Quantifying the quality of forecasts and forecasting systems

Chris Ferro

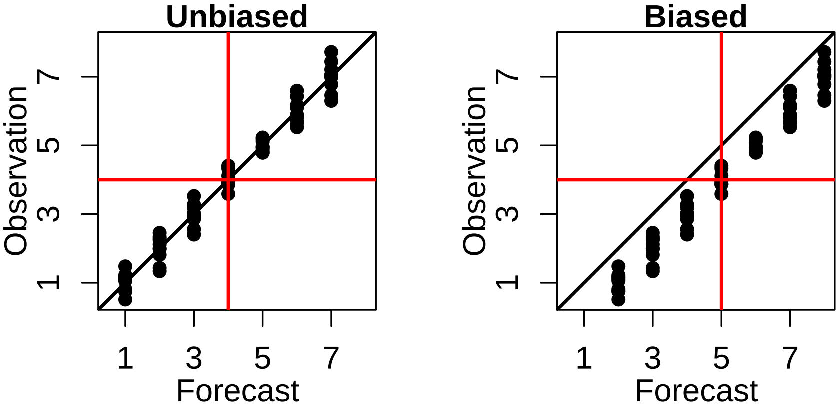
National Centre for Atmospheric Science
University of Exeter

Meteorology Meets Social Science
UK Met Office / 8 June 2007

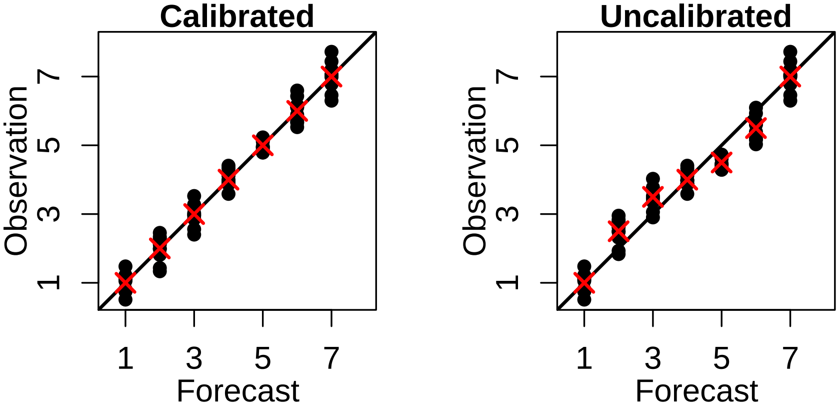
Outline

Outline

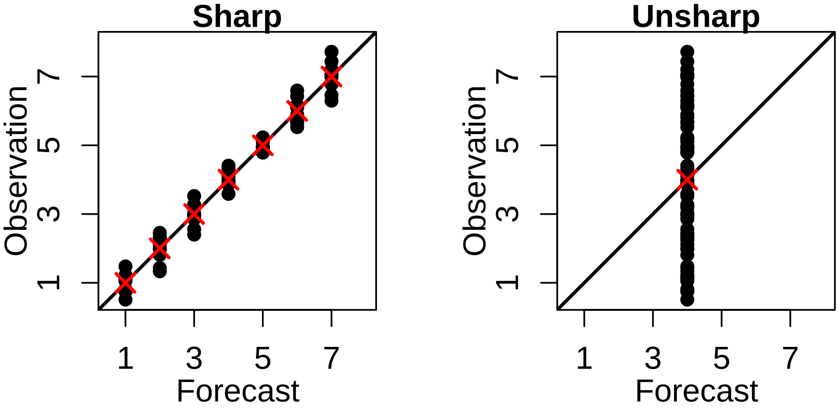
Aspects of forecast quality: bias



Aspects of forecast quality: calibration



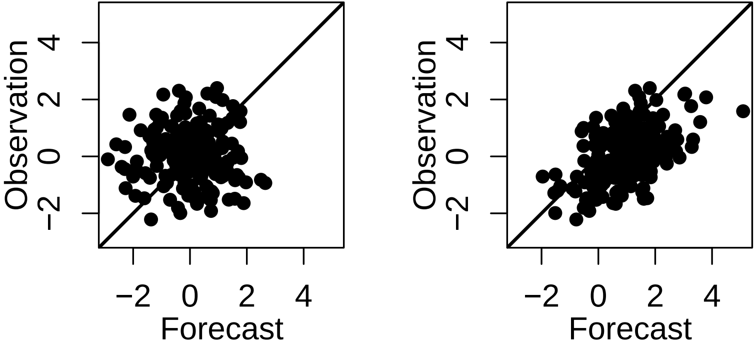
Aspects of forecast quality: sharpness



Mean squared error (MSE)

MSE = variance of observations + variance of forecasts
– 2 x covariance + squared bias

= calibration – resolution + variance of observations



Signal detection theory: ROC analysis

Forecast event if decision variable exceeds threshold.

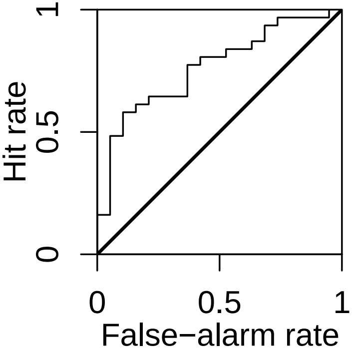
	Observed	Not Obs.	
Forecasted	a	b	$a+b$
Not Forecasted	c	d	$c+d$
	$a+c$	$b+d$	n

Form table and compute

$$\text{Hit rate} = \frac{a}{a+c}$$

$$\text{False-alarm rate} = \frac{b}{b+d}$$

