Interpretation, verification and calibration of ternary probabilistic forecasts

Tim Jupp

University of Exeter

December 13, 2010
Introduction

Thanks to co-authors:
- Rachel Lowe
- David Stephenson
- Caio Coelho (CPTEC, Brazil)

My aims are to:
- interpret forecasts geometrically
- generalise ideas from binary forecasts to ternary forecasts

Paper in preparation for Theme Issue of the Philosophical Transactions of the Royal Society: ‘Climate predictions: the influence of nonlinearity and randomness’
Probabilistic rainfall forecast: current visualisation

- the forecast at each point is a *distribution*
- How should we assign *colours* to *distributions*?
The quantiles of a climatology

(tercile) categories $B$, $N$ and $A$ are observed with equal frequency
Continuous distributions (CDFs)

- Climatology CDF $G(x)$ [darkgrey], and forecast CDF $F(x)$ [black]

- Climatology: $G(x)$, forecast: $F(x)$
**Continuous distributions (PDFs)**

Climatology PDF \( g(x) \) [darkgrey], and forecast PDF \( f(x) \) [black]

- **climatology:** \( g(x) \), **forecast:** \( f(x) \)
Ternary climatology $q$ and forecast $p$

- ternary climatology: $q$ (for terciles $\Rightarrow q = (1/3, 1/3, 1/3)$)
- ternary forecast: $p = (p_B, p_N, p_A)$
Barycentric coordinates

every ternary forecast is a point in the triangle, including
the climatology $\mathbf{q}$
the observed state $\mathbf{o}$
Current visualisation methods

Usually based on a discretisation of ternary forecast space. For example:

1 (Dry): \((p_B > 2/5 \text{ and } p_N < 1/3 \text{ and } p_A < 1/3)\).
2 (Dry or normal): \((p_B > 1/3 \text{ and } p_N > 2/5)\) or \((p_B > 2/5 \text{ and } p_N > 1/3)\).
3 (Normal): \((p_B < 1/3 \text{ and } p_N > 2/5 \text{ and } p_A < 1/3)\).
4 (Wet or normal): \((p_N > 1/3 \text{ and } p_A > 2/5)\) or \((p_N > 2/5 \text{ and } p_A > 1/3)\).
5 (Wet): \((p_B < 1/3 \text{ and } p_N < 1/3 \text{ and } p_A > 2/5)\).

Algebraic descriptions make my head hurt.
Current visualisation methods (EUROBRISA categorical)
Current visualisation methods (EUROBRISA lower)
Current visualisation methods (EUROBRISA upper)

Integrated: Prob. of precip. in the upper tercile (above normal)
Issued: Jun 2009   Valid for JAS 2009

Categories based on $p_A$
Current visualisation methods (IRI)

IRI Multi-Model Probability Forecast for Precipitation for May-June-July 2009, Issued April 2009

`Most likely` categories

IRI colour scheme
A continuum of colours in forecast space

\[ H(p; q) \text{ [dashed], } \theta(p; q) \text{ [solid]} \]

\[ q' = (0.33, 0.33, 0.33) \]

\[ H(p; q) \text{ [dashed], } \theta(p; q) \text{ [solid]} \]

\[ q' = (0.10, 0.20, 0.70) \]

\[
H(p) = \frac{1}{\log \max_i q_i^{-1}} \sum_{i \in \{B, N, A\}} p_i \log \frac{p_i}{q_i}
\]

- \( H(p; q) \) is a measure of the subjective certainty in a forecast
Assignment of colours

Assignment of hues

Assignment of saturation

Normalised Angle \( \left( \theta - \theta_0 \right) \mod 2\pi / 2\pi \)

Subjective certainty \( H(p, q) \)

Designed to be intuitive and colour-blind friendly!
Our proposed colour scheme

Colours when $\theta_0 = 0^\circ$, $m = 1$

- use HSV (hue–saturation–value) colour space
- hue $\propto \theta(p)$
- saturation $\propto [H(p)]^m$ (usually $m = 1$)
Mapping colours $\leftrightarrow$ barplots

