Interpretation, verification and calibration of ternary probabilistic forecasts

Tim Jupp

University of Exeter

December 13, 2010

Outline

Outline

Introduction

Thanks to co-authors:

- Rachel Lowe
- David Stephenson
- Caio Coelho (CPTEC, Brazil)

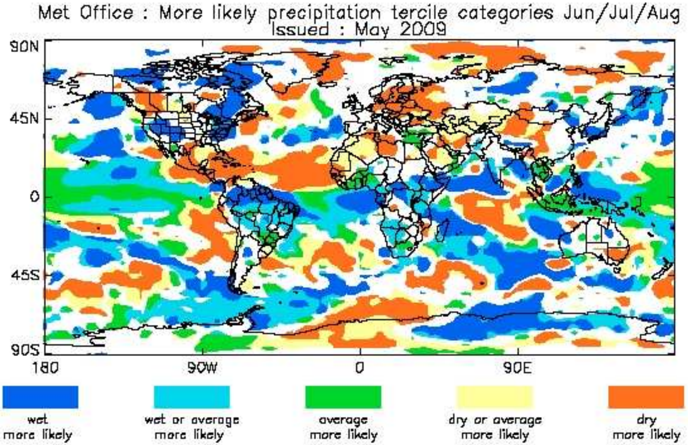
My aims are to:

- 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'**

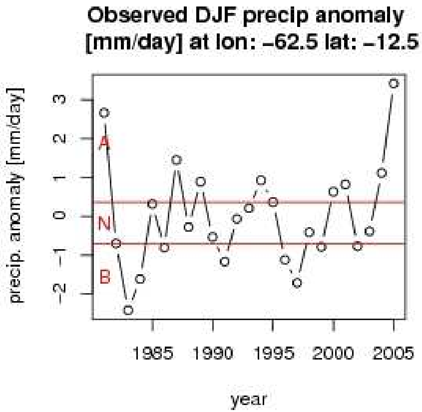
Outline

Probabilistic rainfall forecast: current visualisation



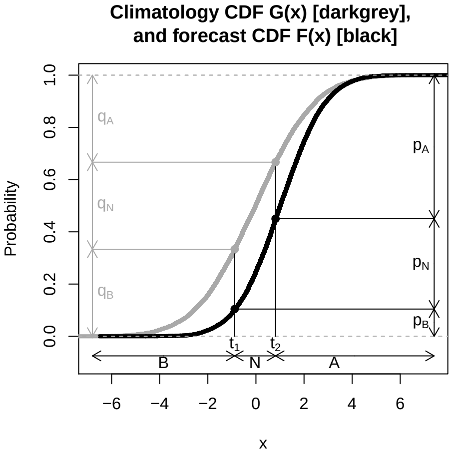
- the forecast at each point is a *distribution*
- How should we assign *colours* to *distributions*?

The quantiles of a climatology



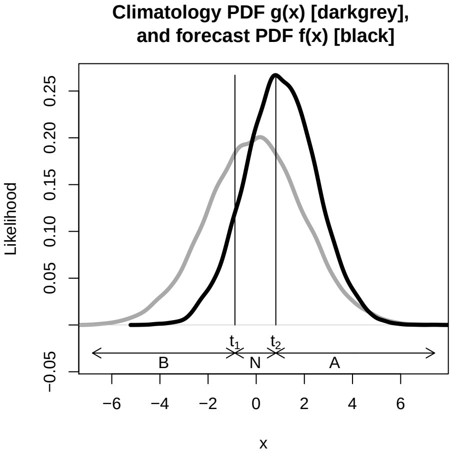
- (tercile) categories B, N and A are *observed* with equal frequency

Continuous distributions (CDFs)



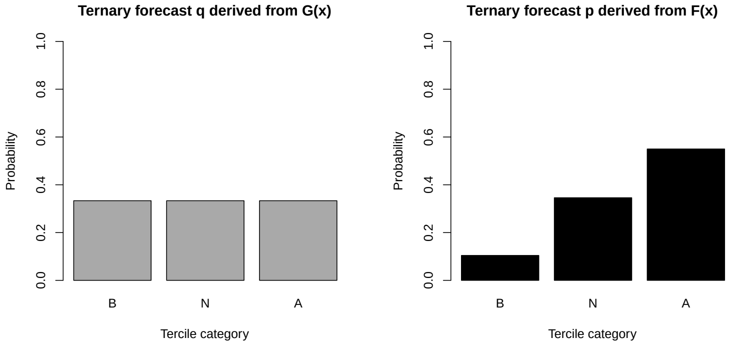
- climatology: $G(x)$, forecast: $F(x)$

Continuous distributions (PDFs)



