

Recommendations on trajectory selection in flight planning based on weather uncertainty

Jacob Cheung*, Alan Hally†, Jaap Heijstek‡, Adri Marsman‡ and Jean-Louis Brenguier†

*Met Office, Exeter, United Kingdom

Email: jacob.cheung@metoffice.gov.uk

†CNRM-GAME (Météo-France, CNRS), Toulouse, France

‡Netherlands Aerospace Center, The Netherlands

Abstract — The chaotic nature of the atmosphere combined with limitations in modelling and an insufficient number of observations means that inaccuracies continue to exist even in the most state-of-the-art Numerical Weather Prediction (NWP) systems. In the world of aviation, Trajectory Prediction (TP) is currently mostly based on deterministic meteorological forecasts and thus does not take into account the probabilistic information available from an Ensemble Prediction System (EPS). One of the main aims of the IMET project¹ is to quantify the predictability of flight planning systems by exploring the impact on TP output of ensemble weather forecast (EWF) generated by the EPS. In this paper, we use Probabilistic TP (PTP) defined by running a TP system n times with n being the number of members in the EWF. This allows an ensemble of trajectories to be created, which provides uncertainty information on flight parameters such as flight duration, and trip fuel cost. The information can be used to support decision making regarding the predicted trajectory. We demonstrate that the three state-of-the-art EPSs used within the IMET project are all capable of capturing relevant weather events observed from a large data sample of AMDAR² measurements, thirty-six hours in advance of take-off.

Index Terms — Trajectory Prediction, Uncertainty, Flight Planning, Ensemble Weather Forecast, Ensemble Prediction System

I. INTRODUCTION

Despite the advancement of forecasting techniques in recent decades, meteorological (Met) forecasts are not perfect and uncertainties remain even in state-of-the-art Numerical Weather Prediction (NWP) systems. Other than limits in modelling and observation techniques, the uncertainty can be attributed to the chaotic nature of weather, in which small errors in the initial state of a deterministic³ NWP model can grow rapidly with time thus yielding a prediction that is very different from the actual weather scenario. As of today, Trajectory Prediction (TP) is based mainly on single deterministic Met forecasts with Airline Operations (AO) preferences, and Air Traffic Management (ATM) constraints imposed. It follows that an inaccurate Met forecast can result in poor estimation of trajectory integrated parameters such as flight duration or trip fuel cost, and thus in a suboptimal

selection of flight paths. Indeed, using a deterministic weather forecast makes it impossible to estimate the uncertainties involved with any route selected; it is still standard practice to use past experiences. In recent years, Ensemble Prediction Systems (EPSs) have been developed to quantify the uncertainties in Met forecasts [1], and have been made operational in various weather centres, such as the European Centre for Medium-range Weather Forecast ECMWF, Met Office, Météo France, and NCEP, the US National Centers for Environmental Prediction. The basis of an EPS is to allow a NWP model to run repeatedly, each time with a different starting state, and/or different physical parametrizations, yielding an ensemble of forecasts. This is illustrated in Figure 1.

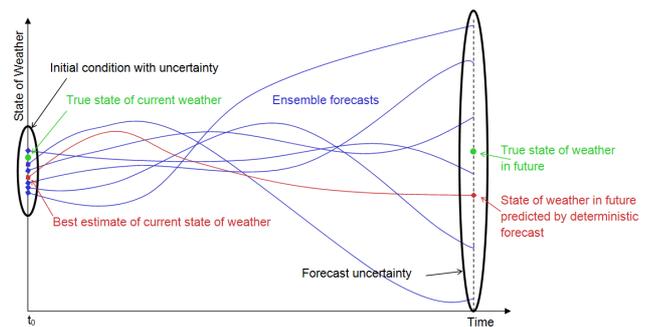


Figure 1 Schematic showing uncertainty captured in an Ensemble Weather Forecast.

The starting conditions of each ensemble member are carefully generated using observations and statistical methods to account for limitations in current modelling and observation techniques, aiming to capture the uncertainty involved in the forecast starting condition. The forecast uncertainty can be quantified by the spread of the end states of each member in the ensemble. In order to maximise the spread of an EWF, and therefore to cover an even greater proportion of possible weather futures, a multi-model “SUPER” ensemble can be constructed by combining the EWFs of different EPSs. Such a large number of ensemble members is more likely to capture Met outliers and give a higher degree of confidence in predicting future atmospheric evolution.

