The Scientific Logic of Extrapolation

Clive Anderson
Sheffield, UK

Ladybower Dam, Derbyshire

Lindley Wood Reservoir, Otley, N. Yorks

Design criteria include:

where a breach could endanger lives in a community

'... withstand 10,000 year flood '

UK Reservoirs Act

Rijkswaterstaat Adviesdienst Geo-Informatie en ICT

... strength of the primary water defences must be assessed every six years (2011, 2017, …) for the required level of protection, which, depending on the area protected, may vary from 250 to 10,000 year loads ...

Dutch Water Act 2009

Aim: step back and look again at

- how/how far is EVT extrapolation justified in scientific terms?
- any limitations? any ways to strengthen?

O wad some Pow'r the giftie gie us

To see oursels as others see us!

It wad from mony a blunder free us, …

Robbie Burns 1786

statistical extrapolation ?

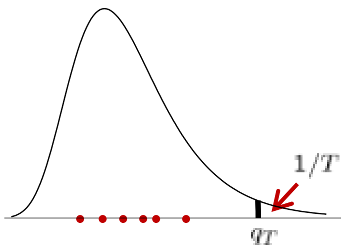
The only function of ~~economic forecasting~~ is to make astrology look respectable.

J K Galbraith

Extrapolation

Estimation of high quantiles of a distribution

beyond available observations



....

$$q_T : \quad \mathcal{F}(q_T) = Pr(X > q_T) = 1/T$$

Outline

1. The Extrapolation Issue
2. The Nature of Mathematical Models
3. The EVT Paradigm
4. But ...
5. Nevertheless ...

2 Mathematical Models

Mathematics is the language of science…

Tobias Dantzig and others

In mathematics we never know what we are talking about,

nor whether what we are saying is true.

Bertrand Russell 'Mysticism & Logic'

Harald Cramér:

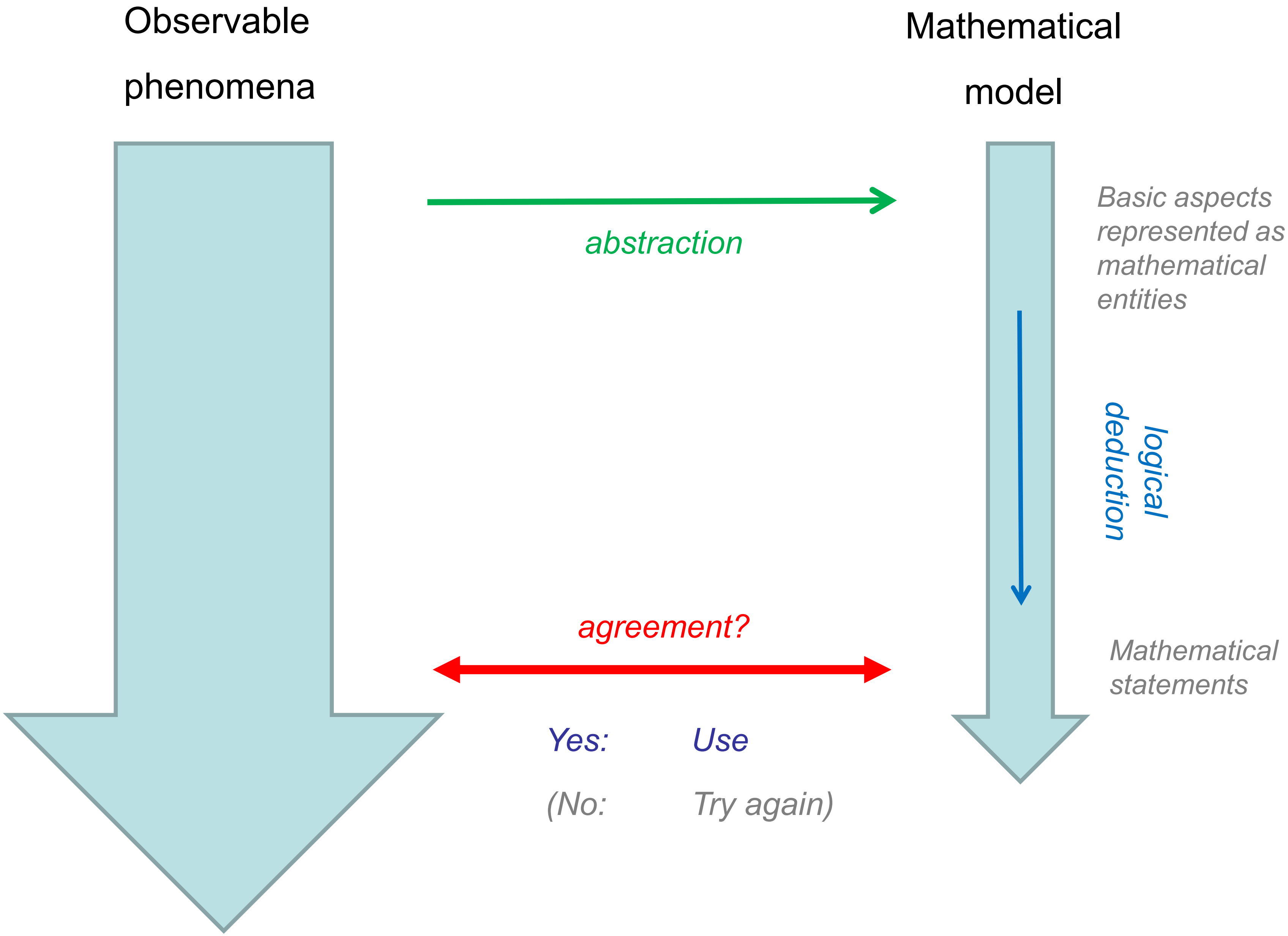
When, in some observable phenomena, we find evidence of a confirmed regularity, we may try to form a mathematical theory … a *mathematical model* of the body of empirical facts which constitute our data.

