Technical
Memo

ECMWF
European Centre for Medium-Range
Weather Forecasts

919

Use and Verification of ECMWF products in Member and Co-operating States (2024)

Tim Hewson and Matthieu Chevallier

September 2024

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:

http://www.ecmwf.int/publications/

Contact: library@ecmwf.int

© Copyright 2024

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

Literary and scientific copyrights belong to ECMWF and are reserved in all countries. The content of this document is available for use under a Creative Commons Attribution 4.0 International Public License. See the terms at *https://creativecommons.org/licenses/by/4.0/*.

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

Abstract

This report summarises Member and Co-operating State reports on use and verification of ECMWF forecast products, focussing on the period January 2022 to May 2024. This period included a key model upgrade, cycle 48r1 in June 2023, which ECMWF particularly wanted feedback on.

Many NMS (National Met Service) submissions were quite comprehensive, providing many verification charts and citing various case studies. There are also many new product requests, despite ECMWF having probably added far more products in the last year than ever before.

The "big features" of cycle 48r1, such as the ENS resolution upgrade, the increased frequency of sub-seasonal forecasts (formerly known as "extended range forecasts") and the multi-layer snow scheme were all very well received. Verification evidence of performance improvements was provided – e.g. for wind speeds over Swiss mountains. Negative meteorological impacts were limited, although some mentioned technical challenges due to the higher volume of data.

In general ECMWF forecast products are liked, respected and widely used, from day 1 out to seasonal timescales, although most usage centres on the medium range – ~day 3 to day 10. Multiple visualisations are used operationally, including ECMWFs OpenCharts and ecCharts offerings.

Users in NMSs are generally very happy with IFS broadscale performance, and with weather parameter outputs. Several countries complimented ECMWF on cyclonic windstorm forecasts, despite some imperfections. Areas where improvements would be desirable, and where verification, survey results and case study data were presented in support, include 2m temperature extremes, winds over mountains, low cloud handling, snowmelt rates, lake and sea ice, and (most of all) convective precipitation and convective hazard prediction.

In general, and relative to the IFS, the skill of Limited Area Models (LAM) deterministic and ensemble systems run by NMSs has risen in recent years. For days 1 and 2 these systems have a performance edge over the IFS, as they should have given their higher resolution (~2.5km versus 9km), and they tend to be used in preference for those leads for weather guidance.

Again, NMSs reported many examples of post-processing and blending, of different types, imparting big improvements to model output, via bias correction and other adjustments. And a number of NMSs are delving into more complex post-processing approaches, related to the worldwide growth of artificial intelligence.

We asked for input on the new breed of data-driven forecast models run in real-time at ECMWF in experimental mode, such as the AIFS. Huge interest was expressed. Though outputs are not currently used in operational production and decision-making, quite a few NMSs mentioned that their forecasters occasionally monitor these outputs. Users want to build trust through better understanding, verification of extremes, seeing more outputs for more parameters and sharing experience.

Uptake of atmospheric composition outputs provided as part of the Copernicus Atmospheric Monitoring Service (CAMS) and Fire-related outputs has grown in recent years, with solar radiation and dust forecasts referenced. Meanwhile CAMS output post-processing, cited a few times, is a new activity.

Plain Language Summary

ECMWF is interested in how, in 2024, forecasters in National Meteorological Services in Europe are using its forecast products, and in how accurate they consider them to be. This report summarises reports sent in by those services to answer those questions. There are special references to recent forecast model changes, such as more frequently issued longer range forecasts, and to how users view a new breed of forecast model, based on artificial intelligence, that in recent years has taken the meteorological world by storm. The memorandum is also peppered with references to recent and upcoming physics-related changes made by ECMWF to its computer models, that continue to make the forecasts even better. Recent unusual weather events in Europe, of different types, that were particularly well forecast or particularly badly forecast are also discussed.

1. Introduction

Every other year ECMWF has been inviting Member and Co-operating State National Meteorological Services (NMSs) to submit reports on the use and verification of ECMWF's forecast products. The newly submitted reports are now available here: https://sites.ecmwf.int/publications/greenbook/. On this occasion the interval since the previous report was three years, not two. This was by design, to allow users in NMSs some time to give feedback on the major cycle 48r1, introduced in summer 2023. We asked users to focus on the period January 2022 onwards.

Reports have been provided by Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Georgia, Greece, Hungary, Ireland, Israel, Italy, Latvia, Lithuania, Luxembourg, Netherlands, Norway, Portugal, Serbia, Slovakia, Spain, Switzerland, Turkey and United Kingdom. This constitutes 26 countries, the same number that responded when this process was last activated 3 years ago.

A summary of the reports is presented below. Content has been combined with some input from official triennial Member State/Co-operating State liaison visits undertaken by ECMWF between July 2021 and July 2024. In using the associated visit reports we only use input from countries that did not send in a Use and Verification report in 2024. These were, in chronological visit order: Sweden (May 2023), Romania (February 2024) and Iceland (March 2024).

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

1. Summary of major highlights
2. Forecast Products
3. Verification
4. Output requests
5. References
6. Additional comments

For section (2), additionally this year, ECMWF solicited also input on atmospheric composition output provided as part of the Copernicus Atmospheric Monitoring Service (CAMS), on fire-related output, on impressions of cycle 48r1, and on the new and rapidly expanding topic of machine learning-based (ML) data-driven forecasting. Feedback regarding cycle 48r1 sits most naturally in section 3 on verification, and so that is where it is included below. Similarly, whilst feedback and figures related to other topics is usually reported on in the section that it was submitted to, there are always overlaps and sometimes it made more sense to relocate to another section in this report. All output/product requests have been collated here in an Appendix.

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 36 months (since the last memorandum of this type) we have introduced the following new cycles: 47r3 on 12th October 2021, 47r3 on the ATOS HPC on 18th October 2022 and 48r1 on 27th June 2023.

Note that the results of ECMWF's own objective verification are considered separately, in ECMWF Technical Memorandum 911 (Haiden et al., 2023), and in ECMWF Technical Memorandum 918 (Haiden et al., 2024).

On the technical side, the process for gathering the NMS reports differed slightly this time: we gave users two reporting options, either fill out a questionnaire style survey, with a figure upload option, or create a standalone word or pdf document in the traditional way. Uptake of the questionnaire option was quite high, and feedback on this was positive, so we plan to retain this option next time.

2. Use and Application of Forecast Products

Here we discuss, with examples, how ECMWF products and outputs of different types are used in many different ways by NMSs. Whilst medium range, sub-seasonal (=extended range) and seasonal forecast output usage has not changed that much, the uptake of CAMS (pollutant and dust modelling) and Fire-related outputs seems to have grown markedly. Similarly the Artificial Intelligence (AI) field is a big growth area.

Access mechanisms for ECMWF data remain similar, and most NMSs discuss these, at least briefly, in their reports. Most commonly an NMS user will visualise disseminated fields in their own workstations, particularly for short and medium range applications. This is often the most flexible option as it allows for overlaying on and intercomparison with output from other models and with real time observational data such as satellite and radar imagery. Sometimes more advanced tools are available on these workstations, to allow forecasts to be subjectively adjusted, or to blend together output or to apply some other mathematical processing. A second access method is to use an intra-organisation website to view pre-prepared outputs that have been specially tailored to meet the needs of specific forecasting tasks – e.g. with postage stamp display from an ensemble, or probabilities of exceeding a warning-related threshold.

The third and fourth access methods are provided and supported directly by ECMWF. The first of these is ecCharts, 15/26 NMS reports refer to this. Many appreciative comments were made regarding this increasingly sophisticated web-based visualisation tool, to which new parameters and outputs of new modelling systems (such as ECMWF's Artificial Intelligence Forecasting Systems) are being added continuously. On the other hand there are still some reports of slow speed, though less widespread than hitherto. The final access method is to use OpenCharts; 9/26 NMS refer to this system. This is like a cut-down version of ecCharts, that lacks pan and zoom and overlay functionalities, although clickable data, such as meteograms, are available. And as with ecCharts (which is the 'engine' behind OpenCharts) new variables, new models and new functionalities are continually being added.

2.1. Medium Range

All NMSs continue to use ECMWF output in various formats for medium range forecasting. The main usage period varies by NMS, linked in part to what other model output they have available – notably Limited Area Models (LAMs) - but the focal point is typically day 3 to day 10. Many do also use ECMWF output for shorter ranges (day 1 and 2) although often users can get better quality forecasts from their LAMs or LAM Ensemble Prediction Systems (LAM-EPS) at such leads (as demonstrated in Section 3), and so usage then will naturally focus on those. Focussed use of ENS sometimes starts later than usage of HRES; for example Sweden say they mainly use ENS from day 5 onwards.

Whilst most services refer to using the long-standing standard, "simple" variables, such as 500mb height, mean sea level pressure, rainfall totals and basic probability products, many of our more sophisticated outputs were singled out for praise. Favourable comments were made by at least two NMSs about the following product classes: EFI/SOT, ecPoint-Rainfall, Precipitation type, Lightning, Convective indices, Cyclone Database products, Weather regimes, Vertical profiles, Tropical-cyclone-related outputs, Meteograms. To this list we can add two products developed in the last 3 years: Visibility meteograms and Clear air turbulence products.

Building on work done at ECMWF and the Hungarian Met Service, Croatia have created a new and useful deterministic product from HRES that represents both precipitation type and intensity, in map format (Figure 1), allowing the user to focus on where amounts (of different precipitation types) are expected to be greatest. A further development could potentially be to build on this to show an ensemble-based equivalent. Whilst ECMWF provides this type of information in meteogram format, the multi-dimensional needs of a map-style product are much more challenging.

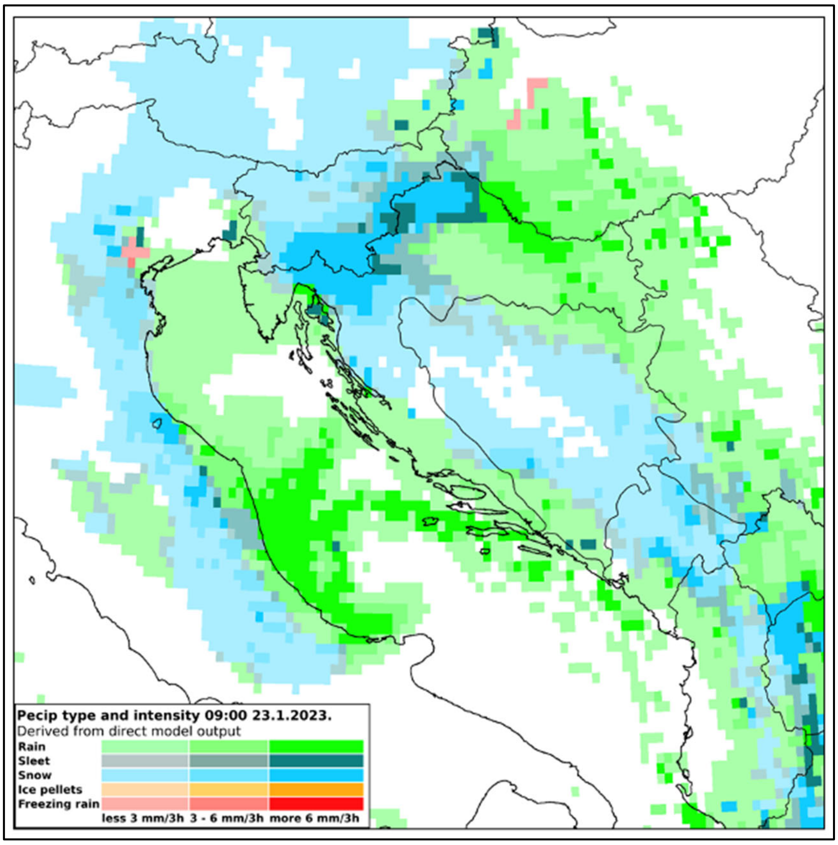


***Figure 1**. Example of a product developed in Croatia, showing instantaneous precipitation type and 3h accumulation from an ECMWF forecast.*

2.2. Sub-seasonal (= Extended range / Monthly)

Most NMSs use ECMWF sub-seasonal forecasts for the public and/or sector-specific customers. Energy was one customer group mentioned, with some NMSs tailoring output to their needs. Commonly NMSs access the output via OpenCharts, or use disseminated data to create bespoke products. Timescales for usage vary; many services still do not actively use the daily updates introduced in cycle 48r1 – e.g. still updating their products twice a week – though on the other hand this initiative has been widely welcomed. This may be a legacy issue. Facilities for simple intercomparison of recent runs using the “4-orange-box” tool in OpenCharts were welcomed, but the same customer also expressed some disappointment at the loss of the old website functionality which allowed the user to easily switch variable. Whilst some users appreciate having more sophisticated tools like regime diagrams, Hovmoeller plots and Stratospheric wind diagrams, most do still focus on using the basic week-long anomaly plots for 2m temperature, total precipitation and mean sea level pressure. Interestingly, there was no direct mention of the quintile and decile probability maps, which can in principle provide useful guidance on unusual anomalies. ECMWF is working on improving the design of sub-seasonal graphical products, aiming to provide information in a more easy-to-interpret format.

Switzerland provide a nice example of a tercile-based combined product (Figure 2). In effect this usefully represents all three terciles on one map, achieved by only depicting probabilities for the most populated tercile, and then only when the said probability is >45%. When no tercile probability exceeds 45% the plot stays white. One can also see on this Swiss product when there is a relatively high probability of normal: in this case greys are shown.

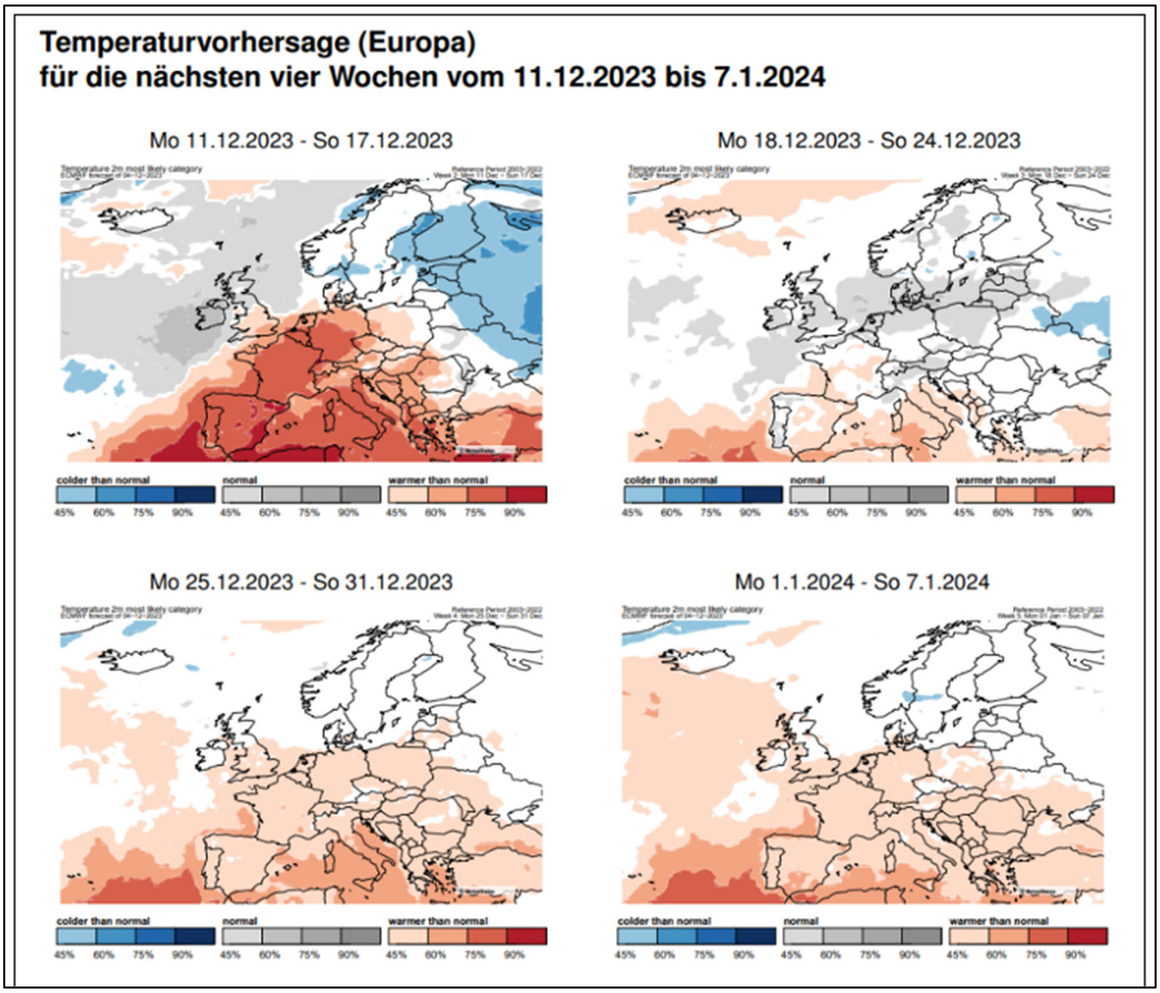


***Figure 2**. Example sub-seasonal forecast product from Switzerland, representing in multi-category tercile probability format, the week-by-week outcomes in a 48r1 forecast from 11 December 2023. The hue (blue, grey, red) represents the most populated tercile category, whilst shading darkness denotes the probability for that category. To be classed as the most populated the probability must exceed 45%; otherwise white is used.*

The UK have welcomed the doubling of ensemble size that took place with cycle 48r1. They say that this helps with their multi-model blending activities (ECMWF output now affords more weight), and also suggest that this leads to less jumpiness, in their regime-style products, between consecutive dates in a given forecast. Figure 3 provides an example – smoother transitions are seen on the right hand panel which used 101 members than on the left panel which used 51.

