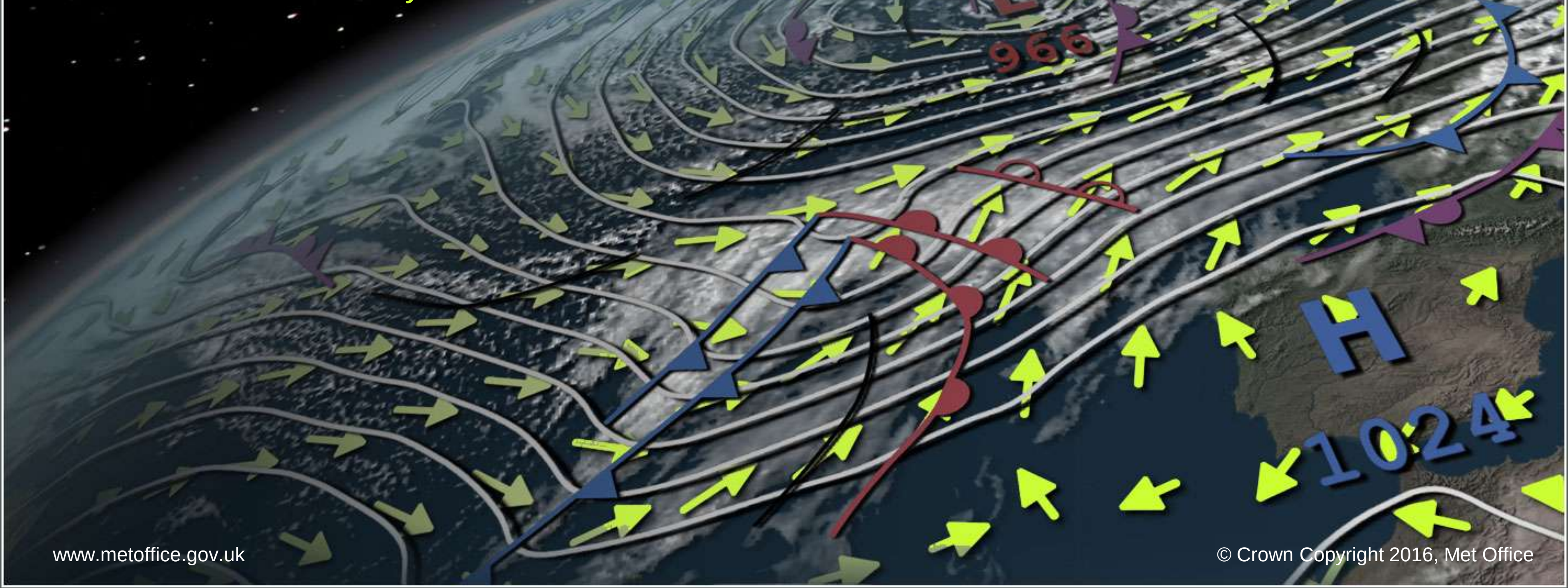
The generalized discrimination score (GDS):
Connections, corrections and potential applications

Roger Harbord roger.harbord@metoffice.gov.uk

7th International Verification Methods Workshop

9 May 2017



References and acknowledgements

- Simon Mason & Andreas Weigel

  "A generic forecast verification framework for administrative purposes." *Monthly Weather Review* 2009; **137**(1): 331–349.

  "The generalized discrimination score for ensemble forecasts." *Monthly Weather Review* 2011; **139**(9): 3069–3074.

- Roger Newson

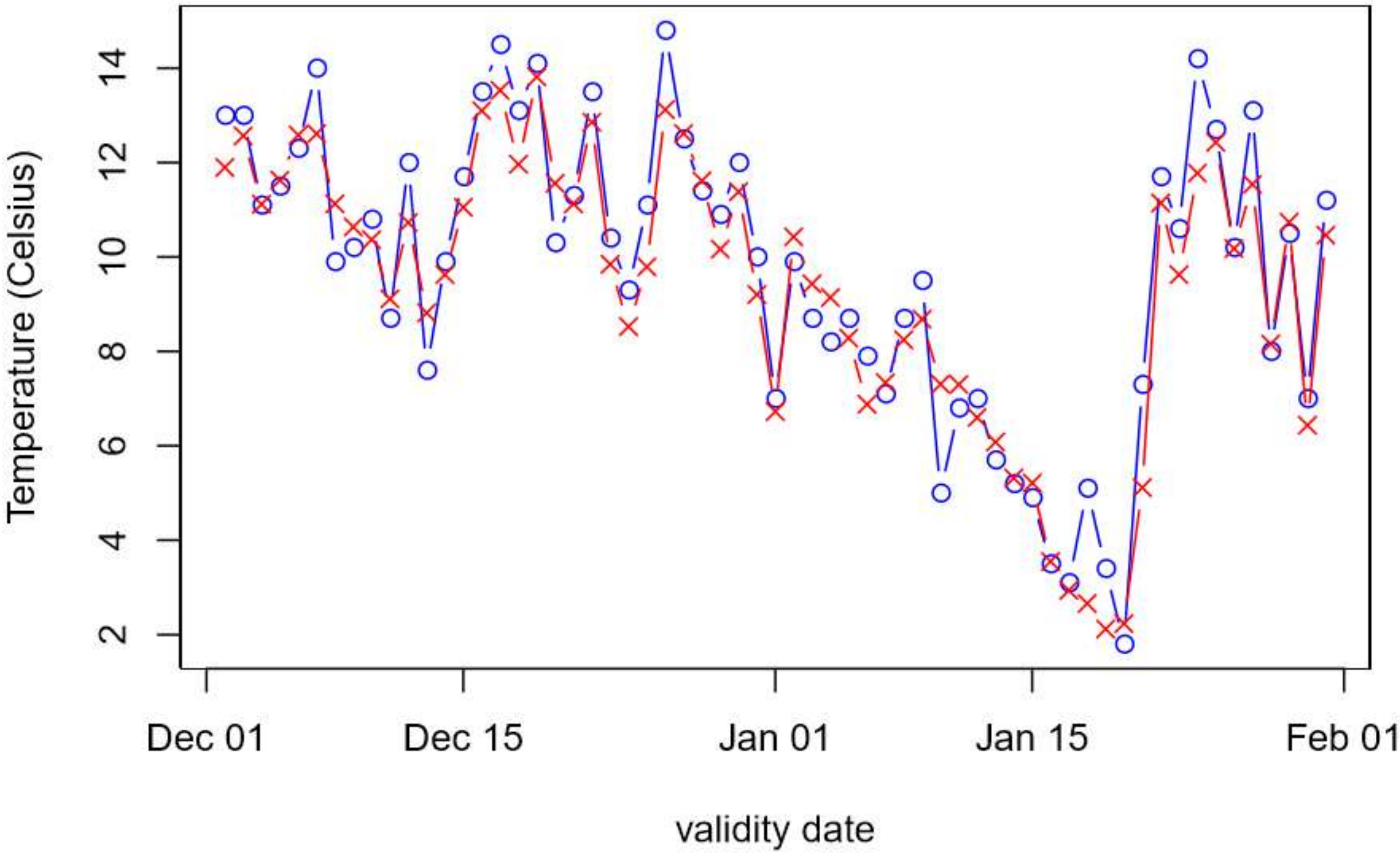
  "Parameters behind 'nonparametric' statistics: Kendall's tau, Somers' *D* and median differences." *Stata Journal* 2002; 2(1): 45–64.