We choose as our starting point some of the most essential and elementary features of the regularity observed. These we express, in a simplified and idealized form, as mathematical propositions which are laid down as the basic *axioms* of our theory. From the axioms, various propositions are then obtained by purely logical deduction, without any further appeal to experience. The logically consistent system of propositions built up in this way on an axiomatic basis constitutes our mathematical theory.

…

In a case where we have found evidence of a more or less accurate and permanent agreement between theory and facts, the mathematical theory acquires a *practical value*, quite apart from its purely mathematical interest. The theory may then be used for various purposes … [including] Description, Analysis and Prediction.

Mathematical Methods of Statistics, 1945


Observable phenomena
Mathematical model
abstraction
Basic aspects represented as mathematical entities
logical deduction
agreement?
Mathematical statements
Yes: Use
(No: Try again)

Harald Cramér, cont.

Every proposition of such a system is *true*, in the mathematical sense of the word, as soon as it is correctly deduced from the axioms. On the other hand it is important to emphasize that no proposition of any mathematical theory *proves* anything about the events that will, in fact, happen. ...

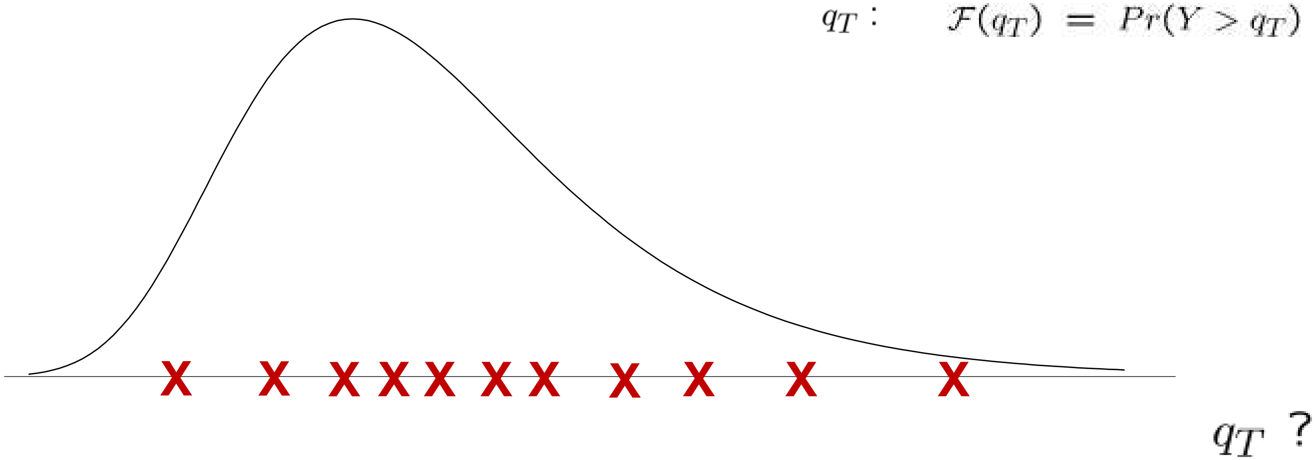
The pure theory ... deals with abstract objects entirely defined by their properties, as expressed by the axioms. For these objects, the propositions of the theory are exactly and rigorously true. But no proposition about such conceptual objects will ever involve a logical proof of properties of the perceptual things of our experience. Mathematical arguments are fundamentally incapable of proving physical facts.