Whilst sub-seasonal forecasts now go out 6 weeks into the future – a change made some years ago – no service mentioned using data beyond “week 4” (other than graphical products like Figure 3). Maybe this is due to lack of anomalies on the “week 5” and “week 6” charts, or due to a perceived or real lack of skill. Indeed, Estonia state that "neither monthly nor seasonal forecasts can be highly trusted in Estonia" which seems to have reduced their uptake there, even for weeks 3 and 4.

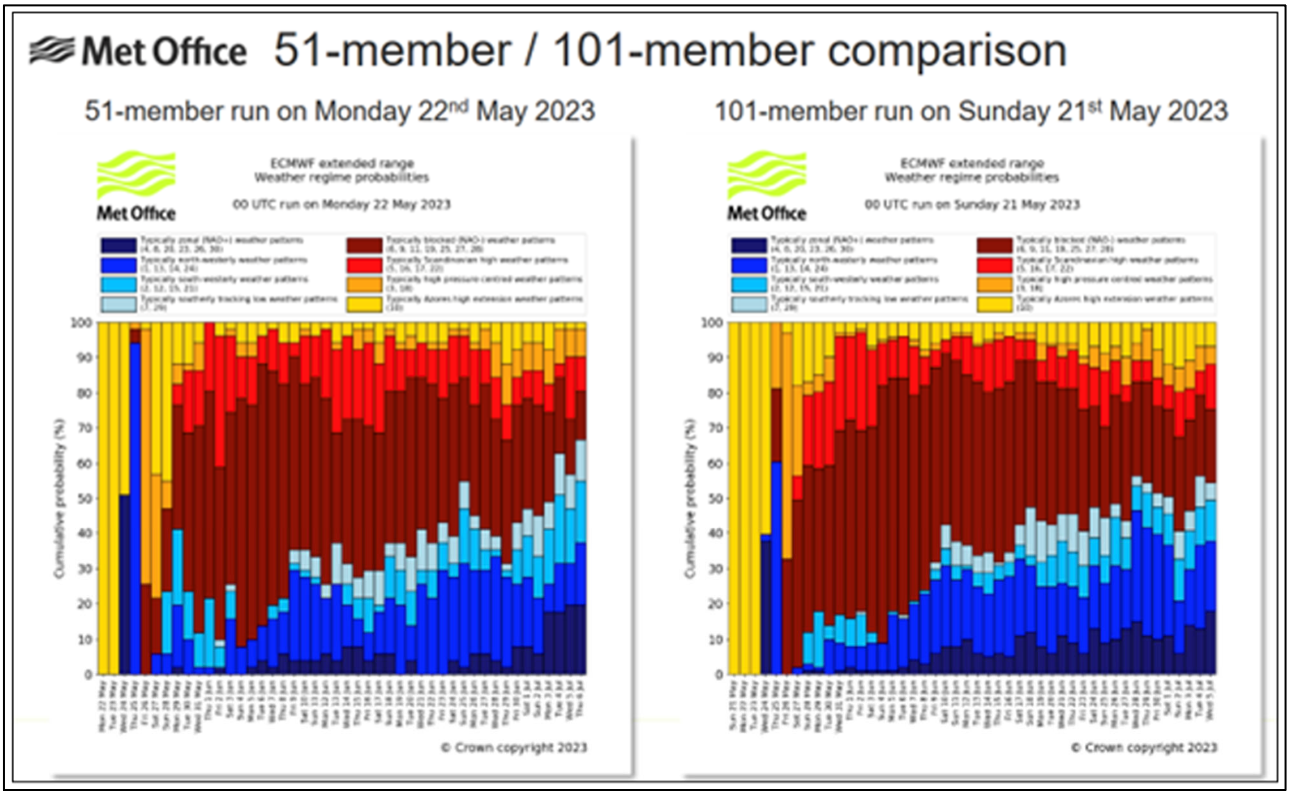


***Figure 3**. Example products from the UK, showing in bar chart format the frequency, in ECMWF sub-seasonal ensemble forecasts, of different UK-centric weather regimes. The two panels are from consecutive dates; the right panel shows a 101-member 48r1 pre-operational forecast from 21 May 2023, the left panel an operational 51-member 47r3 forecast from one day later. Note the somewhat smoother temporal evolution in the 48r1 forecast, believed to be due to the larger number of members.*

2.3. Seasonal (=Long Range)

About 16/26 NMSs regularly use seasonal forecasts for operational purposes, be that for the public or specific customers. In some countries there are no public forecasts; instead, only bespoke customer-specific output is generated. Specific customer groups mentioned in the reports include ice breakers operating on the Baltic (Finland). Hydrological models are also mentioned, whilst Israel note the importance for them of winter-time rainfall forecasts (due to water scarcity and dry summers). Where web charts are used many of the different seasonal forecast products that ECMWF provides get a mention. France and Germany refer to using multi-system data (e.g. the Copernicus Climate Change Service multi-system seasonal forecast[1]), and France state that they look for multi-model agreement in order to place any credence on a given set of forecasts. Norway and Sweden meanwhile note the current lack of skill, with Sweden stating that the 850mb temperature forecasts just follow climatology.

The UK expressed interest in more frequent updates of seasonal forecasts. In this regard, in the next upgrade of ECMWF seasonal forecasting system planned for 2025, forecasts will be initialised twice a month.

---

[1] https://climate.copernicus.eu/seasonal-forecasts

2.4. CAMS / Fire

Use of CAMS outputs, in one form or other, occurs in about 10/26 NMSs. For fire products the proportion is about 8/26.

Dust aerosol optical depth is mentioned several times, due presumably to the periodic incursions into Europe that we see, from Saharan sources, on southerly flow. The UK see these dust forecasts as “reliable”. Outputs related to solar radiation (UV index) are also referenced. Bulgaria use CAMS output to produce a dedicated air pollution product, whilst Hungarian forecasters have access to various CAMS outputs in their “HAWK3” forecaster workstations. Israel are ingesting aerosol data into their COSMO LAM and using this within its radiation scheme.

Some NMSs highlight the importance of fire-related products. There are references to the products delivered as part of the European Forest Fire Information System of the Copernicus Emergency Management System (CEMS), and to specific products accessible via ecCharts used by NMSs contributing to the ARISTOTLE European Natural Hazard Scientific Partnership.

2.5. Derived Products

Using ECMWF raw model output to create derived products is very widespread: 19/26 NMSs reported related activities and gave numerous examples. Only a subset of these can be reported on here.

NMSs continue to want to help their forecasters with thunderstorm forecasting. This is high priority, partly because of associated surface hazards/impacts which can be severe, and related to that also the interests of aviation. To this end thunderstorm-related indices of various types continue to be produced, including for example lapse rates in different atmospheric layers. Latvia take the total lightning outputs from the IFS, which are already a type of calibrated product, and, in conjunction with other data, re-calibrate those for local use.

There is also interest in and activity related to precipitation and drought related indices (in for example Switzerland) and also heatwave indicators. Note that new human thermal comfort indices will be made available as new products from IFS cycle 49r1 later in 2024.

For winter hazards Spain have worked on snowpack indicators, and we saw already precipitation type developments by Croatia in Figure 1.

Meanwhile activities related to weather regime classifications and clustering were already illustrated on Figure 3 but are taking place in other countries too. Spain for example creates clusters for the Balearic Islands and for the Canary Islands on a daily basis, using an ECMWF clustering approach.

Slovakia and Israel derive and use classical development metrics rooted in quasi-geostrophic theory, namely thermal and vorticity advection, presumably to try to infer regions of forced ascent.

IFS fields of 2m temperature and to some extent other variables undergo, in various statistical ways, downscaling, bias correction and blending between models, to achieve seamlessness across lead times and/or more accurate forecasts. Figure 4 from France demonstrates the improvements that have been made to maximum temperatures in both HRES and ARPEGE forecasts with post-processing, and how post-processed outputs themselves can then be blended to deliver further gains. Meanwhile Figure 5 is an example of how Norway now exploit the daily extended range forecasts to create a 3-week product; post-processing and blending activities are also used here.

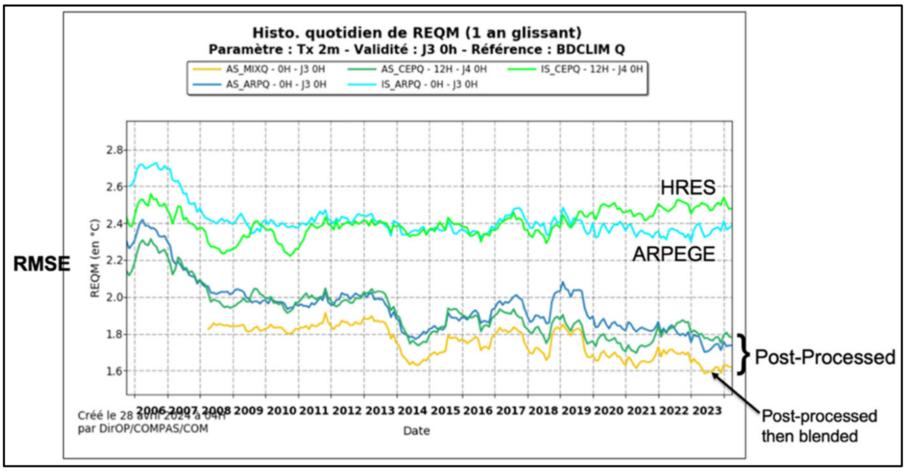
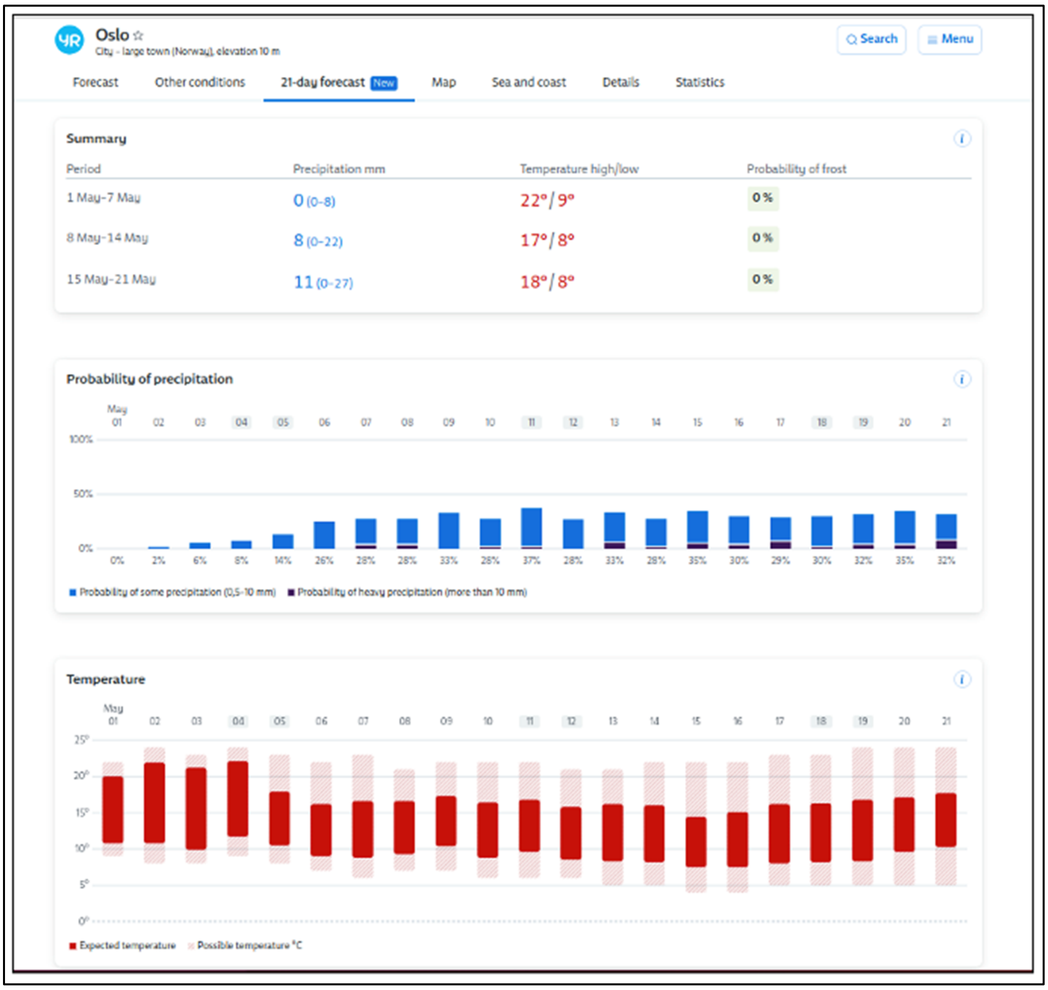


***Figure 4**. Root mean square errors of 2m maximum temperature forecasts, for French stations, for 2006-2023. Green and blue denote, respectively, HRES and ARPEGE global models. Lighter shades show raw model output. Darker shades denote post-processed, which exhibit consistently smaller errors, whilst then blending these outputs achieves further gains (yellow).*



Summary

Period	Precipitation mm	Temperature high/low	Probability of frost
1 May-7 May	0 (0-8)	22°/9°	0 %
8 May-14 May	8 (0-22)	17°/8°	0 %
15 May-21 May	11 (0-27)	18°/8°	0 %

***Figure 5**. Example of a new 3-week single site public forecast for Oslo, from Norway, from the yr.no website. This product is now updated daily using the daily issue sub-seasonal forecasts introduced in cycle 48r1. Week-by-week summary information for rainfall, temperature and frost is shown at the top, whilst probabilistic daily forecasts for rainfall and temperature are shown below.*

2.6. Other modelling activities

This sub-section covers the use of ECMWF model runs to provide initialisation and/or lateral boundary condition data for models run in NMSs, mostly via the long-standing ECMWF Boundary Condition (BC) project. Typically, these NMS-run models are higher resolution LAMs, with spatial resolution of around 2km although that does vary, and whilst the vast majority span regions in and around Europe, some also cover overseas territories. The proportion of NMS that said they do use ECMWF data as a modelling input was 21/26.

Whilst use of ECMWF data to help drive other models is clearly a growth area for NMSs, responses in this class were quite similar to 3 years ago.

Historically NMSs used data from HRES to provide BCs for deterministic LAM runs, but increasingly over the years there has been a trend to drive LAM-EPS systems with ENS data. Such activities received a big boost when ECMWF harmonised the HRES and ENS resolutions at 9km in cycle 48r1 in June 2023. This meant that the LAM-EPS systems benefitted from being able to ingest much more detail at their boundaries than hitherto. When LAM-EPS systems have fewer members than the IFS, as is very often the case, random member selection is one approach used to assign BCs. Norway report that their predictions of high coastal sea surface levels have benefited from the switch from a deterministic to an ensemble-based approach. The LAMs mentioned in reports this time were ALADIN, IREPS, AROME, HARMONIE, HARMONIE-AROME, COSMO, ICON-LAM, INCA, WRF, WRF-NMM, NMMB, C-LAEF, A-LAEF and ALARO.

Besides ‘weather forecasting models’ services are also running ocean, wave, surge, sea-ice, hydrological and agrometeorological (water-balance) models using ECMWF BC data in various ways. Sometimes that data undergoes post-processing from the outset, for example to try to harmonise with the resolution of a local hydrological model.

Additional modelling activities include dispersion and trajectory modelling – e.g. for volcanic ash and potential nuclear emissions, chemical species modelling, marine oil spill drift, and nowcasting activities.

2.7. Artificial Intelligence

2.7.1. Data-driven modelling

Although the use of AI techniques for forecasting purposes was a topic for the last report of this type, 3 years ago, since then there has been a veritable explosion in the use of AI to create 4-dimensional global forecast fields. From a user perspective, and in structural terms, this activity roughly mimics classical physics-based global prediction. However, the transitions between one time frame and the next, whilst they might look very realistic, have typically been ‘learnt’ or inferred, in complex ways, from re-analysis training data, typically provided by ERA5. Partly because of this ERA5 link, this new breed of models also runs at lower spatial resolution than classical global models – e.g. ~25km rather than ~9km. Another current AI model feature is the lower vertical resolution; ordinarily forecasts are only for standard pressure levels, plus some surface variables. The AI models are called “data-driven models” due to their reliance on training datasets.

In 2023, ECMWF started to run in real-time externally developed data-driven models (FourCastNet, FuXi, GraphCast, PanguWeather) initialised with HRES operational analyses. In the autumn of 2023 ECMWF introduced its data-driven forecasting system, AIFS, developed as part of a new Machine Learning project. Some outputs from all those models are made available on ecCharts and OpenCharts. All AIFS fields are also available through the usual data dissemination routes. Note that at the time of the NMS's consultation only deterministic models were available (an AIFS Ensemble forecasting system has been introduced in June 2024). Evaluation has highlighted competitive performance of data-driven models compared to HRES: for example, AIFS deterministic forecast generally scores better than HRES in synoptic scale metrics, and is also better in predicting the track of tropical cyclones (e.g. Haiden et al., 2024).

So, we asked for feedback from NMSs on these very new initiatives, and on their perspective of using outputs from data-driven models in operations. First, we asked whether NMSs were aware of ECMWF AI initiatives; 22/26 said they were. Then we asked for viewpoints. These are summarised with a collection of quotes, most of which were very supportive:

- "We salute ECMWF's leadership here"
- "Potential is huge. ECMWF cannot miss out!"
- "Impressive that you have done this"
- "Very welcome"
- "Interesting"
- "They look surprisingly good"
- "Helps build trust"
- "If it is about AIFS, we found it very needed. The potential of AI model is so great, we cannot miss it!"