Colours when $\theta_0 = 0^\circ$, $m = 1$

$q' = (0.33, 0.33, 0.33)$
Alternative colour schemes

Colours when $\theta_0 = 0^\circ$, $m = 1$

$q' = (0.33, 0.33, 0.33)$

left: $p_B \Leftrightarrow p_A$

right: $RGB \Leftrightarrow CMY$

Colours when $\theta_0 = 60^\circ$, $m = 1$

$q' = (0.33, 0.33, 0.33)$
Climatology $q$ need not be defined by terciles

In all cases: $p = q$ is set to white
An example: seasonal forecasting

ECMWF probabilistic seasonal forecast
Alternative viewpoints: summary

Forecast $\mathbf{p}$, climatology $\mathbf{q}$, observation $\mathbf{o}$ can be regarded as

- vectors in $\mathbb{R}^3$ (non-negative, sum to unity)
- barplots with 3 bars
- vectors in $\mathbb{R}^2$ (i.e. points in a triangle)
- colours
Outline

1. Introduction
2. Probabilistic forecasts
3. Visualisation
4. Verification
5. Calibration
6. Conclusions
How good is a probabilistic forecasting system?

Compare forecasts \( \{ p \} \) with subsequent observations \( \{ o \} \).
Often, observations are treated as ‘certain’, so that

\[
o \in \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}
\]

Conditional averages defined as follows:

\( \bar{o} | p \), mean observation associated with given forecast

\( \bar{p} | o \), mean forecast associated with given observation
Consider skill measures which are \textit{quadratic forms}:

\[
S(\{p; o\}) = \left\langle (p - o)^T L^T L(p - o) \right\rangle
\]

for some matrix $L$.

We want small scores.

Note that

\[
S(\{p; o\}) = \left\langle \|p - o\|^2 \right\rangle
\]

when $\{p; o\}$ is expressed in an appropriate vector space.

In this space, the score is the \textit{mean square distance between forecasts and observations}.
Scores are squares of distances in appropriate simplex

Brier score (left) ; Ranked Probability Score (right)
Decomposition of quadratic scores

By definition $\langle o \rangle = \langle \bar{o} | p \rangle = q$
Discretise forecast space into bins $\{p_k\}$
It follows that:

$$\langle \| p_k - o \|^2 \rangle = \langle \| q - o \|^2 \rangle + \langle \| p_k - \bar{o} | p_k \|^2 \rangle - \langle \| q - \bar{o} | p_k \|^2 \rangle$$

score = uncertainty + reliability − resolution

A better (smaller) score may be obtained if raw forecasts $\{p\}$ are replaced by calibrated forecasts $\{\hat{p}\}$.
The Brier Score and the Ranked Probability Score

$$BS(p; o) = \frac{1}{2} \left[ (p_B - o_B)^2 + (p_N - o_N)^2 + (p_A - o_A)^2 \right]$$

- in $$\mathbb{R}^3$$: $$BS = \text{mean–square–distance difference between PDFs}$$
- in $$\mathbb{R}^2$$: $$BS = \text{mean–square–distance in equilateral triangle}$$

$$RPS(p; o) = \frac{1}{2} \left[ (p_B - o_B)^2 + (p_B + p_N - o_B - o_N)^2 + (p_B + p_N + p_A - o_B - o_N - o_A)^2 \right]$$

- in $$\mathbb{R}^3$$: $$RPS = \text{mean–square–distance difference between CDFs}$$
- in $$\mathbb{R}^2$$: $$RPS = \text{mean–square–distance in right–angled triangle}$$
(reduction in) uncertainty is square of distance

\[ \Delta U(q) \]

\[ q_0' = (0.33, 0.33, 0.33) \]
contour interval: 0.05

\[ U_0 = 0.33 \]

\[ q_0' = (0.50, 0.00, 0.50) \]
contour interval: 0.05

\[ \bar{U} = \bar{U}_0 - \Delta \bar{U} \]

