UUEM Workshop, LSE, January 2014

Testing models as hypotheses: how can we be scientific in the face of epistemic errors?

Keith Beven

Lancaster University and Uppsala University

UPPSALA UNIVERSITET

LANCASTER UNIVERSITY

My background

- I am a hydrologist
- I have worked at the Institute of Hydrology and Universities of Leeds, Virginia and Lancaster, with visiting positions at UC Santa Barbara, EPFL Lausanne, KU Leuven, Uppsala University, and LSE London
- I have worked on many hydrological models (Topmodel, IHDM, SHE, MIPs, DBM….) and 1D/2D hydraulic models (HEC-RAS, ISIS, JFLOW,….)
- Interests in floods, flood forecasting, future change, residence times and travel times of pollutants
- And uncertainty (Generalised Likelihood Uncertainty Estimation, GLUE, methodology)

UPPSALA UNIVERSITET

LANCASTER UNIVERSITY

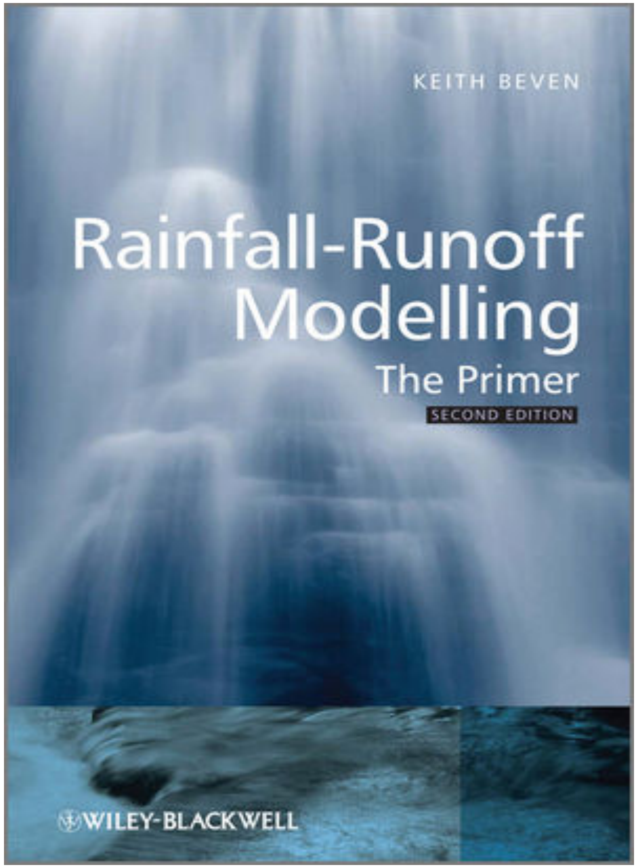
My background

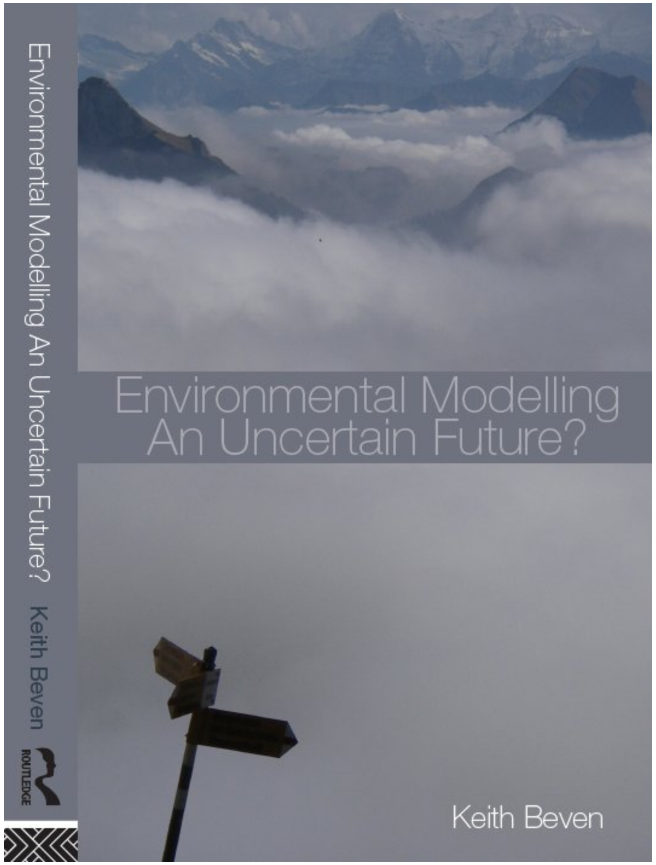
- Started doing Monte Carlo experiments on models at the University of Virginia in 1980 (start of GLUE and equifinality concepts)
- 80's - Used Monte Carlo in continuous simulation for flood frequency estimation
- Moved to Lancaster 1985, continued GLUE work, first publication with Andrew Binley in 1992.
- Most recent thoughts on how to do science given uncertainties in *CRAS Geosciences* 2012 paper and "GLUE 20 years on" paper in *Hydrol. Process.* 2013
- Current CREDIBLE project on uncertainty in risk assessments for natural hazards for NERC
- Work has been summarised in two books.

UPPSALA UNIVERSITET

LANCASTER UNIVERSITY

The books…





www.uncertain-future.org.uk

UPPSALA UNIVERSITET

LANCASTER UNIVERSITY

The Modelling Process

- The Perceptual Model (ideas)
- The Conceptual Model (equations)
- The Procedural Model (code)
- Model Calibration (may be difficult?)
- Estimation of predictive uncertainty

- Model Validation (may be impossible?
  - but a good idea !!)

- **Declare Success ?**
- **Use in decision making?**

Increasing Approximation

Uncertainty in Environmental Models
- Why should we be interested in uncertainty?

- Models are wrong and are *known* to be wrong … (Adam Morton)
- but might still be useful in real applications….(George Box)
- Input and boundary condition data are wrong and are known to be wrong
- Parameter values might have physical significance, but the values we can measure in the lab or field may not be the values that are required to make the model work (therefore some calibration usually needed)
- Observations with which we can check a model are wrong and are known to be wrong
- Result that may be rather difficult to differentiate between different models as hypotheses (the *equifinality thesis*)
- Might there be better ways of doing environmental science?

Uncertainty estimation methods

Three different types of problem:

- Forward uncertainty estimation (no data to assess residuals – must depend totally on assumptions about sources of uncertainty)

- Model conditioning (historical data to assess residuals in calibration period – learn about errors for use in prediction)

- Real-time data assimilation (can assess residuals in real-time and update forecasts and forecast uncertainty – especially used in flood forecasting –)

See Uncertain Future? book

Hydrology as one of the inexact sciences

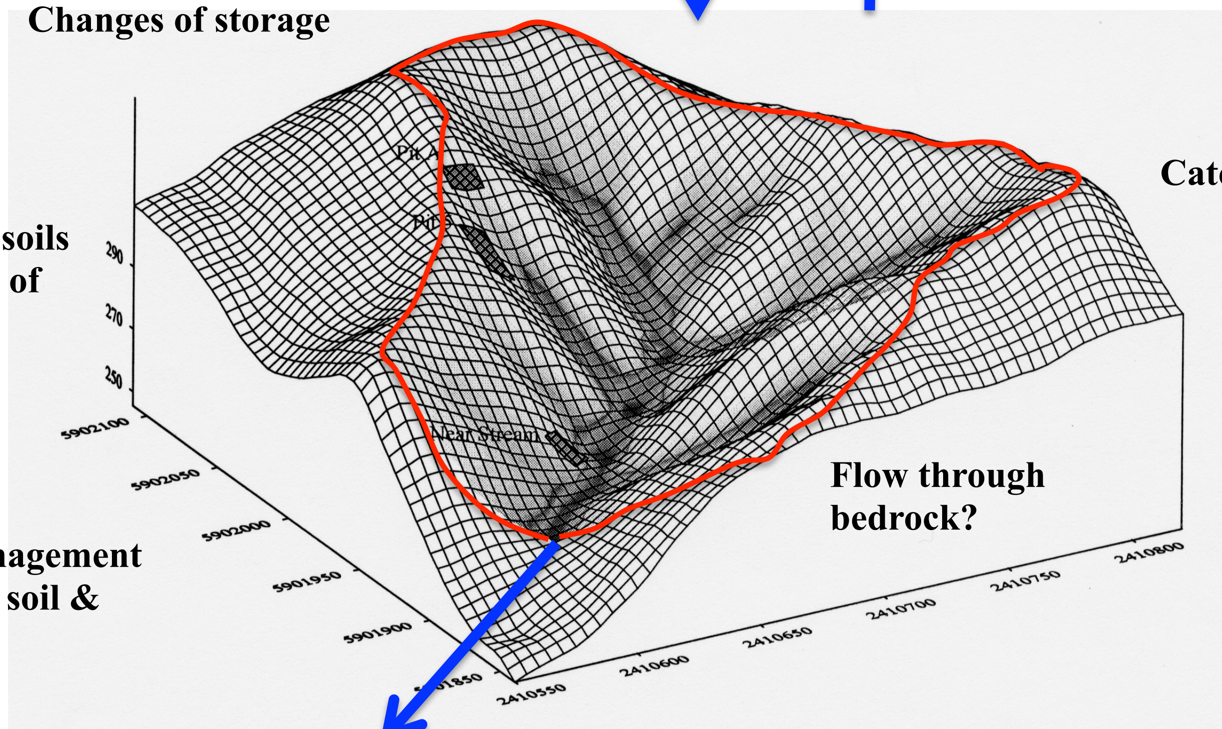
Pattern of rainfall inputs

Pattern of actual evapotranspiration

Changes of storage

Catchment area

Heterogeneity of soils & representation of flow processes



Flow through bedrock?

Land Management Effects on soil & surface

Discharge estimates (requires “rating curve”) A/E/D

Hydrology as one of the inexact sciences

The Water Balance Equation

$$Q = R - E_a - \Delta S$$

All of terms subject to both epistemic and aleatory uncertainties……and there may be other inputs and outputs impossible to measure

Nancy Cartwright *"This Dappled World"* – such principles are capacities rather than truths when applied in practice

UPPSALA UNIVERSITET

LANCASTER UNIVERSITY

Sources of Uncertainty

- Errors in the input and boundary condition data
- Errors in the model structure
- Errors in estimates of parameter values
- Commensurability of modelled and observed variables and parameters
- Errors in the observations used to calibrate or evaluate models
- Errors of omission (not always the unknown unknowns)
- ……

LANCASTER UNIVERSITY

Sources of Uncertainty

- Errors in the input and boundary condition data
- Errors in the model structure
- Errors in estimates of parameter values
- Commensurability of modelled and observed variables and parameters
- Errors in the observations used to calibrate or evaluate models
- Errors of omission (not always the unknown unknowns)

Difficult (impossible) to disentangle different sources of error without making strong assumptions

(Water Science and Technology, 2005)

LANCASTER UNIVERSITY

Types of error and why they are important

- Formal statistical approach to likelihoods (generally) assumes that the (transformed) errors are additive and random (*aleatory error*) conditional on the model being correct

- But in environmental modelling, many sources of error (in model structure, input data, parameter values,….) are a result of lack of knowledge (*epistemic error*) which will result in nonstationarity of error characteristics

- In extreme cases, data available for calibration might even be *disinformative*.