for each possible threshold.



Properties of verification scores

Proper Expected score is optimised by forecasting true, probabilistic belief: discourages hedging.

Consistent Proper, for scores of deterministic forecasts derived from probabilistic forecasts via a rule.

Equitable Expected score is identical for all constant or random forecasts.

Sufficient Forecasts, from which others with equal quality to mine can be derived, score better than mine.

Regular Contours of score on ROC diagram are convex, complete, and pass through (0,0) and (1,1).

Local Score depends on the forecasted probability of the observation only.

How should/do we use verification measures?

Forecast producers

- Systematic assessment can reveal deficiencies...
  ...and possible remedies.
- Prevent hedging: what, why, how?

Forecast users

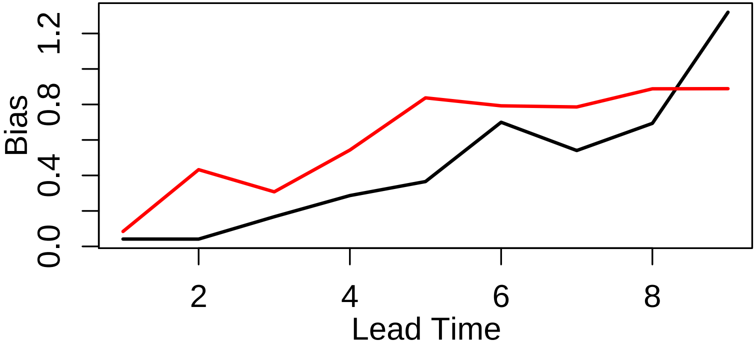
- Some measures can relate directly to value...
  ...perhaps more links can be established.
- How are decisions influenced by overall forecast quality?