An EWF can be applied to existing deterministic TP systems to construct a Probabilistic TP (PTP) in which

¹ IMET: Investigation of the optimal approach for future trajectory prediction systems to use METeoroological uncertainty information.

² AMDAR: Aircraft Meteorological DATA Relay observing system [13].

³ Deterministic models give a specific forecast at a specific time and place with no representation of uncertainty.

uncertainty in the Met input is translated to uncertainty in the PTP output. This is illustrated in Figure 2.

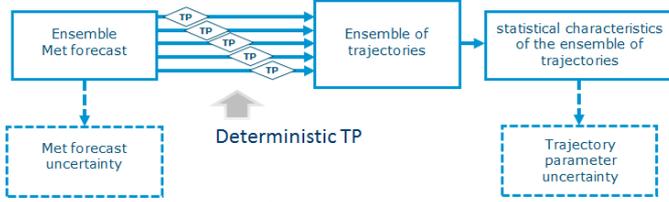


Figure 2 Schematic of Probabilistic Trajectory Prediction based on Ensemble Weather Forecast

More specifically, TP can be applied repeatedly for each member of the EWF, yielding an ensemble of trajectories including parameters such as flight times and routes for a given flight. From an ATM point of view, quantified information on the potential impact of Met uncertainty on TP output, and knock-on effects on TP based decision support tools, provides awareness with respect to the predictability of the TP/Decision Support Tool (DST) output, leading to an increased level of confidence in the predicted trajectory. Using this information, airlines may be able to reduce costs by safely carrying less fuel, or carrying more than originally planned to avoid unforeseen refuelling at an alternate airport. This is being investigated in detail in SESAR WP11.1.

The paper is structured as follows: Section 2 describes the method by which the trajectory ensemble is created; Section 3 deals with model uncertainty and verifying the models' ability to correctly represent real world patterns. Some results and conclusions are touched upon in Sections 4 and 5, respectively.

II. APPROACH

Existing deterministic TP systems can be used in conjunction with ensemble NWP models to quantify the uncertainty in flight planning due to weather. In this study, we consider a single deterministic trajectory T^0 , determined e.g. using a deterministic weather forecast. The question is what the predictability of an integrated parameter of this trajectory is with respect to the ensemble weather forecast. It can be estimated by first creating a trajectory ensemble using trajectory T^0 and each member of the EWF, and then calculating the spread of the trajectory ensemble integrated parameter. The spread is a measure for the uncertainty of the trajectory integrated parameter with respect to the EWF. If the spread is small, the parameter is relatively insensitive to the EWF, and hence highly predictable. If the spread is large, the parameter is highly sensitive to the EWF, and relatively unpredictable. In the latter case, one could seek alternatives which are more predictable.

Given any trajectory, the uncertainty due to weather from a user's perspective can be assessed by creating a Probability Density Function (PDF), which involves calculating the relevant parameter of interest (e.g. flight duration, fuel cost) along the specified route in each possible weather projection. The user is then able to select the trajectory which best suits

his requirement. For simplicity, only time/fuel costs are considered here.

In general, it is assumed that the end user provides the initial trajectory T^0 , and a set of potential alternative trajectories T^1, T^2, \dots, T^m . The set could be constructed in various ways, e.g. taking into account different preferred rerouting options. Each potential alternative trajectory T^j , $j=1, \dots, m$, is examined on predictability with respect to the EWF. To this end, for each j , the corresponding trajectory ensemble $T^{1,j}, T^{2,j}, \dots, T^{n,j}$, and the spread of the trajectory integrated parameter are calculated. This is illustrated in Figure 3.

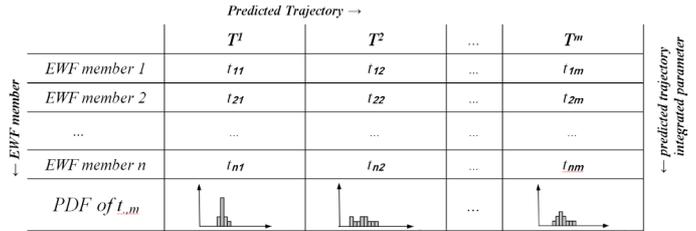


Figure 3 Example PDFs for trajectory integrated parameters.

In IMET, potential alternative trajectories are created by iteratively running the TP system supplied by NLR, reusing the input for the deterministic trajectory T^0 , except for the weather forecast, which is replaced with members of the EWF. Thus, in IMET, we have $m=n$.