Brier score (left) ; Ranked Probability Score (right)
Set radius $\propto 1/BS$
Set radius $\propto 1/BS$ with masking
<table>
<thead>
<tr>
<th>Outline</th>
<th>Introduction</th>
<th>Probabilistic forecasts</th>
<th>Visualisation</th>
<th>Verification</th>
<th>Calibration</th>
<th>Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Probabilistic forecasts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Visualisation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Verification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Calibration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Conclusions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In suitable 2D space (simplex),
Raw forecasts \( \{ p \} \) have score

\[
S(\{ p; o \}) = \langle \| p - o \|^2 \rangle
\]

Calibrated forecasts \( \{ \hat{p} \} \) have score

\[
S(\{ \hat{p}; o \}) = \langle \| \hat{p} - o \|^2 \rangle
\]

Calibrated forecasts \( \hat{p} \) chosen to minimise score.
Calculation of \( \hat{p} \) is a standard \textit{Generalised Linear Model} problem, easily soluble in \( \mathbb{R} \).
Traditional Reliability diagram (binary forecast)

This is a plot of $\bar{o}|p_k$ against $p_k$.

Sharpness diagram: distribution of forecasts $\{p\}$
Ternary Sharpness (all forecasts)

left: raw \{p\}, right: calibrated \{\hat{p}\}
Ternary Sharpness (conditional on $\mathbf{o} = (1, 0, 0)$)

left: raw $\{\mathbf{p}\}$, right: calibrated $\{\hat{\mathbf{p}}\}$
Ternary Sharpness (conditional on $\mathbf{o} = (0, 1, 0)$)

left: raw $\{\mathbf{p}\}$, right: calibrated $\{\mathbf{\hat{p}}\}$
Ternary Sharpness (conditional on $\textbf{o} = (0, 0, 1)$)

left: raw $\{p\}$, right: calibrated $\{\hat{p}\}$
Ternary reliability diagrams

Decomposition of a quadratic score

\[
\langle \| p_k - o \|^2 \rangle = \langle \| q - o \|^2 \rangle + \langle \| p_k - \bar{o} \| p_k \|^2 \rangle - \langle \| q - \bar{o} \| p_k \|^2 \rangle
\]

\[
\text{score} = \text{uncertainty} + \text{reliability} - \text{resolution}
\]
Ternary reliability diagrams

Decomposition of a quadratic score

$$\langle \| p_k - o \|^2 \rangle = \langle \| q - o \|^2 \rangle + \langle \| p_k - \bar{o} \| p_k \|^2 \rangle - \langle \| q - \bar{o} \| p_k \|^2 \rangle$$

score = uncertainty + reliability − resolution
Ternary reliability diagrams: raw forecasts

\[ \langle \| p_k - o \|^2 \rangle = \langle \| q - o \|^2 \rangle + \langle \| p_k - \bar{o} | p_k \|^2 \rangle - \langle \| q - \bar{o} | p_k \|^2 \rangle \]

score = uncertainty + reliability - resolution

threshold = 5  \( \delta p = 1/11 \)
Ternary reliability diagrams: calibrated forecasts

\[
\langle \| p_k - o \|^2 \rangle = \langle \| q - o \|^2 \rangle + \langle \| p_k - \bar{o} | p_k \|^2 \rangle - \langle \| q - \bar{o} | p_k \|^2 \rangle
\]

\[
\text{score} = \text{uncertainty} + \text{reliability} - \text{resolution}
\]
Nonlinear calibration

Raw forecasts in red,
calibrated forecasts in blue
Outline

1. Introduction
2. Probabilistic forecasts
3. Visualisation
4. Verification
5. Calibration
6. Conclusions
Conclusions

- assign unique colour to each ternary forecast
- barycentric coordinates aid interpretation
- white $\Leftrightarrow p \approx q$
- greater subjective certainty $\Leftrightarrow$ stronger colour
- extended verification and calibration concepts from binary to ternary
Special case: climatology $\sim N(\mu_c, \sigma_c^2)$, forecast $\sim N(\mu, \sigma^2)$