Outline

Uncertainty in verification

Incomplete information

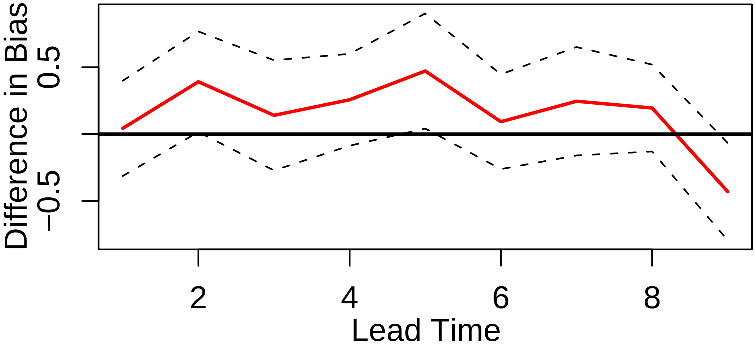
- Assume the sample represents the population
- Compute confidence intervals etc. for the ‘true’ quality
- Avoid using the same data to form and assess forecasts



Uncertainty in verification

Incomplete information

- Assume the sample represents the population
- Compute confidence intervals etc. for the ‘true’ quality
- Avoid using the same data to form and assess forecasts



Uncertainty in verification

Incomplete information

- Better methods for quantifying uncertainty
- What if forecast quality is not stationary?

Observation error

Quality of untried forecasting systems

Quality of systems in untried situations

Other sources of uncertainty?

Summary

- Various aspects of forecast quality
- Careful use of appropriate measures
- Faithful description of uncertainty

Outline

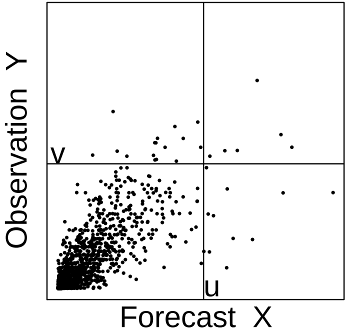
Direct approach

	Observed	Not Obs.	
Forecasted	a	b	$a+b$
Not Forecasted	c	d	$c+d$
	$a+c$	$b+d$	n

$$\text{Hit rate} = \frac{a}{a+c}$$

Forecast if $X > u$

Observe if $Y > v$



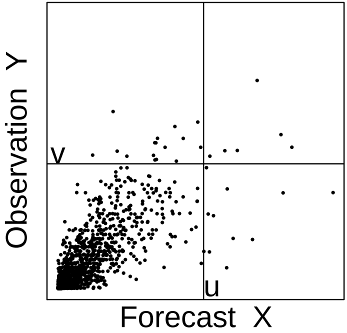
Probability approach

	Observed	Not Obs.	
Forecasted	$\Pr(X > u, Y > v)$	$*$	$\Pr(X > u)$
Not Forecasted	$*$	$*$	$*$
	$\Pr(Y > v)$	$*$	1

Hit rate $= \Pr(X > u \mid Y > v)$

Forecast if $X > u$

Observe if $Y > v$



Probability model

Imagine choosing u so that

$$\Pr(X > u) = \Pr(Y > v) =: p \qquad \text{(base rate)}$$

Extreme-value theory implies

$$\Pr(X > u, Y > v) = \kappa p^{1/\eta} \qquad \text{for small } p$$

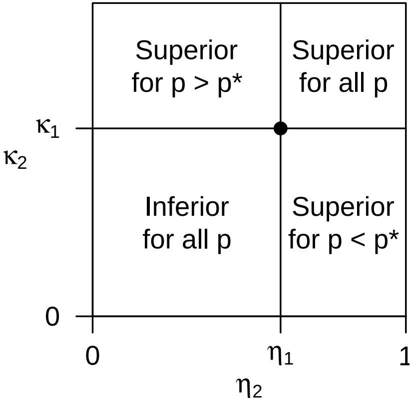
under weak conditions.