- climatology: $g(x)$, forecast: $f(x)$

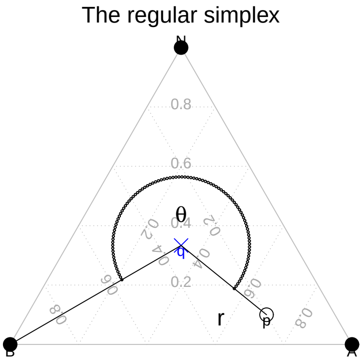
Ternary climatology q and forecast p



- ternary climatology: $\mathbf{q}$ (for *terciles* $\Rightarrow \mathbf{q} = (1/3, 1/3, 1/3)$)
- ternary forecast: $\mathbf{p} = (p_B, p_N, p_A)$

Outline

Barycentric coordinates



- every ternary forecast is a point in the triangle, including
- the climatology **q**
- the observed state **o**

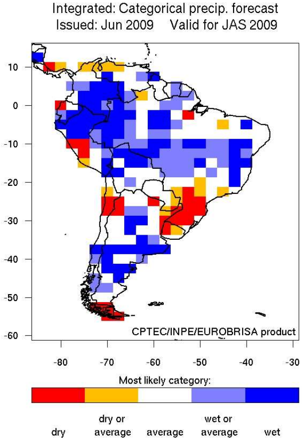
Current visualisation methods

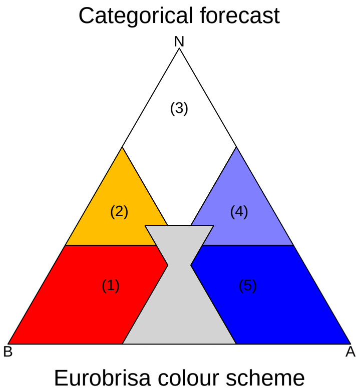
Usually based on a discretisation of ternary forecast space. For example:

- 1 (Dry): ($p_B > 2/5$ and $p_N < 1/3$ and $p_A < 1/3$).
- 2 (Dry or normal): ($p_B > 1/3$ and $p_N > 2/5$) or ($p_B > 2/5$ and $p_N > 1/3$).
- 3 (Normal): ($p_B < 1/3$ and $p_N > 2/5$ and $p_A < 1/3$).
- 4 (Wet or normal): ($p_N > 1/3$ and $p_A > 2/5$) or ($p_N > 2/5$ and $p_A > 1/3$).
- 5 (Wet): ($p_B < 1/3$ and $p_N < 1/3$ and $p_A > 2/5$).

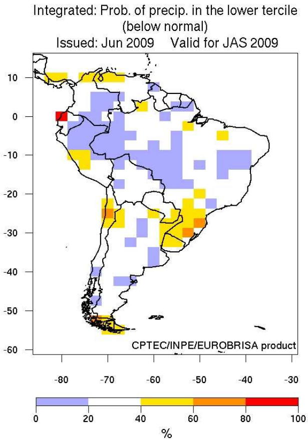
Algebraic descriptions make my head hurt.

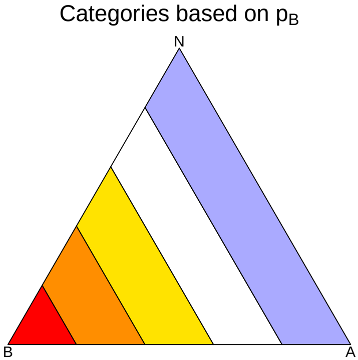
Current visualisation methods (EUROBRISA categorical)



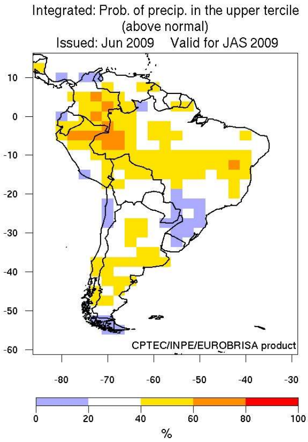


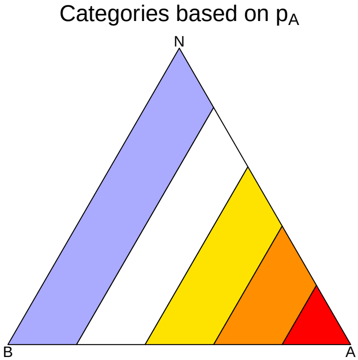
Current visualisation methods (EUROBRISA lower)