For predictable weather conditions, it is likely that the ensemble trajectory ensemble members will be geospatially similar, which implies a small spread, thus simplifying decision making in ATM. In Figure 3 this is illustrated by trajectory T^1 . The column entries t_{kj} denote the time/fuel cost to fly along route T^j in weather scenario k ($k=1, \dots, n$). At the bottom of the corresponding column, an approximation of the PDF of the cost for T^j due to weather (keeping other TP input parameters fixed) is displayed. The end user could decide to use the fuel amount estimated using T^0 .

In contrast, if the weather conditions are unpredictable, which is often the case under severe weather conditions, and assuming TP is able to produce trajectories avoiding severe weather areas (e.g. [2]), geospatially dissimilar trajectories or bifurcations may be observed in trajectory ensembles. This topic is discussed at length in [3]. As each member of the Met ensemble, and hence each member in the trajectory ensemble, is equally probable, as is assumed by design, this presents difficulties in decision making on flight planning without further analysis. At the bottom of the second column in Figure 3, a PDF of the cost for trajectory T^2 due to weather is displayed. The relatively large spread makes it difficult to choose an appropriate amount of fuel. The end user might choose to take (at most) the average value on board, and accept the risk of unforeseen refuelling at an alternate airport, or to take more than the average value on board, to reduce the risk of unforeseen refuelling, or to look for an alternative trajectory leading to a smaller spread. Optimisation of weather forecast for TP has been addressed in detail in IMET. The

approach is also applicable to other ATM areas, e.g. Flow Management ([4], [5]).

III. MODEL DESCRIPTION AND VERIFICATION

In order to demonstrate that the EPSs can reliably reproduce real world wind patterns, before coupling to the TP system, the Met ensembles were to be verified against wind observations. Three different EPSs have been verified: Pr evision d'Ensemble Action de Recherche Petite Echelle Grande Echelle (PEARP), Met Office Global and Regional Ensemble Prediction System (MOGREPS) and the EPS of the ECMWF. A combination of these ensembles is then used to produce a SUPER ensemble, which has also been verified. Each of the ensembles and their verification is discussed hereafter. Two key measures of each ensemble have been analysed, namely the resolution (the ensemble's ability to discriminate between events) and the dispersion (the ensemble's ability to capture the envelope of possible weather scenarios).

A. PEARP

The French operational global ensemble forecasting model, PEARP [6] consists of 35 members (one control plus 34 perturbed members) run twice daily at 0600 UTC (+72 hour forecast range) and 1800 UTC (+108 hour forecast range). PEARP was conceived as a short-medium range (4-5 days) ensemble with a maximum horizontal resolution of 15.5km over France. The perturbed members of PEARP are constructed by using a combination of ensemble data assimilation and singular vectors for the initial conditions (IC) while model uncertainties are represented through a multi-physics approach. The output interval of the model is 6 hours.

B. MOGREPS

MOGREPS [7] has been the Met Office's operational EPS since 2008. MOGREPS consists of 12 members (one control + 11 perturbed) and is run at t=0000, 0600, 1200 and 1800 UTC daily. The IC of each ensemble member is generated using the ensemble transform Kalman filter as described in [8]. Unlike other EPSs (e.g. the one at ECMWF), MOGREPS is designed to represent Met uncertainty in the short range (days 1-2) rather than medium range (days 3-10), which coincides with the time frame in which Reference Business Trajectories (RBTs) are usually determined. The version of MOGREPS used in this study covers the whole of the globe and has a horizontal resolution of N 400 (33km at mid-latitudes) with 70 model levels in the vertical. The output interval of the model is 3 hours.

C. ECMWF

The Integrated Forecasting System (IFS) of the ECMWF is a 51 member ensemble (1 control + 50 perturbed members) has a forecast range of up to 15 days with forecasts being run twice daily at 0000 UTC and 1200 UTC, at a horizontal resolution of 32 km (16 km for the control) and with 91 layers (137 layers for the control) in the vertical. The ensemble has an output interval of 6 hours. The 50 different ensemble members are constructed using a combination of perturbations

upon the ICs (singular vector technique), an ensemble data assimilation approach and by stochastically perturbing the models physical parameterisations [9]. The dates considered in this study are from the 1st of January 2015 to the 31st January 2015 inclusive and applies to all the NWP models described.