UPPSALA UNIVERSITET

LANCASTER UNIVERSITY

Types of Uncertainty (and why they are important)

Aleatory Uncertainty

Epistemic Uncertainty

System Dynamics

Forcing and Response Data

Disinformation

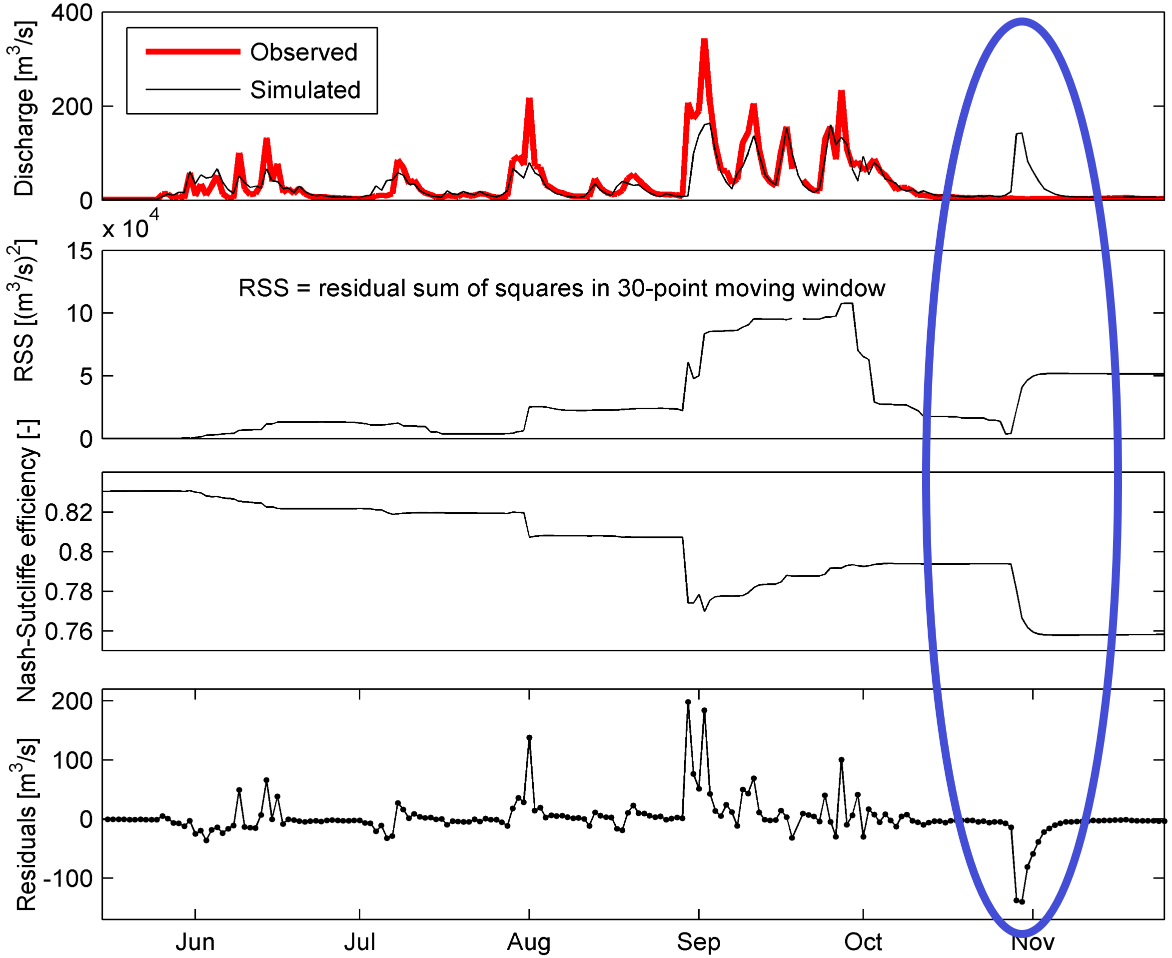
Semantic/Linguistic Uncertainty

Ontological Uncertainty

Type of Uncertainty	Description
Aleatory Uncertainty	Uncertainty with stationary statistical characteristics. May be structured (bias, autocorrelation, long term persistence) but can be reduced to residual stationary random component
Epistemic Uncertainty (system dynamics)	Uncertainty arsing from a lack of knowledge about how to represent the catchment system in terms of both model structure and parameters. Note that this may include things that are included in the perceptual model of the catchment processes but which are not included in the model. They may also include things that have not yet been perceived as being important but which might result in reduced model performance.
Epistemic Uncertainty (forcing and response data)	Uncertainty arising from lack of knowledge about the forcing data or the response data with which model outputs can be evaluated. This may be because of commensurability or interpolation issues when not enough information is provided by the observational techniques to adequately describe variables required in the modelling process.
Epistemic Uncertainty (disinformation)	Analogous to known unknowns (in either system representation or forcing data that are *known* to be inconsistent or wrong. Will have the expectation of introducing disinformation into the modelling processes resulting in biased or incorrect inference (including false positives and false negatives in testing models as hypotheses)
Semantic / Linguistic Uncertainty	Uncertainty about what statements or quantities in the relevant domain actually mean (there are many examples in hydrology including storm runoff, baseflow, hydraulic conductivity, stationarity etc). This can partly result from commensurability issues that quantities with the same name have different meanings in different contexts or scales.
Ontological Uncertainty	Uncertainty associated with different belief systems. Relevant example here might be beliefs about whether formal probability is an appropriate framework for the representation of mode errors. Different beliefs about the appropriate assumptions could lead to very different uncertainty estimates.

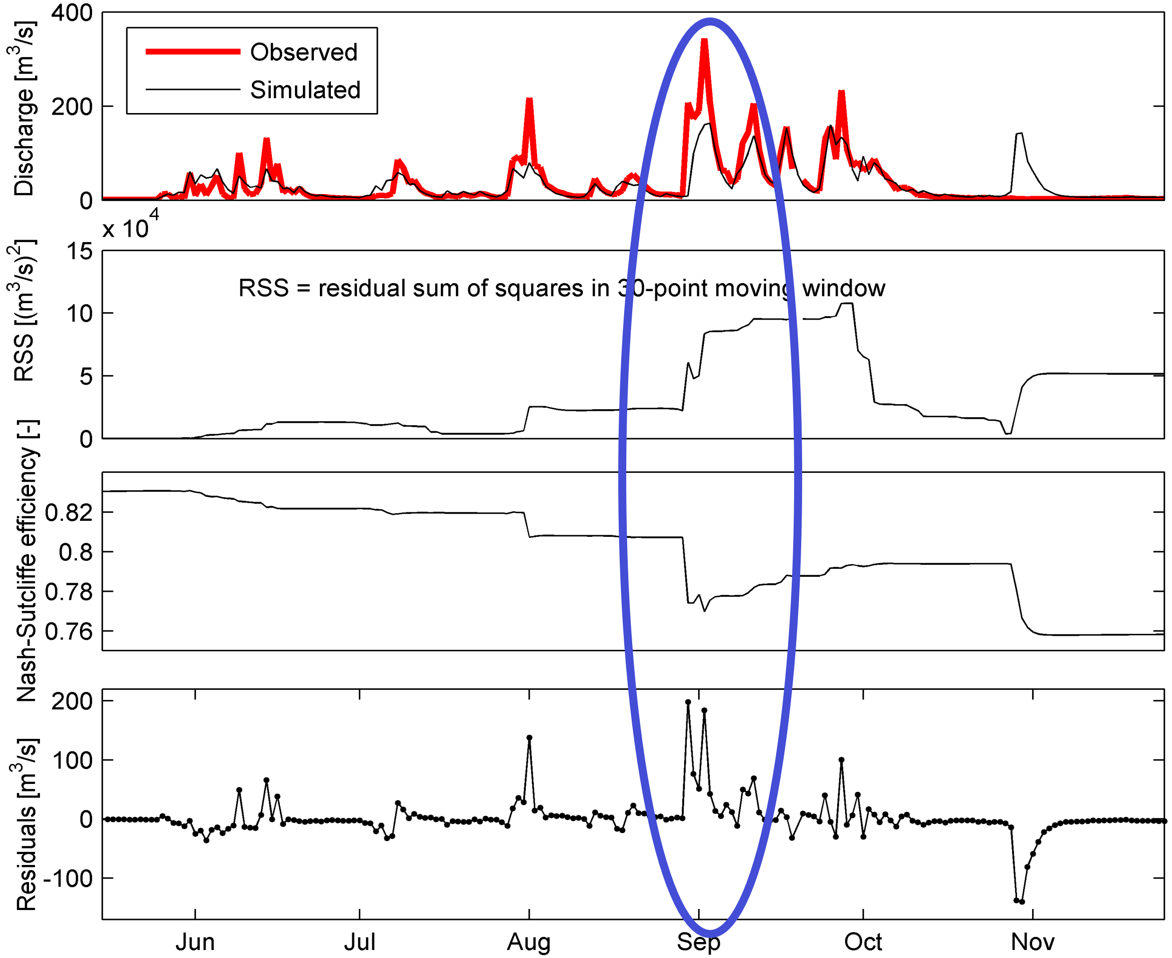
Beven, Leonardo Lecture: Facets of Uncertainty, Hyd.Sci.J. 2014

Disinformation in calibration data



Application of WASMOD to Pasa La Ceiba, Honduras (from Ida Westerberg, Uppsala)

Disinformation in calibration data



Application of WASMOD to Pasa La Ceiba, Honduras (from Ida Westerberg, Uppsala)

Types of error and why they are important

Aleatory (A), Epistemic (E) or Disinformative (D)

- Errors in the input and boundary condition data (A/E/D)
- Errors in the model structure (E/D?)
- Errors in estimates of parameter values (A/E)
- Commensurability of modelled and observed variables and parameters (A/E/D)
- Errors in the observations used to calibrate or evaluate models (A/E/D)
- Errors of omission (sometimes known omissions) (E/D?)
- The unknown unknowns (D?, becoming E/D)

Activity

Make a list of the most important uncertainties in your modelling activity.

Classify these under the following headings:

1. Essentially random
2. Lack of knowledge about how to represent some processes (including different theories/belief systems about how to do so)
3. Lack of knowledge about how to define parameter values at scale of the model
4. Lack of knowledge about boundary conditions or forcing data
5. Lack of knowledge about how observables relate to model variables in model calibration/evaluation/ verification
6. Unknown unknowns

LANCASTER UNIVERSITY

Activity (2)

What effect do you expect the epistemic uncertainties to have on model simulations?

Is there evidence in current model results to suggest they might be important?

Is there evidence that some calibration/ evaluation/verification data might be disinformative?

Uncertainty about uncertainty estimation

- Many sources of uncertainty in the modelling process but can generally only evaluate the differences (residuals) between some observed and predicted variables (e.g. water levels, discharges, ……)
- Leaves lots of scope for different interpretations and assumptions about the nature of different sources of uncertainty
- Model structural error particularly difficult to assess (not easily separated from input and other uncertainties without making strong and difficult to justify assumptions) – often assessed AS IF model is correct
- Therefore lots of uncertainty estimation methods

Statistical Uncertainty Estimation

- Treat the optimal model as if it were the "true" model
- Fit a model to the residuals using appropriate assumptions (e.g. residuals are of zero mean and constant variance and uncorrelated in time/space - or something more realistic, with bias, non-constant variance {heteroscedasticity}, and correlated residuals)
- Nature of error model defines a likelihood function
- Sum of model + error distribution can be used to estimate likelihood (probability) of predicting an observation given the model
- Problem that treating multiple sources of error as if all "measurement error"

Likelihood and Model Evaluation

- Model evaluation normally based on residuals in space and time $\varepsilon(x,t)$

$$\varepsilon(x,t) = O - M(\Theta, I)$$

- Made up of multiple contributions

$$\varepsilon(x,t) = \varepsilon_M(\theta, \varepsilon_\theta, I, \varepsilon_I, x, t) - \varepsilon_C(\Delta x, \Delta t, x, t) - \varepsilon_O(x,t) + \varepsilon_r$$

where $\varepsilon_M(\theta, \varepsilon_\theta, I, \varepsilon_I, x, t)$ is the model error (as affected by parameter and input error

$\varepsilon_C(\Delta x, \Delta t, x, t)$ denotes the commensurability error between observed and predicted values

$\varepsilon_O(x,t)$ is the observation error, and

ε_r is a random(?) error component

Model Calibration and Model Structural Error: formal Bayesian approaches

- As soon as a new observation becomes available then a new residual error can be calculating information from the new observation as

$$\varepsilon = O - M(\Theta, I)$$

- Assuming (for the moment) the model to be unbiased, the contribution to likelihood function from a single residual is assumed (after Gauss) to be given by

$$L(\varepsilon \mid M(\Theta,\mathbf{I})) \propto \exp\left[-\frac{\varepsilon^2}{\sigma_\varepsilon{}^2}\right]$$

Model Calibration and Model Structural Error: formal Bayesian approaches

- Applying Bayes equation over *n* such residuals, assuming independence, the contributions can be multiplied so that:

$$L(\varepsilon \mid M(\Theta,\mathbf{I})) \propto \prod^{n} \exp\left[-\frac{\varepsilon^2}{\sigma_\varepsilon^{\ 2}}\right]$$

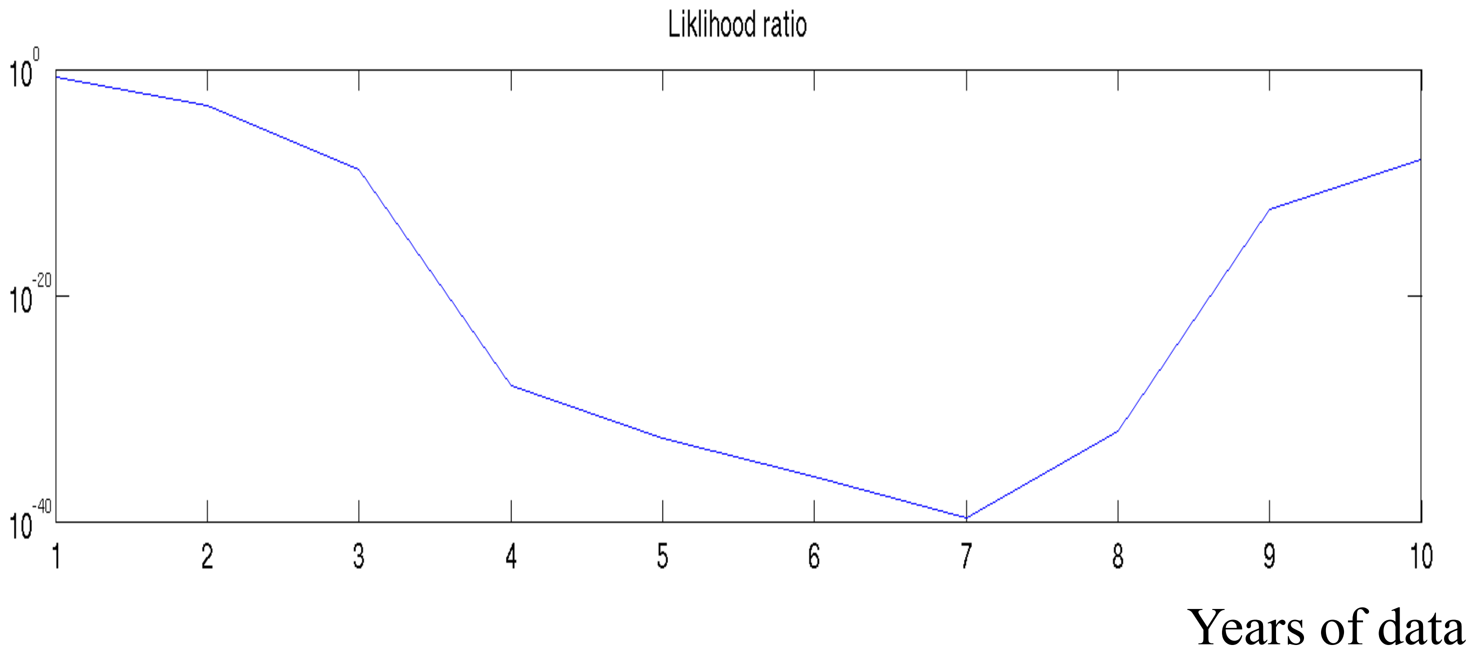
- For the assumption of the Gaussian distributed errors, the final form of the likelihood function is given by.

$$L(\varepsilon \mid M(\Theta,\mathbf{I})) = \left(2\pi\sigma_\varepsilon^{\ 2}\right)^{-n/2} \exp\left[-\frac{1}{2\sigma_\varepsilon^{\ 2}}\left\{\sum_{t=1}^{n}[\varepsilon_t]^2\right\}\right]$$

Do Statistical Error Models lead to Over-Conditioning when Epistemic Uncertainty Important?