Mathematical Methods of Statistics, 1945

3 The Extreme Value Paradigm

An initial idea: the NTEVP

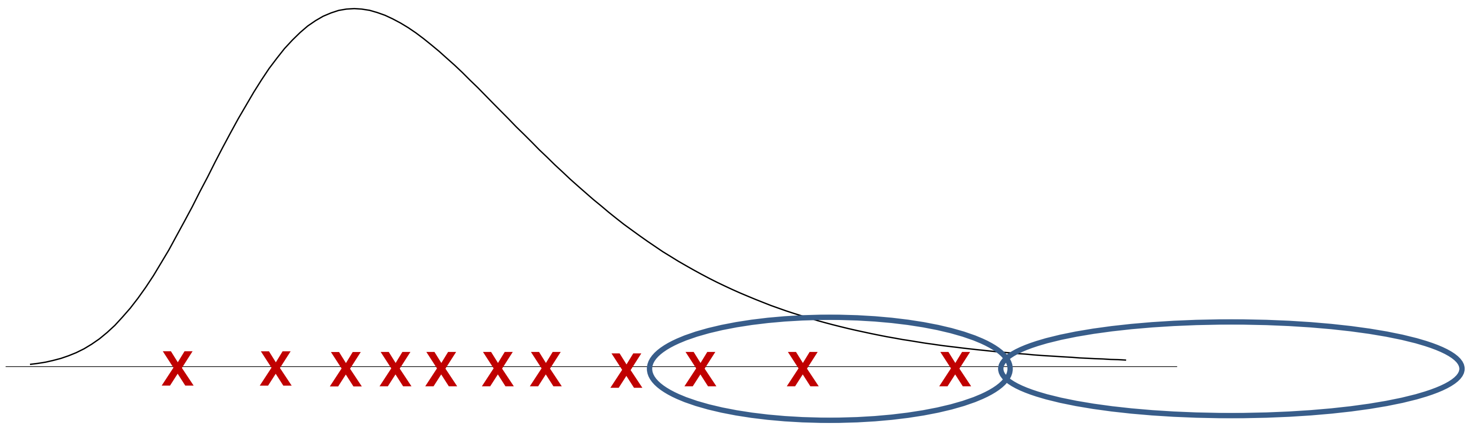
$q_T: \quad \mathcal{F}(q_T) = Pr(Y > q_T) = 1/T$



fit a convenient distribution to the whole data to find $\mathcal{F}$
eg log-normal, gamma, Weibull, . . .

