The Weather Test
A BBC News *collaboration with scientific bodies to produce for a public evaluation of weather forecasts*

1. Background

1.1. In September 2010 BBC News convened a group of scientists from across meteorology and statistics to consider the idea of a public test of the comparative accuracy of weather forecasts. This followed a summer of high profile of media coverage about the value of seasonal forecasts in the UK and the Met Office's decision to cease publication, for the time being at least, of its seasonal predictions.

1.2. The group formulated an idea for an approach that would assess forecasts provided by a range of providers, using a range of techniques, and to do so using methods that whilst being scientifically robust were simple for the general public to understand – this was not an attempt to duplicate the large range of scientific work that is already done in the area of forecast verification, but simply to make this more generally accessible across a common approach to different forecasters.

1.3. This now Steering Group for the project convened two public meetings, under the leadership of BBC News and co-ordinated by the Royal Meteorological Society, on 12 October 2010 and 28 March 2011, and based on these meetings and the input from the Steering Group members, a draft protocol for the project was produced. This protocol was put out to written consultation and following the feedback received, the Steering Group have revised the protocol, which is presented here.

1.4. As with all consultations there were a range of suggestions and proposals, which are not all possible to accommodate because of resource constraints and in some areas a lack of consensus. However, many of the suggestions offered useful improvements to the protocol have been taken on board. The protocol now presents a methodology that many in the community have had the opportunity to help shape and will achieve the aims of the project.

1.5. The assessment phase of this project will begin in earnest in Spring 2012.

2. Introduction

2.1. Forecasts of the weather are produced for many timescales and it is our view that it would not be possible to conduct a true evaluation of these forecasts without including all of the timescales that the public have an interest in. Therefore this project will evaluate forecasts for days 1, 3 and 5 (or as they are referred to in operational meteorology as forecasts at time T for T+24, T+72 and T+120), and seasonal forecasts. For simplicity in what follows we will call our day-1, -3 and -5 forecasts 'daily'.

2.2. There are a number of different elements of the weather that are forecast and any number of these could be evaluated. However, because of a combination of resource constraints, the limited availability of some observations, the lack of consensus on how to measure some aspects of weather, and a desire to focus on the main aspects of weather that affect the public, the project will concentrate on forecasts of rainfall amount, maximum and minimum temperature and wind speed.

2.3. Daily forecasts will be assessed at a number of observation sites across the UK. These have been chosen as key sites where a range of information is already made widely available as part of the public weather service. Because of the nature of both

seasonal forecasts and a desire to maintain scientific integrity, these forecasts will be evaluated for the UK as a whole, rather than at specific observations stations.

2.4. All of the forecasts will be evaluated in a probabilistic framework, and the protocol outlines proposed methodologies that are scientifically robust but that have been chosen for ease of understanding by none specialists in this area. We reiterate that this is not a project to critique the current science of forecast verification but rather to conduct a simple but scientifically robust evaluation which is by design intuitive and that the general public will find easy to understand.

2.5. Scientists from the National Centre for Atmospheric Sciences, whose headquarters is based at the University of Leeds, will conduct the evaluation, independent from any of the forecast providers. The scientists will report their findings back to BBC News via the project Steering Group. All of the outcomes from this evaluation will be made public on timescales that reflect the statistical significance of the assessment data being collected. Further, in order to ensure probity and maintain commercial confidences, the forecasts provided for evaluation will only be released by the project if permission is given by those who produced the forecasts.

3. Submitting forecast information for evaluation

3.1. The evaluation process needs to be automated and as such forecast information will be submitted to the study team electronically (by e-mail or through a password-protected web interface) in a format to be agreed between the provider and the academic lead (who will be undertaking the evaluation work).

3.2. The submission of information has to be timely in order to maintain the integrity and independence of all the different forecast information being provided. With this in mind daily forecast information will be required by 1800 UTC on the day previous to that which the forecast is issues. For seasonal predictions, these forecasts will be submitted no later than 3 days in advance of the beginning of the forecast period.

3.3. Late submissions will be regarded as missing data. If a forecast provider fails to submit at least 95% of the dates for a forecast range, they will be excluded from that particular stage of Weather Test assessment. There will be an annual cycle for the evaluation of daily forecasts and a five-yearly cycle for seasonal forecasts.

4. Data on which assessments will be made

4.1. There has been a long discussion on the suitability of the observations station against which forecasts should be assessed, and there is no simple answer on how to design the optimal network of stations for this project. However the Steering Group agreed with comments from the consultation, that using more stations than the 6 originally proposed will provide an improved quality of assessment. Given the resources and the value in using sites where data is more widely available, it has been decided that the project will use all the 35 stations from the UK's national climate observing network for which data is made available publicly by the Met Office. Summary statistics of these station data can be viewed at: 'http://www.metoffice.gov.uk/climate/uk/stationdata/'.

4.2. Observations will be collated at the end of each calendar month to conduct the evaluation *en bloc* rather than continuously.