Assume standard (aleatory) likelihood with Gaussian information, mean bias, lag 1 correlation.

Likelihood ratio for 2 models with similar error variance

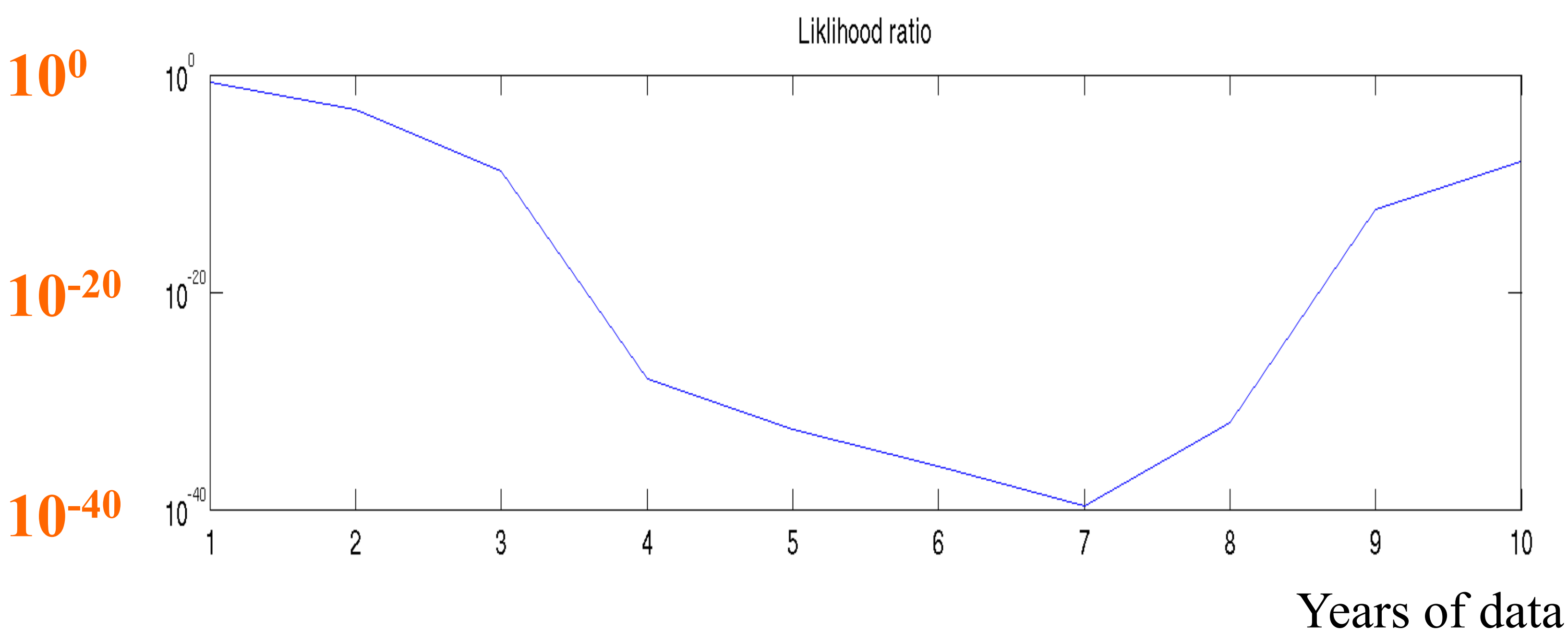


See discussions in Beven JH, 2006; CRAS Geosciences, 2012

Do Statistical Error Models lead to Over-Conditioning when Epistemic Uncertainty Important?

Assume standard (aleatory) likelihood with Gaussian information, mean bias, lag 1 correlation.

Likelihood ratio for 2 models with similar error variance



See discussions in Beven J. Hydrol., 2006; *CRAS Geosciences*, 2012

An aside……

- All these likelihoods based on normal distribution assumptions for the errors derive from Gauss assumption that errors as probabilities are proportional to square of the error (the L_2 norm)

$$L(\varepsilon \mid M(\Theta,\mathbf{I})) \propto \exp\left[-\frac{\varepsilon^2}{\sigma_\varepsilon{}^2}\right]$$

- Earlier work by Laplace (who independently derived a discrete form of Bayes equation in 1816) based on absolute errors (the L_1 norm, see Tarantola, 2005)

$$L(\varepsilon \mid M(\Theta,\mathbf{I})) \propto \exp\left[-|\varepsilon|\right]$$

- Less affected by outliers….but not so convenient for analytical calculations in pre-computer era

Generalised Likelihood Uncertainty Estimation (GLUE)

- Based on rejection of the idea of parameter optimisation: will not be robust to calibration period, performance measure or measurement errors – concept of *equifinality*

  [Beven, J Hydrology, 2006, Manifesto for the Equifinality Thesis]
- Can then only assess the likelihood of different models being good predictors of the system of interest
- Can reject (give zero likelihood) to those models that are not good predictors of calibration data
- Can take account of different model structures as well as different parameter sets
- Can treat complex errors implicitly (so no need for formal error model)

Uncertainty as a likelihood surface in the model space

Basic requirements of a likelihood as belief

- Should be higher for models that are "better"
- Should be zero for models that do not give useful results
- Scaling as relative belief in a hypothesis rather than probability

But how then best to determine weights from evidence given epistemic uncertainties??

Testing models as hypotheses

- Models as multiple working hypothesis about functioning of system – can hypotheses be rejected on basis of uncertain information available?
- Two conflicting requirements (analogy with Type I and Type II errors) – do not want to reject a good model as non-behavioural because of input error & do not want to retain a poor model as behavioural by using a compensatory error model
- JH Manifesto idea – set up limits of acceptability (reflecting observation error, commensurability error and input error) prior to running the model
- But…"Best available" model may not be "fit for purpose" (allowing for uncertainty)

A framework for model evaluation

1. Eliminate obviously disinformative data

2. Set up limits of acceptability (reflecting observation error, commensurability error and input error) *prior* to running the model.

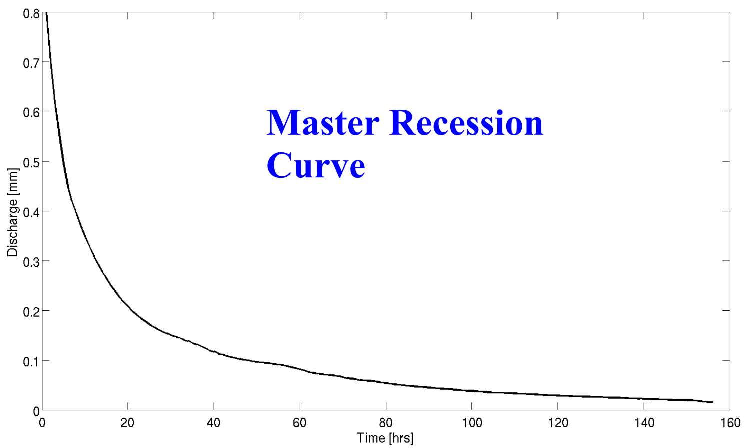
3. For each model run, evaluate performance against limits of acceptability

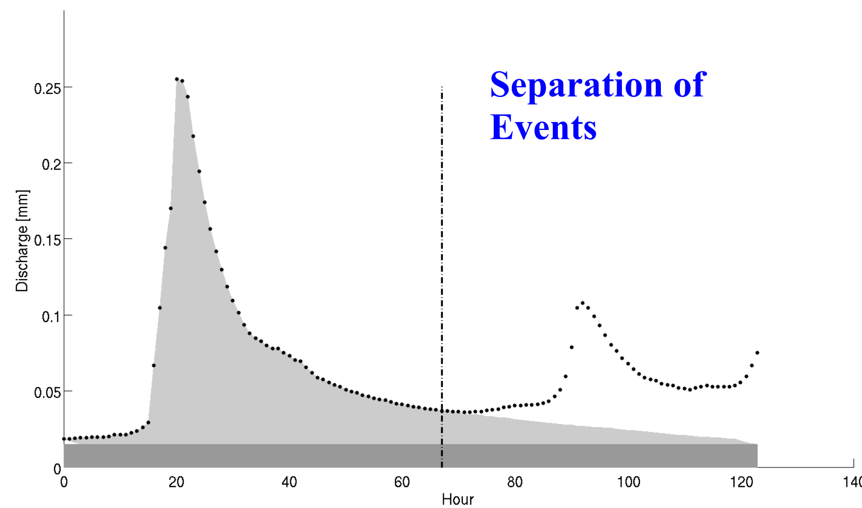
4. Check for error reconstruction to improve predictions / calculate distributions of errors.

Identifying disinformative data

First criterion: Event mass balance consistency (expectation that event runoff coefficient Q / R will be less than one)

But…difficulty of separating events

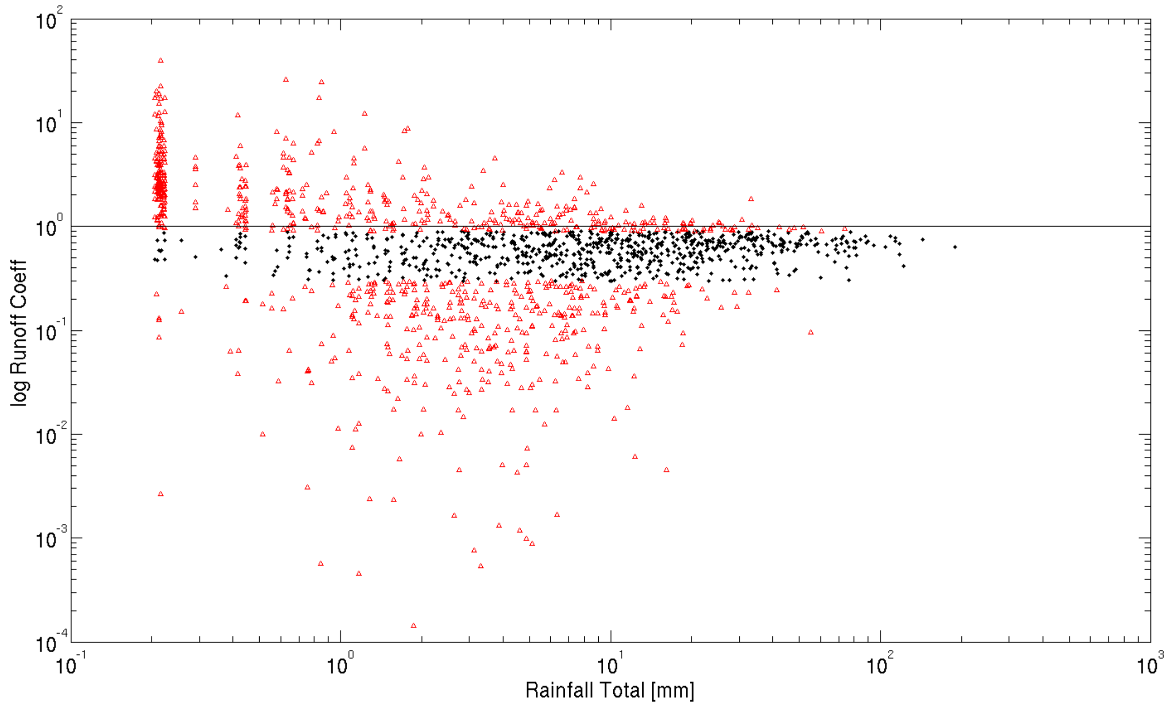




and impact of an inconsistent event on model results might persist for following events, gradually decaying