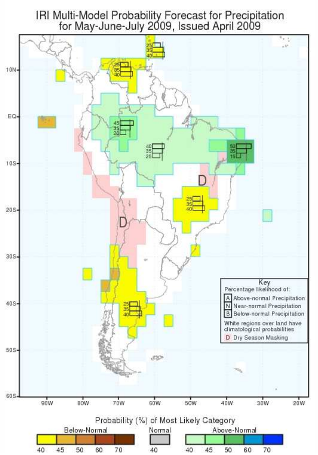


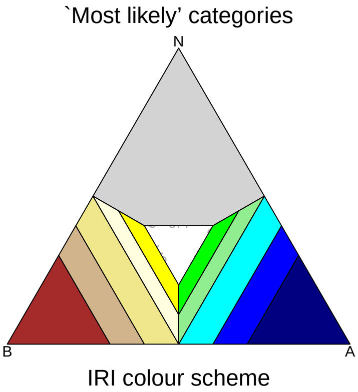
Current visualisation methods (EUROBRISA upper)



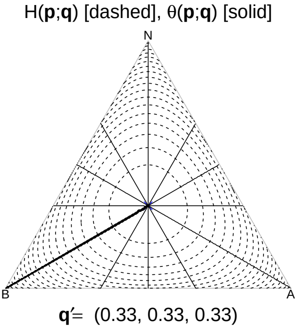


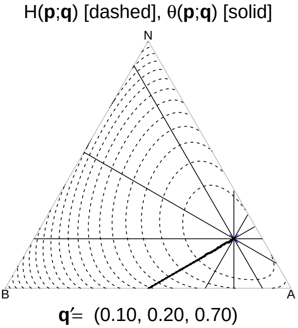
Current visualisation methods (IRI)





A continuum of colours in forecast space

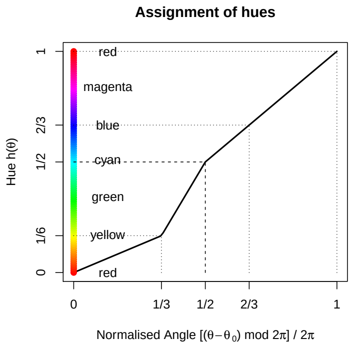


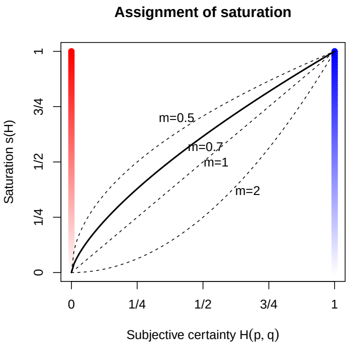


$$H(\mathbf{p}) = \frac{1}{\log \max_i q_i^{-1}} \sum_{i \in \{B,N,A\}} p_i \log \frac{p_i}{q_i}$$

- $H(\mathbf{p};\mathbf{q})$ is a measure of the *subjective certainty* in a forecast

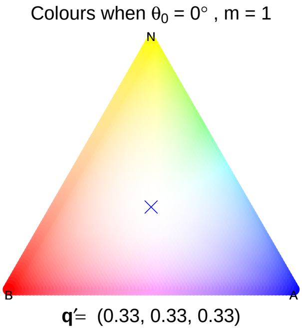
Assignment of colours





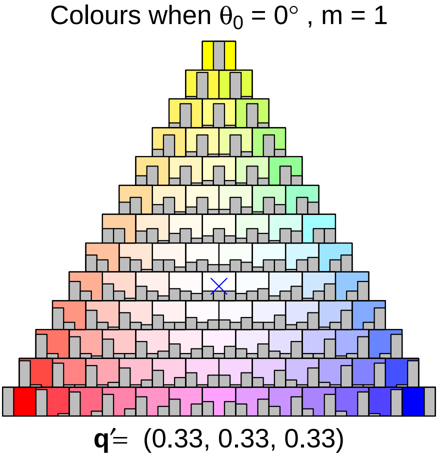
Designed to be intuitive and colour-blind friendly!

