Technical
Memo

European Centre for Medium-Range
Weather Forecasts

860

Use and Verification of ECMWF products in Member and Co-operating States (2019)

Tim Hewson
(Forecast Department)

March 2020

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our website under:
http://www.ecmwf.int/en/publications

Contact: library@ecmwf.int

© Copyright 2020

European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, RG2 9AX, UK

Literary and scientific copyrights belong to ECMWF and are reserved in all countries. This publication is not to be reprinted or translated in whole or in part without the written permission of the Director-General. Appropriate non-commercial use will normally be granted under the condition that reference is made to ECMWF.

The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error or omission or for loss or damage arising from its use.

Abstract

Each summer Member and Co-operating States report on the application and verification of ECMWF's forecast products for the previous year, whilst ECMWF also gathers feedback in other fora. This report summarises the feedback collected between summer 2018 and summer 2019.

Uptake and use of ECMWF products and data appears to have increased this year across the full range of lead times (short range to seasonal). Encouragingly, there also appears to be a greater focus on probabilistic forecasting (as served by the ECMWF ensemble (ENS)), although some forecasters still don't use ENS output.

Positive comments were made about some newly introduced products, notably vertical profiles, the "integrated vapour transport" (IVT) diagnostic and "point rainfall". The ECMWF web-based workstation tool ecCharts was also praised, though some complained about slow speed (a new significantly faster version of ecCharts has since been implemented). Useful ideas for new ECMWF products were also received, such as using ENS-based front-densities to select a "most representative" ensemble member.

Multi-faceted post-processing of ECMWF output continues in many countries, mainly for specific sites, where calibrating observations exist. However, limited area model (LAM) output is increasingly being adopted as a form of "truth" in this sphere.

When performing verification, many compare ECMWF's high resolution runs (HRES) with LAM output, and so usually centre on shorter ranges (up to ~48-60 h). A very common finding, for almost every sensible weather parameter, was that HRES forecast biases have a diurnal cycle, and annual cycles are also often present. Overall, relative and absolute skill levels vary in many ways - e.g. with parameter, geography, weather situation, lead time and model. For low level relative humidity HRES reportedly out-perform LAMs, and through physical linkages this also benefits HRES handling of other weather features, such as low cloud and surface-based convection. Precipitation biases in HRES, for both small and large totals (compared to point observations), were again shown to exceed those of LAMs, but note that the experimental "point rainfall" product introduced in April 2019 helps address this model-resolution-related issue, at least for the ENS.

Large HRES forecast errors for 2m temperatures, in relatively extreme hot or cold situations, are a concern for some member states, notably in Scandinavia where very cold winter nights are demonstrably not nearly cold enough in HRES forecasts. ECMWF continues to work on this model issue. Meanwhile, using the insightful approach of conditional verification (a growth area at ECMWF), Finland show how 2m temperature errors in winter tend to be lowest when the skin temperature (ordinarily of a snow surface) is ~0° C. This is because energy exchanges then involve the latent heat of fusion more than temperature change.

Ensemble-related verification results were limited, but a multi-parameter multi-pressure level comparison by Germany between the ICON and ECMWF ensembles showed ECMWF to be consistently better over time at 48h leads in the northern hemisphere, except in the stratosphere, where ICON was much better. This IFS stratosphere problem will be addressed with a formulation change in cycle 47r1 in 2020.

The pivotal importance of correctly predicting severe events was re-iterated in many reports. For example, extreme Mediterranean cyclones, around 28 September and 29 October 2018, were mentioned by several countries. These cases also appear in ECMWF's severe event catalogue.

1. Introduction

Each summer ECMWF has been inviting 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 2019 reports (24 out of 34), which are available on the ECMWF website (https://www.ecmwf.int/en/publications/search?secondary_title=%22Green%20Book%202019%22)

Reports were provided by Austria, Belgium, Croatia Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Israel, Italy, Latvia, Montenegro, North Macedonia, Norway, Romania, Serbia, Slovakia, Spain, Sweden, Switzerland, Turkey, United Kingdom.

A summary of the reports is presented below. Content has been combined with (i) feedback provided during the "Using ECMWF Forecasts" (UEF2019) workshop held at ECMWF from 3-6 June 2019, and (ii) feedback from official triennial Member State/Co-operating State liaison visits undertaken by ECMWF between July 2018 and June 2019. In chronological order, visits were made to: Hungary, Greece, Slovakia, Czech Republic, Germany, Iceland, Israel, Finland, Denmark, Italy, Portugal. 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 (direct and "other")

c) Verification of products (objective and subjective)

d) Requests for additional output

e) Feedback on ECMWF "forecast user" initiatives

f) References to relevant publications

For the verification section (3), as in 2018, 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, with six contributions received relating to LAM-EPS systems, which is three more than last year.

Note also that the ECMWF IFS is generally 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 one new IFS cycle, 46r1, was introduced, on 11 June 2019.

Note that the results of ECMWF's own objective verification are considered separately, in ECMWF Technical Memorandum 853 (https://www.ecmwf.int/sites/default/files/elibrary/2019/19277-evaluation-ecmwf-forecasts-including-2019-upgrade.pdf).

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 all 25 states represented use IFS data directly in some way to prepare short range forecasts (up to, say, 48–72h ahead). All 25 also use IFS data for medium range forecasting, and 18 of these use ECMWF's monthly (=extended range) and seasonal (=long range) forecasts for operational purposes. For short range, monthly and seasonal the proportions using our forecasts appear to have increased since last year.

In the short range, 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 systems, commonly four times per day but sometimes more often, in, for example, 'rapid update' mode. Denmark are even targeting the creation of update runs every 10 minutes! Ordinarily LAM systems use BCs from the "previous" set of IFS runs - typically 6h old - to ensure product timeliness; this seemingly imperfect set up is probably unavoidable. Some LAM-EPS systems use BCs from more than one global EPS system.

In the medium, extended and longer ranges, ECMWF products continue to be the main or only output used by NMSs, and indeed Germany describe HRES and ENS as providing a "performance benchmark" for their own ICON-model-based global predictions. 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, often despite limited evidence of skill.

Operational short-range limited area modelling efforts within the Member and Co-operating states continue to move to higher and higher resolution. Currently a few run at ~4km (or >4km) horizontal resolution, most run at ~2.5 km, whilst some run at ~1km (Denmark, Greece, Iceland, Montenegro, Spain, Switzerland, UK). There is also increasing evidence of forays into sub-km-scale operational forecasting, usually over very small areas within countries, notably Iceland (0.75km), Denmark (0.5km) and Montenegro (0.5km). Working at much higher resolution than IFS commonly requires a multi-nesting approach, although Switzerland continue to run at 1.1km using HRES BCs directly.

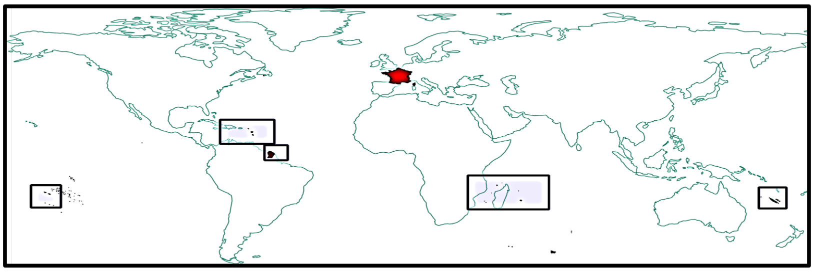

Figure 1: Five overseas domains over which France runs LAMs, based on Arome, 4x per day, using boundary conditions from HRES.

Whilst the main operational focus for all Member and Co-operating states is on local weather, some also have international commitments, for land and sea areas across the world, for which ECMWF forecasts are used in different ways. For example, France has commitments in many different tropical regions (Figure 1) whilst the UK has one forecaster rostered every day to solely handle forecasts for regions worldwide (outside the UK).

2.1. Direct use of ECMWF products

At shorter ranges, most NMS forecasters examine ECMWF products alongside output of their main LAM-based systems, which themselves, as discussed above, usually also incorporate ECMWF input via BCs. Around day 3 or 4 focus shifts onto using mainly or only ECMWF output, and similarly for monthly (extended) and seasonal ranges ECMWF is the only source of forecasts for most NMS. This all means that ECMWF forecasts are vital for a vast range of operational functions in most if not all the Member and Co-operating states. As Norway puts it, our products are "highly valuable for short range forecasting, and indispensable for the medium range".

ECMWF facilitates direct product visualisation through static clickable images on the ECMWF website, within ECMWF's complementary tools "ecCharts", and (related to this) "Dashboard". All three web-based tools continue to undergo upgrades and re-design work to improve usability. ECMWF output can also be viewed in static form on websites internal to NMSs. Some NMSs also continue to use the ECMWF graphics tool Metview to create plots for forecasters. Note that ECMWF has now combined Metview with Python-3, providing even more flexibility, and efficiency savings, for users (see https://confluence.ecmwf.int/display/METV/Metview%27s+Python+Interface).

ECMWF output is often viewed directly by forecasters on independently-developed dedicated forecaster workstation systems (such as "NINJO", "HAWK", "Visual Weather"). Most countries ingest into these a range of ECMWF products (especially from HRES). These workstations offer one major advantage: they allow many forecast relevant products to be overlaid (e.g. output from multiple global and limited area models/ensembles, observations, satellite data). However, one downside of this approach is that it may also 'prevent' access to the full range of ECMWF output: Denmark for example highlight that their NINJO workstation (used also in other countries) does not incorporate any ENS data. Users are again reminded that ecCharts provides a WMS service to facilitate the transfer of ECMWF data into local workstations.

2.1.1. ENS and HRES

ECMWF is now in the process of creating a new 10-year strategy and it is certain that ensemble forecasting will remain central to what we do. Nevertheless, it is again clear this year that our HRES forecasts are still very widely used, and in some NMS, for various reasons, these are used in preference to ENS output. An Icelandic forecaster commented that "HRES is used more often than ENS because it is less confusing". In Israel, the only ENS data that forecasters use is what is shown on Meteograms. And in Greece ENS products are used mainly for days 4 to 7, with HRES being preferred before that. It is also clear that the extra resolution provided by HRES brings clear benefits for many countries, especially those that are topographically complex, and so it is more meaningful then to inter-compare LAM forecasts with HRES than with ENS. The by-design tendency for ENS forecasts to spread evenly on either side of the HRES solution at shorter ranges can also limit ENS usefulness at those leads, although with the introduction of 50 EDA (Ensemble of Data Assimilations) members in cycle 46r1 in June 2019 that became less of an issue.

2.1.2. Severe weather

Another key tenet of ECMWF strategy is to alert forecasters of potential severe events as early as possible, to facilitate timely issue of accurate official warnings. A quote from a German forecaster suggests that we may be achieving that goal: "it is impressive how if there is a severe event the IFS is always the first to show it". The longest lead time at which a warning *can* be issued still varies a lot by country; the maximum seems to be 7 days, in the UK. Because of these variations the extent to which ECMWF output is actively used will also vary, with the output of Nowcasting tools and LAMs inevitably 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 seems to be increasing. Whilst the EFI and SOT are already widely used to alert forecasters to severe events, usage continues to grow. Many NMS reports include favourable comments about uptake and usefulness - Greece describes the EFI as "increasingly popular amongst forecasters". These parameters can provide early warnings for rain, wind, heat, cold and snow, and also, via the CAPE and CAPE-shear EFI/SOT parameters introduced relatively recently, for hazards linked to severe convective outbreaks. These were nicely described by Montenegro as "supercellular disasters"!

A migration over to more impact-based (probabilistic) warnings continues. This 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 for fulfilling international obligations, where forecasters' knowledge of the local climate is often lacking. Meanwhile many NMSs report continuing usage of probability products, which can help more with threshold-based warnings.

In 2018 ECMWF received a lot of positive feedback regarding precipitation type plots (meteograms and maps) that had been introduced in the previous 12 months. These products have again been complimented this year; by the Czech Republic (who presented a poster showing case studies at the UEF meeting), Slovakia and Norway. Norway reports that based on positive tests following introduction the meteogram product is now formally recommended for operational use in winter, for aviation forecasts also.

2.1.3. ecCharts

ecCharts is actively and increasingly used in many countries. Notably this year Latvia and Norway indicate that ecCharts now has official “backup system” status for their forecasting activities. For Norway this was made possible by ECMWF adding vertical profile products. And several countries also made very favourable comments - Italy spoke of “how useful it was for their forecasting”. But as in previous years there have been negative comments too, about the slow speed, from at least five countries; Germany for example says ecCharts is “too slow for operational use”. ECMWF has been targeting this negative feedback with a raft of improvements, which were first released in June 2019 in the “ecCharts-2” test system, as demonstrated at the 2019 UEF meeting. Zooming and panning for example is now faster and users have provided positive feedback about the improvements in ecCharts2 which is now operational. ECMWF also advised again that the way in which ecCharts is used can affect speed and performance; helpful suggestions are provided in online training material and during formal training sessions. Using the companion ‘Dashboard’ tool is one way to mitigate the speed issues.

There is also evidence that the ever-increasing suite of ecCharts fields are being used for increasingly diverse applications. Denmark for example say that specialized parameters for sand/dust storm prediction can be very important for their forecasters working abroad. This presumably refers to the ‘aerosol optical depth’ parameter ingested into ecCharts from CAMS-IFS forecasts.

2.1.4. New Products

Many new products appear first within the ecCharts framework. For example, EFI and SOT fields for the extended ranges have recently been added; CDFs for the same week-long time periods will follow soon. Feedback is also invited on these new facilities.

For the NMS reports, we specifically requested feedback on the new vertical profiles option that was introduced into ecCharts in summer 2018. Although Slovakia reported that it was not using it, six other countries provided positive comments, and most of those found the tool very useful. The UK stated that the “vertical profiles tool has been an extremely welcome addition”.