Ledford & Tawn (1996, Biometrika)

Interpretation

	Observed	Not Observed	
Forecasted	$\kappa p^{1/\eta}$	$*$	p
Not Forecasted	$*$	$1 - 2p + \kappa p^{1/\eta}$	$*$
	p	$*$	1

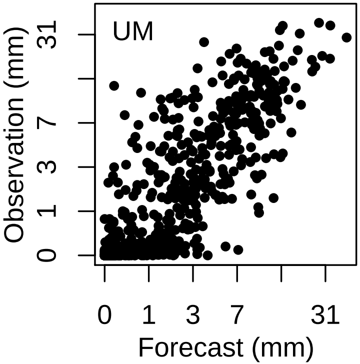
$$\text{Hit rate} \quad = \quad \kappa p^{1/\eta - 1}$$

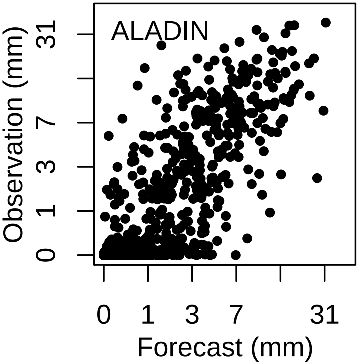


Outline

Daily rainfall: mid-Wales, 1 Jan 05 – 11 Nov 06

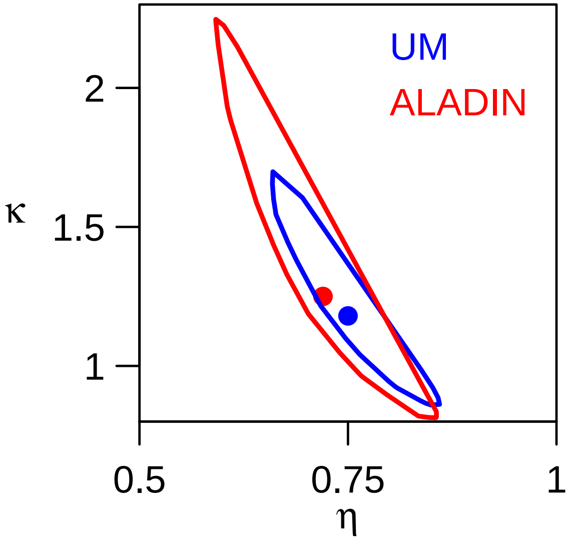
Thanks to Marion Mittermaier



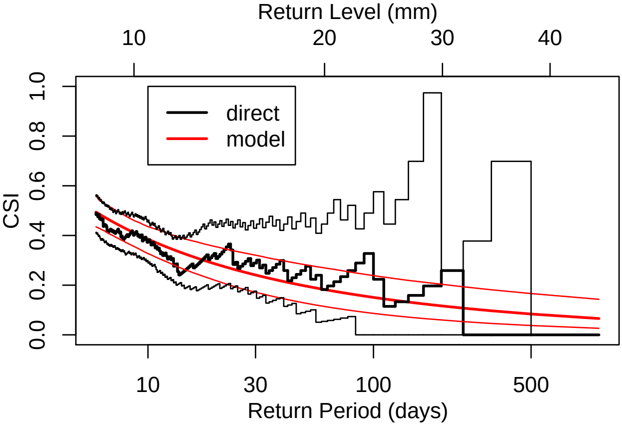


- Maximum-likelihood estimates of η and κ based on ranks
- Threshold choice and model assumptions

Parameter estimates



Verification measures



- Direct estimates degenerate for rare events
- Model estimates change smoothly and are more precise

Summary

- Deterministic forecasts of rare, extreme events
- Only two parameters needed to describe how quality or value of calibrated forecasts changes with base rate
- The model gives more precise estimates of forecast quality

Conclusion

- Statistical models help to identify and measure aspects of forecast quality, their changes and associated uncertainty.
- Why/how should/do producers/users use/do verification?
- Are current methods and procedures adequate?
- Can we verify the quality of decisions?

