Verification of Extreme Weather

Dr David Stephenson and Dr Chris Ferro

University of Reading, U.K.

Thanks to:
Barbara Casati,
Caio Coelho,
Ian Jolliffe,
Paco Doblas-Reyes

Oldest known photograph of a tornado 28 August 1884,
22 miles southwest of Howard, South Dakota

NOAA Historic NWS Collection www.photolib.noaa.gov

What do we mean by “extreme”?

- Large meteorological values
  - Maximum value (i.e. a local extremum)
  - Exceedance above a high threshold
  - Record breaker (threshold=max of past values)

- Rare event

  (e.g. less than 1 in 10 years – p=0.1)

- Large losses (*severe* or *high-impact*)

  (e.g. $200 billion if hurricane hits Miami)

  risk = p(hazard) x vulnerability x exposure

Types of extreme weather forecast

- Occurrence of event

e.g. more than 5mm rain/day tomorrow,
probability of more than 5mm rain/day tomorrow

- Number of events

e.g. expected number of hurricanes to make landfall in a season,
distribution of hurricane counts in a season

- Time-to-next event

e.g. probability of another hurricane within the next week

- Magnitude/intensity of the event

e.g. return level (lower quantile) or return period (1/exceedance probability)

Some issues and possible solutions

Issue: Trivial limits of scores for rarer events

Solution: Develop new scores e.g. Extreme Dependency Score

Issue: Larger sampling uncertainty (less events!)

Solution: Extrapolate from less rare events, Pool events

Issue: Small count problems (e.g. 0's in tables)

Solution: Extrapolate from less rare events, Pool events

Issue: Hedging to avoid missing events

Solution: Issue probability forecasts and use proper scores

Issue: Large outlier values for intensities

Solution: Use resistant scores (e.g. based on rank statistics)