- ECMWF evaluation group

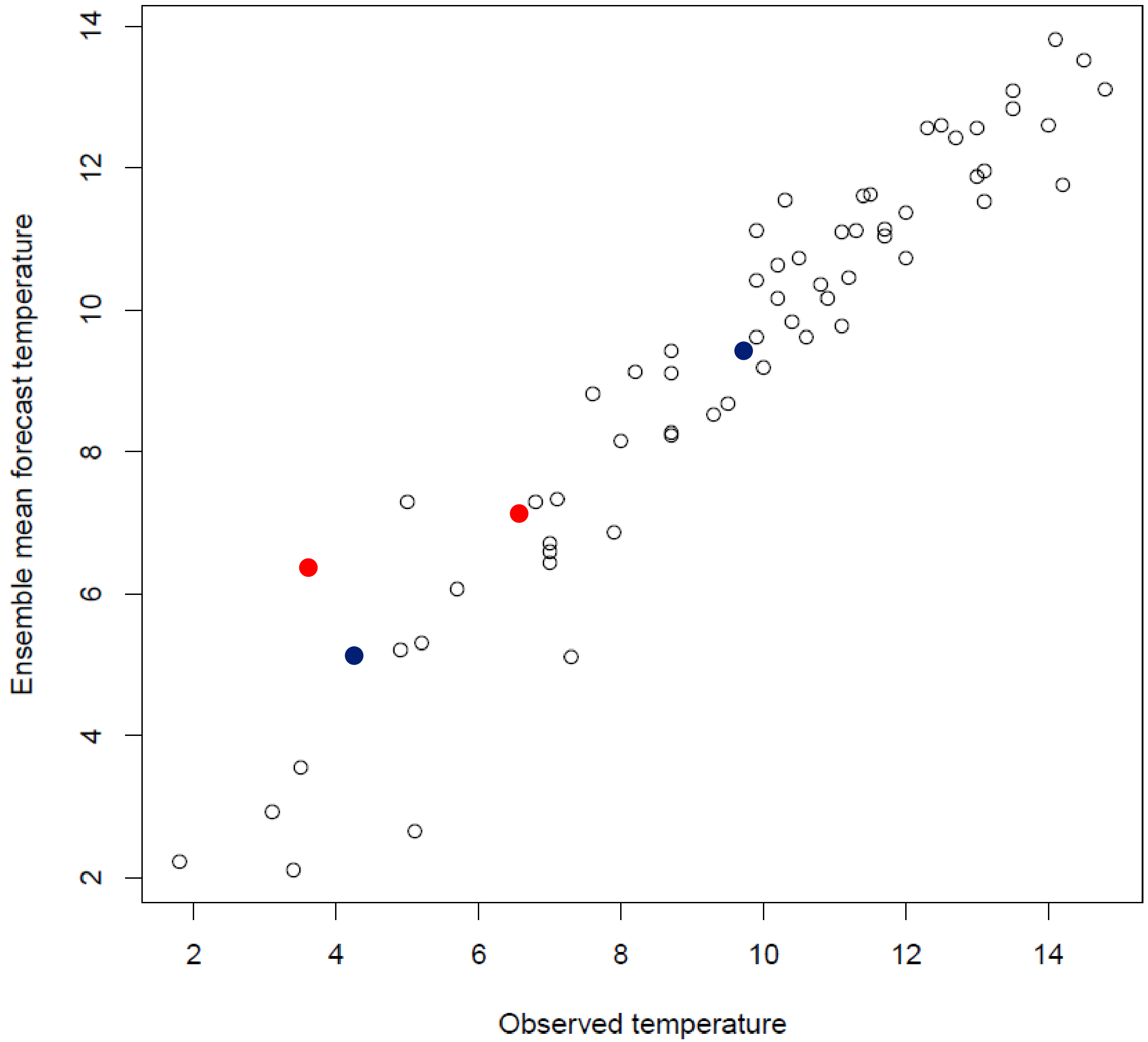
  (especially Zied Ben Bouallegue & David Richardson)

