

# On the visualization, verification and recalibration of ternary probabilistic forecasts

Tim E. Jupp, Rachel Lowe, Caio A. S. Coelho and David B. Stephenson

*Phil. Trans. R. Soc. A* 2012 **370**, 1100-1120

doi: 10.1098/rsta.2011.0350

---

## References

**This article cites 16 articles**

<http://rsta.royalsocietypublishing.org/content/370/1962/1100.full.html#ref-list-1>

**Article cited in:**

<http://rsta.royalsocietypublishing.org/content/370/1962/1100.full.html#related-urls>

## Subject collections

Articles on similar topics can be found in the following collections

[climatology](#) (95 articles)

## Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

# On the visualization, verification and recalibration of ternary probabilistic forecasts

BY TIM E. JUPP<sup>1,\*</sup>, RACHEL LOWE<sup>2</sup>, CAIO A. S. COELHO<sup>3</sup>  
AND DAVID B. STEPHENSON<sup>1</sup>

<sup>1</sup>*Mathematics Research Institute, University of Exeter, Exeter, UK*

<sup>2</sup>*Abdus Salam International Centre for Theoretical Physics, Trieste, Italy*

<sup>3</sup>*Centro de Previsão de Tempo e Estudos Climáticos, Instituto Nacional de Pesquisas Espaciais, Cachoeira Paulista, 12630-000 SP, Brazil*

We develop a graphical interpretation of ternary probabilistic forecasts in which forecasts and observations are regarded as points inside a triangle. Within the triangle, we define a continuous colour palette in which hue and colour saturation are defined with reference to the observed climatology. In contrast to current methods, forecast maps created with this colour scheme convey all of the information present in each ternary forecast. The geometrical interpretation is then extended to verification under quadratic scoring rules (of which the Brier score and the ranked probability score are well-known examples). Each scoring rule defines an associated triangle in which the square roots of the *score*, the *reliability*, the *uncertainty* and the *resolution* all have natural interpretations as root mean square distances. This leads to our proposal for a *ternary reliability diagram* in which data relating to verification and calibration can be summarized. We illustrate these ideas with data relating to seasonal forecasting of precipitation in South America, including an example of nonlinear forecast calibration. Codes implementing these ideas have been produced using the statistical software package R and are available from the authors.

**Keywords:** map; climate; colour

## 1. Introduction

Forecasts are often given in probabilistic terms. For example: ‘There is a 60 per cent chance that rainfall before 12.00 tomorrow will exceed 15 mm’; ‘The odds for the football match are 6-to-1 against a home win and even money for a draw’; ‘The climate model predicts that mean daytime temperature in summer 2100 will be normally distributed with mean  $\mu$  and variance  $\sigma^2$ ’. In each of these examples, the forecast consists of a probability distribution. The sample spaces for the observable events of interest consist of two elements (exceed/not exceed), three elements (home win/draw/away win) and the real line  $\mathbb{R}$ . Consequently, these examples relate to binary, ternary and continuous probabilistic forecasts, respectively.

\*Author for correspondence ([t.e.jupp@ex.ac.uk](mailto:t.e.jupp@ex.ac.uk)).

One contribution of 13 to a Theme Issue ‘Climate predictions: the influence of nonlinearity and randomness’.

This paper is concerned with two types of visualization. In physical space, visualization constitutes colour maps of probabilistic forecasts over a geographical region. In probability space, visualization will constitute a geometrical representation of forecasts and observations in our proposed ternary reliability diagram.

For binary forecasts, visualization tools such as the reliability diagram, the sharpness diagram and the relative operating characteristic (ROC) curve are well known [1]. Our aim in this paper is to develop analogues of these ideas for ternary forecasts and to develop a geometrical intuition for verification and recalibration.

The structure of this paper is as follows. Section 2 contains a discussion of probabilistic forecasting, and illustrates how a continuous forecast distribution can be identified with a ternary forecast. Section 3 then introduces the idea of representing each ternary forecast as a point in barycentric coordinates, and hence assigning to it a unique colour. In §4, the quality of a forecasting system is quantified using quadratic scoring rules. This is then used in §5 to propose an algorithm for recalibrating probabilistic forecasts which is illustrated with seasonal forecasting data. Conclusions are presented in §6.

## 2. Probabilistic forecasting

The generic situation to be considered is as follows: a forecasting system (perhaps a suite of climate models with different initial conditions) can be used to produce probabilistic forecasts (that is, probability distributions) for some variable of interest at a number of points within a geographical area. The subsequent challenge is to display in map form as much as possible of the information contained in this spatial array of probability distributions. The most natural thing to do, perhaps, is to associate with each forecast distribution a set of scalars (such as the distribution's moments or a selection of its quantiles) and produce a separate map for each scalar. The disadvantage of this procedure is that multiple maps are needed to convey the information contained in the forecast distributions. If information on the skill of the forecasting system were available, yet another map would be required.

Consider probabilistic forecasting of a scalar climate variable  $x$ , such as temperature or precipitation. The forecasting system produces a cumulative distribution function (CDF)  $F(x)$  at each spatial location (figure 1). The CDF expresses the forecast probability that the variable of interest will not exceed  $x$ . The forecast distribution  $F(x)$  can be interpreted as the forecasting system's state of knowledge about the future value of the variable  $x$ . In contrast to a deterministic forecast  $x_f$  (which is a scalar), a probabilistic forecast  $F(x)$  is (in principle at least) an analytic function and hence an infinite-dimensional object. For this reason, it is not immediately obvious how to visualize spatial probabilistic forecasts.

Often, the variable  $x$  has associated with it an observed climatology represented by the CDF  $G(x)$  (figure 1). The climatology can be interpreted as the historical distribution of the observable quantity. Typically, a climatology is calculated by aggregating data from an observational record over at least the past 30 years.

It is important to note that both the climatology  $G(x)$  and the forecast  $F(x)$  are continuous probability distributions and so it is quite possible for the forecasting system to issue the climatology  $G(x)$  as its forecast. This is a perfectly valid

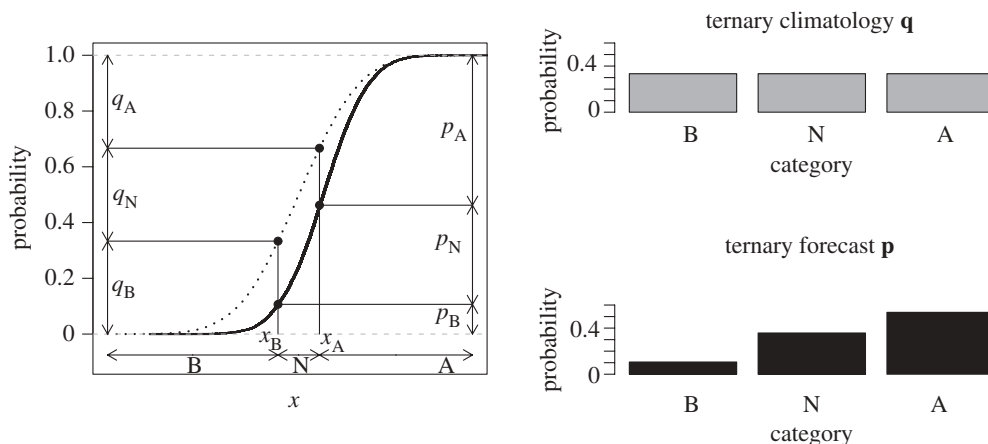


Figure 1. Continuous distribution functions  $F(x)$  (forecast, solid line) and  $G(x)$  (climatology, dotted line) can be associated with ternary forecasts  $\mathbf{p}$  and  $\mathbf{q}$  by defining three categories B, N and A—‘below-’, ‘near-’ and ‘above-’normal (equations (2.1) and (2.2)).

probabilistic forecast but not, perhaps, a very useful one. It is usually hoped that the forecasting system’s state of knowledge about the future extends beyond knowledge of the climatology alone. For this reason, it will be helpful to view the climatology as a benchmark distribution with which all other forecasts should be compared. Thus, in the discussion below, a colour will be assigned to the forecast  $F(x)$  by considering the ‘distance’ between the forecast  $F(x)$  and the climatology  $G(x)$ .

### (a) Ternary probabilistic forecasts

In order to reduce the dimensionality of the problem, it is useful to project the continuous distributions  $G(x)$  and  $F(x)$  onto ternary distributions  $\mathbf{q}' = (q_B, q_N, q_A)$  and  $\mathbf{p}' = (p_B, p_N, p_A)$  (figure 1). We adopt the convention that vectors are column vectors and that  $'$  denotes a transpose. The real line is divided into three ordered categories  $B = (-\infty, x_B]$ ,  $N = [x_B, x_A]$  and  $A = [x_A, \infty)$  whose labels have the following rationale: B, ‘below-normal’; N, ‘near-normal’; and A, ‘above-normal’.

