Statistical Modelling of Groundwater Extremes (STAGE)

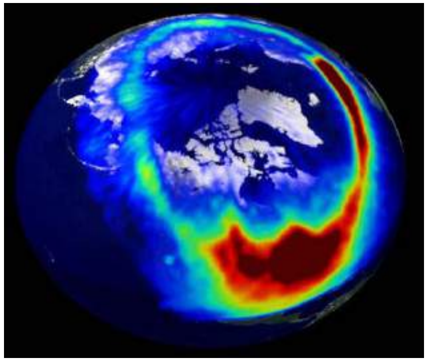
Lancaster University

British Geological Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

SECURE

Aims

- To develop a sustainable collaborative relationship between BGS and LU regarding the study of extreme events in earth science systems
- To demonstrate the use of extreme value statistical methods to model groundwater floods and droughts



Project Team

- Ben Marchant BGS Environmental Statistician

- John Bloomfield BGS Hydrogeologist

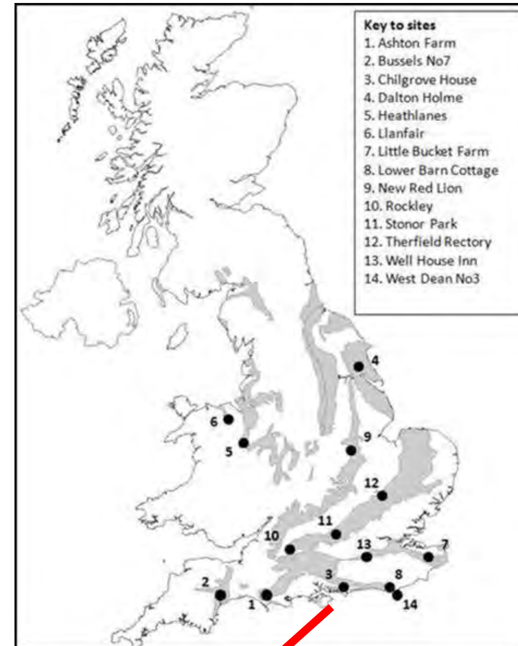
- Emma Eastoe LU EVT Statistician

- Jenny Wadsworth LU EVT Statistician

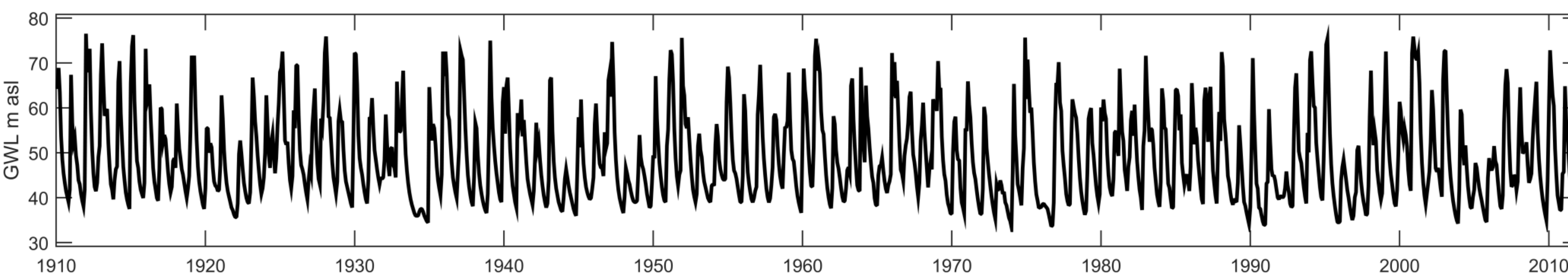
Activities

- A case study exploring extreme events in groundwater hydrographs
- A workshop to discuss other earth science systems where extreme value statistics could be relevant
- Interviews with BGS team leaders about their needs for extreme value statistics
- Projects for the Environmental Pathway of the LU MSc in Statistics

