Challenges of probabilistic turbulence forecasting

Producing and verifying probabilistic aviation turbulence forecasts

Dr Philip G Gill

Turbulence Mitigation Workshop 2, Washington DC, 4th September 2014

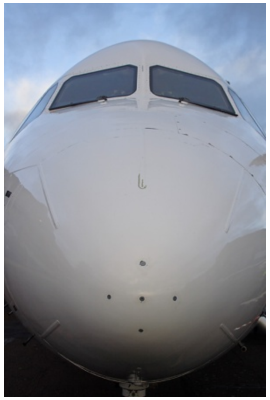
Contents

This presentation covers the following areas

- Introduction
- **Human challenges** - user understanding
- Current Met Office trials
- **Human challenges** - operational integration
- **Technical challenges**
    - Forecast optimisation, calibration and verification
- Summary
- Questions and answers

Introduction

- Turbulence - major cause of aviation incidents & active area of research
- Forecasts routinely produced by UK Met Office - World Area Forecast Centre (WAFC) service (along with WAFC Washington, USA)
- Operational forecasts currently derived from deterministic models
- There is always a degree of uncertainty in deterministic forecasts
- Probabilities generated from ensemble models are a way of communicating that uncertainty



Photos © P Gill

Met Office

Human challenges – user understanding

What is a probability?

Met Office

- A probability is a way of communicating the confidence in an outcome
- Probabilities in a forecast communicate how likely an event is to occur
  - For example a 30% probability of rain means there is a 3 in10 chance that rain will fall in the forecast period. Alternatively this means there is a 7 in 10 chance that it will remain dry.
- Probabilities in routine use in aviation – TAFs
  - PROB30, PROB40 etc.
- Probabilities in routine use for public forecasts (in US)
  - Probability of Precipitation (PoP)

Why use probabilities?

- Allows the **confidence** in a forecast to be communicated
- Studies have shown that probabilistic forecasts can be **more skilful** and have **more value** than deterministic forecasts for turbulence.
- More likely to give an indication of **extreme events**
- Verification of current deterministic WAFC forecasts shows that forecasts for MOG turbulence (max CAT potential>4) occur approx 1 out of 1000 ten-minute flight tracks demonstrating the **uncertainty**.

How are probabilities created?

- **Probabilities** can be created by using a range of different predictors. An estimate of the **probability** can be obtained by calculating the percentage of predictors forecasting turbulence - **GTG** scheme (Sharman et al 2006)
- Alternatively an **ensemble** model can be run giving multiple possible outcomes for a predictor. An estimate of **probability** can be obtained by calculating the percentage of ensemble members that forecast turbulence.
- These two approaches can be combined to create several probabilistic predictors which can then be combined to form a single probability – UK Met Office trials (Gill and Buchanan, 2014)

Example of Ensemble Forecasting in Nature

- DRY forecast

Example of Ensemble Forecasting in Nature (cont.)

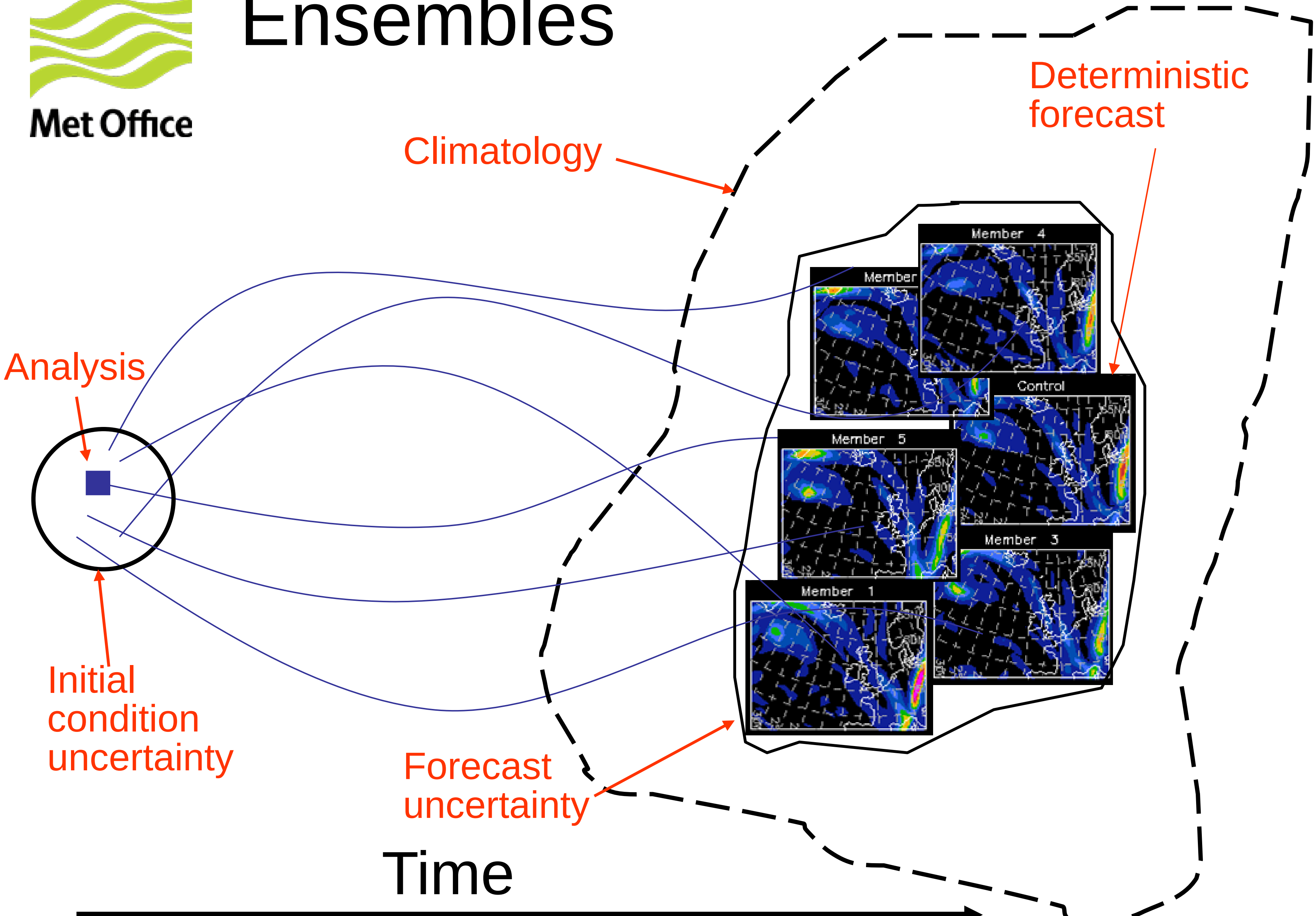
- WET forecast

Photo © C Murrin

Example of Ensemble Forecasting in Nature (cont.)