- Jonathan Flowerdew, Met Office
- Marion Mittermaier, Met Office

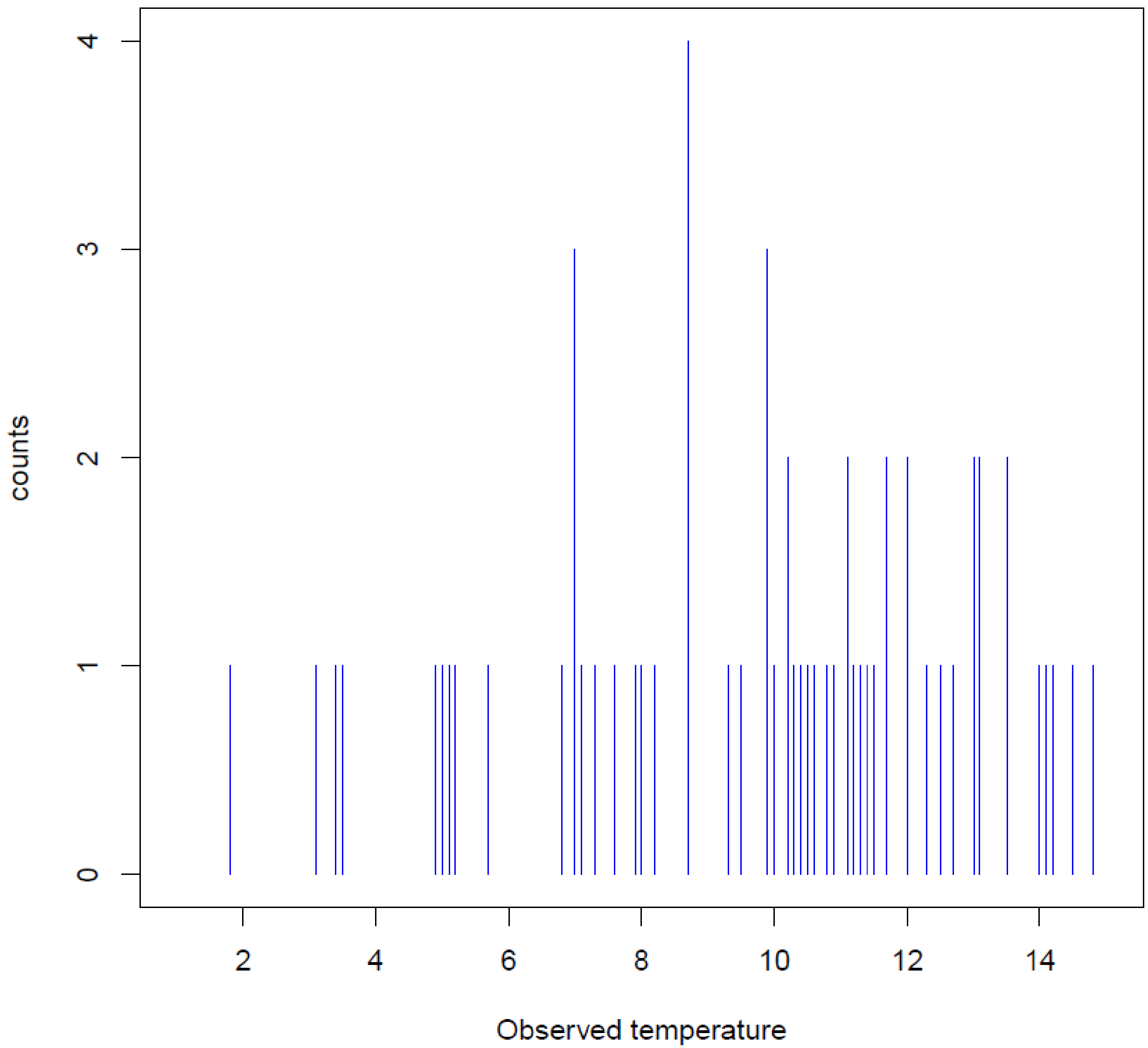
Midday temperature at Filton Airport, Bristol, UK
2 Dec 2015 – 31 Jan 2016
o observations x forecasts (T+24 ensemble mean)



Concordant and discordant pairs



Tied observations



GDS / 2AFC / Harrell's c range [0, 1]

$$GDS = \Pr(\text{pair concordant} \mid \text{obs unequal})$$

(0.5 when both obs *and* fcsts are equal)

