENS and COSMO-E ensemble probability forecasts for more than 1mm/12h precipitation at Swiss stations. The ENS is generally overconfident for this event, while COSMO-E probabilities are too low for values up to 50%. It should be noted that the overconfidence of the ENS is over-emphasized because the processing of the ENS involves averaging over grid boxes. It is nevertheless welcome to see some ENS results included in the model comparison results.

Cloud and screen-level humidity

We have fewer reports in these categories.

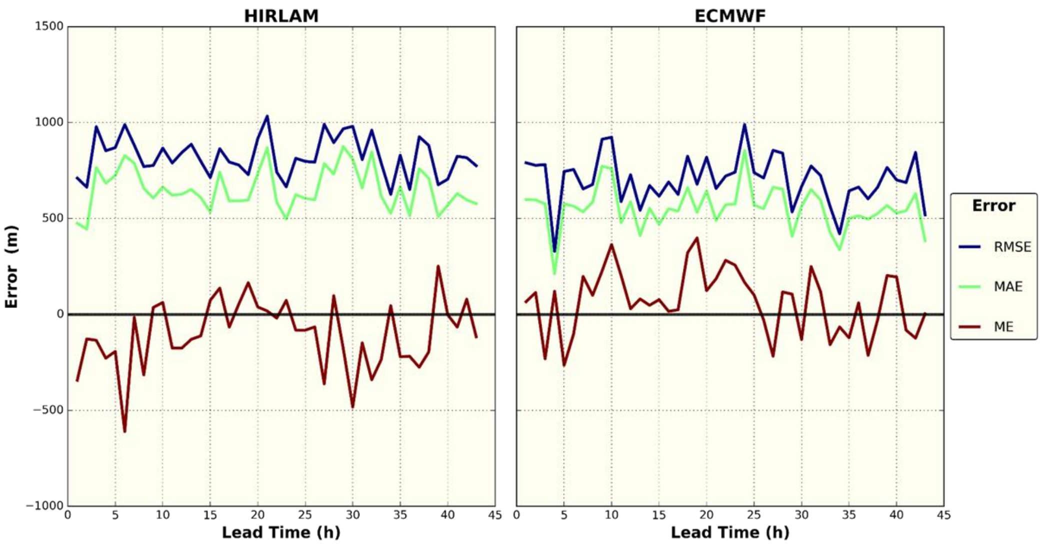
Slovenia report that cloudiness is the most problematic weather parameter to forecast, with significantly larger errors than for other parameters (temperature, wind and precipitation), but note that the problem can be at least partly due to the interpretation with high cloud in the model contributing to total cloud cover, while observations may not record this as cloudy weather. Czech Republic also report that large relative humidity errors tend to be associated with difficulty in handling the inversions sufficiently well. Greece report that HRES and LAM have similar errors for screen-level humidity (dewpoint temperature), while Sweden report HRES as outperforming the LAMs for dewpoint in spring and summer (due to more realistic diurnal cycle).

Hungary report that winter 2016-17 was dominated by cold anticyclonic conditions, with persistent low-level stratiform clouds in the Carpathian Basin, and that HRES and ENS underrepresented the fog and low cloud in many cases, contributing to the large negative bias for the winter. Greece report a small negative bias for HRES cloud cover, and both countries note that the LAMs have opposite (postitive) bias. Greece show that overall RMS error is lower for HRES than for their COSMO LAM. Sweden report HRES as being better than other models for low cloud forecasts.

Czech Republic report on verification of forecast global radiation against observations at 19 Czech stations. They report some underestimation during sunny days in summer. Spain also report on verification of HRES irradiance forecasts (SSRD and FDIR) compared to observations from the Spanish radiation network. They report that ECMWF performs slightly better than AROME-Harmonie.

Although screen-level dewpoint and (equivalently) relative humidity are of less direct relevance to most customers across Europe, compared to say temperature and precipitation, it should be stressed that these are very important parameters for ECMWF, because of close links with surface-based convection as highlighted above. So we would like to encourage Member and Co-operating States to provide feedback on these parameters in reports in future years.