Our proposed colour scheme

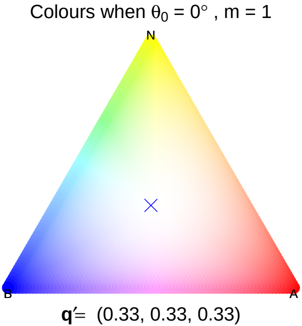


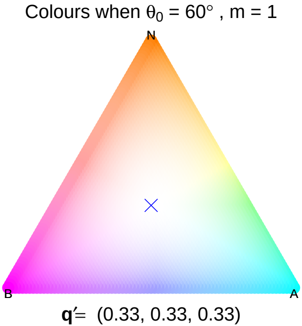
- use HSV (hue–saturation–value) colour space
- hue $\propto \theta(\mathbf{p})$
- saturation $\propto [H(\mathbf{p})]^m$ (usually $m = 1$)

Mapping colours ⇔ barplots



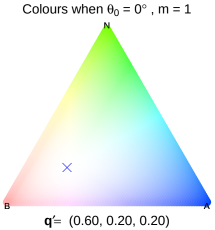
Alternative colour schemes

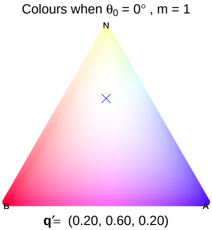


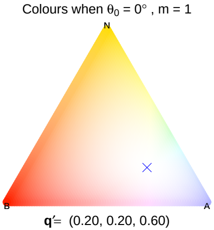


left: $p_B \leftrightarrows p_A$, right: $RGB \leftrightarrows CMY$

Climatology **q** need not be defined by terciles

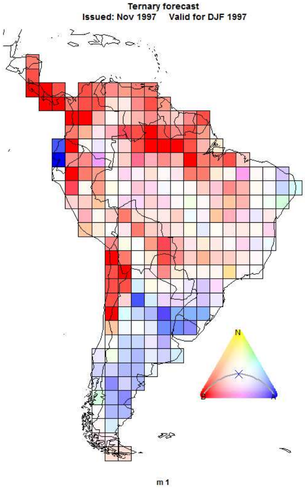






In all cases: $\mathbf{p} = \mathbf{q}$ is set to *white*

An example: seasonal forecasting



ECMWF probabilistic seasonal forecast

Alternative viewpoints: summary

Forecast **p**, climatology **q**, observation **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

How good is a probabilistic forecasting system?

Compare forecasts $\{\mathbf{p}\}$ with subsequent observations $\{\mathbf{o}\}$.
Often, observations are treated as 'certain', so that

$$\mathbf{o} \in \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

Conditional averages defined as follows:

$$\bar{\mathbf{o}}|\mathbf{p}, \text{ mean observation associated with given forecast}$$

$$\bar{\mathbf{p}}|\mathbf{o}, \text{ mean forecast associated with given observation}$$

Quadratic score functions

Consider skill measures which are *quadratic forms*:

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

for some matrix L.

We *want small scores*.

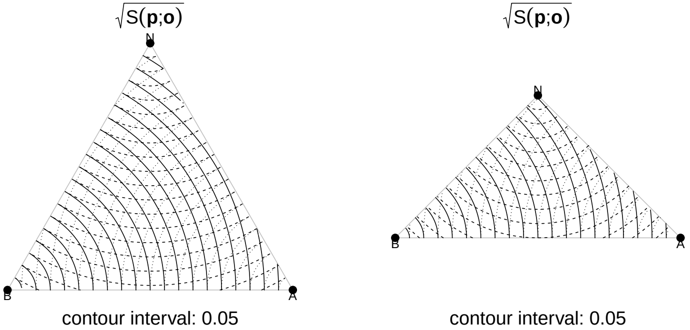
Note that

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

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

In this space, the score is the *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 $\langle \mathbf{o} \rangle = \langle \bar{\mathbf{o}} | \mathbf{p} \rangle = \mathbf{q}$
Discretise forecast space into bins $\{\mathbf{p}_k\}$
It follows that:

$$\begin{array}{ccccccc} \left\langle \|\mathbf{p}_k - \mathbf{o}\|^2 \right\rangle & = & \left\langle \|\mathbf{q} - \mathbf{o}\|^2 \right\rangle & + & \left\langle \|\mathbf{p}_k - \bar{\mathbf{o}}|\mathbf{p}_k\|^2 \right\rangle & - & \left\langle \|\mathbf{q} - \bar{\mathbf{o}}|\mathbf{p}_k\|^2 \right\rangle \\ \textit{score} & = & \textit{uncertainty} & + & \textit{reliability} & - & \textit{resolution} \end{array}$$

A better (smaller) score may be obtained if *raw* forecasts $\{\mathbf{p}\}$ are replaced by *calibrated* forecasts $\{\hat{\mathbf{p}}\}$.