Some NMS also asked for more elaboration:

- "How far out can AI models be skillful?"
- "Need more on reliability and accuracy"
- "...A better overview/comparison of output parameters is needed…"
- "Don't oversell (limited output variables)"

Then on the question of whether data-driven models were used for operational forecasting 25/26 NMSs said no. However, some NMSs mentioned that forecasters already occasionally look at outputs from data-driven models (e.g. Bulgaria, Croatia, Finland, France, Hungary, Ireland, Norway, and UK), comparing broader scale patterns available with other outputs, and/or assessing their relevance. Users were then asked what they would need in order to use AI models in their forecast activities. Training, seminars and experience sharing were all requested, as was information on the pros and cons of these new forecasts versus physics-based NWP forecasts. The ECMWF Forecast User Guide will include guidelines on how to interpret and use outputs from deterministic data-driven models, with extension to the ensemble system next year.

NMSs also requested verification of different types, which we work on (e.g. Ben-Bouallegue et al., 2024), whilst providing already upper-air verification on OpenCharts for AIFS together with the other data-driven models. Surface verification for data-driven models will be added to OpenCharts shortly. One NMS specifically asked for information on performance of data-driven models for extreme events – notably for events in the distribution tails. This we see as pivotal, given that a fundamental goal of NMSs is to issue timely and accurate warnings and a fundamental goal of ECMWF is to support that. We are now exploring this aspect; there are some limitations in the deterministic AIFS forecasts in this regard, which we are hoping to address with the AIFS ensemble. Other users have also asked to see many more parameters, which is understandable given that availability now is just a fraction of what one can get from the IFS. Croatia explicitly mention solid precipitation types. We are working towards a new AIFS version which will include more variables. And finally, Croatia requested higher resolution forecasts, motivated by the topographic complexity of the country itself, and probably also by the benefits they have seen over the years of IFS resolution increases. This is another upcoming target for AIFS development to move towards 9km resolution. Note here that Spain and Denmark reported that they are already investigating the topic of data-driven LAM forecasting, as are Norway[2]. The challenge of staffing this new area was also highlighted by some NMSs.

It is noteworthy that Germany provided comments on AIFS forecast performance for a few parameters. They reported AIFS precipitation forecasts were of good quality but have a strong bias for low rates. They also show, with one month of European verification (April 2024), that AIFS has markedly higher biases and higher RMSEs for 10m wind than HRES, the discrepancy being greatest at night.

Such feedback from Member and Cooperating States is invaluable in a time of rapid change because it can help establish guidelines on strengths and weaknesses relative to physical models (i.e. the IFS). Data-driven modelling can be directly tailored towards users, and feedback on where models can be improved (e.g. specific product additions) can be rapidly incorporated. Notably, data-driven model structure enables more frequent updates than can be achieved with the IFS.

2.7.2. Applying AI techniques to raw NWP output

NMSs also reported examples of where AI techniques were being applied to outputs of ECMWF's physics-based models. Here one needs to bear in mind that there is overlap between what was once called post-processing, and what is now referred to as an AI technique.

France report on using MOS, EMOS and random forests to deliver km-grid forecasts of 2m temperature. They also want to add rainfall and gust parameters, but report that that is on hold due to slow data download speeds. Meanwhile Finland are using gradient-boosting random-forest based error correction for several parameters and Spain are using convolutional neural networks for regime classification. In the "classical method" category, Turkey report on using non-homogeneous Gaussian regression and logistic regression for 2m temperature and precipitation variables, calibrating using reforecasts (see Section 3.2.1 for verification results), whilst Germany highlight the striking improvements they have achieved in air quality forecasts by applying ML/MOS techniques to HRES and CAMS forecasts, using station measurements for training.

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. ECMWF specifically asked users to focus on the period from 1 January 2022 onwards. Of relevance to interpretation are the dates of the most recent upgrades to the IFS:

- Cycle 47r3 became operational on 12th October 2021
- Cycle 47r3 on the ATOS HPC became operational on 18th October 2022
- Cycle 48r1 became operational on 27th June 2023.

---

[2] https://www.ecmwf.int/en/about/media-centre/aifs-blog/2024/data-driven-regional-modelling

This means that in this year's reports bulk verification scores, until some date in 2024, would be expected to correspond mainly to cycle 47r3. On the other hand, we did ask specific questions on the impact of 48r1, addressed in sub-section 3.1.

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 often subtracts from statistics for the operational forecast the equivalent statistics derived from a fixed model version run over the same period (currently based on ERA5), which can help eradicate impacts of this type (see for example Haiden et al., 2024).

And when considering a *fixed verification period*, there are likewise several reasons why one would not necessarily 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 specified.

Furthermore, on the interpolation aspect, note that ECMWF performs some routine comparison of forecast skill for ECMWF and other global centres using the TIGGE data archived at ECMWF (Haiden et al., 2024), in which data is interpolated to grids that have lower resolution than the native models, using techniques which very probably differ between centres (e.g. conservative or nearest neighbour). This interpolation issue is a limitation of such in-house intercomparisons. Therefore, model inter-comparisons that NMSs make between their own models (presumably at full resolution) and IFS output (hopefully also at full resolution) are particularly valuable.

3.1. 48r1 versus 47r3

ECMWF asked users to describe positive and negative impacts for their service, of IFS cycle 48r1, and also to describe any systematic changes they had seen to the output. Only 1/26 NMSs noted a systematic change, and that related purely to the higher resolution grid. For the "impacts" questions there were many more responses, often supported by some verification results. Those responses are summarised here.

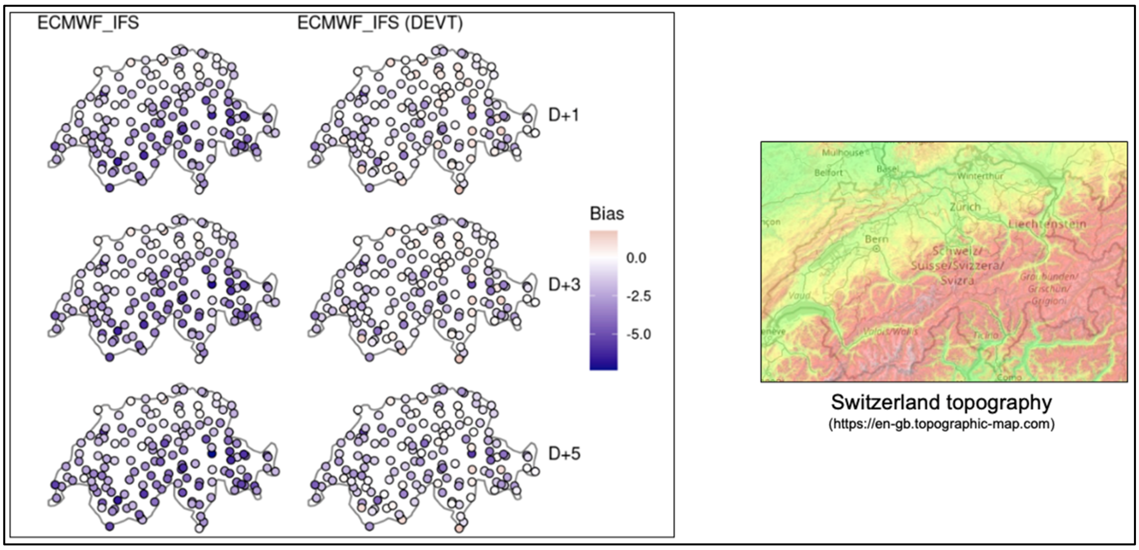
3.1.1. Positive aspects of 48r1

On the structural side, for the medium range, most services indicated that the higher resolution ENS (9km versus 18km previously) had brought positive impacts for them, and some noted that the harmonisation of HRES and ENS resolutions delivered physical consistency. Scrapping the day 15 resolution change between medium and sub-seasonal ranges was also seen as a positive step, as it made things less "messy'. Technically this is true, though on the other side we acknowledge that this change could on occasion reduce continuity across the day 15 barrier (between medium range and sub-seasonal outputs). However, no NMS mentioned that as a problem.

Most NMS found more frequent sub-seasonal forecasts (daily issue instead of twice per week) very helpful. One NMS did however say that day-to-day jumpiness was a minor negative of daily updates. Whilst the simultaneous change to more ensemble members in the sub-seasonal forecast (101 instead of 51) is relatively invisible in many products some NMSs nonetheless expressed appreciation for this upgrade.

Regarding parameter performance in 48r1, increased resolution in the medium range ensemble was expected to lead to a better representation of topographic impacts on weather. Some NMS indeed mentioned better rainfall in mountainous areas in 48r1. And Switzerland noted improved IFS precipitation performance in several metrics in autumn and winter 2023-24 (48r1) when compared to the equivalent times 1 year earlier (47r3), although could not unequivocally rule out the possibility that more predictable weather was the reason. Better rainfall downstream of mountainous areas, also expected with higher resolution due to an increased rain shadow effect in the model, was not reported.

Switzerland provided a nice, striking example of 10m wind speed bias reduction in mountainous areas in 48r1 (Figure 6), using concurrent periods.



(https://en-gb.topographic-map.com)

Figure 6*. Daytime Maximum 10m wind speed bias in (top to bottom) day 1, day 3 and day 5 ENS forecasts for sites in Switzerland, in May and June 2023, from cycle 47r3 (left, 18km resolution) and 48r1 (right, 9km resolution). Marked reductions in biases are seen in the newer cycle, particularly in the most mountainous areas in the south and east of the country (see topographic map to the right). The resolution upgrade is believed to have played a large role in error reduction.*

With regard to cycle-related improvements in general, that are unrelated to resolution changes, Portugal report a small but systematic step change in minimum 2m temperature performance in HRES (Figure 7), using concurrent periods. Finland report reduced biases in very low temperature scenarios in 48r1 (Figure 8, ringed), which is consistent with the positive impact expected from the new multi-layer snow scheme. Having several snow layers modelled (at least when the snow is deep) allows a temperature gradient to develop within the snowpack which usually means much lower temperatures at the top, which can in turn facilitate lower temperatures in the atmosphere. Whilst frequency biases remain below 1 for HRES on the 48r1 panel (right) for values below -15C, they are much closer to 1 than on the 47r3 panel (left). It is also worth noting that runs with the 2.5km resolution MEPS (that uses HRES/ENS for BCs) had better biases than HRES in both periods (red and purple curves).

In other positive comments about 48r1, NMSs noted a better representation of extremes (as defined by the 10th and 90th model climate quantiles) at short lead times, and better visibility forecasts in general.

Despite the positive feedback, some NMSs also said they had not noticed any changes in accuracy with 48r1.

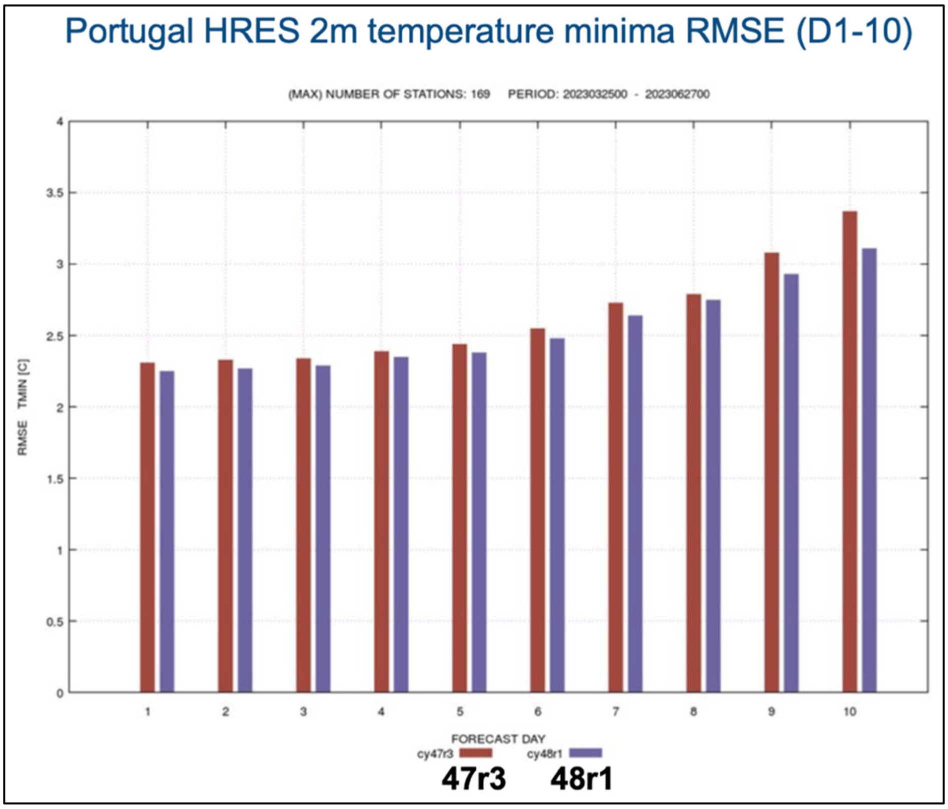


***Figure** 7. A comparison of the root mean square error, in minimum temperature forecasts, between cycles 47r3 and 48r1, for sites in Portugal. Period covered is 25 March to 27 June 2023 in both cases.*

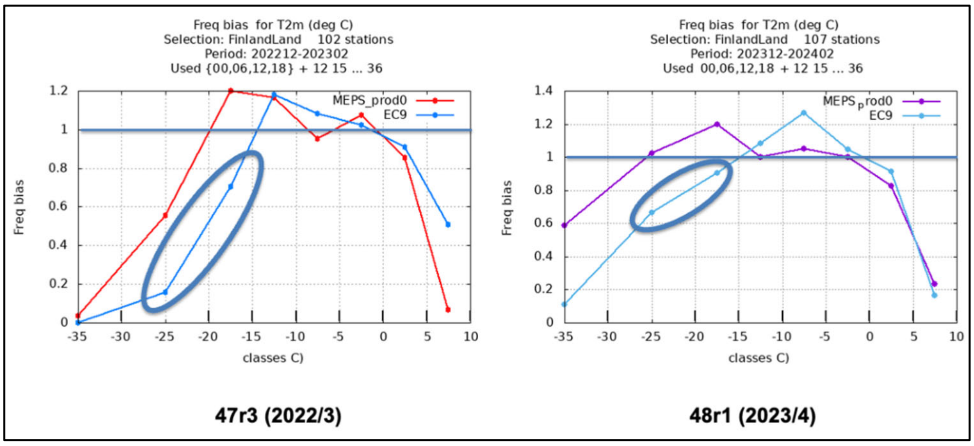


***Figure 8**. Frequency bias of 2m temperature forecasts for Finnish stations, for the winter months (DJF) for cycle 47r3 (left) in 2022-23 and 48r1 (right) in 2023-24. Classes are based on observed 2m temperatures. Forecasts are for T+12-36 (see titles). The high resolution LAM (MEPS) runs are shown in red and purple, whilst ECMWF forecasts are shown in blue. The much less pronounced under-prediction biases, in 48r1 HRES, for very low temperatures (~ -20C) are highlighted with blue ellipses. The multi-layer snow scheme introduced into the IFS in cycle 48r1 is believed to be a contributory factor.*

Representation of 2m temperature in the ENS in coastal, island and peninsula locations has, in general terms, improved markedly in cycle 48r1, due to the resolution increase. Some islands and peninsulas that were just sea points previously in the ENS are now land points. This will have delivered a very large and positive step change on meteogram displays, for example. Moreover, huge inconsistencies between HRES and ENS 2m temperature graphs seen on certain meteograms in cycle 47r3, due to islands being land points in HRES (at 9km) and sea points in ENS (at 18km), will have disappeared. Nonetheless, we still get negative feedback for some sites. A case in point is the resort of Nida, situated on the Curonian spit (a very narrow peninsula) in Lithuania, which Lithuania highlighted (see Figure 9). The meteogram here has not improved; it continues to show 2m temperatures that relate purely to over-sea conditions. This is because the spit is only about 4km wide, so not quite large enough to justify a land point at 9km resolution. More than 50% land is required for land point designation.