Case Study: Groundwater Extremes



Chilgrove House

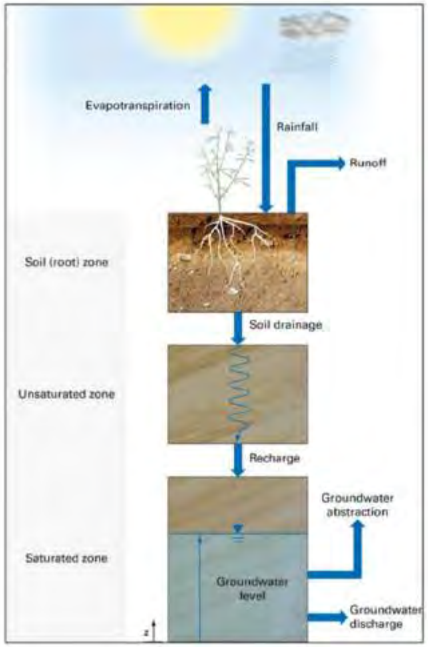


Key Questions

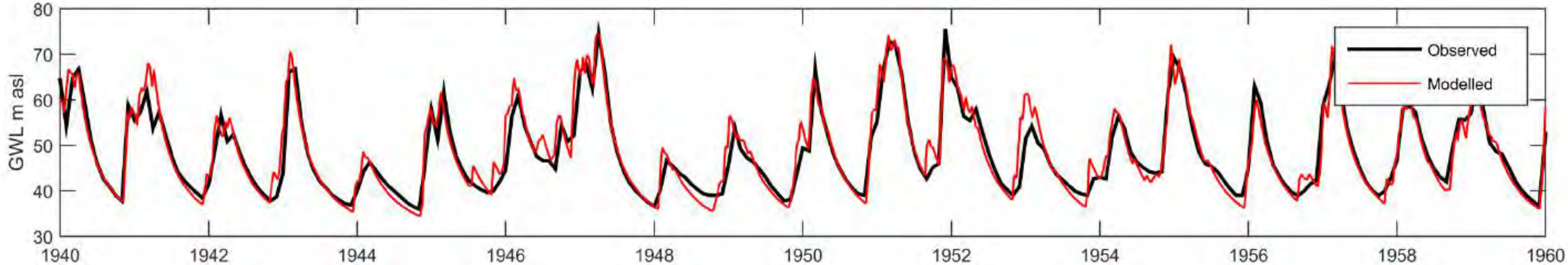
- What GWLs can we can expect to exceed only once every 100/500/1000+ years?
- How many floods or droughts might we expect every 100/500/1000+ years and how long will they last?
- What might the total water surpluses or deficits be during these events?
- Will these characteristics change in a changing climate?

Groundwater Modelling

British Geological Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL



- Calibration ensures that mean behaviour is well approximated
- Extremes might be poorly predicted



Extreme value theory

Methods to:

- Estimate the probability that a particular threshold is exceeded
- Estimate the distribution of the size of the exceedances
- Simulate successive exceedances and determine characteristics of the complete drought/flood episode
- Relate the above to weather parameters (e.g. observed rainfall and potential evapotranspiration)

Modelling probability that an event occurs

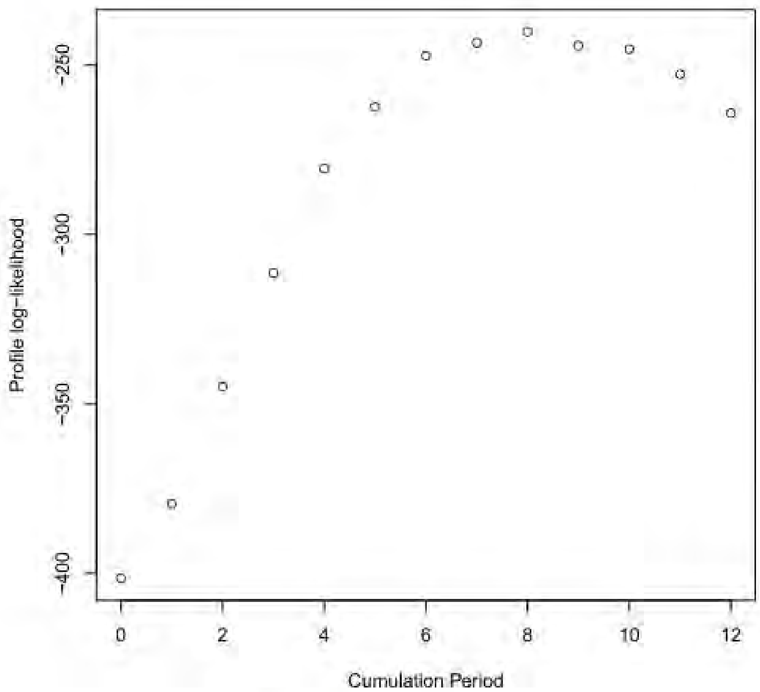
- Initial estimate for the probability of an extreme event