But ... distributions with different tail behaviours
can fit equally well but lead to very different q_T

The Extreme Value Paradigm



- Use evidence from large observations

and

- Use asymptotically-motivated models to relate large observations to the further tail

'Large observations' traditionally – block maxima, or
– exceedances of high thresholds

'Asymptotically-motivated models' come from limit theory:

$$F^n(b_n + xa_n) \to \text{lim} \quad \text{(GEV Dist)} \qquad F = 1 - \mathcal{F}$$

or

$$Pr(X > u + xg(u) \,|\, X > u) \to \text{lim} \quad \text{(GP Dist)}$$

Equivalently:

$$\boxed{\frac{\mathcal{F}(u + xg(u))}{\mathcal{F}(u)} \to \text{lim} \quad \text{(GP Dist)} \quad \text{as } u \nearrow}$$

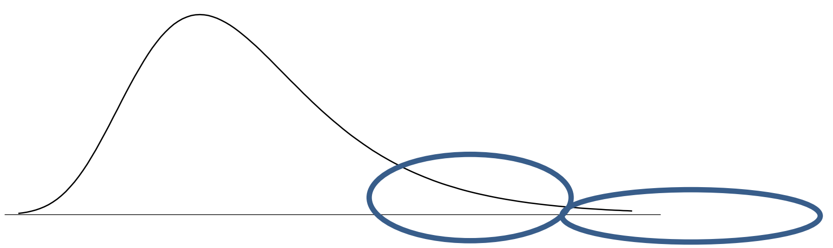
some function $g \geq 0$

Advantages

- convenient for extrapolation:

$$\frac{1}{T} = \mathcal{F}(q_T) \approx Pr(Y > u)\ Pr(Y > q_T \mid Y > u)$$

from observations ≤ u from GPD



- internally consistent

GPD form preserved under increases in the threshold u

- semi-parametric

avoids assumption of specific distributional form for $\mathcal{F}$ – only general tail property $\frac{\mathcal{F}(u + xg(u))}{\mathcal{F}(u)} \to$ a limit, for some g

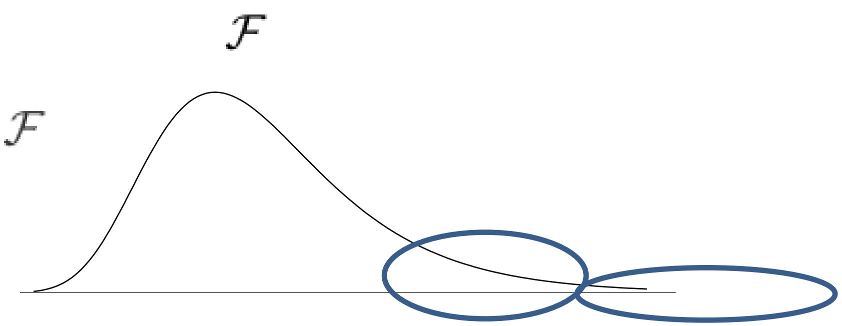
Inevitability of the EV Paradigm

Actually,

$$\frac{\mathcal{F}(u+xg(u))}{\mathcal{F}(u)} \longrightarrow \text{a limit, for some } g \qquad (\dagger)$$

iff limit essentially GPD

($\dagger$) a condition on regularity of decay of $\mathcal{F}$



Status of EV Paradigm:

- extrapolation via GEV or GPD reasonable $\longleftrightarrow$ regularity of decay of $\mathcal{F}$ as in ($\dagger$)

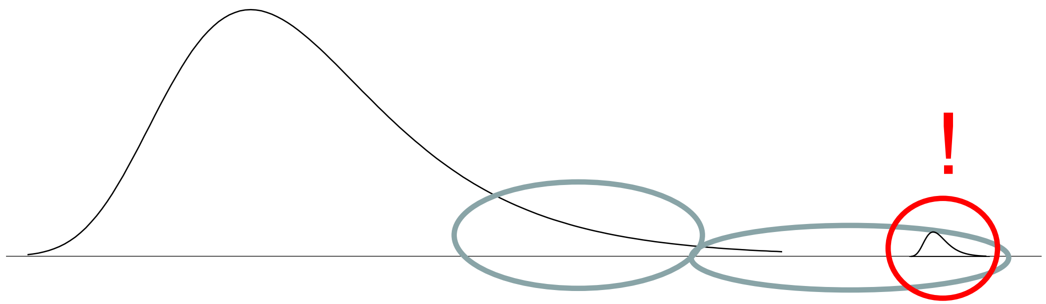
moreover,

- for any reasonable extrapolation to be possible $\longleftarrow$ *some* relation between $\mathcal{F}(u)$ and $\mathcal{F}(v),\ v > u$ if that relation is ($\dagger$), then should use EV Paradigm.