Setting Limits of Acceptability prior to running a model

Results of runoff coefficient determination for River Tyne at Station 23006 – plotted against rainfall totals over catchment area as estimated from 5 gauges (black – range 0.3 to 0.9)



Beven et al, HESS, 2011

Limits of acceptability

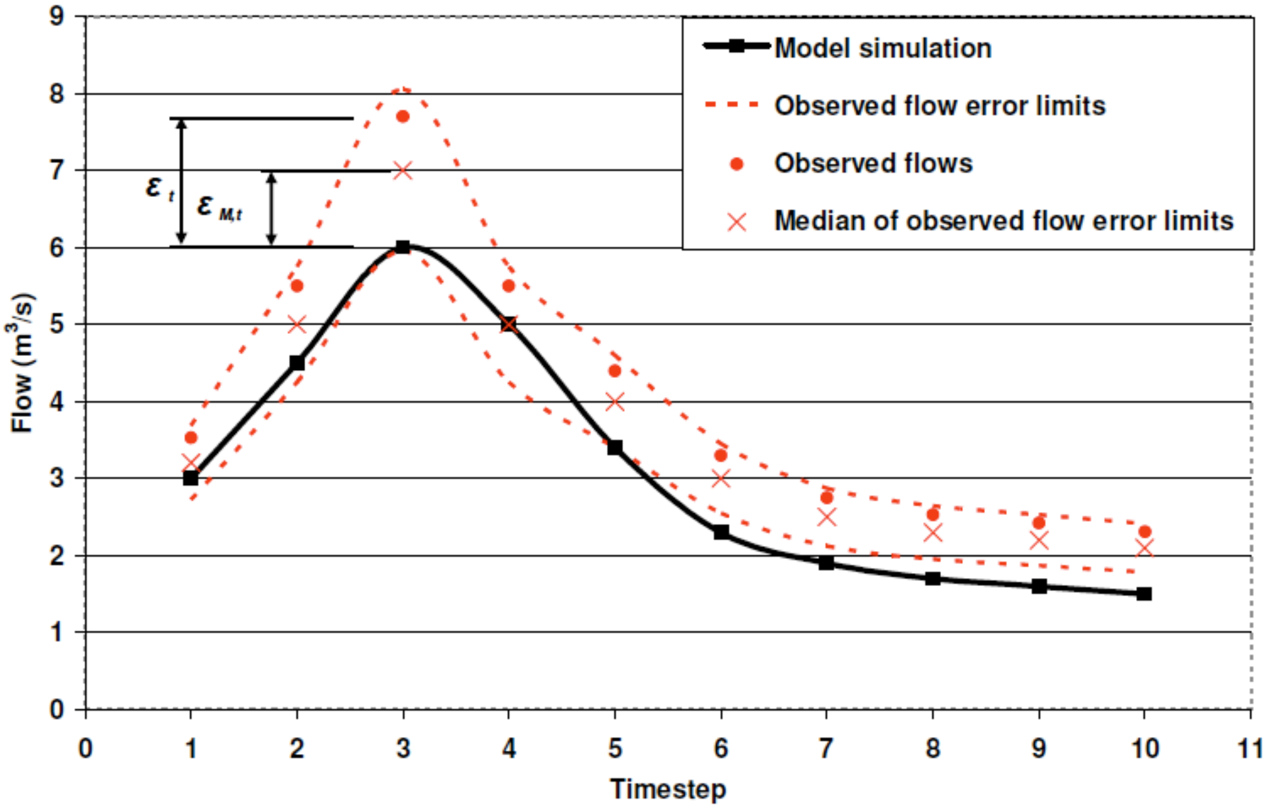
- The question that then arises within this framework is whether, for an particular realisation of the inputs and boundary conditions, $\varepsilon_M(\theta, I, \varepsilon_I, x, t)$ is acceptable in relation to the terms $\varepsilon_O(x,t) + \varepsilon_C(\Delta x, \Delta t, x, t)$. This is equivalent to asking if the following inequality holds:

$$O_{min}(x,t) < M(\theta, I, \varepsilon_I, x, t) < O_{max}(x,t) \text{ for all } O(x,t)$$

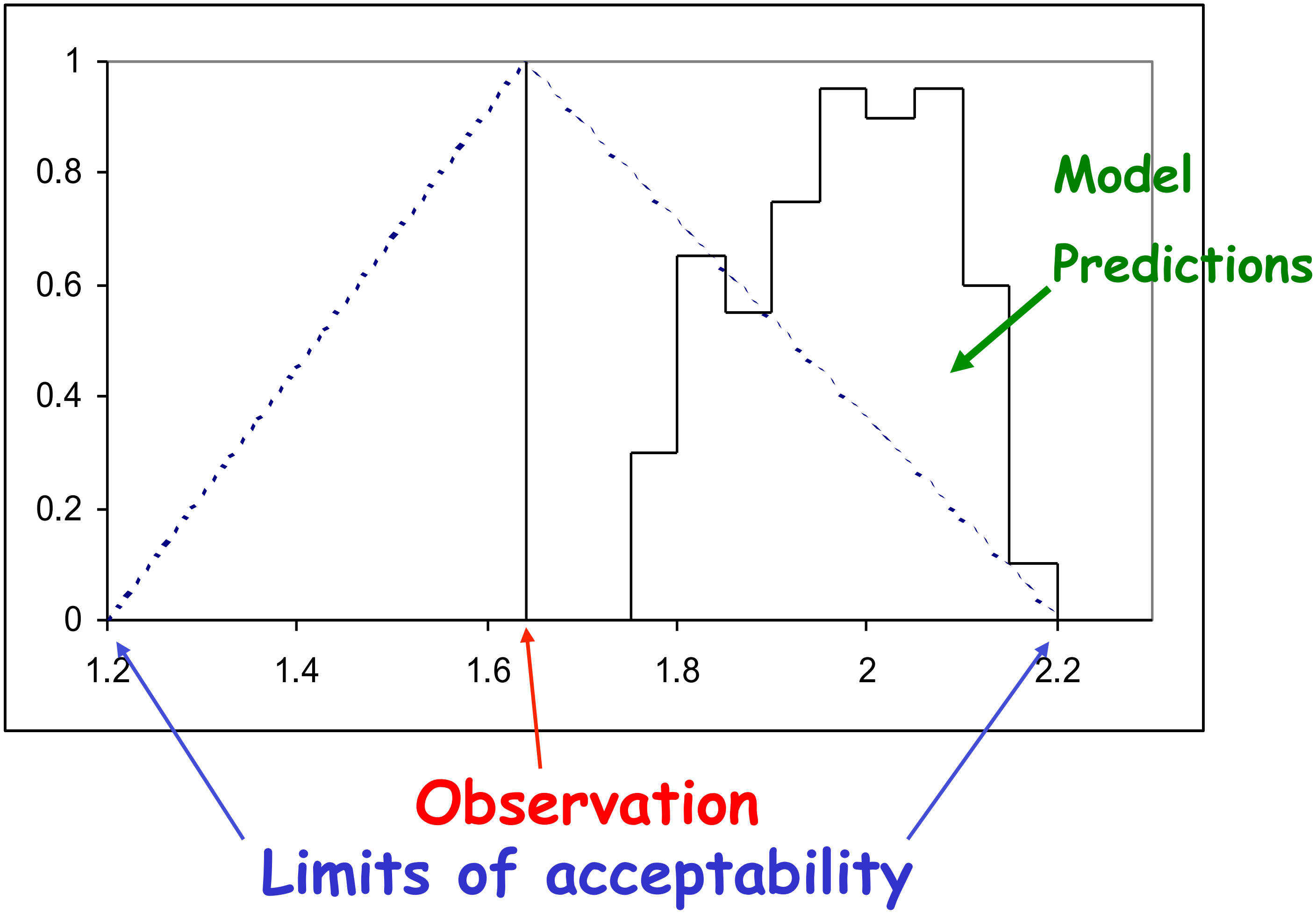
  where $O_{min}(x,t)$ and $O_{max}(x,t)$ are acceptable limits for the prediction of the output variables *given* $\varepsilon_O(x,t)$ and $\varepsilon_C(\Delta x, \Delta t, x, t)$
- Idseally, limits of acceptability should be evaluated prior to running the model (but note I, ε_I in $M(\theta, I, \varepsilon_I, x, t)$)

Model Evaluation using Limits of Acceptability

Likelihood can be developed based on scaled deviation away from observation, with zero value at any time step that prediction lies outside limits.



Predictive distribution over all behavioural models: what if predictions do not encompass new observation

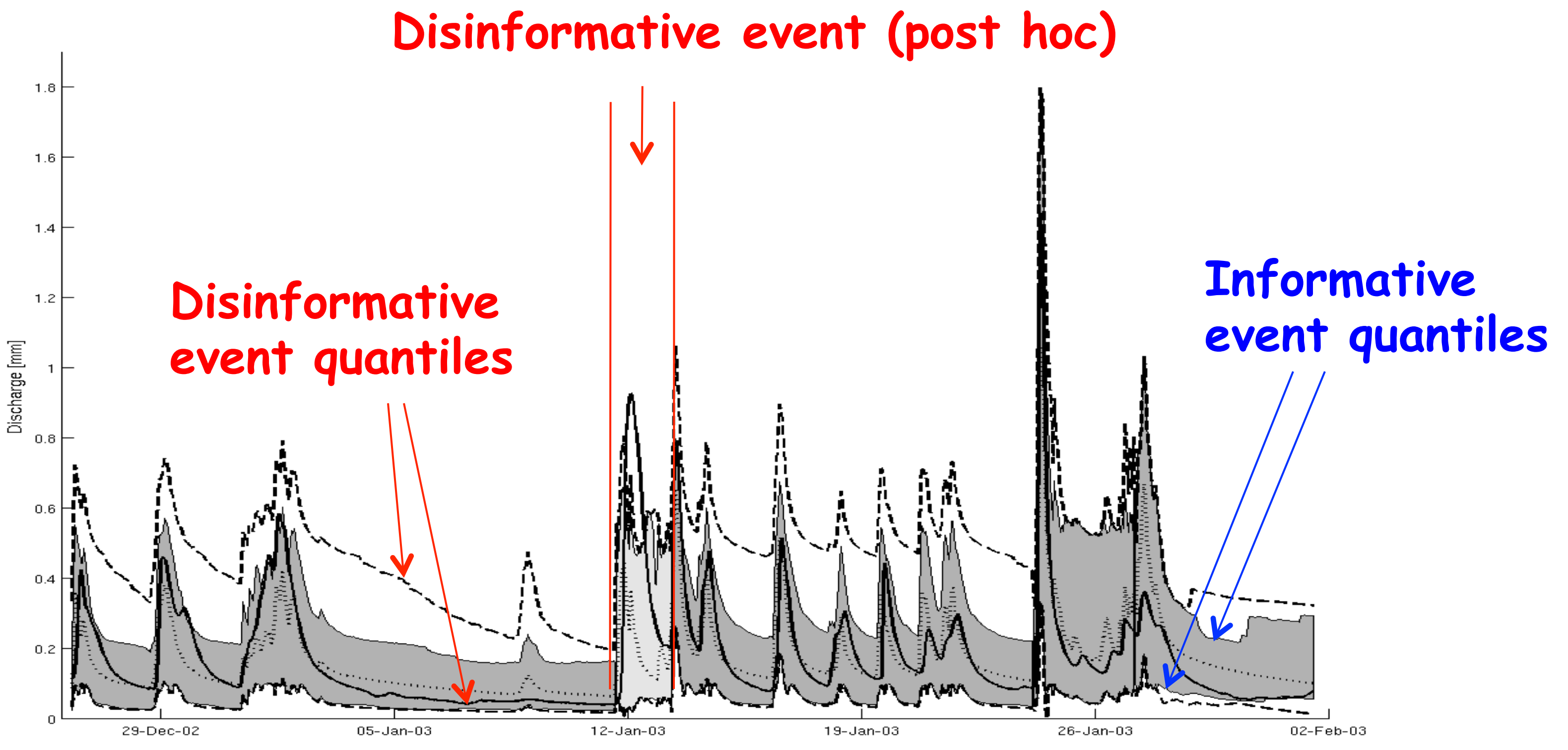


Allowing for disinformation

If disinformation can be identified in calibration, what to do in prediction?