The ternary distributions  $\mathbf{q}(G)$  and  $\mathbf{p}(F)$  encode the probabilities with which the variable is forecast to lie in the categories B, N and A (figure 1), and so all of their elements are non-negative with  $q_B + q_N + q_A = p_B + p_N + p_A = 1$ .

For given quantiles  $x_B$  and  $x_A$ , it follows that the ternary climatology  $\mathbf{q}$  and ternary forecast  $\mathbf{p}$  are given by

$$\mathbf{q}' = (q_B, q_N, q_A) = (G(x_B), G(x_A) - G(x_B), 1 - G(x_A)) \left. \vphantom{\mathbf{q}'} \right\} \quad (2.1)$$

and

$$\mathbf{p}' = (p_B, p_N, p_A) = (F(x_B), F(x_A) - F(x_B), 1 - F(x_A)). \left. \vphantom{\mathbf{p}'} \right\}$$

An equivalent interpretation is that for a specified ternary climatology  $\mathbf{q}$  the categories B, N and A are defined by the quantiles

$$x_B = G^{-1}(q_B) \quad \text{and} \quad x_A = G^{-1}(q_B + q_N). \quad (2.2)$$

A natural choice for the ternary climatology is the uniform distribution  $\mathbf{q}' = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ . In this case—which we shall regard as the default—the categories B, N and A are defined by the terciles of the climatology  $G(x)$  and observations over a long period would be expected to lie with equal frequency in each of the three categories. It should be stressed, however, that any choice of quantiles can be made when defining the ternary climatology. For example, the choice  $\mathbf{q}' = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$  would imply that categories B and A had been chosen to be the lower and upper quartiles of the climatology.

### 3. Barycentric coordinates

In this section, a geometrical representation of a ternary forecast  $\mathbf{p}$ , ternary climatology  $\mathbf{q}$  and ternary observation  $\mathbf{o}$  is considered.

#### (a) Scoring functions

Scoring functions quantify the past skill of a forecasting system. The idea is to compare a ternary forecast  $\mathbf{p}$  with the corresponding ternary observation  $\mathbf{o}$  made after the event. Since the categories are exclusive and complete, it follows that a ternary observation  $\mathbf{o}$  can take one of three values  $\mathbf{o}'_B = (1, 0, 0)$ ,  $\mathbf{o}'_N = (0, 1, 0)$  and  $\mathbf{o}'_A = (0, 0, 1)$  according to whether the observable is found to lie in category B, N or A.

The skill of a forecasting system can be quantified for a set of forecast–observation pairs by a score function  $S$ . The score is a measure of the difference between forecasts and observations. It follows that the lower the score, the more skilful the forecast.

In this paper, attention is restricted to scores defined by quadratic forms

$$S = \overline{(\mathbf{p} - \mathbf{o})' L L (\mathbf{p} - \mathbf{o})}, \quad (3.1)$$

where the 3-by-3 matrix  $L$  defines the particular scoring rule being used (appendix Aa),  $L'L$  is assumed to be positive definite and the overbar denotes an average over all forecast–observation pairs. The Brier score (appendix Aa) and the ranked probability score (appendix Ab) are well-known examples of quadratic scores [2–6].

The scoring matrix  $L$  can be used to define a 2-by-3 matrix  $\hat{M}$  (equation (A 4)) which maps a ternary forecast  $\mathbf{p} \in \mathbb{R}^3$  to a corresponding point  $\mathbf{P} = \hat{M}\mathbf{p} \in \mathbb{R}^2$  in the plane. This transformation maps ternary forecasts to points within a triangle and so each scoring matrix  $L$  induces a corresponding triangle in  $\mathbb{R}^2$ . In this paper, the matrix  $\hat{M}$  is defined so that a score  $S$  corresponds to a (mean) squared distance between forecasts  $\mathbf{P}$  and observations  $\mathbf{O}$  (considered as points inside an appropriate triangle):

$$S = \overline{\|\mathbf{P} - \mathbf{O}\|^2}. \quad (3.2)$$

The corners of this triangle are the points  $\mathbf{O}_B = \hat{M}\mathbf{o}_B$ ,  $\mathbf{O}_N = \hat{M}\mathbf{o}_N$  and  $\mathbf{O}_A = \hat{M}\mathbf{o}_A$  associated with the three possible values of a ternary observation. The point  $\mathbf{Q} = \hat{M}\mathbf{q}$  associated with the climatology lies within the triangle. The default choice for the scoring matrix is that associated with the Brier score, in which case the associated triangle is an equilateral triangle with unit sides. In the case

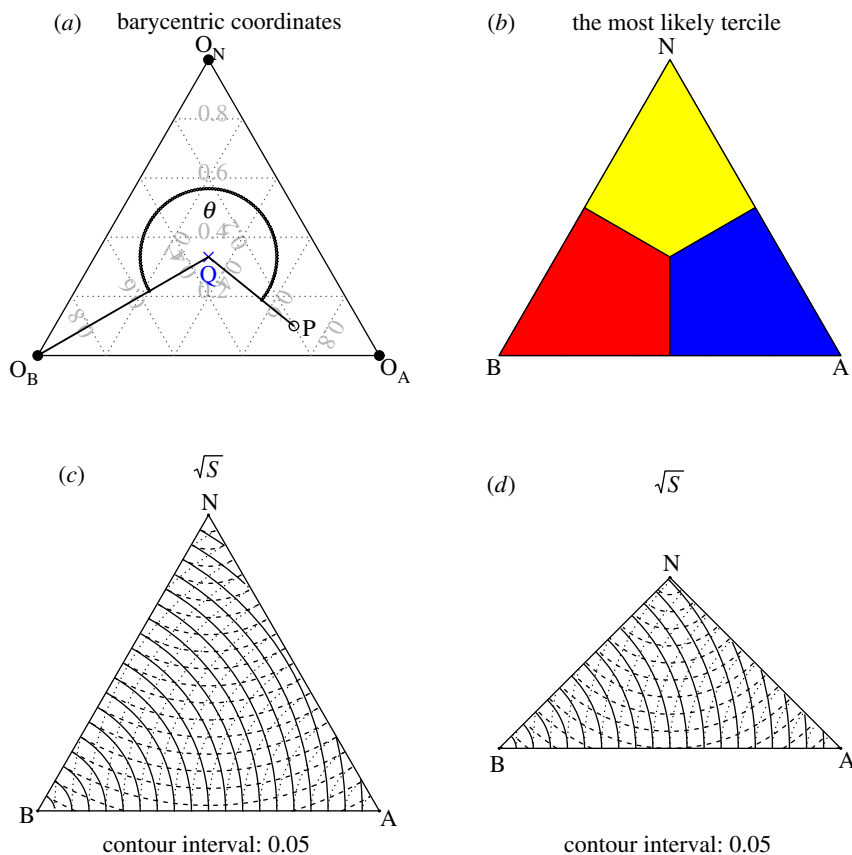


Figure 2. Visualization of ternary forecasts and observations in barycentric coordinates using the categories B, N and A defined in figure 1. (a) Observations  $\mathbf{O}$  can take one of three values and lie at the corners. The angle  $\theta$  can be used to compare an arbitrary forecast  $\mathbf{P}$  with the climatology  $\mathbf{Q}$ . (b) A simple way to assign colours to ternary forecasts is to consider the most likely category. (c) For the Brier score, the root-score  $\sqrt{S}$  corresponds to distance in an equilateral triangle. Contours of  $\sqrt{S}$  are shown for  $\mathbf{o} = \mathbf{o}_B$  (solid),  $\mathbf{o} = \mathbf{o}_N$  (dashed),  $\mathbf{o} = \mathbf{o}_A$  (dotted). (d) As (c) but for the right-angled triangle induced by the ranked probability score.

of the default climatology  $\mathbf{q}' = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ , the point  $\mathbf{Q}$  lies at the centre of this triangle (figure 2a). This sort of visualization receives different names in different disciplines. In mathematics, it is known as a plot in barycentric coordinates while in applied sciences it is known as a ternary phase diagram.