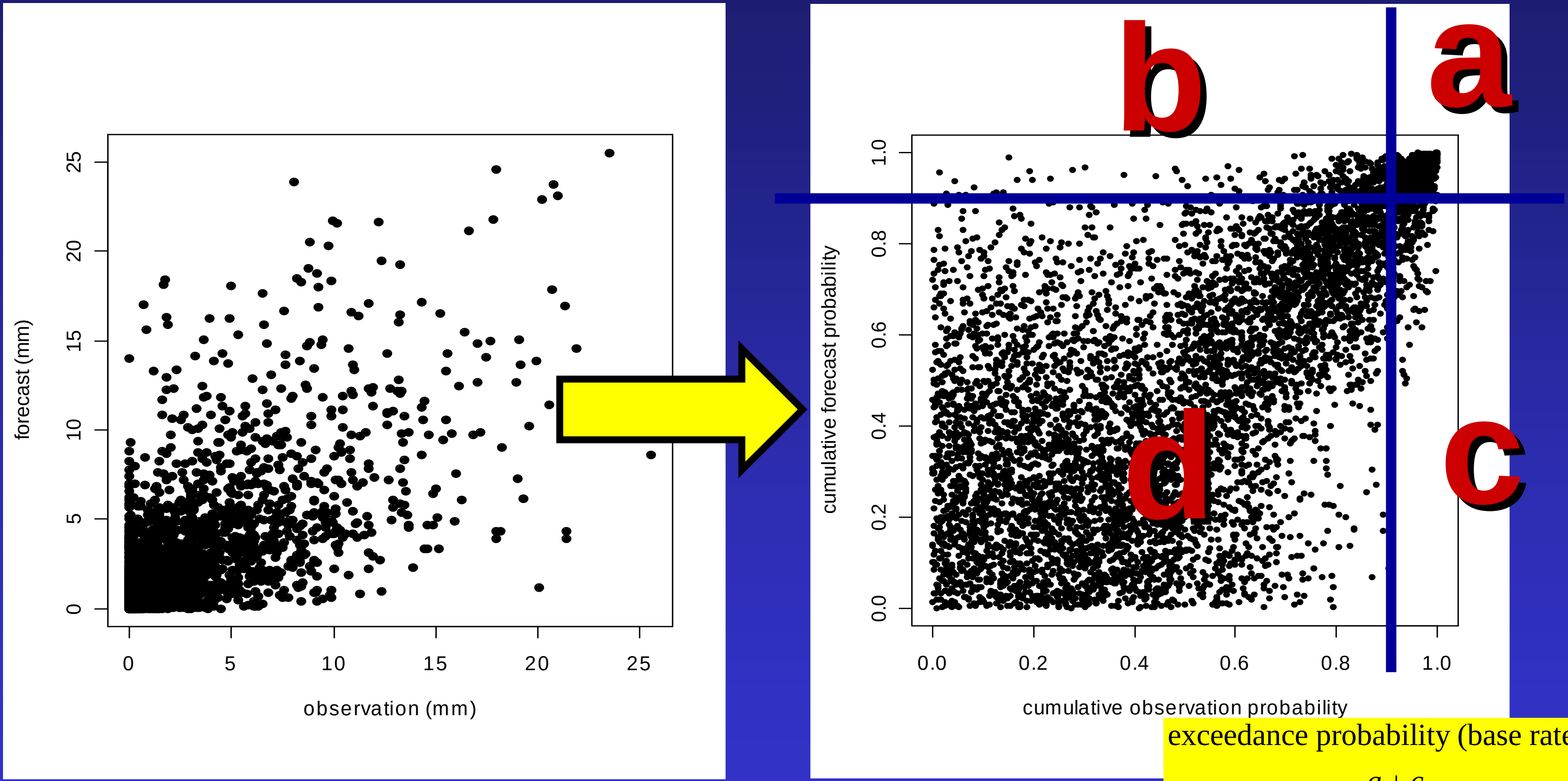
Some scores for extreme weather

Forecast type	Deterministic	Probabilistic
Occurrence of event	Threat scores, EDS, ROC	Brier Score, Logarithmic score
Number of events	MSE of sqrt(counts)	???
Time-to-next-event	MSE of sqrt or log of interval	???
Magnitude/intensity of event	MSE (non-resistant!)	Coverage, Interval scores

→ Need to develop scores for extreme weather forecasts!!!

Example: rainfall forecasts versus observations

Met Office mesoscale model forecasts of 6h ahead 6h Eskdalemuir precipitation amounts (4x times daily)   Total sample size n=6226



exceedance probability (base rate):

$$\Pr(X > 5mm) = \frac{a+c}{n} = p$$

→ note dependency for extreme events in top right hand corner

Trivial limits of scores

$$\text{Proportion correct} = (a+d)/n \sim 1 - p(1+B) \to 1$$

$$\text{Threat score} = \frac{a}{a+b+c} \sim \frac{hp^k}{1+B-hp^k} \to 0 \text{ as } p \to 0$$

$$\text{Heidke skill score} = \frac{a+d-a_r-d_r}{n-a_r-d_r} = \frac{2(H-pB)}{1+B-2pB} = \frac{2ETS}{1+ETS} \to 0$$

$$\text{Peirce skill score} = \frac{a}{a+c} - \frac{b}{b+d} = H - F \sim hp^k - Bp \to 0$$

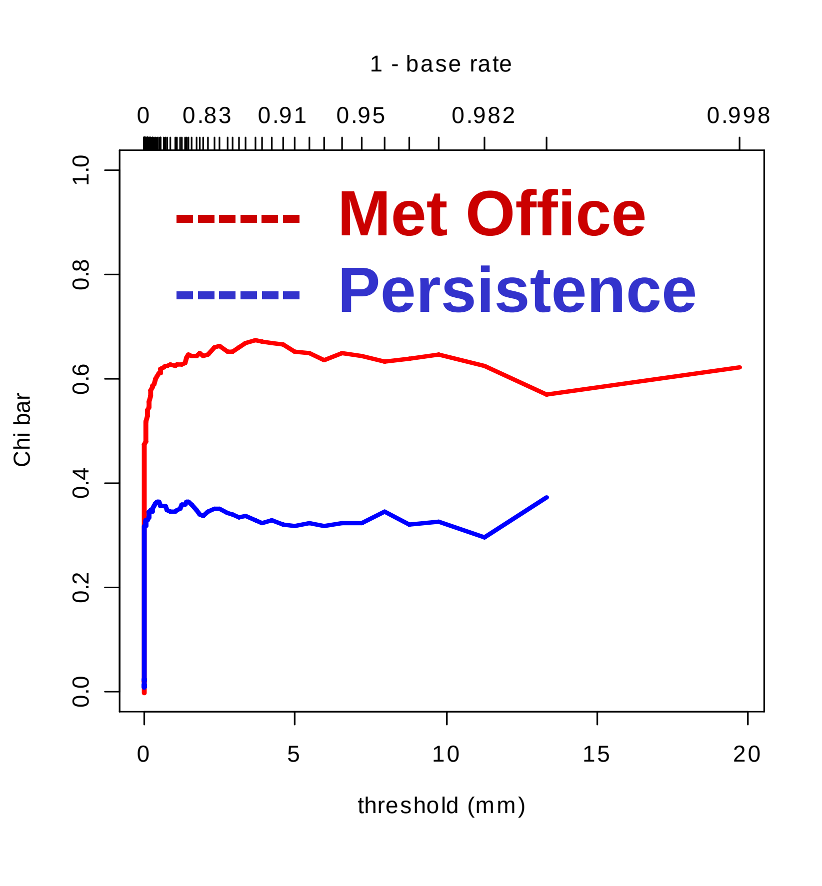
Does this mean that extreme events are really more difficult to forecast than small magnitude events?