- what's actually needed is closeness of the two sides of ($\dagger$) at relevant levels

4. But ... some reservations

- the unobserved rogue occurrence



Probability that unknown unknown not observed but affects extrapolation to $T = k\times$ sample size ($k = 2, 3, \ldots$) can be up to 0.38, 0.47, 0.54 . . .

Jennison (1990)

(a case in which limit theory, if applicable at all, not applicable at relevant levels)

- transformations, scale of analysis:

$$\mathcal{F}(u + xg(u)) \sim \mathcal{F}(u) \times \text{GPD}$$

is a condition on slope or curvature of $\mathcal{F}$ at high levels

(eg under differentiability conditions, that relative slope or relative curvature behave regularly: $(\log \mathcal{F})' \sim \frac{\text{const}}{x}$, or $\left(1/(\log \mathcal{F})'\right)' \to 0$, etc)

If it holds for X, it holds often also for transformations of X: log, polynomials

Extrapolate in terms of the original variable or of a transformation of it?

Results of extrapolation not invariant to such transformations

- why *linear* normalization and *simple* scaling?

Other normalizations:

If F (cts) is df of X, and G is **any** df, then

$$Pr\left(t_n(\max_{i\leq n} X_i) \leq x\right) = G(x) \qquad \textit{exactly}$$

for transformations $t_n(y) = G^{-1}(F^n(y))$.

Reason for preferring linear/simple? F unknown

linear procedure convenient

But why should nature be convenient?

- hidden covariates, pre-asymptotics

eg Poisson-exponential exceedances but with randomly varying rate

Exponential variation of rate gives

annual max $\frown$ logistic distribution

$$Pr(\max X \leq x) = \frac{1}{1 + e^{-\frac{x-\alpha}{\beta}}}$$

Similarly, seasonal heterogeneity $\rightarrow$ mixture distributions

if unrecognized $\rightarrow$ *!*!

- finite endpoints

- moderate extremes

5. Nevertheless ...

`There is so much good in the worst of us, and so much bad in the best of us, that it behoves all of us not to speak ill of the rest of us.' *R L Stevenson or his grandmother*

Some mitigation steps …

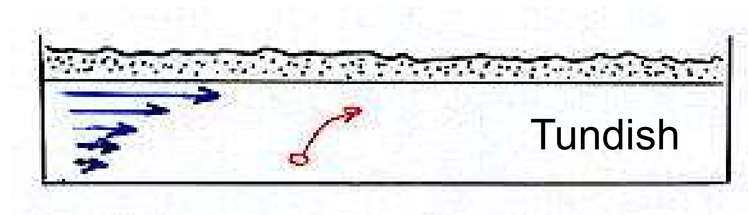
- health warnings
- pragmatic transformations
- more data
- extra information more generally
  - build process knowledge into modelling and inference?

Large Inclusions in Clean Steels

… sizes of large inclusions are important to safety-critical fatigue properties

Removal of inclusions by flow of molten steel through tundish

Mechanism: flotation according to Stokes Law



$$f_{out}(d) \propto \beta(d) \times f_{in}(d)$$

$d = \text{diam}$

spherical inclusions

$f_{out}(d)$: inclusion size pdf on exit

$\beta(d)$: prob. inclusion does not reach slag layer

$f_{in}(d)$: inclusion size pdf on entry

Simple laminar flow: $\beta(d) \propto (d_0^2 - d^2)^{1/3}, \quad d < d_0$

So $f_{out}(d) \overset{\propto}{\sim} f_{in}(d_0) \times (d_0 - d)^{1/3}, \quad d \nearrow d_0$