Barycentric coordinates yield an intuitive geometrical interpretation of ternary forecasts. As well as allowing scores to be visualized as squared distances they allow any scheme for colour assignment to be visualized as a triangular colour palette. For example, figure 2b illustrates the algorithm which assigns colours according to which of the three possible outcomes is considered in the forecast to be the most probable. Here and subsequently, it is assumed for illustration that the forecast variable  $x$  is precipitation. It therefore seems sensible to assign the ‘dry’ colour red to category B and the ‘wet’ colour blue to category A. Note that both the ternary forecasts  $\mathbf{p}' = (1, 0, 0)$  and  $\mathbf{p}' = (0.34, 0.33, 0.33)$  would be

assigned the colour red in figure 2 even though they differ greatly. In the former case, the forecasting system is *certain* that  $x$  will lie in category B while in the latter case, the forecast is barely different from the climatology  $\mathbf{q}' = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ .

Since scores are squares of distances within the triangle, it is helpful to consider the root-score  $\sqrt{S}$ . Figure 2c shows contours of constant root-score in the case of the Brier score, when the observation  $\mathbf{o}$  takes each of its three possible values. These contours show the set of forecasts which are equally good (in terms of score) as predictions of the subsequently observed value. Figure 2d shows contours of constant root-score in the case of the ranked probability score, for which the induced triangle is a right-angled triangle with sides  $1/\sqrt{2}$ ,  $1/\sqrt{2}$  and 1 (appendix Ab).

### (b) Current colour schemes for ternary forecasts

Current methods for assigning colours to ternary forecasts discretize barycentric coordinates into a finite number of regions, meaning that information present in the forecast is not present in the assigned colour [7]. For example, consider the ‘most likely tercile’ scheme of figure 2b used for displaying precipitation forecasts produced in the EURO-Brazilian Initiative for improving South American seasonal forecasts (EUROBRISA) project.<sup>1</sup> Precipitation forecasts produced in EUROBRISA are Gaussian distributions obtained from a calibration and combination procedure known as forecast assimilation [8]. These ‘integrated’ forecasts are a combination of four coupled ocean–atmosphere general circulation models (European Centre for Medium Range Weather Forecasting (ECMWF) [9], UK Met Office [10], Météo France [11] and Centro de Previsão de Tempo e Estudos Climáticos (CPTEC) [12]) and an empirical model that uses Pacific and Atlantic sea surface temperatures as predictors for South American precipitation [13].

A feature of current visualization methods is that the algorithms by which colours are assigned tend to be described algebraically. In the example considered here, colours are assigned to forecasts based on the following algebraic definitions of five regions of forecast space:

1. (dry): ( $p_B > \frac{2}{5}$  and  $p_N < \frac{1}{3}$  and  $p_A < \frac{1}{3}$ ).
2. (dry or normal): ( $p_B > \frac{1}{3}$  and  $p_N > \frac{2}{5}$ ) or ( $p_B > \frac{2}{5}$  and  $p_N > \frac{1}{3}$ ).
3. (normal): ( $p_B < \frac{1}{3}$  and  $p_N > \frac{2}{5}$  and  $p_A < \frac{1}{3}$ ).
4. (wet or normal): ( $p_N > \frac{1}{3}$  and  $p_A > \frac{2}{5}$ ) or ( $p_N > \frac{2}{5}$  and  $p_A > \frac{1}{3}$ ).
5. (wet): ( $p_B < \frac{1}{3}$  and  $p_N < \frac{1}{3}$  and  $p_A > \frac{2}{5}$ ).

The use of barycentric coordinates allows the meaning of these definitions to be visualized easily (figure 3). It also reveals something that is not immediately obvious from the algebra. There is a region in barycentric coordinates (here coloured grey) that is not included in these definitions. This region of ternary forecast space—at the base of the triangle—corresponds to ternary forecasts in which category N is assigned a low probability, but the outlying categories B and A are assigned relatively high probability. Such forecasts can arise when the forecast distribution  $F(x)$  has variance much greater than that of the climatology  $G(x)$  (§3e).

<sup>1</sup><http://eurobrisa.cptec.inpe.br>.

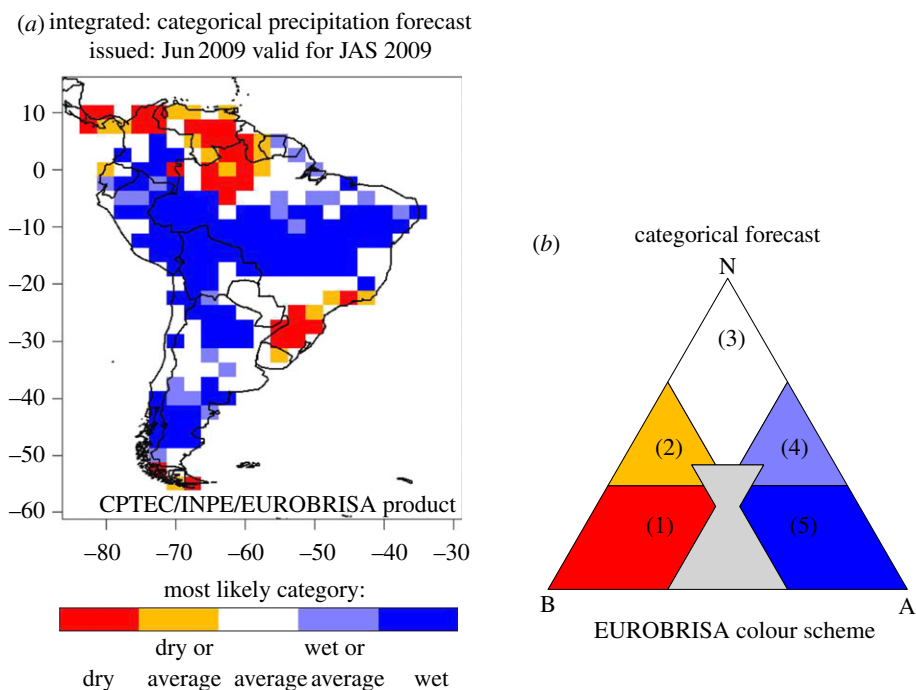


Figure 3. (a) Example of a current ternary forecast map (from EUROBRISA). (b) Plotting the colour palette in barycentric coordinates is a useful aid to interpretation.

In summary, many current methods for assigning colours to forecasts lose information by discretizing ternary forecast space. The same colour is assigned to more than one ternary forecast and the colours do not convey a sense of how much the forecast differs from climatology. A ternary forecast close to climatology provides little gain in information—the forecast has not told us much that we did not already know. On the other hand, a ternary forecast far from climatology provides a large gain in information. In such a case, the forecasting system assigns high probability to the variable lying in one particular category and not the others.

The aim now, therefore, is to assign a continuum of colours to ternary forecasts, viewed as points in barycentric coordinates. This will retain all information present in the ternary forecast and so it will be possible to identify a ternary forecast uniquely from its assigned colour. The choice of colour assignment will, in particular, take account of the information gain in each ternary forecast.

### (c) Comparing forecasts with the climatology

Probabilistic forecasts are regarded here as measures of belief and so a measure based on entropy [14] is preferred. A sensible measure of information gain  $E(\mathbf{p}; \mathbf{q})$  is

$$E(\mathbf{p}; \mathbf{q}) = \left[ \frac{1}{\log \max_i q_i^{-1}} \right] \sum_{i \in \{B, N, A\}} p_i \log \frac{p_i}{q_i}. \quad (3.3)$$



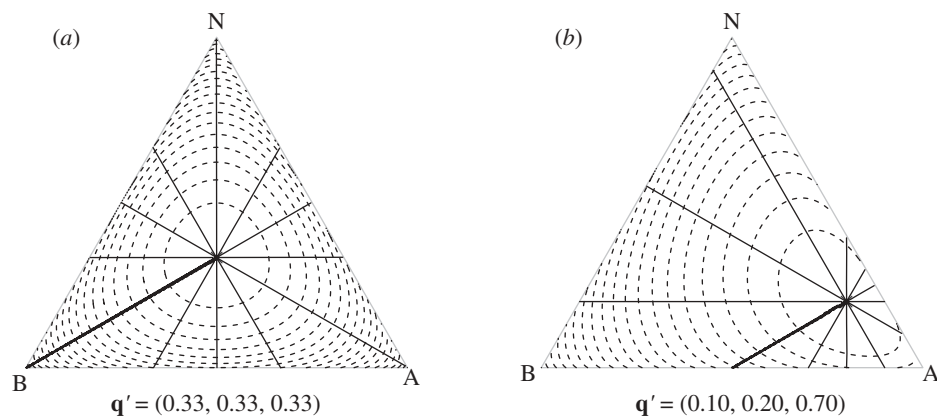


Figure 4. (a) Proposed coordinate system in the case when  $\mathbf{q}' = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ . Information gain  $E(\mathbf{p}; \mathbf{q})$  (equation (3.3); dashed lines) ranges from 0 (at the centre) to 1 (at the corners). Dominant category  $\theta(\mathbf{p}; \mathbf{q})/2\pi$  (solid lines) ranges from 0 to 1 moving clockwise from thick line. (b) As (a) but with  $\mathbf{q}' = (0.1, 0.2, 0.7)$ .