- Given only input data we do not know if the next event will be consistent or disinformative (can only differentiate a postiori when response is observed)
- We might still be surprised (the 2008 crash as a “25σ event”)

Informative and disinformative prediction bounds in validation



See Beven & Smith, JHE, accepted

River Eden: January 2005 event

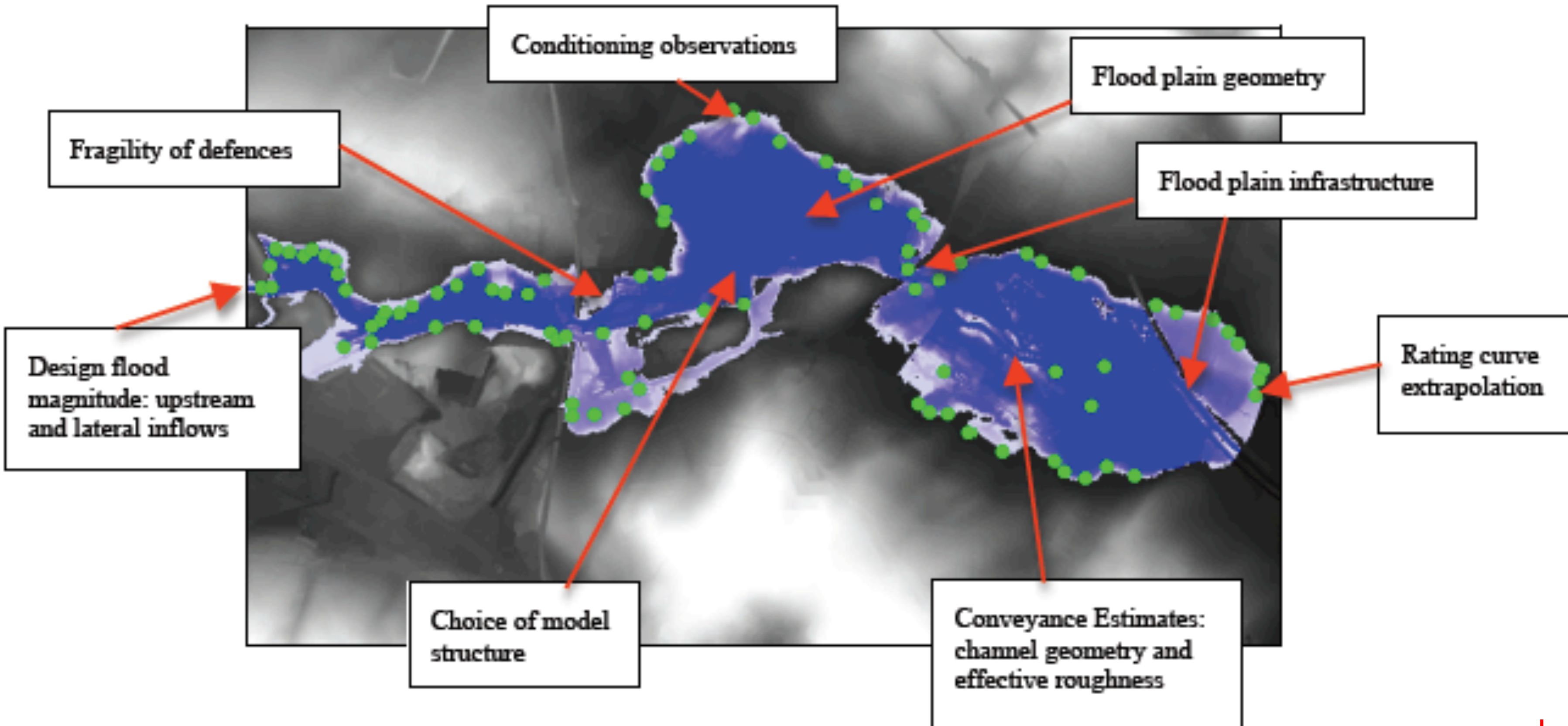
Upstream at Appleby

Emergency Centre
at Carlisle

Public response
at Carlisle

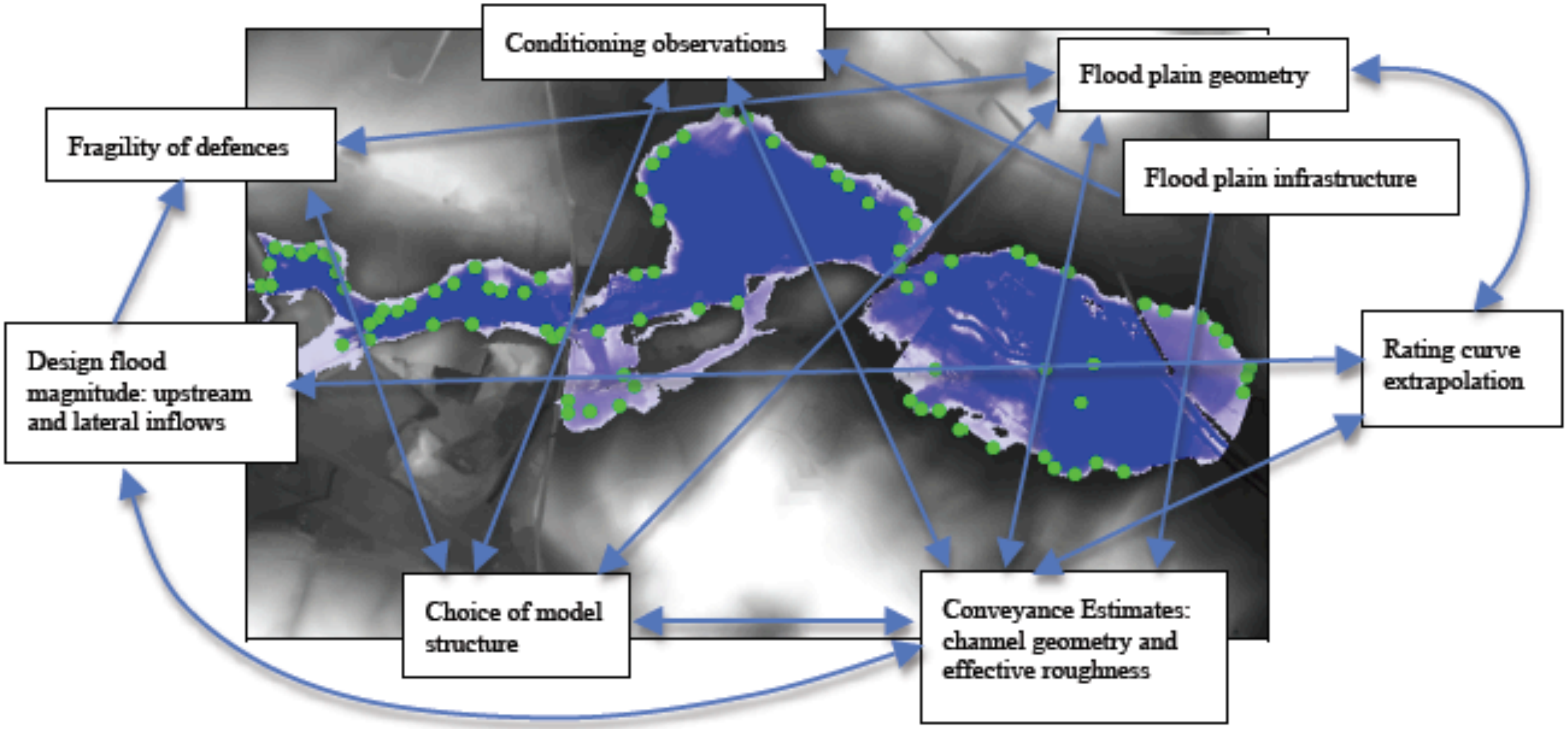
LANCASTER
UNIVERSITY

Sources of Uncertainty in Flood Risk Mapping



LANCASTER UNIVERSITY

Interactions between Sources of Uncertainty



LANCASTER
UNIVERSITY

Types of error in flood risk mapping

Source of uncertainty	Aleatory errors	Epistemic nature
Design flood Magnitude	What is the range of sampling variability around underlying distribution of flood magnitudes?	Are floods generated by different types of events? What frequency distribution should be used for each type of event? Are frequencies stationary? Will frequencies be stationary into the future?
Conveyance estimates	What is the random sampling variability around estimates of conveyance at different flood levels?	Is channel geometry stationary over time? Do conveyance estimates properly represent changes in momentum losses and scour at high discharges? Are there seasonal changes in vegetation in channel and on floodplain? Is flood plain infrastructure, walls, hedges, culverts etc. taken into account?
Rating curve interpolation and extrapolation	What is standard error of estimating the magnitude of discharge from measured levels?	Is channel geometry stationary over time? What is estimation error in extrapolating rating curve beyond the range of measured discharges? Does extrapolation properly represent changes in momentum losses and scour at high discharges?

UPPSALA UNIVERSITET

(from Beven and Alcock, *Freshwater Biology*, 2011)

LANCASTER UNIVERSITY

Types of error in flood risk mapping

Source of uncertainty	Aleatory errors	Epistemic nature
Flood plain Topography and Infrastructure	What is the standard error of survey errors for flood plain topography? What is the random error in specifying the positions of elements, including elevations of flood defences?	Are there epistemic uncertainties in correction algorithms in preparing digital terrain map? How should storage characteristics of buildings, tall vegetation, walls and hedges in geometry be treated? Are there missing features in the terrain map (e.g. walls, culverts)?
Model structure		How far do results depend on choice of model structure, dimensions, discretisation and numerical approximations?
Observations used in model calibration/ conditioning	What is the standard error of estimating a flood level given post-event survey of wrack marks or gauging station observations?	Is there some potential for the misinterpretation of wrack marks surveyed after past events? Are there any systematic survey errors?

UPPSALA UNIVERSITET

LANCASTER UNIVERSITY

Types of error in flood risk mapping

Source of uncertainty	Aleatory errors	Epistemic nature
Future catchment change		What process representations for effects of land management should be used? What future scenarios of future change should be used? Are some scenarios more likely than others?)
Future climate change	What is the variability in outcomes owing to random weather generator realisations?	How far do results depend on choice of model structure? What process representations in weather generators should be used? What future scenarios of future change should be used? Are some scenarios more likely?
Fragility of Defences Consequences/ Vulnerability	What are the probabilities of failure under different boundary conditions? What is the standard error of for losses in different loss classes?	What are the expectations about failure modes and parameters? What knowledge about uncertainty in loss classes and vulnerability is available?

UPPSALA UNIVERSITET

LANCASTER UNIVERSITY

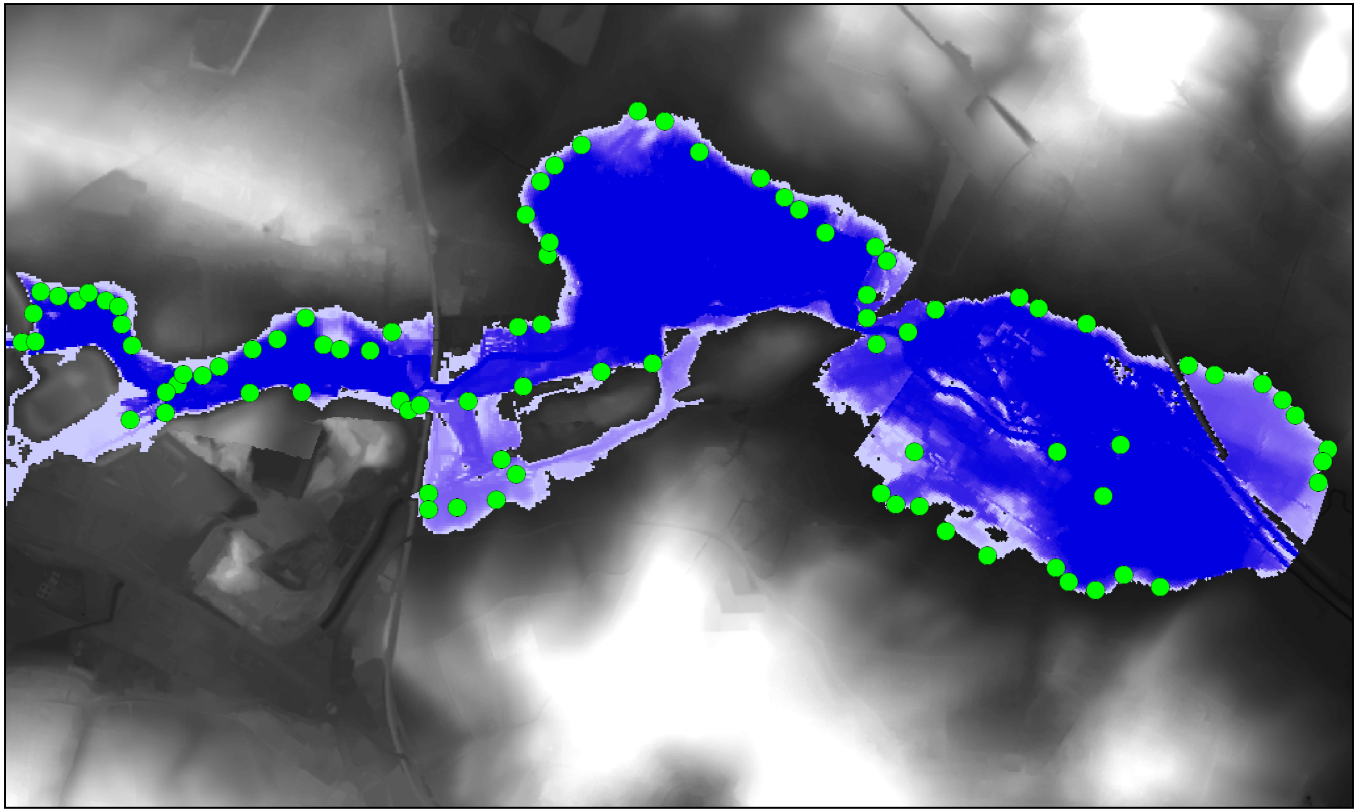
The GLUE methodology

1. Decide on limits of acceptability prior to running model
2. Decide on prior distributions for parameters
3. Run the model with many parameter sets chosen randomly from priors
4. Evaluate the model against available observations
5. Reject models that are "non-behavioural"
6. Weight the remaining models by some likelihood measure
7. Use ensemble of models in prediction (with implicit or explicit handling of residual errors)