$$\phi = \Pr[Y_t > u]$$

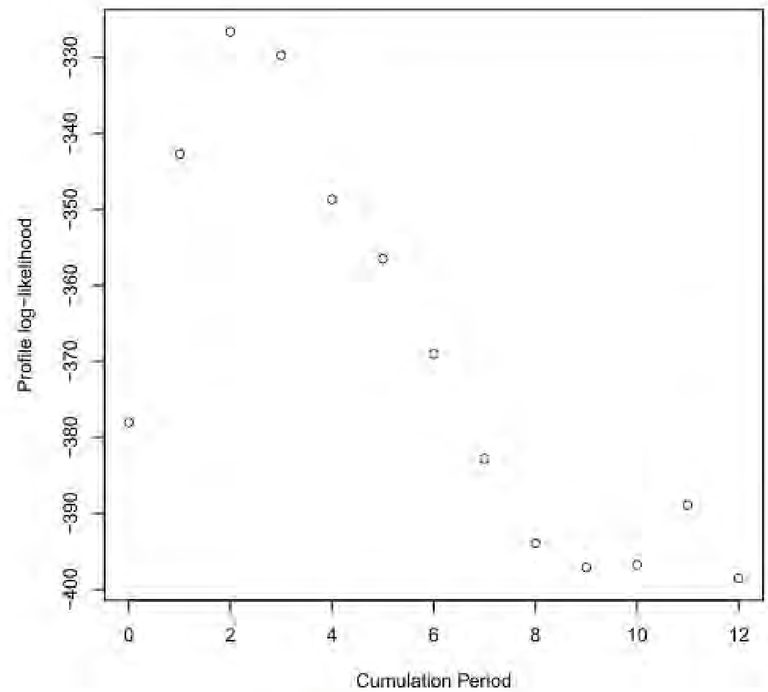
  is given by the observed proportion of extreme events.
  - This assumes that the probability of an extreme event is the same across time.
- Instead can include covariates to allow this probability to change using logistic regression

$$\text{logit}(\phi) = \log \frac{\phi}{1-\phi} = \phi_0 + \phi_1 \text{Rain} + \phi_2 \text{PE}$$

- Model floods and droughts separately.
- Use cumulated PE and rainfall as covariates. Cumulation periods unknown. Fix the same period for both variables, but allow this to be different for flood and drought models.
- Cumulation period is estimated by profile likelihood.