Somers' D range [-1, 1]

$$D = \Pr(\text{pair } \textit{con}\text{cordant} \mid \text{obs unequal}) - \Pr(\text{pair } \textit{dis}\text{cordant} \mid \text{obs unequal})$$

$$D = 2c - 1$$

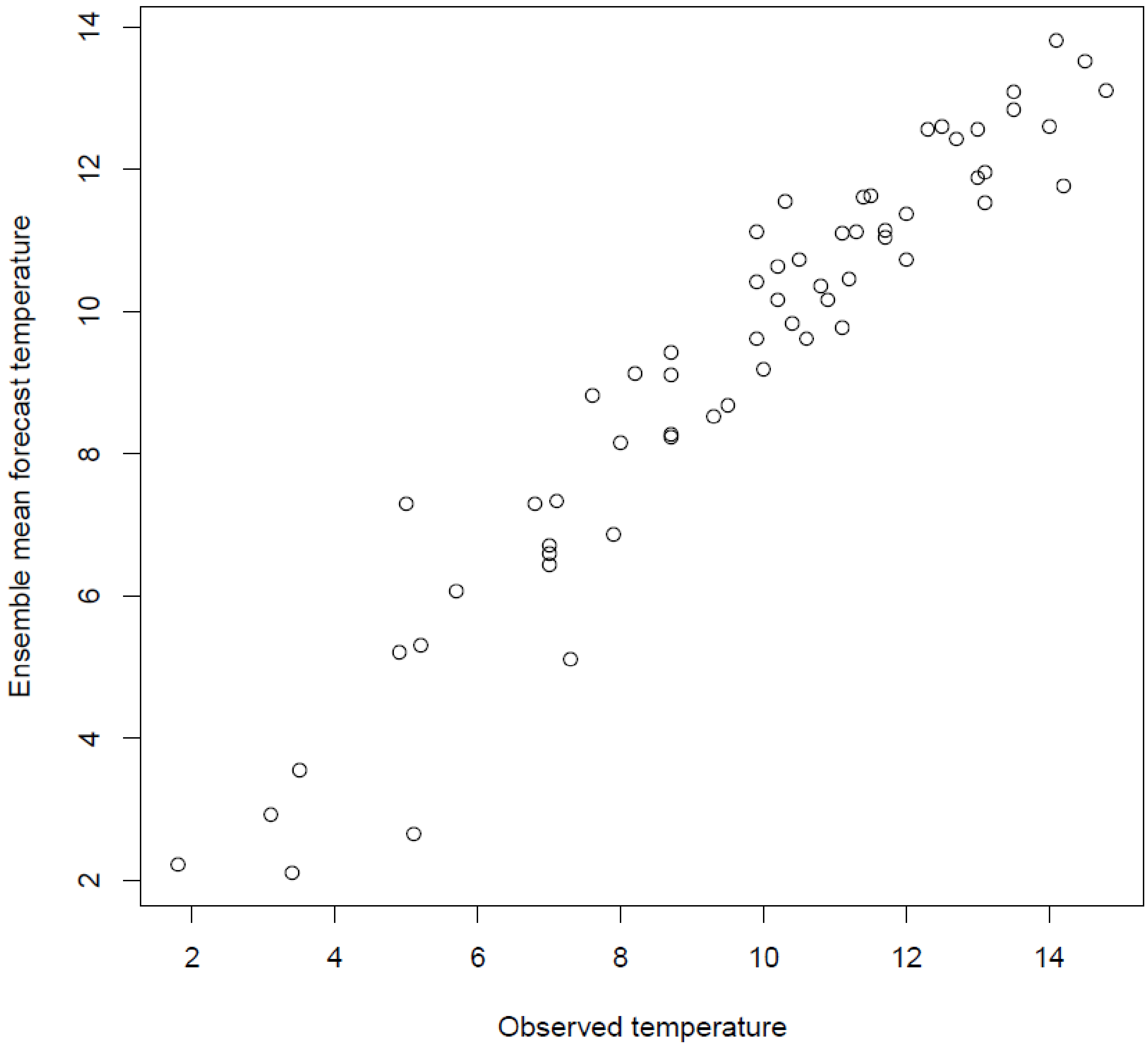
Computation of Somers' *D*

$$\mathrm{sign}(o_i - o_j) \times \mathrm{sign}\left(f_i - f_j\right),$$

averaged over all pairs (i, j) of data points with unequal observations $(o_i \neq o_j)$

