

Verification of extreme events



b. casati @ gmail.com

AKNOWLEDGEMENTS: ERIC GILLELAND & RICK KATZ



Talk outline

1. Introduction: extremes and their characteristics
2. Extremes in climate studies: Extreme Value Theory
3. Extremes and Forecast Verification:
 - Behavior of categorical scores for extremes
 - The Extreme Dependency Score
 - Complex extreme events
4. Conclusions

What is an extreme event?

Extreme events can be defined by:

- Maxima/minima
- Magnitude
- Rarity
- Impact/losses

Definition: **extreme/rare events** are the events in the tail of the distribution

“Man can believe the impossible, but man can never believe the improbable.” - Oscar Wilde

Gare Montparnasse, 22 October 1895



Why focus on extreme events ?

- **User prospective:** high social and economical impacts (\$)
- **Climate change projections:** increase in extreme frequency and intensities
- **Weather forecast:** model resolution got to finer temporal and spatial scale, which can resolve (some) extreme events

Verification of extreme events:

- understand capabilities of our models in predicting extreme events
- help decision maker and mitigate (hopefully) losses
 - extreme event forecast + communication + reaction = time constraints make challenging (e.g. Katrina)
 - climate change impacts and adaptations

Should forecasts of extreme events be probabilistic ?

YES: large uncertainties + value/risk assessment by cost-loss model

Extreme events statistical challenges

Rarity:

- **small sample, large uncertainty:** data **pooling** in time and space can help, however need to account for inhomogeneities and non-stationarity (Vincent and Mekis 2006); Regional Frequency Analysis (Fowler and Kilsby 2003, Wigley et al 1984).
- use of **moderate extremes: infer behavior** of more-extreme events from larger sample (e.g. ECMWF seasonal forecasts, use exceedence of 0.15 and 0.85 percentiles)
- small or zero counts when stratifying events in categories (Agresti 2002): **statistics unstable behavior, oversensitivity to bias, non-informative asymptotic limits**

Rarity + Magnitude:

small samples + large values **need robust statistical approaches**

Extremes in climate: indices

Model and obs trends:

1. Definition of indices related to extreme weather (Fritsch et al 2002)

2. Evaluation of indices trends (present vs future model run), the geographic location (e.g. Tebaldi et al 2006) and indices distribution (e.g. Alexander et al 2006)

3. evaluation of indices trends at obs location (e.g. Vincent and Mekis, 2006)

Temperature Indices

frost days = n of days with $T_{min} < 0$

cold days = n of days with $T_{max} < 10\%ile$

cold nights = n of days with $T_{min} < 10\%ile$

summer days = n of days with $T_{max} > 25C$

warm days = n days with $T_{max} > 90\%ile$

warm nights = n days with $T_{min} > 90\%ile$

Diurnal T range = average $T_{max} - T_{min}$

Standard deviation of T_{mean}

Precipitation Indices

total annual snowfall accumulation

total snow to total precipitation ratio

days with precipitation

days with rain

average precipitation intensity for precip days

average rainfall intensity for rain days

max number of consecutive dry days

highest 5-days precip accumulation amount

Very wet days = n of days with $P > 95\%$

Heavy precipitation days ($P > 10 \text{ mm}$)

Extremes in Climate: Extreme Value Theory (EVT)

Strength: extreme events are rare, small samples
EVT enables to fit theoretical distributions to these small samples:
robust approach which enables to infer properties of very large
extremes (even if not belonging to the actual sample)

Key References for EVT:

Coles (2001) *An introduction to Statistical Modeling of Extreme Values*, Springer

Naveau et al (2005) *Statistical methods for the analysis of climate extremes*, Comptes Rendus Geoscience, 337: 1013-1022

Katz et al (2002) *Statistics of Extremes in hydrology*, Advances in Water Resources, 25: 1287-1304

R. Smith, Lecture notes on environmental statistics, chapter 8

Rick Katz and Eric Gilleland web page

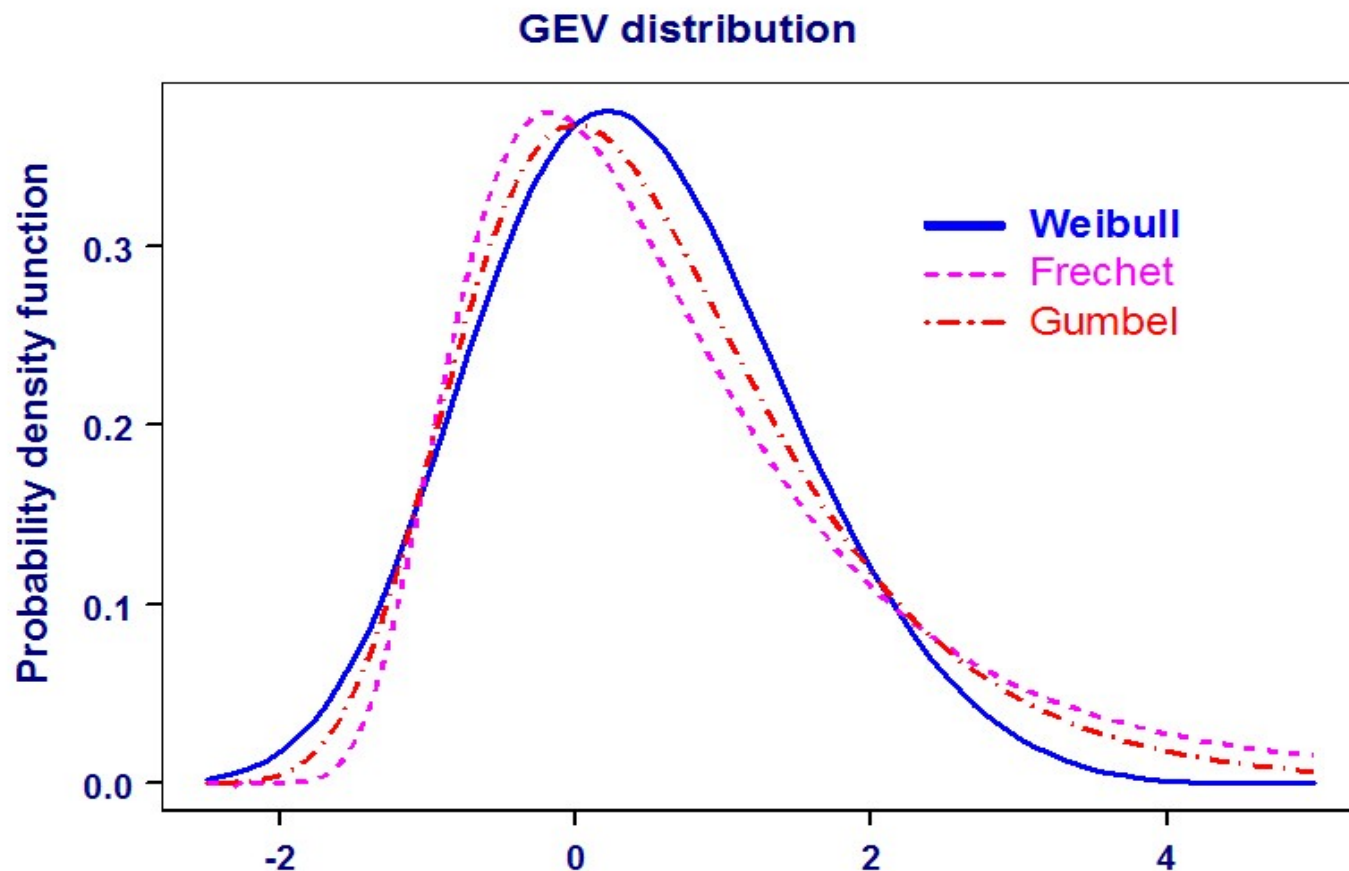
<http://www.isse.ucar.edu/extremevalues/extreme.html>

Block maxima
(e.g. max annual
temperature) are
distributed as a
Generalized
Extreme Value
(GEV) distribution

NOTE: the larger
is the block, the
more extreme is
the value, the
faster the
convergence (but
the sample gets
smaller): trade-off !