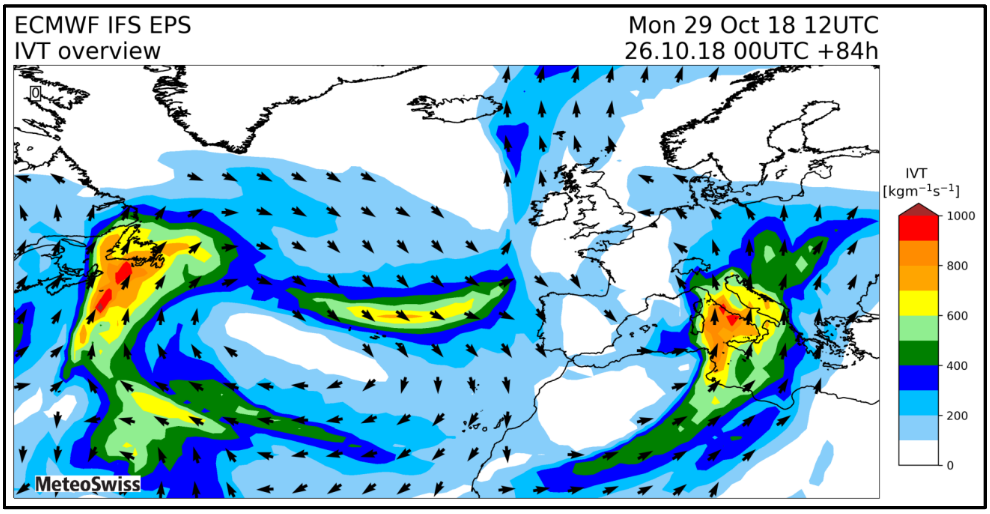

Figure 2: Example ENS IVT visualization from MeteoSwiss for a situation with potential for flooding in late October 2018. Although southern Switzerland was not affected in the event, flooding occurred in various areas in northern Italy.

Switzerland and Norway both compliment the new IVT (integrated water vapour transport) diagnostic output introduced into the IFS in June 2018. This is especially useful in mountainous areas, generally providing a better pointer to the potential for orographically-enhanced rainfall than can be provided by IFS rainfall output on its own (see Figure 2). It is significant that the two countries cited are two of the most mountainous countries in Europe. Norway put in a strong request to see EFI and SOT for IVT, highlighting how this could help "prevent devastating flooding" and "minimize income loss", particularly as preparatory actions need to start 2-5 days in advance. Accordingly, ECMWF is pleased to announce that EFI and SOT for IVT were introduced as operational IFS output fields with cycle 46r1 in June 2019, and related static web products were added at the end of the summer.

In April 2019 ECMWF introduced a new experimental 'point rainfall' product into ecCharts (from a suite called 'ecPoint'). This is a post-processed product that aims to deliver reliable and skilful probabilistic forecasts of rain-gauge-measured rainfall. This is done by adjusting for expected weather-dependant biases and sub-grid variability, that together make raw ENS forecasts of gridbox-average rainfall less valid at points, especially in convective situations. Although these products have only been in existence for a few months Spain, Norway, Hungary and the UK all make positive references in reports. Norway say that the output compares well with their LAM-EPS forecasts, and with rain-gauge measurements, and adds value because it picks up signals up to 5 days in advance. At the UEF meeting, where point rainfall was showcased, several attendees requested access to the related grib files through MARS. These will become available in due course; first ECMWF needs to progress data governance procedures that require WMO approval.

2.1.5. Other Products

The UK again highlights how Cyclone Database products (showing automated fronts and cyclonic features) continue to be regularly used by forecasters, and make a related new proposal (Section 2.2.2

below), whilst Norway illustrates how spaghetti fronts plots are “a big help for the forecasters as they efficiently visualise the spread in the ensemble”.

Miscellaneous references are again made to many other ECMWF web products. It seems that Meteograms in their different formats remain as popular as ever.

2.2. Other Uses of ECMWF Output

2.2.1. Post-processing

A new initiative reported on this year is post-processing of ECMWF forecasts via a calibration that incorporates LAM-based analyses (at ~1km resolution) as “truth”, albeit with additional input from observations. The major attraction of this must be that it allows gaps between observation sites to be comprehensively filled, in grid-like fashion, capturing considerable topography-related detail. The downside is that the post-processed output could be contaminated by LAM systematic errors. However, if these are small and/or successfully offset using observations, the final product may still be accurate at least for some parameters. Austria report on such an initiative (which they refer to as SAMOS, i.e. “Standardized Anomaly Model Output Statistics”), which is now at a development stage (see Figure 3). Israel are involved in a similar activity, based on the same INCA system (output is also used for nowcasting). Meanwhile Norway use LAM climatologies, operationally, for 2m temperature and wind adjustments. ECMWF is currently collaborating on a related proof-of-concept initiative using neural networks to downscale 100 m winds; currently this uses ERA-5 analyses (31 km) as the driver, and HRES analyses (9 km) to train, and as a target to verify against.

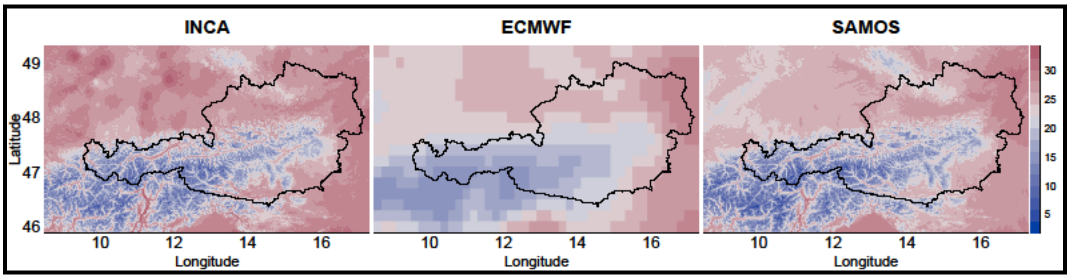

Figure 3: A SAMOS temperature forecast example for T+15 = VT 1st July 2016 15 UTC, with a verifying INCA analysis on the left, the raw ECMWF ENS mean forecast in the middle and the post-processed SAMOS forecast on the right. Whilst the main input to SAMOS is the ENS forecast, the local topographic influence missing from ENS is nonetheless captured.

Other post-processing initiatives, often referred to generically as “MOS”, are as wide-ranging as ever but have the common goal of systematically improving upon raw model output. Benefits would ideally be seen in *all* weather situations, although that is more challenging because rarely experienced scenarios, that may be conducive to extremes, provide little or no training data. Techniques include, but are not limited to, linear regression, multiple linear regression, polynomial regression, logistic regression, non-homogeneous Gaussian regression, quantile-quantile mapping, use of Kalman filters, gamma-distribution fitting (for rain), discriminant analysis, and simple altitude corrections. Certain countries perform post-processing specially tailored for local needs: for example, Norway target 2m temperature errors caused by their intricate coastline, Finland use special algorithms for determining precipitation type and Spain dress ENS irradiance output at short ranges to increase spread and so improve reliability.

Forecasts are generally provided for a set of sites. The larger NMS's site sets often cover the world, whilst smaller NMSs focus on Europe or just their own country. Sometimes a higher level of sophistication is incorporated for forecasts for the originating country (e.g. on the Norwegian yr.no website). In some instances, the number of predictors used is extremely large: Austria quote 37, which include not only raw model output but also derived parameters such as relative vorticity.

Some post-processing techniques are applied to just HRES. Germany's short-range auto-TAFs (Terminal Aerodrome Forecasts) for example are based on a weighted MOS using the last two HRES runs, out to a 41h lead. Increasingly however ENS provides an additional or alternative input to HRES, especially beyond short ranges. Sometimes IFS output is also combined statistically with output from LAMs, and other global models; in Belgium, Greece, the UK and Germany for example. Indeed, the UK actively produce multi-model products, and would like ECMWF to collaborate in this area.

In some countries such as Germany and the UK special post-processed products are designed to assist with early warnings. Germany have even developed their own type of EFI, known as the EWI (Extreme Weather Index).

2.2.2. Derived Fields

Derived products are generated locally in NMSs for many reasons: often for use by NMS forecasters, also for specific societal or economic applications, and sometimes for use by the public. Whilst many countries reference again products covered in previous years' reports, the range of outputs and uses seems to be growing, as does the level of ingenuity applied. Often the objective is to provide products that historically ECMWF has not, although on occasion ECMWF fields are re-derived in a way that is more tailored to local needs.

There can be merit in creating new products via simple re-gridding or smoothing techniques. Indeed Sweden adopt a "neighbourhood approach" for all their model output (except ENS means), by applying a 20 km radius template around each gridpoint, and by then computing percentiles from values at all gridpoints that fall within. These percentiles are then assigned to the said gridpoint for plotting. This type of approach seems most appropriate for convection-resolving models, and ECMWF will have to consider something similar as its future model versions enter the grey zone, and as the huge number of ENS members then needed to cover all convective possibilities becomes unaffordable.

Convection-related indices continue to be generated in many Member and Co-operating States for forecasters (e.g. in Italy, Latvia. Czech Republic). This evidently reflects the lack of a discrete representation of the hazard at global model resolution and the severe impacts that can arise. More specifically, at longer ranges forecasters need to identify a broadscale risk, and then, at shorter ranges, to predict the hazards themselves, ideally giving guidelines on timing and location. Whilst ECMWF currently provides CAPE and CAPE-shear EFI fields to flag up areas at risk from convective hazards *in general*, Austria report on an initiative that goes a step further, distinguishing which areas are at risk from high convective gusts, which from heavy rain and which from hail. This they call the "Sweetspot Index"; it is empirical and uses HRES-based convective indices as input. Note that ECMWF's ENS-based point rainfall fields, discussed above (Section 2.1.4), can help identify areas at risk from extreme localised rainfall, whilst our recently introduced lightning density products, which received favourable comments in reports from Norway and Switzerland, aim to predict thunderstorm activity levels directly. In the coming years we plan to investigate in addition how the risk of hail might be more directly assessed from IFS fields. This could also be of interest for hail suppression activities carried out in some central and eastern European countries (e.g. as reported by Serbia).

It seems this year that some IFS surface-based parameters are attracting increased attention linked to specific customer needs. Spain perform a 'snowpack analysis' using the differences between successive HRES analyses as a guide to what snow has fallen. A word of caution is perhaps necessary here - wide-ranging factors such as the IFS assumption of uniform snow density, erratic observing practices, incomplete observational coverage and miscellaneous model issues with snow (and sleet) accumulation and melting have the potential together to make this assessment somewhat problematic. However, ECMWF's new multi-layer snow scheme, due for implementation in cycle 48r1, should help. Meanwhile, within an ESA project, Romania are using gridded soil moisture fields to see how pre-conditioning affects landslide occurrence at regional scale, whilst Latvia continue to use IFS-based derived parameters to predict river ice formation and break up, and associated flood risk. And Italy continue to create sea state codes from our wave model output.

ECMWF recognises that its inevitably broadscale approach to clustering ENS output does not necessarily suit the needs of every country, and this year the UK again highlight how it's 'Decider' system, that delivers many multi-model UK-centric synoptic-scale-cluster-related products, continues to be regularly used in the medium and extended ranges. The reader is referred to the 2019 UK report, and the 2018 version of this report for more discussion and examples. Meanwhile, Spain have adopted the clustering approach developed by ECMWF to create their own clusters separately for two regions, Iberia and the Canary Islands.

The UK provide examples of an imaginative alternative to the front spaghetti charts, which are 'front probability' maps. They use alongside this a score-based system to identify the most representative member (Figure 4). This has exciting potential for a number of applications. Firstly, for forecasters that need to "hang their hat" on a most probable outcome, this is an ideal product. Secondly, for verification and to develop forecaster guidelines it would also be very interesting to see how often and at what lead times HRES and/or Control attain/retain the status of most-representative member(s). And thirdly there may be potential to investigate links between these outputs and ECMWF's cluster products, which also show, for each cluster, the most representative member. We are therefore very keen to collaborate and investigate further how these ideas might be used at ECMWF.

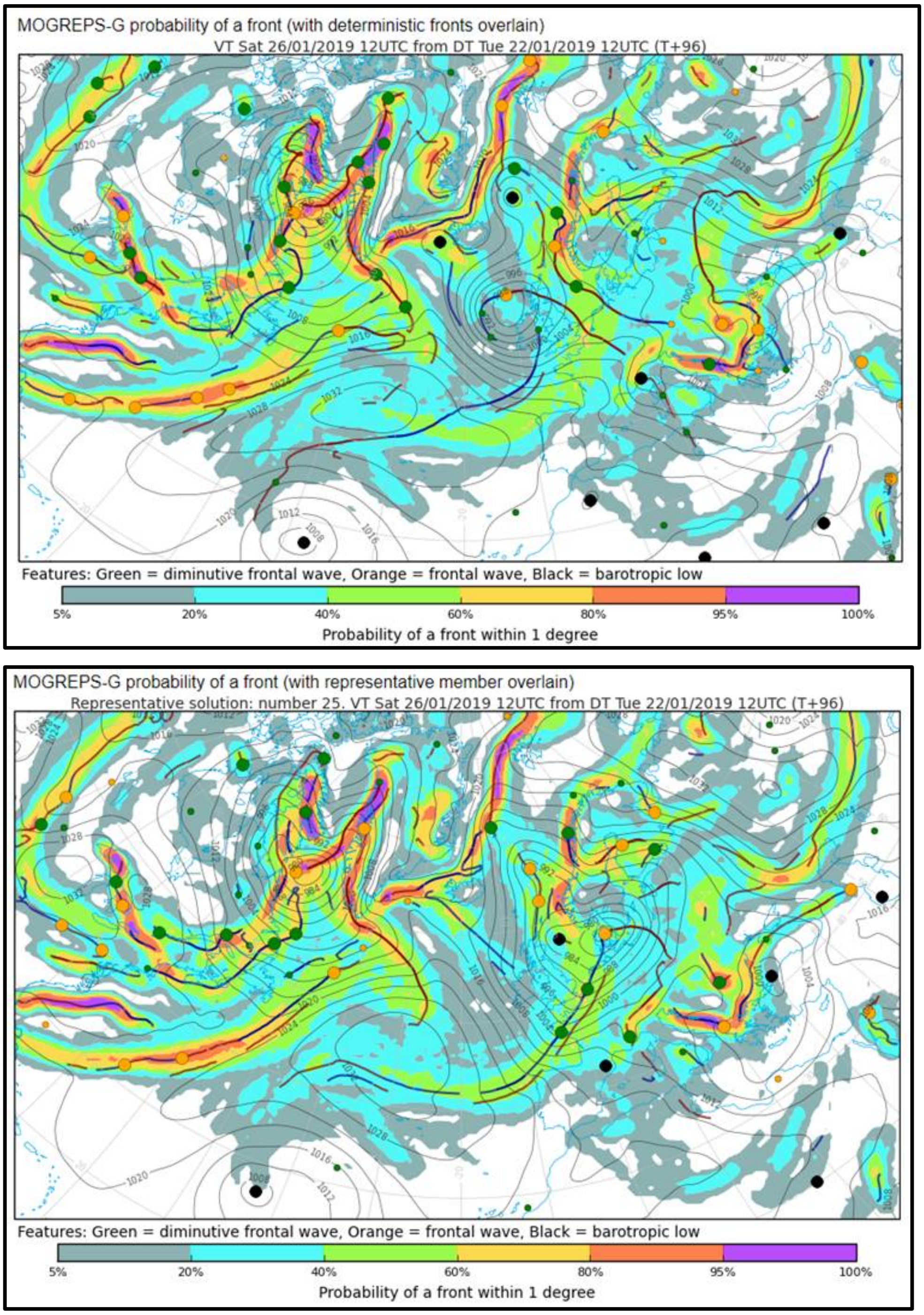

Figure 4: An example of the objective fronts probability map (shading) for T+96h. Top: Fronts from the deterministic model (represented by blue and red lines) can be seen to be an outlier compared to the ensemble consensus (coloured areas) over both the Atlantic Ocean and Northwest Europe. Bottom: Fronts shown are from the most representative member for this lead time. Plots are from the Met Office global ensemble (MOGREPS-G) but could equally be applied to ECMWF data.

