SafeWind

Collaborative project funded by the European Commission under the 7th Framework Program, Theme 2007-2.3.2: Energy

"Multi-scale data assimilation, advanced wind modelling & forecasting with emphasis to extreme weather situations for a safe large-scale wind power integration"

Grant Agreement N°: 213740

Deliverable Dp-5.4

"Suitability and feasibility of integrating Limited-Area EPS (LEPS) for WPF applications"

Appendix J

Skill assessment of ECMWF IFS & EPS

DOCUMENT TYPE	Deliverable
DOCUMENT NAME:	swind_deliverable_Dp-5.4_V1.1_Apdx_J.doc
VERSION:	V1.1
DATE:	2011.02.18
CLASSIFICATION:	R0: General Public
STATUS:	Approved/Released

Appendix J: Skill assessment of ECMWF IFS & EPS

Most European LEPS platforms comprise a Control Forecast (CF) besides EPS members. Validating different LEPS platforms, a basic reference system's performance should be established. Such a system has been the ECMWF EPS. The idea behind this is that in order for a LEPS scheme to be considered as an appropriate system to be used in WPF, it should reflect to higher skill scores than SafeWind's backbone system EPS. Furthermore for each LEPS scheme the performance of the corresponding operational component is also assessed as in case of EPS and its corresponding high-resolution operational IFS component.

1. Details of the 2007-2009 period verification (model configuration before 26 January 2010)

Verification statistics for all ECMWF deterministic and probabilistic components were estimated for two years (731 days), e.g., from December 2007 to November 2009, using two different resolution data sets. The low regular 2.5 x 2.5 degree resolution was utilized for surface global fields (EPS model level 62), while a higher resolution 0.5 x 0.5 degree was used for various European areas of interest, both at surface and at various model levels. The model levels used in this study are: ml 61 (~35 meters), ml 60 (~70 meters) and ml 59 (~110 meters), shown in Figure 2.1.1. During the two-year verification period (December 2007 to November 2009), ECMWF IFS had a horizontal resolution of 25 km with a 91 vertical level scheme, while the EPS has a horizontal resolution of 50 km and a vertical resolution of 62 levels. As mentioned already, the EPS combines one control run (CF), started from the same analysis as the IFS HR, and 50 perturbed members. The chosen common forecast horizon of IFS and EPS predictions has been chosen as the S-R and early M-R (e.g., maximum horizon of T+120 hours: 5 days), with all forecasts initiated from 00 UTC daily analysis fields and verifying on both the 00 & 12UTC analysis cycles, i.e., implementing a separation (verification) step of 12 hours. It is important to point out that besides surface variables (as winds at 10 meters), additional variables at different model levels were considered corresponding to different heights at the lower part of the Surface Layer (~0.5 to 1km).

- *Verification methodology*

A new verification package developed at ECMWF is used to verify probabilistic and deterministic forecasts against analyses. This new verification methodology is now used to monitor the performance of the operational EPS [J.1]. The main differences with "older" packages used for the verification of IFS and EPS are:

- the use of a new long-term climatological distribution to define events and skill scores; and
- probabilistic scores can also be computed for the EPS Control Forecast (CF), Ensemble Mean (EM) and the High-Resolution deterministic operational forecast (HR), by constructing appropriate "synthetic" Probability Density Functions (PDFs).

The climate used in the new probabilistic verification is based on ERA-40 [J.2] analyses, which is expected to provide the most accurate available, consistent, long-term description of the atmosphere. The new climatology consists of daily fields of the mean, standard deviation and quantiles of anomalies. An example of such a "mean" field is given in Figure J.1 for wind speed at 125 meters.

In order to obtain good estimates globally, the climate is based on the years 1979–2001 during which satellite data constrain the analysis well in the Southern Hemisphere. For each day of the year, statistics are based on a 61-day window centred on the day of interest. The statistics are computed with weights that get maximum values at the window centre and gradually decrease to zero at days. Thus, dates contribute to the climate statistics of one day. These variable weights are superior to constant weights in terms of resolving the annual cycle and in filtering high-frequency sampling uncertainty. Using the same methodology, corresponding fields of mean, standard deviation of anomalies, and quantiles of anomalies are constructed for various vertical levels of the numerical scheme (model) used in the Reanalysis project.

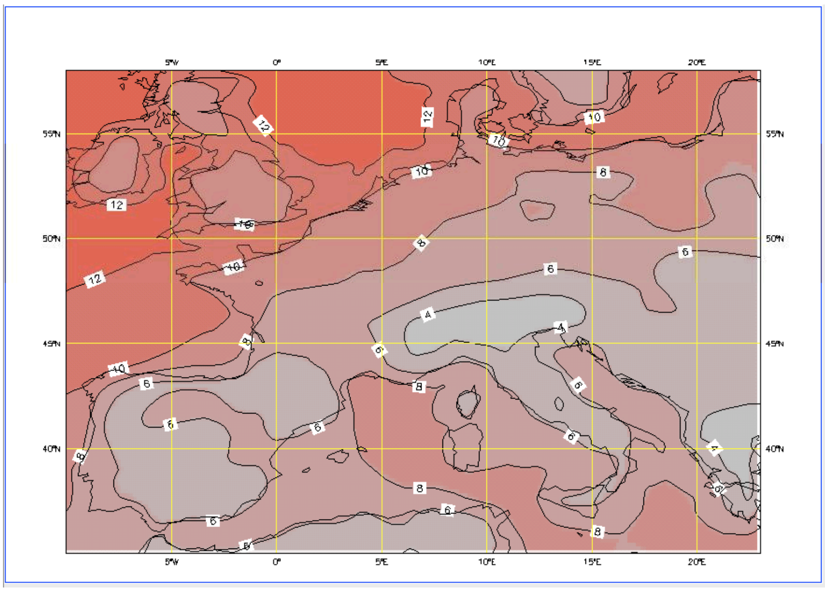


Figure J.1: *New climatological field of "mean" wind speed at Reanalysis level 57 (at about 125 meters) valid for the 1st day of January.*

It becomes obvious why this new verification (package) is considered of great importance. We should keep in mind that ensemble forecasts should represent the "true" distribution of probabilities that can develop from an initializing analysis, because these probabilities should reflect to the possibility of an early warning in case of an extreme wind event.

- *Verification areas*

Since international partners as Australia, India & Korea are participating in the SafeWind project, an appropriate selection of different verification areas had to be made in both the *low-res* and *high-res* verification modes. In global (low-res) mode, a selection of 20 verification areas was made, while in European (high-res) mode, a selection of 15 verification areas was decided (as shown in Figure J.2).