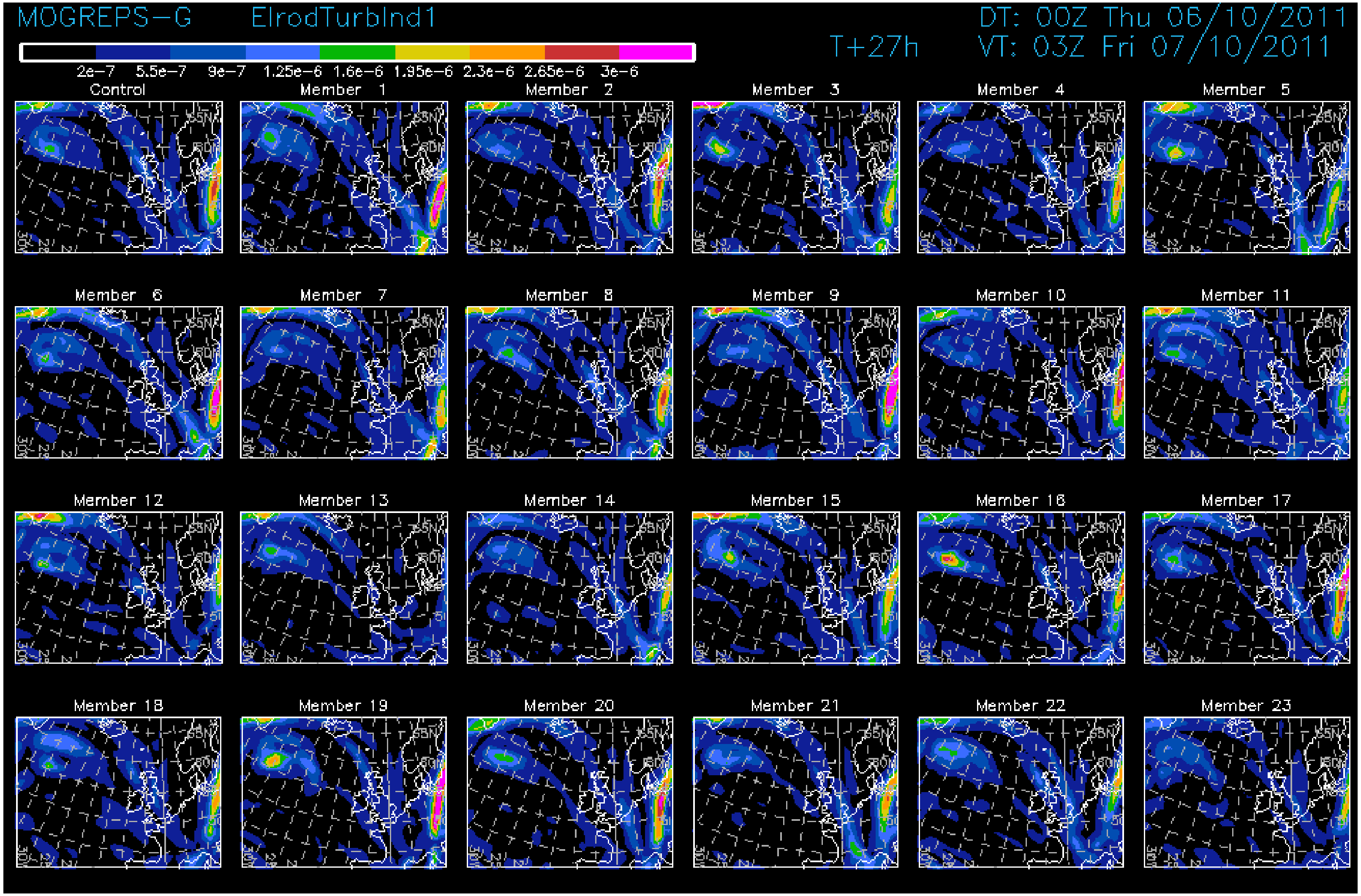
- Uncertain precipitation forecast

Ensembles



Climatology
Deterministic forecast
Analysis
Initial condition uncertainty
Forecast uncertainty
Member 4
Member
Control
Member 5
Member 3
Member 1
Time

MOGREPS-G “postage stamp” plots



MOGREPS-G

Met Office Global and Regional Ensemble Prediction System
4 cycles per day | 12 members per cycle
24 member products by lagged averaging of last 2 cycles
Operational since 2008 following 3 years of trials

- **Global Component (MOGREPS-G)**
- 33km, 70 Levels (N400L70)
- T+7 days
- Run at 00Z, 06Z, 12Z and 18Z
- ETKF for IC perturbations
- Stochastic physics (SKEB2) and random parameters for model physics

Met Office

Met Office probabilistic turbulence trials

Philip Gill and Piers Buchanan

Ensemble turbulence trials

- Over two years of trials from November 2010 to December 2012.
- **Objective** verification of deterministic and probabilistic model forecasts against automated aircraft observations.
- **Five** thresholds used on each predictor to generate probability forecasts.
- **Eight** numerical predictors and climatology combined using weightings derived from performance in previous 12-months and verified.
- **Near-operational** production of probabilistic turbulence forecasts using **MOGREPS-G** global ensemble since December 2013.

Turbulence predictors

Turbulence can come from different sources – **wind shear, convection, mountain-wave**

Windshear related:

- Ellrod TI1, Ellrod TI2
- Brown
- Dutton
- Lunnon

Convection related:

- Convective rainfall rate
- Convective rainfall accumulation

Both wind shear and convection: Richardson number

Turbulence climatology

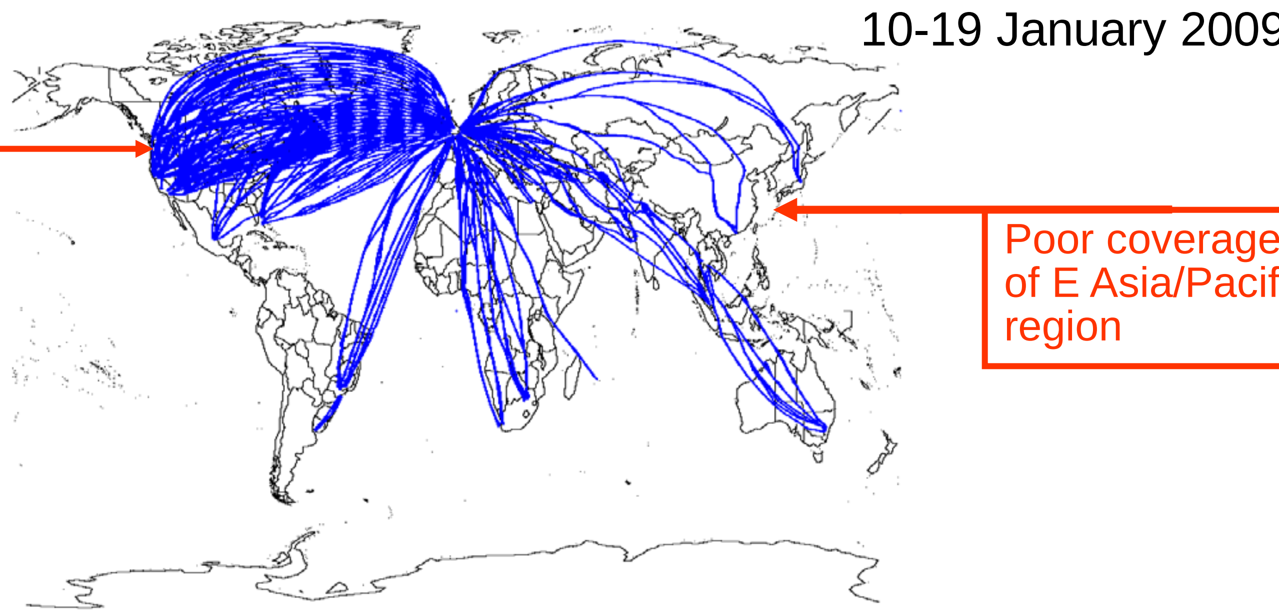
- Gridded field of observed turbulence frequency produced from aircraft observations from previous year
- Frequency of light or greater and moderate or greater turbulence climatology produced

Aircraft observations

Met Office

- **Global** coverage, but flights mainly over northern hemisphere
- **Automated** aircraft observations available every 4 seconds