Italy are using a decision tree approach to create a meaningful present weather coding (e.g. “drizzle”, “thunderstorm”) from HRES output. Similar activities have been previously reported by other countries.

Applications are often in media or “app” forecasts, that require one word or symbol to describe the weather, although in the case of Italy the output is reported to have a higher-level goal, namely impact prediction.

Other miscellaneous derived fields generated from IFS output include the following: height of the -10C surface (Serbia) and a turbulence parameter (Spain), both for aviation purposes; visibility based on relative humidity, precipitation and, curiously, latitude (Sweden), as an alternative to ECMWF’s equivalent (in spite of positive verification results for the latter that Sweden reported last year); advection variables and tropopause height (Serbia); Meteograms showing many more parameters than ECMWF’s, including mean sea level pressure and convective precipitation (Italy); fire weather indices (Denmark, Czech Republic); rime, snow drift, ventilation and soil drought parameters (Czech Republic), and a highly tuned UV index (Israel). Israel consider the UV index “very important” and had several related exchanges with ECMWF (they also praised the CAMS UV index, which they say has improved considerably in recent years).

Norway and Latvia also report on the use of re-analysis data (ERA-Interim, just retired, and ERA-5) for specific customer applications. In the case of Norway, it is used as a “weather-generator” for hydropower planning.

2.2.3. Limited Area Modelling

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 the atmosphere and/or the earth’s surface for those models too. This process is also known as “dynamical downscaling”. Mostly this happens via the Optional “Boundary Conditions” Programme which provides additional HRES and ENS forecasts from 06 and 18Z data times, now at hourly intervals (see document ECMWF/TAC/51(19)16).

The limited-area models that use ECMWF data via the Boundary Conditions programme, and that are referenced this year, are ALADIN, AROME, ALARO, HARMONIE, COSMO, WRF. WRF-NMM, NMMB. Typically these run 4x per day, although there is a lot of variability. The LAM-EPS systems, which are essentially based on versions of the above, have the acronyms: ALADIN-LAEF, COSMO-E, MEPS, COSMO-ME EPS, C-LAEF, COMEPS. Sometimes these LAM-EPS systems use ENS data for boundary conditions, although creative alternatives also exist. Often the EPS versions have somewhat lower resolution than their “deterministic” counterparts, because of computational resources, though sometimes this challenge is instead met by running the ensemble much less frequently - e.g. just once per day. Whilst it is difficult to arrive at a representative figure the average maximum lead time for LAM and LAM-EPS systems seems to be about 57h, although the range is large, from 12h in a nowcasting mode (24x/day), out to 144h for one 4km model. Operational domains naturally vary, both in size and location. Most inevitably centre on European countries, though France run also for five overseas domains (Figure 1) and Norway run a model for the Arctic.

Last year we highlighted that formal co-operation between different countries, in merging and developing model formulations, was growing. This trend has continued, with for example Latvia now partnering with Sweden, Norway and Finland for one EPS system (MEPS). To help make these cooperations work different NMS sometimes take responsibility for different parts of the model formulation. For example, Israel focuses on cloud-radiation interactions within the COSMO consortium.

The other modelling class where ECMWF forecast data is used, often much more directly, is trajectory, dispersion and air quality modelling. Activity in this field appears to have grown in the last year. It remains largely deterministic in nature, although Austria in particular have investigated the use of ENS, with the direct technical assistance ECMWF provided being acknowledged in their report.

The Metview trajectory- and dispersion-related modules FLEXTRA and FLEXPART are used operationally by Italy, Switzerland and Austria to for example predict contaminant dispersion in the case of chemical or nuclear accidents. Switzerland and Austria use alongside these the LAGRANTO and TAMOS systems. Meanwhile to handle accidental releases France and Spain use the French MOCAGE-ACCIDENT model framework. The Czech Republic are also active in this area. Then for air pollution modelling in general France and Spain use MOCAGE, whilst for finer-scale requirements France utilise the PERLE system. Austria and Slovakia use WRF inputs for air quality modelling, respectively in the 'Chem' and CMAQ systems. Norway meanwhile employ the EMEP model, which in effect allows them to gauge 'who is polluting whom', by defining the source locations for poor air quality events for cities around Europe. For details of these various systems, of how they have evolved in the last year, and of the various ways in which IFS data is utilised in each case please see the individual NMS reports.

The remaining modelling activities reported on lie in 3 categories: hydrological, oceanic and surge, and sea pollution modelling. Slovakia, Latvia, Serbia, Israel and the Czech Republic all state that IFS data, including ENS, is a key input in their own hydrological models. For ocean modelling Sweden refer again to using NEMO, which uses HRES for the upper BCs whilst Norway run TOPAZ daily out to 10 days, for seas and sea ice in the Arctic. Denmark refer to running several wave, ocean and sea-ice models (e.g. HYCOM-CICE) over different domains, using IFS output in different ways. France meanwhile run the internationally competitive MFWAM wave model, using HRES winds and sea ice, and currents from a Copernicus CMEMS source. Surge modelling is performed in France, Denmark and Norway and some of these suites have improved this year; IFS atmospheric fields provide input. In Norway and France there is an ensemble approach of sorts, but only in Norway is ENS output actually used. For Norway this is a change from last year; they say the change has been beneficial because it delivers more consistent warnings. Oceanic pollutant drift is predicted operationally by Greece, using MOTHY, and by Denmark and Norway.

There is also some interest in fire modelling, as noted during the visit to the Greek NMS (see also this UEF2019 presentation about the tragic event in Greece in July 2018 that killed more than 100 people:

https://events.ecmwf.int/event/119/contributions/565/attachments/117/206/UEF2019-Giannaros.pdf

https://ecmwf.adobeconnect.com/po9q9lme9fk6/?launcher=false&fcsContent=true&pbMode=normal)

3. Verification of ECMWF products

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 43r3 became operational 11 July 2017

Cycle 45r1 became operational 5 June 2018

Cycle 46r1 became operational 11 June 2019

This means that in this year's reports verification corresponds mainly to cycle 45r1, although in some the verification period used is 2018 which will encompass cycles 43r3 and 45r1 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 (e.g. see several plots in ECMWF Technical Memorandum 853 (https://www.ecmwf.int/sites/default/files/elibrary/2019/19277-evaluation-ecmwf-forecasts-including-2019-upgrade.pdf).

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 help resolve some of the issues listed above, and some results deriving from that are presented below.

3.1. Direct ECMWF model output (HRES and ENS), and other NWP models

Many reports focus on comparing HRES with LAMs, and for this reason usually centre on the shorter ranges (up to about 48/60h). Again, 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.

Precipitation biases in LAMs, for both small and large totals (versus point observations), continue to be mostly smaller than for HRES, as expected. 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 this year is that in the short ranges at least (where the bulk of the comparisons were performed) a multi-model approach to forecasting continues to have considerable merit. This would be particularly true if one could vary weightings according to known synoptically-varying performance characteristics. This concept is undoubtedly being used in subjective fashion by forecasters across Europe.

Details, by parameter, are given below. Some of the IFS issues arising here were known about, and most of these are also listed in the ECMWF's publicly accessible 'Known IFS Forecasting Issues' web page at: https://confluence.ecmwf.int/display/FCST/Known+IFS+forecasting+issues , which continues to be updated several times per year.

In the short and medium ranges, reports on ENS-related verification were as follows: surface and deep atmosphere parameters (Germany, for ICON EPS benchmarking), 2 m temperature (Norway and Belgium), multiple surface parameters (Hungary and Denmark), precipitation totals (Finland and Austria), cloud ceiling (Israel), irradiance (Spain), Baltic ice cover (Finland), regimes (UK), jumpiness and skill versus HRES (France, subjective), post-processed 2 m temperature (Switzerland and Finland), post-processed wet-bulb globe temperature (Switzerland). Discussion of the remaining ENS-related NMS contributions is incorporated into the parameter-specific sub-sections just below. On the topic of ENS handling in general Denmark make one particularly noteworthy remark - they say "ENS … continues to be severely under-dispersive for short-range forecasts of most key forecast variables". Whilst this can be indicative of genuine problems - e.g. with low cloud in winter -ECMWF has investigated and believes that the main reason is representivity (i.e. sub-grid variability); when this is accounted for the problem mostly goes away.

Feedback in the category of 'subjective verification', that often reflects forecasters' experiences, has also been blended, this year, into the sub-sections just below. This is a logical step because perceptions usually mirror features seen in verification statistics.

3.1.1. 2 m temperatures

Some NMSs (e.g. Italy, Latvia) show that HRES errors are typically larger in winter in Europe, which is broadly consistent with ECMWF's own verification. Several countries comment on different aspects of HRES bias; Austrian results show a strong diurnal cycle in bias, which has a larger amplitude in summer (1.8° C versus 1.4° C in winter), whilst both Spain and Sweden illustrate large negative biases, generally, in spring. Several countries (e.g. Greece) show that HRES maxima are too low in summer and minima a bit too high in winter, which seems to be a general result, although regionally (and inevitably with synoptic situation) the situation can vary a lot.

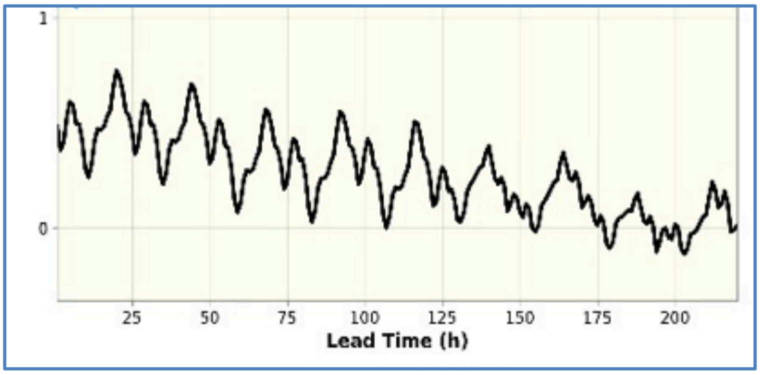

Figure 5: HRES 2m temperature bias as a function of lead time, from Latvia's report

One plot from Latvia (Figure 5) hints at spurious drift in HRES output. Note how the bias reduces with lead time; this trend is seen only in quarter 4 (Oct-Dec). Even though longer leads look better we know that snow accumulates too quickly on the ground as we move into winter, and this trend may well be a manifestation of the detrimental impact that has on 2 m temperature. One wonders also to what extent such a drift might affect absolute values forecast in the extended ranges.

One significant IFS problem re-iterated again this year by Finland, which impacts on winter-time bias metrics generally, is the insufficiently low minima forecast by HRES in extreme winter situations. Sweden's report supports this; showing that very cold nights are "missing" from HRES, and meanwhile very hot days in summer are underpredicted; they quote frequency biases for <-30° C and >30° C to be respectively 0 and 0.5. Meanwhile the AROME 2.5 km model used in Sweden has corresponding values of 0.3 and 1.4 respectively. So, although under-prediction of "temperatures in heatwaves" has been an IFS characteristic for some time, it is still quite possible for biases to be of the opposite sign, indicating that IFS errors are not so straightforward to eradicate.

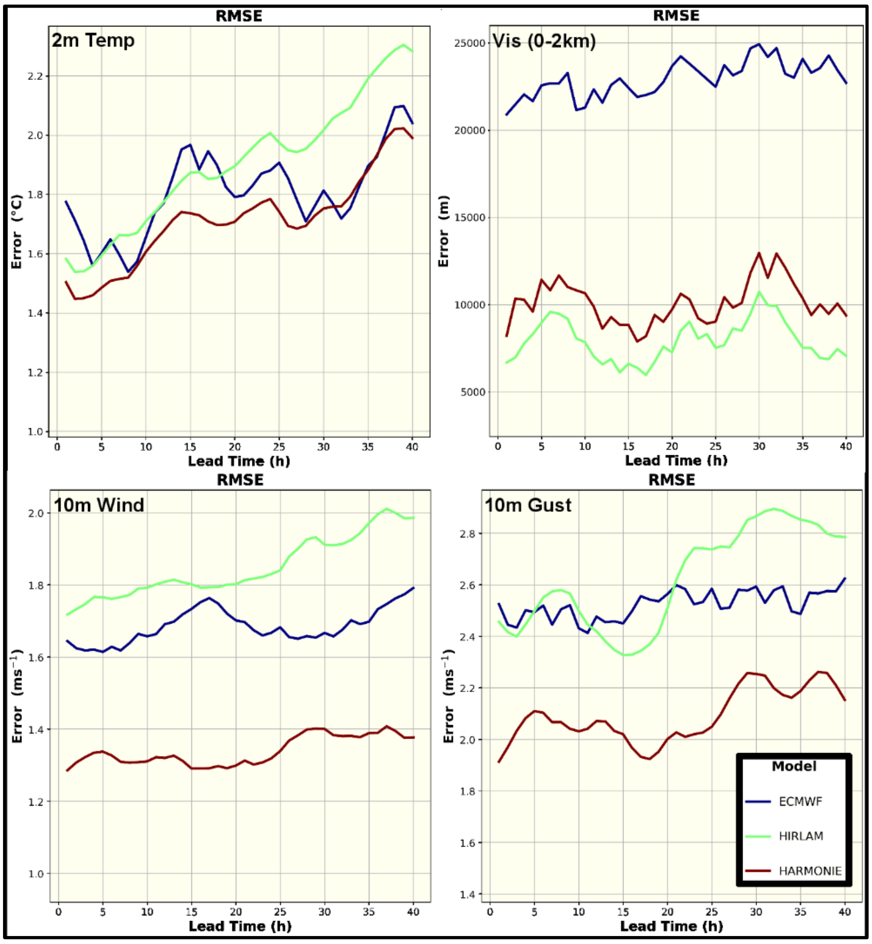

Figure 6: Comparison of RMSEs in 2018, over Latvia, for two LAMs (Harmonie and HIRLAM) and HRES, for 4 surface weather parameters: 2m temperature (top left), Visibility (0-2km range) (top right), 10m wind (bottom left) and 10m wind gusts (bottom right)

Examples of HRES performing as well as or better than LAM counterparts are commonplace. Romania and Serbia each show HRES to be as accurate as or more accurate than three different LAMs, that have resolutions of ~7 km and 4 km in each country respectively. In both cases domains are focussed on E Europe. Meanwhile, Slovakia show forecasts from HRES to be as good as forecasts from three LAMs with resolution in the range 2-9 km, Hungary show HRES to be a bit better than AROME 2.5 km, and Greece show HRES to be marginally better, overall, than 1 and 4 km versions of COSMO, the main differences being in winter.