Pooling or
r-block maxima

EVT: block maxima

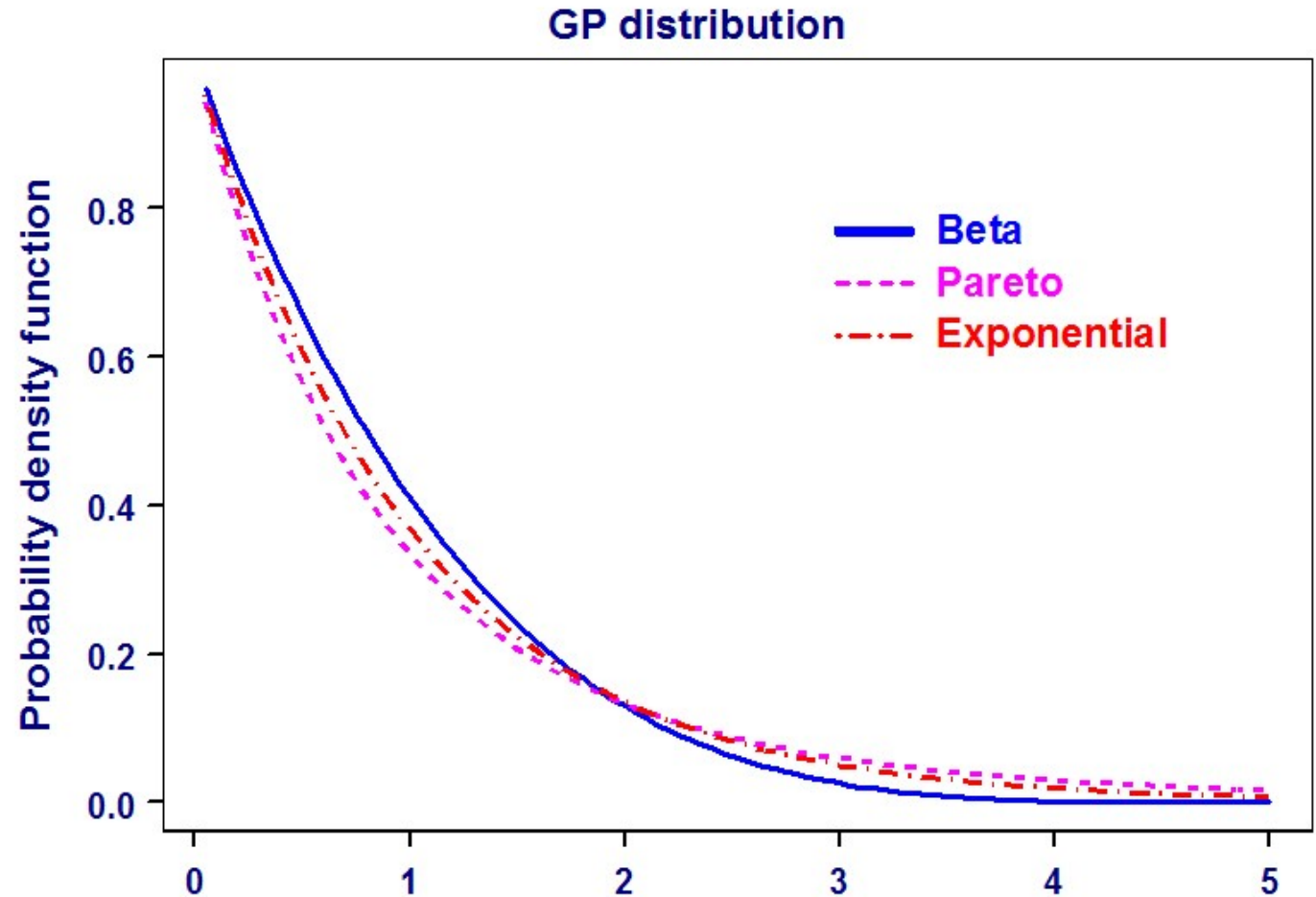


GEV distributions are characterised by a
shape (tail), **location** and **scale** parameter

EVT: Peak over Threshold (PoT)

Tail of the distribution: values exceeding a high thresholds are distributed as a Generalized Pareto (GP) distribution

NOTE: more information is available, however we need to account for dependence between events (EVT assumes iid events): clustering



GP distributions are characterised by a **shape** (tail) and **scale** parameter (optimal exceedence threshold can be determined: lowest threshold for which shape and scale are constant)

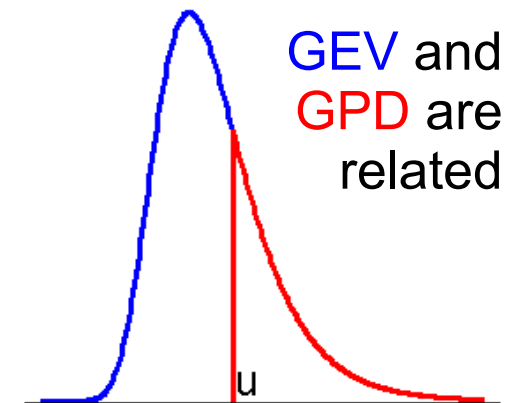
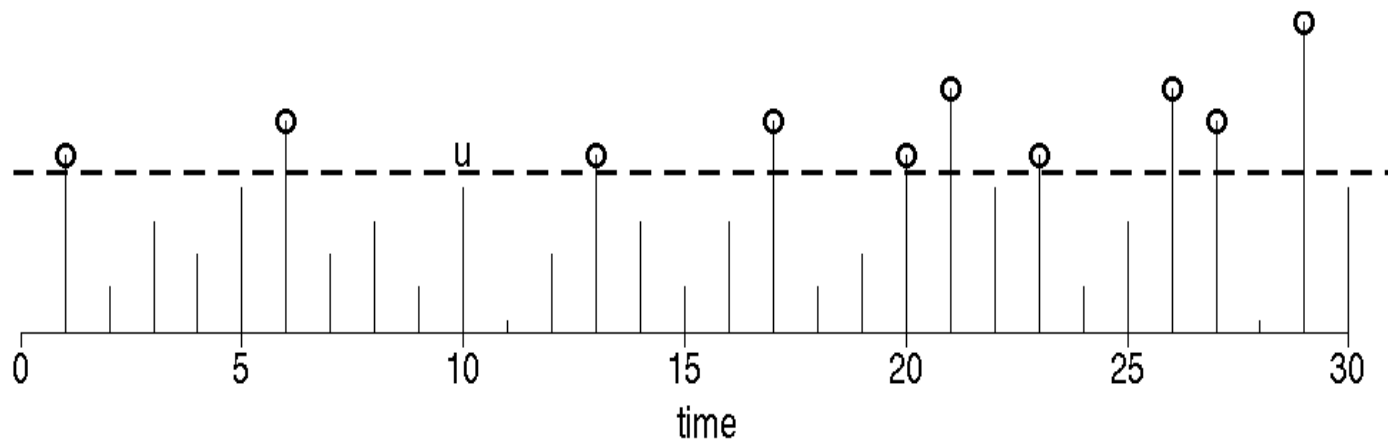
Block maxima or Peak over Thresholds ?