D. SUPER

A multi-model ensemble system was constructed by combining the three previously introduced EPSs. The 35 members of PEARP, the 12 members of MOGREPS and the 51 members of the IFS were mixed together in order to form a 98 member EPS. This ensemble was initialised at 18UTC, just like the PEARP and MOGREP EPSs and 6 hours after the initialisation time of the ECMWF EPS. This meant that the SUPER ensemble had a forecast range of +42 hours. The output interval of the model was 6 hours as this was the interval common to all component models.

E. Model verification

A comparison of the inherent uncertainty that exists for each of the NWP models introduced above was carried out by comparing the capacity of each of the models to forecast the observed wind values at a fixed flight level of FL340 (corresponding to 250hPa). Wind observations from the observational Aircraft Meteorological Data Relay (AMDAR) database for a domain encompassing much of Western Europe, the North Atlantic and North America as far as the Midwest (75N-10N, 105W-15E) were used. Such a large geographical zone and relatively long time period meant that for each model validation time, thousands of AMDAR observations were available and thus a statistically robust calculation of the Met uncertainty could be undertaken.

The ensemble forecasts utilised covered a time window from the initialisation time (1800 UTC for PEARP, 1800 UTC for MOGREPS and 1200 UTC for ECMWF) to 48 hours ahead, thus falling well within the RBT requirement of having a reliable probabilistic forecast 36 hours in advance of take-off time. However, since only a forecast window of +42 hours is common to all ensembles, this is the window reported upon here within.

In order to determine the resolution of the different models, an ensemble skill score called a Relative Operating Characteristic (ROC) was used. A complete description of the score and its meaning can be found in [10]. The ROC score determines the ensemble's ability to correctly reproduce hits (when an observed event is forecast), correct misses (when an event is not observed and not forecast), false alarms (when an event is not observed but forecast) and misses (when an event is observed but not forecast) for a prescribed observational threshold. These four factors allow a contingency table to be produced and thus the Probability of Detection (POD) and False Alarm Rates (FAR) to be calculated. An example of a contingency table is given in Figure 4. The area underneath this curve (a score value between 0.5 and 1, with 1 being a perfect ensemble) of POD's versus FAR's is then used as a measure of the usefulness of an ensemble.

The models' spread was measured using a score called the Reduced Centred Random Variable (RCRV). This score compares the average of each ensemble to the observational data taking into account the observational and model error. A value of 1 for the dispersion of the RCRV indicates a perfect level of ensemble dispersion while a value greater than (less than) 1 indicates under-dispersion (over-dispersion).

An important factor to be taken into account in the RCRV's calculations is the observational error σ_0 . Following the example of the σ_0 value used in the ARPEGE Four-dimensional Variational Assimilation (4D-VAR) data assimilation process, the observational error for an AMDAR observation was taken to be equal to 2.3m/s . A more complete description of the score can be found in [11] and [6].

Along with the ROC and RCRV scores of each of the operational ensemble products introduced in the previous sections, the scores related to a multi-model ensemble (referred to as SUPER), comprised of a mix of the ECMWF, MOGREPS and PEARP ensembles, were also measured. This was done in order to underline the advantages of having as

Event Forecast	Event Observed	
	Yes	No
Yes	a	b
No	c	d

Figure 4 Example contingency table from which a ROC curve can be plotted.

many ensembles at ones' disposal as possible.

Figure 5 shows the area underneath the ROC curve for each of the three operational ensembles, along with the SUPER ensemble, at a horizontal resolution of 0.5 (56km) and for a lead time of +42 hours after initialisation time. The values plotted represent the median value of the score while the extremes of the error bars represent the 5th (lower bound) and 95th (upper bound) percentile values, after performing a statistical test known as a boot-strap test [11].

The scores were calculated for an observational wind threshold of 55 m/s (average observed for January 2015, +10%). Since in real-time it takes approximately 10 hours after initialisation to have access to all three ensembles (for example, a run of PEARP at 1800 UTC becomes available at around 0400 UTC the following day), only the scores from the +12hr lead time onwards are relevant for TP. The previous time steps are included for completeness, and to give an impression of the evolution of the ensemble scores in relation to lead time.

What Figure 5 illustrates is that for a lead time of +36 hours, one obtains scores in the range 0.82 to 0.93, indicating an excellent level of model resolution in all models. As the lead times increase, the area under the ROC curve score of the SUPER ensemble becomes the best amongst the four ensembles.