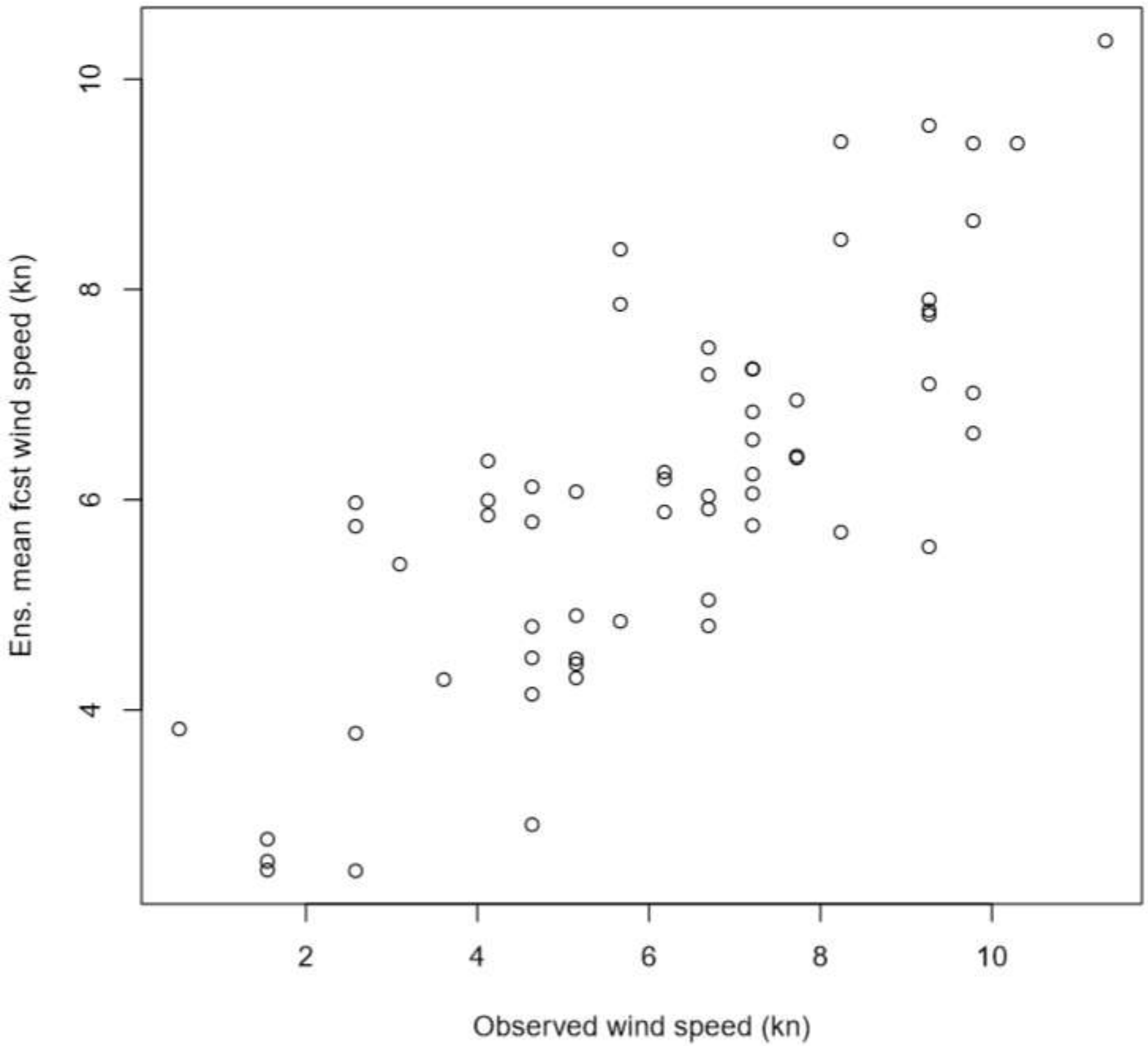
where o_i is the observation and f_i is the forecast for data point i.

Temperature *GDS* = 0.92 (*D* = 0.83)