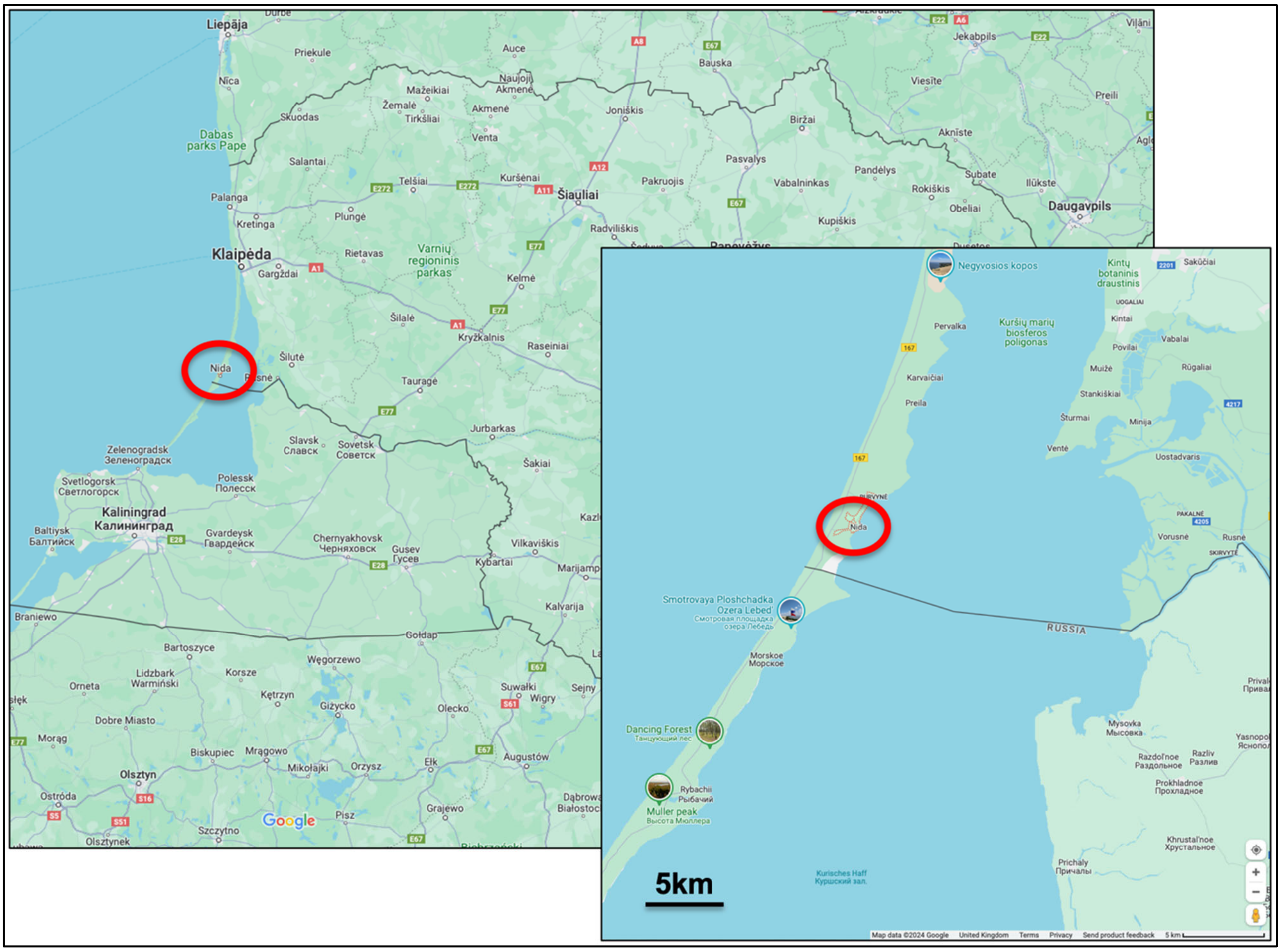


***Figure 9**. Google maps extract showing the geographical location of a resort (Nida) for which Lithuania reported poor 2m temperature forecast performance, even after the resolution increase in 48r1.*

3.1.2. Negative aspects of 48r1

Only 8/26 NMSs reported anything negative about 48r1. Some simply expressed disappointment that certain aspects had not improved. For example, continuing "issues" with convective rainfall were noted, with Italy's rainfall example for HRES in Figure 10, for summer seasons, being an example of this. Note how the frequency bias picture for the 2023 summer (mainly 48r1, right panel) has a similar structure to the one for 2022 (47r3, left panel). We would re-iterate that whilst some improvements to this picture are possible, there is a fundamental problem when comparing gauge data at points, with model data which is a gridbox average. This comes to the fore in convective situations (e.g. as prevail in summer in Italy) when sub-grid variability, that can be mostly responsible for the “mismatch”, is typically large (see also Section 3.2.3).

Turkey have highlighted that the timing and location of extreme precipitation seem to be a problem for 48r1, which is a slightly different issue to the above and could be a real degradation. However, when looking at ENS products such a perception could arise from the double penalty effect: higher resolution (of 48r1 ENS) innately delivers more detail, making misplacements more obvious.

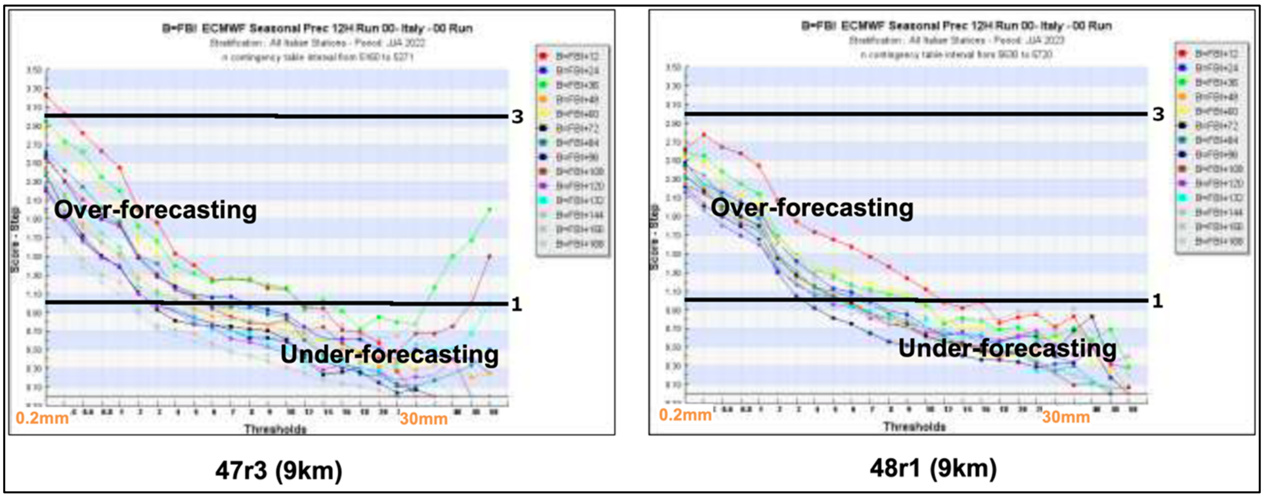


***Figure 10**. Frequency bias for threshold exceedance (x-axis) for 12h rainfall totals forecast by HRES, for Italian sites, for JJA 2022 (left, cycle 47r3) and JJA 2023 (right, 72% cycle 48r1). Different colours denote different lead times, going up to day 7. Small totals are over-forecast whilst large totals are under-forecast. Italy report little difference between the two cycles (mainly because HRES resolution did not change in 48r1).*

Ireland express disappointment that the too-slow melt rate for snow on the ground has not gone away in 48r1. The new multi-layer snow scheme introduced in cycle 48r1 is expected to have significant impact only when the snow depth becomes large enough for the extra snow layers to come into play. For the small depths typically experienced over Ireland there will have been just one layer, as previously. Cycle 49r1 will include a physics change that will enhance the melting process for small amounts, and we would encourage users to check how this behaves in real situations in the coming winter.

Lithuania report that visibility in modest convective wintry precipitation now drops too low, which is surprising. We are not aware of changes in 48r1 that would specifically alter visibility levels in wintry precipitation, and moreover the IFS does not reference the convective precipitation component when assessing visibility. The latter point means that during modelled convective snow, visibility levels in IFS output should rather be too high. To back up any reports of perceived IFS behaviours ECMWF would always encourage users to provide examples.

The slight differences between HRES and the ENS Control run have been noted as a negative. It is true that both were not bit identical in 48r1, and miscellaneous small differences in auxiliary fields, such as orographic height, had a tendency to have a growing impact with lead time on forecast evolution, via “chaos in action”. Note that ECMWF did report on these slight differences before cycle 48r1 implementation. However, systematic differences shown in Hungary’s report and shown in Section 3.2.3 are more surprising. With the next cycle, 49r1, HRES and ENS Control runs will be bit identical.

On a technical level, the higher volumes of data in the 48r1 ENS presented challenges that have been mentioned by some NMSs. Finland reported having related technical difficulties which have slowed down the post-processing of the data. As a trade-off, they mention that for the ensemble forecast they have elected to use full resolution data only for surface parameters.

France mention that one of their post-processing suites was altered with the upgrade to 48r1 in June 2023 (Figure 11) – note the jump of about 1C in RMSE values at the start of July 2024. A re-calibration would have been necessary prior to the cycle implementation, through a re-training on a set of data from the new chain. Accordingly, France mention the need to access a sufficiently long dataset (~1 year) from an upcoming cycle sufficiently in advance of cycle implementation in order to assess impact and to recalibrate their post-processing suites.

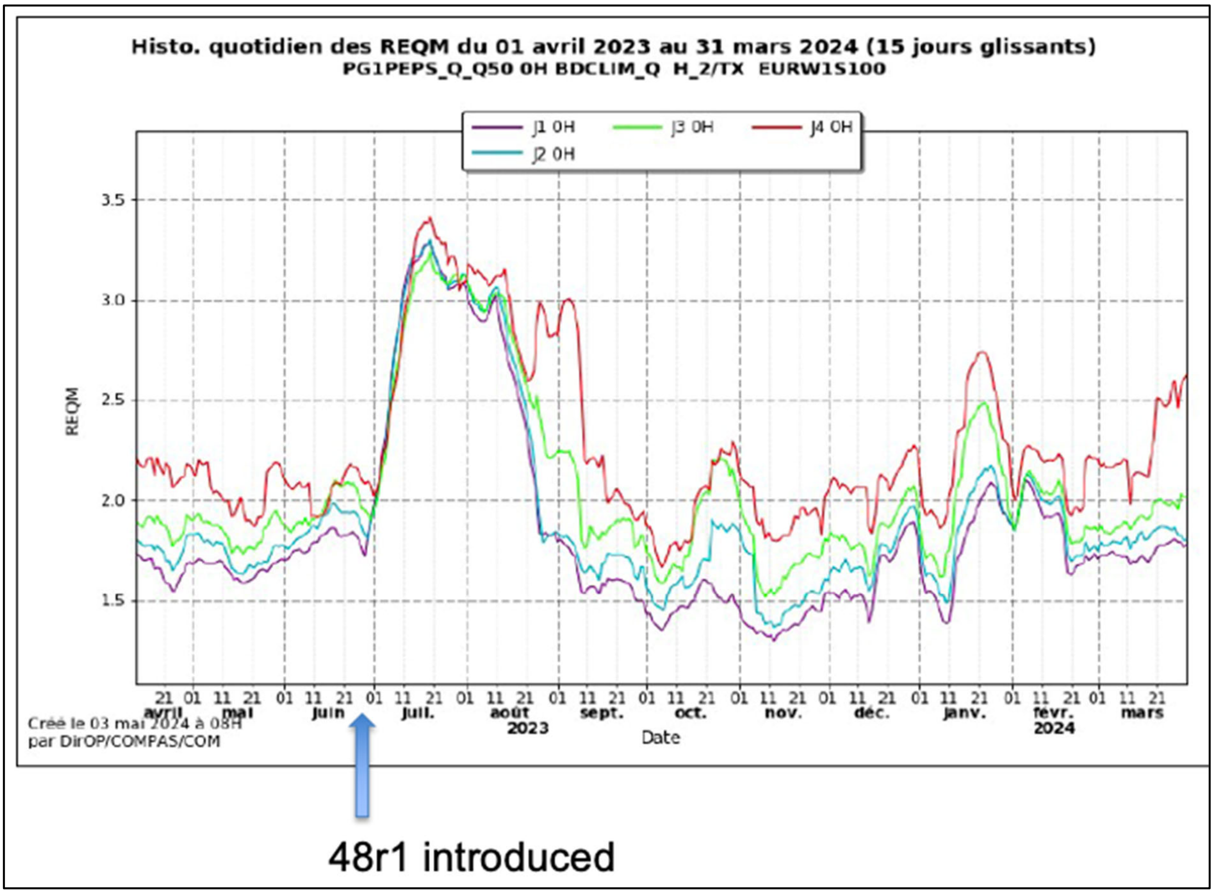


***Figure 11**. Timeseries of the RMSE of post-processed (median) 2m maximum temperature forecasts for French stations between April 2023 and March 2024, based on ENS outputs (15-day rolling averages). The forecast for day 1 is shown in purple, day 2 in blue, day 3 in green and day 4 in red.*

3.2. ECMWF and LAM output, Raw Model and Post-processed

NMSs present a lot of verification data, far more than can be summarised here. Readers are referred to the country reports to gain more local insights. Key messages from the verification are bulleted below; some evidence is then presented in sub-sections that follow:

- HRES/ENS are outperformed by LAMs at short leads for many surface parameters, though not all. This is different to the situation 3 years ago – LAM skill is rising in relative terms. Figure 12 is an example of this.
- The main reasons for the striking LAM skill seem to be high spatial resolution, the fact that IFS boundary conditions are often used, and increasingly co-ordinated LAM modelling efforts.
- Whilst LAMs can sometimes be too wet, they clearly have a better frequency bias for rainfall than the IFS.
- There is some evidence that broadscale patterns and pressure level values are better represented at short leads in the IFS than in LAMs
- The German ICON global model has become increasingly competitive, and leads ECMWF in the verification results presented by DWD
- Extremes of heat and cold tend to not be extreme enough in the IFS
- Post-processing continues to add noteworthy value across the product ranges

In general, the impact of the ENS resolution upgrade to 9 km is not fully felt in the results presented. We would also note that the UK view HRES data as a good representation of truth, given that they use this to verify their own global model.

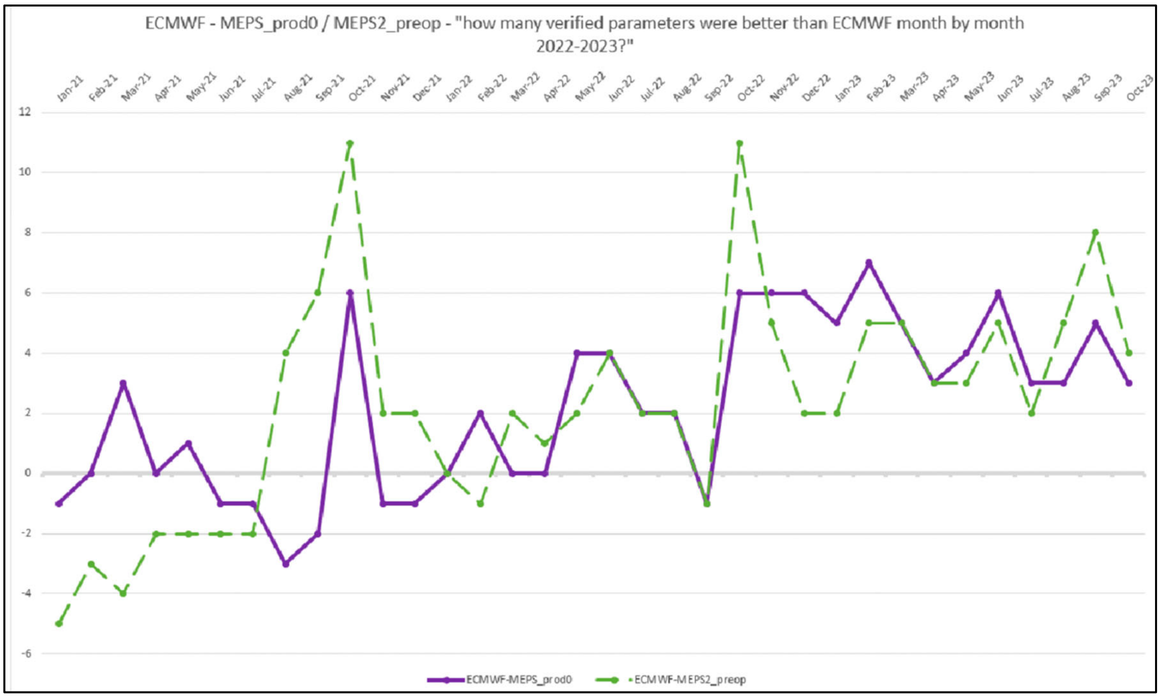


***Figure 12**. Evolution of MEPS performance versus ECMWF, from January 2021 to October 2023 (x-axis). Lines show how many weather parameter scores, from routine Finnish verification, showed better performance in MEPS model output (2.5km resolution) than in IFS output (y axis). Zero means an equal number in each, positive values show MEPS leads. Purple line denotes the operational MEPS forecasts, green is pre-operational.*

3.2.1. 2m temperature

Figure 13, from Finland, illustrates the tendency of raw output from the IFS (specifically HRES here) to underestimate maxima that are unusually high for the domain of interest – Finland in this case (lower right panel). Meanwhile the 2.5km resolution MEPS runs do a better job, exhibiting negligible bias overall (lower left panel). The results for minima are similar in the sense that HRES minima are not low enough on very cold nights (top right); in such instances MEPS mitigates this tendency but does not entirely eradicate it (top left). These types of IFS characteristics have been reported by numerous countries, and in some regions, in some seasons, these "shortfalls" in both minima and maxima are seen not just for extreme values but more generally. Sometimes LAMs have larger biases than the IFS, even in spite of higher resolution, but usually they are better, as on Figure 13.

The shortfall for minima in the IFS is believed to relate primarily to the difficulties of representing (very) stable boundary layers, and surface wind speeds under such conditions. Commonly the IFS shows very light winds when observations show even lighter winds or calm conditions. These problems can be amplified over snow cover, as is commonplace over Finland in winter. At the heart of the issue is the fact that in these circumstances tiny amounts of energy can be associated with very large temperature discrepancies. Some upcoming changes in cycle 49r1 should help with these issues, notably the use of 2m temperature observations in assimilation.