The Brier Score and the Ranked Probability Score

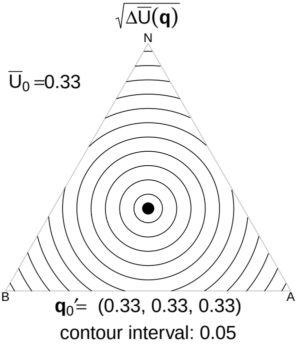
$$BS(\mathbf{p};\mathbf{o}) = \frac{1}{2}\left[(p_B - o_B)^2 + (p_N - o_N)^2 + (p_A - o_A)^2\right]$$

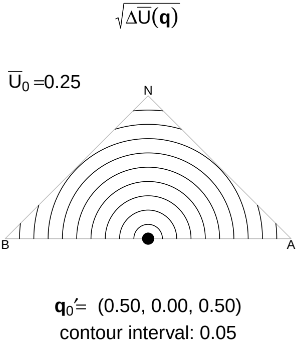
- in $\mathbb{R}^3$: BS = mean–square–distance difference between PDFs
- in $\mathbb{R}^2$: BS = mean–square–distance in equilateral triangle

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

- in $\mathbb{R}^3$: RPS = mean–square–distance difference between CDFs
- in $\mathbb{R}^2$: RPS = mean–square–distance in right–angled triangle

(reduction in) uncertainty is square of distance

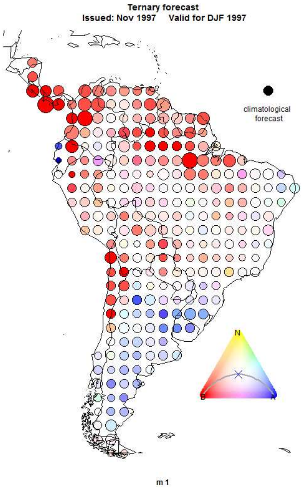




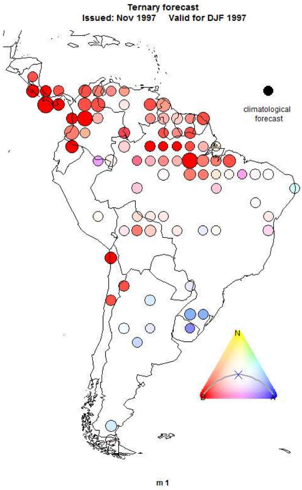
$$\bar{U} = \bar{U}_0 - \Delta\bar{U}$$

Brier score (left) ; Ranked Probability Score (right)

Set radius $\propto 1/BS$



Set radius $\propto 1/BS$ with masking



Outline

Calibration

In suitable 2D space (simplex),
Raw forecasts $\{\mathbf{p}\}$ have score

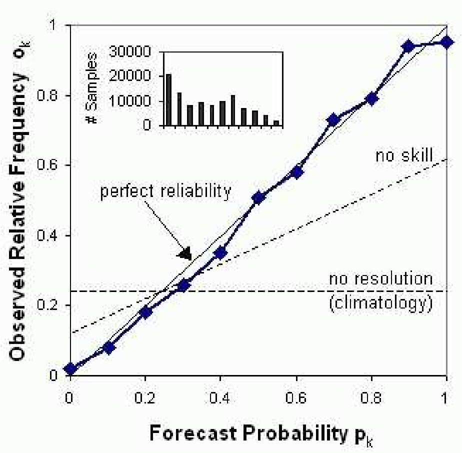
$$S(\{\mathbf{p};\mathbf{o}\}) = \langle \|\mathbf{p}-\mathbf{o}\|^2 \rangle$$

Calibrated forecasts $\{\hat{\mathbf{p}}\}$ have score

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

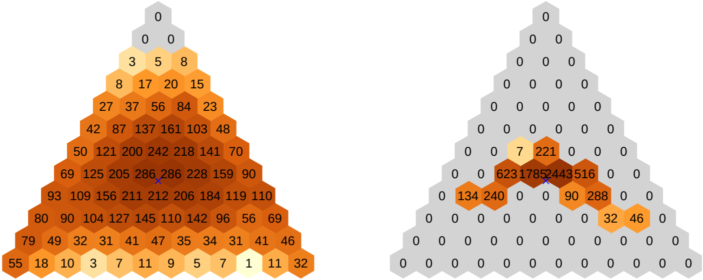
Calibrated forecasts $\hat{\mathbf{p}}$ chosen to minimise score.
Calculation of $\hat{\mathbf{p}}$ is a standard *Generalised Linear Model* problem, easily soluble in `R`.