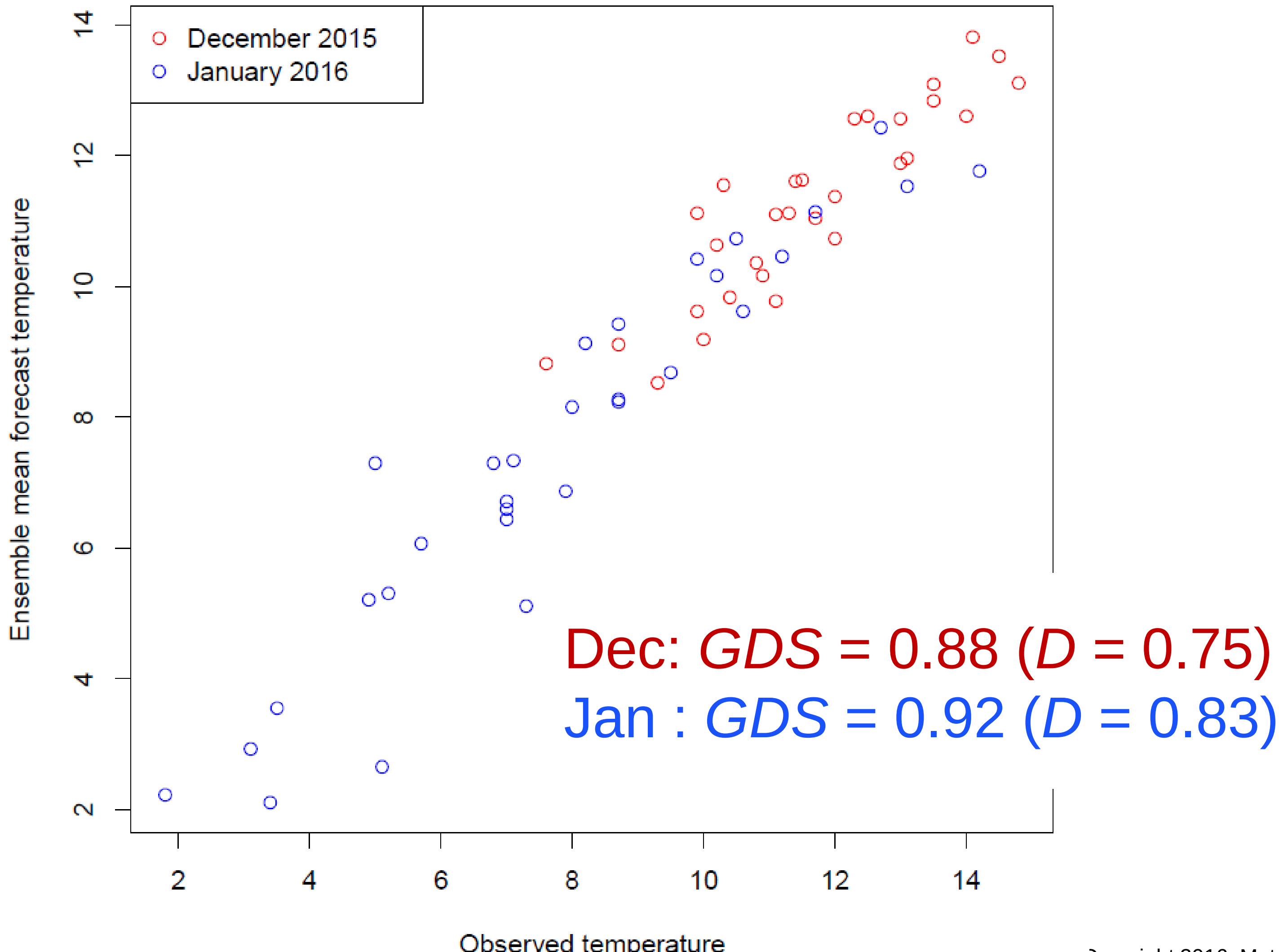


www.metoffice.gov.uk

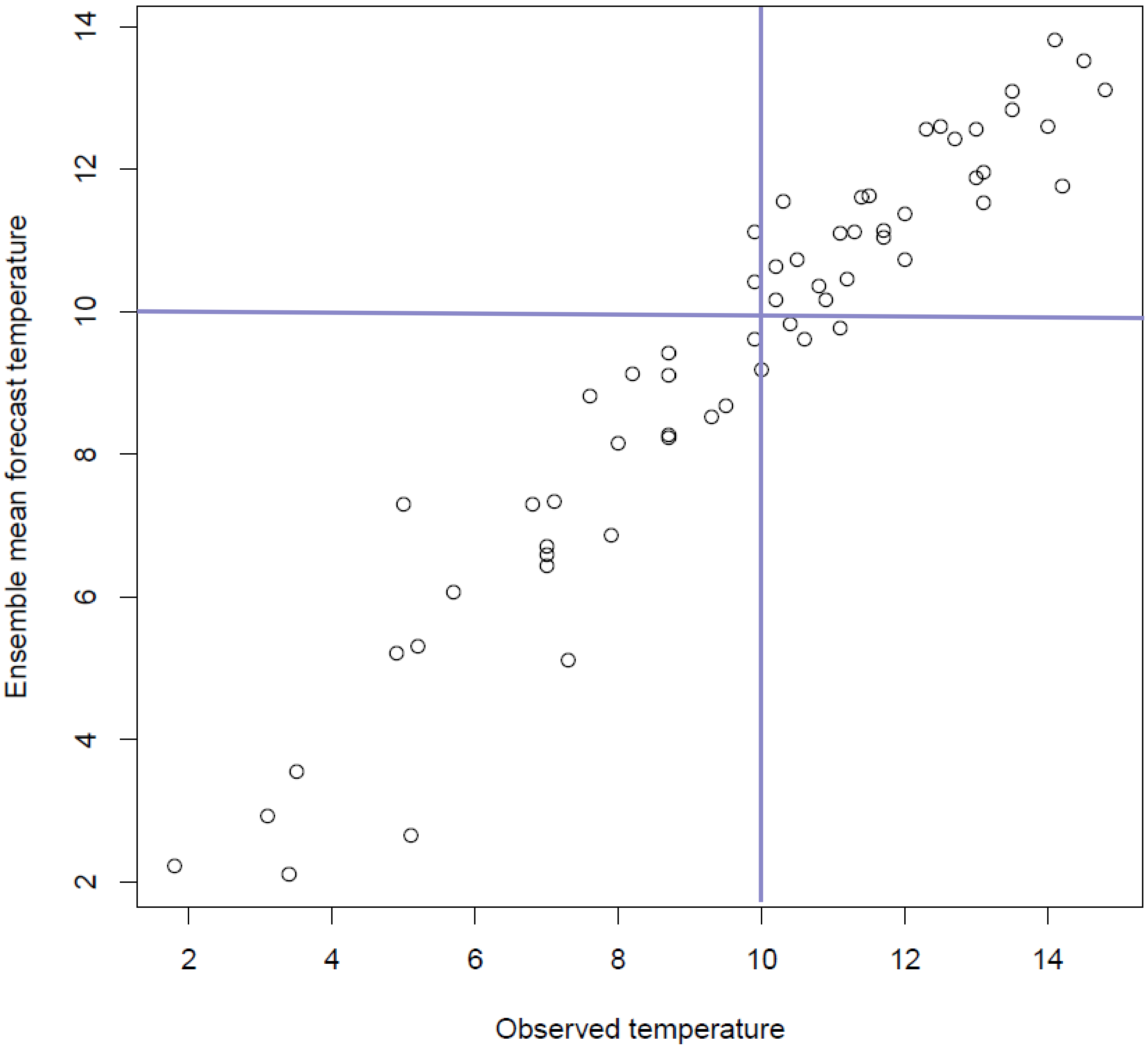
Wind speed: *GDS* = 0.81 (*D* = 0.62)



“December 2015 – an exceptionally mild month in the UK” (Burt & Kendon, *Weather*, 2016)



Dichotomous variables



Dichotomous variables

- If the observations are dichotomous, but the forecasts are continuous,

  GDS is the area under the ROC curve (AUC)
  Somers' *D* is the ROC skill score (ROCSS)

  (Can compute using Mann-Whitney-Wilcoxon rank-sum statistic)

- If both observations *and* forecasts are dichotomous,

  Somers' *D* is the Peirce skill score
  (hit rate – false alarm rate)

Extension to probabilistic forecasts (including ensemble forecasts)