This can be interpreted as a scaled version of the Kullback–Leibler divergence between  $\mathbf{p}$  and  $\mathbf{q}$ . Contours of constant  $E(\mathbf{p}; \mathbf{q})$  are plotted in the unit triangle in figure 4. Note that the climatology  $\mathbf{q}$  corresponds to an information gain of 0 ( $E(\mathbf{q}; \mathbf{q}) = 0$ ) and that the corner furthest from the climatology corresponds to an information gain of 1 ( $E(\mathbf{o}; \mathbf{q}) = 1$  for some  $\mathbf{o}$ ).

Motivated by the idea of assigning colours according to the most likely tercile, the continuous angular measure  $\theta(\mathbf{p}; \mathbf{q})$  will be referred to as the dominant category. This measures the angle in barycentric coordinates of the forecast  $\mathbf{p}$  with respect to an origin at the climatology  $\mathbf{q}$ . It follows that the information gain  $E(\mathbf{p}; \mathbf{q})$  and the dominant category  $\theta(\mathbf{p}; \mathbf{q})$  define an alternative coordinate system for ternary forecast space (figure 2).

#### (d) A new colour scheme for ternary forecasts

In the red–green–blue (RGB) representation of colour (used in colour televisions), colours are represented by sets of three numbers in the range  $[0, 1]$  corresponding to the brightness of the three primary colours. Thus,  $\text{RGB} = (1, 0, 0)$  is bright red,  $\text{RGB} = (0, 1, 0)$  is bright green, and so on. In this paper, colours are assigned using the more intuitive hue–saturation–value (HSV) representation. Geometrically, the RGB system defines a unit cube in colour space in which shades of grey occur on a grey-line running from the black corner  $\text{RGB} = (0, 0, 0)$  to the white corner  $\text{RGB} = (1, 1, 1)$ . The HSV system describes colours as points in a cylindrical coordinate system with this grey line as its axis. In this system, value  $\in [0, 1]$  is a measure of distance parallel to the grey-line axis, saturation  $\in [0, 1]$  is a measure of distance perpendicular to the axis and hue  $\in [0, 1]$  is an angular measure around this axis.

Here, a colour is assigned to the forecast  $\mathbf{p}$  by associating the dominant category  $\theta(\mathbf{p}; \mathbf{q})$  with the hue and the information gain  $E(\mathbf{p}; \mathbf{q})$  with the saturation. The

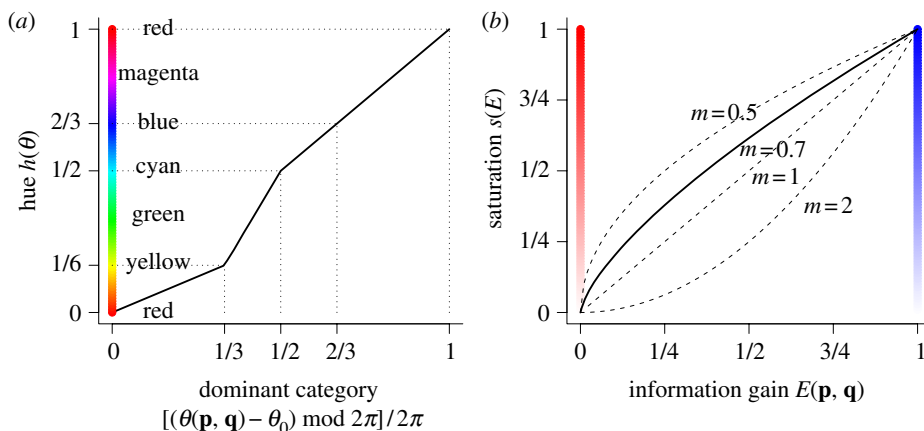


Figure 5. The functions of equation (3.4). (a) Hue as a function of dominant category  $\theta(\mathbf{p}; \mathbf{q})$ . (b) Saturation as a function of information gain  $E(\mathbf{p}; \mathbf{q})$ .  $m = 0.7$  and  $\theta_0 = 0$  are used throughout this paper.

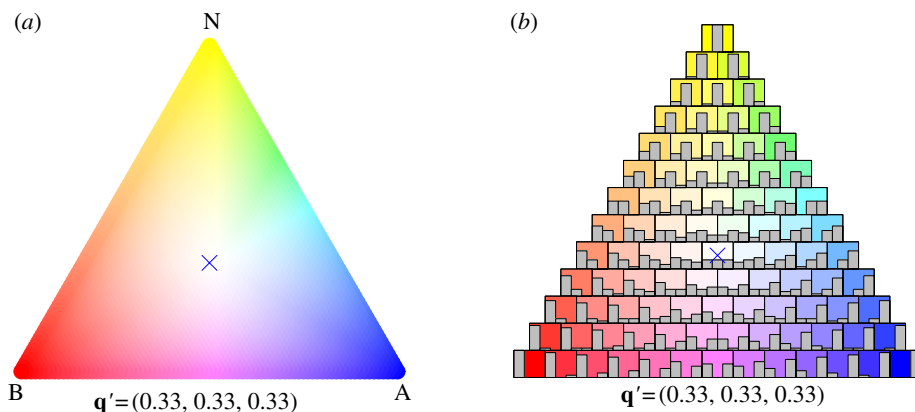


Figure 6. (a) Palette (when  $\theta_0 = 0^\circ$ ,  $m = 0.7$ ) of colours assigned to ternary forecasts (equation (3.4)). Climatology  $\mathbf{q}$  indicated by blue cross. (b) As (a) but with barplots indicating ternary forecasts  $\mathbf{p}$ .

proposed algorithm for colour assignment is

$$\left. \begin{aligned} \text{hue} &= h\left(\frac{(\theta(\mathbf{p}; \mathbf{q}) - \theta_0) \bmod 2\pi}{2\pi}\right), \\ \text{saturation} &= (E(\mathbf{p}; \mathbf{q}))^m \\ \text{and} \quad \text{value} &= 1, \end{aligned} \right\} \quad (3.4)$$

where the functional forms  $h(\theta)$  and  $s(E) = E^m$  are illustrated in figure 5. Suggested default choices for the parameters are  $m = 0.7$  and  $\theta_0 = 0$ , which together yield the colour palette shown in figure 6. Note that the climatology is always assigned the colour white by this system. The nonlinear hue function

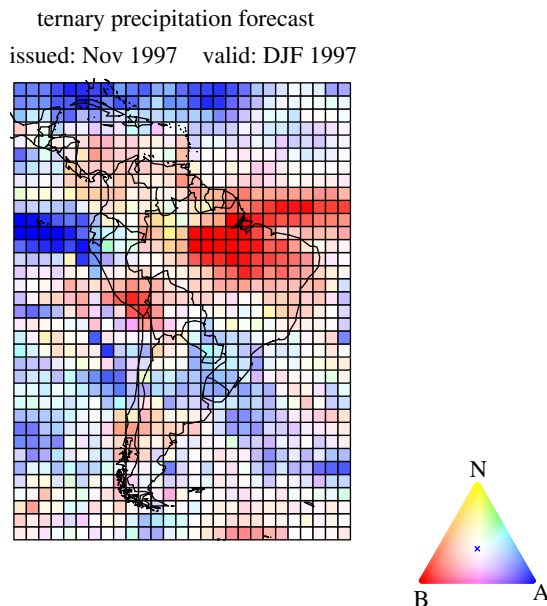


Figure 7. A probabilistic forecast map produced using the colour scheme proposed here (equation (3.4)). The continuous colour palette (bottom right) conveys the probabilities assigned to ‘below-’, ‘near-’ and ‘above-normal’ precipitation categories (figure 1). For example, colours close to white indicate a forecast similar to climatology, strong red indicates a high probability of below-normal precipitation, and strong blue a high probability of above-normal precipitation.

$h(\theta)$  has been chosen to minimize the region of barycentric coordinates that is assigned the colour green and hence to minimize difficulty for readers with green-weak colour blindness [15,16]. The parameter  $\theta_0$  can be chosen in order to rotate the palette in the triangle about the climatology, while the exponent  $m$  controls the rate at which colour saturation changes away from the climatology.