Examples of better LAM performance were as follows: Italy showed COSMO 5 km to perform better than HRES in spring and autumn, Slovakia showed AROME 2 km to be a bit better generally, whilst Austria show that AROME 2.5km can predict better than HRES, but only over "flatlands" in summer. Meanwhile HARMONIE (2.5 km) improves upon HRES at most short range leads over Latvia (Figure 6).

France illustrate that over overseas island territories AROME-OM with a 2.5 km resolution performs better than HRES, which is probably not surprising given that island sizes may not be that much greater than HRES resolution. Similar "resolvability" reasons probably account in large part for HRES and ENS being systematically too cold, by 1-2C, at different times of year in different parts of Norway (which Norway have successfully addressed using post-processing).

As last year, France report that using HRES, rather than their ARPEGE global model, to deliver BCs for a Western Europe AROME LAM version makes little difference to skill (for most surface parameters).

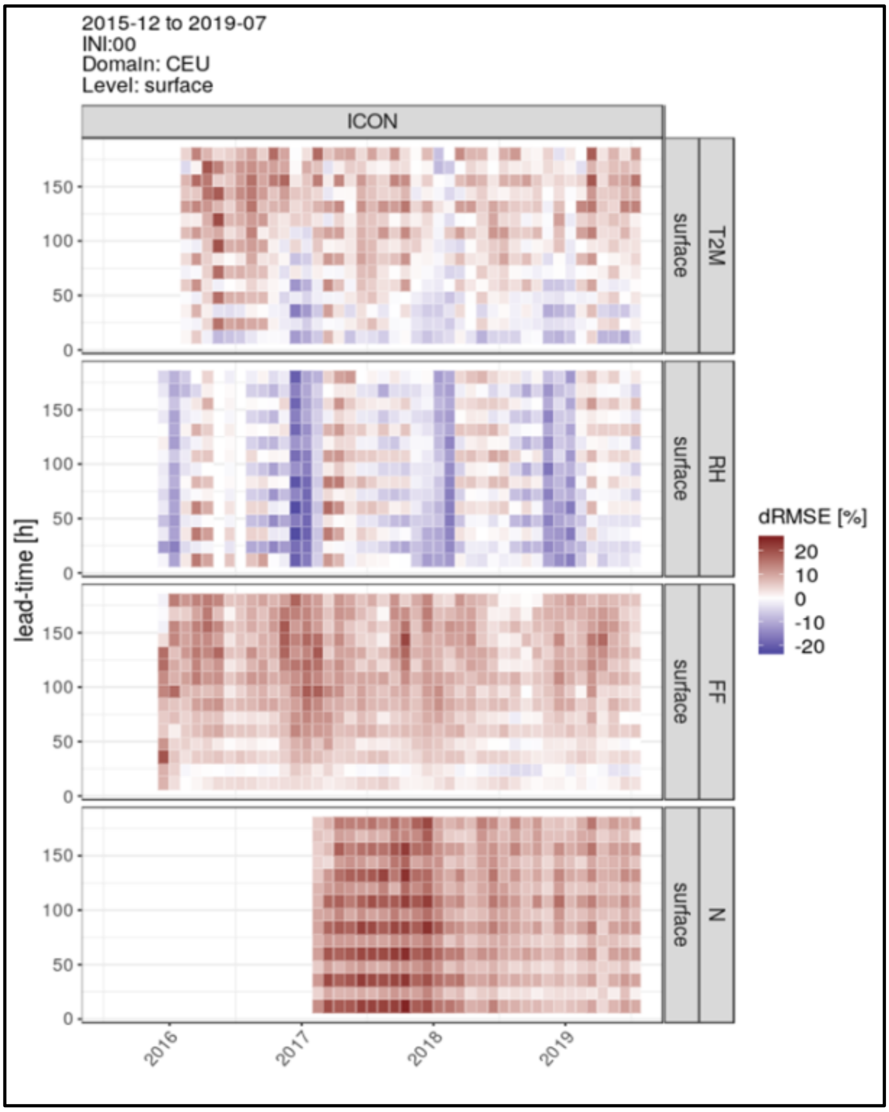

Figure 7: Deterministic model verification showing relative difference in root mean square error (dRMSE [%]) of ICON minus HRES OOUTC runs, for a European domain, using SYNOP observations as truth. Brown (blue) colours show better scores for IFS (ICON), for different lead times (y-axis), for multiple years and months (x-axis). The 4 blocks denote 4 parameters - from top down: 2m temperature, relative humidity, wind speed and cloud cover. The horizontally-striated appearance is symptomatic of differences between ICON and HRES in the diurnal cycles they exhibit in RMSE (and bias).

Germany compare their ICON global model with HRES over Europe. At some short lead times over the last three years ICON has been performing slightly better (Figure 7), with biases at those leads being of comparable magnitude (not shown), but at longer leads HRES is the clear winner. Another interesting result is that the annual cycles in bias between the two models look very different, which probably reflects differing good and bad points of the physics in each, and hints that using collaboration to help improve formulations could be quite efficient.

Belgium provided the only report referencing ENS 2 m temperature verification directly - they examined leads up to day 4, for five sites, and in that dataset found minimal difference in the accuracy of ENS mean versus HRES (the same applied for other surface parameters too).

3.1.2. 10 m wind

It is consistently reported that, *relative to observations*, HRES mean speeds are stronger by night than by day. But whether HRES forecast values, at each of these times, are positively or negatively biased seems to depend on country and on whether the sites are mountain sites. We have repeated strong evidence now of a general HRES mean-speed under-prediction over mountains, from Norway and Iceland, and this year also from Denmark (referencing Greenland, where there is a "systematic failure to predict strong windstorms") and Hungary (in a synoptically different region, remote from storm tracks). So probably in results from other countries, where mountain and non-mountain data is pooled, this will also affect the perceived bias. Even though Spain have subjectively assessed IFS wind gusts to be systematically too strong over elevated areas, and demonstrate with an example, this is not necessarily inconsistent - in case studies we sometimes observe a very large difference between concurrent mean and gust speed forecasts over mountains.

For lowland sites Hungary report that speeds are on average over-predicted at all clock times, but the night-time discrepancy is bigger. In Finland winds tend to be too strong at night and too light by day. Meanwhile Italy report net underprediction throughout (averaging 0.5-1m/s) with the biggest deficits by day and in summer. Perhaps this net under-prediction is not a characteristic of Italian lowland sites but owes its existence to inclusion of mountain sites?

Overall, performance metrics reported this year do not clearly show either HRES or LAMs to be better. Hungary report that over mountains HRES and a 2.5km LAM are comparable for mean winds, but that HRES gusts, which have a bias ~0 there, are better. In Latvia HARMONIE (2.5km) has RMSEs that are 25% smaller (for mean and gust) than they are for HRES (Figure 6), whilst in Sweden and Italy respectively AROME (2.5km) and COSMO-ME (5km) perform marginally better, overall, than HRES. The better performance in Italy relates in part to smaller summer biases in COSMO-ME with e.g. stronger daytime winds. In Hungary, over lower terrain, HRES and AROME 2.5km performance has been similar. Meanwhile Spain report that HRES is better than their HARMONIE-AROME (2.5km) LAM, and Serbia report that HRES beats their various 4 km models.

France again show that AROME 2.5 km has better winds than HRES over international Island domains, whilst in a global model inter-comparison ICON and HRES exhibit only minor RMSE differences for mean speeds at short leads (Figure 7). Verification of ICON-EPS alongside ENS suggests that in the Tropics the ICON gust forecasts are much better, though results are only shown for June 2019.

A curious marked increase with lead time in HRES gust bias is again seen in results from Latvia this year.

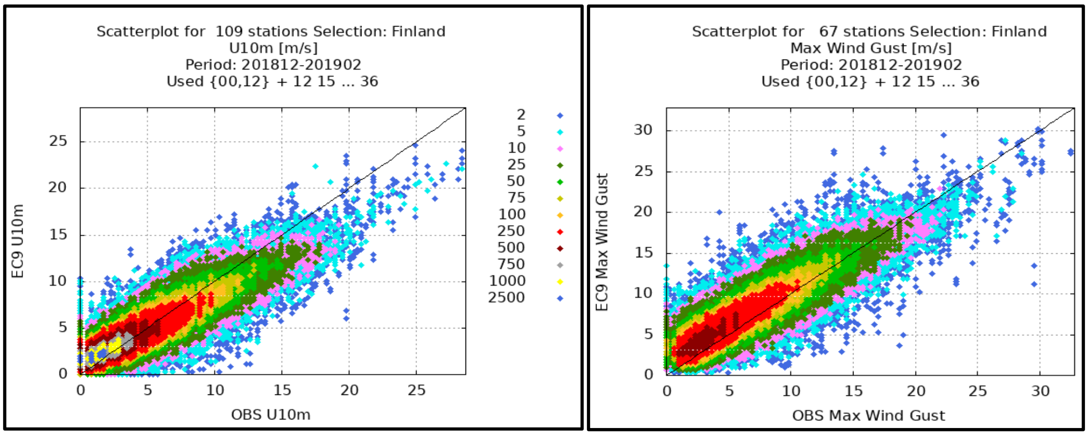

Figure 8: Scatterplots for HRES forecasts of 10m wind speeds on left and max wind gust on right versus observations from Finnish stations in winter (DJF) 2018/19, for lead times of 12-36h. Units are m/s.

Finland present interesting plots that they say show, at short leads, a dependence of HRES forecast errors on observed speeds (Figure 8). Very probably there is overprediction of both means and gusts at low speeds, and underprediction of high mean speeds, although whether there is under-prediction of high gusts (right panel) is less clear. The apparent discrepancy could instead be due to the contaminating effect on the plot of random errors. The larger are the random errors, the further away from the diagonal, and the closer to the horizontal, will be the apparent best-fit line. Note how the number of observed and forecast values around 25m/s is about the same. Spain noted that HRES under-predicted for observed events of >10m/s, though whether this could have been due to random errors is not clear. Meanwhile Sweden praised the more extreme HRES gust forecasts produced for last winter; this was significant for them because their winter was relatively stormy.

3.1.3. Precipitation

Many new verification results related to precipitation mirror those presented last year, so discussion in the 2018 summary report remains relevant.

When forecast gridbox totals are verified against point observations, as is very commonly done (at ECMWF also), the scale mismatch can result in somewhat negative conclusions regarding frequency bias (FB), that ordinarily do not reflect true “model issues”. Indeed, LAMs will usually exhibit a “better” FB than global models, but that will be mainly because their gridbox scale is closer to the observation scale.

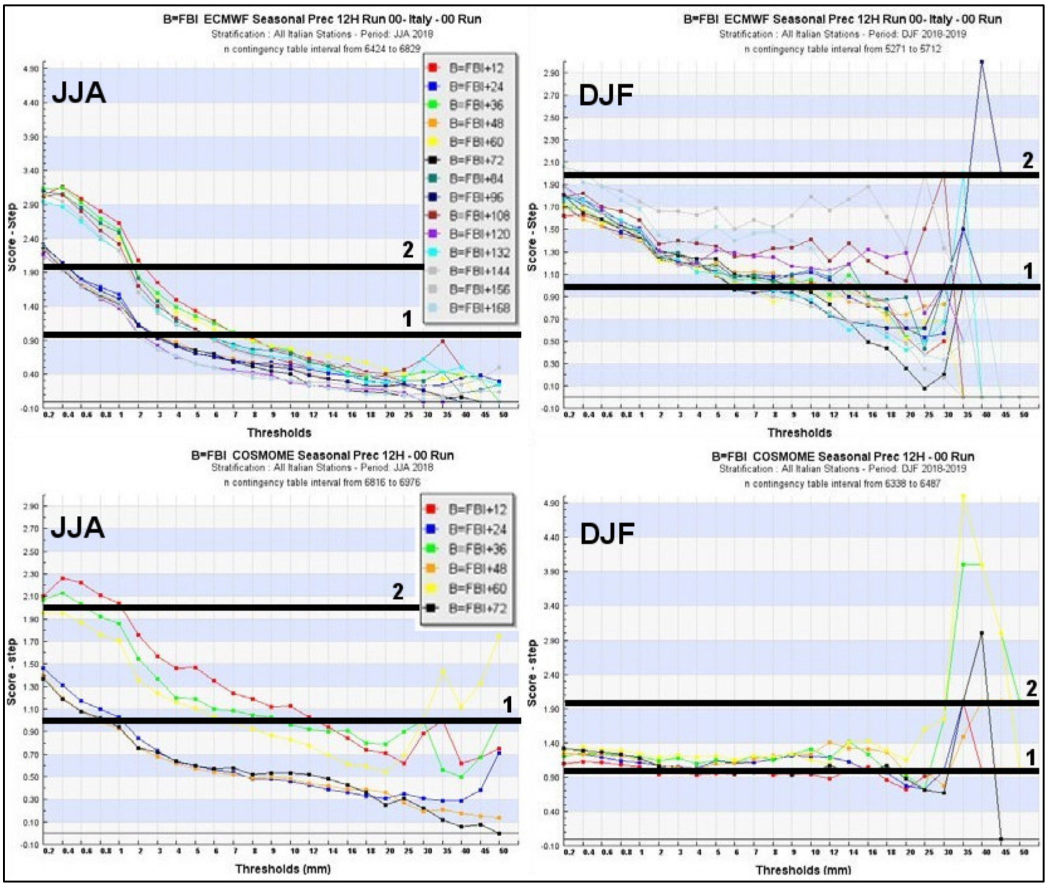

Figure 9: Frequency bias (FB) as a function of rainfall threshold (mm) for 00UTC forecasts from HRES (top) and COSMO-ME (5km) (bottom) during summer 2018 (left) and winter 2018-19 (right) using gauge observations across Italy as truth. High total FBs are noisy due to small sample size. Thick black lines highlight FBI=1 and 2 for comparisons (scales differ). Lead-time legends apply to their respective rows.