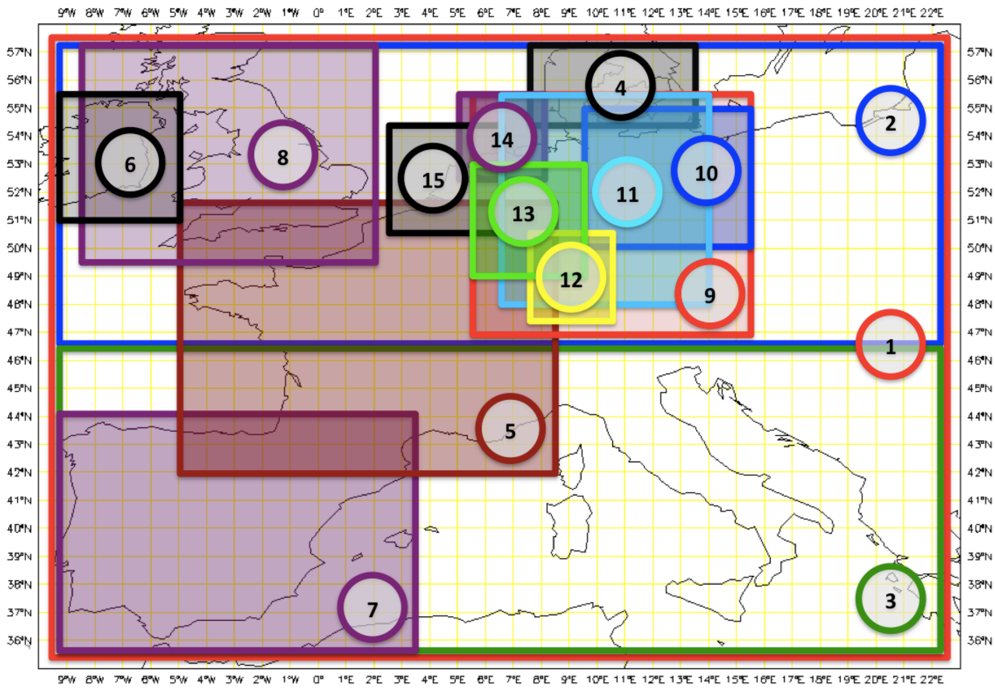


Figure J.2: *European (high-res mode) 15 verification areas.*

In more details the 20 global (low-res) areas are listed below. Most of European verification areas are listed in both the low- and high-res categories.

- northern hemisphere extra-tropics (G.1)
- southern hemisphere extra-tropics (G.2)
- northern hemisphere mid-latitudes (G.3)
- northern hemisphere sub-tropics (G.4)
- greater europe (75N/12.5W/35N/42.5E) (G.5)
- mediterranean (G.6)
- asia (G.7)
- australia (G.8)
- india (G.9)
- korea (G.10)
- europe (58N/10W/35N/23E) (G.11)
- Northern europe (58N/10W/46.5N/23E) (G.12)
- southern europe (58N/10W/35N/23E) (G.13)
- denmark (G.14)
- france (G.15)
- Ireland (G.16)
- Spain (G.17)
- United kingdom (G.18)
- germany (G.19)
- fino1 platform area (55.5N/ 5E/52.5N/8E) (G.20)

In European (high-res) mode, the selection of 15 verification areas was made as shown below:

- europe (E.1)
- northern europe (E.2)
- southern europe (E.3)
- denmark (E.4)
- france (E.5)
- ireland (E.6)
- spain (E.7)
- united kingdom (E.8)
- germany (E.9)
- germany vaten energy subarea (E.10)
- germany eon energy subarea (E.11)
- germany enbw energy subarea (E.12)
- germany rwe energy subarea (E.13)
- fino1 platform area (E.14)
- cabauw: 54.5N/2.5E/50.5N/6.5E (E.15)

- *Verification statistics*

A set of different score statistics were used focusing on different aspects of the skill of IFS & EPS components (as shown in Table J.1). For binary events (probabilistic mode), the Brier Score, the Ignorance Score and the Area under the Relative Operating Characteristic (ROCA) Curve were computed [J.3]. A selection of different events was made reflecting to positive anomalies larger than 0, 0.5, 1, 1.5 “typical” climatological standard deviations. Additional thresholds corresponding to the 90%, 95% & 99% percentiles were also considered referring to extreme/rare wind events. For multiple categories (probabilistic mode), the RPS (Ranked Probability Score) and CRPS (Continuous Ranked Probability Score) were computed. These are equivalent to the arithmetic mean of the Brier Scores for exceeding thresholds that separate predefined categories [J.4]. Ten selected climatologically equally likely categories were used to define RPS/CRPS thresholds. In that sense, RPS/CRPS constitutes the arithmetic mean of Brier Scores for exceeding the 1st, 2nd, …, 9th climatological decile.

Probabilistic scores were computed for the ensemble (EPS: perturbed forecasts and control with equal weights), control forecast (CF), Ensemble Mean (EM), IFS high-resolution forecast (HR) and the climatogical forecast (CLIM). Skill scores were computed on monthly and seasonal basis using the skill

of the long-term climatological forecast as reference.

Table J.1: *List of score statistics.*

Score	Name
ACC	Anomaly Correlation Coefficient
RMSE	Root Mean Square Error
BS	Brier Score
BSS	Brier Skill Score
IS	Ignorance Score
ISS	Ignorance Skill Score
ROCA	ROC Area
RPS	Rank Probability Score
RPSS	Rank Probability Skill Score
CRPS	Continuous RPS
CRPSS	Continuous RPS (Skill)
TRH	Talagrand Rank Histogram

Besides the standard deviation (spread) of the EPS, the RMSE and ACC of the CF, EM and HR were estimated. Talagrand Rank Histograms (TRH diagrams) statistics were also computed. A subset of the vast range of verification statistics is presented to indicate the overall performance. For selected figures shown below, emphasis has given to the performance of IFS & EPS over Europe (high-res mode) at EPS model 59 (110 meters) for 2008-09.

- *ACC & Useful Forecast Interval (UFI)*

The Anomaly Correlation Coefficient (ACC) is defined as the correlation between the forecast and analysed deviations from climate. ACC values (deterministic mode) for CF, EM & HR for Australia (solid) and India (dashed lines) over two-year period (2008-09) at 10-meter height (low-res mode) are shown in Figure J.3.

Figure J.3: *ACC values for Australia and India (at 10 meters).*

Interestingly, EM for Australia appears to be more skilful than HR from as early as 28 hours, while EM's skill for India becomes comparable to HR's at about 84 hours. Investigating scores as ACC, one can look at the forecast lead-time at which the score drops below a certain useful threshold. In that way the so-called Useful Forecast Interval (UFI) is defined. For deterministic forecasts, it is a tradition to consider the lead-time at which the anomaly correlation reaches 0.6 (value). Values of UFI for different areas over globe (10-meter winds) are presented in Table J.2. Entries equal to max correspond to 5 days. Inter-comparisons among different values of UFI for CF, EM & HR (Table J.2) reveal that EM keeps its usefulness for longer forecast horizons (shaded values).

Table J.2: *Values of Useful Forecast Interval (UFI) for Control Forecast (CF), Ensemble Mean (EM) and High-Resolution IFS (HR).*

#	Area G10	CF	EM	HR
01	n.hem (ex-trop)	4.55	max	4.75
02	s.hem (ex-trop)	4.80	max	4.90
03	n.hem (mid-lat)	4.25	max	4.45
04	n.hem (sub-trop)	max	max	max
05	g.europe	4.25	max	4.50
06	mediterranean	4.50	max	4.90
07	asia	4.10	4.95	4.40
08	australia	max	max	max
09	india	3.85	4.90	4.80
10	korea	2.90	3.70	3.35

Furthermore, Table J.3 was constructed in a similar way to Table J.2, but for selected subareas of Europe (high-res mode) for surface and 110-meter winds.