Visibility



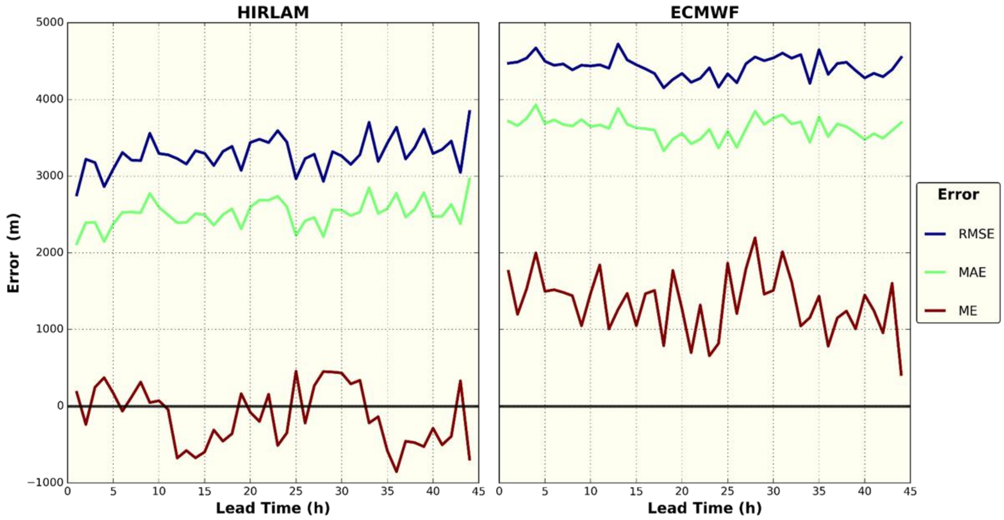


Figure 3: Temporal verification results for visibility from Latvia, comparing HRES and HIRLAM forecasts for the period 1 November 2016 to 1 July 2017. Upper panel shows verification results for all model and observational data below 2 km visibility; lower panel shows verification results for all model and observational data below 10 km visibility. Verification is against station observations in Latvia and surrounding area

With cycle 41r1 in May 2015 ECMWF introduced in a new visibility diagnostic, which is now also available in ecCharts. Latvia report that have recently introduced verification for the forecast visibility. They present results for the period from November 2016 – June 2017 verified against observations from Latvia (and some from Estonia and Sweden). Overall HRES forecasts have similar errors to HIRLAM for the visibility threshold of 2km, but has larger errors for the 10km threshold, mainly due to a larger bias (too high visibility) in the HRES (Figure 3). This is consistent with subjective reports from the UK that visibility parameter tends to be either low visibility or clear, with few cases of more moderate visibility. Note that the visibility output parameter was revised in July 2017 (cycle 43r3) to use the new seasonally varying aerosol climatology (based on CAMS aerosol re-analysis) and also now includes a dependence on relative humidity. These changes will help to address the reported issues.

Little other feedback has been received this year. We would encourage comments in future years.

SST and Sea Ice

Though no routine verification of these parameters has been performed some issues have been recorded.

Iceland report that in November 2016, 2-metre temperature forecasts by the west coast of Iceland were far too low, in both HARMONIE-AROME and HRES. The reason for these erroneous forecasts turned out to be an incorrect sea ice cover by the west coast of Iceland in the HRES analysis, resulting from incorrect sea ice detection by the OSISAF product. This meant that while the SST around Iceland were high, 8-10°C, and to the north of Iceland more than 4°C above the 1981-2010 average, in fjords in west and north west Iceland the SST was below freezing. Consequently the 2-metre temperatures by the coast in these regions were forecasted far below realistic values. The underlying problem was not resolved until March 2017, when the UK Met Office updated the OSTIA SST data set with a new filtered product from OSISAF. ECMWF implemented this and the fictional sea ice cover disappeared from the ECMWF analysis on 22 March. In the meantime, Iceland implemented a short term solution - from 7 December 2016 the HARMONIE-AROME forecasts were initialised with sea ice and SST from the ocean model

MyOcean. Although the SSTs in the fjords were still lower than observed, they became closer to reality and the gradient to the open sea was decreased.