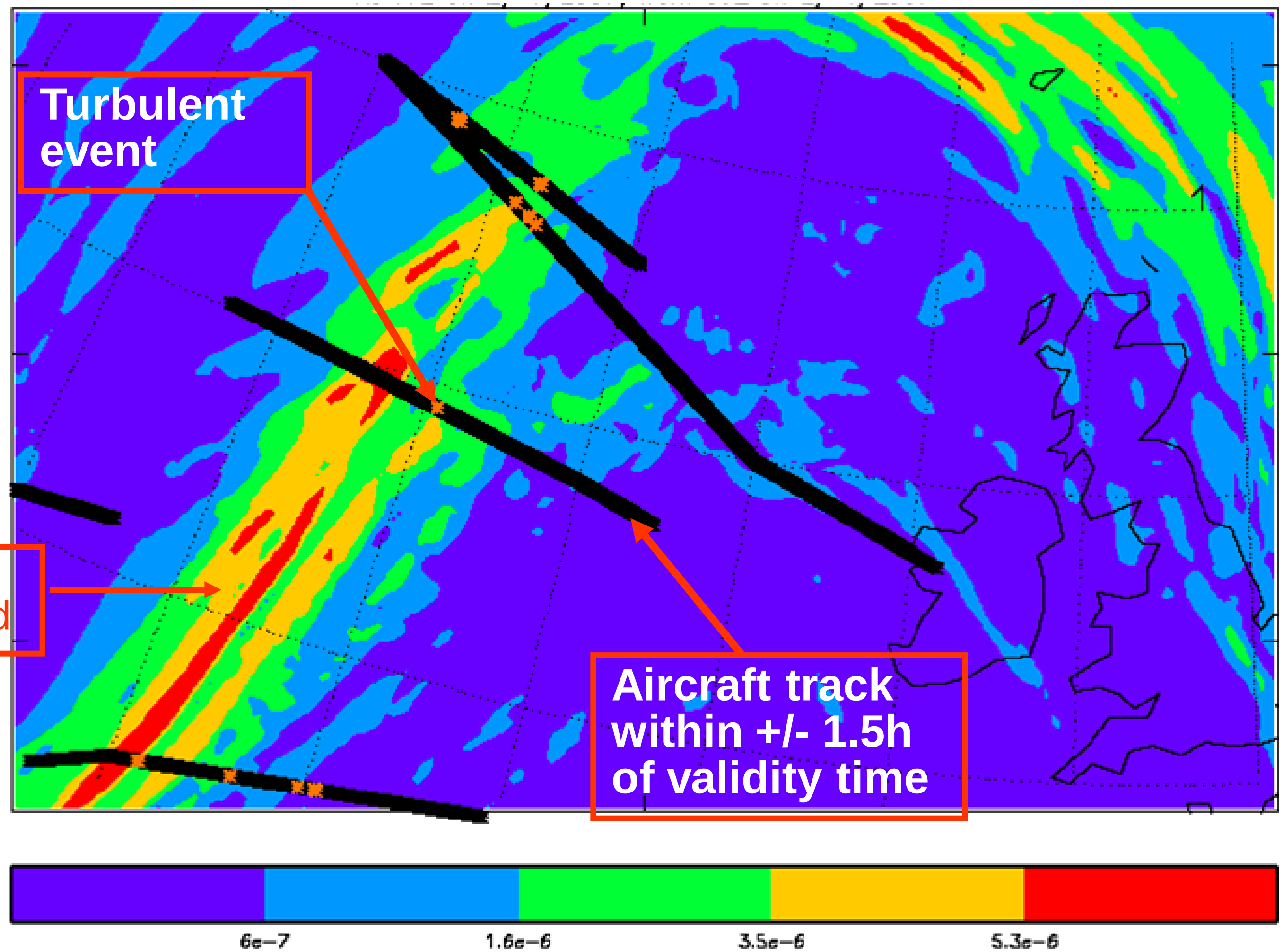
Good coverage of N Atlantic, US and Europe



Poor coverage of E Asia/Pacific region

- Derived Equivalent Vertical Gust (DEVG) – Measurement of observed turbulence derived from **vertical acceleration**, **aircraft mass, altitude and airspeed**

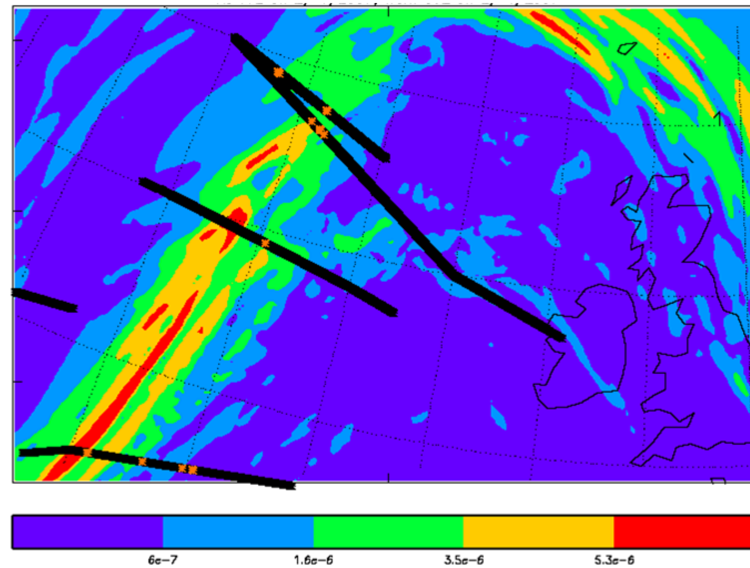
Verification methodology



Ellrod TI1 – Currently used by WAFC for gridded forecasts

Forecast assessment

- Turbulent/non turbulent event defined on **10min aircraft track** ~120km - approx grid size of WAFC grid
- Forecast probability of exceeding a certain threshold for given turbulence indicator
- Observed (moderate or greater) turbulent event - **DEVG>=4.5m/s**
- Construct 2x2 contingency tables for each threshold
- Sum entries in contingency tables over the verification period



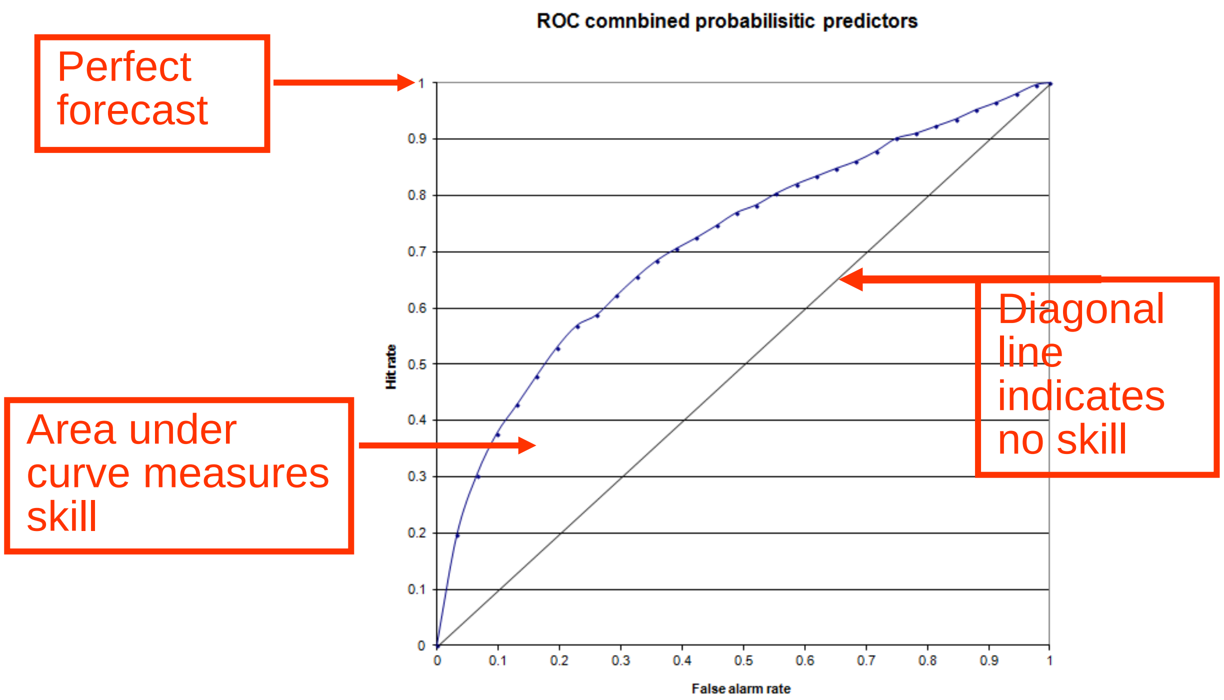
	Turbulence observed	No turbulence observed
Turbulence forecast	Hit	False alarm
No turbulence forecast	Miss	Correct rejection