Block Maxima = GEV is simpler (iid, no threshold); L-moments parameter estimation is robust and fast. Bootstrap for Confidence Intervals (CIs)

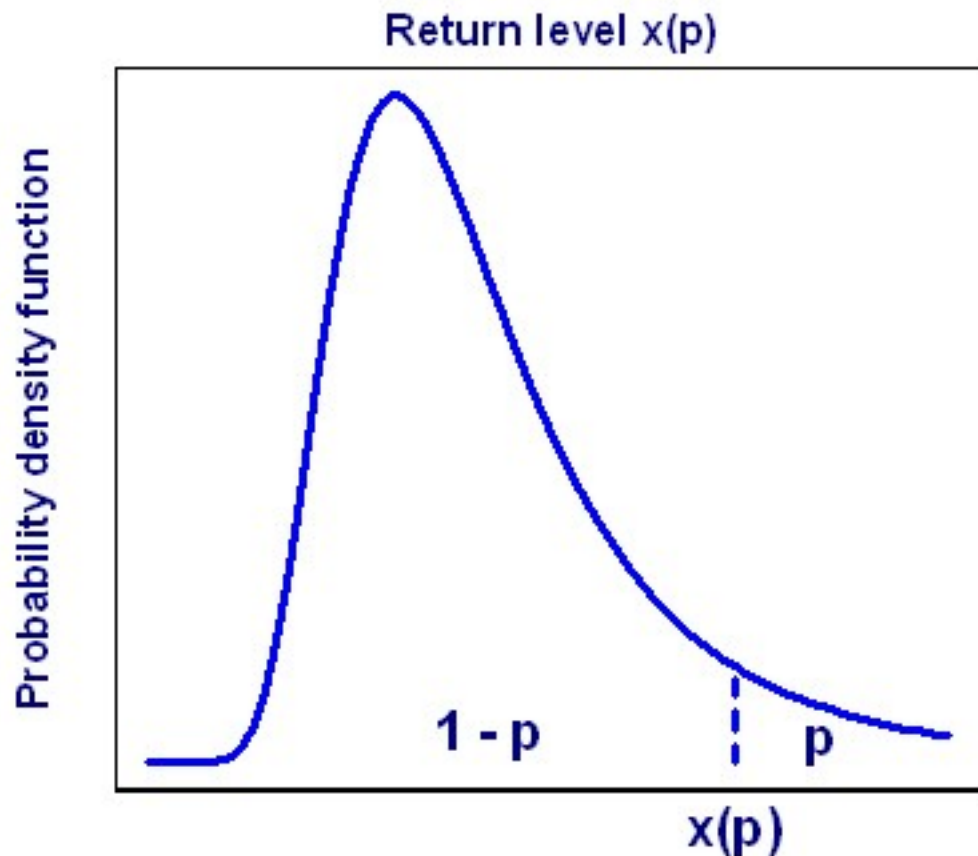
PoT = GPD use more information than block maxima (more robust results), but:

1. PoT = GPD is more complex theoretically

- Point process: event occurring with frequency k (e.g. exceedence of threshold u)
- Poisson distribution: model occurrence of exceedence of threshold u
- GPD: model the magnitude of the exceedence of the threshold



2. need to account for dependency in sample data: clustering or model dependency in parameter estimation (e.g. Montecarlo for time dependency). Maximum Likelihood Estimation (MLE): more complex than L-moments, but allow data dependency, physical covariates, model annual cycles, time evolution of parameters: more appropriate for Climate Change. Moreover, MLE allow probability model for CIs



A return value corresponding to a return period of 25 years is the magnitude of the event occurring with probability $p = 1/25$

Return Values and Return Periods

- Fit GEV to extreme events for present and future climate
- Evaluate return values for different return periods
- Analyze return values trends for different geographical locations
- Uncertainty on trends bootstrapping change in return values from a multimodel ensemble

e.g. Fowler et al (2007), Karhin, Zweirs, Zhang (2007)

Summary: extremes in climate

Use of **Extreme Value Theory** enable **robust diagnosis of extreme behaviors** (e.g. extreme intensity and frequency increase)

GEV or GPD provide information on marginal distribution, no joint distribution. Spatial matching is assessed by regional analysis: good enough for Climate, but **weather forecasting need bivariate distribution from EVT**

NOTE: not many studies perform observations and climate model direct comparison (all assess trends in either obs or model projections). **Climate model resolution ~ 50 km: certain kinds of extremes are not really resolved !** Up-scaling or downscaling needed for direct comparison ...

Extremes in weather forecasts: categorical scores for extreme and rare events verification

Analysis of the behaviour of categorical scores in extreme rare event situations:

Schaefer (1990): ETS converge to TS

Doswell et al (1990): HKD converges to H

Marzban (1998): bias over-sensitivity for extremes

Goeber (2004): trivial limits as events get rarer, odds ratio

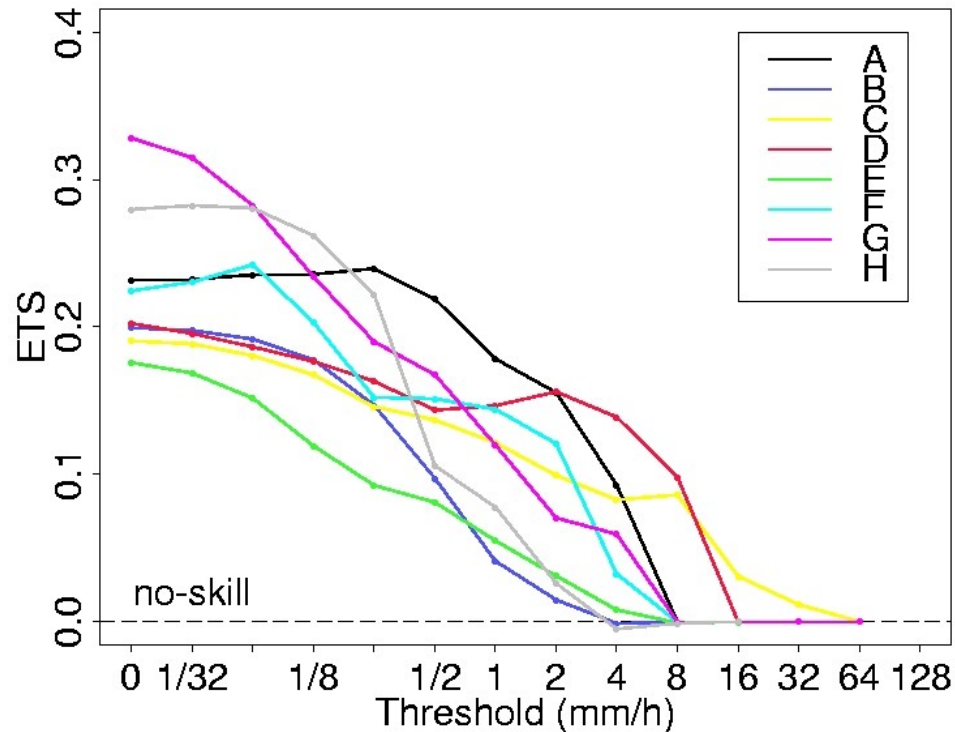
Coles et al (1999): bivariate distributions in extreme value theory

Stephenson et al (2008): the Extreme Dependency Score (EDS)