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, France report that calibration substantially reduces the bias in the diurnal cycle of temperature over France and reduces the RMSE by up to 30%. For Turkey, the improvement is 5-20% for max and min temperature (using a Kalman Filter approach). Norway report that although HRES 2m temperature forecasts have significantly improved (cycle 41r2), statistical calibration can still make a substantial improvement. Post-processing the ECMWF surface solar radiation (FDIR) and adding aerosol information from CAMS improves forecasts of DNI over Spain.

Denmark report on new (not yet operational) calibration of the ENS 2m temperature and 10m wind speed forecasts for Europe. The calibration improves the match between spread and RMSE of the ensemble mean, by increasing the spread. However, the RMSE of the calibrated forecast is very similar to that of the raw forecast. It should be noted that when verification of surface weather parameters is made against station observations, the observation "representativeness error" can be a significant component of the apparent under-dispersiveness of the ENS, and should be taken into account.

The UK apply their own tropical cyclone tracker to the ECMWF, Met Office and NCEP ensembles. Verification for 2016 shows 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 4).

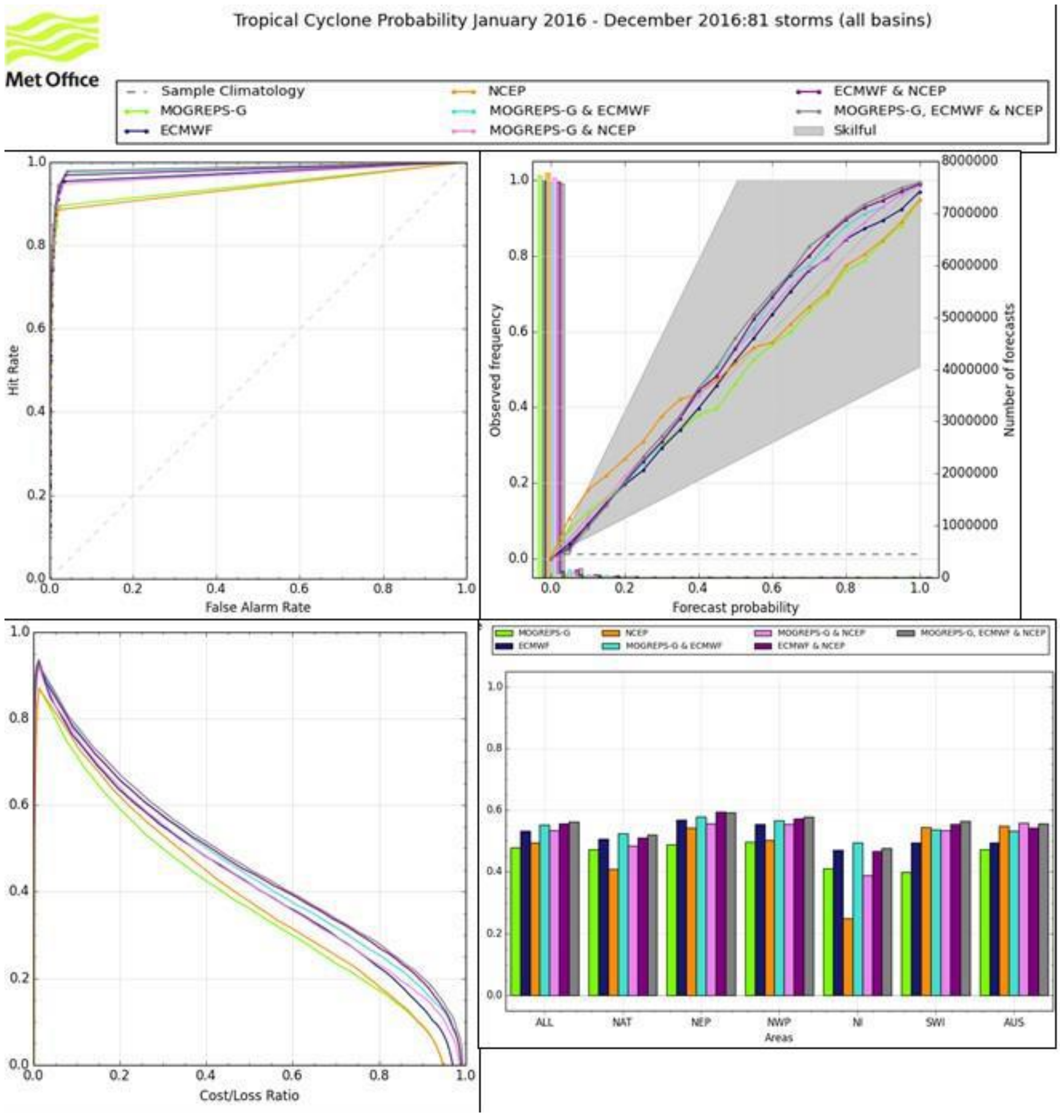


Figure 4: Verification plots from the UK comparing strike probability forecasts from January 2016 to December 2016 in 7 different ensemble models or multi-model combinations. Relative Operating Characteristic (ROC) plot (top left) assesses the skill of the forecast. A forecast with perfect skill would produce a curve from bottom left to top left to top right. The points along the curve are the corresponding hit rates and false alarm rates for each probability bin. Reliability plot (top right) displays how well the predicted probabilities correspond to their observed frequencies. Perfect reliability would be a diagonal line from (0, 0) to (1,1). Economic value plot (bottom left) indicates the relative improvement in economic value between the sample climatology and a perfect forecast for all cost loss ratios. Brier Skill Score (bottom right) for each basin, using CLIPER Climatology and Persistence forecasts as the reference in the skill score.

The UK also produce a Global Hazard Map (GHM) using ENS and the Met Office MOREPS-G ensemble. The 99th centile of the ECMWF model climate (M-Climate) from the operational reforecasts is used to define hazards for precipitation and wind. Verification of the precipitation forecasts again shows that while ECMWF is overall the best single-model ensemble, the multi-model combination has the highest skill.

Miscellaneous