For HRES (and ENS) the profile of FB versus totals over a period (typically 6, 12 or 24h) suggests over-prediction for low totals, which drops to under-prediction for high totals, meaning that there is a crossover total where FB=1 (as on Figure 9). Greece, Hungary, Italy, Turkey and Switzerland all demonstrate this behaviour for HRES. The crossover might be ~10mm/24h, though varies quite a lot according to region and time of year (and of course the rainfall accumulation period). In summer when more rainfall is convective the crossover value is typically smaller and more sharply defined. These effects are clearly illustrated on the top two panels in Figure 9. The lower panels for a COSMO (5 km) model show similar behaviour, although FB values there are closer to 1 overall, notably in winter. These COSMO plots perhaps give a flavour of what FB we can expect for unadjusted output from a future IFS, when horizontal resolution has become finer.

Whilst underprediction of the frequency of high totals is not necessarily a model bias, systematic overprediction compared to point totals, a characteristic demonstrated by Serbia and Hungary for their LAMs (e.g. 1.4 for 30-40mm/24h), very probably does denote a model problem.

Another noteworthy feature of Figure 9 is the FB profile split for both models in summer. The split is between validity times of 00-12UTC (top cluster), and 12-24UTC (bottom cluster) and arises for HRES because of documented errors in the diurnal cycle of convection (presumably the same applies for COSMO). In relative terms we have over-prediction early in the day (higher FBs) and under-prediction later in the day (lower FBs), because convective rainfall tends to develop too soon and decay too soon in models with parametrised convection. Results from Greece, Spain and Sweden also highlight this issue; Spain say that the anomalous shift in the diurnal rainfall peak is 3h, whilst Sweden say also that the amplitude of the rainfall diurnal cycle is too great.

Assessing the performance of IFS versions alongside LAMs remain challenging, more so than for other parameters, because there are more metrics one can use, and these may give different answers. One bias-independent measure of skill is SEDI, which can be used to better intercompare model skill for high totals. Hungary again show SEDI plots comparing HRES with 2.5km and 8km LAMs, and show that up to ~25mm/24h HRES skill is comparable; only after that does it drop away. Whilst this looks much better than the 10mm/24h they reported last year, sampling implies large error bars on these values. Meanwhile Serbia suggest that HRES beats their 4 km models for 2 and 20mm/24h totals, and Norway state that (subjectively) HRES is better than the MEPS (2.5 km) LAM EPS system for convection over the sea. However Finland, which also use MEPS, say that it is often better than ENS for precipitation in general. Using the Pierce score France indicate that over international island domains AROME-OM is worse than HRES for small totals, but better for >6mm/6h.

Whilst Spain say that HRES appears to have an upper limit of 60mm/12h we would re-iterate that the IFS has no inbuilt ceiling on totals, although underestimates of point-scale peaks will be commonplace in convective regimes.

As in recent years Sweden use the Fractions Skill Score, and pooled high-density observations, to compare AROME and HRES with rainfall averaged over *gridboxes* of differing size (≥20x20km). Significantly this shows AROME to perform better for all sizes and all thresholds considered.

Austria have this year also used as verification totals averaged over regions (measuring 20x20km, over Austria, from their INCA system), to assess AROME (2.5km) and HRES. They sub-divide cases by flow regime: SW'ly, NW'ly and slack flow. The best forecasts occurred with slack flow (presumably because hard-to-capture orographic effects are then minimised), whilst in SW'ly regimes HRES typically overestimated large totals, perhaps because of too much moisture spilling over the Alps. Overall, AROME was better for large totals and HRES for small (using the equitable threat score).

Austria also compared the performance of their 2.5 km 17-member LAM EPS system with ENS, for July 2016, for 1h precipitation up to T+33, using 1km INCA analyses as truth. Although this is old data and an old IFS version the ENS resolution then was the same as it is now. Using Brier scores for >0.3, 3 and 5mm/h, and the Ranked Probability score they showed the LAM EPS to be clearly better than ENS, although the difference was more marginal for the highest threshold. This performance difference is as one would intuitively expect; lack of members and lack of post-processing are likely more of an issue for the LAM EPS at the higher threshold. It would be interesting to also see how the systems compared in the recent past.

So, the picture we get from comparing HRES with LAM performance is mixed. The Swedish results are quite compelling, in showing one LAM to be uniformly better, but these are not consistently backed up in other reports. Recall also that ECMWF introduced a new experimental 12h point rainfall product into ecCharts in April this year. This aims to correct for frequency bias issues (e.g. as shown on Figure 9) and discriminate better when very large totals are actually possible. The current operational version caters only indirectly for topographic influences and does not account for diurnal cycle errors. However, a new 6h point rainfall product under development in the MISTRAL project accounts directly for topographic complexity for large scale rainfall, and also uses local solar time to try to address diurnal cycle errors. Initial verification results over Italy suggest that for large totals skill levels for the 12h product exceed those of raw LAM EPS output. So, with the new refinements for the 6h product we hope to see even bigger improvements. Thereafter it will be interesting to see how *post-processed* LAM output compares.

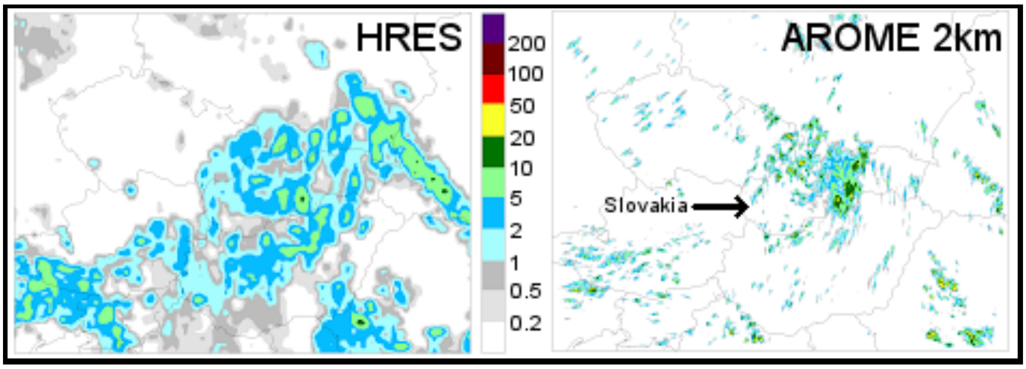

Figure 10: Precipitation totals (mm) for 3h ending T+36 = VT 12UTC 19 June 2019, from HRES (left) and the AROME 2km LAM (right). Areal extent of rainfall is too large in HRES but more realistic in AROME.

As subjective feedback, Finland make many comments on convection. They say that precipitation from elevated convection is difficult for the IFS to handle, which mirrors feedback regularly provided of late from several semi-arid regions around the world. Special investigations have been carried out in-year at ECMWF and continue. Representing the *localised* sub-grid scale lower-tropospheric moistening that allows precipitation to reach the surface in such situations is challenging. Ultimately there may be scope to use lightning diagnostics to help anticipate such errors, subjectively and/or automatically via post-processing. Finland also re-iterate that the inability of the IFS to move convective features such as MCSs is limiting, and may even affect the next day's forecast. Also highlighted is the proliferation of overly large areas of convective rainfall. This concurs with feedback from Slovakia (see their Figure 10 example), Norway, Israel and Italy. Finland recommend using an ingredients-based approach to predict convective rainfall. We agree that this can be helpful, and accordingly provide related diagnostics such as the CAPE-shear EFI, and we recommend referencing the point rainfall in addition. In spite of the (understandable) negative feedback points from Finland, it should be highlighted that they nonetheless still view the forecasts of major convective outbreaks quite highly, stating that: "ECMWF performance regarding deep moist convection is rather good".

Greece again comment on several perceived systematic biases in IFS rainfall, of different sign, in different synoptic situations; for details see their report. Spain meanwhile note the known, and somewhat inevitable IFS issue of underestimation of orographic enhancement. And Israel re-iterate how the known and innate IFS tendency to not propagate SST-triggered convective cells inland can lead to very large errors for them. In one January example with onshore flow there is a clear coast-parallel band of under-prediction inland, with a forecast:observed ratio as low as 15mm:60mm (24h totals).

Norway state that HRES is over-forecasting snowfall, in terms of horizontal extent and (presumably related to this) the snow fraction in marginal snow-rain situations. One factor may be that snow fraction in deep, narrow valleys (and fjords) is commonly overdone because true valley base altitude will be lower than it is in the IFS. The issue is also made more apparent by valleys being where people live. In any case ECMWF would welcome a case list from Norway to support their comments.

3.1.4. Screen-level humidity

Six of the NMS reports referenced screen-level humidity verification, as is or via dewpoint. Although this is not a huge sample the regions covered were reasonably diverse: Sweden, Hungary, Greece, Germany, Belgium and French island territories overseas.

The reports painted a very favourable picture regarding HRES performance. Biases were generally reported to very small, typically ~0.5° C in dewpoint temperature, and generally slightly negative. By day in Greece (outside winter) the bias drops to -1° C. Sweden were very complementary, highlighting "small systematic errors during the whole year … valuable guidance regarding 2m moisture".

For Greece forecasts provided by HRES are substantially better than their local LAMs; whilst for Sweden and Hungary HRES performs a bit better. For the French islands the LAMs fare a little better than HRES, whilst Germany show that ICON has slightly smaller RMSEs overall, the most notable difference being in winter (See Figure 7).

3.1.5. Cloud and Irradiance

This year we combine cloud and irradiance into one section to recognise that solar power in the growing renewables sector depends on radiation, and that instrumentation and analyses are increasingly incorporating irradiance measurements to infer cloud cover. Four out of the eight countries that contributed some cloud verification refer to use of radiation measurements/forecasts in one form or another.

In the vicinity of mountainous areas (here the Alps) both Austria and Switzerland demonstrate spring/summer over-prediction of amounts. Within Austria the overprediction is much greater over the "flatlands", though is present also in mountainous regions. The IFS' diurnal convection time-shift mentioned previously is thought to be partly responsible. Austria comment that in the mountains the daily build-up is more regular and predictable, and that this is well-handled. Probably this build-up is largely fuelled by repeated moistening from a relatively abundant and persistent local moisture supply (which in turn comes partly from snowmelt). Figure 11 from Switzerland provides better evidence of a diurnal cycle shift (~2h difference in the daytime peaks), and indeed shows an overprediction bias around-the-clock.

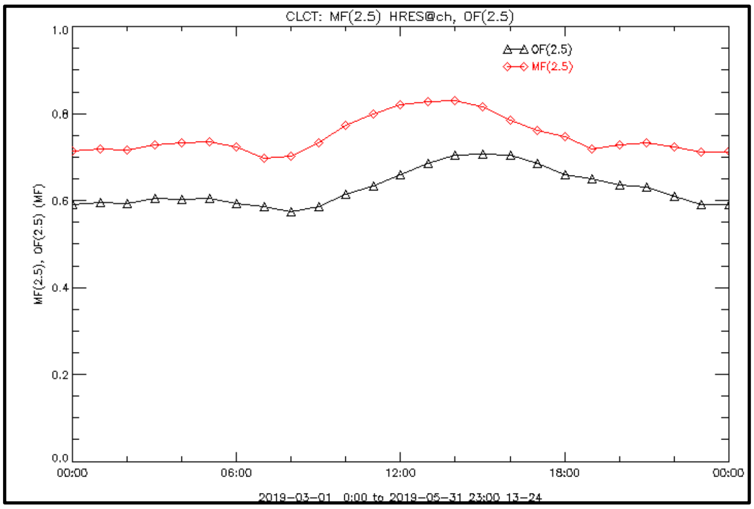

Figure 11: Diurnal cycle of the frequency of occurrence for hourly cloud amounts above 2 oktas: HRES (red) and observations (black), spring 2019, at Swiss stations.

In the 100-day sequence from Austria (not shown) forecasts from LAMs (over two 100x100km boxes) were overall remarkably similar to those of HRES, and indeed the performance levels are similar. For HRES this seems to be a particularly noteworthy achievement, given its relative lack of topographic detail.

Results from Belgium and Greece have similar characteristics. Biases were very small - e.g. <5%, with just a slight underprediction bias for the Greek stations. Belgium do not compare with LAM output, but Greece show HRES to have biases and RMSEs that are mostly better than their LAMs. These LAMs exhibit a large diurnal cycle in bias which is not present in HRES.

Using irradiance metrics Spain show that the 2.5 km HARMONIE-AROME model and HRES have similar performance, although they also state that their LAM EPS has a much better spread at short range than ENS which is under-dispersed for cloud cover.

Norway comment that HRES handles cloud cover better than the MEPS 2.5 km LAM EPS, especially in onshore flow, due to MEPS having dry bias and HRES not.

The aforementioned results are mostly very positive for HRES, but conversely Hungary report again this year that that for lowlands HRES has much worse RMSEs than their 2.5 and 8 km LAMs, partly because of a big underprediction bias (~20%). They pool together data for the whole year, so it is difficult to understand this bias, although one can speculate that it may be largely due to winter stratus under-prediction. This would be consistent with the assertion in Germany's report that cloud cover is under-predicted by HRES in winter.

Meanwhile in their global model inter-comparison Germany show that HRES beats ICON for cloud cover at all leads and all seasons, the net difference being larger by day (Figure 7).