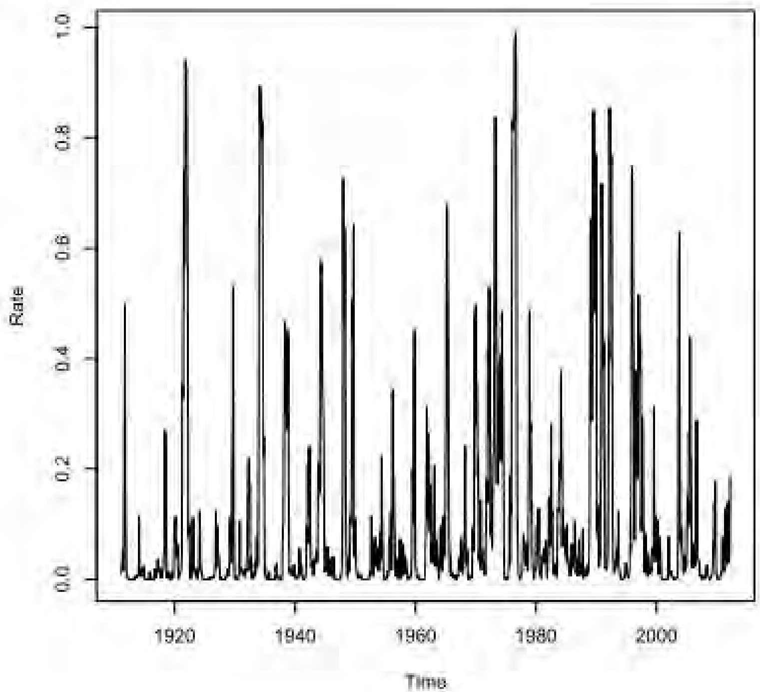
(a) Droughts



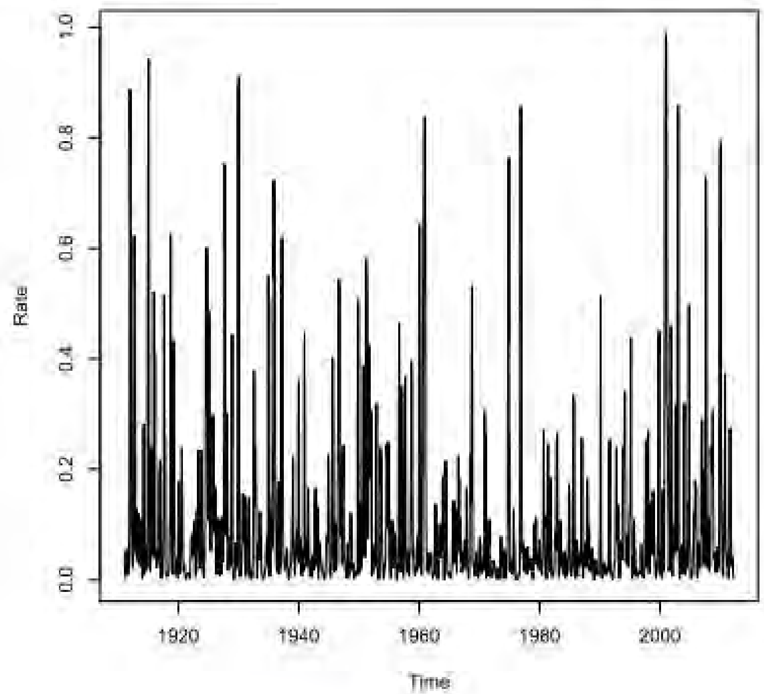
(b) Floods

Probability of extreme events

Estimated probabilities of a flood or drought event:



(a) Droughts



(b) Floods

The probability of a drought/flood event changes considerably over the time period, varying between 0 and 1.

Modelling sizes of events

- Use the generalised Pareto distribution with scale $\psi > 0$ and shape ξ parameters to model the sizes of excursions Y above the threshold u (excesses):

$$\Pr[Y \leq y | Y > u] = 1 - \max\left(0, \left[1 + \xi\left(\frac{y-u}{\psi}\right)\right]\right)^{-1/\xi}, \quad y \geq u.$$

- Distribution covers all rates of tail decay.
- Parameters can be written as linear functions of covariates *e.g.*

$$\log \psi = \psi_0 + \psi_1 \mathrm{Rain} + \psi_2 \mathrm{PE}$$

where ψ_0, ψ_1 and ψ_2 are parameters to be estimated.

Dependence model

Assume a bivariate obsevation (X, Y).