LANCASTER UNIVERSITY

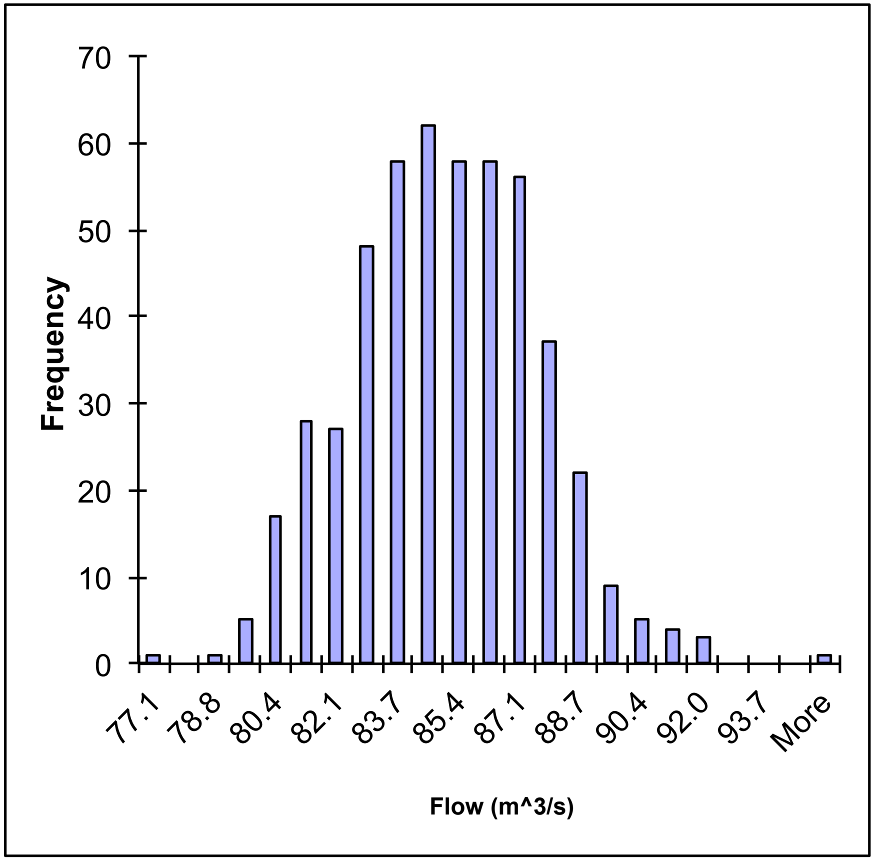
Mexborough: Summer 2007

Mapped maximum inundation and model predicted flow depths for Summer 2007 floods at Mexborough, Yorkshire using 2D JFLOW model



Mexborough Risk Mapping: Defining Input Uncertainties

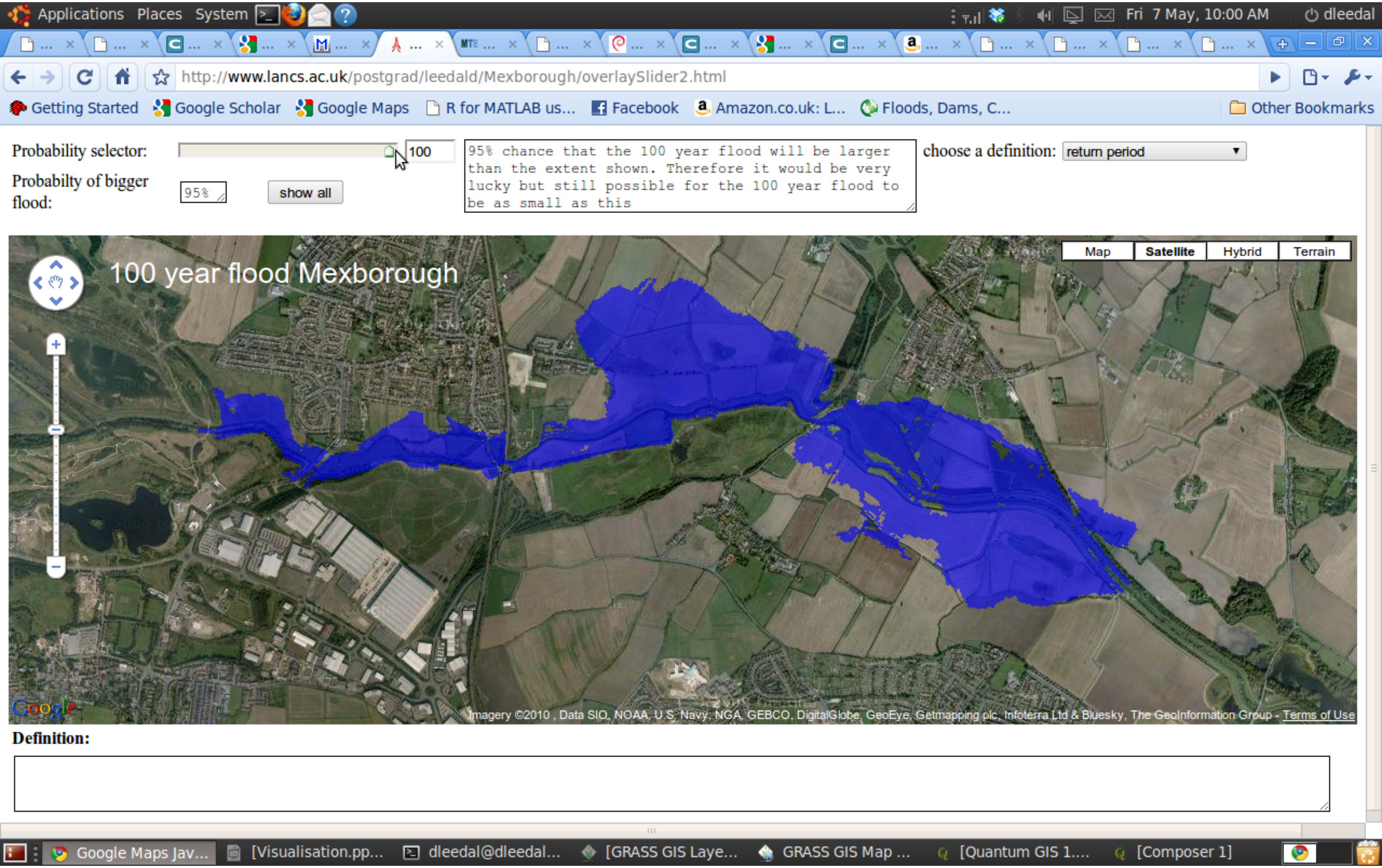
WinFAP estimate of 0.01 AEP (T100) flood peak at Adwick



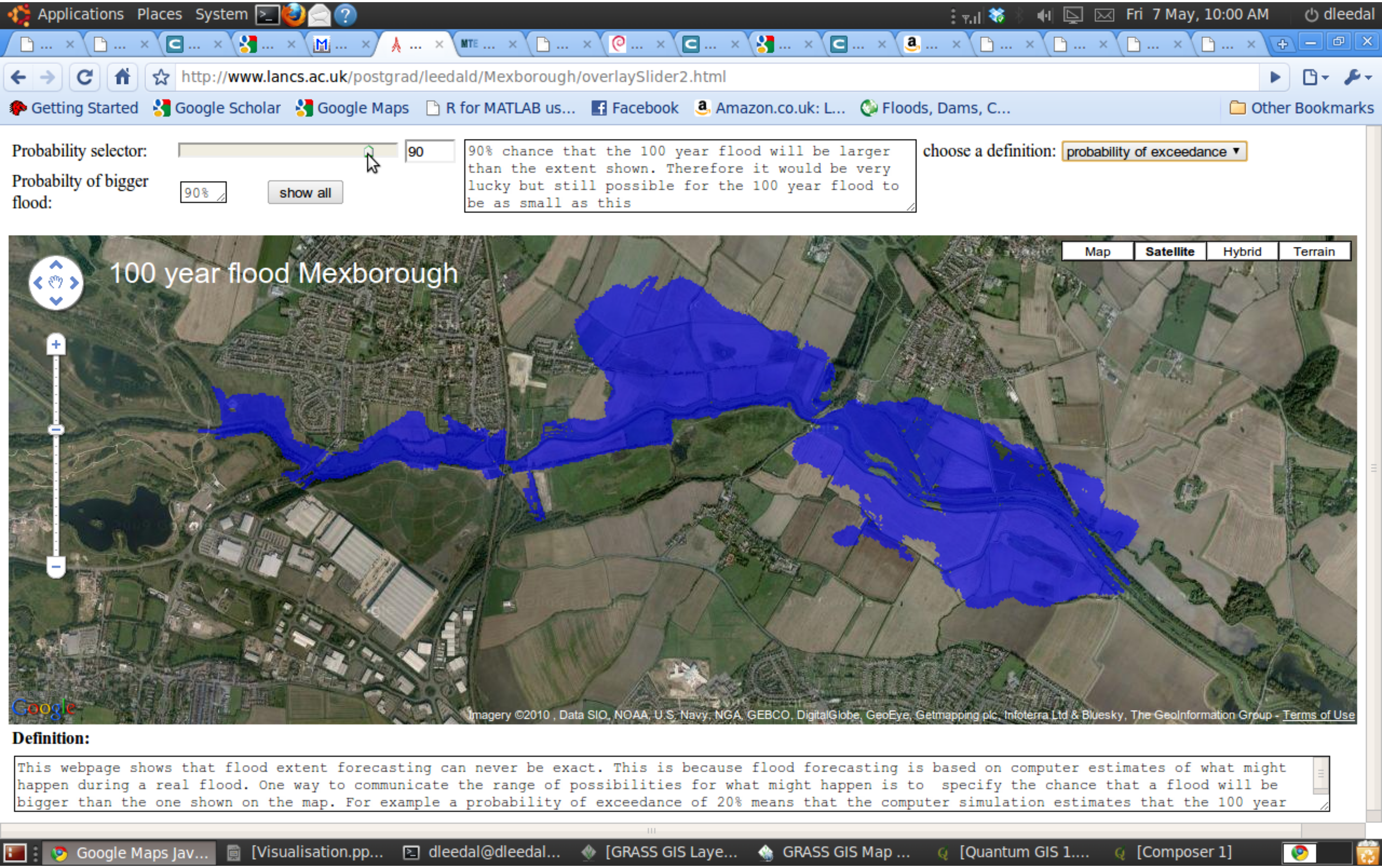
Mean: 86.6 (m^3s^{-1})

Var: 6.25 (m^3s^{-1})