Table J.3: *UFI & MOI values, where G10: global low-res mode (10-meter winds) – H10: European high-res mode (10-meter winds) – H110: European high-res mode (110-meter winds).*

		CF (UFI)			EM (UFI)			HR (UFI)			EM > HR (MOI)		
#	AREA	G10	H10	H110	G10	H10	H110	G10	H10	H110	G10	H10	H110
01	europe	4.37	4.41	4.35	max	max	max	4.64	4.77	4.63	2.60	2.97	2.38
02	n.europe	4.10	4.09	3.97	4.89	4.39	4.72	4.27	4.30	4.16	1.54	2.05	1.39
03	s.europe	4.24	4.16	4.09	max	max	4.81	4.58	4.66	4.51	2.94	3.72	3.31
04	denmark	3.17	2.94	2.74	3.38	3.32	3.10	3.31	3.36	2.94	3.09	3.41	1.80
05	france	3.87	4.08	3.96	4.77	max	4.76	4.15	4.50	4.33	1.80	3.26	2.53
06	ireland	2.36	3.21	2.96	3.06	4.20	3.44	2.44	3.39	3.08	1.18	2.36	1.38
07	spain	4.12	3.80	3.82	4.83	4.50	4.47	4.40	4.48	4.37	3.33	4.44	4.12
08	england	3.40	3.35	3.23	4.05	4.10	3.74	3.62	3.63	3.40	2.10	3.29	1.50
09	germany	3.66	3.62	3.62	4.15	4.14	3.99	3.80	3.88	3.73	1.48	2.59	1.95
10	fino1	2.68	2.56	2.39	3.11	3.01	2.72	2.92	3.05	2.45	0.77	1.43	1.19

In Table J.3, the first three main categories (CF - EM - HR) correspond to UFI values; while the last category (EM > HR) represents the minimum time interval that EM becomes better than HR, noted as MOI (Minimum Overtaking Interval). For forecast horizons longer than MOI, EM provides better guidance than HR. Gray-shaded values represent local maximum values of UFI for different subareas. These values that reflect to higher values of UFI, seem to pile under EM category, revealing that EM is able to provide useful guidance for longer time intervals. In addition, it seems that EM is more skilful at the lowest model level (G10 & H10 categories). This proved to be the case for CF and HR also, suggesting that ECMWF seems to forecast slightly better at its lowest model level (deterministic mode).

On the other hand, yellow-shaded values represent local minimum values of MOI. The fact that most of minimum MOI values fall in G10 category, suggests that EM can cope slightly better at the lowest model level being in low-res mode, but exceptions do exist (i.e., yellow-shaded values in the H110 category). Furthermore, MOI values vary for different areas, model levels and resolution modes. For instance, the mean value of MOI for Europe is 2.65 days (averaged value over G10, H10 and H110 categories), compared to a considerable lower value (1.66) for the subarea of northern Europe. This means that EM represents a better option over HR for northern Europe if forecast horizon becomes longer than 1.66 days (about 40 hours).

Figure J.4: *Seasonal ACC values for EM-HR difference over Europe at 110 meters (high-res mode).*

For relatively small areas (such as FINO1 area), MOI values as low as 0.77 can be found, suggesting that EM is a better option than HR from as early as 18.5 hours. Furthermore, the corresponding UFI values for EM and HR found to be relatively limited (short), i.e., 3.11 & 2.92 respectively. It is worth to note that in the G10 low-res category, only four (4) grid points fall into the FINO1 platform area. Seasonal ACC scores vary in both the time and forecast horizon in such way that they form a rather complex surface. Such a surface is shown in Figure J.4, where seasonal ACC values of the EM-HR difference are plotted for Europe (at model level 59).

The superiority of EM over HR during EM-R is obvious since the EM-HR surface crosses the zero value plane at about the end of S-R interval. Negative ACC values belonging to the S-R interval reflect to HR's superiority, while positive values belonging to the EM-R reveal that EM provides a better-

forecast guidance.

- *Root Mean Square Error*

The Root Mean Square Error (RMSE) score provides a measure of the magnitude of errors (range: 0 to infinity, perfect score: 0). RMSE measures "average" error, weighted according to the square of the error. It does not indicate the direction of the deviations. The RMSE puts greater influence on larger than smaller errors, which may be a good thing if large errors are especially undesirable. RMSE values for European subareas are shown in Figure J.5. Results are for EM-HR difference formulation for ten different areas (same as in Table 3). Interestingly, a considerable number of EM-HR lines change sign early enough (between 24 and 36 hours) suggesting that EM becomes superior than HR in less than 1.5 days.

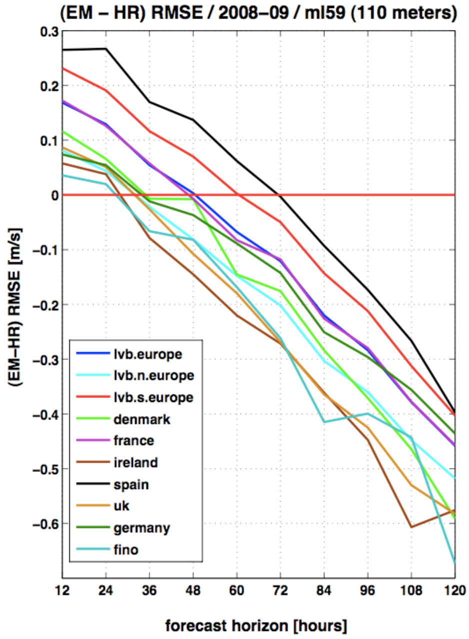


Figure J.5: *RMSE values of EM-HR difference for European subareas.*

- *Brier Score & Brier Skill Score*

In probabilistic mode, the Brier Score (BS) represents the most common verification method for probabilistic forecasts. The BS is similar to the RMSE, measuring the difference between a forecast probability of an event (p) and its occurrence (o), expressed as 0 or 1 depending on if the event has occurred or not. The lower the Brier score the "better".

Figure J.6: *BSS values for Europe (at EPS model level 59).*

The Brier Skill Score (BSS) is conventionally defined as the relative probability score compared with the probability score of a reference forecast. For this study, BSS was considered as a measure of the improvement of the probabilistic forecast relative to the long-term climatology. In Figure J.6, BSS values for different Probability Density Functions (PDFs) based on EPS, CF, EM and HR are shown for Europe for 2008-09 (EPS model level 59).

The threshold for defining bin categories was set to the 90% climatological percentile value. From Figure J.6, it is obvious that the BSS for EPS is more skilful than the BSS for CF, EM and HR. This was true for the rest of verification areas. Furthermore, the superiority of HR over CF is evident. Comparing EM to HR, the outcome in almost all cases is that HR is superior to EM for the S-R. On the other hand, EM provides better guidance than HR in EM-R. The mean MOI value is estimated at about 2.5 days.

- *Ignorance Score & Ignorance Skill Score*

The Logarithmic or Ignorance Score (IS) is "strictly proper", which means that if a forecaster believes the probabilities of each outcome occurring are p_i (i=1,...n) then that forecaster will minimize their expected logarithmic score by issuing a forecast $f_i = p_i$. The BS is also strictly proper, but unlike BS, the logarithmic score is local since it depends only upon the probability assigned to the outcome, which occurs, and not to any of the probabilities assigned to the other outcomes.