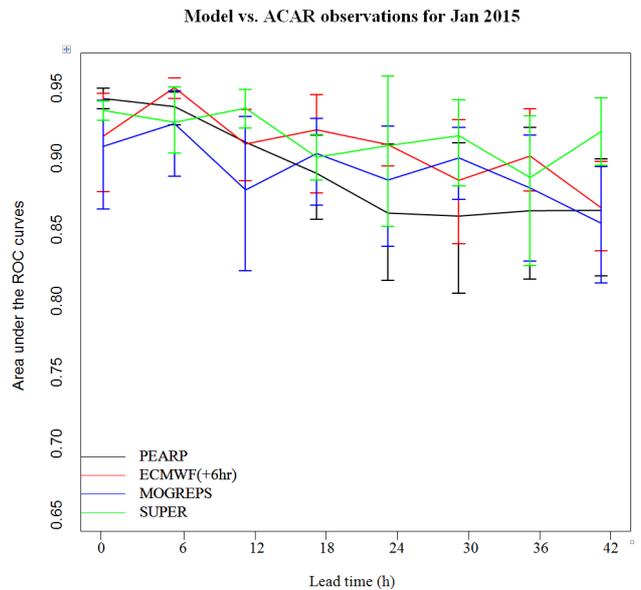


Figure 5 Area under the ROC curve for each operational ensemble and for the SUPER ensemble for a lead time of up to +42 hour after initialisation for the month of January 2015.

This not only underlines the importance of using multi-model ensembles, but also demonstrates that each of the ensembles does a good job in reproducing the observed situation and could be relied upon to capture the envelope of possible weather futures. Figure 6 displays the dispersion of the RCRV, and gives an indication of the spread in each of the models.

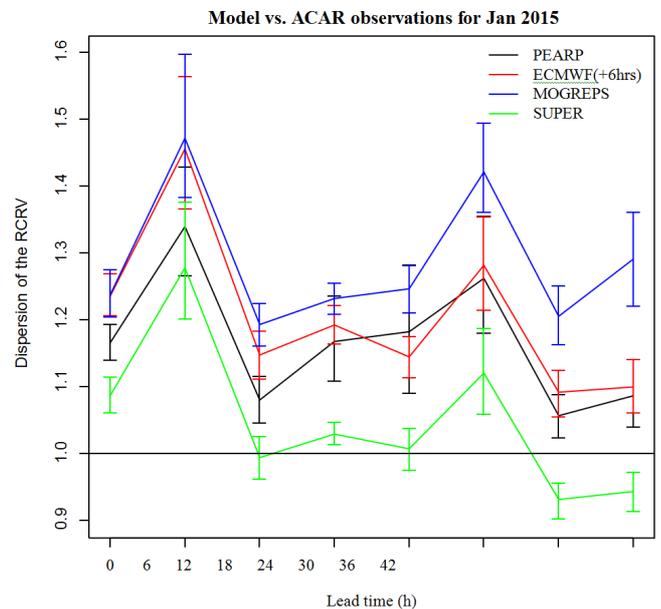


Figure 6 The dispersion of the RCRV for all operational models and for the SUPER ensemble for the month of January 2015. As in Figure 5, the error bars represent the 5th and 95th percentile values of the score.

In order to compliment the ensemble spread illustrated by the RCRV, each model's bias is shown in Figure 7. Zero bias means that the ensemble neither under- nor over-forecasts the wind at altitude with air pressure 250hPa, while negative

values suggest a slight under-forecast, with positive values indicating an over-forecast. The dispersion of the RCRV in Figure 6 illustrates that at early lead times, all models tend to under-disperse somewhat. However from the +12 hour lead time onwards (the first realistically usable ensemble output), the dispersion of all models increases illustrated by the values getting closer and closer to 1. Indeed for the SUPER ensemble, the perfect score of 1 is attained at +12 hour lead time while it also displays the best levels of dispersion of all the models for all other lead times. This underlines again the clear advantages of a multi-model ensemble approach for TP as even at lead times of +36 hour the SUPER ensemble gives an excellent level of dispersion and thus is sure to capture the entire envelope of possible weather futures.