Israel deal with the separate topic of "cloud ceiling", which technically means the base of the lowest clouds that cover at least half of the sky, which can be very relevant for aviation. Their results, for leads of 24-83h, for HRES and Control runs, show, for the Bet Dagan ceilometer site near Tel Aviv, poor agreement with observations (Figure 19). Observed values are commonly in the range 500-1000m, whilst forecast values are more wide-ranging, and higher on average. ENS is also shown to exhibit poor skill in forecasting ceiling at this site, using several metrics; climatological forecasts (for this ceilometer) are in fact better, Whilst cloud base can be challenging to predict even at relatively short ranges, and whilst these results relate to older cycle 43r3, and whilst proximity to the coast (8km away) and local topography (~50m) may be of relevance here the large discrepancies seem to warrant further investigation by ECMWF.

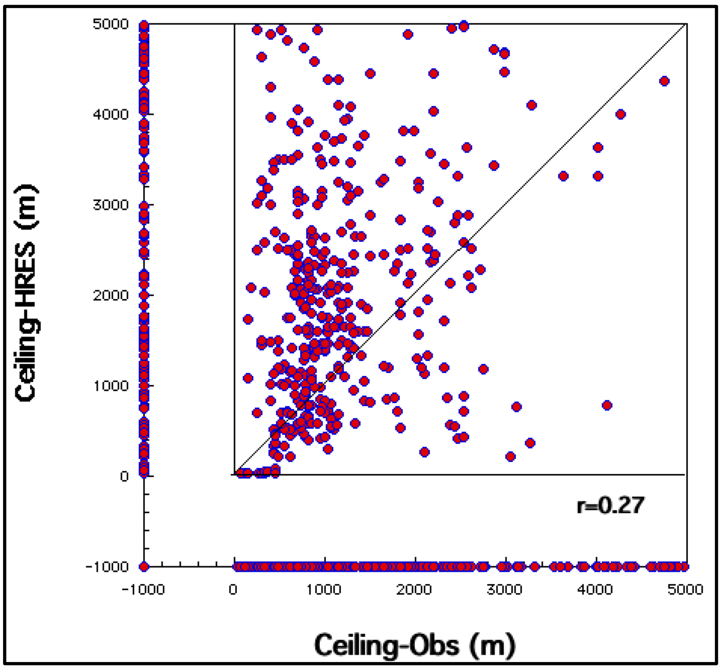

Figure 12: Comparison of HRES and ceilometer ceiling. Each dot represents an hourly ceilometer data centered at the model forecast time. Period represented: Jan through May 2018. Ceilometer is located near model grid center. Clear sky and values above 5000 meters are represented by a value of -1000m. Correlation coefficient is calculated for positive values.

3.1.6. Visibility / Fog

We have received much less feedback on visibility and fog than last year: statistics have been provided by Finland and Latvia, whilst Sweden and Israel make brief reference.

Historically, forecasters have used the concept of a "fog point" to signify the level to which 2 m temperatures has to drop for fog to form, so clearly temperatures and visibility are intimately linked. Thus the visibility results from Finland, where we know there can be systematic large positive 2 m temperature errors in settled conditions, when fog is more likely (see sub-section 3.1.1 above), may be heavily impacted. On the one hand Finland report that HRES "visibility is often forecast to be better than observed" (backed up by a figure, not shown) and that "fogs are frequently missed", which is of course undesirable but is at least consistent with the 2 m temperature errors. On the other hand, they later say that their post-processed visibility product exhibits errors of the opposite type - visibilities (at T+24 = 04UTC) in autumn and winter are more than twice as likely to be too low as too high. Probably this apparent contradiction is highlighting how difficult it is to get visibility forecasts correct, and to adjust for errors with post-processing, especially when temperature errors can also be very large. It is reasonable to expect that in other countries where the temperature errors are less visibility forecasts might as a result be somewhat better. Sweden comment that fog (and low cloud) in the AROME 2.5 km LAM are over-forecast in spring because dewpoints are too high, and that HRES does not have this issue. The spring bias difference in dewpoint between AROME and HRES is however only 0.4C. So, again, we have some evidence of high sensitivity of visibility to small changes in another variable. Conversely Israel appear disappointed with our fog forecasts, attributing errors to IFS mis-representation of different types of inversion.

Latvia's results suggest that Harmonie visibility forecasts are better than HRES (Figure 6; a result supported by separate analysis by Iceland last year). They also demonstrate that visibility errors in HRES tend to grow quite rapidly with lead time, which is as expected, although their plots show a positive bias that does the same, which is more surprising. A complication in visibility forecast assessment in general, that may be affecting results here, is the scaling; to address this, users could consider working with the logarithm of visibility rather than absolute values.

3.1.7. Sea Ice and Snow Cover

Finland again refer to sea ice issues in the Baltic, highlighting how, in winter 2017–18, ice cover forecast by ENS at extended ranges fell systematically well short of the observed coverage. They suggest that this is likely due to analysis issues. We would agree. However, in cycle 45r1 that went live in July 2018 (after Finland's example) ECMWF introduced weakly coupled data assimilation, which can go a long way towards correcting these analysis errors. Figure 13 below, adapted from a Summer 2019 Newsletter article by Browne et al (https://www.ecmwf.int/sites/default/files/elibrary/2019/19165-coupled-ocean-atmosphere-data-assimilation-ecmwf.pdf), clearly shows how much better the analysed sea ice in the Baltic can be when this assimilation approach is used.

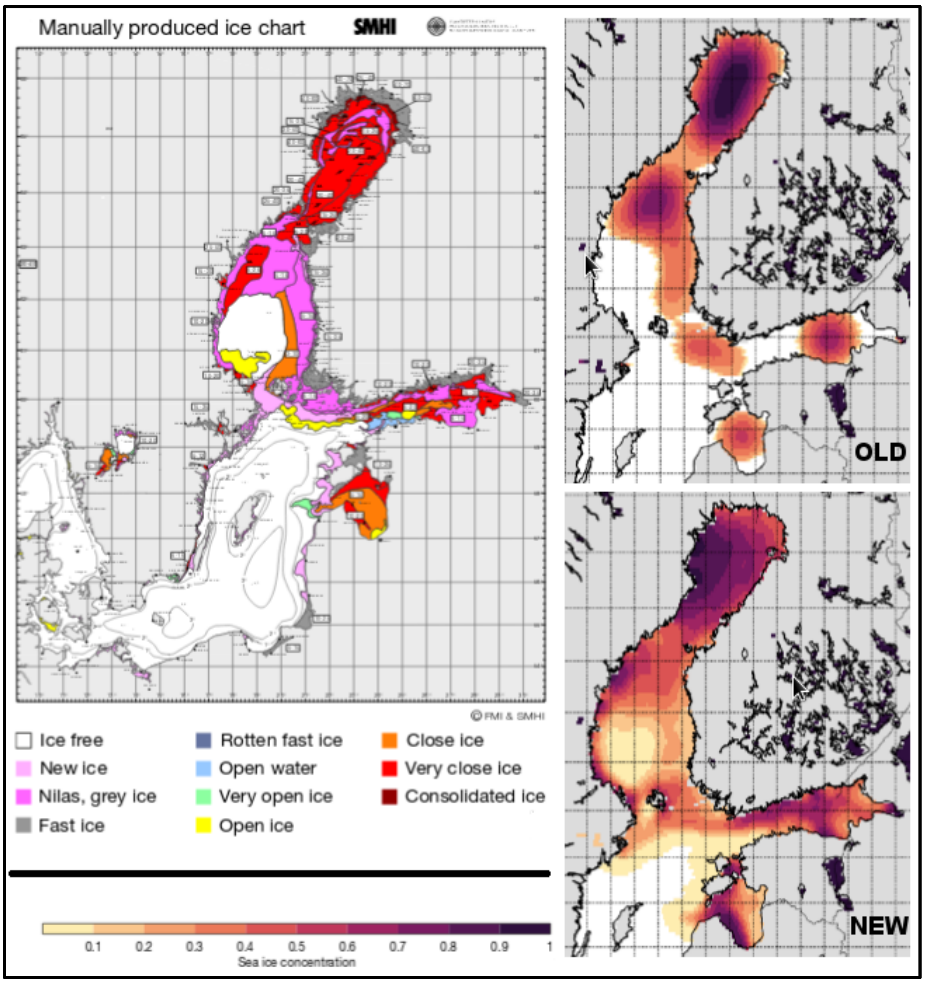

Figure 13: Data for 5 March 2018. Left: a manually produced Finnish-Swedish ice chart of the Baltic Sea (copyright FMI and SMHI, 2018); top right: uncoupled ECMWF sea ice concentration analysis; bottom right: ECMWF sea ice concentration analysis using weakly coupled data assimilation.

Finland also made a confusing, but interesting and potentially pivotal comment about snow depth reporting. They state "we know that during spring even when the observation shows 30 cm it means that e.g. fields and all open areas are snowless". If the observation is manual then should it not be representative of open areas? There is also a state-of-ground indicator in SYNOP messages that can denote the proportion of ground that is snow covered (e.g. > or <50%). Is it acceptable for that to indicate <50% when the depth is 30 cm? Answers to these questions have the potential to help ECMWF better assess snow depth, and in turn address a co-located IFS spring-time cold bias that could be explained by model snow cover being too extensive. In one example Finland show that HRES minima and maxima were systematically up to 8C too cold over several days. A spring-time screen-level dry bias that they reference separately may also relate.

3.1.8. Other

This sub-section relates primarily to broader-scale metrics that define the synoptic pattern.

Greece show that for mean sea level pressure forecasts over Greece HRES clearly outperforms their two LAMs, for bias and RMSE. In winter for example HRES 72h forecasts are, remarkably, as good as the LAM forecasts at 18h. There appears to be drift with lead time in LAM mslp values, in spite of IFS BCs being used.

Germany provide a nice “deep atmosphere” comparison between ICON-EPS and ENS for 2018-2019 for the N Hemisphere (Figure 14, for T+48h). The most prominent features are better performance of ENS in the troposphere, for all variables shown, and worse performance of ENS in the stratosphere for most variables. The latter aspect relates to well-documented stratospheric problems in the IFS, which will be addressed using “quintic interpolation” in cycle 47r1 in 2020 (see: https://www.ecmwf.int/en/elibrary/19084-control-stratospheric-temperature-ifs-resolution-and-vertical-advection).

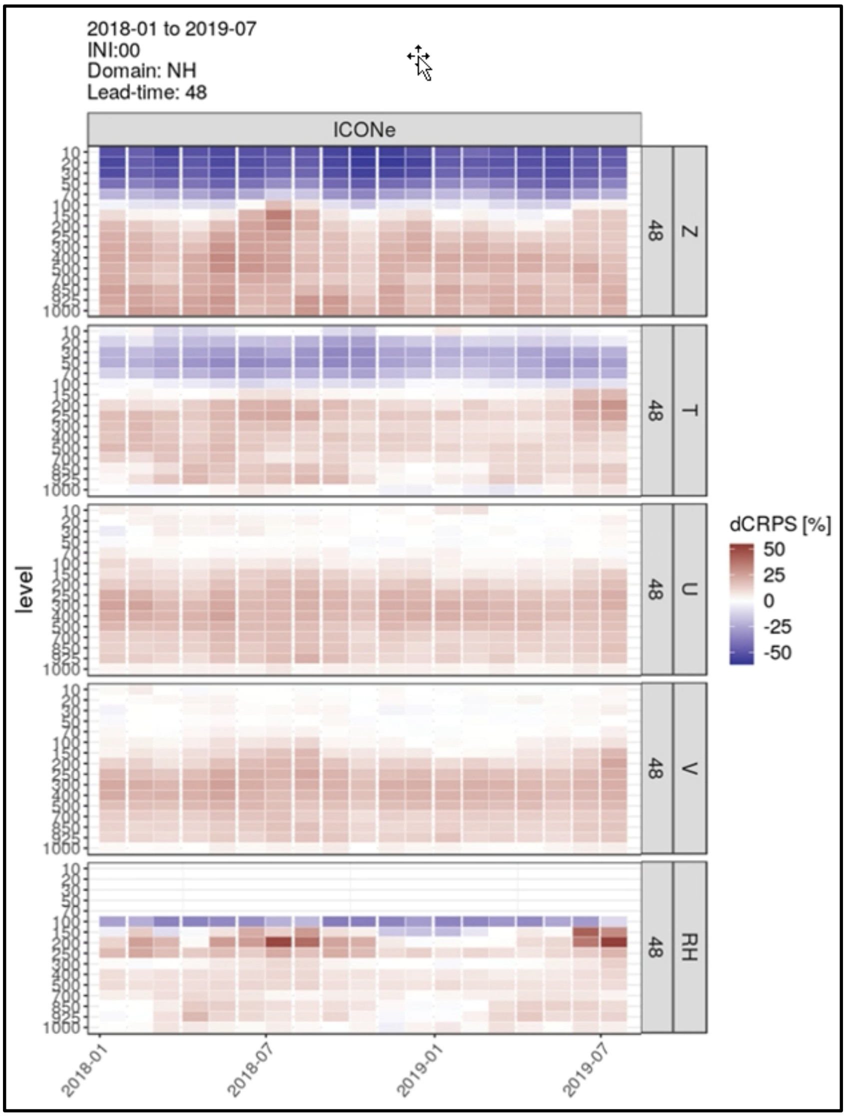

Figure 14: Monthly relative differences in CRPS between 00UTC ICON-EPS and ENS runs in the N hemisphere for different levels [hPa] at T+48h. Brown denotes ENS better, blue ICON-EPS better.

Germany also note that relative to the equivalent IFS components, ICON EPS has been catching up less over time than ICON deterministic. Indeed, during the two years shown on Figure 14, no trends are apparent in the EPS differences.

Using their “Decider” system the UK have compared the ability of two different EPS systems to capture different UK-relevant weather regimes over days 1-15. They show that ENS performs slightly better overall than the American “GEFS” system. Also noted is the relative ease or difficulty of predicting different regimes, which seem to be common to the two systems. The “most predictable” seems to be “blocked NAO-”, whilst two of the least predictable are the “Scandinavian blocking high”, and “Anticyclonic westerly with Azores extension”. They also note that, for both models, and according to their metrics, skill is better in winter than in summer.

We continue to see adverse comments regarding IFS jumpiness, throughout the year, in different fora, and in this year’s reports from Spain, Finland and Norway. Whilst ECMWF continues to investigate jumpiness issues the criticisms we receive are not always justified, and users are referred to the Forecast User Guide pages on jumpiness (https://confluence.ecmwf.int/display/FUG/7.2+Jumpiness) which might help clarify some aspects.