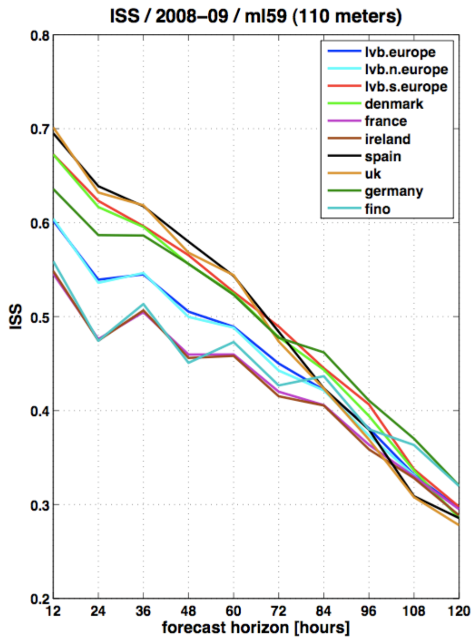


Figure J.7: *ISS values for all European subareas (at EPS model level 59).*

IS (defined with a negative sign) cannot be negative and smaller values of the score are better. In the ISS formulation (used here), higher values reflect to higher skill scores. The threshold of 90% climatological percentile value was utilised as in the BSS case. EPS ISS values for different subareas are shown in Figure J.7. It is evident that EPS shows considerable skill over all verification areas.

- *ROCA Scores*

ROCA stands for the Area Under the ROC Curve. ROC (Relative Operating Characteristic) measures the ability of the forecast to discriminate between two alternative outcomes, thus measuring resolution. It is not sensitive to bias in the forecast, so it reveals nothing about reliability. A biased forecast may still have good resolution and produce a good ROC curve, which means that it may be possible to improve the forecast through calibration. The ROC can thus be considered as a measure of potential usefulness. Furthermore, ROC is conditioned on the observations (i.e., given that Y occurred, what was the corresponding forecast X?)

Figure J.8: *Area Under the ROC Curve for Europe (at EPS model level 59).*

The Area Under the ROC Curve is frequently used as a score. Range: 0 to 1 (perfect score), while values below 0.5 indicate no skill. In Figure J.8, ROCA values for EPS, CF, EM and HR are plotted for Europe (at 110-meter height). ROCA for EPS showed the best characteristics, indicating considerable skill from 12 hours to the max forecast horizon (5 days). ROCA for HR and CF showed similar characteristics but not so skilful as EPS's, while ROCA for EM scored the poorest results.

- *Ranked Probability Score & Ranked Probability Skill Score*

The Ranked Probability Score (RPS) measures how well the probability forecast is predicting the category that the observation falls into. Ranked Probability Skill Score (RPSS) represents a measure of the improvement of the multi-category probabilistic forecast relative to a reference forecast such as the long-term climatology.

A selection of 10 climatologically equally likely categories (as analyzed in the Verification Statistics Section) were used to estimate RPS and corresponding RPSS scores. In Figure J.9 (left), RPSS values are shown for Europe (at 110 meters). It becomes obvious that EPS is superior to CF, EM and HR. HR found superior to CF, while EM keeps its superiority over CF for both the S-R and EM-R interval. IFS HR found superior to EM during S-R, but not during EM-R, where EM provides better-forecast guidance than HR.

Figure J.9: *RPSS values for Europe at EPS model level 59 (left) and at level 62 (right).*

In the same Figure J.9 (right), RPSS values are shown for Europe at 10-meter height. As for the previous 110-meter case, RPSS values for EPS are better than corresponding values for CF, EM and HR. For S-R, HR represents the best option over CF and EM, but for EM-R, EM is better than both the CF and HR. It is interesting to note that scores in Figure J.9 (right) show a clear dependency on the analysis cycle with poorer scores corresponding to 12UTC reflecting to possible difficulties of the model to simulate the daily cycle in a realistic way. Furthermore, it seems that ECMWF model can forecast slightly better (probabilistic mode) at 110 meters (compared to 10-meter forecasts).

- *Continuous Ranked Probability Score & Continuous Ranked Probability Skill Score*

The RPS (presented in the previous Section) utilizes a scoring rule based on the quadratic scoring approach, that seeks to reward forecasters for assigning probability, not only to the verified event, but also to near the verified event. RPS's continuous formulation, i.e., the Continuous Ranked Probability Score (CRPS) corresponds to the integral of the Brier Score (BS) for the associated binary probabilistic forecasts at all real value thresholds.

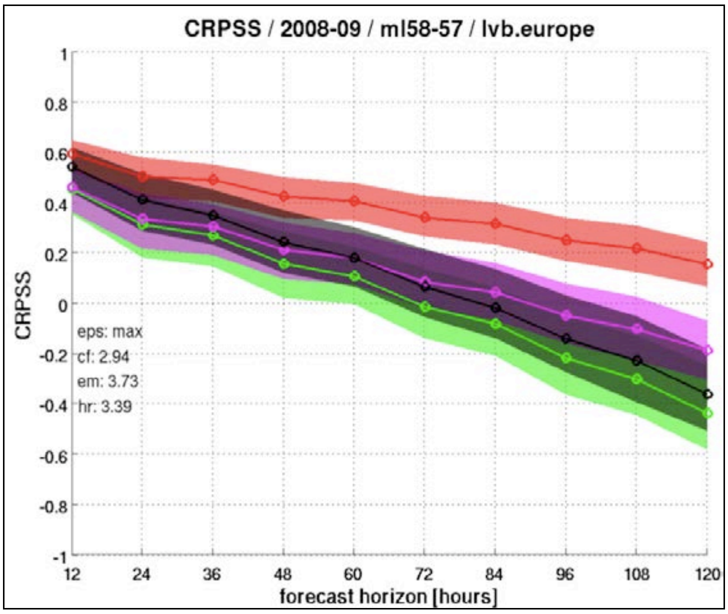


Figure J.10: *CRPSS values for Europe at EPS model level 59 (110 meters).*

In Figure J.10, CRPSS values are plotted for EPS, CF, EM and HR for Europe (at 110 meters). In the same plot, seasonal maximum and minimum CRPSS values are also plotted defining different colour-shaded areas for IFS & EPS components. The superiority of the red-shaded EPS is obvious (being on top), reflecting to higher CRPSS values. EM (in magenta) represents the second best option, while HR (in black) and CF (in green) show poorer performance.

- *Talagrand Rank Histograms*

Due to the limited number of EPS members, the verifying analysis may lay outside the ensemble range. For the current ECMWF EPS with 50 members this will normally happen 2/51 (~4%) of the time. If the percentage is higher, this means that the EPS does not spread out sufficiently. A usual way of analyzing the EPS spread is to construct a so-called TRH (Talagrand Rank Histogram).

In an ideal EPS system the long term "Talagrande distribution" [J.5] should be flat with equally many verifications in each interval. 3-D TRH graphs are utilized here to estimate the number of analysis values that lay outside EPS "cloud" for different areas and model levels. In Figure J.11, an example of TRH is shown for Europe (at ml59). The exact number of outliers is studied by plotting the sum of the left and right bin outliers. Such a graph is embedded in Figure J.11, where the total number of outliers for both the surface level (ml62) and 110-meter (ml59) is shown. It seems that EPS can produce a more sufficient spread for the 110-meter level, since its corresponding number of outliers is smaller.