When making dependence assumptions on the data, it is helpful to standardize marginal distributions; here choose exponential,

$$\Pr(X > x) = \Pr(Y > x) = e^{-x}.$$

For a large variety of dependence structures the normalized variables

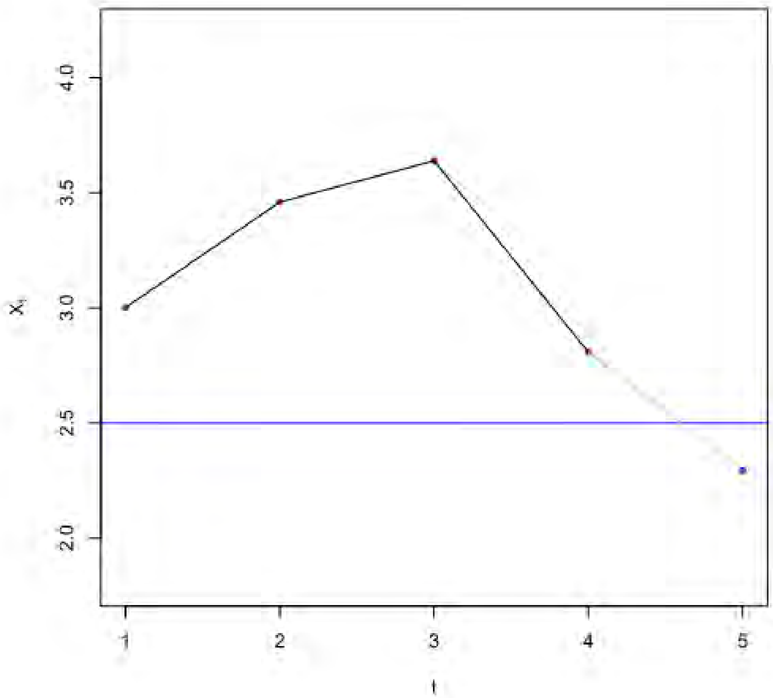
$$Z = \frac{X - \alpha Y}{Y^{\beta}}, \quad \text{and} \quad Y - u$$

are approximately independent for large u, given $Y > u$ (Heffernan and Tawn, 2004).

Regression-type equation:

$$X = \alpha Y + Y^{\beta} \times Z, \quad Y > u.$$

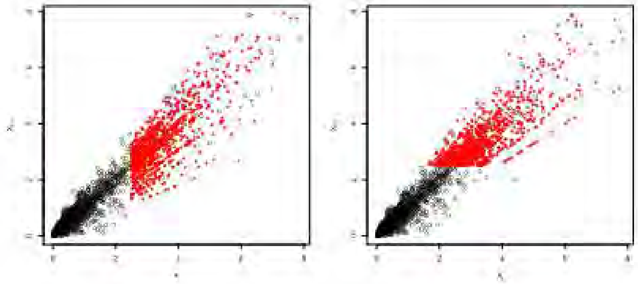
Model fitting and cluster simulation



Generate simulations of drought/flood episodes to determine expected duration and severity

Comments

- Could go forwards or backwards
- Slight asymmetry in data: if X_{t+1} high (GWL low) then X_t likely to be high (GWL low) as drought builds up slowly. If X_t high, then other effects (high rainfall) could make X_{t+1} low.
- Backwards simulation may be preferable on this basis (it was for Chilgrove House)
- Higher order Markov assumptions needed to really capture the features here: some extensions to the methodology developed to capture this