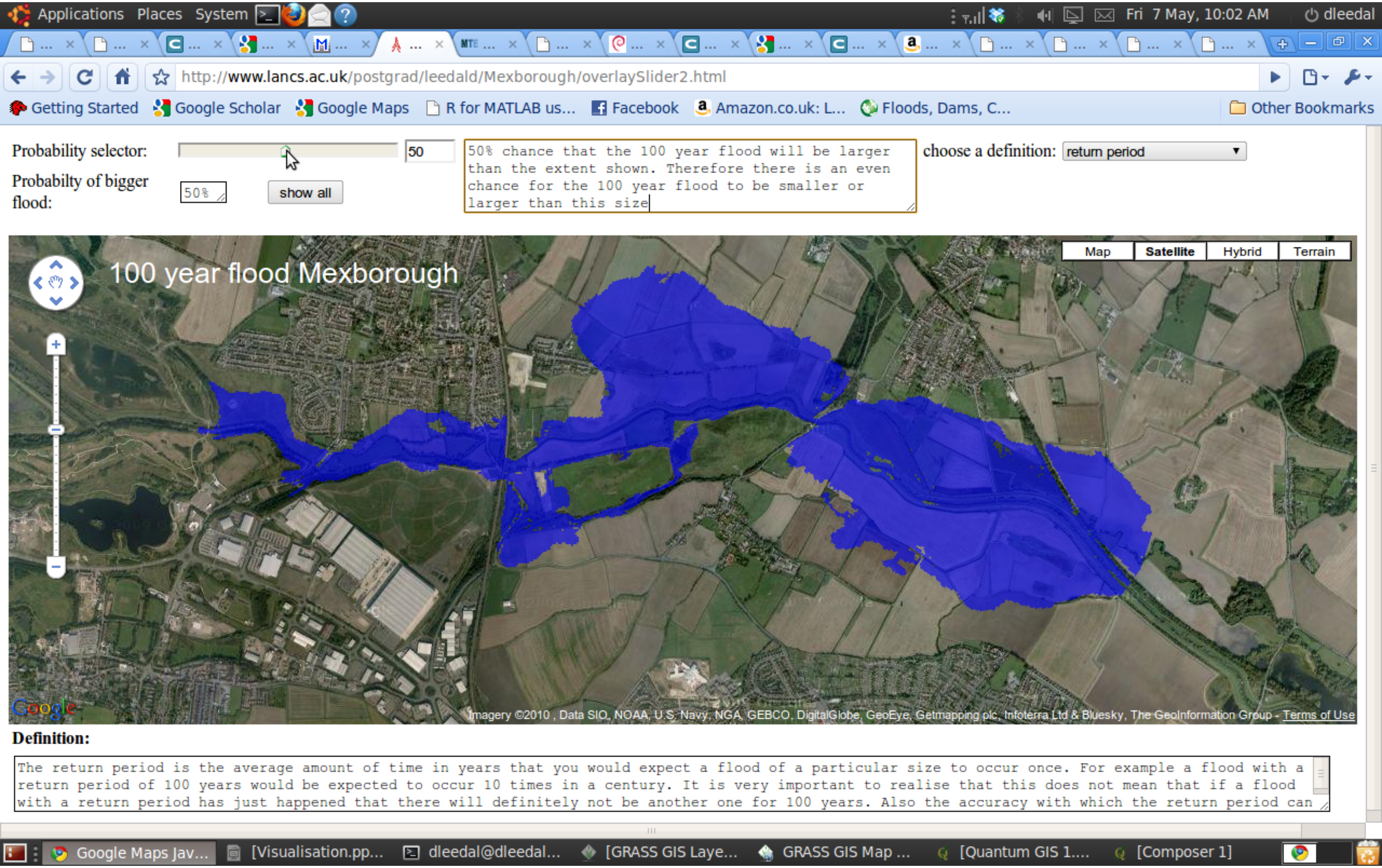
Google maps API



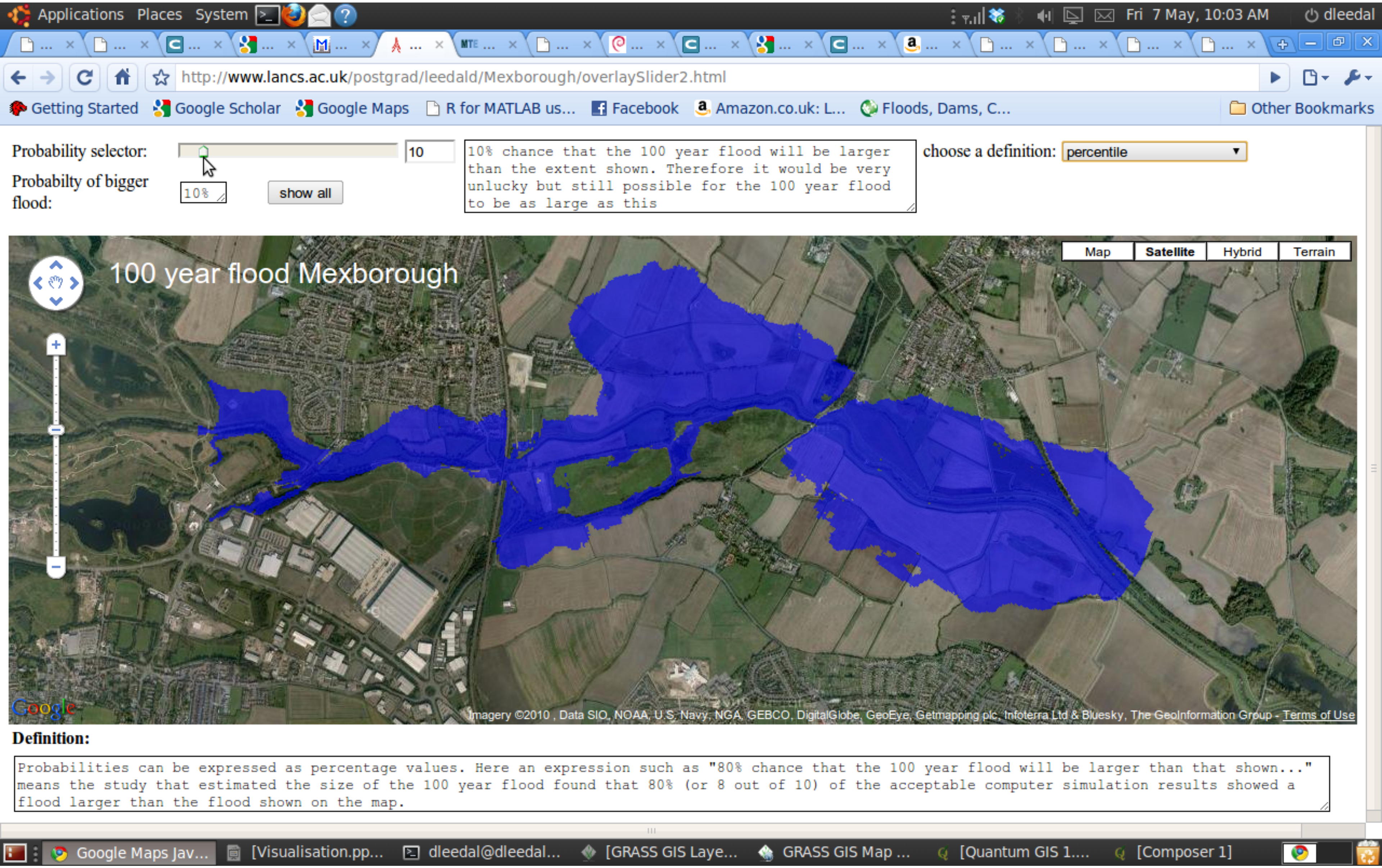
Google maps API



Google maps API



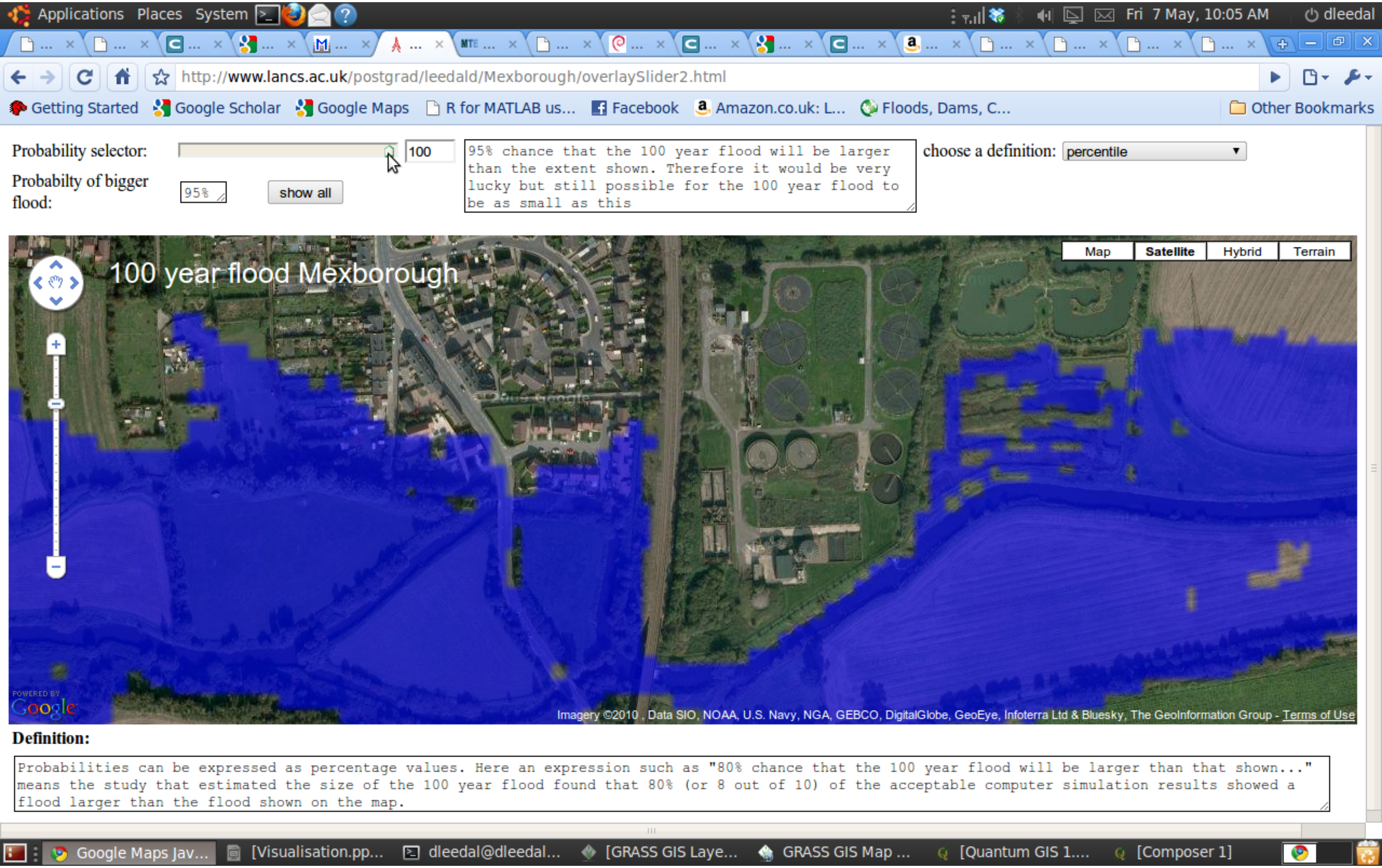
Google maps API



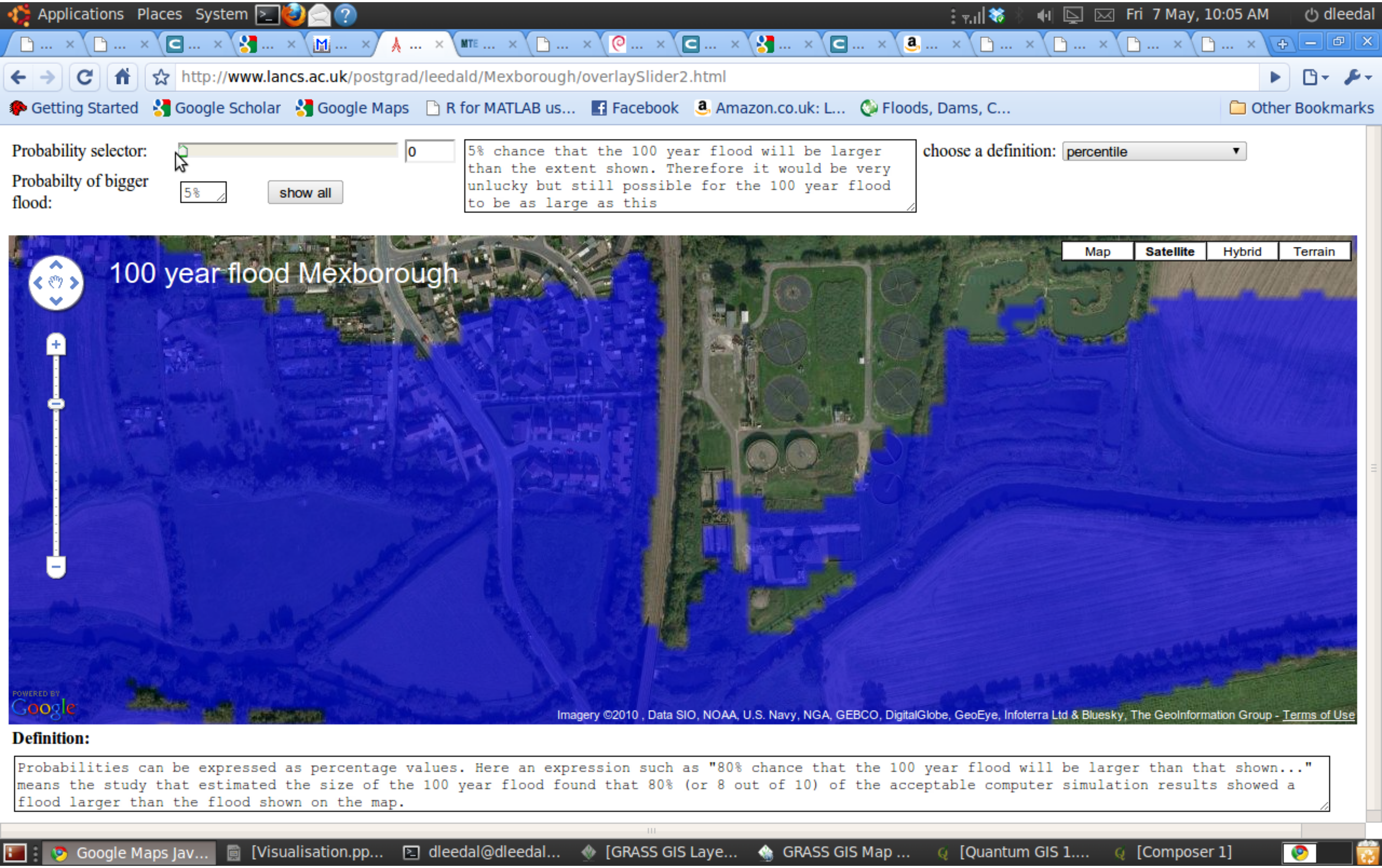
Google maps API

The image cannot be displayed. Your computer may not have enough memory to open the image, or the image may have been corrupted. Restart your computer, and then open the file again. If the red x still appears, you may have to delete the image and then insert it again.