An example of a forecast map using the proposed colour scheme is shown in figure 7. In this illustration, the data consist of integrated seasonal precipitation forecasts for South America produced by the EUROBRISA project. At each spatial location, the probabilistic forecast is compared with the terciles of the local climatology. In other words, the ternary climatology  $\mathbf{q}' = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$  (shown by a cross in the colour palette at bottom right) has been used as the benchmark ternary forecast. The map shows clearly that high probability is assigned to low rainfall in northern Brazil (red) while there are regions in southeast Brazil and the Southern Ocean for which the forecast barely differs from climatology (white). There are also regions where rainfall in the near-normal category is forecast with high probability (yellow) and low probability (purple).

(e) *Special case: Gaussian distributions*

To gain insight into the colours assigned by equation (3.4), it is helpful to consider the special case in which the climatology  $G(x)$  and the forecast  $F(x)$  are both Gaussian distributions. The space of Gaussian distributions is

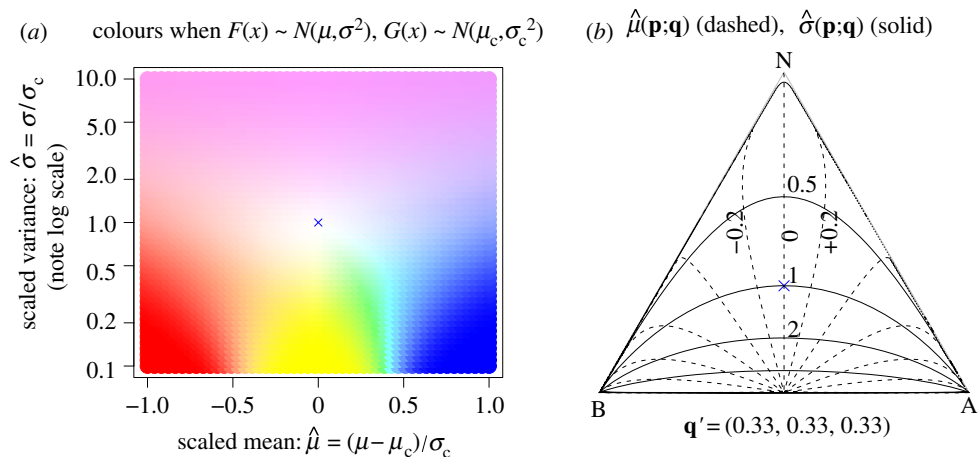


Figure 8. (a) Colours assigned when both forecast  $F(x)$  and climatology  $G(x)$  are Gaussian distributions. The case where the forecast equals the climatology is indicated by a cross. (b) Contours of scaled mean  $\hat{\mu} \in \{0, \pm 0.2, \pm 0.5, \pm 1, \pm 2, \pm 5\}$  and scaled standard deviation  $\hat{\sigma} \in \{0.2, 0.5, 1, 2, 5\}$  in barycentric coordinates (equation (3.6)).

two-dimensional (each distribution defined by its mean and variance) and so there is a natural one-to-one mapping between a (Gaussian) forecast  $F(x)$  and its ternary representation  $\mathbf{p}$ .

Specifically, suppose that the climatology is  $N(\mu_c, \sigma_c^2)$  and the forecast is  $N(\mu, \sigma^2)$ . It follows that

$$G(x) = \Phi\left(\frac{x - \mu_c}{\sigma_c}\right) \quad \text{and} \quad F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right), \quad (3.5)$$

where  $\Phi(z)$  is the CDF of the standard normal distribution  $N(0, 1)$ . It is helpful to normalize the forecast mean and standard deviation with respect to those of the climatology

$$\hat{\mu} = \frac{\mu - \mu_c}{\sigma_c} \quad \text{and} \quad \hat{\sigma} = \frac{\sigma}{\sigma_c}. \quad (3.6)$$

It follows that the forecast is more sharply peaked than the climatology when  $\hat{\sigma} < 1$  and less sharply peaked than the climatology when  $\hat{\sigma} > 1$ . The ternary forecast  $\mathbf{p}$  can be calculated from equation (2.1) by noting that

$$F(x_B) = \Phi\left(\frac{\Phi^{-1}(q_B) - \hat{\mu}}{\hat{\sigma}}\right) \quad \text{and} \quad F(x_A) = \Phi\left(\frac{\Phi^{-1}(q_B + q_N) - \hat{\mu}}{\hat{\sigma}}\right). \quad (3.7)$$

Figure 8a illustrates the colours that are assigned to a forecast and climatology when both are Gaussian. When the forecast is much more sharply peaked than the climatology ( $\hat{\sigma} \ll 1$ ) the ternary forecast assigns nearly all of its probability mass to one of the three categories. When the forecast mean is much less than the climatological mean ( $\hat{\mu} \ll 0$ ) this probability is assigned to category B and

the assigned colour is a strong red. Similarly, a strong yellow is assigned when the means are comparable ( $\hat{\mu} \approx 0$ ) and a strong blue is assigned when the forecast mean significantly exceeds the climatological mean ( $\hat{\mu} \gg 0$ ).

The situation is rather different when the forecast has much higher variance than the climatology ( $\hat{\sigma} \gg 1$ ). In this case, the ternary forecast assigns little probability mass to the central category N but rather splits it approximately evenly between the two extreme categories. In this case, the assigned colour is purple. This sort of situation might arise in an ensemble climate forecast in which some members of the ensemble predict that the climate variable increases while other members predict that it decreases. The purple colour might also arise if the forecasting system predicted an increase in the frequency of extreme events in both tails of the distribution.

Figure 8*b* shows  $\hat{\mu}$  and  $\hat{\sigma}$  in barycentric coordinates, and hence the Gaussian distributions of climatology and forecast that would yield given a ternary forecast  $\mathbf{p}$ . Consider the case in which the forecast and the climatology have similar variances, as appears often to be the case. It follows that  $\hat{\sigma} \approx 1$  and so the ternary forecast  $\mathbf{p}$  lies close to the contour  $\hat{\sigma} \approx 1$  in figure 8*b*. The forecast mean  $\mu$  then controls the colour assigned to the forecast, with red assigned when  $\hat{\mu} \ll 1$ , white assigned when  $\hat{\mu} \approx 1$  and blue assigned when  $\hat{\mu} \gg 1$ .

#### 4. Verification of ternary probabilistic forecasts

The previous sections have considered the visualization of a set of probabilistic forecasts. The aim in this section is to visualize the difference between a set of forecasts and the corresponding set of observations. This is known as forecast verification [1]. In the next subsection, the standard score decomposition of Murphy [17] is reviewed and re-interpreted geometrically.

##### (a) Decomposition of the score

Suppose that the scoring matrix  $L$  has been specified so that the score of an individual forecast–observation pair is given by a squared distance in the appropriate triangle (equation (3.2)). As in previous sections, the Brier score is the default choice but the results hold for any quadratic scoring function.

The triangle representing ternary forecast space can be discretized into a finite number of bins with centres  $\mathbf{P}_k$ . For simplicity, each forecast  $\mathbf{P}$  that lies in the bin with centre  $\mathbf{P}_k$  will be reassigned the central value  $\mathbf{P}_k$  [18]. Following Murphy [17], consider the identity

$$\mathbf{P}_k - \mathbf{O} = (\mathbf{P}_k - \bar{\mathbf{O}}|\mathbf{P}_k) - (\mathbf{Q} - \bar{\mathbf{O}}|\mathbf{P}_k) + (\mathbf{Q} - \mathbf{O}), \quad (4.1)$$

where  $\bar{\mathbf{O}}|\mathbf{P}_k$  is the mean observation associated with forecasts in bin  $\mathbf{P}_k$ . If the climatology is defined to be the mean of all the observations  $\mathbf{Q} = \bar{\mathbf{O}}$ , it follows that a decomposition of the mean score with binned forecasts is

$$\left. \begin{aligned} \|\mathbf{P}_k - \mathbf{O}\|^2 &= \|\mathbf{Q} - \mathbf{O}\|^2 - \|\mathbf{Q} - \bar{\mathbf{O}}|\mathbf{P}_k\|^2 + \|\mathbf{P}_k - \bar{\mathbf{O}}|\mathbf{P}_k\|^2, \\ \text{score} &= \text{uncertainty} - \text{resolution} + \text{reliability} \end{aligned} \right\} \quad (4.2)$$

$$\text{and} \quad S = U - Z + R.$$

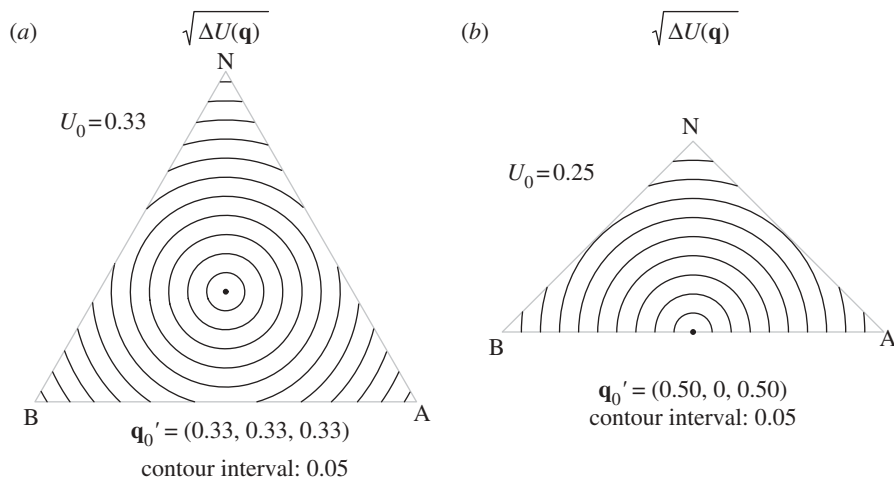


Figure 9. The root-uncertainty-reduction  $\sqrt{\Delta U}$  (equation (4.3)) for the (a) Brier score and (b) ranked probability score interpreted as distances within the appropriate triangles.