- $\mathrm{sign}(f_i - f_j)$ is no longer meaningful.
- Instead of considering whether the point value of one forecast is greater or less than the value of a second forecast,
- consider the probability that a draw from the first forecast *distribution* is greater than or less than a draw from the second forecast distribution

$$\mathrm{sign}(\Pr[f_i > f_j] - \Pr[f_i < f_j])$$

Extension to probabilistic forecasts
Mason & Weigel 2009 Appendix A

APPENDIX A

The 2AFC for Probabilistic Forecasts

Let $p_{0,i}(x)$ represent the probability density of the ith forecast for an event when an event did not occur, and let $p_{1,j}(x)$ be the jth forecast probability density when an event did occur. The quantities $P_{0,i}(x)$ and $P_{1,j}(x)$ are the corresponding cumulative distributions. Let $X_{0,i}$ be a random sample drawn from $p_{0,i}(x)$, and let $X_{1,j}$ be a random sample drawn from $p_{1,j}(x)$. The probability that $X_{1,j} > X_{0,j}$ $[p(X_{1,j} > X_{0,i})]$ is given by

Extension to probabilistic forecasts

Mason & Weigel 2009 Appendix A:

JANUARY 2009 MASON AN

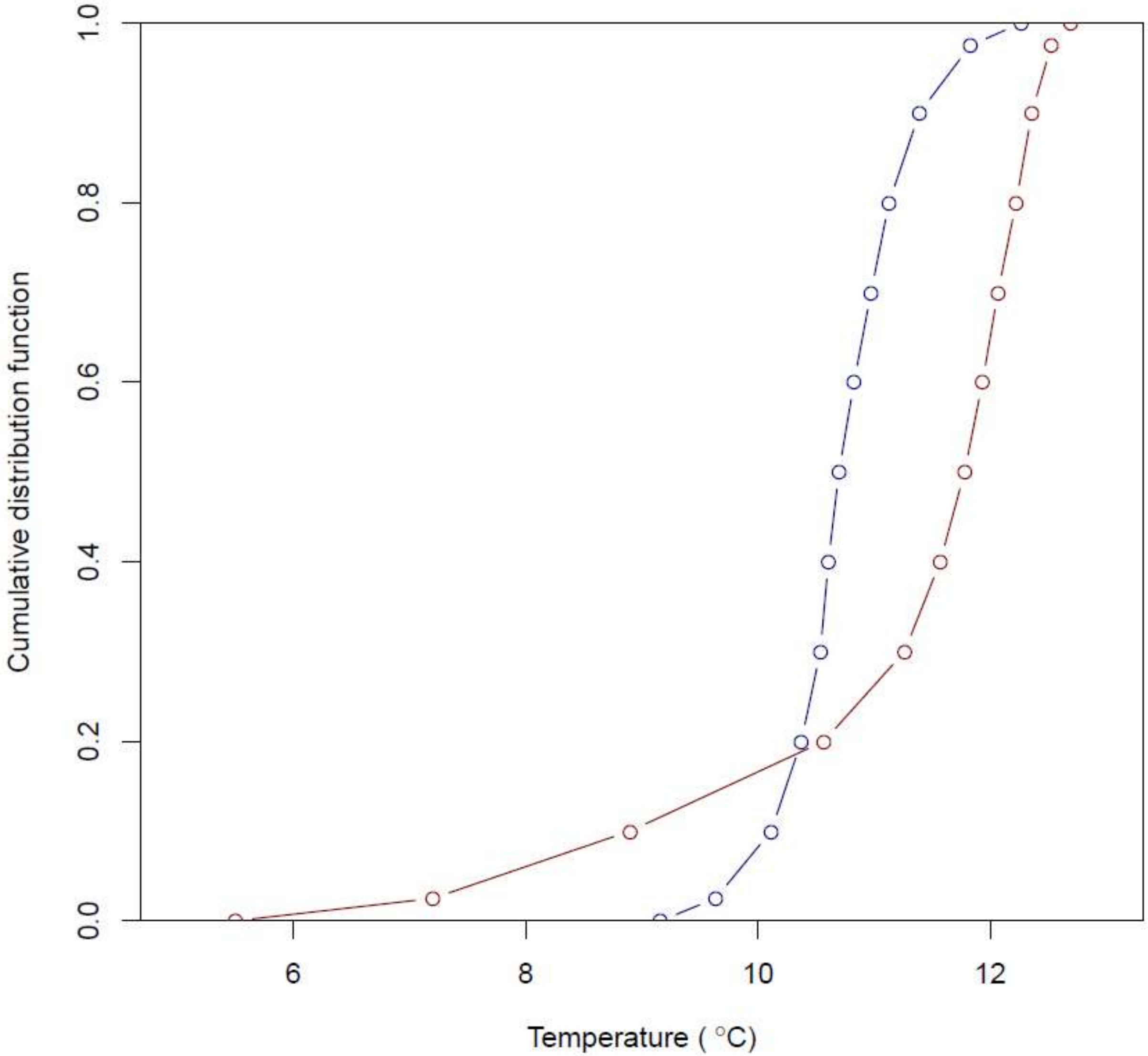
$$p(X_{1,j} > X_{0,i}) = 1 - p(X_{1,j} \le X_{0,i})$$
$$= 1 - \int_{-\infty}^{\infty} p_{0,i}(x) \int_{-\infty}^{x} p_{1,j}(y)\, dy\, dx$$
$$= 1 - \int_{-\infty}^{\infty} p_{0,i}(x) P_{1,j}(x)\, dx. \qquad \text{(A1)}$$

Thus, $p(X_{1,j} > X_{0,i})$ is an obvious basis for a 2AFC score. However, one must consider that, by the nature of the 2AFC test, it is known a priori that the two observations can be discriminated. Therefore, $p(X_{1,j} > X_{0,i})$ needs to be conditioned on the prior knowledge that $X_{1,j} \ne X_{0,i}$. Using Bayes's theorem,