Traditional Reliability diagram (binary forecast)



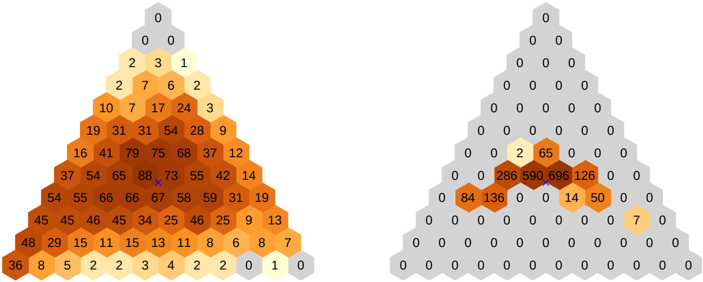
This is a plot of $\bar{\mathbf{o}}|\mathbf{p}_k$ against $\mathbf{p}_k$.
Sharpness diagram: distribution of forecasts $\{\mathbf{p}\}$

Ternary Sharpness (all forecasts)



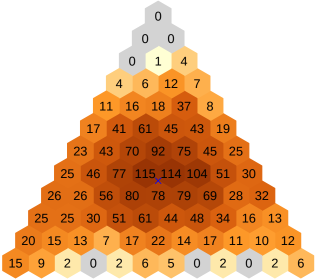
left: raw $\{\mathbf{p}\}$, right: calibrated $\{\hat{\mathbf{p}}\}$

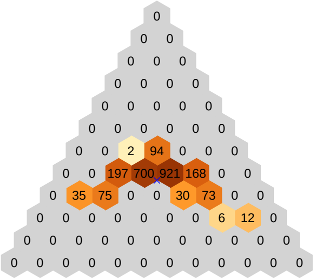
Ternary Sharpness (conditional on $\mathbf{o} = (1, 0, 0)$)



left: raw $\{\mathbf{p}\}$, right: calibrated $\{\hat{\mathbf{p}}\}$

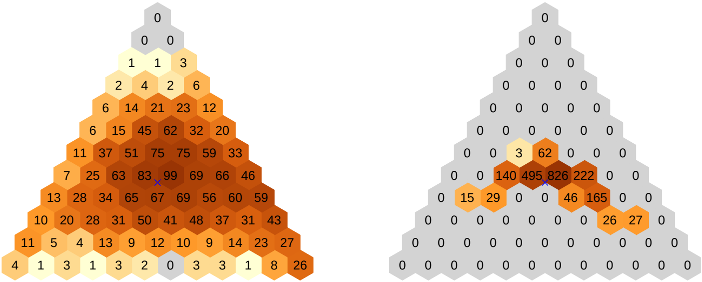
Ternary Sharpness (conditional on $\mathbf{o} = (0, 1, 0)$)





left: raw $\{\mathbf{p}\}$, right: calibrated $\{\hat{\mathbf{p}}\}$

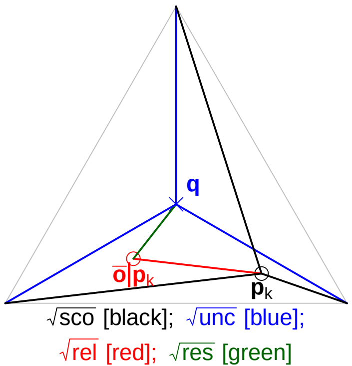
Ternary Sharpness (conditional on $\mathbf{o} = (0, 0, 1)$)