Google maps API



Google maps API



Differences between formal Bayes and GLUE

Formal Bayes

- Uncertainty has to be expressed as probability of predicting an observation conditional on the model (still underlying idea of finding "the model")
- A suitable formal likelihood to represent total error or error sources can always be found (and is consistent across model parameter sets)
- Data is always coherent and information content of new data can be evaluated by looking at change in discrepancy measure
- Prediction bounds should bracket data to required 90/95% level
- Poor use of likelihoods (e.g. ignoring autorcorrelation or non-Gaussian residuals) is poor practice

Differences between formal Bayes and GLUE

GLUE

- Acceptance of equifinality in face of model error, input error and observation error (epistemic errors in general)
- Uncertainty is not a probability of predicting an observation but expressed as empirical prediction limits for behavioural models after weighting by informal likelihood measure / degree of belief in a model - residuals associated with each behavioural model handled implicitly
- Prediction limits may then not bracket observations to any specified level depending on model and input error (but if model can span the observations they might do)
- Formal likelihoods can be used as a special case, error model simply becomes an additional model component
- Poor interpretation of likelihood weights is poor practice

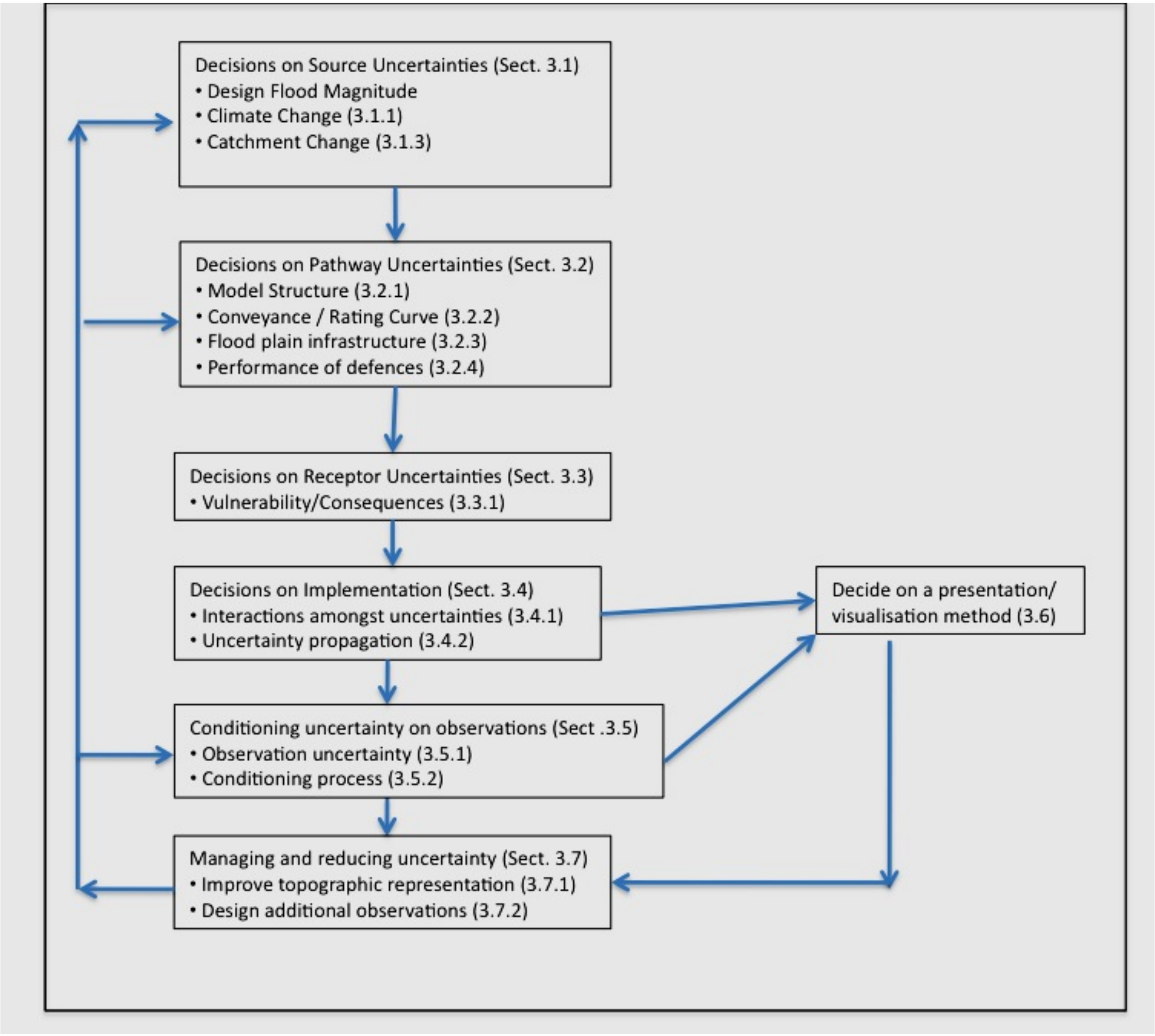
Risk Communication

- Decision makers are not interested in uncertainties (?)
- But uncertainty might make a difference to the decision that is made (being more risk averse where consequences might be catastrophic)
- But how best to convey meaning of uncertainty estimates (Faulkner et al., *Ambio*, 2007)
- AND the assumptions on which those estimates are based (Beven and Alcock, *Freshwater Biology*, 2011)

UPPSALA UNIVERSITET

LANCASTER UNIVERSITY

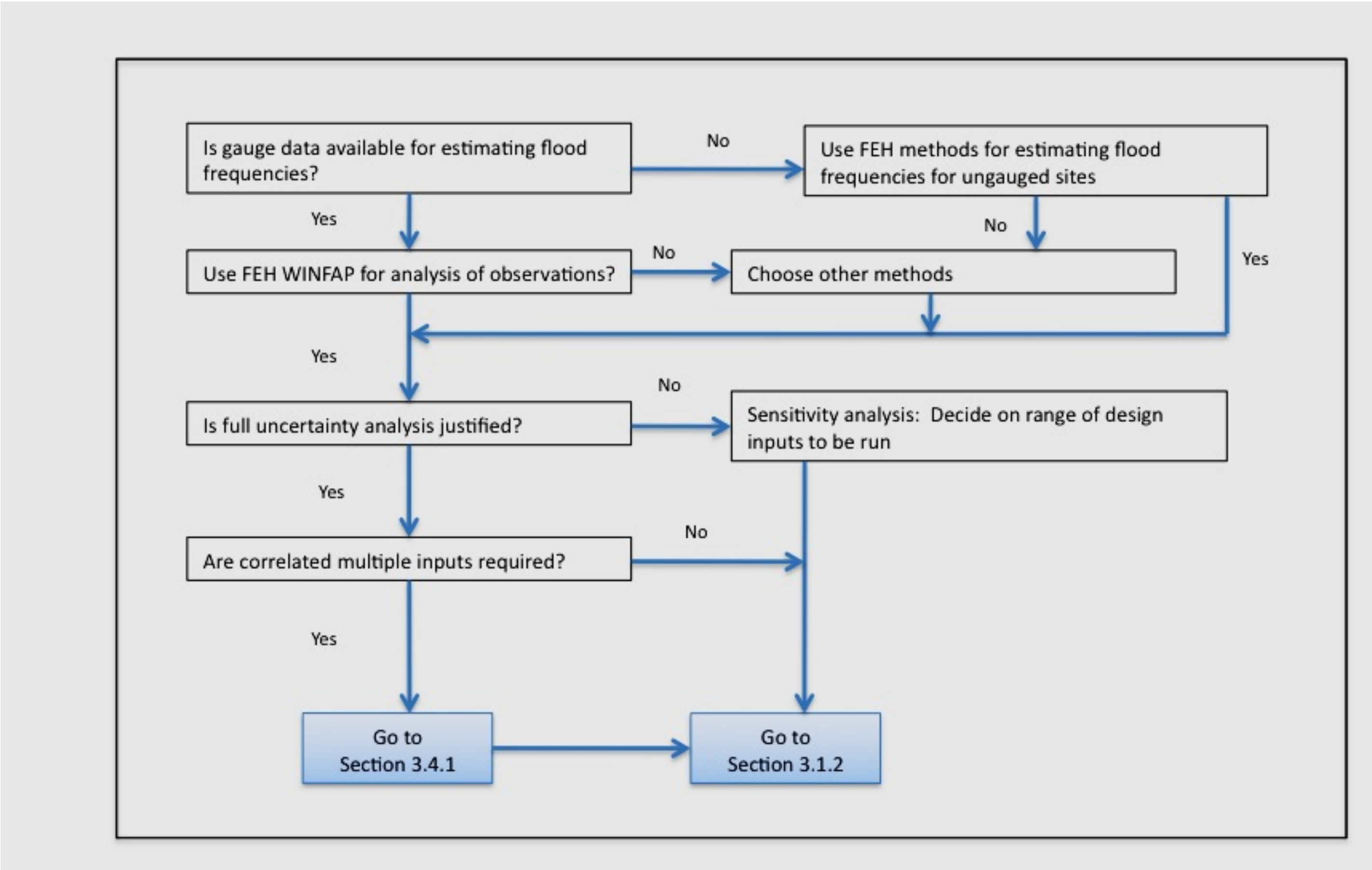
Communicating the meaning of uncertainty estimates: Case of Flood Inundation



LANCASTER UNIVERSITY

Guidelines for Good Practice as Condition Trees

e.g. Decision Structure for design flood estimation



LANCASTER UNIVERSITY

Summary

- May be impossible to separate out sources of error from series of model residuals
- Epistemic sources of uncertainty result in non-stationarity in error characteristics (and potential for surprise in prediction)
- Treating all uncertainties as aleatory can lead to dramatic over-conditioning
- GLUE approach - allows for equifinality of model structures and parameter sets
- Limits of acceptability approach as an alternative to statistical testing of models as hypotheses
- Discussion and agreement regarding assumptions of analysis provide a basis for communication of concepts

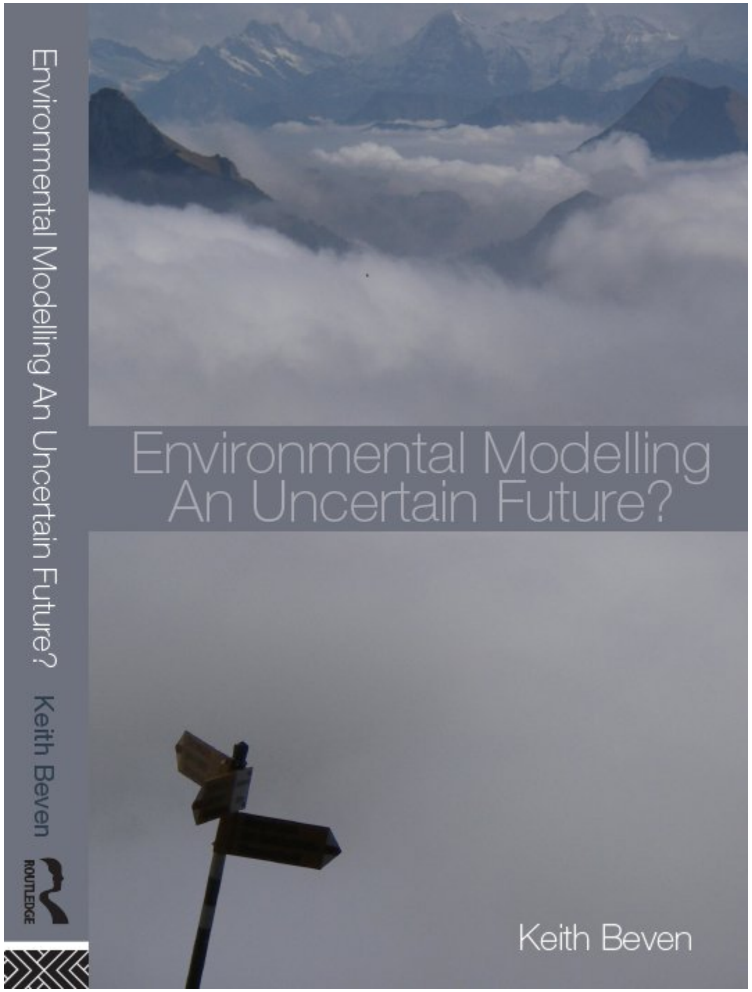
A paradox to think about in your own work…

- Generally, the more physical understanding that is built into a model, the more parameter values must be specified to run the model
- The more parameter values that cannot be estimated precisely, the more degrees of freedom that will be available in fitting the observations (we cannot measure effective parameters everywhere).
- Therefore the more physical understanding that is built into a model, the greater the problem of equifinality is likely to be.
- A “perfect” model with unknown parameters is no protection against equifinality

Still be done

1. Reducing uncertainty depends more on better observation techniques than better model structures. Better model structures might still be achieved - need more tests for both flow <u>and</u> transport
2. Will still have to accept that models of both hazard and consequences will have limited accuracy – so uncertainty estimation will remain important (better likelihood measures and need to identify disinformative data)
3. Should look more carefully at epistemic uncertainties that might lead to surprise – particularly in respect of future non-stationarities (are we looking at a wide enough range of scenarios?)
4. Need to communicate the meaning of predictions (and uncertainties) to decision makers – it may make a difference to the decision process.

More on these techniques……



in
Environmental Modelling: An Uncertain Future?

Routledge, 2009
ISBN: 0-415-46302-2

More information at
www.uncertain-future.org.uk