Some countries verify or at least comment upon forecaster performance compared to models. Whilst performance gaps are small, there is evidence from Denmark, Hungary, and Turkey that forecasters continue to add a little value, on average. For Hungary the added value is primarily in maximum and minimum temperature forecasts (up to 10%). In Denmark forecasters add up to 5% to skill of temperature forecasts (up to day 5) compared to direct use of HRES (the ENS is one of the additional inputs used by the forecasters).

3.2. Subjective verification

3.2.1. Subjective conditional verification

Feedback has also been received concerning forecast biases related to specific synoptic situations, most of which were inferred subjectively. We can call this ‘subjective conditional verification’. Caution is needed given that these impressions have not been subject to objective testing, but we record some of the comments below.

Forecasters in Finland have noted a number of improvements since the resolution increase in spring 2016. They report improved wind direction for sea breeze and see improved 2m temperature forecasts, but note that there is still room for improvement, for example ECMWF is too cold in Fohn situations (HIRLAM does better). Low level cloud forecasts have improved (for both location and amount) and are now used in aviation forecasts (previously only HIRLAM was used), but aviation forecasters report some problems in northern Finland (mainly cases of too little low stratus, although some cases reported of too much). UK also note overall a slight clear bias for low cloud (though with some cases of too much convective infill in shallow convective boundary layer situations in spring). UK forecasters report that visibility tends to be either too dense or nothing at all (consistent with the objective results reported by Latvia); the revision of the visibility output parameter in July 2017 (cycle 43r3) will help address this.

Spain reported timing issues with fronts and convection and also suggest that meso-lows in the Mediterranean may be over-forecast (although Greece have previously praised the ability of HRES to predict small-scale Mediterranean cyclones). Greece report a general overestimation of precipitation totals, notably snow, in the northeast of mainland Greece, and one can imagine that this too could be closely linked to Mediterranean cyclone handling.

3.2.2. Synoptic studies

Several centres report on specific severe weather events that have affected them recently – notably Italy, Lithuania, Norway, Portugal and UK.

Very brief summaries of a few of these are given below but for more details please refer to the Member State/Co-operating State reports, and to the UEF2017 meeting presentations and posters at: http://www.ecmwf.int/en/learning/workshops/using-ecmwfs-forecasts-uef2017.

In November 2016, an unusually persistent south-westerly flow caused the second highest rainfall event recorded in Spitsbergen. This was well predicted by ECMWF which gave consistent signals for unusually high temperature and heavy precipitation, although as may be expected the higher resolution (2.5km) AROME-Arctic better matched the observed rainfall totals which were too low in HRES.