2x2 contingency table

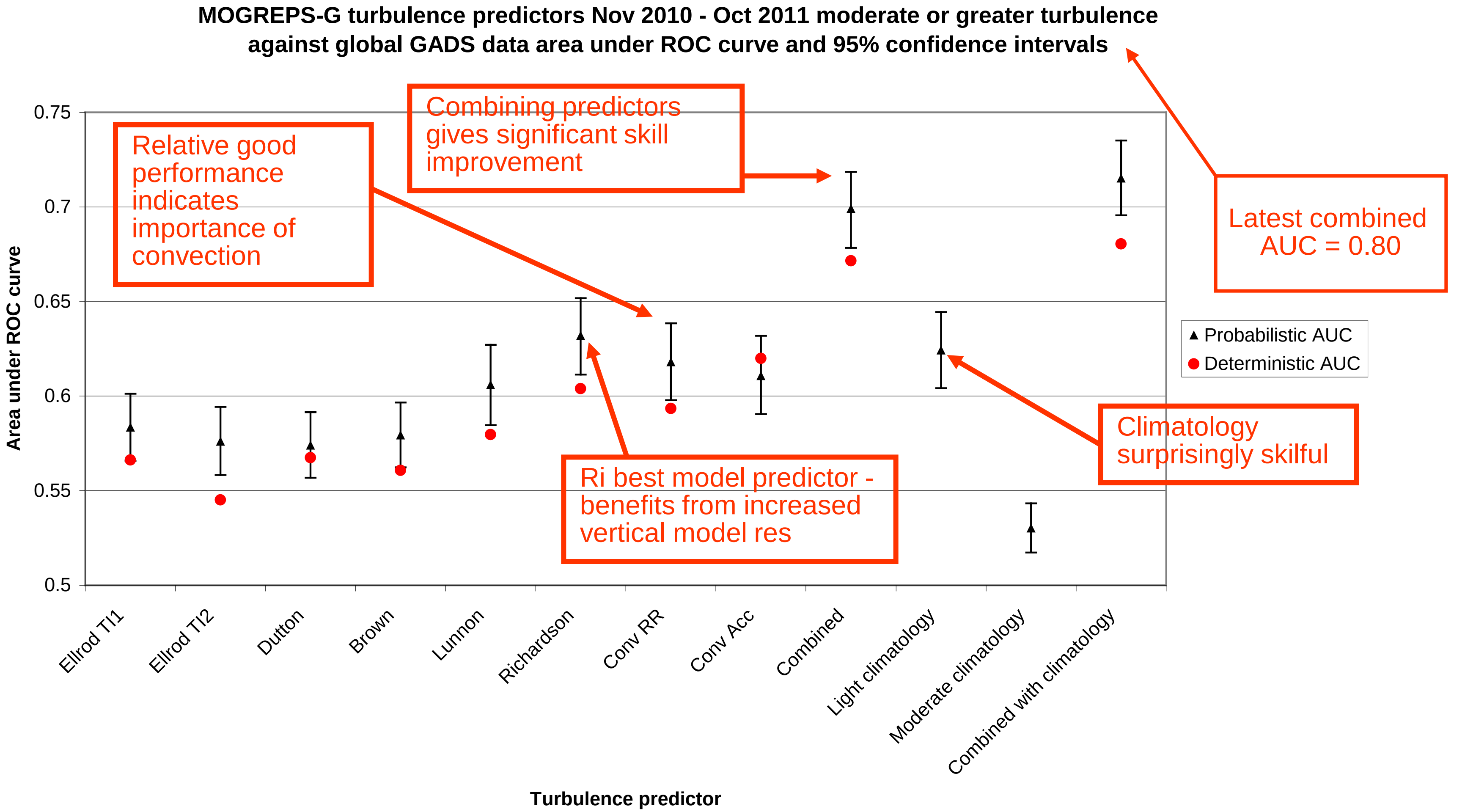
Gill PG. 2014. “Objective verification of World Area Forecast Centre clear air turbulence forecasts”, *Meteorological Applications.* **21:** 3-11

Verification measures - Skill

- **Relative Operating Characteristic** (ROC) curve – created by plotting the hit rate against false alarm rate for each threshold. The area under the ROC curve is a measure of skill. Useful for both **deterministic** and **probabilistic** forecasts.



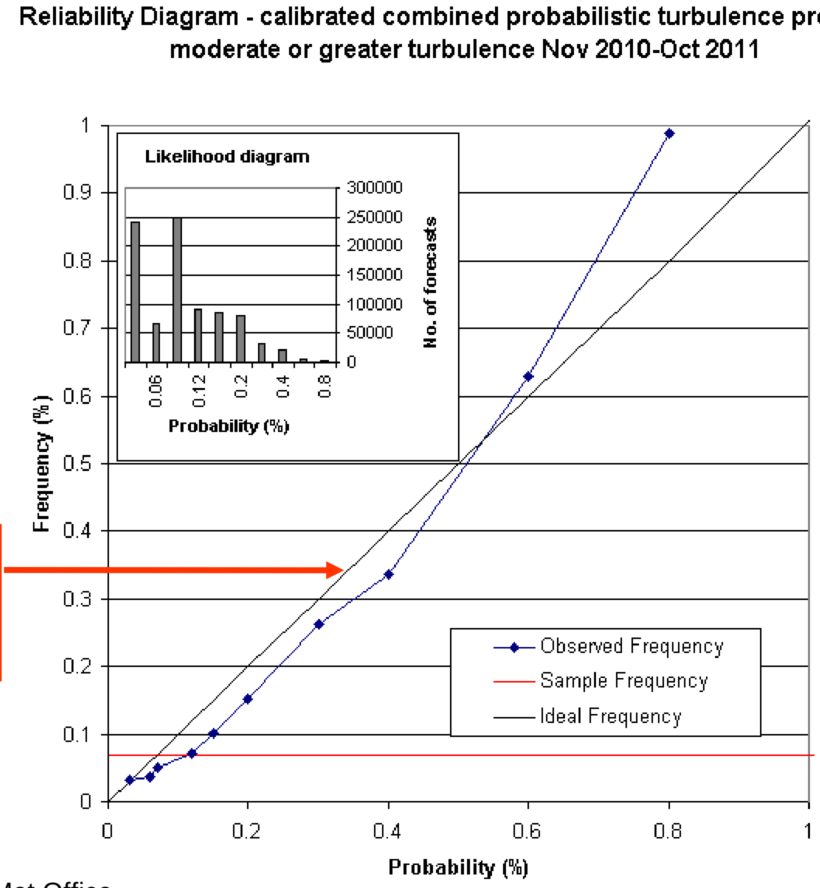
Verification measures – comparing skill of predictors



Gill PG, Buchanan P. 2014. "An ensemble based turbulence forecasting system", *Meteorological Applications.* **21:** 12-19.

Verification measures - Reliability

- **Reliability Diagram** by plotting the forecast probability against the frequency of occurrence



Perfect reliability on diagonal

Low probabilities but significant compared to background frequency

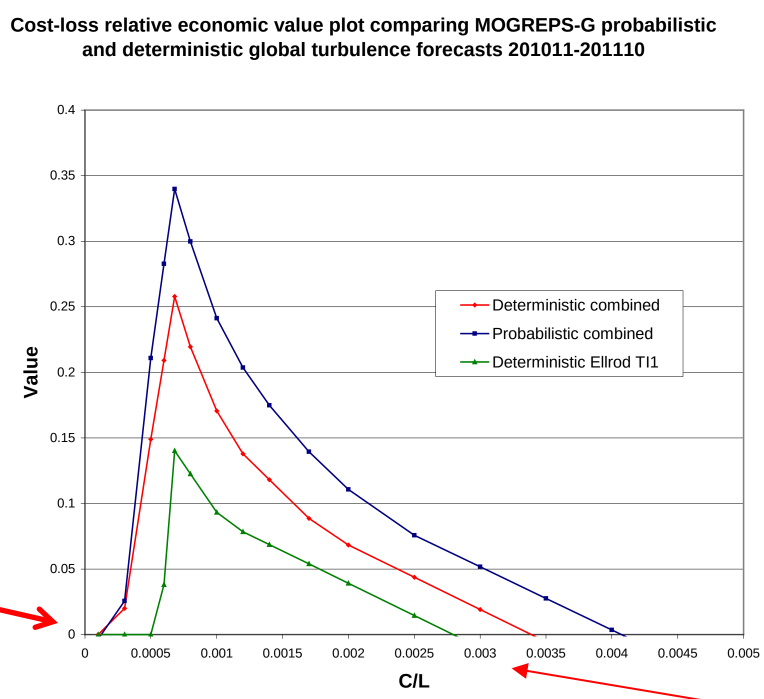
Verification measures - Value

- **Relative economic value** (Richardson, 2000) by calculating the value for a range of cost/loss ratios. Useful for both **deterministic** and **probabilistic** forecasts.