4.3. The following definitions apply for this protocol:

- maximum temperature is defined as the maximum temperature occurring between 0900 UTC and 2100 UTC;
- minimum temperature is defined as the minimum temperature occurring between

2100 UTC and 0900 UTC the next day;

- daily rainfall is defined as the total rain in mm falling during the 24h period ending 0900 UTC the next day;
- wind speed against the threshold of 20knts is the number of instances in any one hour, throughout the 24h period ending 0900 UTC the next day, when wind speed is measured as greater than 20 knts (defined in the Beaufort Scale as Beaufort Number 6, signifying a Strong Breeze).

5. The evaluation methodologies for daily forecasts

5.1. The providers will be asked to provide a forecast for daily maximum (Tmax) and minimum (Tmin) temperature, daily rainfall total (Ptotal), wind speed exceeding 20 knts in any one hour (W20knts), as defined in 4.3 above, for each of the 35 observation stations defined in 4.1. An example submission format is shown in Section A.1 of Appendix A.

5.1.1. For temperature

The forecast Tmax and Tmin will be compared directly with the observed values. The number of days when Tmax and Tmin are within ±2 deg C and the number of days when these are greater ±5 deg C will be recorded.

5.1.2. For rainfall

Forecasts of daily rainfall totals for each of the observation station sites will be provided in terms of the probability that it would fall into each of the three categories, < 2mm, 2-5 mm, > 5 mm. The evaluation of these forecasts will be based on the classical Brier score defined as follows:

$$B = (1/n)\ \Sigma\ (p_i - x_i)^2$$

where p_i is the forecast occurrence probability and x_i equals 1 for occurrence and 0 for non-occurrence. B is therefore zero for a perfect forecast.

The Brier Skill Score is then defined as:

$$BSS = 1 - B/B_{ref}$$

where B is the Brier score of the forecast and B_{ref} is the Brier score of the background 30-year 1971–2000 climatology as provided by the UK's National Climate Information Centre. This way BBS equals one for a perfect deterministic forecast and zero for a forecast exactly as skilful as the 30-year climatology. Negative values occur for forecasts with no skill.

5.1.3. For wind

Forecast will be provided for the number of occasions in any one hour that wind speed exceeds 20 knts. The number of forecast occurrences will be compared directly with the number of observed occurrences.

6. The evaluation methodologies for seasonal forecasts

6.1. The nature of seasonal forecasts are such that they will be assessed over the UK as a whole rather than at specific observing stations. Forecasters will be asked to provide percentage probabilities for temperature and rainfall across the three categories of lower, average or higher than average as compared with the 30-year 1971–2000 climatology for the UK, provided by the UK's National Climate Information Centre. An example submission format is shown in Section A.2 of Appendix A.

6.2. These probability forecasts will be evaluated using Relative Operating Characteristics (ROC) curves. These curves are common for evaluating probabilistic information and are well suited to analysis of the type of information that will be submitted for seasonal forecasts. These curves show a plot of the proportion of events that were forecast and actually occurred against the proportion that were forecast but did not occur, sometimes referred to as the hit rate and the false alarm rate respectively. The area under the ROC curve provides a measure of the forecast skill.

6.3. In addition Reliability Diagrams will be derived. These plot the observed occurrence against the forecast probability and again are a very useful graphical indication of forecast skill.

7. Publication of results

7.1. A digest of the results from these evaluations will be made available to the public on the BBC News website and the Today programme. Full details will be on theWeather Club website. The evaluation of the daily forecasts will be released at the end of each full year of assessment. The results from the evaluation of the seasonal forecasts at the end of five years of full evaluation. The reason for taking longer to consider seasonal forecasts is to ensure that there are sufficient numbers of forecasts to make the outcome scientifically meaningful.

7.2. It is not the purpose of this project to provide a league table of how accurate certain providers are against each other. Many other factors would need to be reviewed and a much fuller verification programme undertaken in order for that to be a fair and reasonable assessment. The aim of this project is to demonstrate to the public how accurate different types of forecasts are when produced by different methods and approaches. With this in mind, alongside of the evaluation outcomes we will also be providing some information about how forecasts are produced. We very much hope that this will help the public to understand more about weather forecasts, to be able to use them more effectively, to gain a greater understanding of how to make use of forecast information which has a degree of uncertainty associated with it.