ie GPD with ξ = -3/4 almost irrespective of entry pdf

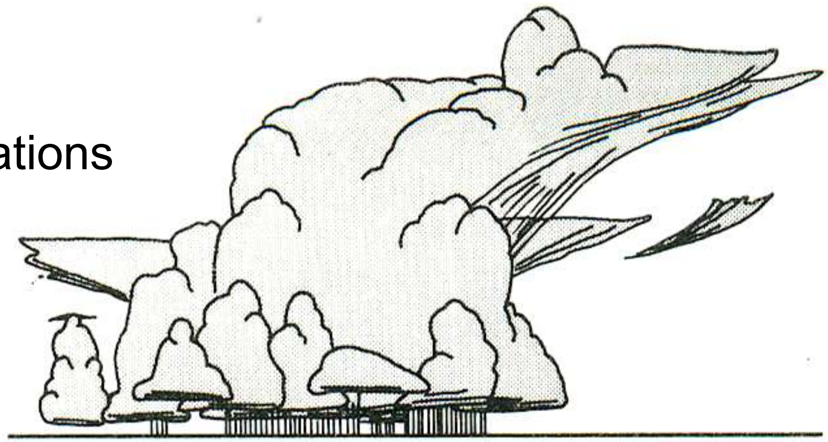
Extreme Rainfall

Could estimate high quantile from immediate data x via EVT

But much meteorological knowledge encapsulated in concept of

PMP = 'largest rainfall physically possible'

- based on physical structure of storms
  in the atmosphere:
  -- vertical temperature & humidity relations
  -- forced ascent over orography
  -- local surface heating (thermals)
  -- mesoscale convergence.