Perfect forecast would score 1

Relative economic value between 0 and 1

Climatology forecast would score 0



Can be used for comparing the value to a user for different forecasts

Ratio of cost/loss for user

© Crown copyright Met Office

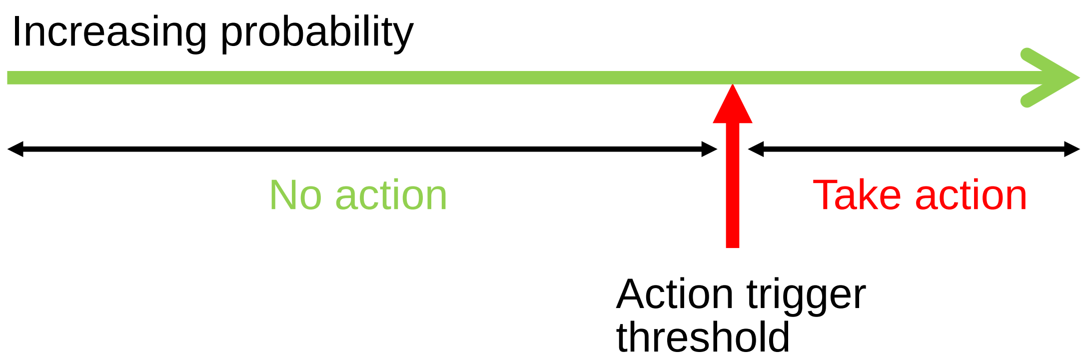
Human challenges – operational integration

When to take mitigating action?

With a binary forecast action is clear:

Event forecast – take action
Event not forecast – take no action

With a probabilistic forecast a decision needs to be made on which probability value to act on:

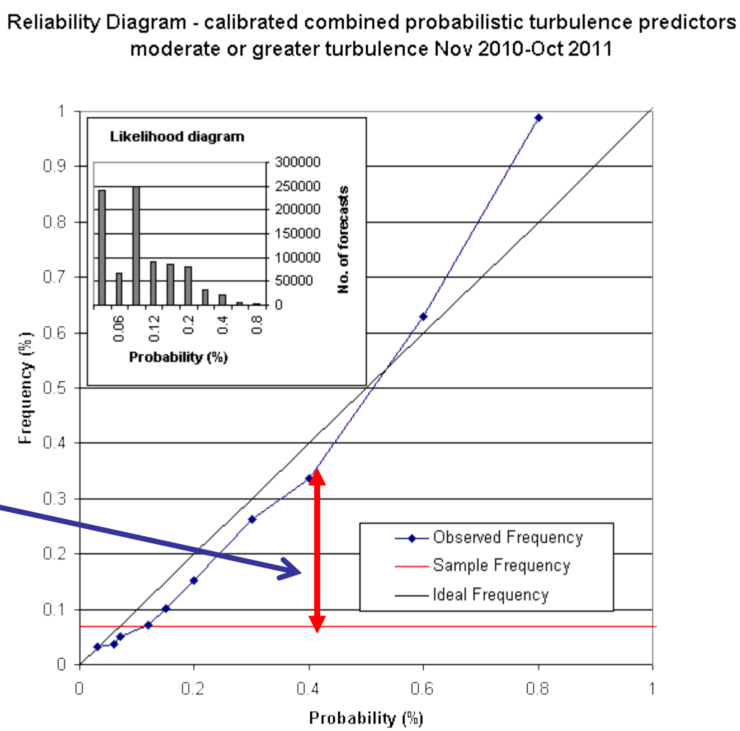


How do we determine the trigger threshold?

Selecting a trigger threshold – comparing to climatology

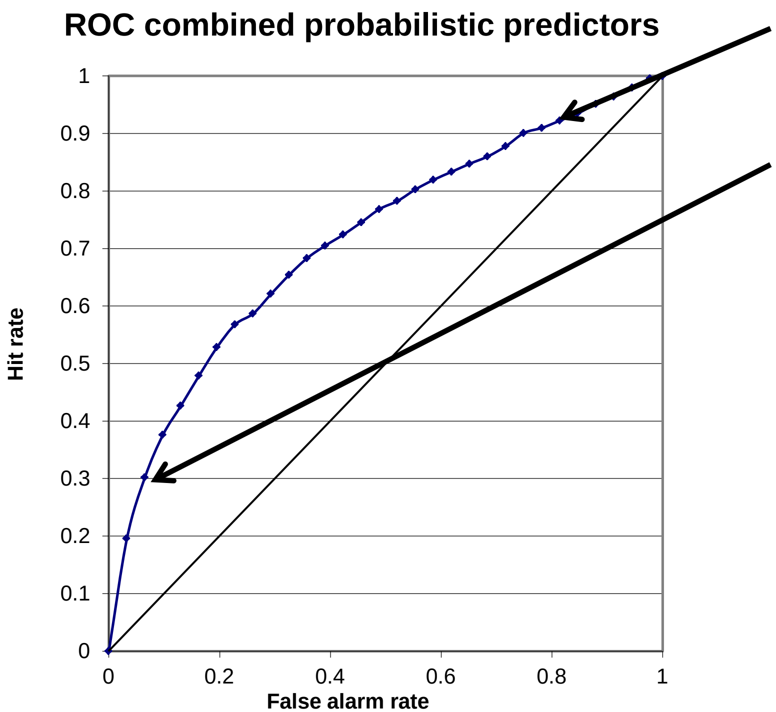
- Probabilities need to be considered with some knowledge of the background frequency of encountering turbulence.
- If the climatological probability of encountering turbulence is around 0. 05 % then a 0.5 % probability is actually a significant increase (10 times).

Difference from climatology



Selecting a trigger threshold – using verification

- Verification of past events can be used to help the user make a more informed decision.
- The user can view recent verification statistics to decide on a threshold that will give an acceptable balance between hits and false alarms



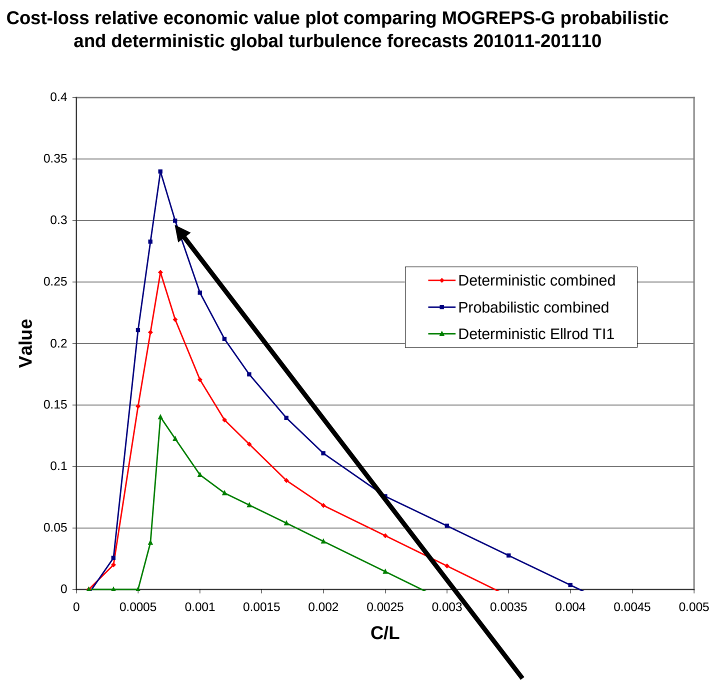
A low probability threshold will increase the number of hits but also increase the number of false alarms