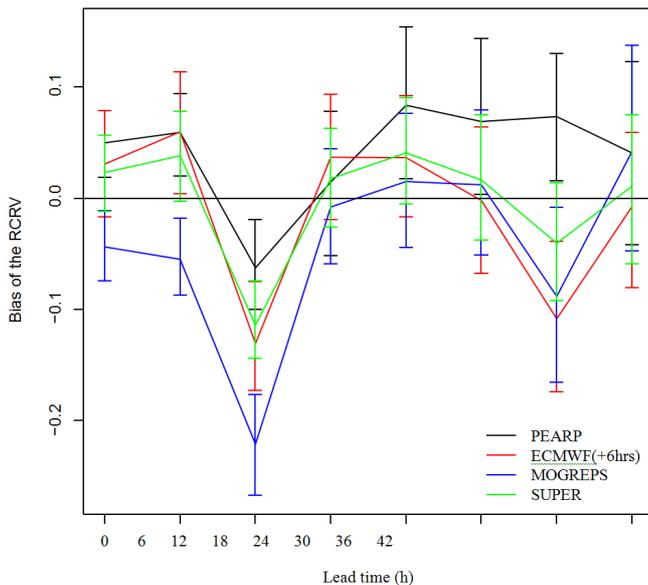


Figure 7 The bias of the RCRV for all operational models and for the SUPER ensemble for the month of January 2015. As in Figure 5 and Figure 6, the error bars represent the 5th and 95th percentile values.

Figure 7 displays the bias of the RCRV, and further verifies this point. All models give very low levels of bias with values varying between 0.2 and 0.1. This alone points to a reliable representation of the observed situation. However, as was the case in Figure 6, it is the SUPER ensemble bias which follows the most satisfactory evolution throughout the different lead times and thus underlines its ability to more accurately capture the observed variability than simply one operational ensemble would on its own.

All of the scores illustrate a very high degree of confidence in the different models' ability to accurately and reliably forecast the observed situation. However, a SUPER ensemble utilising ensemble products from many operational centres seems to be the most reliable and would thus be a preferred option for the TP.

IV. RESULTS

As an example, the MOGREPS ensemble was used as an input to a Dijkstra-based [12] TP in order to demonstrate quantitatively the benefits of using ensemble TPs. The case study of a flight from London (EGLL) to New York (KJFK) on the 25th of January 2015 was used. This could also have been done using the ECMWF, Météo France and SUPER ensembles, but for the purposes of this study, solely the results obtained using the MOGREPS model, are reported upon.

Figure 8 provides a graphical representation of the trajectory matrix for this case. The trajectories T^i are shown in individual panels and a PDF of the flight times for each trajectory is shown in the bottom right of each panel. This (Gaussian) PDF has been constructed using the mean value and the spread of the 12 flight times obtained from the trajectory ensemble. In this particular example, it is observed that the predicted (sub-)optimal trajectories are very similar for each member of the MOGREPS ensemble, with the exception of two members (shown on rows 2 and 3 of column 2, referred to as outliers) which picked a high latitude route.

Assuming the predicted trajectories shown are indeed optimal for the corresponding member of the ensemble, the outliers T^i have a PDF with a long tail, which implies that they might be the quickest to fly under certain weather projections and that it will take much longer to fly along other weather projections of the ensemble. In this case study, decision making is relatively easy as the outliers have a long average flight time (location of the peak of PDF) with large spread (width of PDF) in flight times. The user can choose any trajectory shown as they are all very similar to each other, both in terms of the trajectory taken and the PDF of spread times. The fact that the majority of the trajectories and their corresponding PDFs are similar in this case study provides an additional level of confidence.

V. CONCLUSIONS

The objective of this paper was to assess a PTP system incorporating state-of-the-art ensemble weather forecasts into an existing deterministic TP system. The methodology considered is applicable to any deterministic flight planning support TP system. Using ROC and RCRV as metrics, it can be shown, as has been illustrated here for TP, that all of the state-of-the-art EPSs considered (ECMWF, MOGREPS and PEARP) are capable of capturing specific nominal weather events observed from AMDAR measurements 36 hours before take-off time.