- PMP = function of physical constants, rates, …

$$PMP = PMP(\zeta)$$

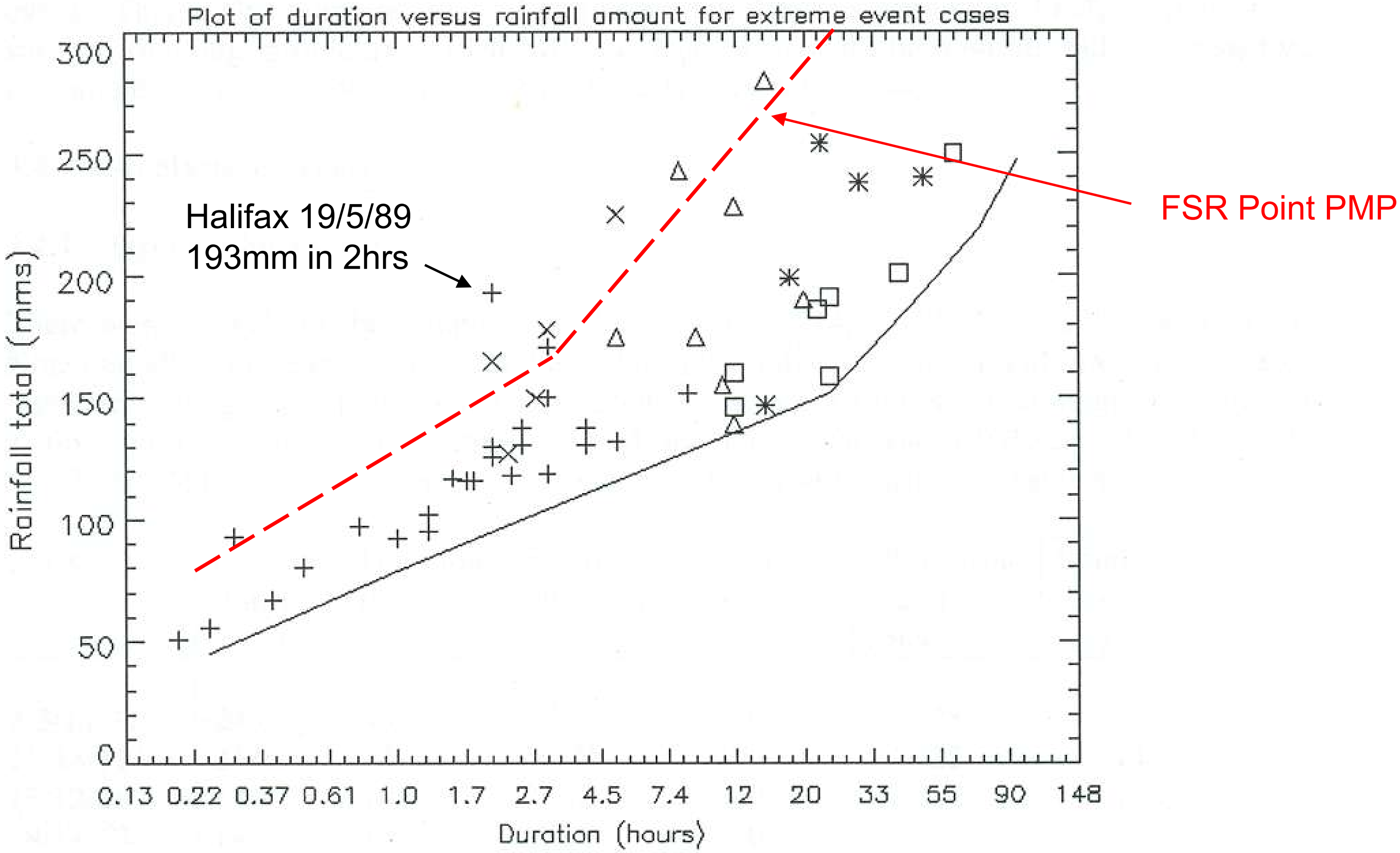


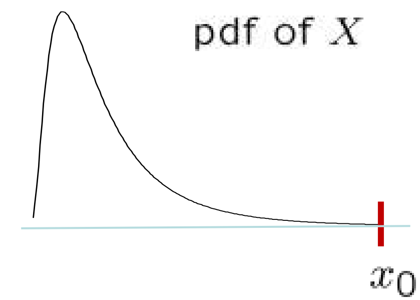
Figure 4. *Plot of rainfall amount (mm) versus duration (h) (on a logarithmic scale) for each of the five event categories listed in table 3. '+' = convective , 'X' = convective*** (frontal forcing), * = orographic, Δ = frontal*** (with embedded instability) and 'square' = frontal. The solid line plot indicates the lowest threshold used for extreme event classification as in table 1.*

From Collier, Fox & Hand, DEFRA Report FD2201, 2002

Recognize that PMP is uncertain

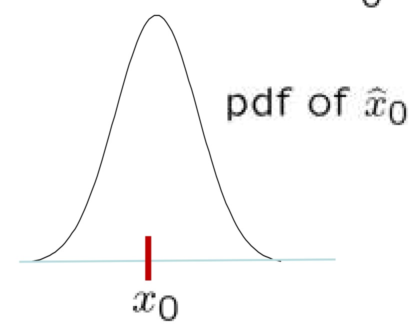
– really it's $\widehat{x}_0$, an *estimate* of true upper limit (end-point) x_0 of distribution of X

Cox (2003)



Possibilities then:

- study distribution of $\widehat{x}_0$ in relation to x_0
  - eg analytically or via Monte Carlo: $\widehat{x}_0 = PMP(\zeta)$



- combine information from $\widehat{x}_0$ with information from directly observed data $\boldsymbol{x}$ to make inferences about x_0 (and other aspects of the X distribution)

-- via likelihood

$$L_{full}(x_0, \phi) = L_{data}(x_0, \phi; \boldsymbol{x}) \times L_{PMP}(x_0; \widehat{x}_0)$$

$\phi =$ other parameters

-- via Bayes

under independence

$$[x_0 \mid \boldsymbol{x}, \widehat{x}_0] \propto \int [\boldsymbol{x}, \widehat{x}_0 \mid x_0, \phi]\, [x_0, \phi] d\phi$$

$$= \int [\boldsymbol{x} \mid x_0, \phi] \times [\widehat{x}_0 \mid x_0] \times [x_0, \phi] d\phi$$

Significant Wave Height Offshore

eg for design of sea defences, offshore structures

Observations on H_s $\longrightarrow$ EVT estimation of high quantiles

Oceanography, hydrodynamics: H_s *depth- limited*

How incorporate oceanographic /hydrodynamic knowledge into EV analysis?

Final Comments

... bring wider evidence into extrapolations?

Assumption of EV-paradigm-type regularity is also an appeal to wider evidence – but more generalized. Can that be studied, quantified?

-- for what kind of systems is EV-paradigm-regularity reasonable?

-- how likely is it?

-- can the weight of evidence for it be incorporated into the analysis?