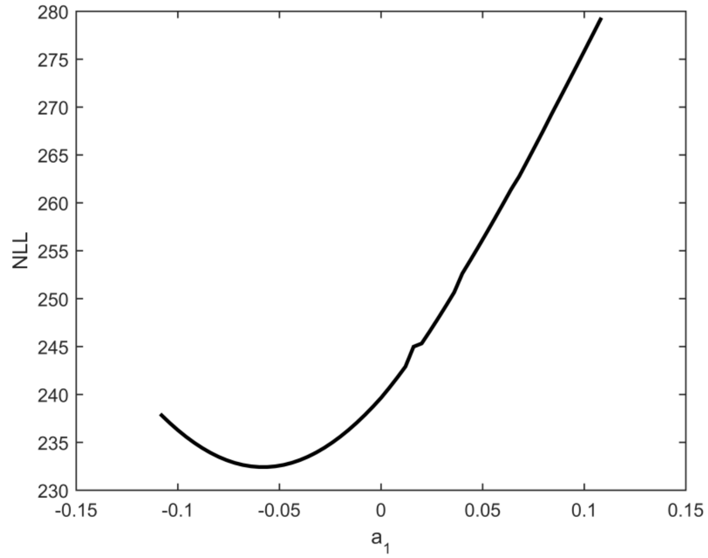
On-going groundwater work

- Including weather covariates in the Heffernan & Tawn model

- Applying to more boreholes

- Exploring model parameter uncertainty and data requirements

- A paper



Project Workshop

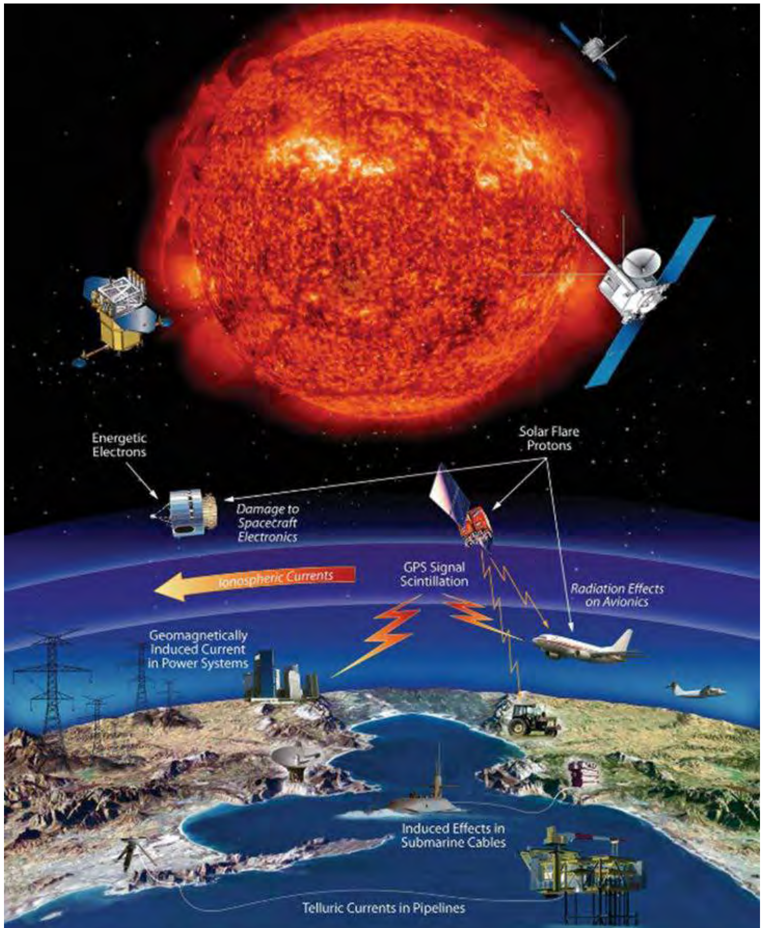
- Presentation by Jenny and Emma about extreme value statistics and earth sciences
- Meetings with BGS teams
- Identified potential projects on space weather, soil contamination, seismology & glaciology

Space weather

BGS Space Weather @BGSspaceWeather · Sep 15

Next 24hrs - QUIET. Geomagnetic activity expected to remain QUIET, although there remains a slight possibility of an isolated ACTIVE period.