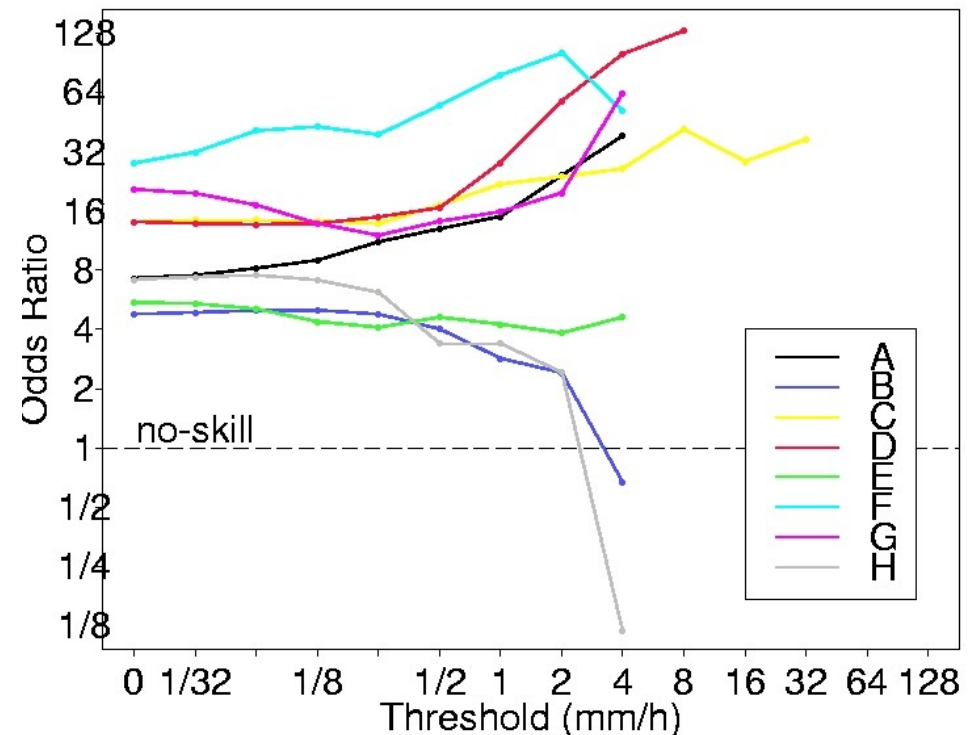
Ferro (2007): probability model for the EDS from EVT

Present a summary following Casati (2004), PhD thesis, chapter 6

Categorical scores versus threshold



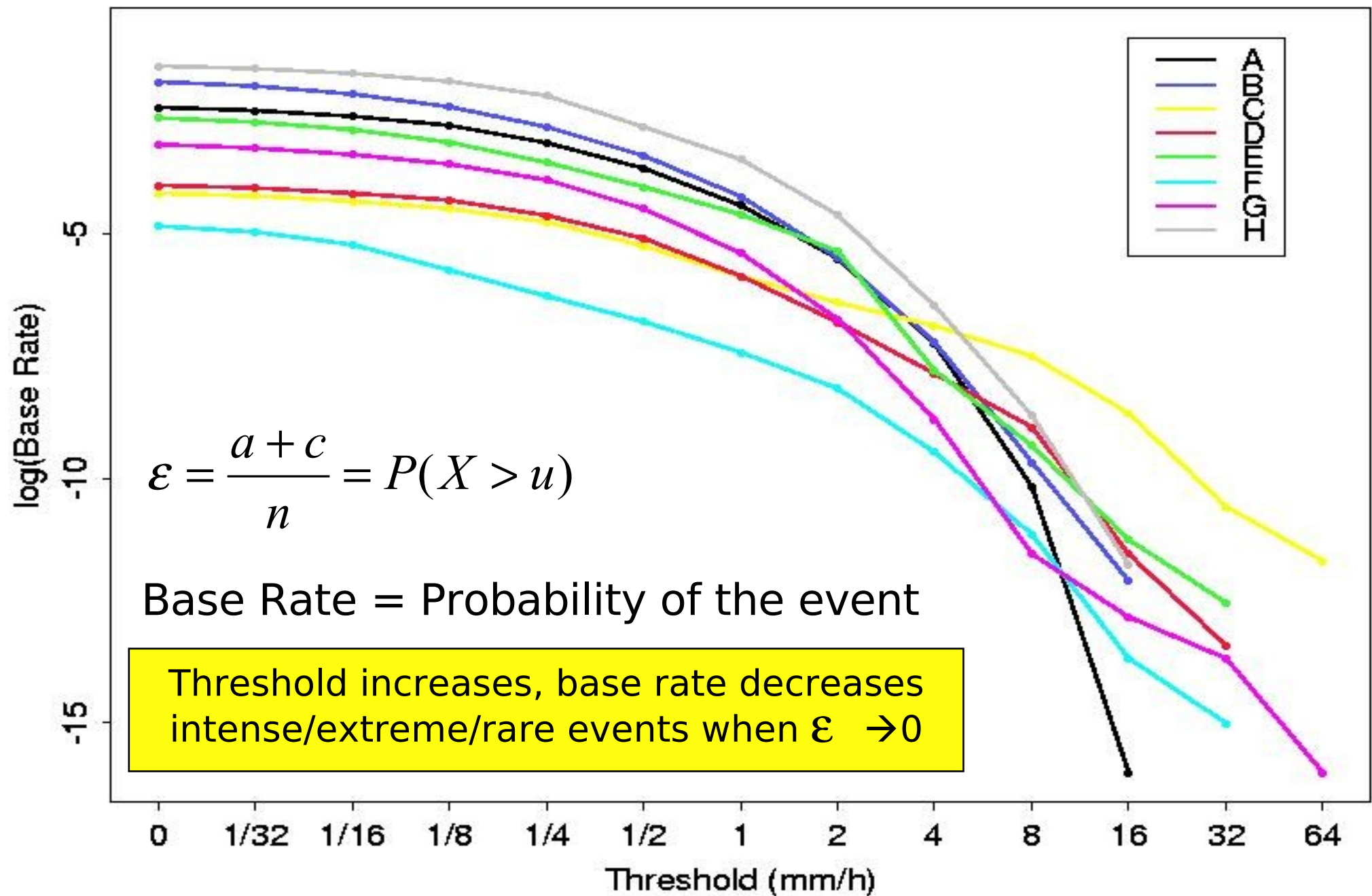
As the threshold increases TS,ETS,HSS,KSS converge to zero (no skill) for all the cases