8. The Selection of participants

8.1. In order to ensure that there are a sufficient range of methods, approaches and complexity amongst those taking part in the project, the Steering Group have selected a list of participants it would like to invite. These are (listed alphabetically):

- Joe Bastardi – a professional meteorologist in the USA who makes active use of generally available weather model predictions and a collection of indices that track the characteristics of the weather around the globe;
- David King – an amateur observer and forecaster using a combination of weather lore and methodologies he has established based on historical observations of nature;
- Met Office – the UK's National Meteorological Service and provider of the public weather service. The Met Office runs its own numerical weather prediction models on global and local scales and has a operational forecasting centre with human input to the first guess forecasts provided by the models;
- MeteoGroup – Europe's largest commercial weather provider, making forecasts with its own forecasters and numerical prediction forecast models, driven by boundary conditions from the European Centre for Medium-range Weather Forecasting (ECMWF);
- Metra – a commercial provider arm of the New Zealand National Meteorological Service. The models used are similar in concept but different in application to

those of the Met Office, and these are utilised with the input of forecasters to provide services in the UK;

- Positive Weather Solutions – a more recently-established smaller commercial weather forecast provider, who forecast for Europe by taking weather model output and indices to create a range of forecast products;
- Weather Action – a company that make forecasts of the weather and its impact using solar and lunar patterns and cycles.

8.2. This group of providers represents the broad spectrum of methods and forecasting techniques and models currently used within Europe.

8.3. Within the limits of resources we would not wish to exclude any provider of weather forecasts who feel that they can usefully add to the diversity of methods and approaches that the selected list represents. Therefore the project will consider proposals from additional provider who would like to take part. However they will need to demonstrate that they offer something significantly different from the list of active participants.

9. References

Jolliffe, I.T. and D.B. Stephenson, 2003. Forecast verification: A practitioner's guide in atmospheric sciences. Wiley.

Nurmi, P., 2003: Recommendations on the verification of local weather forecasts. ECMWF Technical Memorandum No. 430, ECMWF, Reading, UK, 19pp.

World Meteorological Organisation, 2007: Standardized verification system (SVS) for long-range forecasts (LRF). Attachment II-8 to the Manual on the GDPFS (WMO-No. 485), Volume I.

The document has been compiled by the Royal Meteorological Society on behalf of the project Steering Group from feedback received at the written and oral consultations. Any correspondence should be addressed to Prof Paul Hardaker at 'chiefexec@rmets.org'.

APPENDIX A: Sample data submission formats

A.1. Possible example of a daily forecast submission file (every day at 1800 UTC):

```
Provider:       X Weather
Date, Time:     15/03/2011,18:00:00 UTC

DAILY 15/03/2011 09 UTC to 16/03/2011 09 UTC
DAY-1
          Tmax      Tmin         Prcp              Wind
                                <2  2-5 >5     1  2  3  4  .. 24
Stn 1     12.8      6.4        0.2 0.8 0.0    0  0  2  1  .. 0
Stn 2     12.4      5.4        0.1 0.8 0.1    0  0  1  0  .. 0
   …
Stn 6     12.1      6.0        0.3 0.7 0.0    0  0  0  3  .. 0

DAY-3
          Tmax      Tmin         Prcp              Wind
                                <2  2-5 >5     1  2  3  4  .. 24
Stn 1     11.8      3.4        0.0 0.8 0.2    0  0  1  2  .. 0
Stn 2     11.4      4.4        0.1 0.8 0.1    0  0  0  1  .. 0
   …
Stn 6     11.1      3.0        0.0 0.7 0.3    0  0  0  2  .. 0

DAY-5
          Tmax      Tmin         Prcp              Wind
                                <2  2-5 >5     1  2  3  4  .. 24
Stn 1     10.8      4.4        0.8 0.2 0.0    0  0  0  1  .. 0
Stn 2     10.4      3.4        0.9 0.1 0.0    0  0  5  0  .. 0
   …
Stn 6     10.1      4.0        0.7 0.3 0.0    0  0  1  2  .. 0
```

A.2. Possible example of a seasonal forecast submission file (3 days before the start of every month):

```
Provider:       X Weather
Date, Time:     29/03/2011,18:00:00 UTC


SEASONAL 01/04/2011 0900 UTC to 01/07/2011 0900 UTC
          LOW       AVERAGE        HIGH
TEMP       30         45            25
PRCP       40         35            25
```

APPENDIX B: Those associated with the project

Project Sponsor:	Fran Unsworth	Head of Newsgathering, BBC
Project Initiator and Steering Group Chair:	Roger Harrabin	The BBC's Environment Analyst
Academic Lead:	Dr Peter Knippertz	School of Earth and Environment, University of Leeds
Project Secretariat:	Prof Paul Hardaker	Royal Meteorological Society
Project Website:	www.theweatherclub.org.uk/features/article/be-a-part-of-the-bbc-s-weather-test	Hosted by theWeather Club

Those attending Steering Group Meetings:

Roger Harrabin	BBC Environment Analyst
Dominic Groves	BBC Today Programme
Prof Lenny Smith	London School of Economics
Dr Matt Owens	Royal Astronomical Society
Prof Tim Palmer FRS	Royal Meteorological Society
Prof Paul Hardaker	Royal Meteorological Society
Dr Martin Dougherty	Royal Statistical Society
Alice Henchley	Royal Society
Dr Peter Knippertz	University of Leeds
Dr Sarah Norris	University of Leeds
Philip Eden	Weather UK and Chiltern Observatory
Dr Matthew Foote	Willis Research Network
Tricia Holly-Davis	Willis Research Network