We have also illustrated that the Met performance can be further improved by combining the different EPSs to form a so-called SUPER ensemble. As an alternative to a deterministic TP system, we presented in this study an ensemble TP system from a single trajectory and the members of the ensemble Met forecast. Each member in the trajectory ensemble represents the optimal path from the origin to the destination predicted by TP using the corresponding weather

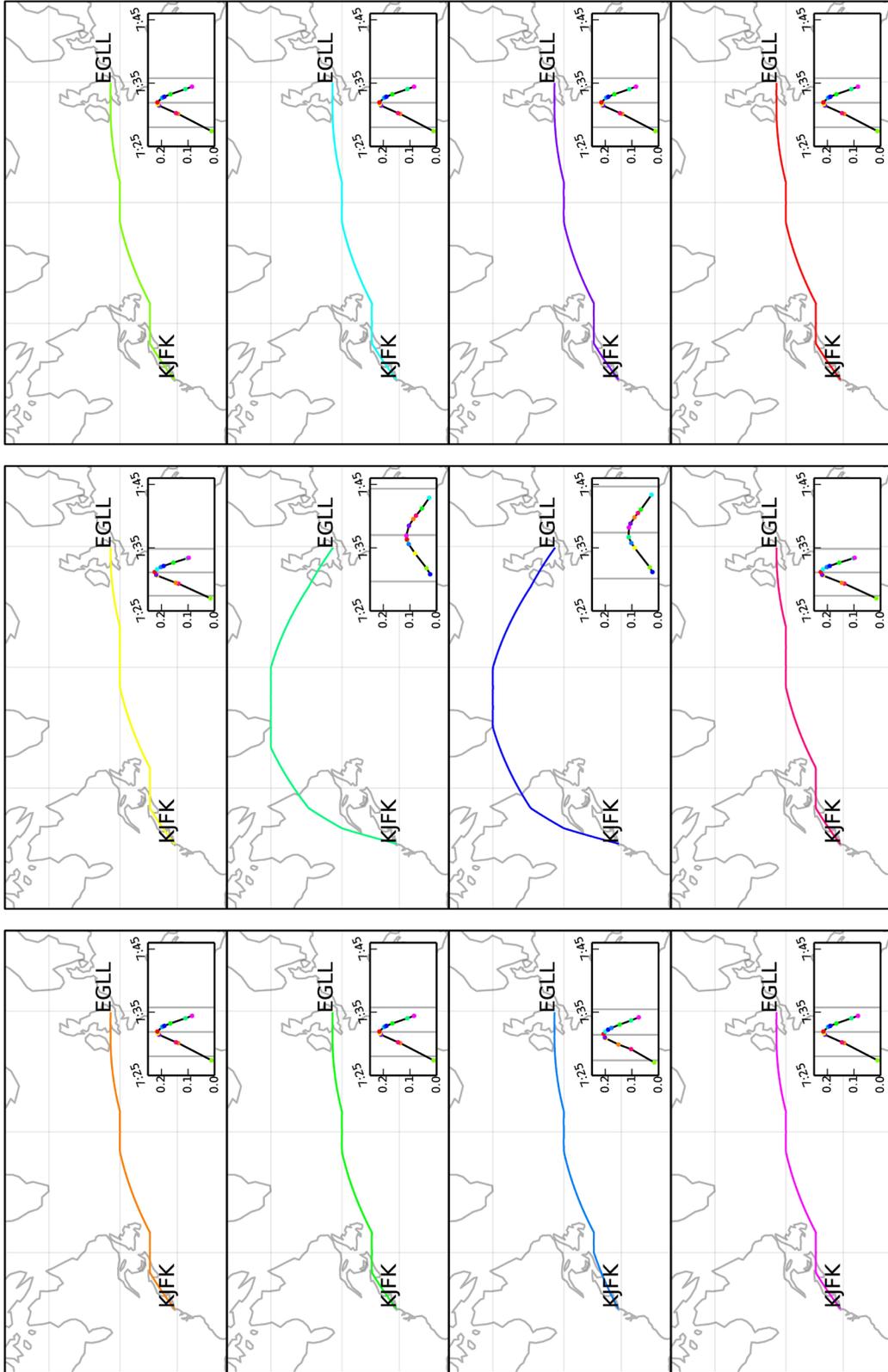


Figure 8. Graphical representation of the trajectory matrix. Each panel shows the (sub-)optimal trajectory calculated for a different Met scenario from the same ensemble forecast. A PDF is shown in the bottom right of each panel to show the spread of flight times for an aircraft to fly along the specified trajectory under different realisations of the Met ensemble forecast. Each Met scenario is represented by a unique colour and applies to both the trajectories and PDFs shown. The vertical grey lines in the PDFs show the standard deviation intervals. The x-axis and y-axis of the PDFs show the flight time [hh:mm] and probability respectively. Data shown is a flight from EGLL to JFK departing at 25th January 2015 0600UTC using a forecast with an analysis time of 23rd January 2015 1800UTC (t+36hr).