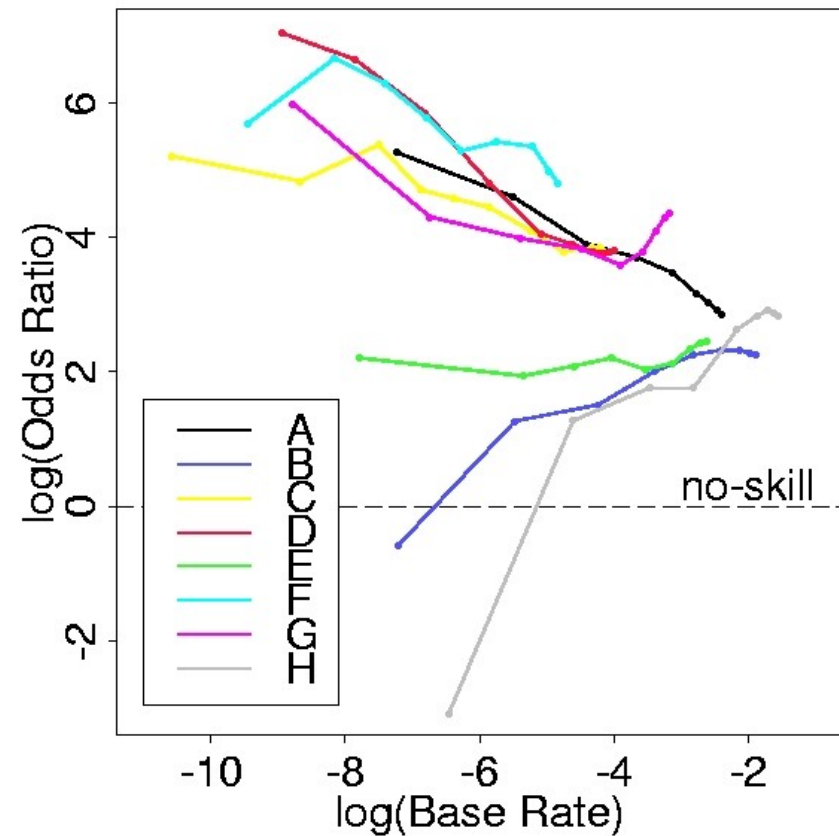
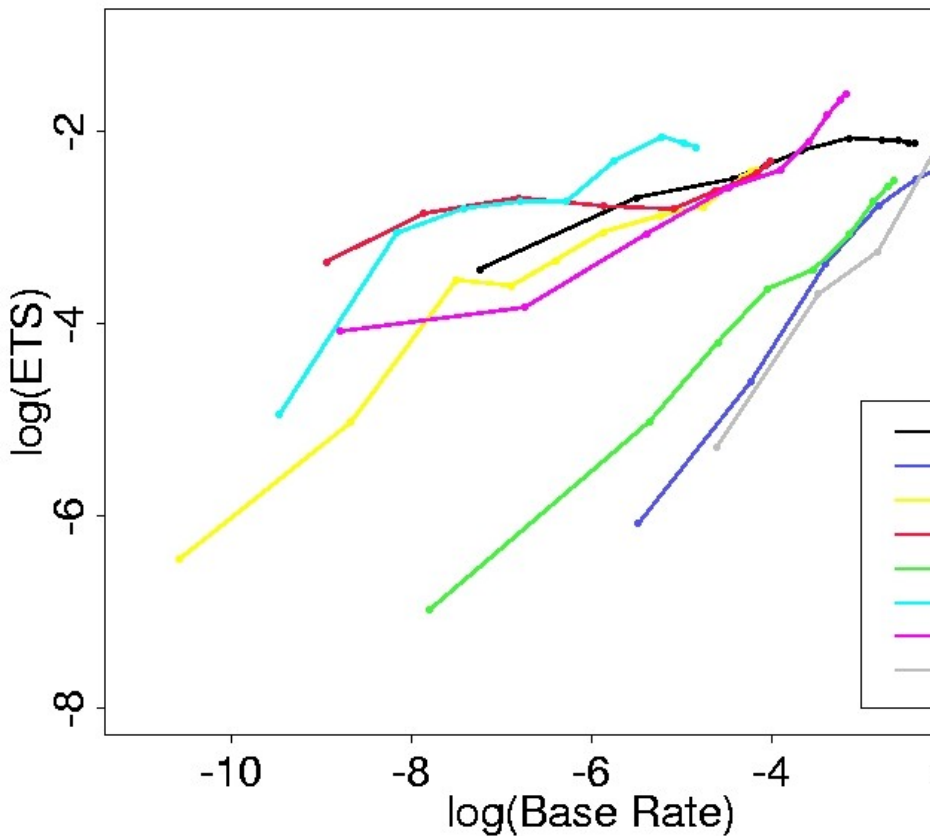
As the threshold increases odds' ratio, ROC, Yule's Q, separates the cases

Why do the scores behave differently ?

Base rate versus threshold

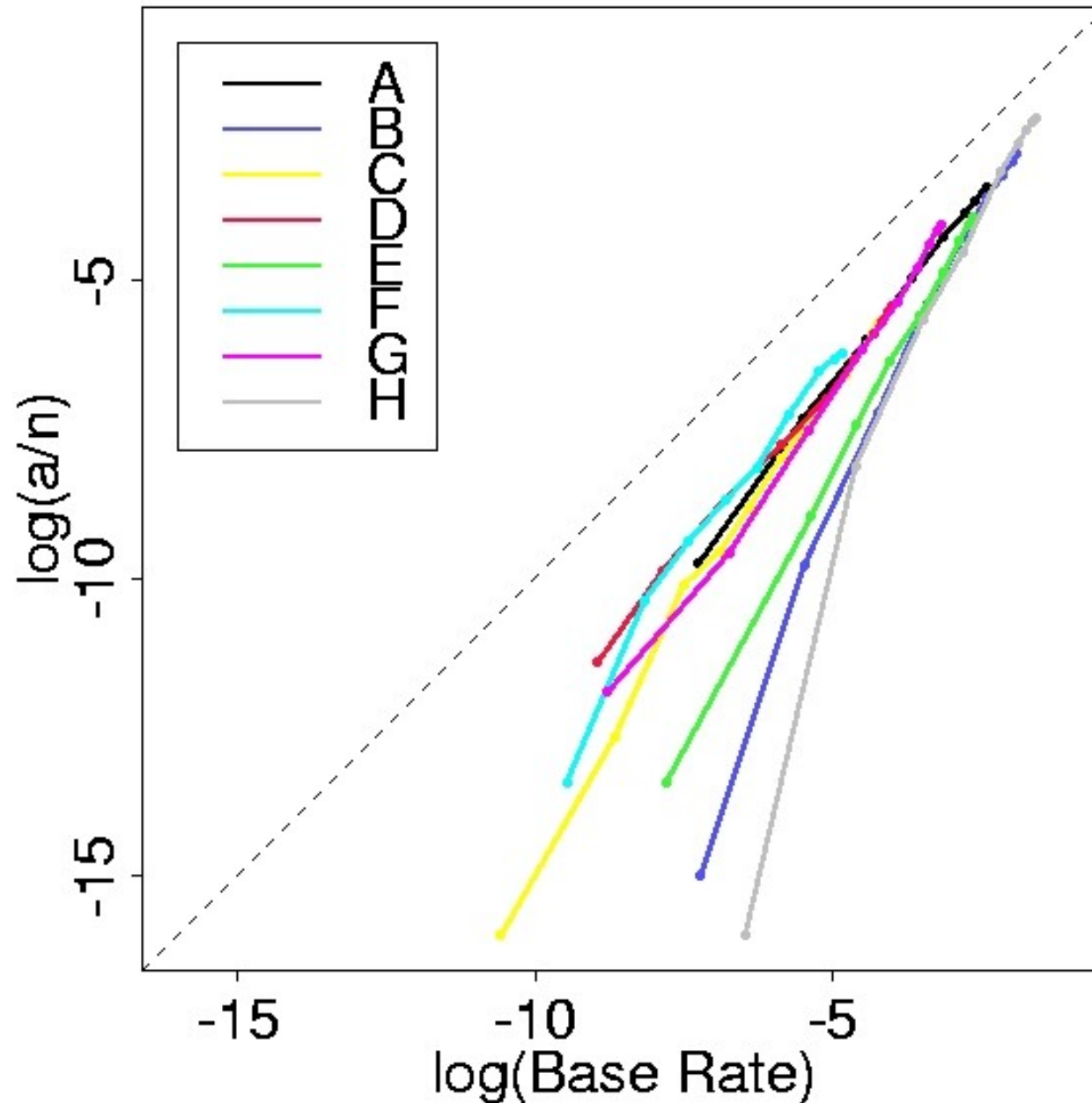


Categorical scores versus base rate



Plots in logarithmic scale: the rate of convergence of the statistics plays a key role in discriminating the NIMROD cases