Equation (4.2) constitutes the standard decomposition of the score  $S$  into uncertainty  $U$ , reliability  $R$  and resolution  $Z$  [1]. It is important to stress that this decomposition applies only when all forecasts within a bin are assigned the central value [19].

It has already been remarked that the score is the mean square distance in the triangle between forecasts and observations, and it is clear that the uncertainty is the mean square distance between observations and climatology. In order to see how uncertainty varies with climatology  $\mathbf{q}$  (figure 9), it is helpful (appendix A c) to consider the particular climatology  $\mathbf{q}_0$  (shown by a dot in figure 9) for which the uncertainty gains its maximum value  $U_0 = U(\mathbf{q}_0)$ . The uncertainty  $U(\mathbf{q})$  is then

$$U = U_0 - \Delta U, \quad (4.3)$$

where  $\Delta U(\mathbf{q})$  represents the reduction in uncertainty between  $\mathbf{q}_0$  and  $\mathbf{q}$ . Figure 9 illustrates that the root-uncertainty-reduction  $\sqrt{\Delta U}$  can be visualized as the distance in the triangle between the climatology of interest  $\mathbf{q}$  and the climatology  $\mathbf{q}_0$  of maximum uncertainty.

Similarly, the resolution in equation (4.2) is the mean square distance between climatology  $\mathbf{Q}$  and the mean observations conditional on the forecasts  $\bar{\mathbf{O}}|\mathbf{P}_k$ . The reliability is the mean square distance between forecasts  $\mathbf{P}_k$  and the mean observations conditional on the forecasts  $\bar{\mathbf{O}}|\mathbf{P}_k$ .

It follows that the root-score  $\sqrt{S}$  is the r.m.s. distance between forecasts and observations, the root-uncertainty  $\sqrt{U}$  is the r.m.s. distance between climatology and observations, the root-resolution  $\sqrt{Z}$  is the r.m.s. distance between climatology and mean conditional observations and the root-reliability  $\sqrt{R}$  is the r.m.s. distance between forecasts and mean conditional observations. These geometrical interpretations as r.m.s. distances are illustrated in the triangle in figure 10.



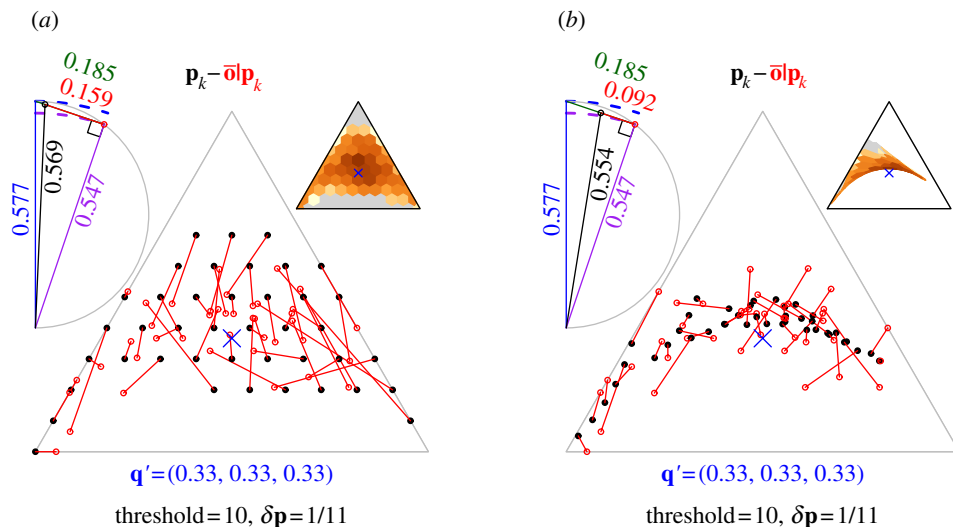


Figure 11. Examples of proposed ‘ternary reliability diagrams’. (a) A set of poorly calibrated ‘original’ forecasts has poor reliability (long red dipoles). (b) Recalibration of these forecasts by ‘moving the black dots’ improves reliability (shorter red dipoles). Decomposition diagram (top left) shows that recalibrated score is close to the best achievable by recalibration. Sharpness diagram (top right) illustrates the recalibration geometrically.

line. This represents the worst performance that might be expected from a skilful forecasting system. On the other hand, the best possible performance is indicated by the purple dashed line. This would occur for a forecasting system with perfect root-reliability  $\sqrt{R} = 0$ .

### (b) Ternary reliability diagrams

The ideas of the preceding subsection can now be illustrated with an example. The data in this example are seasonal precipitation hindcasts at one month lead-time from the EUROBRISA integrated forecast model. The data relate to the years 1981–2005, the season January–February–March and a spatial region  $72.5^\circ\text{W}$ – $42.5^\circ\text{W}$ ,  $12.5^\circ\text{S}$ – $2.5^\circ\text{N}$  for which seasonal forecasts show reasonable skill.

Figure 11a contains one of our proposed ternary reliability diagrams. The large triangle contains information on the reliability of the forecasts. Bins of size  $\delta\mathbf{p} = \frac{1}{11}$  were chosen and the central forecast in each bin  $\mathbf{P}_k$ —to which all forecasts in that bin are set equal—is plotted in the triangle as a black circle. The mean observation conditional on a forecast  $\bar{\mathbf{O}}|\mathbf{P}_k$  is plotted as a red circle and the two are joined by a red line in order to form a ‘dipole’. The length of each dipole represents the root-reliability of the forecasts in each bin. The number of forecasts in each bin is shown graphically by the ternary sharpness diagram at the top right. (Dark colours, high density; light colours, low density; grey, no data.) The climatology is shown by a blue cross. Dipoles for bins containing fewer than 10 forecasts (indicated in the diagram by the text ‘threshold = 10’) are omitted for clarity. An advantage of this sort of visualization of the data is that coherent patterns in the reliability dipoles are immediately obvious.



Such patterns would indicate regions of ternary forecast space in which the forecasting system consistently assigned the wrong probability to one of the three categories.

The decomposition diagram (top left) illustrates geometrically the decomposition of the root-score in this dataset. Thus, the root-score  $\sqrt{S} = 0.569$  is seen to be rather poor compared with the best-possible root-score  $\sqrt{U - Z} = 0.547$  that could be attained through recalibration.

(c) *Forecast maps including verification data*

The map of figure 7 contains information about forecasts but not about the skill of the forecasting system. In order to incorporate verification data into a forecast map, the array of coloured squares can be replaced by an array of coloured circles whose radii are a measure of forecast skill in a verification dataset. One possibility would be to set radius proportional to  $(Z - R)/Z$  which is a ‘standard’ definition of skill. In this paper, however, the convention has been to consider quantities like  $\sqrt{Z}$  and  $\sqrt{R}$  which can be interpreted as distances in a triangle rather than squared distances (figure 10) and so  $(\sqrt{Z} - \sqrt{R})/\sqrt{Z}$  is preferred here. (It is straightforward to relate these two possibilities via the identity  $(Z - R)/Z = 1 - (1 - (\sqrt{Z} - \sqrt{R})/\sqrt{Z})^2$ .) Making the choice

$$\text{circle radius} \propto \frac{\sqrt{Z} - \sqrt{R}}{\sqrt{Z}} \quad (4.4)$$

it follows that larger circles will be plotted in regions of high skill, and no circle will be plotted if the forecasting system has less skill than a climatological forecast (i.e. when the assigned radius is negative). A map produced in this way is shown in figure 12. This map shows clearly that, for the season shown, the skill of this forecasting system is greatest near the northern coast of South America. Maps of this type should prove useful in communication of operational forecasts because the reader’s eye is drawn to areas where the forecasting system has been shown to perform well.