A high probability threshold will decrease the number of false alarms but also decrease the number of hits

	Turbulence observed	No turbulence observed
Turbulence forecast	Hit	False alarm
No turbulence forecast	Miss	Correct rejection

Selecting a trigger threshold – considering cost-loss ratios

- Verification presented in terms of a cost-loss ratio could be used to maximise the user's value of a forecast

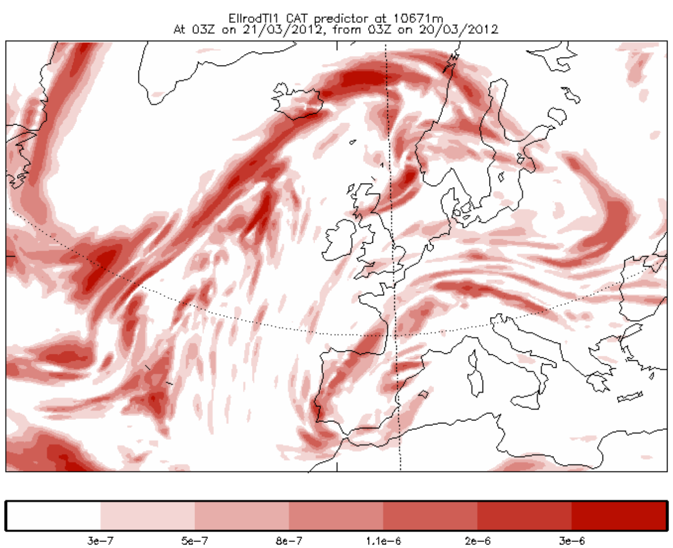


	Turbulence observed	No turbulence observed
Turbulence forecast Action taken	Hit **COST +** **REDUCED LOSS**	False alarm **COST**
No turbulence forecast No action taken	Miss **LOSS**	Correct rejection

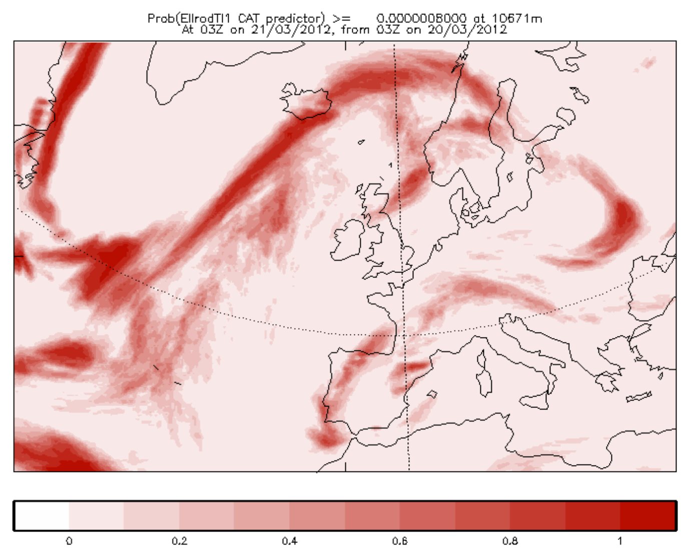
An estimate of the user's cost/loss ratio can be used to estimate the value of the forecast and determine the optimum operating threshold.

Visualisation – contour plots

- Contour plots could be simple outlines – similar to existing WAFC SIGWX charts
- Contour plots could show a range of probability values – similar to WAFC gridded products
- Could present a contour plot as the difference of the forecast from local or global climatology