Lithuania report that all extreme events affecting the country in 2016 were well predicted by ECMWF at least 108 hours in advance, with an unusual case of June frost captured up to 10 days ahead.

Wind storm Urd (December 2016) was consistently forecast in the ENS up to a week ahead (Norway), although the highest wind speeds were under-forecast; the maximum wind was better represented in the post-processed LAM.

First indications for storm Doris (February 2017) were provided over a week ahead from the ENS (via the Met Office synoptic-type Decider application). Although the location of the main area at risk of severe winds was not captured at early lead times, consistency between subsequent ENS and Met Office ensembles gave forecasters confidence in the changing signals to a more southerly risk area.

Portugal reported a case from April 2017 where the ENS short-range forecast gave high probability for rain over Lisbon, whereas in the event the precipitation did not extend as far north-westwards as forecast.

Italy provide a detailed study of several days of intense rainfall over north-west Italy in November 2016 (this event is also included in the ECMWF Severe Weather Catalogue). The event was well forecast by the HRES and (together with more detail at very short range from the local COSMO-ME model) enabled timely earnings to be issued together with the Civil Protection authorities. The study highlights the role of warm conveyor belts (WCB) in the event and the association between the heavy precipitation and the WCB. ECMWF has recently initiated diagnostic activity focusing on WCBs including their role in magnifying uncertainty and in regime transitions as the main focus of its work with ECMWF Fellow Heini Wernli at ETH, Zurich.

In other cases Italy report that maximum 6-hour rainfall totals are closer to the observed values than they have seen in previous years, consistent with the resolution upgrade.

4. Monthly and seasonal forecasts

Demand for, and use of, monthly and seasonal forecasts continues to grow year on year. 17 out of 22 reports contain some reference to use of ECMWF monthly and/or seasonal products. In many instances a key motivation is demand from the energy sector, for example for hydropower.

Finland report on a new project CLIPS to raise public awareness of weather impacts - prototype products are being developed based on the ENS (up to 46 days forecasts) and verification is being developed in parallel. FMI is collaborating closely with ECMWF in this development. Norway are working with their flood forecasting authorities to use the ENS monthly forecasts for flood risk scenarios. Switzerland are setting up an early warning system for heat stress based on the ECMWF monthly forecasts (as part of the H2020 Heat-Shield project). They use the forecast temperature and dewpoint, bias corrected using a quantile mapping to compute a wet-bulb globe temperature (WBGT) as a heat-stress index. Verification shows positive correlation up to 30 days and CRPSS remains positive to 20 days for WBGT, which is slightly more skilful than the underlying variables (Figure 5 shows maps of results for day 15).