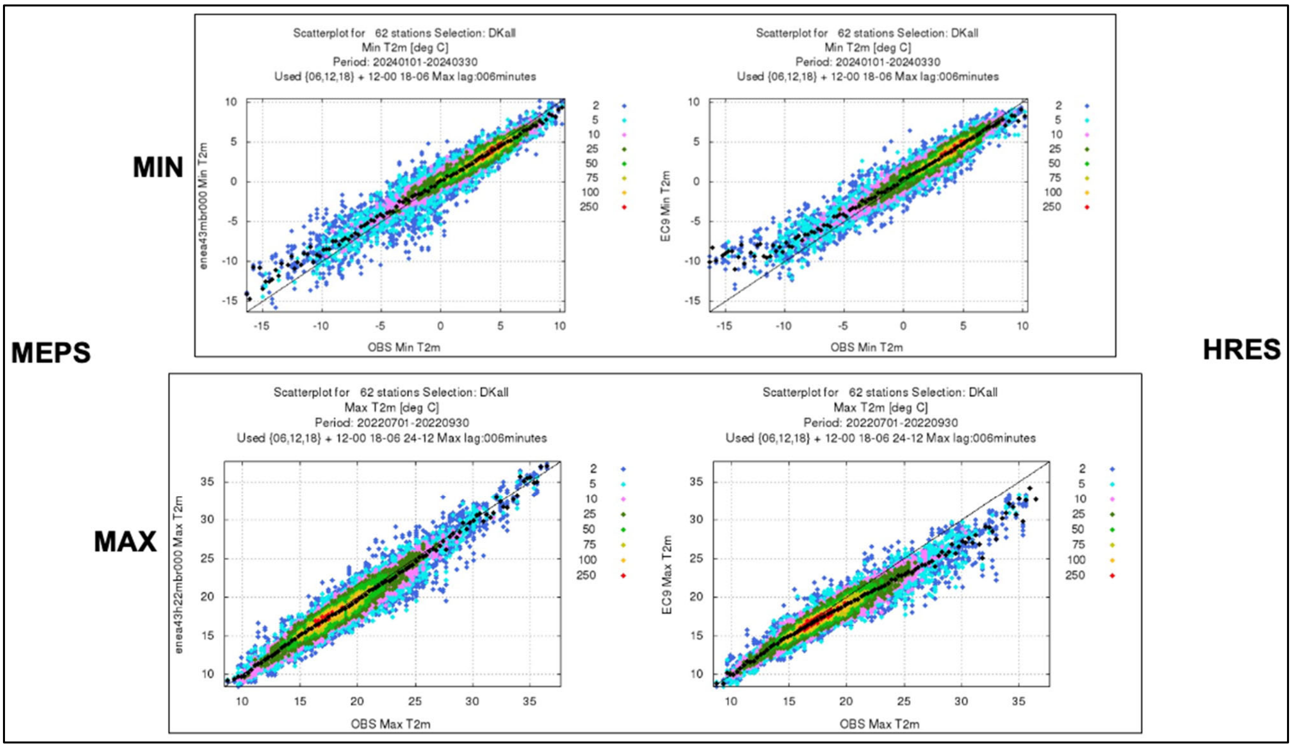


***Figure 13**. Day 1 verification scatterplots for 2m temperatures, for January to March 2024 for minima (top row) and July to September 2022 for maxima (bottom row), for stations in Finland, for HRES (right) and the higher resolution MEPS LAM (left). Case counts are shown by spot colours. Evidently when extreme minima or extreme maxima occurred MEPS performed better than HRES; HRES most commonly showed values that were not sufficiently extreme.*

The shortfall in maxima is somewhat harder to disentangle because the amounts of energy involved, per degree Celsius temperature change, are usually orders of magnitude greater than in the minima case. Probably the degree to which the IFS allows superadiabatic lapse rates in the heat of the day is important. Another (related) factor is the impact of urbanization. Urban tiles will be introduced into the land surface scheme in 49r1 and will sometimes help increase temperatures in large urban areas, which may help with error reduction in those locations.

The problems of topography-related 2m temperature biases are mentioned in several reports, notably for countries which are topographically complex. As well as using higher resolution LAMs another way to address problem is to post-process ECMWF output, using local methods. Croatia report, for example, that their “simple post-processing algorithm makes ECMWF temperature forecasts at seaside and in mountains very usable”. Related to this issue, in a collaborative initiative related to the German Waves-to-Weather project, we have been exploring how to improve upon the fixed lapse rate assumption ECMWF uses to generate the widely used site-specific meteogram forecasts.

To illustrate the benefits of post-processing in general (not specifically related to topographic effects) Figure 14 from Turkey shows, on the left and centre panels, the impact of using different types of non-homogeneous Gaussian regression. There are clearly improvements right out to day 15.

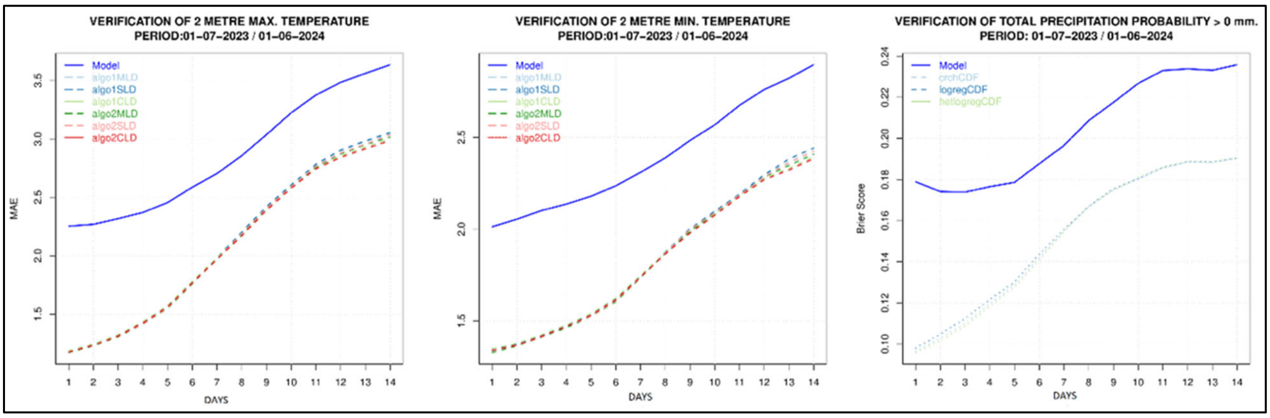


***Figure 14**. 11-month verification of ENS mean 2m maximum and minimum temperatures (MAE), and 24h ENS-based precipitation probability of >0mm (Brier score) – left, centre and right respectively - for stations in Turkey, for July 2023 to May 2024. Blue line shows raw model output, dashed lines denote output post-processed using Non-Homogeneous Gaussian Regression variants.*

Regarding the relative performance of global models, and consistent with the assessment presented in Haiden et al. (2023, 2024), Germany shows that ICON has been systematically catching up with HRES over the years in terms of RMSE, and indeed now exhibits smaller errors at T+24 (Figure 23, top panel). This relates in part to the introduction of an online parameter adjustment that is based on analysis increment values (see Zängl, 2023).

3.2.2. 2m dewpoint

Surface dewpoint temperature (Td) is a variable that mainly assumes importance because of related implications for other parameters, such as surface-based convection, fog and low cloud. And because of some direct connection with surface fluxes, notably in the summer half of the year, and thereby soil moisture (a variable not usually assimilated by models), correct representation is challenging. IFS performance in predicting Td has comparable skill to that of higher resolution models, such as ICON-GR (2.5km resolution) as reported by Greece, and MEPS (also 2.5km resolution) as reported by Estonia. And for a landlocked country such as Hungary – where surface fluxes probably assume even more importance - we see in Figure 15 that in 2023 HRES had smaller biases and smaller RMSEs than either AROME (2.5km resolution) or Aladin (8km resolution).

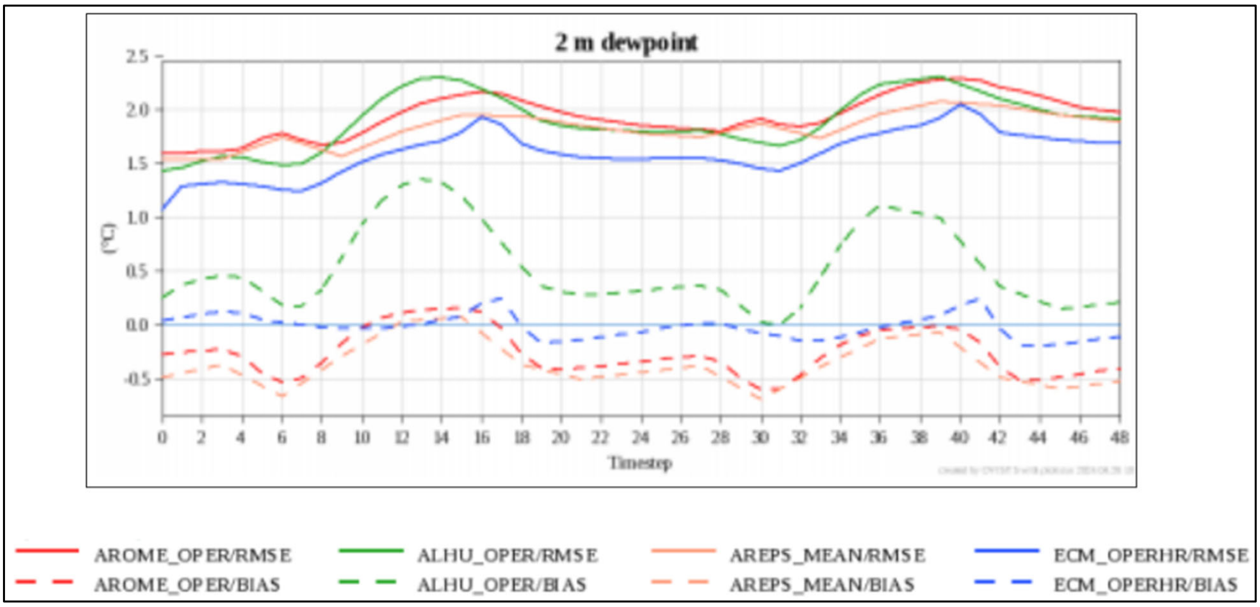


Figure 15. *RMSE (solid) and bias (dashed) for 2m dewpoint forecasts, for Hungarian SYNOP stations below 400m altitude, for 20203, for 00UTC runs of HRES (blue), AROME LAM (red, ~2.5km resolution), AROME EPS mean (peach, ~2.5km resolution) and ALADIN (green, ~8km resolution). X-axis shows timestep.*

3.2.3. Rainfall

Rainfall handling by the IFS is of particular concern for NMSs, because it is implicated in many high impact weather events (flash flooding and riverine flooding, naturally) and because correct prediction, particularly of local convection-related extremes, remains very challenging, as NMSs highlighted.

From the reports received, it is again clear that rainfall forecast performance in the IFS is better in winter than summer, due to the higher prevalence of convective rain in summer. This is even true at northerly latitudes, as Norway note.

Whilst there were a few positive comments about IFS output in reports, the fact that HRES and ENS under-predict the frequency of occurrence of both dry conditions and high totals, whilst over-predicting the frequency of small rainfall totals, were stressed many times.

Evidence for the small total and large total biases can be clearly seen on Figure 16 (from Hungary, blue and turquoise lines) and Figure 17 (from Italy, right panel). These figures also illustrate LAM behaviour. On Figure 17, even allowing for the different vertical scales (note lines labelled 1 and 3) we see a flatter profile for the COSMO 5km model – meaning biases closer to 1, generally. This is also true, strikingly, for Aladin (green line on Figure 16) even though its resolution (8km) is similar to that of the medium range IFS. Meanwhile AROME clearly had a strong over-prediction bias for very large totals, indicating that achieving a frequency bias of 1 by increasing resolution can be non-trivial. Such a bias, if unchecked by forecasters, would lead to a disproportionate number of false alarms.

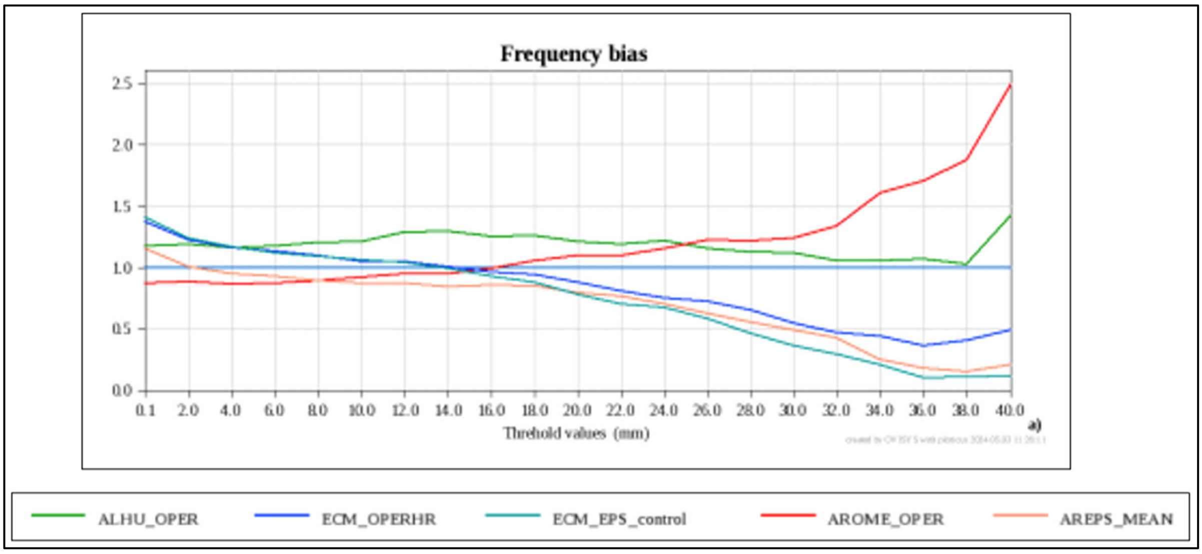


***Figure 16**. Frequency bias for threshold exceedance for 24h rainfall forecasts, for Hungarian sites, for Aug to December 2023. Lines show ECMWF HRES (blue), ECMWF control (turquoise), AROME (~2.5km resolution, red), the AROME EPS mean (peach) and ALADIN (~8km resolution, green). Striking features are the overprediction of large totals by AROME, the large total discrepancy between ECMWF HRES and ENS Control, and the flat profile for ALADIN, despite its "lower" resolution.*

Another feature of Figure 16 is the unexpected discrepancy between HRES and Control runs for totals greater than about 16mm/24h. This is hard to explain.

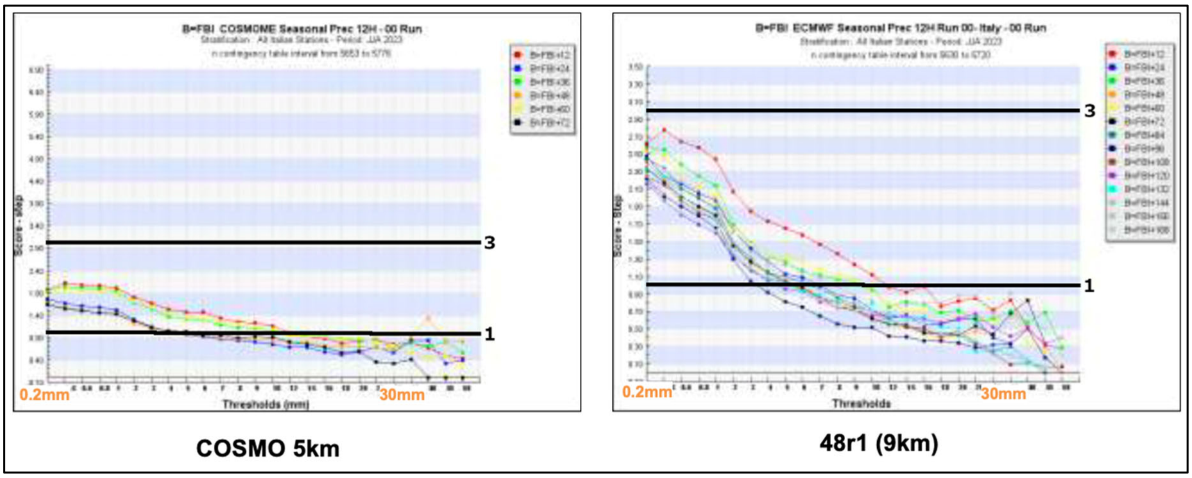


***Figure 17**. Frequency bias for threshold exceedance (x-axis) for 12h rainfall totals, for Italian sites, for JJA 2023, for forecasts by HRES (right, mainly cycle 48r1, also shown on the right panel of Figure 10), and 5km resolution COSMO LAM runs (left). Different colours denote different lead times, going up to day 7 (HRES) and day 3 (COSMO). In both models small totals are over-forecast whilst large totals are under-forecast, although these discrepancies are smaller in COSMO, very probably due to its higher resolution.*

Calibrated products can help overcome the frequency bias issues and improve also the discrimination ability of forecasts. For instance, ECMWF's ecPoint-Rainfall output (Hemri et al, 2022, and Gascón et al, 2024), available on ecCharts and OpenCharts, addresses not only sub-grid variability, but also compensates for weather-situation-dependant gridscale biases at the same time. Italy highlight the benefits and utility of ecPoint in their report.