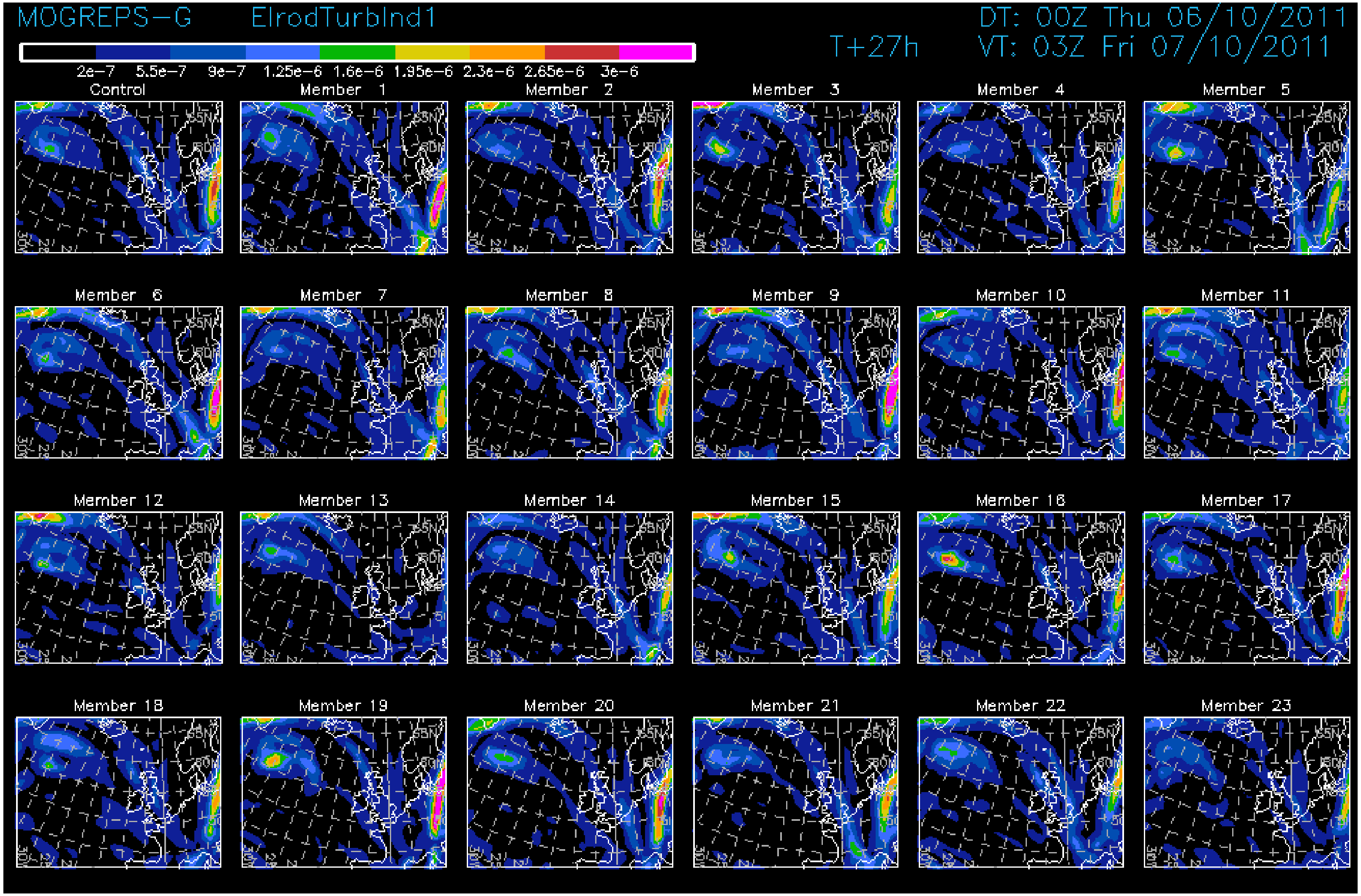


Deterministic



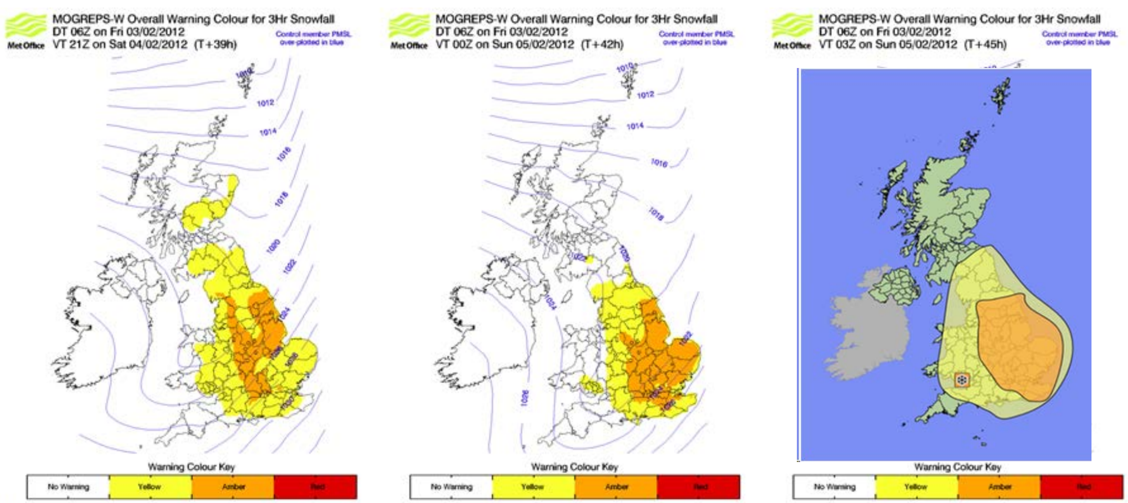
Probabilistic

Visualisation – postage stamp plots

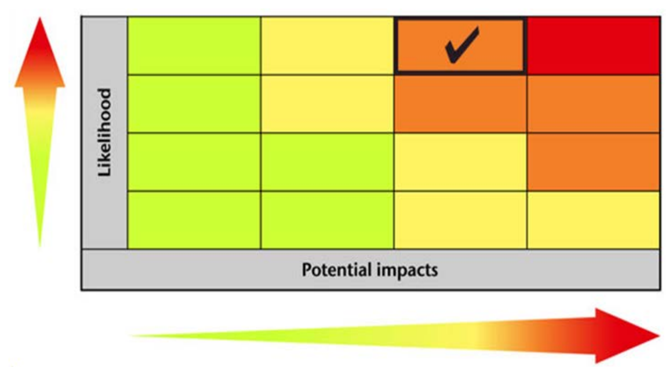


Visualisation – hazard matrix

- Currently in use for UK Public Weather Service severe weather warnings
- Shows the likelihood and impact of severe weather at a particular point or region
- Ensemble guidance for forecasters to write warnings (EPS-W)



Hazard matrix



Neal *et al*. 2014. Ensemble based first guess support towards a risk based national severe weather warning service, *Meteorological Applications*. **21**: 563-577

Met Office

Technical challenges – optimisation

Lisa Murray and Philip Gill

Combining predictors

Met Office

- Combining turbulence predictors has been shown to increase forecast skill (**Sharman** *et al*, 2006)
- Studies including convective predictors show a further increase in skill (Gill and Stirling, 2013)
- Met Offfice trials currently use weights derived from verification using ROC area to combine shear and convective turbulence predictors, using an iterative scheme.
- Predictors combined using a weighted sum

Optimisation

Met Office

- An efficient way needs to be found to use the probabilities from a range of turbulence predictors, ensemble members and ensemble systems
- Performance based weighting – simple to implement, fast to compute, may not be optimal
- Iterative schemes – improved performance, can be computationally expensive, may not be optimal
- **Logistic regression** – statistical process, scalable to larger number of predictors, may not be optimal
- Currently trials are ongoing looking at each of the above methods. Results can be computationally expensive and may not converge to the optimal blend.

Technical challenges – calibration

Piers Buchanan and Philip Gill

Calibration

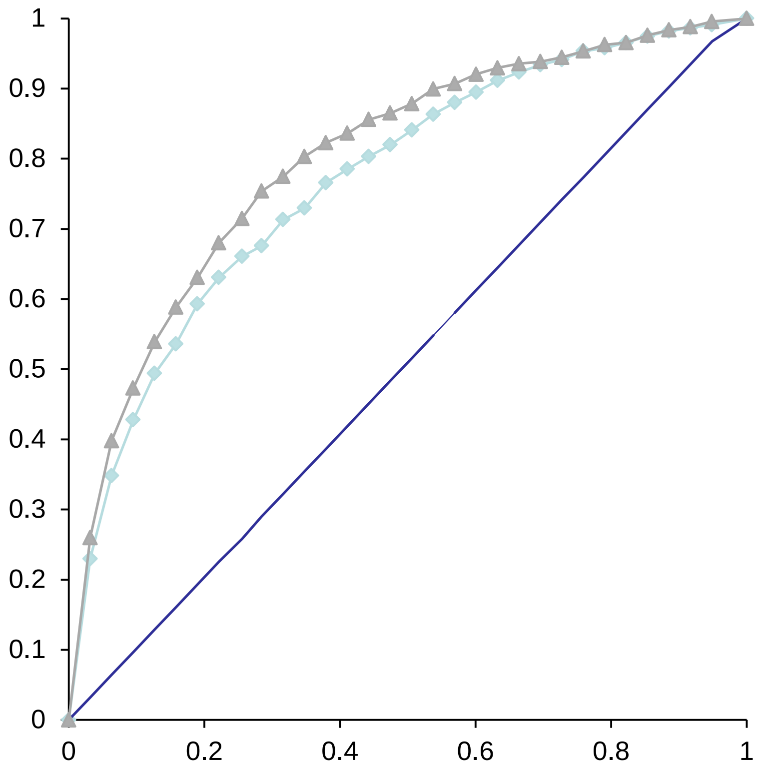
- A **reliability** diagram visualises the frequency an event is observed in each probability category. Eg. Need to make sure that a 30% forecast actually occurs 30% of the time it is forecast.
- Simple **calibration** can **improve the reliability** of the forecast and **maintain skill**. Use of verification data can determine whether the forecast frequency and observed frequency are equivalent in each probability band.
- Resulting probabilities are low but still **significant** compared to the climatological frequency.

Time Periods

- First Year (Training)
  - Nov. 2010 to October 2011
- Second Year (Assessment)
  - Nov. 2011 to October 2012
- Major MOGREPS-G upgrades in period
  - 28th March 2012- upgrade to run 12 members every 6 hours instead of 24 members every 12 hours
  - (Just beyond period) 16th Jan 2013: Upgrade from 60km horiz. res in mid latitudes to 33km horiz. res. in mid latitudes.