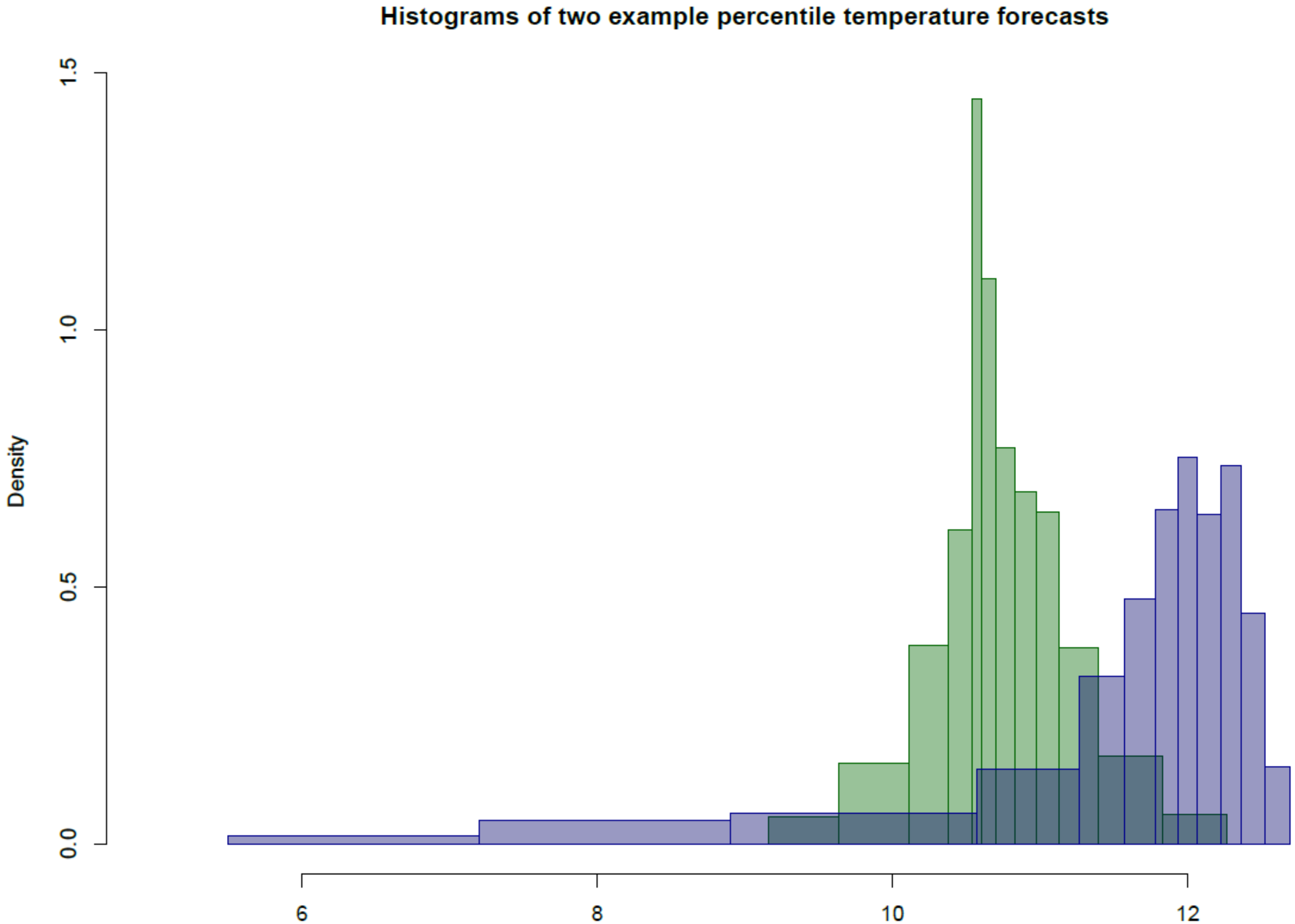
Assumption of independence of probability forecasts

- Not true for forecasts of nearby locations
- Not necessary for ensembles
  - can evaluate $\mathrm{Pr}[f_i > f_j]$ by in the usual way: evaluate $f_i > f_j$ (0/1) in each ensemble member and calculate the mean over ensemble members
- But the dependence information may be lost in post-processing
  - e.g. if we store only quantile-based (percentile) forecasts for each space-time forecast point

Percentile forecasts: cdfs $F(x)$ for two example days



Convert to pdfs *f*(x)



Results (values of *GDS*)
Forecasts of temperature

Forecasts		Observations: Dichotomous	Observations: Continuous
Dichotomous	Ens. median > 10 °C	0.919	—
	50^{th} percentile > 10 °C	0.919	—
Continuous	Ensemble median	0.975	0.918
	50^{th} percentile	0.966	0.921
Probabilistic	Ensemble	0.975	0.917
	Percentile	0.969	0.924

Results (values of *GDS*)
Forecasts of wind speed

Forecasts		Observations: Dichotomous	Observations: Continuous
Dichotomous	Ens. median > 6 m/s	0.805	—
	50th percentile > 6 m/s	0.761	—
Continuous	Ensemble median	0.873	0.807
	50th percentile	0.842	0.780
Probabilistic	Ensemble	0.878	0.810
	Percentile	0.843	0.783

Met Office

Summary

- A rank-based measure of discrimination
- Insensitive to extremes
- Prone to spurious skill (like other discrimination measures)
- Not sensitive to spread of probabilistic forecasts
  - (probabilistic version is not ‘proper’)
- Is it be useful?
  - As a check that post-processing increases it, or at least does not decrease it?