Extreme Dependency Score $\overline{\chi}$

$$EDS = \frac{2\log((a+c)/n)}{\log(a/n)} - 1$$

$$\sim \frac{1-k}{1+k} \text{ as } p \to 0$$

- does not tend to zero for rare events
- not explicitly dependent on bias B
- measure of the exponent:
  k=(1-EDS)/(1+EDS)



Coles, S. Heffernan, J. and Tawn, J. (1999) Dependence measures for extreme value analyses, *Extremes* **2** 339-365.
http://www.maths.lancs.ac.uk/~currie

Why issue probability forecasts?

Advantages:

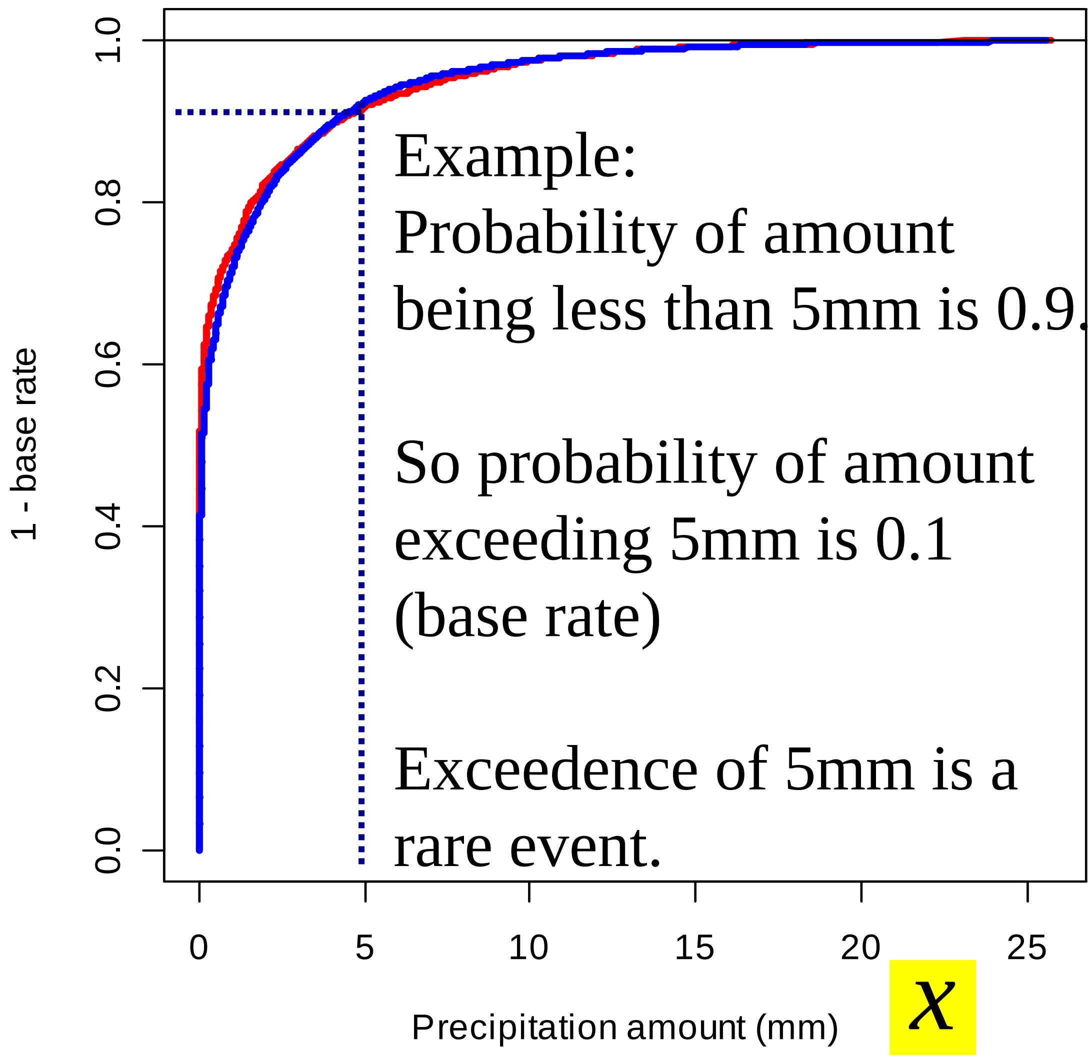
- Honest statement about the UNCERTAIN future
- Allows decision-makers to quantify their risk
- Harder to hedge than deterministic forecasts
  Murphy, A., 1991: “Probabilities, Odds, and Forecasts of Rare Events”, Weather and Forecasting, Vol. 6, 302-307