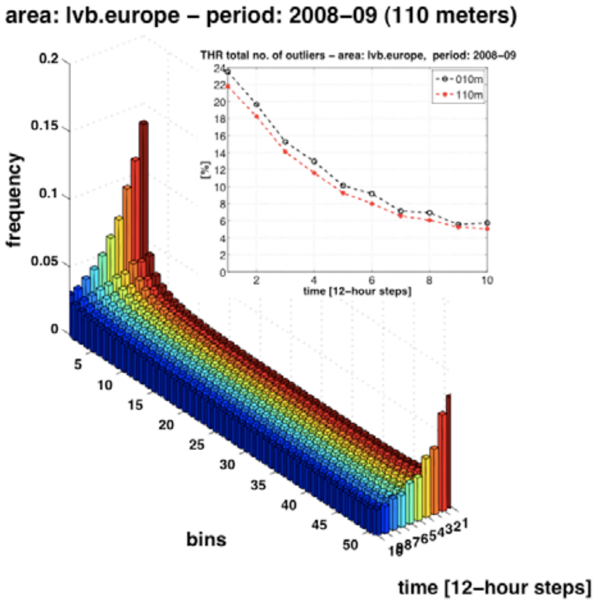


Figure J.11: *TRH (ml59) & number of outliers (for ml59 & ml62).*

Overall, the number of outliers typically gets its minimum value at the end of the forecast horizon reaching its optimum value (~4%). Interestingly, for relatively small areas (such as FINO1 subarea), the optimum value can be reached as soon as 2.5 days.

- *Comparison of EPS Spread from Ensemble Mean and Skill of Ensemble Mean*

Concerning EPS, one basic quantity of interest is the second moment of the ensemble – the spread. In a perfect EPS, the time-mean ensemble spread about the ensemble-mean equals the time-mean RMS error of the ensemble-mean [J.6]. The relatively large number of outliers found in TRH diagrams in the S-R was further investigated by plotting the difference between mean EM's error from mean EPS spread values. In Figure J.12 such a relationship is shown for selected subareas (at ml59).

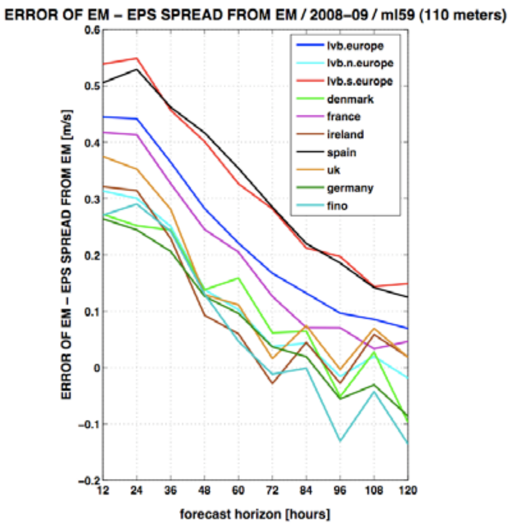


Figure J.12: *Differences of mean EM's error minus mean EM's spread for selected subareas (m/s).*

From Figure J.12, it is evident that all European subareas are characterized from relatively small spread values especially during S-R. On the other hand, considerable variations can be seen especially over relatively smaller areas. Areas such as Ireland or FINO1 reach zero value (i.e., perfect balance between EPS spread and EM's error) at about 72 hours. For Europe, such a balance seems not possible, since difference values remain greater than zero during both the S-R and EM-R intervals.

2. Details of the DJ & FM 2010 verification (model configuration after 26 January 2010)

On 26 January 2010 ECMWF upgraded the horizontal resolution of the deterministic forecasting system and Ensemble Prediction System (EPS) that form components of the Integrated Forecasting System (IFS) [J.7]:

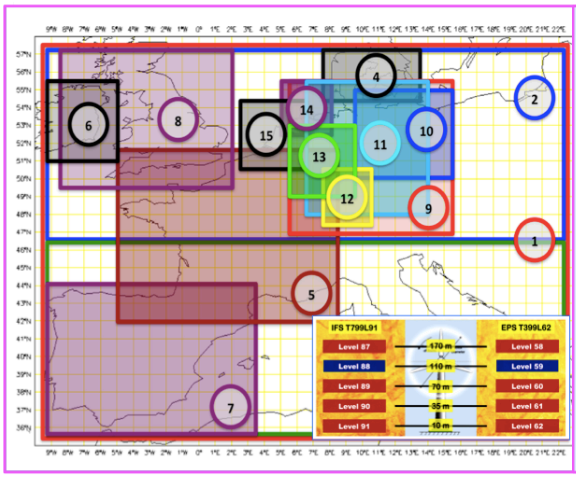
• for the deterministic forecast and analysis the horizontal resolution increased to T1279 (~16 km grid spacing) from T799 (~25 km)

• the resolution of the inner loops of the 4D-Var analysis changed to T159/T255/T255 (~125 km/78 km/78 km) from T95/T159/T255 (~210 km/125 km/78 km)

• the EPS resolution increased to T639 (~32 km) from T399 (~50 km) for leg1 (the first 10 days of the forecast) and to T319 (~65 km) from T255 (~80 km) for leg2 (day 9 to day 15 and day 32 for the monthly forecast)

• the coupled ocean wave model resolution was upgraded to 0.25° (~28 km) from 0.36° (~40 km) in the deterministic system, and to 0.5° (~56 km) from 1° (~111 km) in the EPS.

Furthermore, the vertical resolution remained unchanged at 91 levels for the deterministic system and at 62 levels for the EPS. The various resolution increases have been implemented as IFS Cycle 36r1. A correction to the interaction of short-wave radiation with clouds is also included in this cycle.



01.	europe :	57.5 / -9.5 / 35.5 / 22.5	11,481 grid points
02.	n.europe :	57.5 / -9.5 / 46.5 / 22.5	5,805 grid points
03.	s.europe :	46.5 / -9.5 / 35.5 / 22.5	5,805 grid points
04.	denmark :	57.5 / 7.5 / 54.5 / 13.5	325 grid points
05.	france :	51.5 / -5.0 / 42.0 / 8.5	2,145 grid points
06.	ireland :	55.5 / -9.5 / 51.0 / -5.0	361 grid points
07.	spain :	44.0 / -9.0 / 36.0 / 3.0	1,617 grid points
08.	uk :	57.5 / -8.5 / 49.5 / 2.0	1,419 grid points
09.	germany :	55.5 / 5.5 / 47.0 / 15.5	1,435 grid points
10.	vattenfall :	55.0 / 9.5 / 50.0 / 15.5	525 grid points
11.	eon :	55.5 / 6.5 / 47.0 / 14.0	1,085 grid points
12.	enbw :	50.0 / 7.5 / 47.5 / 10.5	143 grid points
13.	rwe :	53.0 / 5.5 / 49.0 / 9.5	289 grid points
14.	fino1 :	55.5 / 5.0 / 52.5 / 8.0	169 grid points
15.	cabauw :	54.5 / 2.5 / 50.5 / 6.5	289 grid points

1 Dec 2009 to 25 Jan 2010:
DJ period (56 days)

Old IFS (25 km) & EPS (50 km)
New COSMO (7 km)

26 Jan to 22 Mar 2010:
FM period (56 days)

New IFS (16 km) & EPS (32 km)
New COSMO (7 km)

Figure J.13: *Definition of various European subareas used for verification over DJ and FM periods.*

Extensive experimentation for along list of model parameters accompanies a major operational change such as those described above. Nevertheless, verification results that follow focus mainly on the assessment of wind forecasts (at 10 meters) over two distinct periods of 56 days each (e.g., DJ & FM) for the extended winter period of DJFM 2010 (e.g., December 2009 – January – February – March 2010). Obviously DJ period refers to the old (before 26 January 2010) model configuration, while FM refers to the new one (details shown in Figure J.13).