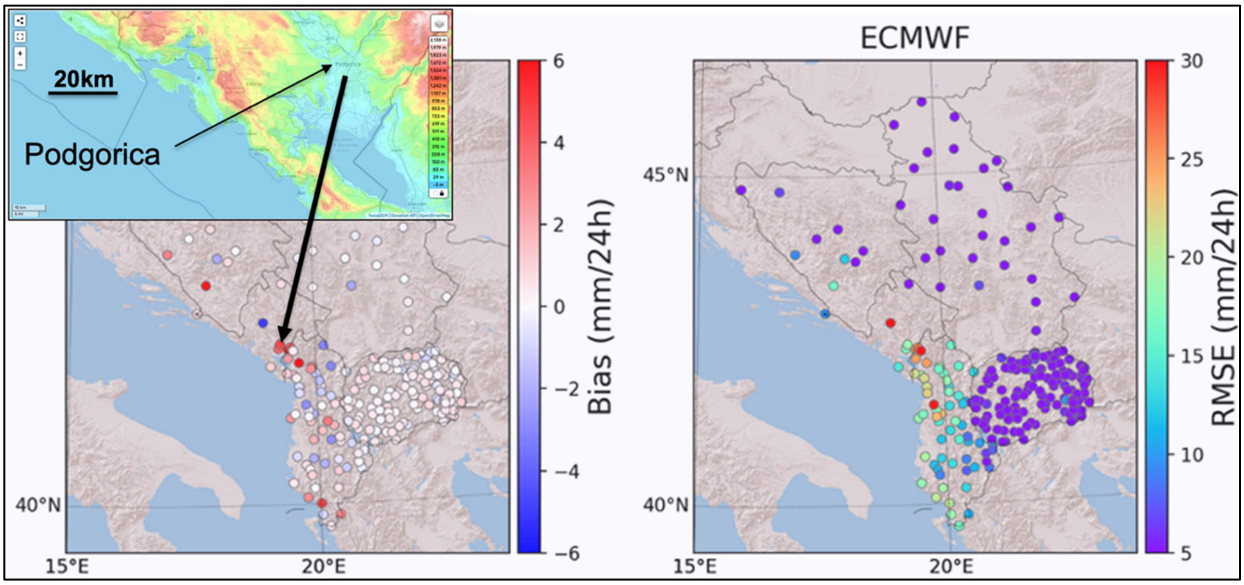


Figure 18*. HRES 24h rainfall forecast bias (left) and RMSE (right), as reported by Croatia (the verification period was not provided). The inset map shows the topography in and around a cluster of observations in Montenegro (near Podgorica) where there is a strong over-prediction bias.*

Topographic influences on precipitation can be quite profound, particularly (but not exclusively) in stronger wind situations, due to upslope enhancement, rain shadow effects and some overspill into the rain shadow area, all of which are potentially modulated by ice-to-liquid phase changes. Therefore, it is not surprising that biases and errors arise when the mountains themselves, and even hills, are not represented well because of model resolution. To correctly assess any issues, high density observations are essential. Data exchanged in the SEEMHEWS project has high density in some areas - see Figure 18, provided by Croatia. Amongst the countries depicted, the largest biases and RMSEs, and the biggest local variations in these, seem to be in Albania, which Croatia attribute to it being the most topographically complex.

Around Podgorica, in Montenegro, close to the Albanian border, we see large over-prediction biases. Possible reasons are indicated on the topography inset. Given that the main rain-bearing wind direction is likely to be S to SW'ly, a key factor there is probably rain shadow underestimation, to the lee of the narrow coastal mountain range to the S and SW (the highest peak there, Rumija mountain, reaches 1600m). Seasonal size variations of lake Skadar, NE of these mountains (140km^2 to 200km^2) may also contribute to model error, through flux modulation. Lake size is currently fixed in the IFS. So this is a nice example of how multiple factors can play into rainfall predictions issues, and reinforces the importance of high quality, high density data for process understanding and model improvement.

Switzerland and Croatia have invested in neighbourhood (signal spreading) post-processing approaches for precipitation forecasting, which are particularly appropriate for LAMs. These show the expected benefits of this post-processing, alleviating the double penalty effect somewhat, although at the same time Switzerland discuss the challenge of finding the right neighbourhood scale to use for comparison purposes. Later in their report Switzerland then show that their COSMO LAM ensemble outperformed IFS for both CRPS and MAE of the median, for 6h precipitation, with post-processing tending to double the advantage over ECMWF. That said, there was a sharp reduction in the benefits of raw and post-processed, relative to the IFS ENS, as lead times advanced from day 1 to day 4.

3.2.4. 10m wind

Whilst six NMSs report better wind speed performance from their respective LAM or LAM-EPS systems overall, two effectively rank them about equal to the IFS (Hungary and Ireland) and for one the IFS is superior (Serbia). Note that ECMWF expects significant improvements in perturbed ENS member handling of 10m winds in cycle 49r1, due to the introduction of a new stochastic physics scheme, which facilitates larger variability in surface variables.

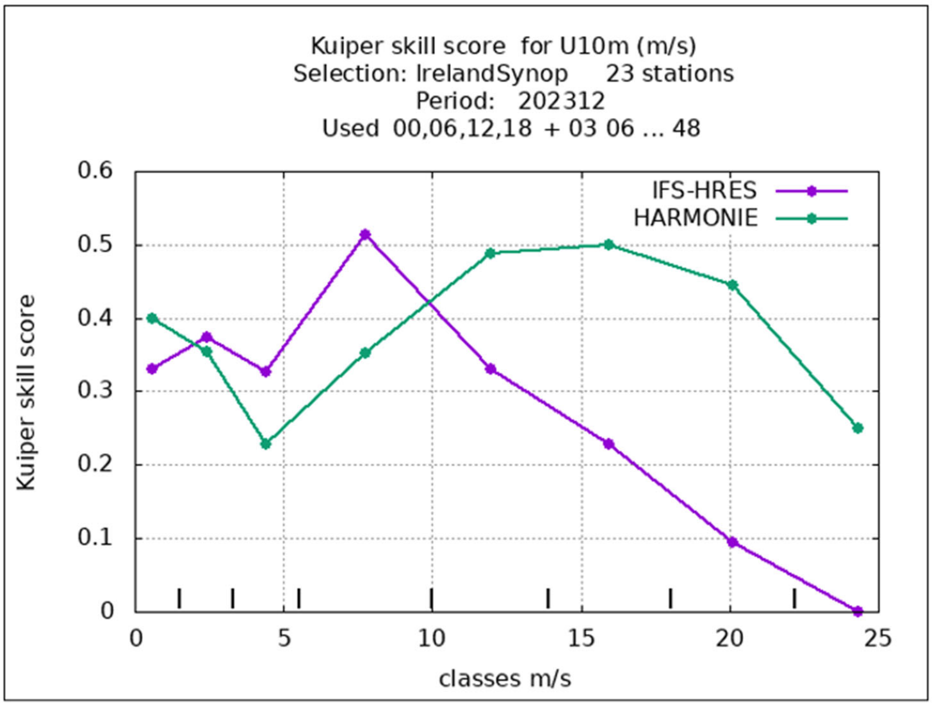


***Figure 19**. Kuiper skill scores for 10m wind speed for SYNOP stations in Ireland, in December 2023, for HRES (9km resolution) and Harmonie-AROME (2.5km resolution), using different wind speed classes (x-axis). Lead times verified are 3-hourly from T+0 to T+48.*

Of particular relevance for warnings is the handling of high wind speeds. Whilst the case count is inevitably small, Ireland (Figure 19), Denmark and Norway all note that very strong winds tend to be underestimated somewhat by the IFS, whilst LAMs do better. This probably relates to the narrowness of some strong wind swathes, and/or underrepresented sting jet phenomena, around cyclonic storms. Figure 20 shows the Met Office analysis of storm Rolf over Denmark that Denmark refer to: note the fine scale of the very strong gradients over the north of the country. In similar situations of cyclonic storms, the higher resolution (4.4km) Continuous Extremes Digital Twin model run for Destination Earth delivered stronger peak 10m winds (and gusts), within a narrow footprint, than the corresponding 9km HRES. This seems consistent with a resolution-based explanation, although it will also depend on storm size. Smaller cyclones are likely to be handled worse, in relative terms, by a lower resolution model, whilst for larger ones representation may be OK. Despite the verification issues, positive feedback on cyclonic windstorm handling also featured in reports (see section 3.4).

There is also evidence of a strong wind issue over sea areas, in ENS relative to the MEPS system, in ROC area scores from Finland in Figure 21. A new ocean wave parametrisation feature in cycle 49r1 will reduce the roughness of seas in very stormy conditions, and accordingly provide a slight but systematic increase in 10m wind output in the IFS under such conditions, all other things being equal.

Meanwhile, Figure 19 shows generally better performance of HRES for speeds <10m/s, which is encouraging, although at the same time the net downturn in ENS skill for marine gales over the last 3 years on Figure 21 is concerning.

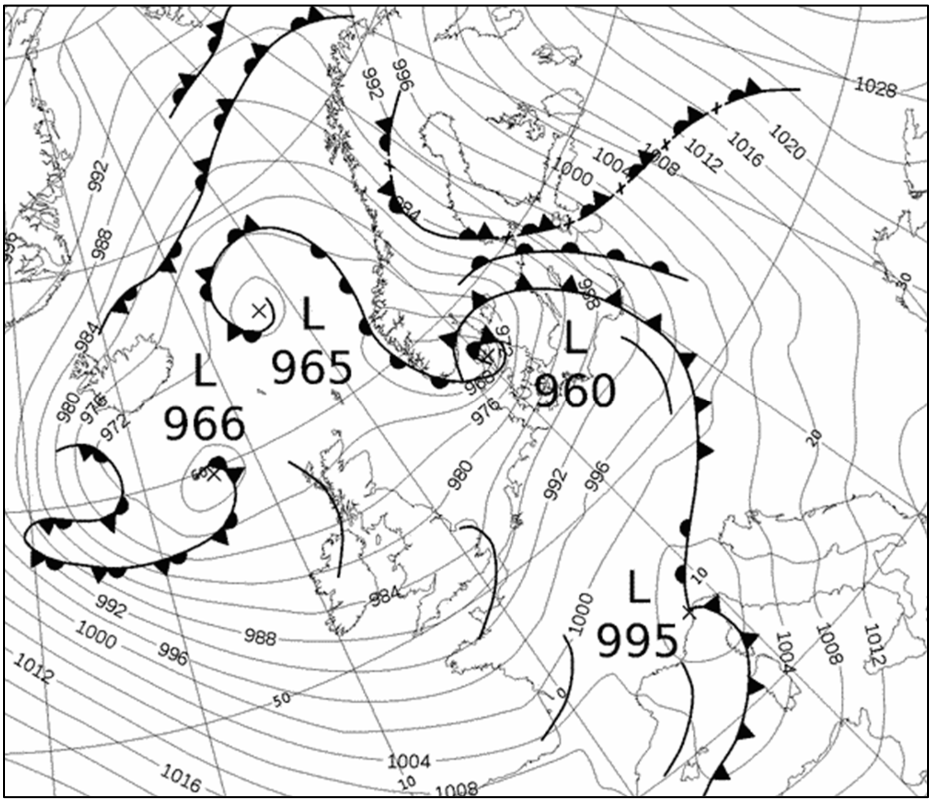


Figure 20. *Segment of a Met Office surface analysis chart for 06UTC Friday 23 Feb 2024, showing storm Rolf affecting Denmark.*

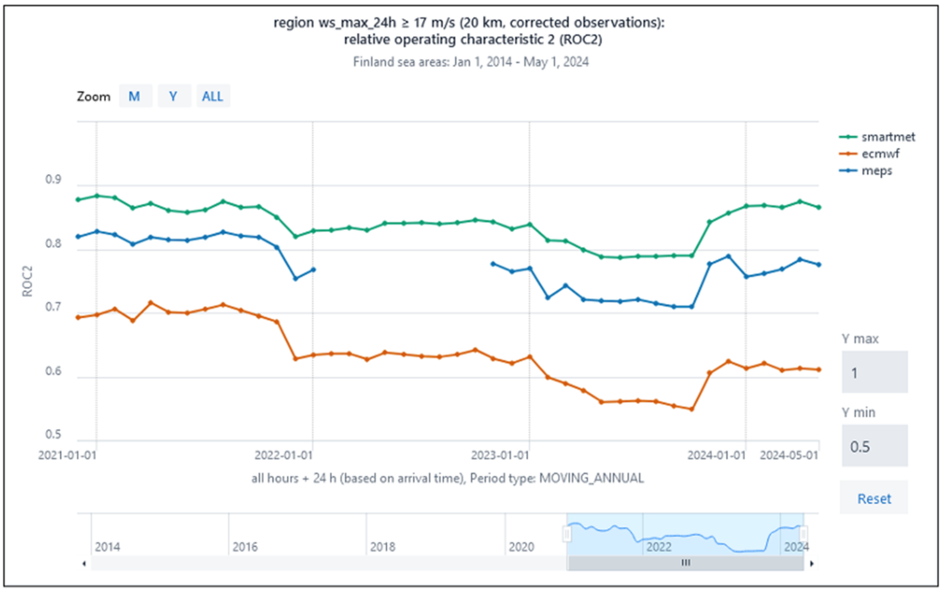


Figure 21. *ROC area verification scores for ensemble forecasts of gale force 10m mean winds for Finland sea areas, for 2021-2023 (x-axis), for ECMWF ENS (orange), the 2.5km resolution MEPS ensemble (blue), and warnings generated by the forecasters (smartmet, green). Lead time information was not provided. Values are 12-month moving averages.*

Norway, Spain and Iceland again note the systematic and long-standing negative 10m wind speed bias in IFS output over mountains, across multiple speed classes, already highlighted in previous reports. At the same time Iceland find that the wind EFI is not reliable, which is slightly surprising as this should normalise systematic errors. Interestingly, Latvia show that with post-processing it is possible to systematically reduce RMSEs for 10m wind speed, in both HRES and their HARMONIE LAM system, by more than 50%.

3.2.5. *Other parameters / levels*

With regard to representing the vertical structure of the atmosphere, only limited results were provided by the NMSs. However, results from Turkey for T+24, 48 and 72, for 3 months in 2022, on Figure 22, suggest that over the lower and middle troposphere HRES temperatures exhibit systematically smaller biases and smaller RMSEs than do two LAMs that they run. At levels of 300mb and above results are a bit more mixed.

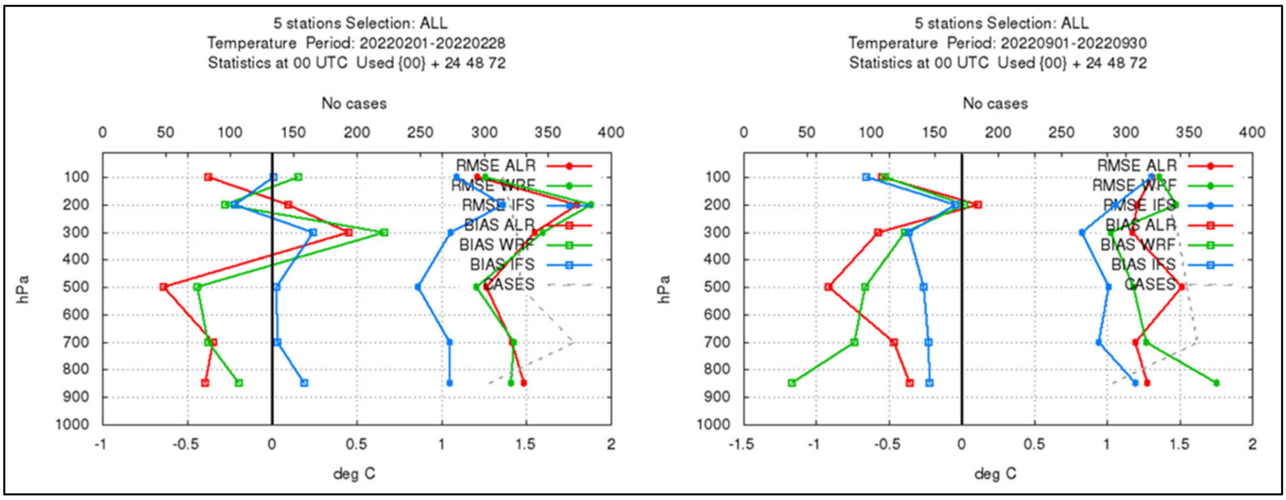


***Figure 22**. Bias (open squares, left groupings on each panel) and RMSE (closed squares, right groupings) for temperature forecasts at different pressure levels (y-axis), for January and February 2022 (left plot) and September 2022 (right plot), for radiosonde stations in Turkey, for 24, 48 and 72h forecasts, from HRES (blue), an ALARO LAM (red) and a WRF LAM (green). Thick black line is the reference point for zero bias and zero RMSE.*

Similarly, feedback on aspects like cloud cover, cloud base and visibility was limited, although the Finland report suggests visibility and cloud base tend to be better handled by IFS than by MEPS, which is noteworthy for aviation applications. Germany show a 7.5 year verification score sequence for cloud cover (oktas coverage from SYNOPs) for HRES versus their global ICON model (Figure 23, lower panel), and note the steady improvement, in relative terms, in ICON, such that it now performs better at short leads than HRES, and about the same at longer leads. They also note the apparent degradation in HRES scores over the years. This degradation is in large part due to the comprehensive new physics package introduced in cycle 47r3 in October 2021. Radiation at the surface was systematically improved in this cycle, which, from an energy standpoint, we view as more important for the overall integrity of the forecast. The lower scores shown relate mainly to the handling of high cloud which became more often 0 or 100% than hitherto. The dividing line between full cover and partial cover for high cloud can be difficult to define precisely. Nonetheless ECMWF is investigating this aspect with a view to future improvements.