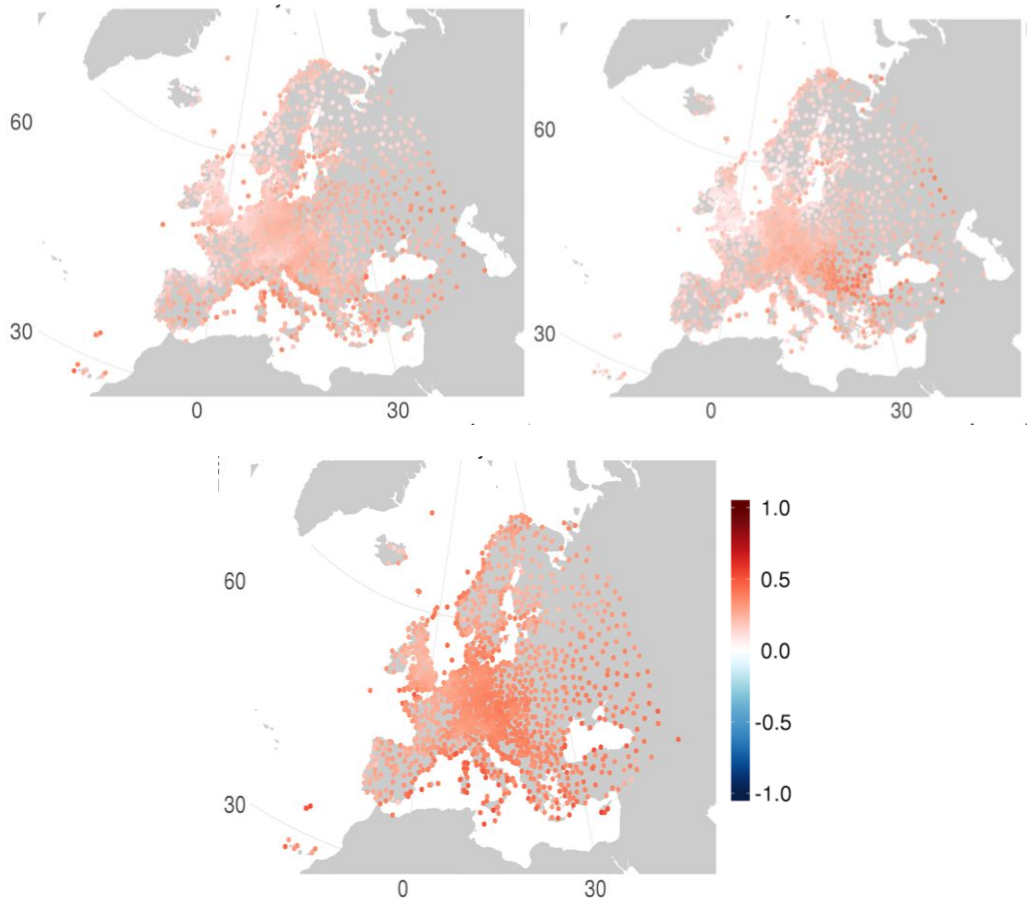


Figure 5: Verification from Switzerland for heat stress forecasts from ENS. Ensemble mean correlation of surface temperature (top left), dewpoint temperature (top right), and WBGT (bottom) at a lead time of 15 days for 1996-2015 hindcasts initialized between May and July.

Hungary, Switzerland and France provide some verification data. Forecasts of temperature are always found to be more accurate than forecasts of precipitation. Hungary show that there is only skill in month 1 of seasonal precipitation forecasts. For temperature some skill is seen out to month 6, although it is possible that this is a climate change signal. Since 2005 France have subjectively verified the monthly 2 m temperature anomaly forecast for their country, for weeks 1, 2, 3 and 4. The highest skill is in winter, probably from the teleconnections (MJO) and impact of large-scale weather regimes on temperature. They highlight that in winter scores are similar for weeks 3 and 4 (around 40% “good” forecasts) and this suggests weeks 5 and 6 may also have some useful skill (though the scoring system has not yet been applied to these forecasts). Extended-range forecasts, including the maps for weeks 5 and 6, have recently been added to ecCharts and feedback from users on these would be valuable feedback for ECMWF.

Finland produce outlooks for Baltic sea-ice for up to 6 months ahead (used to estimate number of icebreakers needed in the coming months), using the ECMWF seasonal forecast along with observations and analogues. Last winter sea-ice was over-forecast because negative NAO signal in the ECMWF forecasts was not realised. However, the positive NAO in February 2017 was well forecast in the ENS 46-day extended range forecasts.

5. Interactive feedback facility for forecast users

In June 2014 ECMWF introduced a new portal for interaction with its forecast users (https://software.ecmwf.int/wiki/display/FCST/Forecast+User+Home). The portal contains two main sections, one about known forecasting issues and one about severe weather events, which are both regularly updated through the year. As last year, we asked for feedback on whether countries were aware of these facilities, and how useful they are. There were fewer responses than last year; those that were received are summarised in Table 1.

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 (10)	0 (4)	2 (0)	2 (1)	4 (6)
"Severe event catalogue"	13 (11)	0 (4)	5 (1)	0 (2)	3 (3)

It seems that users are becoming more aware of these facilities, and comments were made that forecasters are learning about this when attending the ECMWF training courses and that the portal is now easier to locate from the main web site. The main reason given for lack of use by some countries was the pressure of the operational schedule for forecasters and a lack of time to explore the pages. However it was generally considered that this is valuable information that should be available to users.

6. References

Haiden, T, Janousek, M, Bidlot, J, Ferranti, L, Prates, F, Vitart, F, Bauer, P and Richardson, DS, 2017. Evaluation of ECMWF forecasts, including 2017 model upgrades. ECMWF Technical memorandum no. 817.