scenario (i.e. the member of the Met ensemble) as input. Assuming that there is no update to the observations available, it is impossible to determine which weather scenario represents the future the best, as EPSs are designed in such a way that each projected weather scenario is assumed equally probable. To support decision making, the uncertainty involved in each member of the trajectory ensemble was visualised. That is, for each member the calculation of the cost (e.g. flight duration, fuel usage) was repeated for each weather projection, yielding a PDF of the cost involved. This would allow TP users to select the trajectory that meets their optimum cost distributions (i.e. full/time constraints). For instance, if low uncertainty in the Required Time of Arrival (RTA) is mandated (e.g. at congested airports) the total fuel cost may necessarily increase. In contrast, minimising total costs at the expense of higher flight time uncertainties may be more appropriate for other flights.

The IMET approach is currently being validated in SESAR WP11.1, validation exercise 791.

ACKNOWLEDGEMENT

The authors thank SESAR Joint Undertaking for providing the co-funding which made this work possible.

REFERENCES

- [1] D. WILKS, Statistical methods in the atmospheric sciences, 3 red., Academic Press, 2011.
- [2] J. KROZEL, C. LEE and J. MITCHELL, „Turn-Constrained Route Planning for Avoiding Hazardous Weather,” *Air Traffic Control Quarterly*, vol. 14, nr. 2, pp. 159-182, 2006.
- [3] J. HEIJSTEK, A. HALLY, R. OSINSKI, J. CHEUNG, E. BLOEM, E. KNAPEN, R. ROOSIEN, J.-L. BRENGUIER and H. WELLS, „Report on optimisation of probabilistic TP and Met input, Deliverable D2.3 of WP-E project 02.40 IMET,” SESAR Joint Undertaking, Brussels, 2015.
- [4] J. KROZEL, J. MITCHELL, V. POLISHCHUK and J. PRETE, „Capacity estimation for airspace with convective weather constraints,” in AIAA Guidance, Navigation and Control Conference and Exhibit, Hilton Head, South Carolina, August 2007.
- [5] M. STEINER, R. BATEMAN, D. MEGENHARDT, Y. LIU, M. POCERNICH and J. KROZEL, „Translation of ensemble weather forecasts into probabilistic air traffic capacity impact,” *Air Traffic Control Quarterly*, 18(3), p. 229 – 254, 2010.
- [6] L. DESCAMPS, C. LABADIE, A. JOLY, E. BAZILE, P. ARBOGAST and P. CE' BRON, „PEARP, the Météo-France short-range ensemble prediction system,” *Quarterly Journal of the Royal Meteorological Society*, 2014.
- [7] N. E. BOWLER, A. ARRIBAS, K. MYLNE, K. B. ROBERTSON and S. E. BEARE, „The MOGREPS short-range ensemble prediction system,” *Quarterly Journal Of The Royal Meteorological Society*, 134, p. 703–722, 2008.
- [8] C. E. B. M. S. BISHOP, „Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects,” *Monthly weather review*, vol. 129, nr. 3, pp. 420-436, 2001.
- [9] R. BUIZZA, M. MILLER and T. N. PALMER, „Stochastic representation of model uncertainties in the ECMWF ensemble prediction system,” *Quart. J. Roy. Meteor. Soc.*, vol. 125, nr. 560, p. 2887–2908, 1999.
- [10] S. J. MASON and N. E. AND GRAHAM, „Areas beneath the relative operating characteristics ROC and relative operating levels (ROL) curves: Statistical significance and interpretation,” *Quarterly Journal of the Royal Meteorological Society* 128, 584, p. 2145–2166, 2002.
- [11] G. CANDILLE, C. CÔTÉ, P. HOUTEKAMER and G. PELLERIN, „Verification of an ensemble prediction system against observations,” *Monthly Weather Review* 135, 7, pp. 2688-2699, 2007.
- [12] E. W. DIJKSTRA, „A note on two problems in connexion with graphs,” *Numerische Mathematik* 1, 1, p. 269–271, 1959.
- [13] World Meteorological Organization, „The WMO AMDAR Observing System,” [Online]. Available: https://www.wmo.int/pages/prog/www/GOS/ABO/AMDAR/index_en.html