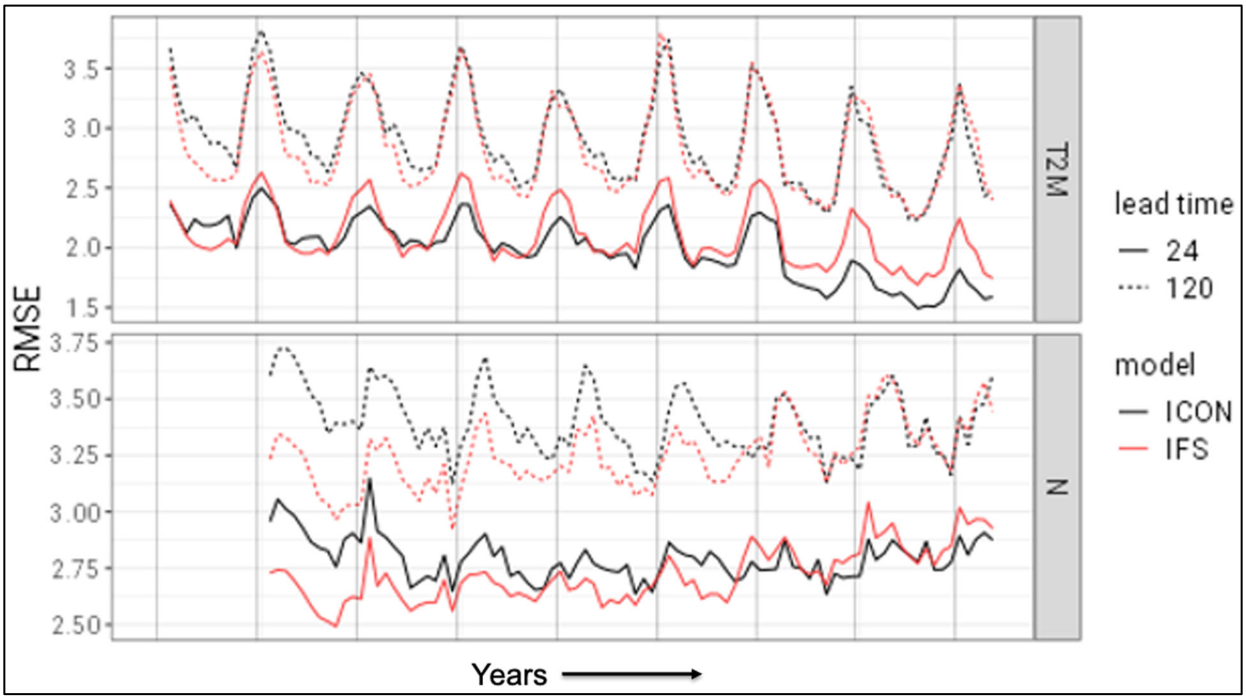


***Figure 23**. RMSE performance of HRES (red) and ICON (black) from 2016 to 2024 (x-axis), for 2m temperature (top graph) and total cloud cover (bottom), for lead times of 24h (solid) and 120h (dotted), for Northern Hemisphere stations, as reported by Germany. The reducing skill in HRES cloud cover forecasts was highlighted, as were the better ICON forecasts for short leads in recent years. This lead was replicated for many other surface variables (not shown here).*

Clear Air Turbulence (CAT) is an aviation parameter recently introduced by ECMWF; indeed ENS-based CAT probabilities only became available in ecCharts and OpenCharts in summer 2024. In one comment Switzerland, who reference our CAT fields when considering aviation SIGMETs, question why they deliver significantly lower values on average than the DWD equivalent (for example) and ask if the product has been verified. Bechtold et al (2021) describes the approach, with relevant verification.

Romania indicate that there can be curious oscillations in predictions of rainfall, snow and humidity between the intermediate (06 and 18UTC) and the main (00 and 12UTC) IFS runs, which causes them messaging issues.

Iceland note that IFS is good at handling low pressure centres, which can be critical given their location and vulnerability to extreme winds, and in this vein they regularly consult ECMWF cyclone database products, from about 7 days out.

3.2.6. Sub-seasonal and Seasonal

With two exceptions no substantive verification results were provided for these forecasting systems. Hungary provided month 1 to month 6 verification metrics for max, min and mean temperatures and precipitation in Hungary, for all of 2023 for the SEAS5 system. There was some positive skill, on average, for all temperature variables for all lead times (i.e. months), but for precipitation for all these leads the outcome was on average negatively correlated with the forecast. These are in line with expectations – some modest skill in 2m temperature forecasts due mainly to capturing the climate change signal; negligible skill for rainfall which can by chance manifest itself as a slightly negative correlation. Very limited results provided by Israel give a similar picture for Israel.

3.3. Subjective Verification

Figure 13 in the previous version of this memorandum, Hewson (2021), showed results of a 2021 forecaster survey in Estonia, regarding IFS performance for different parameters, as assessed subjectively by forecasters. Results of a 2024 repeat of this survey (albeit with some minor differences) are shown in Figure 24. Mostly rankings are similar, although forecasts of large scale precipitation now have a much better rank. Convective precipitation remains a problem area however. Visibility and ice conditions are viewed as the least satisfactorily forecast parameter, once again, with badly forecast ice conditions over the Baltic and over large lakes again indicated to have adverse knock-on effects for temperatures and wind forecasts in the vicinity. A new sea ice model in cycle 50r1 should go some way to helping with forecasts of ice over the Baltic.

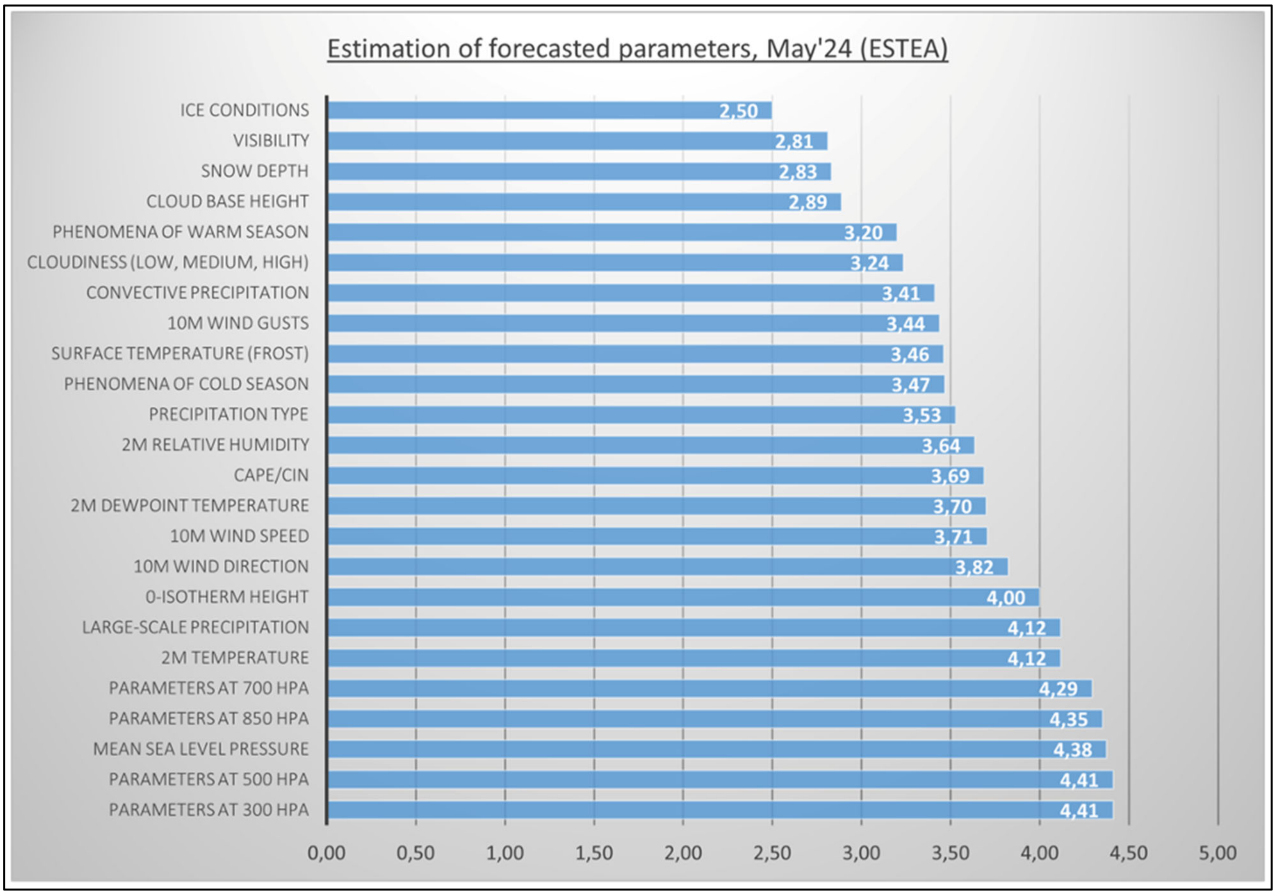


***Figure 24**. Results of a forecaster survey in Estonia in May 2024. Each forecaster could subjectively rate ECMWF forecast performance, for a variety of variables (y-axis), on a scale from 1 to 5, where 1 indicated "very bad," 2 "bad," 3 "average," 4 "good," and 5 "very good". All 17 forecasters responded; mean scores are shown above, by rank, with the worst performers at the top. This is in effect a recent update to Figure 13 in Hewson (2021).*

Finland forecasters have noted a tendency for stratus to be too widespread in winter-time anticyclones. This may relate to the known difficulties we have with mixed phase clouds in the Arctic. They also highlight an absence of coastal stratus formation in southerly flow in autumn - apparently an issue with "all models".

We note that a number of services have reported systematic biases that are already known about and documented on the "Known IFS forecasting issues" web page in the Forecast User Portal. We encourage our users to regularly refer to this webpage.

Finally, the UK make an interesting point linked to the free availability of an increasing number of IFS fields on OpenCharts (and through other websites). They say that “there have been occasions this last winter when media attention has been drawn to ECMWF snow depths yet Met Office expectation was that these depths would be overdone, thus making Met Office messaging of the forecast harder”. Clearer pointers to ECMWF web pages such as the “known IFS forecasting issues”, or the “forecast user guide” may help addressing this issue that will keep on increasing as the ECMWF open data roadmap is implemented.

3.4. Case Studies

Case studies are often triggered by poor model performance for a significant weather event that has led to both impacts and media interest. So, case study examples can be very helpful for ECMWF when deciding on priorities for model improvement and for product development.

This year we received many reports of cases, which we summarise here, with a few examples. Far more details can be found in the NMSs reports.

The online questionnaire version of this year’s survey allowed for explicit reference to up to 5 cases, good, bad or mixed. Most returns covered between 0 and 5, although some NMSs provided far more – notably Croatia who separately describe over 30! The vast majority of cases were in the medium range class, discussed in more detail below.

For sub-seasonal ranges Croatia noted how one period of colder weather in late February / early March 2023, and two in January 2024 (approximately 6th-10th and 16th-20th), that they link to sudden stratospheric warming events, were well captured in sub-seasonal output. For seasonal ranges Germany suggest that multi-model output (including SEAS5) for winter 2022/23, issued in autumn 2022, and in winter 2022/23, matched up quite well with the observed conditions. On the other hand, Croatia highlighted consistently poor SEAS5 forecasts for March 2022 in the lead up, including the February issue. Warm was predicted when the outcome was cold, linked to long-lasting high pressure.

For the many medium range cases, Table 1 classifies according to weather type, with Croatia’s contribution shown separately. By design there is likely to be a bias towards reporting poor IFS handling.

Evidently poorly handled cases of (severe) convection are quite commonplace, although reassuringly a few examples of good handling were also reported. The poor handling of severe convection in a global model is nothing new but note that ECMWF does try to provide additional targeted guidance products, such as vertical profiles and the CAPE-shear EFI, and is actively continuing its collaborations with ESSL with these issues in mind. Screen level temperatures can also be problematic, in ways that sometimes link to poor cloud cover handling, which itself seems to sometimes be a significant issue. Countries that are climatologically different – Finland, Sweden and Israel – all reported cases where low cloud cover and cloud ceiling were too pessimistic. Whilst reasons for these poor forecasts may not be the same, the negative impacts, for aviation, may be similar. Snow details and heavy rainfall (likely unrelated to convection) also feature. Spain provide a dramatic example of 2m deep snowfall measured (above 1200m) when none was forecast! This case from late February 2023 was over Majorca, where model topographic heights are limited by spatial resolution, although some snow did fall even at sea level. The most positive feature of Table 1 is probably the favourable handling of cyclonic windstorms, reported 6 times by NMSs. Whilst there are still some issues with extreme low-level winds around these, as reported in Section 3.2.4, it seems that forecasters have found IFS guidance relatively useful for recent cases.

NMS cases excluding Croatia (**bad forecasts**)		Croatian cases (**bad forecasts**)	
(Severe) convection	6*	Convection	10
2m temperatures	5	Jumpiness	8
Snow details	3	Heavy rain	4
Low clouds / ceiling (too pessimistic)	3	Low cloud / fog	4
Freezing rain	1	Cloud generally	2
Extreme rainfall	1	2m temperature	2
Jumpy forecasts	1	Freezing rain	1
NMS cases excluding Croatia (**good forecasts**)		Bora wind	1
Extra-tropical cyclone and related gusts	6	Timing issue	1
(Severe) convection	4	Lightning	1
Freezing rain	2*	Snow cover	1
Cyclone and extreme rainfall	2*		

Table 1*. Case study classifications and counts, from Croatia (right side) and other NMSs (left), for medium-range forecasts. Croatia intrinsically focussed on poorly handled cases; more reports of good performance were provided in the "other NMS" category. Stars denote classes for the 3 cases depicted in Figures 25-28.*

Latvia provided two examples of well-handled freezing rain events. One is shown in Figures 25 and 26. Whilst the distribution of freezing rain in the forecast shown is imperfect, values were at about the right general level compared to the observations. It is good to see that the freezing rain accumulation output, that ECMWF developed specifically for users in NMSs, helps to combat impacts.

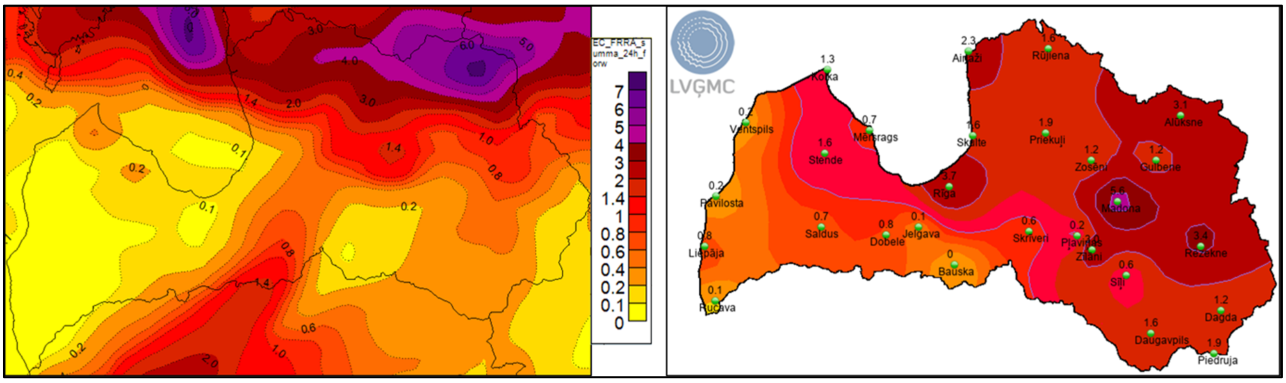


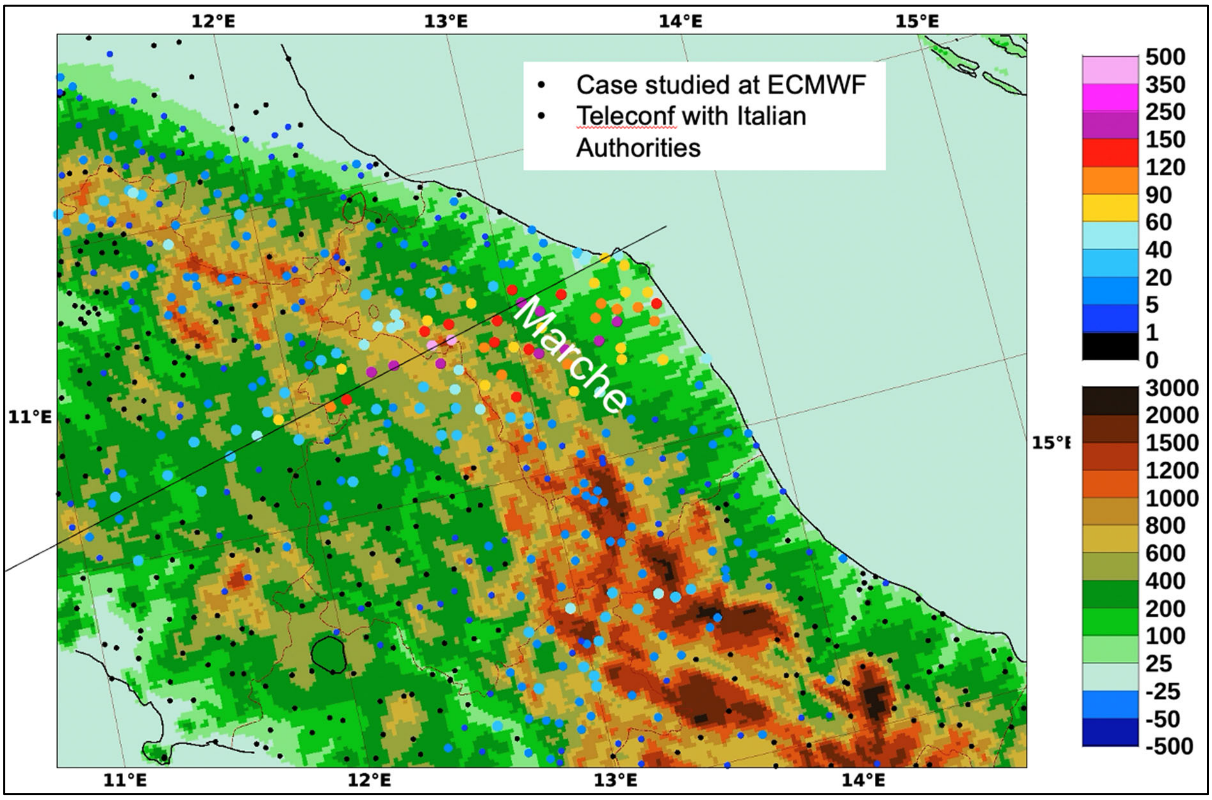
***Figure 25**. 72h forecast of freezing rain accumulation (in mm) over Latvia, from 18UTC HRES run on 10 February 2024 (left panel), with verifying observations (right panel).*

***Figure 26**. Ice accumulation from the freezing rain event.*

Figure 27 relates to a tragic case, highlighted by Italy, when IFS forecasts of convective rainfall totals over mainland Italy were very misleading. Over 300mm accumulated in 6 hours, locally, with storm activity substantially spilling over to the lee (northeastern) side of the Apennines when the IFS placed its heavy rainfall (admittedly with smaller totals) on the upwind side. This case underwent detailed investigations and further exchanges between ECMWF and Italy in the following months.