Behavior of the hits



**Asymptotic
model:**

$$a/n \sim \varepsilon^\eta$$

η = slope
parameter

$\eta > 1$ for ROC
curve regularity

$\eta = 2$ for random
forecast:

- $\eta > 2$ no skill
- $\eta < 2$ skill

Asymptotic behaviour of the joint distribution (un-biased forecast)

	$P(X > u)$	$P(X < u)$	
$P(Y > u)$	ε^η	$\varepsilon - \varepsilon^\eta$	ε
$P(Y < u)$	$\varepsilon - \varepsilon^\eta$	$1 - 2\varepsilon + \varepsilon^\eta$	$1 - \varepsilon$
	ε	$1 - \varepsilon$	1

2 degrees of freedom: (η, ε) fully describe the joint distribution

Express joint prob. and verification statistics as functions of η, ε

Analyze statistics asymptotic behavior when the base rate $\varepsilon \rightarrow 0$

Scores asymptotic behaviour (no bias)

$$TS = \frac{\varepsilon^\eta}{2\varepsilon - \varepsilon^\eta} \rightarrow 0$$

$$ETS = \frac{\varepsilon^\eta - \varepsilon^2}{2\varepsilon - \varepsilon^\eta - \varepsilon^2} \rightarrow 0$$

$$HSS = KSS = \frac{\varepsilon^\eta - \varepsilon^2}{\varepsilon - \varepsilon^2} \rightarrow 0$$

$$OR = \frac{\varepsilon^\eta(1 - 2\varepsilon + \varepsilon^\eta)}{(\varepsilon - \varepsilon^\eta)^2} \rightarrow \begin{cases} 0 & \text{if } \eta > 2 \\ 1 & \text{if } \eta = 2 \\ +\infty & \text{if } \eta < 2 \end{cases}$$

TS, ETS, HSS, KSS magnitude converges to zero as $\varepsilon \rightarrow 0$ either for skilful, random or worse than random forecasts (ε acts as a diminishing factor)

Odds Ratio exhibits different asymptotic behaviours depending on whether a/n converge to zero faster, at the same rate or slower than a/n for a random forecast

Asymptotic behaviour of the joint distribution (biased forecast)

	$P(X > u)$	$P(X < u)$	
$P(Y > u)$	$B\varepsilon^\eta$	$B(\varepsilon - \varepsilon^\eta)$	$B\varepsilon$
$P(Y < u)$	$\varepsilon - B\varepsilon^\eta$	$1 - \varepsilon - B\varepsilon + B\varepsilon^\eta$	$1 - B\varepsilon$
	ε	$1 - \varepsilon$	1

3 degrees of freedom: (B, η, ε) fully describe the joint distribution

Express joint prob. and verification statistics as functions of B, η, ε

Analyze statistics asymptotic behavior when the base rate $\varepsilon \rightarrow 0$

Analyze statistics sensitivity to the bias B in the limit $\varepsilon \rightarrow 0$

Scores asymptotic behaviour (bias)

$$TS = \frac{B\varepsilon^\eta}{\varepsilon(1+B) - B\varepsilon^\eta} \sim \left(\frac{B}{1+B} \right) \frac{\varepsilon^\eta}{\varepsilon}$$

$$ETS = \frac{B(\varepsilon^\eta - \varepsilon^2)}{\varepsilon(1+B) - B\varepsilon^\eta - B\varepsilon^2} \sim \left(\frac{B}{1+B} \right) \frac{\varepsilon^\eta - \varepsilon^2}{\varepsilon}$$

$$HSS = \frac{2B(\varepsilon^\eta - \varepsilon^2)}{\varepsilon(1+B) - 2B\varepsilon^2} \sim \left(\frac{2B}{1+B} \right) \frac{\varepsilon^\eta - \varepsilon^2}{\varepsilon}$$

$$KSS = \frac{B(\varepsilon^\eta - \varepsilon^2)}{\varepsilon(1-\varepsilon)} \sim B \frac{\varepsilon^\eta - \varepsilon^2}{\varepsilon}$$

$$OR = \frac{B\varepsilon^\eta(1 - B\varepsilon - \varepsilon + B\varepsilon^\eta)}{B(\varepsilon - \varepsilon^\eta)(\varepsilon - B\varepsilon^\eta)} \sim \frac{\varepsilon^\eta}{\varepsilon^2}$$

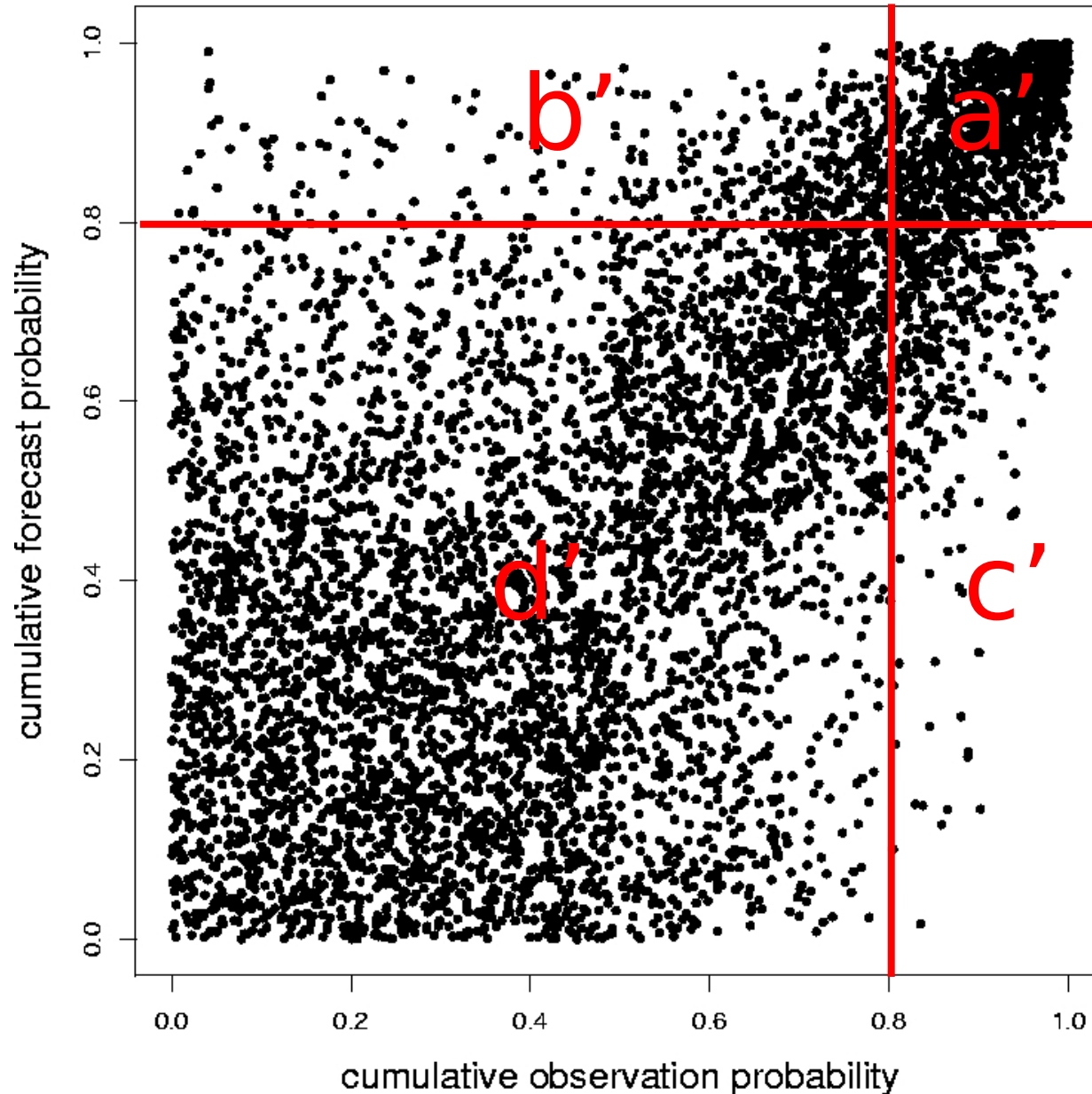
magnitude of **TS**,
ETS, **HSS**, **KSS**

monotonically
increase as B
increases:

**encourage over-
forecasting !!**

**The odds' ratio
is not affected
by the bias**

Coles (1999): extreme dependency



forecast and obs values
are transformed into
empirical cumulative
probabilities

uniform marginal
distributions, no bias

cumulative probability:

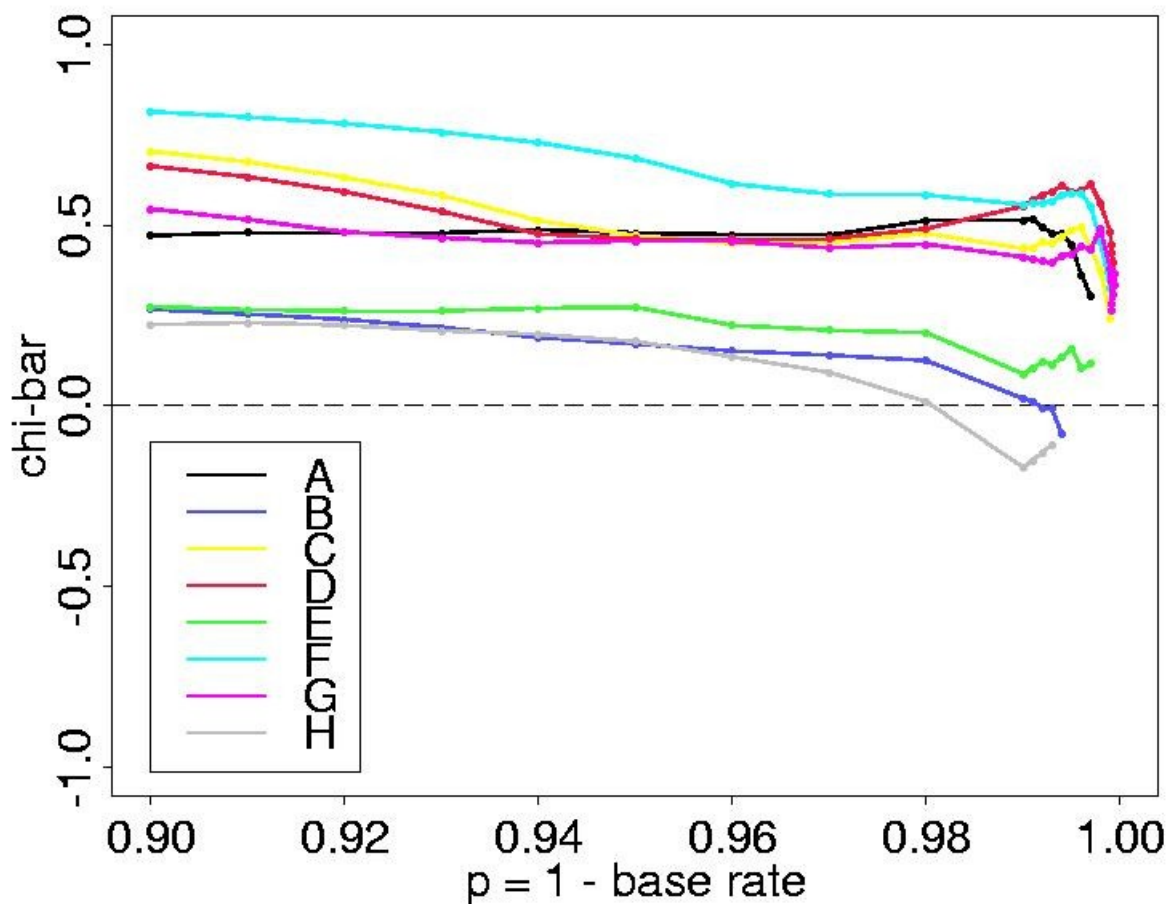
$$p = 1 - \text{base rate}$$

as $\varepsilon \rightarrow 0$, then $p \rightarrow 1$
extreme rare events

Extreme Dependency Score

- does not depend on the base rate
- is not affected by the BIAS
- it depends only on the parameter η (rate of convergence of the joint probability a/n to zero, as the events get rarer)
- separate the case studies

$$EDS = \lim_{p \rightarrow 1} \frac{2 \ln((a' + c') / n')}{\ln(a' / n')} - 1 \rightarrow \frac{2}{\eta} - 1$$



EDS measures
forecast and obs
extreme dependency

Categorical scores and extremes: summary

- TS, ETS, HSS, KSS magnitude converges to zero either for skilful, random or worse than random forecasts: not suitable to verify extremes
- Odds ratio, Yule's Q, ROC are more suitable for detecting the skill in forecasting extreme/rare events
- TS, ETS, HSS, KSS are overly sensitive to the bias in extreme event situations and encourage over-forecasting
- the odds ratio, Yule's Q, ROC are not affected by the bias when verifying extreme/rare events
- **The Extreme Dependency Score provides a bias and base rate independent measure of extreme dependency: very suitable to verify extremes**
- **EDS comes from bivariate Extreme Value Theory: strong statistical framework (Ferro, 2007)**

Complex extremes events

Real-world weather events are complex phenomena consisting of many complexly related attributes, e.g. variable temporal duration and spatial scales, more than one non-independent variable

SEVERE WEATHER INDEX:

$$\text{TMPV} = (\text{IND_GREL} + \text{IND_RAF} + \text{IND_VIL} + \text{IND_ZSUR} + \text{IND_VEF} + \text{IND_MESO}) * 100/135$$

IND_GREL = hail events (5)

IND_RAF = wind gusts (10)

IND_VIL = torrential rain (15)

IND_ZSUR (25)

IND_VEF (30)