3 5



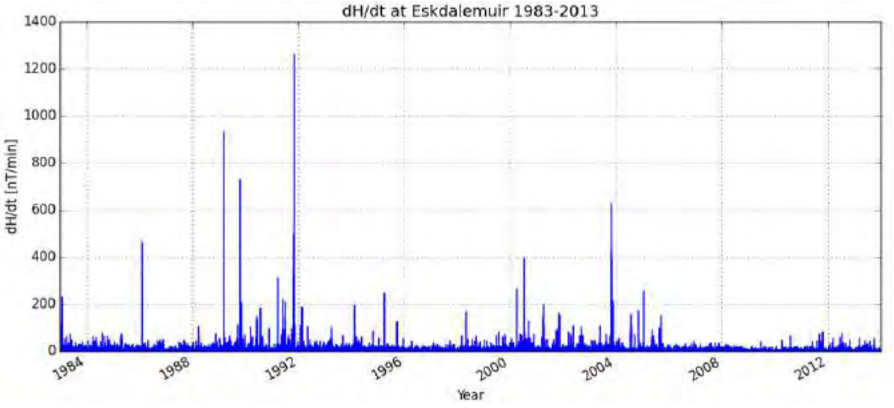
National risk Register

Impact	Between 1 in 20,000 and 1 in 2,000	Between 1 in 2,000 and 1 in 200	Between 1 in 200 and 1 in 20	Between 1 in 20 and 1 in 2	Greater than 1 in 2
5				Pandemic influenza	
4			Coastal flooding Effusive volcanic eruption		
3	**Major transport accidents**	**Major industrial accidents**	Other infectious diseases Inland flooding	Severe space weather Low temperatures and heavy snow Heatwaves	
2			Animal diseases Drought Public disorder	Explosive volcanic eruption Storms and gales	
1			Severe wildfires	**Disruptive industrial action**	

Relative likelihood of occurring in the next five years

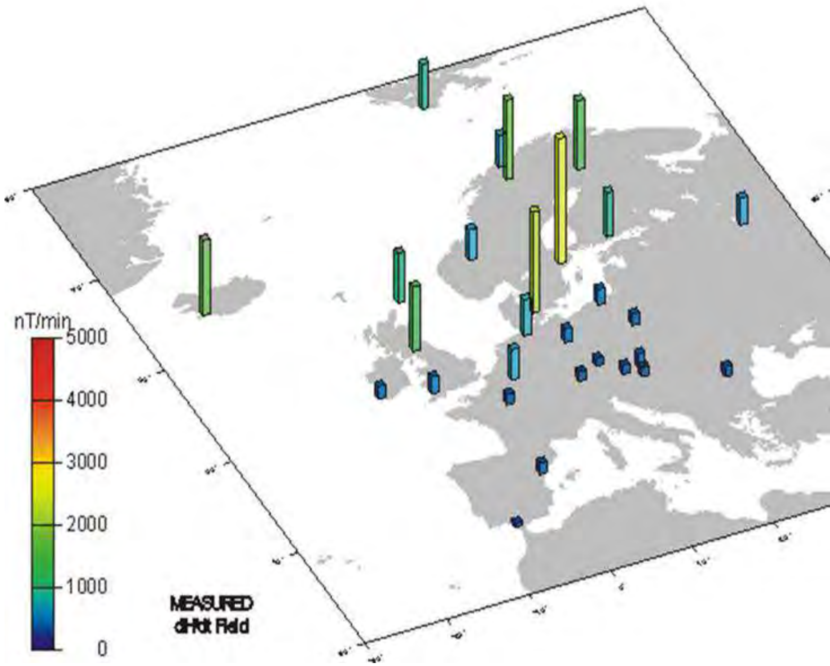
Geomagnetic Field Variations

- 30 years of 1-minute data
- 28 sites across Europe



Issues

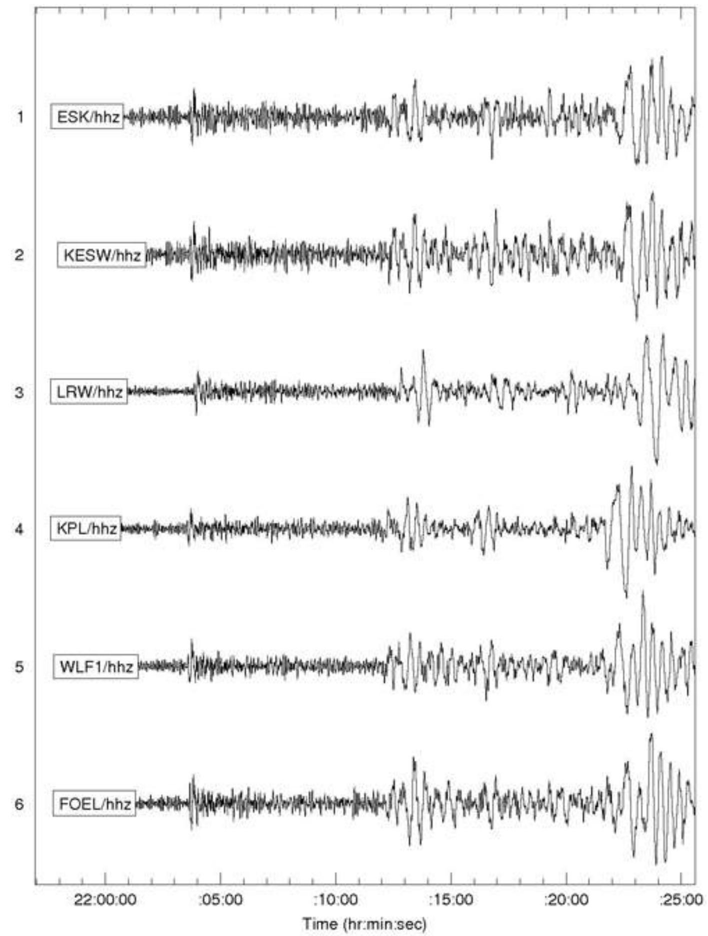
- Temporal clustering of storms
- Spatial extent of storms
- Relationship with solar cycle
- Uncertainty of return times