France provide interesting verification statistics, relating in part to jumpiness, based on forecasters’ subjective assessments for one data time per day (Figure 15). The frequency of ENS jumpiness (Figure 15a) is considered greater at longer leads, and if one counts red as genuinely “jumpy”, then at day 4 one gets jumpy forecasts about 7x per year (2x1%), increasing to once every 5 days at day 8-9. The day 8-9 frequency seems quite high and, one would hope, links primarily to a sharp but correct reduction in spread as the outcome becomes more certain. Figure 15b quantifies the practical advantages of using ENS rather than HRES. Beyond day 6 ENS gives the user a lead time gain over HRES of about 1 day. Moreover, when a new HRES forecast comes in (before the new ENS) the *previous* ENS is still more likely to represent the correct outcome for lead times ≥ day 6. This seems to be a result that could be usefully incorporated into the Forecast User Guide.

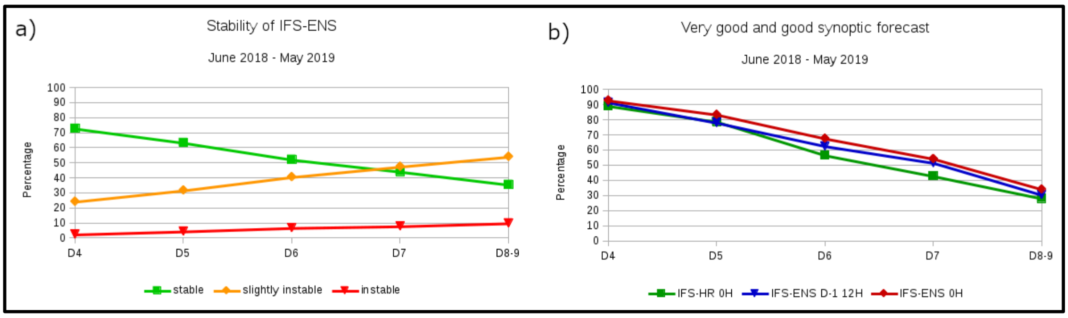

Figure 15: Subjectively verified IFS synoptic scale performance over France from Jun 2018 to May 2019 as a function of valid day. (a) Stability between consecutive 12UTC and 00UTC ENS runs (≈ jumpiness) as ranked by forecasters, in 3 categories (see legend). (b) Percentage of “very good” and “good” marks for HRES 00UTC (green), ENS D-1 12UTC (blue) and ENS 00UTC (red).

Greece comment at length on model performance in different synoptic situations, highlighting for example potential issues with IFS handling of over-ocean transitions between stable and unstable air. Please refer to their report for more details.

3.1.9. Forecaster Impact

A few countries verify or at least comment upon forecaster performance compared to models. Whilst performance gaps are small, Hungary again show that forecasters continue to add a little value, on average. This added value in a composite, multi-parameter index continues to equate 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 about day-3. At longer leads "ENS mean" and "forecaster" are about equally good, and remain the best options on average. The forecaster improves cloud cover forecasts at shorter lead times. However, they add most value overall to maximum and minimum temperatures in the summer half of the year; the gain diminishes with lead time but remains positive even at day 6. To achieve these longer lead benefits the forecasters are mainly removing systematic biases, which can be large.

Turkey say that their subjective verification shows that continuing improvements in HRES are helping their forecasters deliver better precipitation forecasts. Denmark meanwhile show that their forecasters are continuing to provide better 2m temperature forecasts than HRES out to day 5.

3.1.10. 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 (i) model, (ii) reality or (iii) external data (such as site altitude, or population density). It underpins many insightful subjective comments provided by forecasters over the years - e.g. "in this situation the model tends to get this wrong". Objective conditional verification is a growth area that holds a key position in ECMWF's plan of work for verification. If model-based subsets (i) are used, as in ecPoint (see sub-section 2.1.4), value can be added through post-processing, and physical meaning can be given to certain types of bias/RMSE behaviour, which can help with improving the model formulation, particularly the physics. If reality (ii) is used, physical insights can also be gained, though in self-referential conditional verification (e.g. "winds are under-predicted when *they are* strong") caution is required to not over-interpret results (see sub-section 3.1.2). In this category, but on the topic of the wave model, Spain report that when wave height is over 6 metres near the Atlantic coast our wave model forecasts underestimate. Reference to external factors (iii) can demonstrate when sub-optimally represented aspects of the physical environment are leading to errors, such as when the ability to predict 10 m winds and gusts over "mountainous areas" is being assessed (again see sub-section 3.1.2).

Whilst noting that conditional verification topics are scattered through this report, here we refer directly to just one study, by Finland, relating to the large winter-time 2 m temperature errors that they experience (see sub-section 3.1.1 above).

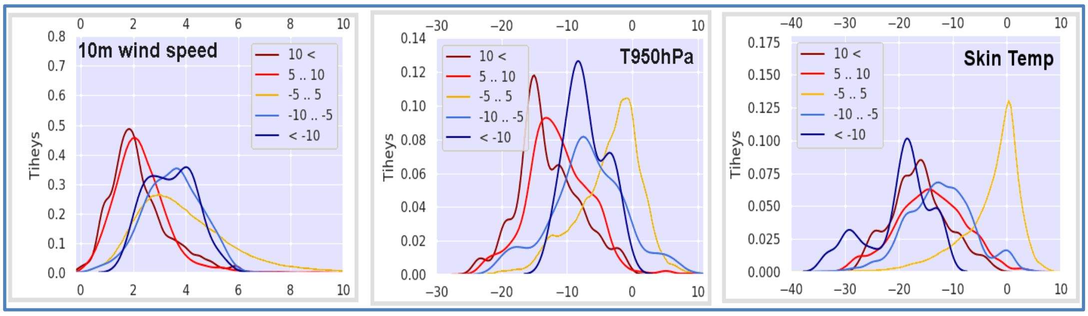

Figure 16: Density plots for 10m wind speed, temperature at 950hPa and skin temperature, for 5 different 2m temperature forecast bias categories; where blue curves denote model forecasts too cold and red denote model forecasts too warm. Data is for southern Finland, spanning 5 years, lead times <48h.

Figure 16 shows that large positive temperature errors (red) in southern Finland most commonly occur when 10 m wind speeds are light, and the 850hPa temperatures are low, which together suggest settled, cold airmass conditions, when more extreme "unmodelled" inversions are more likely to occur. The rightmost panel, for skin temperature (which is one of the components from which 2 m temperature, which is a diagnostic quantity, is derived), shows a strong signal for small errors to preferentially occur when skin temperature is around 0C. This will be due "temperature pegging" during episodes of melting snow, when energy is used not to change temperature but instead for the latent heat of fusion. So, there is nice evidence here not only that 2 m temperature errors - random and bias - depend on other variables, but also that one could predict these errors a priori using forecast values of those other variables. This has parallels again with the ecPoint initiative, and statistical temperature correction in a similar way will be explored within the ECMWF component of the EU-funded HIGHLANDER project due to start in October 2019.

3.2. Post-processed Products and End Products delivered to users

Within the NMS reports five countries highlight verification activities in this category; most show notable positive benefits although Finland's results show that forecasts can be degraded too.

Switzerland report on a system to deliver automated post-processed forecasts for any point in Switzerland, using ENS and COSMO EPS and deterministic forecasts as input. Two years of 2 m temperature verification show impressive results. For example, day 15 post-processed ENS forecasts are on average better than day 1 raw ENS forecasts at the verifying sites. Meanwhile for days 1-5 CRPS values for ENS and COSMO EPS average ~3.4 and ~1.9° C respectively; whilst for the multi-model post-processed product this reduces to ~1.1° C. They show that both systems contribute, vindicating the multi-model approach. Although the UK provide no results this year, in many of their activities, at multiple time ranges, they are also actively pursuing a multi-model approach.

In other initiatives Spain report that post-processing of ENS irradiance fields delivers much better CRPS scores in the short ranges, that are then comparable to their LAM EPS system, whilst Turkey report that Kalman filtering improves forecasts of 2 m temperature extrema by 5-25%. Meanwhile Germany indicate that they verify "autoTAFs" that use IFS inputs, though provide no new scores.

Finland have performed Europe-wide calibration of 2 m temperature and wind forecasts, using for this "Gaussian" and "Box-Cox t-" distributions respectively, and using as predictors, from the last 30 days, ENS mean, ENS spread, analysis clock time, lead time and station elevation. Forecasts are better over Europe as a whole, and a Finland example they provide, out to day 15 during April 2019, shows reduced errors and an improved spread-skill relationship (as reported last year). However, at other times of year in some regions, such as Finland, a large bias is unfortunately made even larger through post-processing, as illustrated on Figure 17. Perhaps inclusion of other factors such as snow cover/depth could help, although in Europe-wide calibration the sample size might be limiting.

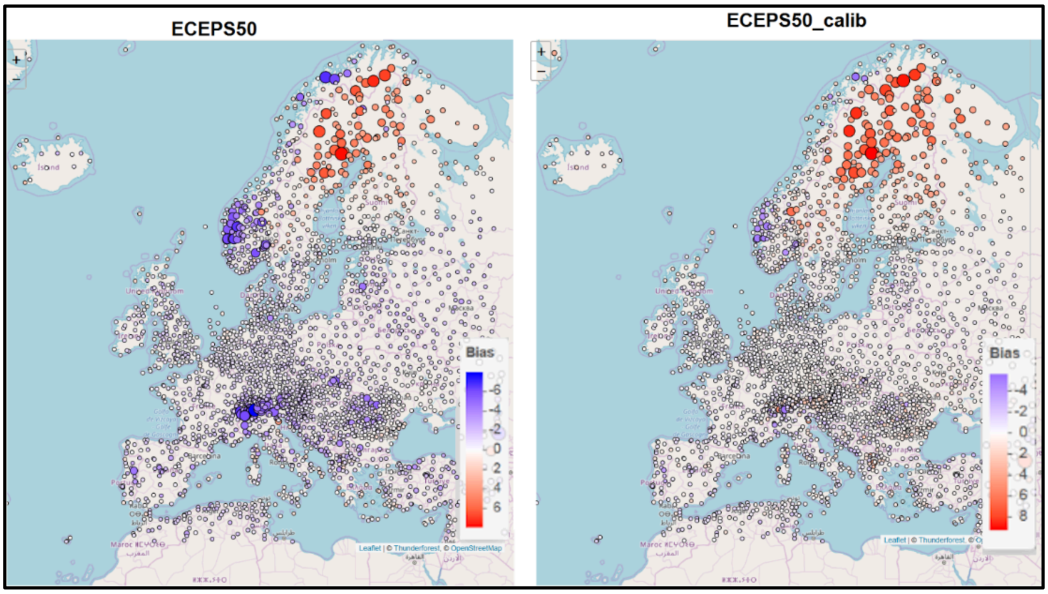

Figure 17: Station-based mean bias for raw (left) and calibrated (right) 2m temperature forecasts (°C) in January 2019 (note that the scales differ).

3.3. Extended Range (Monthly) and Seasonal Forecasts

In feedback gathered prior to UEF 2019 interest shown in these topics was sufficiently high for one of four breakout groups to be devoted to them. However, in formal NMS reports there was less feedback than last year in these categories (possibly because of the reduced page limit). NMSs generally commented on monthly forecasts alone, or monthly and seasonal together - hence the section sub-division below.

Relating to Extended Range (Monthly) only

The UK use both ENS and their own GLOSEA modelling system for formal extended range guidance, and products regularly used from both include UK-centric regimes from the 'Decider' post-processing system (see also sections 2.2.2 and 3.1.8). The UK have compared verification metrics for these. Overall ENS forecasts perform better than those from GLOSEA, with the Brier Skill Score zero line reached around day 16 in ENS and day 14 in GLOSEA. Some regimes seem more predictable than others, and the two systems broadly "agree" on which they are. The UK also say that inter-comparison is problematic because of different update configurations. In this regard note that ECMWF is looking into the possibility of a different configuration with more regular updates in future, though nothing is decided yet. The UK also cite the "Beast from the East" at the end of winter 2018 as being a nice example of predictability sometimes extending well beyond day 15 (see the 2018 version of this report for more discussion). We would add that the end of winter 2019 provided a counter example of a strong multi-model signal for somewhat similar conditions beyond day 15 that proved wrong.

Switzerland produce Europe-wide forecasts, for weeks 1 to 4, for European workers in the H2020 Heat-Shield project, by post-processing ENS forecasts of wet bulb globe temperature. Quantile mapping delivers some skill beyond climatology out to week 4 (days 26-32), whereas raw model output skill drops to zero by week 2 (days 12-18) - Figure 18. This is another example of how post-processing can add value to temperature-related forecasts, aided here by having available the 20 years of ENS re-forecasts that are used for calibration. Note also that quantile mapping outperformed EMOS.

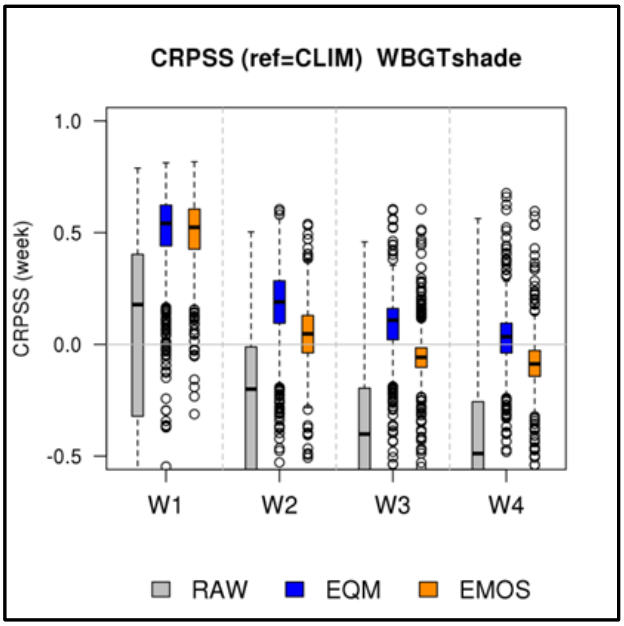

Figure 18: CRPSS for forecasts of wet-bulb globe temperature for April to September 2018 for ~1500 European stations. Reference point is climatology. ENS forecasts of daily maximum and dew-point temperature were calibrated with empirical quantile mapping (blue) and EMOs (orange) using all available re-forecasts. Box-and-whiskers denote skill for individual stations. Week 1 means day 5-11.