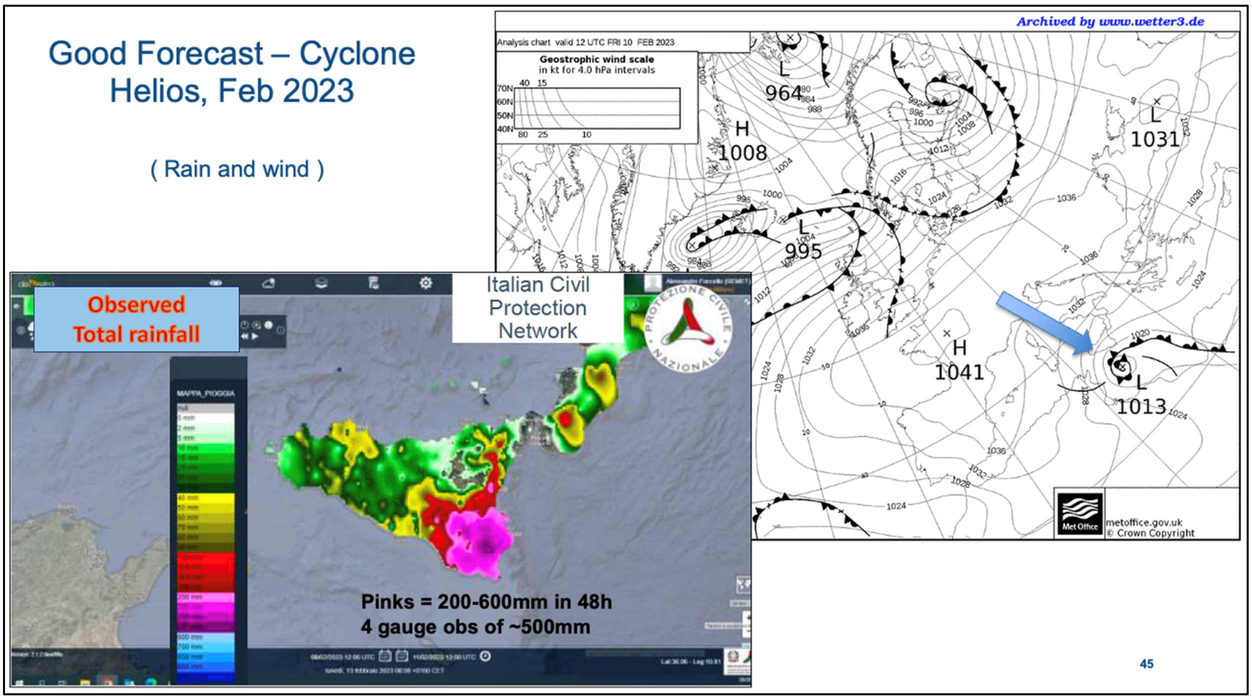
Conversely, Figure 28 shows the well forecast case of cyclone Helios, that delivered extreme rainfall and flooding to Sicily early in 2023. According to Italy, extreme winds in the circulation of Helios were also well predicted. When heavy rainfall is tied to a significant surface cyclonic feature, like this, rather than being driven by slow-moving convective cells or the thermodynamic upscaling of convective activity (as in e.g. mesoscale convective systems or vortices), the IFS can often perform rather well.

Since 2014, the Severe Event Catalogue in the Forecast User Portal has displayed a collection of materials for specific weather events, focusing on ECMWF meteorological and environmental forecast performance, with inputs from NMSs. Some events reported by NMSs as case studies are covered in the Severe Event Catalogue. We encourage NMSs to refer to this resource and provide ECMWF with feedback and inputs for any cases relevant to them.



- Case studied at ECMWF
- Teleconf with Italian Authorities

Figure 27. *ECMWF-generated data for the Le Marche floods of 15 September 2022, which killed at least 11, highlighted by Italy as an example of poor quality ECMWF forecasts. On the map spots show 12h rainfall in mm up to 00Z 16th, whilst shading represents topography at 1km resolution; fine lines are regional boundaries.*



Good Forecast – Cyclone Helios, Feb 2023

(Rain and wind)

Observed Total rainfall

Italian Civil Protection Network

Pinks = 200-600mm in 48h
4 gauge obs of ~500mm

Figure 28. *Plots relate to a case study from 10-11 February 2023, of cyclone Helios that brought extreme rainfall, flooding and very strong winds to southernmost Italy, highlighted by Italy as an example of "excellent ECMWF forecast performance". The Met Office synoptic chart for 12UTC on 10th shows Helios (an extratropical cyclone, not a medicane), whilst the second panel shows 48h rainfall totals.*

References

Bechtold, P., Bramberger M., Dörnbrack A., Leutbecher, M. & Isaksen, L. (2021) Forecasting clear-air turbulence. ECMWF Newsletter, **168**. 32-37. https://www.ecmwf.int/sites/default/files/elibrary/072021/20109-newsletter-no-168-summer-2021_1.pdf

Ben Bouallègue, Z., Clare, M. C., Magnusson, L., Gascon, E., Maier-Gerber, M., Janoušek, M., ... & Pappenberger, F. (2024). The Rise of Data-Driven Weather Forecasting: A First Statistical Assessment of Machine Learning–Based Weather Forecasts in an Operational-Like Context. *Bulletin of the American Meteorological Society*, *105*(6), E864-E883.

Gascón, E., Montani, A. & Hewson, T.D. (2024) Post-processing output from ensembles with and without parametrised convection, to create accurate, blended, high-fidelity rainfall forecasts. *Quarterly Journal of the Royal Meteorological Society*, 1–29. Available from: https://doi.org/10.1002/qj.4753

Haiden, T., Janousek, M., Vitart, F., Ben-Bouallegue, Z. and Prates, F. (2023). Evaluation of ECMWF forecasts, including the 2023 upgrade. ECMWF Technical Memorandum 911. https://doi.org/10.21957/d47ba5263c

Haiden, T., Janousek, M., Vitart, F., Tanguy, M., Prates, F., Chevallier, M. (2024). Evaluation of ECMWF forecasts. ECMWF Technical Memorandum 918. https://doi.org/10.21957/52f2f31351

Hemri, S., Hewson, T., Gascón, E., Rajczak, J., Bhend, J., Spirig, C., ... & Liniger, M. A. (2022). *How do ecPoint precipitation forecasts compare with postprocessed multi-model ensemble predictions over Switzerland?*. ECMWF Technical Memorandum 901. https://doi.org/10.21957/hy89j7svk

Hewson, T., 2021. *Use and verification of ECMWF products in member and co-operating states (2021)*. European Centre for Medium Range Weather Forecasts. https://doi.org/10.21957/vp4z0x4yo

Lang, S., Alexe, M., Chantry, M., Dramsch, J., Pinault, F., Raoult, B., ... & Rabier, F. (2024). AIFS-ECMWF's data-driven forecasting system. *arXiv preprint arXiv:2406.01465.*

Zängl, G. (2023) Adaptive tuning of uncertain parameters in a numerical weather prediction model based upon data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **149**(756), 2861–2880. https://doi.org/10.1002/qj.4535

NMS Reports that contributed to this technical memorandum: https://sites.ecmwf.int/publications/greenbook/

The ECMWF “Forecast User” portal: https://confluence.ecmwf.int/display/FCST/Forecast+User+Portal

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

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

ECMWF Severe Event Catalogue: https://confluence.ecmwf.int/display/FCST/Severe+Event+Catalogue

Appendix – Requests for New Products

Users were asked to list new product requests/ideas, allowing for inclusion of up to 5 items. Most NMSs provided between 1 and 5. Responses were quite wide ranging, a few going a little beyond a strict definition of "product requests", but nonetheless providing useful ideas and insights. All responses have been reproduced in the table below, assigning to different classes where common themes were identified. Whilst the requests still need to be discussed and managed internally at ECMWF, a few initial comments have been added in the third column.

#	Class and Title of request	Request Details	Comment
	Convection-related		
1	Mixed layer cape	In ecCharts	
2	Lapse rate in different layers; hail products	It would be beneficial to have „Lapse rate" product in different layers, same as „Bulk wind shear" product, and some product for hail probability or possible size.	Hail products are in development
3	Lightning forecast documentation	Commonly, the HARMONIE lightning post-processed forecast differs from the ECMWF one. A link to the ECMWF documentation regarding this product would be appreciated.	
4	Automated Severe Weather Guidance	Severe weather probabilities based on IFS ENS and algorithm similar to ESSL, if possible. Take a look at https://www.stormforecast.eu/	
	Geographical		
5	New map area	We would like a predefined map that covers the Nordic countries including the Spitsbergen area. The Northwest Europe map do not cover Spitsbergen.	
6	Errors in "cities" dataset	We would like to ask you to update the base map of the EcCharts, where our cities are written incorrectly e.g. Bat'umi which should be Batumi, without apostrophes, this is just one example, there are a lot of locations with incorrect names. Also, there are missing very important cities/resorts not included. We would like to update coordinates, which are much better after the big update which happened several months ago, but there are some places missing or incorrect.	
	Reforecasts		
7	30y reforecasts	DWD presents subseasonal, seasonal and decadal predictions on a common website (www.dwd.de/climatepredictions). For a better comparison between all timescales of this 'seamless prediction' (e.g. anomalies w.r.t. to common reference period based on hindcasts) and better verification of extended range forecast we would need a reforecast with lead times up to 46-days for 30 years. Currently, the reference period for seasonal and decadal predictions is 1991-2020 as recommended by WMO.	

#	Class and Title of request	Request Details	Comment
8	Data sets over long periods of time	For a variety of uses, Météo-France needs replays/reforecasts over long periods, on the order of a year. Hereafter are two examples of use. Météo-France needs at least one year's worth of data on chemical boundary conditions to prepare for the MOCAGE OPER switchover. This replay must be available no later than 4 months before the switchover, to allow time for evaluations to be carried out and for production to be adapted if necessary. This need has already been indicated to CAMS Global as part of the CAMS2_40 (CAMS regional) contract, and concerns all 11 regional models participating in this service. At the medium-range horizon, forecast data for a long enough time period are needed for proper calibration of model outputs before a new ECMWF model goes live. Otherwise, it can lead to systematic errors in the MOS technique and introduce critical biases.	
	Vertical profiles		
9	Sounding forecasts	Sounding forecast at least one week in advance.	Currently stops at T+120.
10	Adding vertical profiles to open data charts	Adding vertical profiles to open data charts in the same manner that users can access meteograms would be a welcome additional capability.	Already available: stand-alone and clickable options.
11	Vertical profiles every 3 hours	Vertical profiles are currently depicted only every 6 hours on ecCharts (0, 6, 12 and 18 UTC). It would be really helpful to have them every 3 hours, especially at 15 UTC when diurnal convection is usually at its peak of activity.	
12	ecCharts profiles	When using vertical profiles in ecChart you get overloaded by data and it’s hard to detect details which can be quite important when using this in aviation forecasting and in severe precipitation predictions.	
13	tephigrams in Metview	Metview is used to plot forecast (HRES and AROME) and observed (radiosonde) tephigrams. It would be most useful if one could have the values of the geopotential height shown on the right vertical axis of the plot.	
	Accessibility issues		
14	Response times of ecCharts / openCharts	It would be very much appreciated if the response/refresh time of the ecCharts would be quicker.	
15	MARS request issues	ENS outputs are too large to be stored on a single magnetic tape. As a result, ENS data requests on MARS usually fail in timeout.	
	Dynamics-related		
16	EC Charts: Positive Vorticity Advection (PVA)	...@ several levels, especially at high levels (e.g., 500mb, 300mb, 200mb).	

#	Class and Title of request	Request Details	Comment
17	Q-vector divergence field	Q-vector divergence field on various pressure levels (like ESSL uses in their weather data displayer)	
	EFI-related		
18	Extending EFI/CDF point product beyond day 5	The EFI output has been one of the more useful products, as well as the EFI/CDF point product which allows users to look at the lead time variation of EFI and the CDF for temperature, wind and precipitation. What would be a useful development is whether this could be extended to longer lead times as this only starts plotting at day 5. This would allow emerging signals towards the beginning of the period at which the Met Office can issue public service (NSWWS) warnings (day 7) to be picked out. The example below is the day 5 EFI/CDF for Reading and there are only two model runs plotted which means that trends may not become meaningful/confident to users until day 2 or 3.	
19	Floating 24 hour period not 00UTC to 00UTC for EFI	As was discussed during our Member State visit in October 2023 our forecasters would like to be able to access (via OpenCharts) a floating 24-hour period version of the EFI product.	
	Regimes		
20	Finer-grained classes of weather regimes for IFS-EXT	e.g. following the works by former ECMWF fellow Christian Grams, see. E.g. https://www.ecmwf.int/en/newsletter/165/meteorology/how-make-use-weather-regimes-extended-range-predictions-europe	
21	Additional weather regimes in extended range and seasonal forecast	If possible, extension of weather regimes probabilities to Greenland Anticyclone, Summer Atlantic Low, Summer Zonal, Summer Blocking.	
	Temperature-related		
22	Universal Thermal Climate Index	Operational output of the UTCI (deterministic and EPS)	Coming with cycle 49r1.
23	Heating/cooling degree days	If possible, heating/cooling degree days on all temporal scales.	
24	Maximum/minimum 2-m temperature probability at 24-h intervals	Some time ago this output was available at 24-h intervals, in addition to the 6- and 12-h intervals. However, some months ago the interval was restricted to 6 h. From an operational perspective, this is not very useful, since we are usually interested in forecasting the maximum and minimum temperatures of the day. I would suggest returning to the 24-h interval, so that we can estimate the probability of exceeding certain temperature thresholds on a day (this is closely related to the issuance of warnings).	
	Fire Products		
25	FWI sub-indices	It would be nice to be able to consult a product for each FWI subindex, as well as the anomalies of each one in ecCharts.	

#	Class and Title of request	Request Details	Comment
26	Fire layers documentation	We cannot find the documentation relating to the new fire layers. For example, it would be good for us to know the ranges and levels of the indices.	
	ecCharts (miscellaneous)		
27	EC Charts: Liquid Water Content (LWC)	...@ several levels, especially at low levels (e.g., 1000mb, 950mb).	Total column water is available and may be of some use here?
28	EC Charts: Satellite Airmass Simulation	Similar to the existing Vis, IR, WV products.	
29	comparison of fields in ecCharts Dashboard	It would be interesting to have the possibility to compare fields, to see more than one field at once, in the ecCharts Dashboard.	This is already possible (albeit without much interactivity).
	Products (miscellaneous)		
30	Radar reflectivity simulation	Radar reflectivity simulations (max dBZ) for HRES and reflectivity threshold probability from ENS, if possible. Particularly in the future in case of higher IFS spatial resolution.	
31	Horizontal visibility	It would be helpful to have a product with the horizontal visibility.	Already available.
32	Crossing Point Forecasts	A valuable product to disseminate in the future would be the Crossing Point Forecasts	
33	Significant weather	A kind of extension of precipitation type product with a fog and T-showers/storms if possible.	
34	Including the 800 hPa level in the dissemination	This level is really useful in places like the Canary Islands. It is used operationally for determining possible weather warnings.	
35	Lead time aligned IFS-EXT tercile maps	e.g. W2 tercile maps with W2 being W2 w.r.t. to forecast reference time) along with the existing calendar week aligned tercile maps (with W2 being next calendar week running from Mo to Su).	
	Other		
36	ENS uncertainty statistics in dissemination	It would be really useful for us to access some fields related to the uncertainty of forecasts. Some percentiles, standard deviation, interquartile range or some other metrics related to these ones, would be interesting that they could arrive directly by dissemination.	

#	Class and Title of request	Request Details	Comment
37	ERA5-Land parameters	ERA5-Land is a high-resolution reanalysis and can therefore be applied well to regional processes and impact models. Unfortunately, some important parameters are not provided, for example: sea level pressure, maximum and minimum temperature or 10m wind gust	