Two completed MSc projects

Seismology

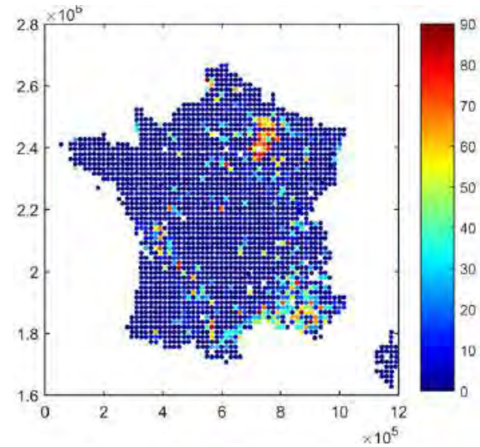
"On all accounts, methods which utilize all available data give superior estimates of the parameters of seismicity than do extreme value methods" Knopoff & Kagan, 1977.

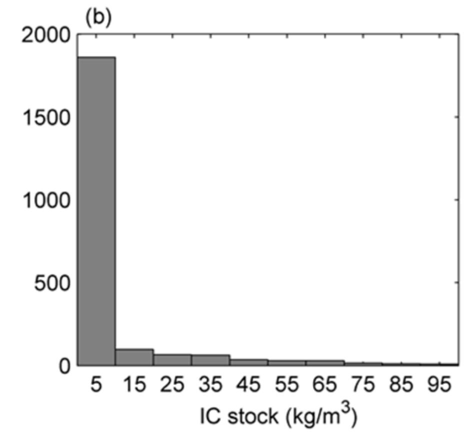


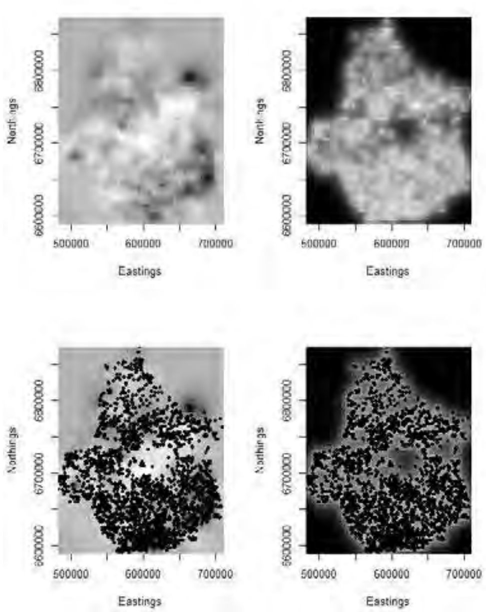
Future MSc project

Soil science

- Mapping highly skewed soil properties
- Model marginal with a mixture distribution with a GPD tail
- A completed MSc project exploring the relationship between soil carbon concentration and soil texture







STAGE outcomes

- EVT capability for BGS
- Challenging data for LU
- Better understanding of groundwater extremes and geomagnetic storms
- More general geostatistical models
- Future funding opportunities
- PhD proposals

Acknowledgements