- *ACC (IFS & EPS) scores for December-January (DJ) versus February-March (FM) 2010*

Concerning the inter-comparison of IFS & EPS skill during DJ period (56 days with old model configuration) versus FM period (56 days with new model configuration), COSMO-LEPS has been used to justify that no significant trend in skill has taken place during the extended DJFM 2010 winter season. COSMO-LEPS configuration during DJFM 2010 has remained constant. From Figure J.14, it becomes obvious from the skill of Ensemble Mean and its corresponding spread (from the ensemble mean) that no significant trend has taken place during DJFM 2010. It is worth to point out that skill assessment uses its own COSMO operational analysis.

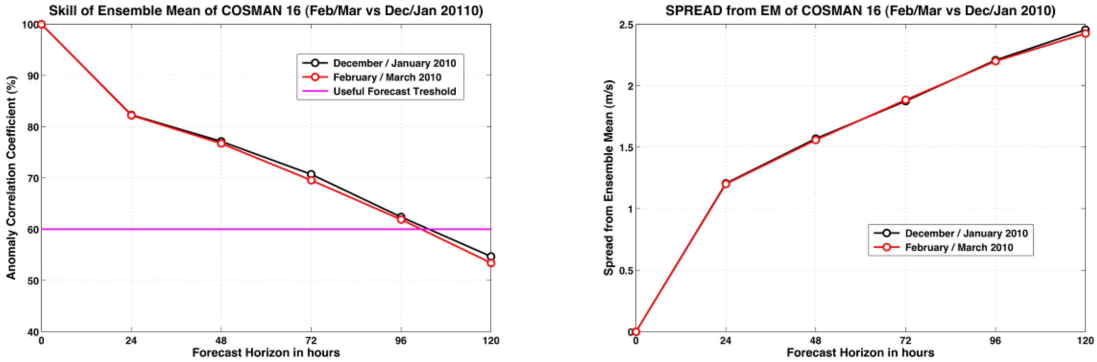


Figure J.14: *COSMO-LEPS Ensemble Mean ACC values (left) and its corresponding Spread in RMSE terms (right) valid for DJ (black) & FM 2010 (red line) over Europe.*

Concerning the impact of increased resolution to IFS over Europe overall (Figure J.15 / left) results are not positive, since a slight deterioration of skill is recorded for almost all forecast horizons. Nevertheless, what seems important is the positive (local) impact of increased resolution to IFS in various European subareas, such as France, UK or EnBW (German energy subarea) concerning mainly S-R horizon as clearly shown in Figure J.15 (right) and Figure J.16.

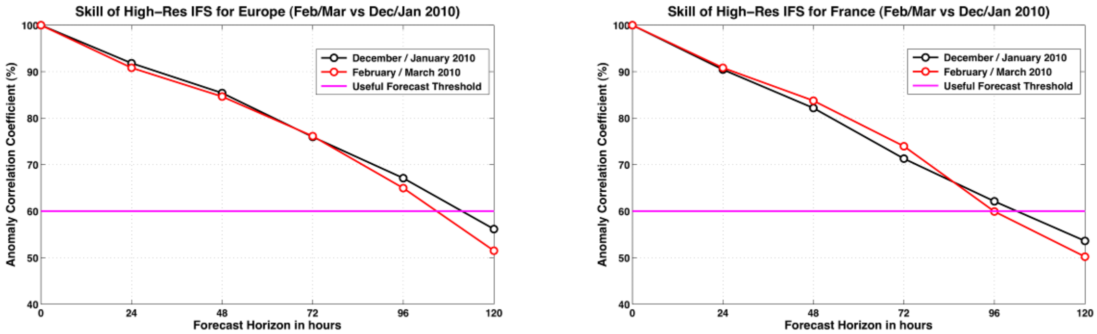


Figure J.15: *IFS ACCs for Europe (left) and France (right) valid for DJ (black) & FM 2010 (red line).*

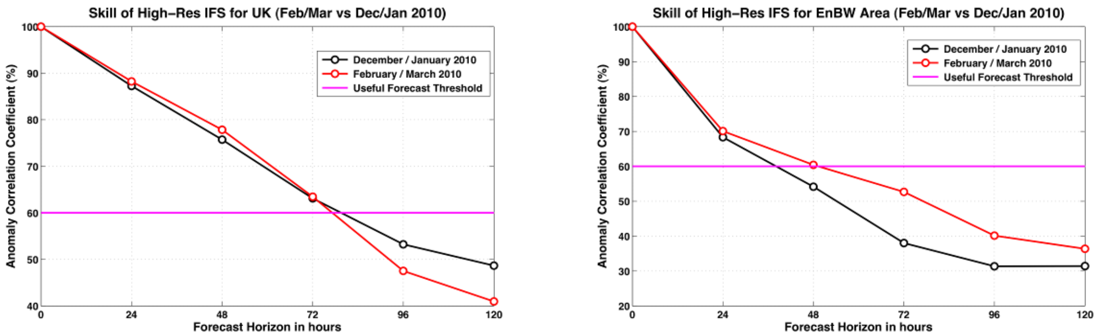


Figure J.16: *IFS ACC for UK (left) and EnBW subarea (right) valid for DJ (black) & FM 2010 (red line).*

Concerning the impact of increased resolution to EPS over Europe overall results are both positive, and promising, since an improvement of skill is recorded for EPS in almost all forecast horizons. More specifically, an improvement is obvious for both the skill of Control Forecast (Figure J.17 / left) and Ensemble Mean (Figure J.17 / right) over Europe mainly for S-R horizons.

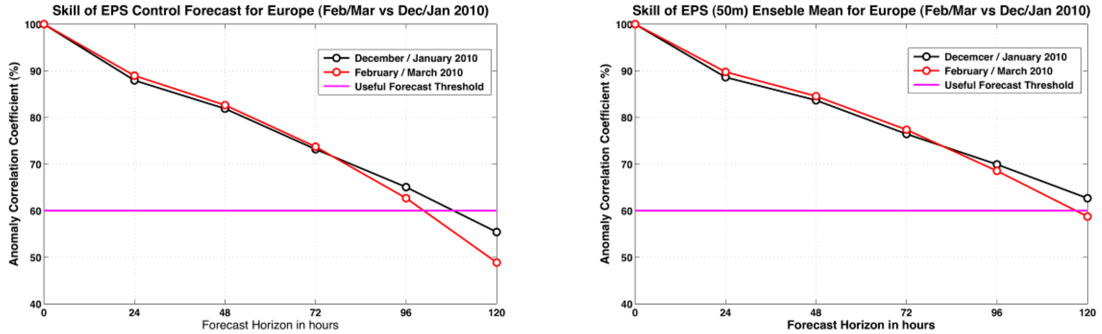


Figure J.17: *EPS Control (left) and EM (right) ACCs for DJ (black) & FM 2010 (red line) over Europe.*

- *CRPSS scores for DJ versus FM 2010*

As mentioned already, the Continuous Ranked Probability Score (CRPS) corresponds to the integral of the Brier Score (BS) for the associated binary probabilistic forecasts at all real value thresholds. CPRSS values shown in Figure J.18 (left) reflect to a significant improvement of EPS skill not only over Europe overall but for many other subareas as shown in Figure J.18 (right) for Spain. EPS seems to get a clear benefit from the new (increased) resolution.