## 5. Recalibration of ternary probabilistic forecasts

The decomposition diagram in figure 11*a* illustrates that significant improvement can sometimes be made to the mean-score by recalibrating the forecasts. The score of the original forecasts is  $\|\mathbf{P} - \mathbf{O}\|^2$  and so the aim is to produce recalibrated forecasts  $\tilde{\mathbf{p}}$ , calculated from the original forecasts via some specified functional form, whose score  $\|\tilde{\mathbf{P}} - \mathbf{O}\|^2$  is minimized. As an example, consider recalibrated ternary forecasts which are a quadratic function of the original ternary forecasts (appendix A*d*). Given a dataset of forecasts and observations, standard numerical optimization techniques can be used to find coefficients in the quadratic function which minimize the score of the recalibrated forecasts. The recalibration function can be interpreted by comparing figure 11*a, b*. Recalibration minimizes the score by changing the reliability (that is, the r.m.s. length of the red dipoles). It is important to note that recalibration changes the forecasts (black dots) but not the observations (red circles) and so the r.m.s. length of

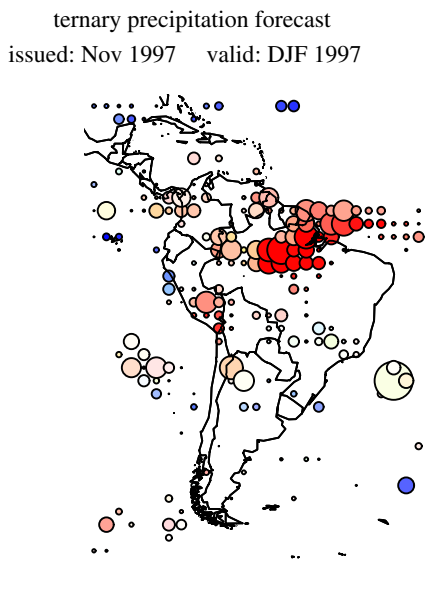


Figure 12. Probabilistic forecast map including verification data. Larger circles correspond to forecasting system performing well historically. No circle is plotted where forecasting system has performed worse than a climatological forecast. Colours assigned as in figure 7, circle radius via equation (4.4).

the red dipoles has been reduced from  $\sqrt{R} = 0.159$  to  $\sqrt{R} = 0.092$  solely by moving the position of black dots. The effect of this reduction can be seen in the decomposition diagrams (top left). Recalibration changes neither the root-uncertainty  $\sqrt{U} = 0.577$  nor the root-resolution  $\sqrt{Z} = 0.185$ , but reduces the root-score from  $\sqrt{S} = 0.569$  to  $\sqrt{S} = 0.554$ .

In this case, it is immediately clear that the recalibrated forecasts (figure 11*b*, black dots) are clustered around an arc passing close to the climatology (blue cross). It is interesting that this arc is similar to the contour  $\hat{\sigma} = 1$  in figure 8*b* which applies when both the climatology and the forecast are Gaussian with equal variance.

Finally, the sharpness diagram (figure 11*b*, top right) shows the distribution of the recalibrated forecasts. It is produced by passing the sharpness diagram of the original forecasts (figure 11*a*, top right) through the quadratic recalibration function (appendix A*d*). Regions of the sharpness diagram coloured grey indicate possible forecasts that were not issued originally. Regions coloured white indicate forecasts that can never be issued under this recalibration.

## 6. Conclusions

This article has outlined a novel procedure for visualizing ternary probabilistic forecasts. The proposed maps convey through colour all of the information present in a ternary forecast. A colour palette printed next to the map can be used to deduce the ternary forecasts directly from the colour used.

The proposed colour scheme always assigns the colour white to a forecast close to climatology, and strong colours to forecasts that differ greatly from climatology. The examples considered in this paper have concerned precipitation and so the ‘wet’ colour blue has been assigned to the above-normal category A and the ‘dry’ colour red to the below-normal category B. It is straightforward to reverse these conventions if desired for other forecast variables (e.g. temperature). The default palette (with  $\theta_0 = 0$ ) assigns strong red, yellow and blue, respectively, to forecasts that assign high probability to one of the three categories. Yellow has been chosen in preference to green in order to assist colour blind readers. A variety of palettes can be created by varying the parameters  $m$  and  $\theta_0$  in equation (3.4).

A novel visual interpretation of ternary forecasts has also been suggested for verification and recalibration under quadratic scoring functions. It has been shown that Brier scores correspond to squared distances in an equilateral triangle and ranked probability scores correspond to squared distances in a right-angled triangle. Thus, the root-score, root-uncertainty, root-resolution and root-reliability can all be interpreted as r.m.s. distances. Verification data can be visualized using the proposed ternary reliability diagrams (figure 11), which incorporate a decomposition diagram (top left) and a sharpness diagram (top right) to aid interpretation. The geometrical interpretation can also be applied to nonlinear recalibration of probabilistic forecasts.

It is hoped that the procedures outlined here will prove useful in the interpretation and communication of operational weather and climate forecasts. When no information on skill is available a map like that of figure 7 can be produced. If skill information is available, it can be incorporated as in figure 12.

This work was supported by the EUROBRISA network project (F/00 144/AT) kindly funded by the Leverhulme Trust. The dynamical ensemble forecast data were kindly provided by ECMWF as part of the EUROSIP project. Three forecasting centres are the partners in EUROSIP, these are ECMWF, the UK Met Office and Météo-France. C.A.S.C. was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), processes 2005/05210-7 and 2006/02497-6. R.L. and D.B.S. thank Aidan Slingsby of City University for useful discussions.

## Appendix A

Let the triangle in  $\mathbb{R}^2$  induced by the matrix  $L$  have vertices BNA (figure 13) and sides of length  $b$ ,  $n$  and  $a$ :

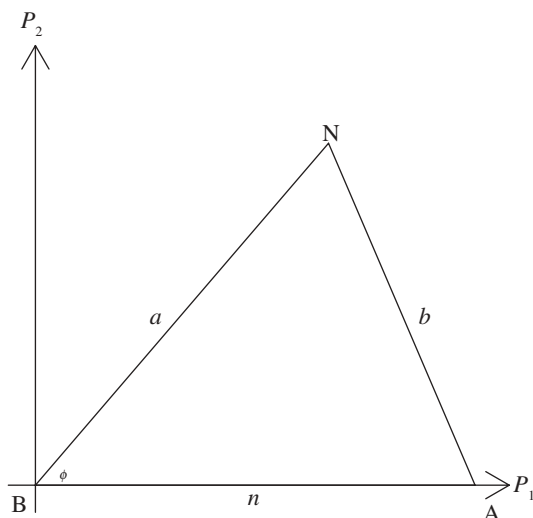
$$b = \sqrt{S(\mathbf{o}_N; \mathbf{o}_A)}, \quad n = \sqrt{S(\mathbf{o}_A; \mathbf{o}_B)} \quad \text{and} \quad a = \sqrt{S(\mathbf{o}_B; \mathbf{o}_N)}. \quad (\text{A } 1)$$

From the cosine rule, the angle  $\phi$  is given by

$$\cos \phi = \frac{n^2 + a^2 - b^2}{2an}. \quad (\text{A } 2)$$

It follows that a ternary forecast  $\mathbf{p} \in \mathbb{R}^3$  and the associated point  $\mathbf{P} \in \mathbb{R}^2$  within the induced triangle are related by

$$\mathbf{P} = \hat{M}\mathbf{p} \quad \text{and} \quad \mathbf{p} = M\mathbf{P} + \mathbf{o}_B, \quad (\text{A } 3)$$

Figure 13. The triangle in  $\mathbb{R}^2$  induced by the score matrix  $L$ .

where the transformation matrices are defined by

$$\hat{M} = \begin{pmatrix} 0 & a \cos \phi & n \\ 0 & a \sin \phi & 0 \end{pmatrix} \quad \text{and} \quad M = \frac{1}{an \sin \phi} \begin{pmatrix} -a \sin \phi & a \cos \phi - n \\ 0 & n \\ a \sin \phi & -a \cos \phi \end{pmatrix}. \quad (\text{A } 4)$$

(a) *Brier score*

The Brier score is defined by

$$S(\mathbf{p}; \mathbf{o}) = \frac{1}{2} [(p_B - o_B)^2 + (p_N - o_N)^2 + (p_A - o_A)^2]. \quad (\text{A } 5)$$

It follows that, for the Brier score,

$$L = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \hat{M} = \begin{pmatrix} 0 & \frac{1}{2} & 1 \\ 0 & \frac{\sqrt{3}}{2} & 0 \end{pmatrix} \quad \text{and} \quad M = \begin{pmatrix} -1 & \frac{-1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{3}} \\ 1 & \frac{-1}{\sqrt{3}} \end{pmatrix}. \quad (\text{A } 6)$$

The triangle induced by the Brier score is an equilateral triangle with unit sides  $b = n = a = 1$ .