Disadvantages:

- More information to communicate
- Can sound imprecise and confusing
- Harder to understand (risk understanding/perception)
- User has to make a decision on when to act

Empirica n

$$F(x) = \Pr(X \le x)$$
$$= 1 - p$$
$$p = \text{"base rate"}$$



→ can use E.D.F. to map values onto probabilities (unit margins)

History of Threat Scores

$$TS = \frac{a}{a+b+c}$$

- Gilbert (1884) - “ratio of verification”(=Threat Score)

  “ratio of success in forecasting”(=ETS)
- Palmer and Allen (1949) - “threat score” TS
- Donaldson et al. (1975) - “critical success index”(=TS)
- Mason (1989) – base rate dependence of CSI(=TS)
- Doswell et al. (1990) – HSS→2TS/(1+TS)
- Schaefer (1990) – ETS=HSS/(2-HSS)
- Stensrud and Wandishin (2000) – “correspondence ratio”

→ Threat score ignores correct rejections (d) and is strongly dependent on the base rate.

Asymptotic model for contingency table

	Obs=Yes	Obs=No	Marginal $\sum$
Fcst=Yes	a=npH	b=np(B-H)	a+b=npB
Fcst=No	c=np(1-H)	d=n-np(1+B-H)	c+d=n(1-pB)
Marginal $\sum$	a+c=np	b+d=n(1-p)	n

p = prob. of event being observed (base rate)

B = forecast bias (B=1 for unbiased forecasts)

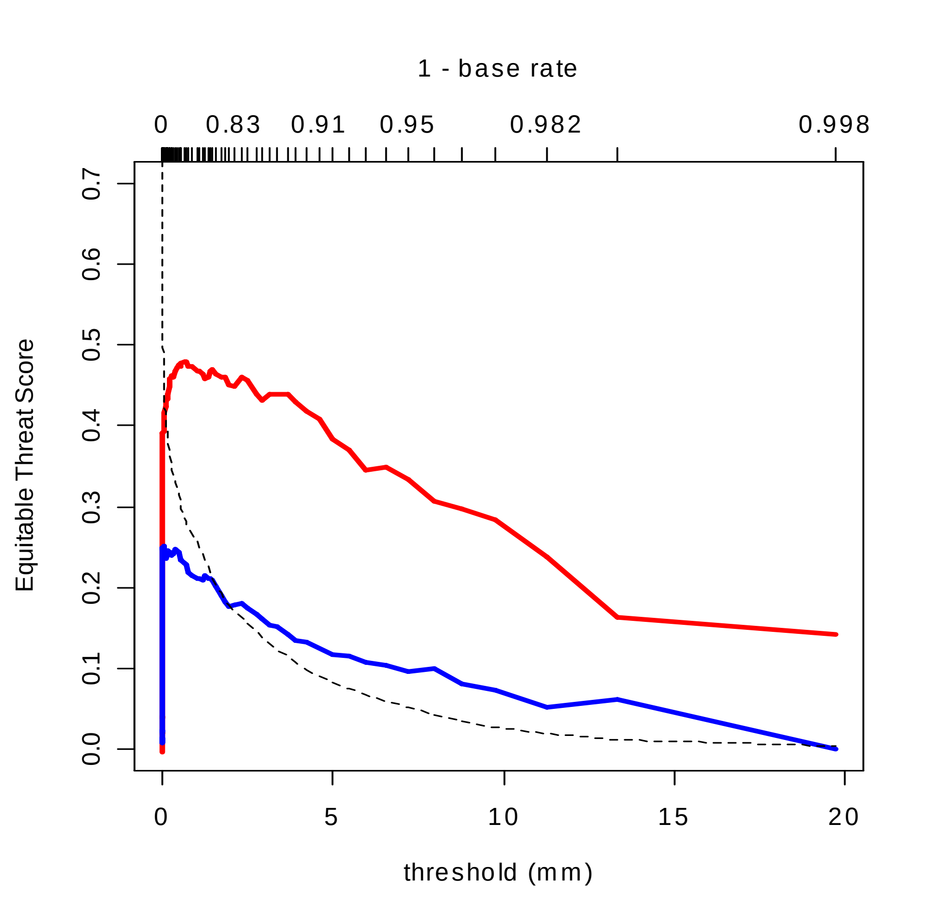
H = hit rate = a/(a+c)