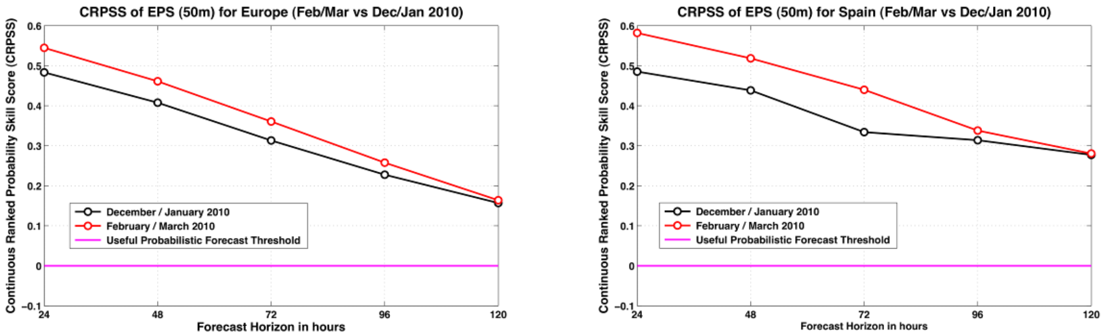


Figure J.18: *EPS CRPSSs for Europe (left) and Spain (right) valid for DJ (black) & FM 2010 (red line).*

- *Comparison of EPS Spread from Ensemble Mean and Skill of Ensemble Mean*

An improvement of the Spread-Skill relationship is obvious not only for Europe as shown in Figure J.19 (left) but for many other subareas as France, Spain & UK (Figure J.19 / right & Figure J.20).

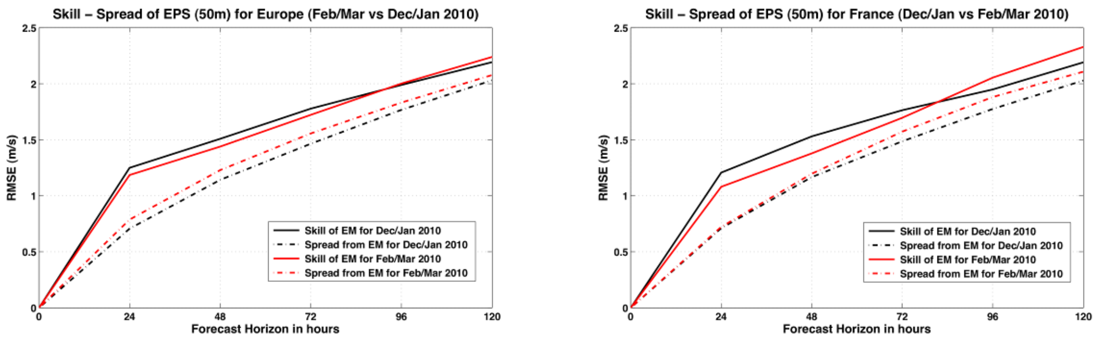


Figure J.19: *Skill-Spread for Europe (left) & France (right) valid for DJ (black) & FM 2010 (red lines).*

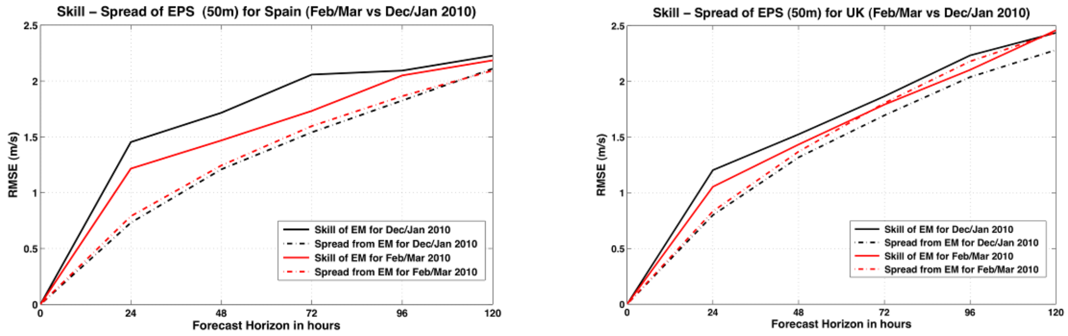


Figure J.20: *Skill-Spread for Spain (left) and UK (right) valid for DJ (black) & FM 2010 (red lines).*

- *Talagrand Rank Histograms (TRHs)*

Although an improvement in the Spread-Skill relationship is obvious (as clearly seen in Figure J.19 & Figure J.20) there still exist some (although smaller) "gap" between the EPS spread and the skill of the Ensemble Mean. A closer investigation focusing on TRHs show a significant reduction of outliers for both sides at almost all forecast ranges as for T+024 (Figure J.21) and for T+72 (Figure J.22).

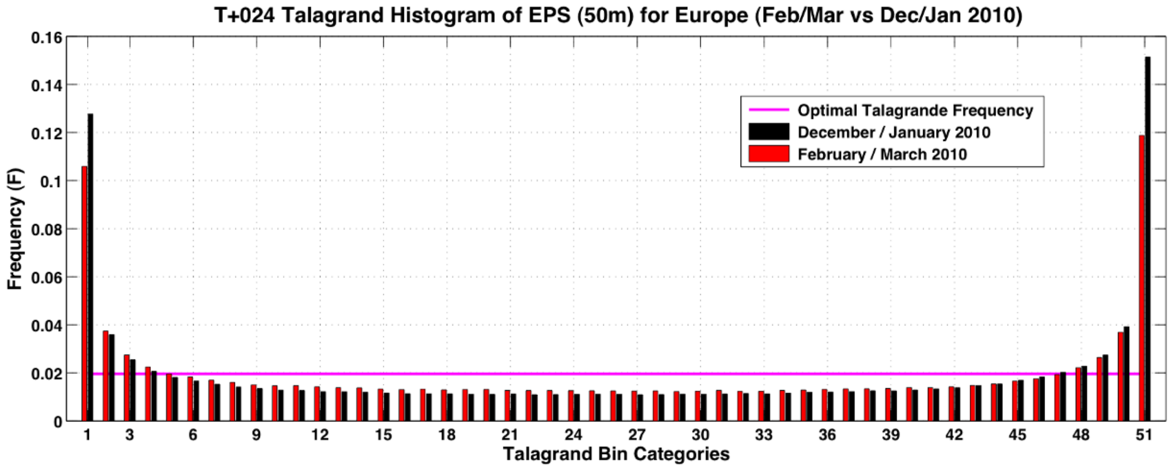


Figure J.21: *T+024 Talagrand Rank Histogram of EPS (50 members) for DJ (black) & FM 2010 (red boxes) over Europe.*

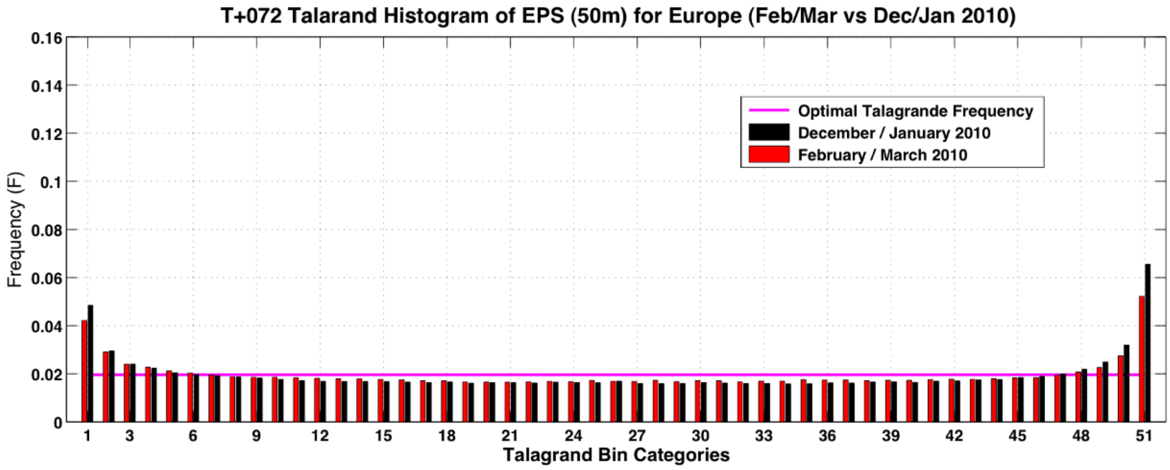


Figure J.22: *T+072 Talagrand Rank Histogram of EPS (50 members) for DJ (black) & FM 2010 (red boxes) over Europe.*