Germany generate some of their own monthly forecast indices (e.g. a "dry spell index") whilst Serbia have created a suite of Python tools (using also ecCodes and Magics) to adapt monthly forecasts to local needs. Readers are again reminded of the new ECMWF "Metview-Python" framework (which includes Magics and ecCodes) which should be well suited to activities such as this.

Meanwhile Norway report that they provide monthly forecasts to energy suppliers and flood forecasting authorities, with a short text component added by forecasters. In common with several other countries they also say they like the look of new extended range products.

Relating to Extended (Monthly) and Long Range (Seasonal)

Romania is very active in generating output pertaining to monthly and seasonal timescales. They create many of their own indices, and also weather regime metrics for both ranges, and focus operationally on regime transitions (a key tenet in ECMWF's 10-year strategy). They also perform 4-member dynamical downscaling of seasonal forecasts using a 10 km LAM, for various purposes. Latvia are also active in downscaling at the extended ranges; they create hydrological simulations based on ECMWF forecasts that extend as much as 1 year ahead. Denmark are also very active in using extended range outputs and make several related requests - see section 4.

Finland use monthly and seasonal output for their sea ice prediction services (for the Baltic) whilst Austria, though not big users, do create their own climagrams for specific points.

Hungary are active in both spheres and provide graphical examples of locally generated meteograms for monthly time scales, including weekly-mean wind speed and cloud cover that are not available from ECMWF. They have also assessed seasonal forecasts for Hungary verifying in 2018. Rainfall exhibited no skill even in month 1. Temperature forecasts were relatively good in month 1 and retained a modicum of skill (possibly related to climate change) out to month 6. This was best for maximum temperature.

3.4. Case Studies

Several centres report on specific severe weather events that have affected them in the last year or so. Many make very positive comments about IFS performance, whilst others focus on less satisfactory aspects. One case each is highlighted by France (extreme Mediterranean cyclone), Austria (extreme Mediterranean cyclone), Spain (gust over-prediction in strong easterly flow), Slovakia (over-estimation of areal extent of precipitation - Figure 10), Romania (downscaling for a flooding event), Norway (snow and strong winds); 2 cases by Turkey, at the UEF meeting (warm front snowfall, a new 24h rainfall record) and also by Greece (medicane, unsatisfactory precipitation forecasts); 3 cases by Montenegro (tornadic convective outbreak, broadscale floods, flash floods); and 11 cases by Italy (flash floods with some hail: 4, general precipitation over/under-estimation: 3, medicanes: 2, broadscale errors: 1, extreme Mediterranean cyclone: 1). Italy, Austria and France referenced the same extreme Mediterranean cyclone event at the end of October, whilst Italy and Greece referenced the same medicane event at the end of September (see Figure 19). A brief overview of these two events is given below. For more details see also the ECMWF Severe Event Catalogue: (https://confluence.ecmwf.int/display/FCST/Severe+Event+Catalogue), where they are both recorded, and for the other events please refer to the Member State/Co-operating State reports, and to UEF2019 posters at: https://events.ecmwf.int/event/119/page/42-posters .

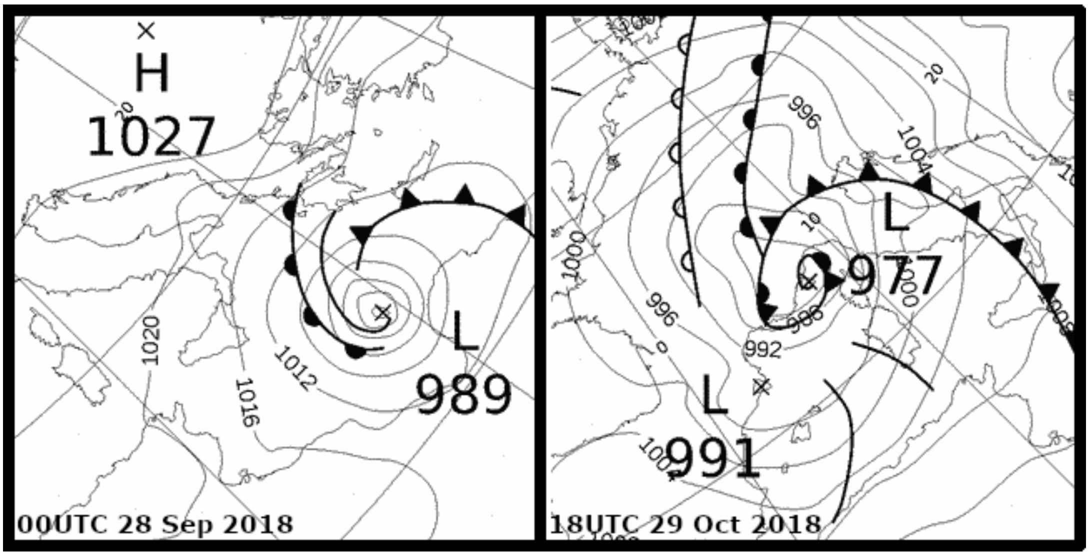

Figure 19: Analysed surface charts from the Met Office for the medicane at the end of September 2018 (left, as referenced by Greece and Italy), and the extreme Mediterranean cyclone at the end of October 2018 (right, as referenced by Austria, France and Italy).

The medicane at the end of September (Figure 19, left) developed tropical cyclone characteristics during its lifecycle, and was characterised by inconsistent IFS forecasts, as Greece highlight. Extreme weather that at one point seemed destined to affect Italy in the end affected Greece more. Some forecast issues may relate to the system originating in a data sparse area near the northern coast of Libya. In addition, correct modelling of air-sea interaction that drives such systems can be challenging. However, HRES was at least better equipped to model these processes than hitherto, following introduction of HRES ocean-atmosphere coupling in June 2018.

For the Mediterranean cyclone at the end of October 1018 (Figure 19, right) France show that their Arome 2.5 km model forecasts were superior to HRES, in capturing extreme winds over Corsica as it passed by, although even these were not strong enough. Figure 20 shows a cyclone 6hPa deeper in Arome, and mean winds about 15 kts stronger, reaching hurricane force (≥64 kts). France say that the explosive development of this system, with a tiny region of extreme winds, presented a particularly acute forecast challenge that only very high resolution models could potentially overcome. Whilst this assertion is not unreasonable, it is extremely rare to see an extreme extra-tropical cyclone for which 9 km horizontal resolution falls short of what is needed to capture the wind structure. For the vast majority 10-20 km is sufficient. France highlight also very large spread in ENS output even at T+12h; so if nothing else this would have urged caution by highlighting a sensitive and “dynamically charged” situation. Italy meanwhile say that HRES actually performed better than their COSMO LAM. Austria, whilst happy overall with IFS performance, also show the limitations of HRES resolution in respect of orographic enhancement; the maximum event total rainfall in the south of the country was ~ 700mm whilst in HRES it was ~ 400mm. Italy describe the event as marking the beginning of a ‘nightmare week’ for forecasting; there were 32 casualties, periods of extreme (tornado-like) winds, exceptional waves, widespread frontal lightning, critical under and over-estimations of extreme rainfall events by IFS and later on a major storm surge in the Adriatic. Red warning issue meant major disruption as public services such as airports and schools were then forced to close. Extreme events such as these provide some justification for the resolution increases targeted in ECMWF strategy.

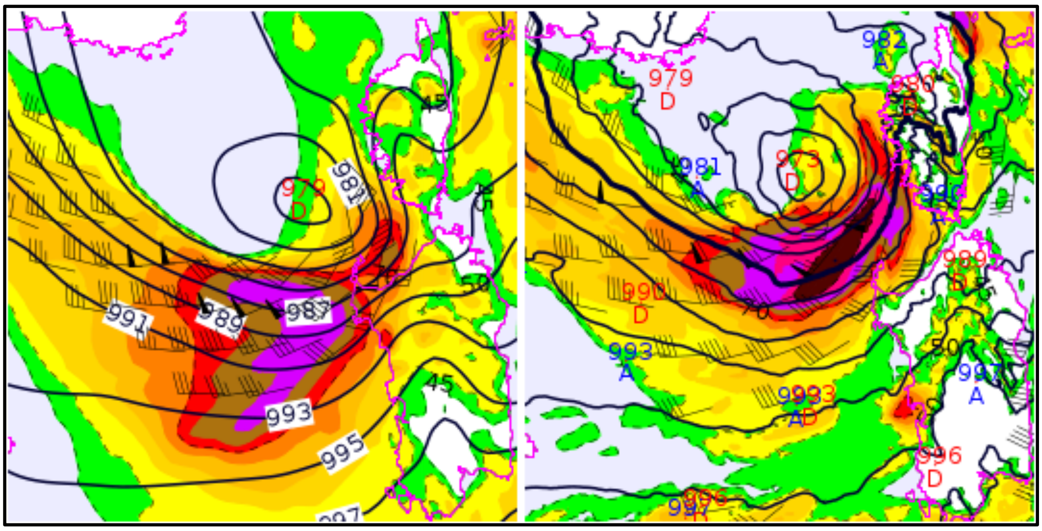

Figure 20: 15h forecasts for 15UTC 29th October 2019, from HRES (left) and 2.5km Arome (right), showing mean winds as flags/shading, and mslp as contours. Corsica and Sardinia are on the right.

Several events in the case list above were flash floods. Readers are reminded that point rainfall products, introduced into ecCharts in April (admittedly after most of the reported events took place) should in future be able to help with probabilistic flash flood prediction.

4. Requests for Additional Output

ECMWF has collated together specific requests highlighted in questionnaire responses submitted for UEF2019, in its User Request Management System (URMS), and is now working on some of these, and prioritising others. Requests that appear in the NMS reports are outlined below. Meanwhile Austria reported that they are happy with our already comprehensive range of products.

Some requests have already been fulfilled by ECMWF, notably Norway's wish to have EFI/SOT fields for IVT and Denmark's desire to see extended range products in ecCharts.

Israel have asked for weekly updates of the seasonal forecast, saying that "once a month is not good enough". It is debatable whether much more regular updates of a costly system with low skill would deliver forecasts that were genuinely more useful.

There were again requests to improve and better document ECMWF's CAPE diagnostics, in reports from France and Slovakia. Following similar feedback from many quarters in previous years ECMWF embarked on a collaborative project with ESSL (the European Severe Storms Laboratory) to look into the representation of CAPE and CIN. This project delivered its recommendations in August and these are now being considered by ECMWF. In addition, the project was successful in identifying a bug in ECMWF's CIN computation which has been rectified for operational introduction in cycle 47r1 in 2020.

Many requests relate to ecCharts. Greece have asked us to facilitate the sharing of saved products (which would be welcomed by many), UK would like Normand's Point construction to be facilitated on HRES vertical profiles, Montenegro, Israel and Italy would like vertical cross-sections, Israel would like adjustable contour settings, Montenegro would also like a tabulation of parameters represented in the Dashboard, Norway want to see legends ported to the Dashboard (without this they understandably say that maps there are of limited use), Israel would like adjustments to animations in Dashboard and Iceland have asked for better help button facilities. Some other ecChart additional product requests, such as BC runs (Portugal), hourly data (Montenegro), archived data (Greece) and seasonal (Denmark) are more problematic for a variety of reasons. It may, however, be viable to show monthly means (from seasonal forecasts), also requested by Denmark, as static web charts.

France have asked us again to try to deliver more continuous surface pressure fields around islands. The reason for the noise currently seen is that the spectral representation of island topography creates spurious 'topographic ripples' in the model representation of sea surface height nearby, which can even impact upon the convective rainfall distribution. These ripples are an artefact of IFS model design, which would require local post-processing to remove, which may not be viable.

Spain have asked us to improve and consolidate the new point rainfall products, bringing in more directly topographic influences. As reported above in Section 3.1.3 we are already starting to do this.

Sweden have again asked us to re-visit our definitions of medium, low and high cloud, to make them match WMO definitions. This seems eminently reasonable.

Israel had several miscellaneous requests, such as hourly radiation fields out to day 15, access to an online trajectory tool, an extension to our standard postage stamp domain so that it includes Israel and inclusion of objective fronts/cyclones in ecCharts. The first two of these will prove challenging; however the third request is relatively straightforward, and Python re-coding that will make the fourth request viable is progressing well.

One of the biggest "wish lists" came from the UK. They would like to collaborate in developing multi-model EPS-related tools, and would also like multi-run products to see how forecasts are evolving with data time. Already ECMWF provides multi-run CDF plots, but of course much more could be done, looking for example at synoptic scales rather than just local weather. Less viable, perhaps, would be satisfying their request to interactively compare real satellite imagery with pseudo imagery, and provide monitoring tools for developing hazards. Both relate more to nowcasting which is beyond ECMWF's remit. And the UK have again asked ECMWF to run a storm surge model and provide related products. Italy also request this. This would be a substantial undertaking.

Two requests relate to the use of re-analyses and reforecasts. Finland have asked for cyclone tracks to be created from ERA5. In principle this could be done using existing cyclone database code, but would require significant manpower. The UK have asked for improved statistics of extremes, to be able to better contextualise aspects of the current forecast. M-climate construction, as used for EFI and SOT, has been active at ECMWF for many years, although it is true that the range of parameters and lead times could be extended. Currently we are looking at computing longer period snowfall M-Climates from re-forecasts. Within the HIGHLANDER project we will also be creating global point rainfall climatologies from ERA5, with which point rainfall forecasts can be compared.

Finally, Germany have asked for more timely documentation of new products, in time for their release in an e-suite. This is entirely reasonable and is something we will strive hard to achieve.

5. Feedback on ECMWF Forecast User Initiatives

Out of the six NMSs responding to our request for feedback on the Forecast User Portal and the online Forecast User Guide, five provided very positive, appreciative comments (Norway, Spain, UK, Hungary, North Macedonia). Portugal were more negative about the Forecast User Guide, saying that it was not so much for the day-to-day needs of a forecaster. We asked for clarification in an email. The initial brief response suggested they would like, for example, more detailed guidance on how the fog product performed at airports. This could certainly be done but arguably constitutes a move away from ECMWF's medium range remit into short range forecasting.

References

NMS Reports that contributed to this technical memorandum:
https://www.ecmwf.int/en/publications/search/?secondary_title=%22Green%20Book%202019%22

Presentations and Posters from the 2019 UEF meeting:
https://www.ecmwf.int/en/learning/workshops/using-ecmwfs-forecasts-uef2019

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+Portal

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