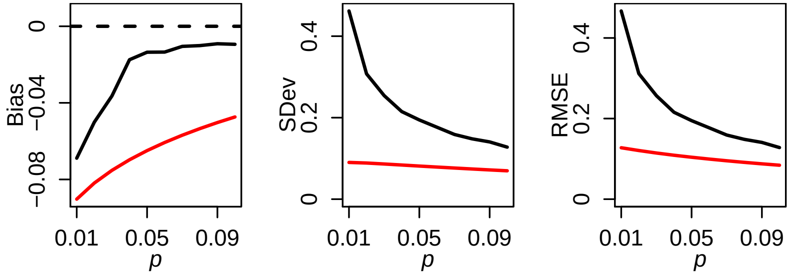
Papers, code and slides available at

www.secam.ex.ac.uk/people/staff/ferro

c.a.t.ferro@exeter.ac.uk

Simulation study

- Bivariate Normal data: correlation 0.8
- Direct and model estimates of hit rate



Model theory – 1

Imagine choosing u so that

$$\Pr(X > u) = \Pr(Y > v) =: p \qquad \text{(base rate)}.$$

Define
$$\begin{aligned} \tilde{X} &= -\log[1 - F(X)] \quad \text{where} \quad F(x) = \Pr(X \leq x) \\ \tilde{Y} &= -\log[1 - G(Y)] \quad \phantom{\text{where}} \quad G(y) = \Pr(Y \leq y) \end{aligned}$$

Then $\tilde{X}$ and $\tilde{Y}$ are Exponential with unit means and

$$\begin{aligned} \Pr(X > u, Y > v) &= \Pr(\tilde{X} > -\log p, \tilde{Y} > -\log p) \\ &= \Pr(Z > -\log p) \end{aligned}$$

where $Z = \min\{\tilde{X}, \tilde{Y}\}$.

Model theory – 2

For $\tilde{X}$ and $\tilde{Y}$ Exponential with unit means and $Z = \min\{\tilde{X}, \tilde{Y}\}$,

$$\Pr(Z > z) = \begin{cases} \exp(-z) & \text{if } \tilde{X} \equiv \tilde{Y} \\ \exp(-2z) & \text{if } \tilde{X} \perp\!\!\!\perp \tilde{Y} \end{cases}$$

Assume

$$\Pr(Z > z) \sim \mathcal{L}(e^z) \exp(-z/\eta) \qquad \text{as } z \to \infty,$$

where $0 < \eta \leq 1$ and $\mathcal{L}(rt)/\mathcal{L}(r) \to 1$ as $r \to \infty$ for all $t > 0$.

e.g. $(X, Y) \sim$ Normal has $\eta = [1 + \mathrm{cor}(X, Y)]/2$.

Ledford & Tawn (1996, Biometrika)

Model theory – 3

$\Pr(Z > z) \sim \mathcal{L}(e^z)\exp(-z/\eta)$ where $\mathcal{L}(rt)/\mathcal{L}(r) \to 1$ as $r \to \infty$.

For a high threshold w_0,

$$\begin{aligned} \Pr(Z > w_0 + z) &\approx \mathcal{L}(e^{w_0+z})\exp[-(w_0+z)/\eta] \\ &\approx \mathcal{L}(e^{w_0})\exp[-(w_0+z)/\eta] \end{aligned}$$

so model

$$\Pr(Z > z) = \kappa \exp(-z/\eta) \qquad \text{for all } z > w_0$$

i.e.

$$\Pr(Z > -\log p) = \kappa\, p^{1/\eta} \qquad \text{for all } p < \exp(-w_0).$$

Limiting behaviour of measures

$$\text{Hit rate} = \frac{a}{a+c} \sim \kappa p^{1/\eta - 1} \to \begin{cases} 0 & \text{if } \eta < 1 \\ \kappa & \text{if } \eta = 1 \end{cases}$$

$$\text{PC} = \frac{a+d}{n}, \qquad \text{PSS} = \frac{ad-bc}{(a+c)(b+d)}, \qquad \text{OR} = \frac{ad}{bc}$$