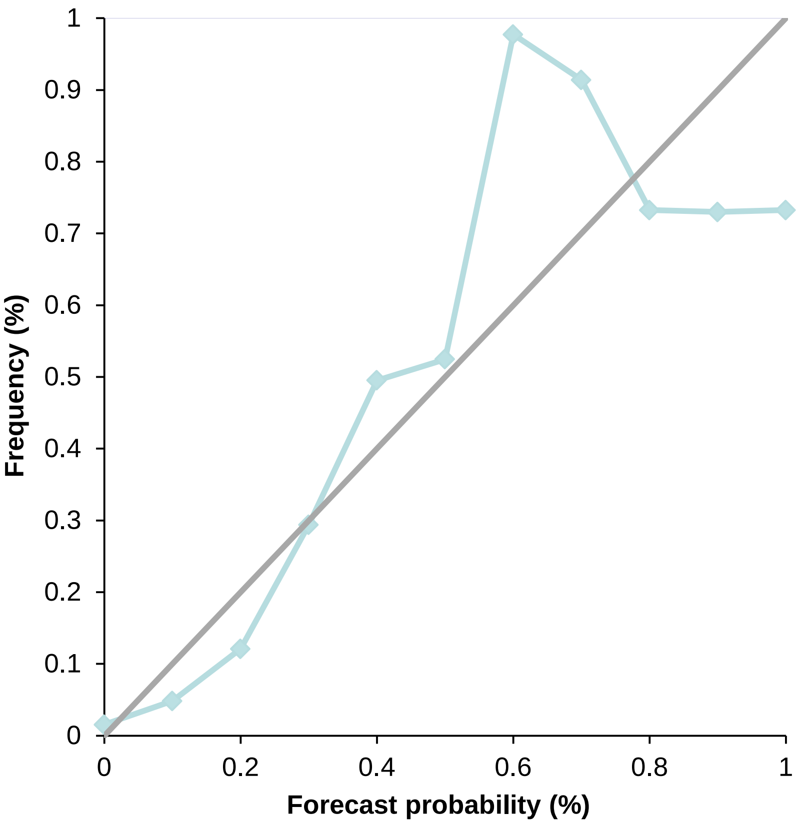
Second year of trial

ROC curve



Probabilistic AUC = 0.796
Deterministic AUC = 0.769

Reliability diagram



Technical challenges - verification

Verification issues to resolve

Met Office

- Do we have observations in the right place?
    - Aircraft observations are difficult to get access to.
    - Aircraft don't sample the atmosphere uniformly.
- Is the quality control adequate to remove suspect data?
    - Analysis has shown that a range of errors can be seen in the data.
    - More rigorous quality control is needed.
- What are the most suitable measures to use?
    - Important to look at a range of measures to build up a full picture.
    - Working with users will help to identify the most appropriate ways of measuring forecast performance.

Partial ROC area

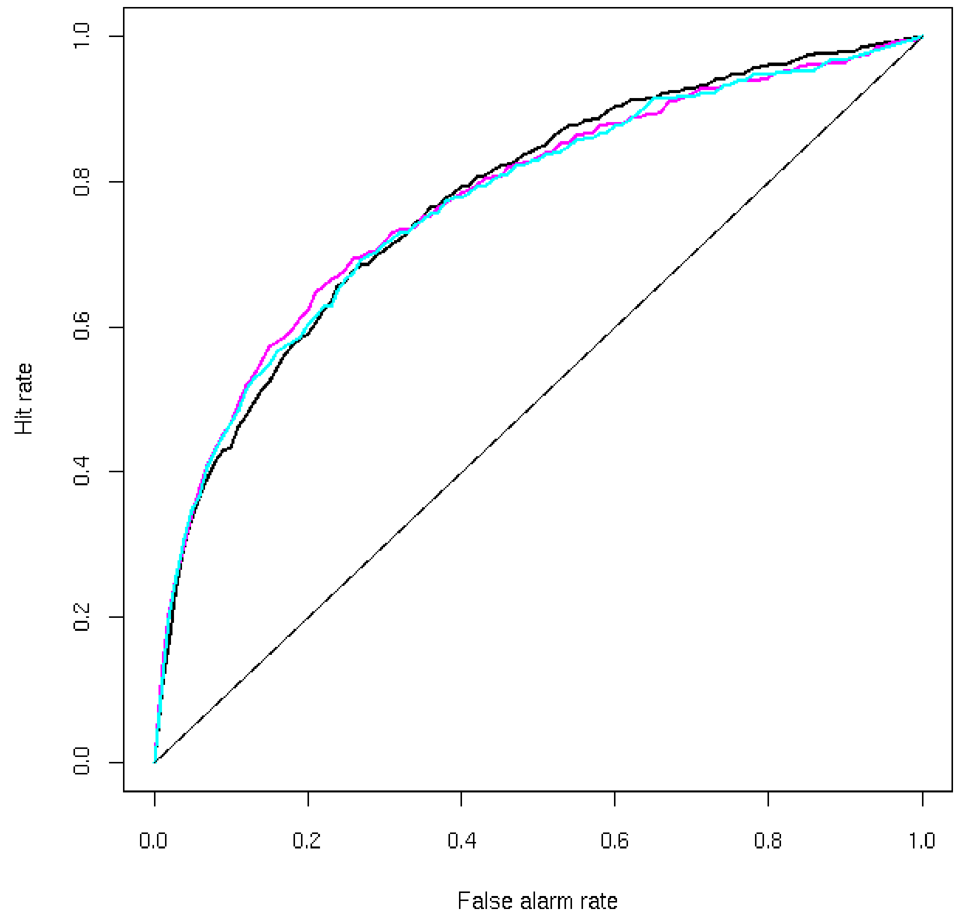
As most users are unlikely to accept very high false alarm rates or very low hit rates then an acceptable performance area can be identified and the skill at forecasting in this range measured using the **partial area under the ROC curve**.

The AUC scores for each of the curves in the opposite figure are:

Black - 0.7759
Light Blue - 0.7709
Pink – 0.7756

If you imagine the separation between the pink and black curves to be larger at both the lower and upper ends of the false alarm rate, their AUC scores would still be similar. However, if the models had been optimised using the **partial area under the curve** (pAUC), would the model represented by the pink curve have proven to be considerably better, the same or worse?

This is an example of where the AUC score can be misleading, and future work is planned to trial optimising techniques based on the pAUC up to a false alarm rate of 0.2.



Summary

Summary

- Benefits of probabilistic turbulence forecasts
  - Confidence can be communicated with every forecast
  - Significant increase in skill
  - Increased economic value of forecast
- Human challenges
  - Education in using and interpreting probabilistic forecasts is critical
    - Working more closely with users is important to enable this
- Technical challenges
  - Current work on optimisation, calibration and verification of forecasts is promising
    - Further research is needed to fully utilise the benefits of probabilistic turbulence forecasting

Future plans

Met Office

- Implement **calibration** method to apply to near-operational system by March 2015.
- Investigate using a **multi-model ensemble** for WAFC turbulence forecasts by March 2016 using UK and ECMWF ensembles (NCEP to follow in collaboration with WAFC Washington)
- Extend verification to include additional probabilistic predictors (CAPE, Mountain wave routinely produced)
- Investigate **optimisation** using different techniques and measures including partial area under the ROC curve
- Seek **additional EDR observations** to give greater coverage to the verification

Acknowledgements

- Thanks to the UK **Civil Aviation Authority** for funding this project

- The content of this presentation appears in the following papers:

Gill PG, Stirling A. 2013. "*Including convection in global turbulence forecasts." Meteorological Applications. **20**: 107-114.*

Gill PG, Buchanan P. 2014. "An ensemble based turbulence forecasting system", *Meteorological Applications.* **21:** 12-19

Gill PG. 2014. "Objective verification of World Area Forecast Centre clear air turbulence forecasts An ensemble based turbulence forecasting system", *Meteorological Applications.* **21:** 3-11

Any Questions?

philip.gill@metoffice.gov.uk

Questions & answers

Derived Equivalent Vertical Gust

Met Office

Turbulence severity	DEVG (ms^{-1})
None	DEVG ≤ 2
Light	2 ≤ DEVG < 4.5
Moderate	4.5 ≤ DEVG < 9
Severe	DEVG ≥ 9

$$DEVG = \frac{Am|\Delta n|}{V}$$

Where $|\Delta n|$ = peak modulus value of fractional deviation of aircraft normal **acceleration** from 1*g* in units of g.

m = total aircraft **mass** in metric tonnes.
V = calibrated **airspeed** at the time of occurrence of the acceleration peak, in knots.
A = An aircraft specific parameter which varies with flight conditions, and may be approximated by the following formulae:

$$A = \bar{A} + c_4(\bar{A} - c_5)\left(\frac{m}{\bar{m}} - 1\right) \qquad \bar{A} = c_1 + \left(\frac{c_2}{c_3 + H(kft)}\right)$$

H = **altitude** in thousands of feet
$\bar{m}$ = reference **mass** of aircraft in metric tonnes.

The parameters depend on the aircraft and values appropriate for the B747-400 were used (Truscott, 2000).