(b) *Ranked probability score*

The ranked probability score [3,4,20] is defined by

$$S(\mathbf{p}; \mathbf{o}) = \frac{1}{2} [(p_B - o_B)^2 + (p_B + p_N - o_B - o_N)^2]. \quad (\text{A } 7)$$

It follows that, for the ranked probability score,

$$L = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}, \quad \hat{M} = \begin{pmatrix} 0 & \frac{1}{2} & 1 \\ 0 & \frac{1}{2} & 0 \end{pmatrix} \quad \text{and} \quad M = \begin{pmatrix} -1 & -1 \\ 0 & 2 \\ 1 & -1 \end{pmatrix}. \quad (\text{A } 8)$$

The triangle induced by the ranked probability score is a right-angled triangle with sides  $b = a = 1/\sqrt{2}$ ,  $n = 1$ .

(c) *The uncertainty of a quadratic score*

For a quadratic scoring rule (equation (3.1)), the uncertainty  $U(\mathbf{q})$  is equal to the expected score when the climatology  $\mathbf{q}$  is issued as the forecast. It follows that

$$U(\mathbf{q}) = \mathbf{v}'\mathbf{q} - \mathbf{q}'L'L\mathbf{q}, \quad \text{where } \mathbf{v} = \text{diag}(L'L). \quad (\text{A } 9)$$

It follows that

$$U(\mathbf{q}) = \frac{1}{2}\mathbf{q}'_0\mathbf{v} - (\mathbf{q} - \mathbf{q}_0)'L'L(\mathbf{q} - \mathbf{q}_0), \quad \text{where } \mathbf{q}_0 = \frac{1}{2}(L'L)^{-1}\mathbf{v}. \quad (\text{A } 10)$$

In the particular cases of the Brier and ranked probability scores:

$$\bar{U} = \begin{cases} \frac{1}{2}(1 - \mathbf{q}'\mathbf{q}) & \text{(Brier score)} \\ \frac{1}{2}(q_B(1 - q_B) + q_A(1 - q_A)) & \text{(ranked probability score)}. \end{cases} \quad (\text{A } 11)$$

(d) *Quadratic recalibration*

An original forecast  $\mathbf{p}' = (p_B, p_N, p_A)$  is defined by  $p_B$  and  $p_A$  since  $p_N = 1 - p_B - p_A$ . Define a recalibrated forecast  $\tilde{\mathbf{p}}' = (\tilde{p}_B, \tilde{p}_N, \tilde{p}_A)$  by

$$\left. \begin{aligned} \tilde{p}_B &= C_1 + C_2p_B + C_3p_A + C_4p_B^2 + C_5p_Bp_A + C_6p_A^2, \\ \tilde{p}_A &= C_7 + C_8p_B + C_9p_A + C_{10}p_B^2 + C_{11}p_Bp_A + C_{12}p_A^2 \end{aligned} \right\} \quad (\text{A } 12)$$

and  $\tilde{p}_N = 1 - \tilde{p}_B - \tilde{p}_A$ .

Numerical values for the parameters  $C_1, \dots, C_{12}$  can be found by minimizing the mean score  $S(\tilde{\mathbf{p}}, \mathbf{o})$  of the recalibrated forecasts.

## References

- 1 Jolliffe, I. T. & Stephenson, D. B. 2003 *Forecast verification: a practitioner's guide in atmospheric science*. Chichester, UK: Wiley.
- 2 Brier, G. W. 1950 Verification of forecasts expressed in terms of probability. *Mon. Weather Rev.* **78**, 1–3. (doi:10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2)
- 3 Epstein, E. S. 1969 A scoring system for probability forecasts of ranked categories. *J. Appl. Meteorol.* **8**, 985–987. (doi:10.1175/1520-0450(1969)008<0985:ASSFPF>2.0.CO;2)
- 4 Murphy, A. H. 1969 On the ranked probability skill score. *J. Appl. Meteorol.* **8**, 988–989. (doi:10.1175/1520-0450(1969)008<0988:OTPS>2.0.CO;2)
- 5 Murphy, A. H. & Staël von Holstein, C.-A. 1975 A geometrical framework for the ranked probability score. *Mon. Weather Rev.* **103**, 16–20. (doi:10.1175/1520-0493(1975)103<0010:AGFFTR>2.0.CO;2)
- 6 Staël von Holstein, C.-A. & Murphy, A. H. 1978 The family of quadratic scoring rules. *Mon. Weather Rev.* **106**, 917–924. (doi:10.1175/1520-0493(1978)106<0917:TFOQSR>2.0.CO;2)

- 7 Slingsby, A., Lowe, R., Dykes, J., Stephenson, D. B., Wood, J. & Jupp, T. E. 2009 A pilot study for the collaborative development of new ways of visualising seasonal climate forecasts. In *Proc. 17th Annu. Conf. of GIS Research UK, Durham, UK, 1–3 April 2009*.
- 8 Stephenson, D. B., Coelho, C. A. S., Doblas-Reyes, F. J. & Balmaseda, M. 2005 Forecast assimilation: a unified framework for the combination of multi-model weather and climate predictions. *Tellus* **57A**, 253–264. (doi:10.1111/j.1600-0870.2005.00110.x)
- 9 Anderson, D., Stockdale, T., Balmaseda, M., Ferranti, L., Vitart, F., Molteni, F., Doblas-Reyes, F. J., Mogensen, K. & Vidard, A. 2007 Development of the ECMWF seasonal forecast system 3. Technical report 503, ECMWF, Reading, UK.
- 10 Graham, R. J., Gordon, M., McLean, P. J., Ineson, S., Huddleston, M. R., Davey, M. K., Brookshaw, A. & Barnes, R. T. H. 2005 A performance comparison of coupled and uncoupled versions of the Met Office seasonal prediction general circulation model. *Tellus* **57A**, 320–339. (doi:10.1111/j.1600-0870.2005.00116.x)
- 11 Guérémy, J.-F., Déqué, M., Braun, A. & Peidelèvre, J.-P. 2005 Actual and potential skill of seasonal predictions using the CNRM contribution to DEMETER: coupled versus uncoupled model. *Tellus* **57A**, 308–319. (doi:10.1111/j.1600-0870.2005.00101.x)
- 12 Nobre, P., Malagutti, M., Urbano, D. F., De Almeida, R. A. F. & Giarolla, E. 2009 Towards an integrated seasonal forecasting system for South America. *J. Climate* **22**, 5686–5697. (doi:10.1175/2009JCLI2757.1)
- 13 Coelho, C. A. S., Stephenson, D. B., Balmaseda, M., Doblas-Reyes, F. J. & van Oldenborgh, G. J. 2006 Towards an integrated seasonal forecasting system for South America. *J. Climate* **19**, 3704–3721. (doi:10.1175/JCLI3801.1)
- 14 Jaynes, E. T. 2003 *Probability theory: the logic of science*. Cambridge, UK: Cambridge University Press.
- 15 Light, A. & Bartlein, P. J. 2004 The end of the rainbow? Color schemes for improved data graphics. *Eos* **85**, 385–391. (doi:10.1029/2004EO400002)
- 16 Stephenson, D. B. 2005 Comment on ‘Color schemes for improved data graphics’, by A. Light and P. J. Bartlein. *Eos* **86**, 196. (doi:10.1029/2005EO200005)
- 17 Murphy, A. H. 1973 A new vector partition of the probability score. *J. Appl. Meteorol.* **12**, 595–600. (doi:10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2)
- 18 Doblas-Reyes, F. J., Coelho, C. A. S. & Stephenson, D. B. 2008 How much does simplification of probability forecasts reduce forecast quality? *Meteorol. Appl.* **15**, 155–162. (doi:10.1002/met.50)
- 19 Stephenson, D. B., Coelho, C. A. S. & Jolliffe, I. T. 2008 Two extra components in the Brier score decomposition. *Weather Forecast.* **23**, 752–757. (doi:10.1175/2007WAF2006116.1)
- 20 Murphy, A. H. 1971 A note on the ranked probability skill score. *J. Appl. Meteorol.* **10**, 155–156. (doi:10.1175/1520-0450(1971)010<0155:ANOTRP>2.0.CO;2)