	$\eta < \frac{1}{2}$	$\eta = \frac{1}{2}$	$\eta > \frac{1}{2}$	$\eta = 1$
PC	$1 - 2p \uparrow 1$	$1 - 2p \uparrow 1$	$1 - 2p \uparrow 1$	$1 - 2\bar{\kappa}p \uparrow 1$
PSS	$-p \uparrow 0$	$-\bar{\kappa}p \updownarrow 0$	$\kappa p^{\delta - 1} \downarrow 0$	$\kappa - \bar{\kappa}p \uparrow \kappa$
OR	$\kappa p^{\delta - 2} \downarrow 0$	$\kappa - 2\kappa\bar{\kappa}p \updownarrow \kappa$	$\kappa p^{\delta - 2} \uparrow \infty$	$\kappa/(\bar{\kappa}^2 p) \uparrow \infty$

where $\delta = 1/\eta$ and $\bar{\kappa} = 1 - \kappa$

Limiting behaviour of measures

$$\text{Hit rate} = \frac{a}{a+c} \sim \kappa p^{1/\eta - 1} \to \begin{cases} 0 & \text{if } \eta < 1 \\ \kappa & \text{if } \eta = 1 \end{cases}$$

$$\text{PC} = \frac{a+d}{n}, \qquad \text{PSS} = \frac{ad - bc}{(a+c)(b+d)}, \qquad \text{OR} = \frac{ad}{bc}$$

	$\eta < \frac{1}{2}$	$\eta = \frac{1}{2}$	$\eta > \frac{1}{2}$	$\eta = 1$
PC	$1 - 2p \uparrow 1$	$1 - 2p \uparrow 1$	$1 - 2p \uparrow 1$	$1 - 2\bar{\kappa}p \uparrow 1$
PSS	$-p \uparrow 0$	$-\bar{\kappa}p \updownarrow 0$	$\kappa p^{\delta-1} \downarrow 0$	$\kappa - \bar{\kappa}p \uparrow \kappa$
OR	$\kappa p^{\delta-2} \downarrow 0$	$\kappa - 2\kappa\bar{\kappa}p \updownarrow \kappa$	$\kappa p^{\delta-2} \uparrow \infty$	$\kappa/(\bar{\kappa}^2 p) \uparrow \infty$

where $\delta = 1/\eta$ and $\bar{\kappa} = 1 - \kappa$

Limiting behaviour of measures

$$\text{Hit rate} = \frac{a}{a+c} \sim \kappa p^{1/\eta - 1} \rightarrow \begin{cases} 0 & \text{if } \eta < 1 \\ \kappa & \text{if } \eta = 1 \end{cases}$$

$$\text{PC} = \frac{a+d}{n}, \qquad \text{PSS} = \frac{ad - bc}{(a+c)(b+d)}, \qquad \text{OR} = \frac{ad}{bc}$$

	$\eta < \frac{1}{2}$	$\eta = \frac{1}{2}$	$\eta > \frac{1}{2}$	$\eta = 1$
PC	$1 - 2p \uparrow 1$	$1 - 2p \uparrow 1$	$1 - 2p \uparrow 1$	$1 - 2\bar{\kappa}p \uparrow 1$
PSS	$-p \uparrow 0$	$-\bar{\kappa}p \updownarrow 0$	$\kappa p^{\delta-1} \downarrow 0$	$\kappa - \bar{\kappa}p \uparrow \kappa$
OR	$\kappa p^{\delta-2} \downarrow 0$	$\kappa - 2\kappa\bar{\kappa}p \updownarrow \kappa$	$\kappa p^{\delta-2} \uparrow \infty$	$\kappa/(\bar{\kappa}^2 p) \uparrow \infty$

where $\delta = 1/\eta$ and $\bar{\kappa} = 1 - \kappa$

Contradictory skill scores?

ERA-40 daily rainfall forecasts: $\eta = 0.81$, $\kappa = 1.16$