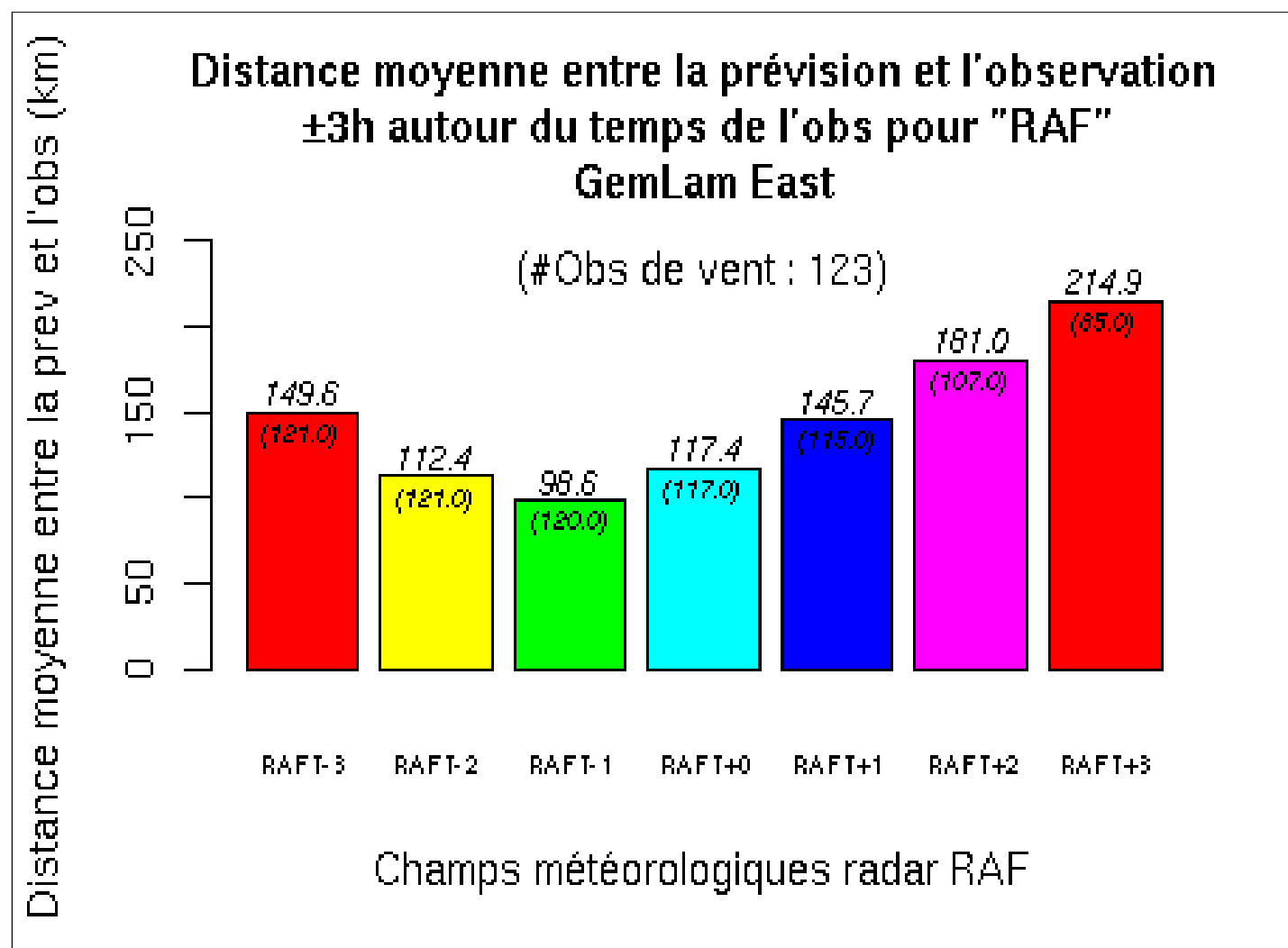
IND_MESO (50)

} associated
to intensity of
thunderstorms

Roy, Turcotte, Chartier (2007) *Verification des algorithmes radars du GemLam 2.5*, internal report, National Severe Weather Lab, Montreal, Canada

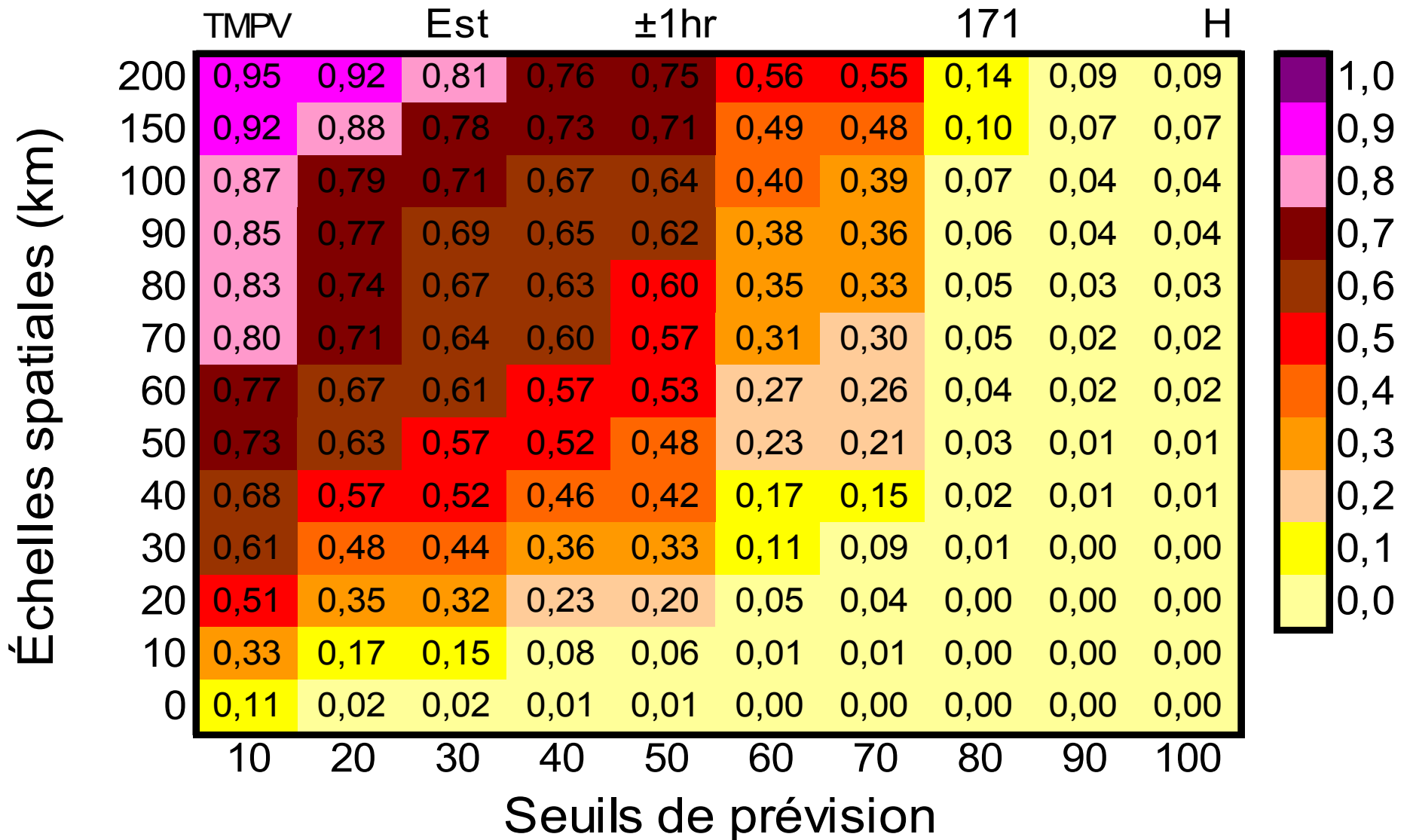
OPTIMAL RADIUS

- search the 5 closest forecast severe events near an observed severe weather report, and average the distances
- the search is performed in a space AND time window



Fuzzy Categorical Approach

... and this method is not contrary to the way a meteorologist would forecast the weather



Conclusions

- **Challenges in extreme event verification** are associated to their **rarity: small sample, large uncertainty**
- **Extreme Value Theory** provide a robust, well established framework to for extreme value analysis
- Research in **Climate and Weather ought to converge** (e.g. weather forecasting could start with GEV and GPD, before jumping to bivariate EVT)
- Close **collaboration between the climate/weather and statistics community** is needed to address un-solved issues in EVT, e.g. regional extreme value analysis
- **Fuzzy verification** approaches are suitable for extreme events evaluation since relax time-space match requirements: **can we combine them with EVT ?**