The improvement in the reduction of outliers for T+024 (Figure J.21) and for T+72 (Figure J.22) has shown that recent changes in resolution were in the right direction concerning the S-R and early M-R forecast horizons, besides the fact that this reduction is not clear for T+120 horizon (Figure J.23). For T+120 horizon, the reduction of outliers is clear only for the left side of the TRH diagram.

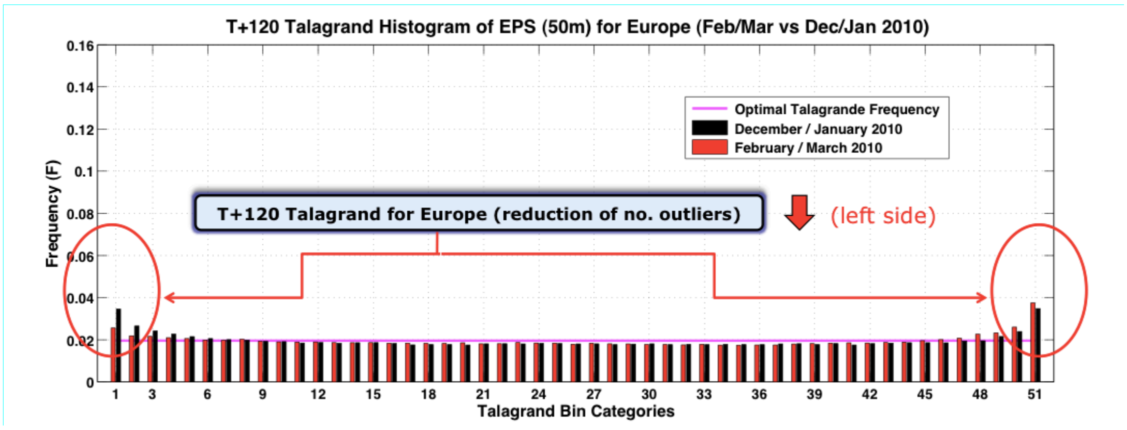


Figure J.23: *T+120 Talagrand Rank Histogram of EPS (50 members) for DJ (black) & FM 2010 (red boxes) over Europe.*

The one side reduction of outliers to the left at T+120 should be investigated more, since the right side of the diagram is the most important since this side reflects to extreme values (linked to extreme events). Nevertheless the total number of outliers (when left and right side outliers summed together) found to be smaller during FM period (new model configuration) compared to DJ period (old model configuration) as clearly shown in Figure J.24.

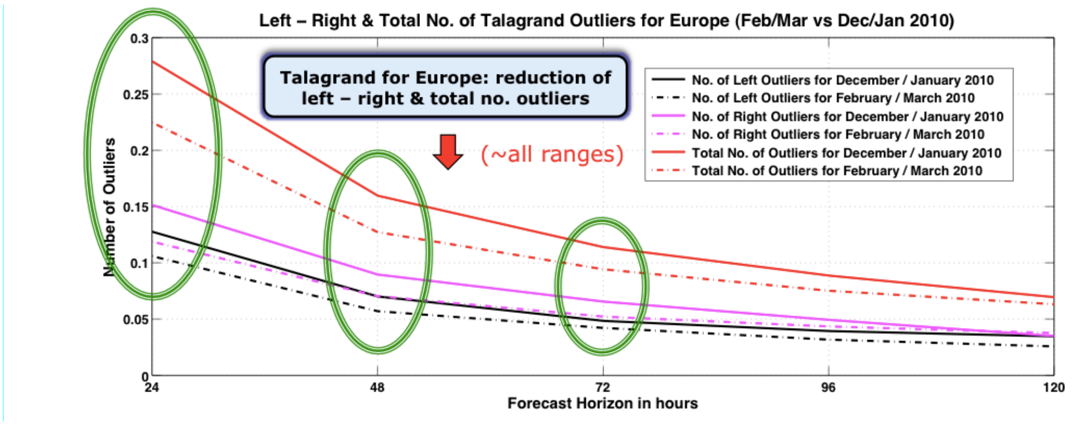


Figure J.24: *Talagrand Rank Histogram left, right and total number of EPS outliers for DJ (solid) & FM 2010 (dashed lines) verification periods over Europe.*

References

[J.1] Palmer T.N., R. Buizza, M. Leutbecher, R. Hagedorn, T. Jung, M. Rodwell, F. Vitart, J. Berner, E. Hagel, A. Lawrence, F. Pappenberger, Y-Y. Park, L von Bremen and I. Gilmour, 2007: The Ensemble Prediction System - Recent and Ongoing Developments. TM 540, ECMWF, U.K.

[J.2] Uppala S.M., Kallberg P.W., Simmons A.J., Andrae U., da Costa Bechtold V., Fiorino M., Gibson J.K., Haseler J., Hernandez A., Kelly G.A., Li X., Onogi K., Saarinen S., Sokka N., Allan R.P., Andersson E., Arpe K., Balmaseda M.A., Beljaars A.C.M., van de Berg L., Bidlot J., Bormann N., Caires S., Chevallier F., Dethof A., Dragosavac M., Fisher M., Fuentes M., Hagemann S., Holm E., Hoskins B.J., Isaksen L., Janssen P.A.E.M., Jenne R., McNally A.P., Mahfouf J.-F., Morcrette J.-J., Rayner N.A., Saunders R.W., Simon P., Sterl A., Trenberth K.E., Untch A., Vasiljevic D., Viterbo P. and J. Woollen, 2005: The ERA-40 re-analysis. Q.J.R.Meteorol. Soc., 131, 2961-3012.

[J.3] Swets J. and R.M. Pickett, 1982: Evaluation of Diagnostic Systems: Methods from Signal Detection Theory. NY: Academic Press.

[J.4] Candile G. and O. Talagrand, 2005: Evaluation of probabilistic prediction systems for a scalar variable. Q.J.R. Meteor. Soc., 131, 2131-2150.

[J.5] Persson A. and F. Grazzini, 2007: User guide to ECMWF forecast products, available at http://www.ecmwf.int/products/forecasts/guide/index.html

[J.6] Palmer T.N, R. Buizza, R. Hagedorn, A. Lawrence, M. Leutbecher and L. Smith, 2006: Ensemble prediction: a pedagogical perspective. Newsletter 106, ECMWF, U.K.

[J.7] Miller M., Buizza R., Haseler J., Hortal M., Janssen P. and A. Untch, 2010: Increased resolution in the ECMWF deterministic and ensemble prediction systems. Newsletter 124, ECMWF, U.K.