Assume that in the p→0 limit : H→0 and so **$H \sim hp^k$**

E.g. random forecasts: H=Bp so h=B and k=1

Equitable Threat Score (Gilbert Skill Score)

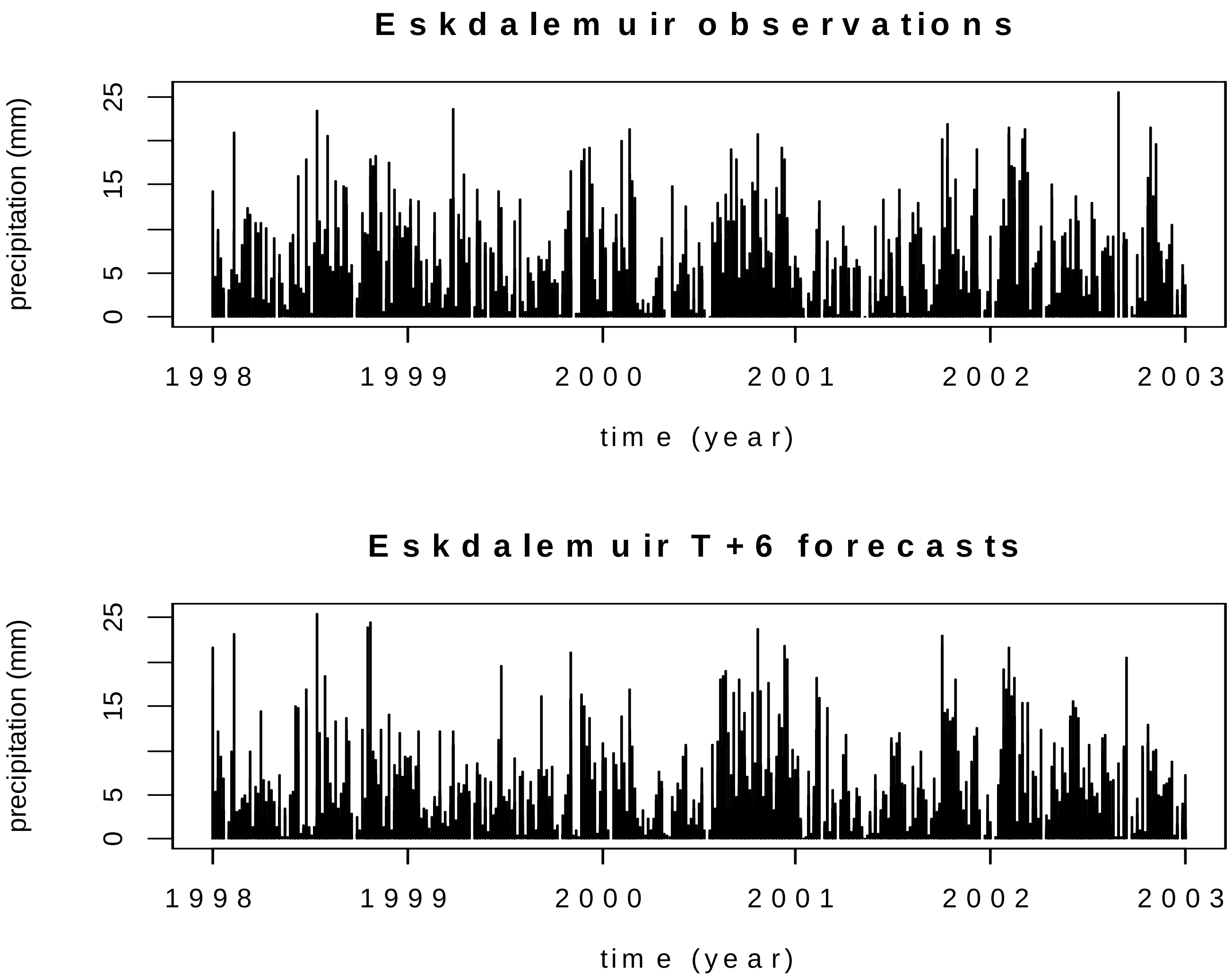
$$ETS = \frac{a - a_r}{a + b + c - a_r} \sim \frac{hp^k - pB}{1 + B - hp^k - pB} \rightarrow 0$$



→ ETS tends to zero as base rate p→0 → no skill for extremes

Forecasts of 6 hourly rainfall totals

In collaboration with Dr Barbara Casati & Dr Clive Wilson



Eskdalemuir, Scotland