left: raw $\{\mathbf{p}\}$, right: calibrated $\{\hat{\mathbf{p}}\}$

Ternary reliability diagrams

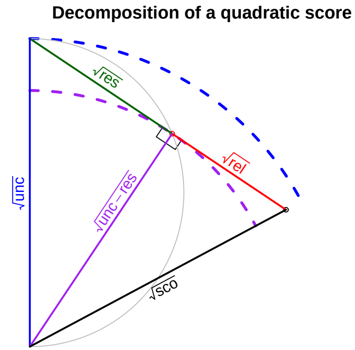
Decomposition of a quadratic score



$$\left\langle \|\mathbf{p}_k - \mathbf{o}\|^2 \right\rangle = \left\langle \|\mathbf{q} - \mathbf{o}\|^2 \right\rangle + \left\langle \|\mathbf{p}_k - \bar{\mathbf{o}}|\mathbf{p}_k\|^2 \right\rangle - \left\langle \|\mathbf{q} - \bar{\mathbf{o}}|\mathbf{p}_k\|^2 \right\rangle$$

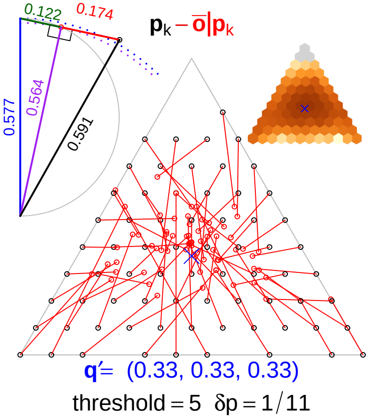
$$\textit{score} = \textit{uncertainty} + \textit{reliability} - \textit{resolution}$$

Ternary reliability diagrams



$$\begin{array}{ccccccccc} \left\langle \|\mathbf{p}_k - \mathbf{o}\|^2 \right\rangle & = & \left\langle \|\mathbf{q} - \mathbf{o}\|^2 \right\rangle & + & \left\langle \|\mathbf{p}_k - \bar{\mathbf{o}}|\mathbf{p}_k\|^2 \right\rangle & - & \left\langle \|\mathbf{q} - \bar{\mathbf{o}}|\mathbf{p}_k\|^2 \right\rangle \\ \textit{score} & = & \textit{uncertainty} & + & \textit{reliability} & - & \textit{resolution} \end{array}$$

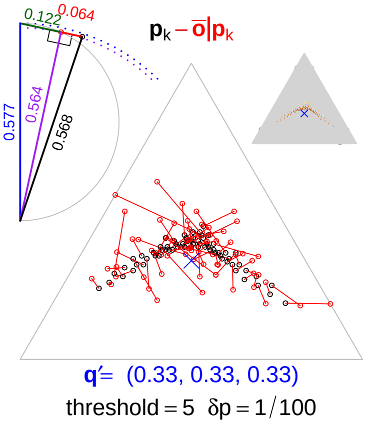
Ternary reliability diagrams: raw forecasts



$$\left\langle \|\mathbf{p}_k - \mathbf{o}\|^2 \right\rangle = \left\langle \|\mathbf{q} - \mathbf{o}\|^2 \right\rangle + \left\langle \|\mathbf{p}_k - \bar{\mathbf{o}}|\mathbf{p}_k\|^2 \right\rangle - \left\langle \|\mathbf{q} - \bar{\mathbf{o}}|\mathbf{p}_k\|^2 \right\rangle$$

score = *uncertainty* + *reliability* − *resolution*

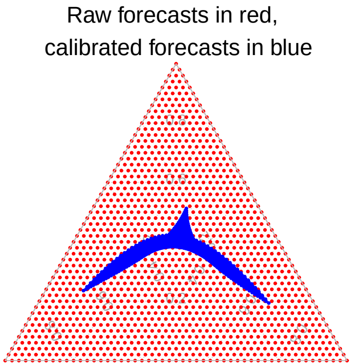
Ternary reliability diagrams: calibrated forecasts



$$\langle \|\mathbf{p}_k - \mathbf{o}\|^2 \rangle = \langle \|\mathbf{q} - \mathbf{o}\|^2 \rangle + \langle \|\mathbf{p}_k - \bar{\mathbf{o}}|\mathbf{p}_k\|^2 \rangle - \langle \|\mathbf{q} - \bar{\mathbf{o}}|\mathbf{p}_k\|^2 \rangle$$

score = *uncertainty* + *reliability* − *resolution*

Nonlinear calibration



Outline

Conclusions

- assign unique colour to each ternary forecast
- barycentric coordinates aid interpretation
- white $\Leftrightarrow \mathbf{p} \approx \mathbf{q}$
- 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 $\sim N(\mu, \sigma^2)$