

On measures of model behaviour (with a hydrological modelling bias - sorry!)

Jim Freer



Introduction - Aims of the talk

- Broad perspective - A set of comments
- Deliberately Controversial, the 'devils advocate' - I hope people will disagree!
- Food for Thought - Get us all thinking I hope!
- Take a step back and think about our abilities to quantify the real world!
- Reminder: The concepts are geared towards the goals of PUB

Start with a quote.....

Clay Brandow from Watershed Management Council (*apparently*
- Advancing the art and science of watershed management)

“Fact is that with computer models we can easily trick ourselves into thinking we know something about which we are still profoundly ignorant. That's worse than ignorance, that's arrogance, and it often comes before the fall...

...file that under the Empire Strikes Back, the former British Empire that is”.

Clearly he's not advancing diplomacy then!

What Constitutes a Behavioural Model?

- Depends on the Application (Environmental Models)
- How to Define a Behavioural Model - Is it:

Fit for purpose?

A model that is consistent with the available observations?

It worked OK so I didn't really try too hard to falsify the model!

The problem is that we can always reject conceptual / process models if we try hard enough (i.e. perceptual weaknesses)

Behavioural Models?

- So we should accept that our model is in error to some extent, the question is what still makes it behavioural?

Forecasting - Use the minimum variance solution

For Simulation Experiments (conceptual etc) - A model that is consistent with the observations, but not overly bias by the potential for error in those same observations

Also..... If we do not have enough measurements to close the water balance should our models?

Verification of Models?

- When is a model verified. Consider the hierarchal verification system of Klemes (1986):

1) split sample test (the lowest level)

2) proxy-area test

3) differential split sample (perturbation)

4) proxy-area differential split sample (PUB!)

Can a model be verified (Oreskes et al. 'confirmed')? Are these categories necessary?

Assessing Model Performance (huge topic!)

- There are many standard PM used in Hydrological Modelling (r^2 , SAE, HMLE, AMLE, RMSE, BIAS, etc). Each measure has / is:

Sensitive towards performance for different parts of the data

Assumptions regarding the nature of the error structures (i.e. zero mean, constant variance, normality and independence)

Difficulties in saying what constitutes an acceptable level of performance for different types of data

- Can we make PM more robust using a variety of methods

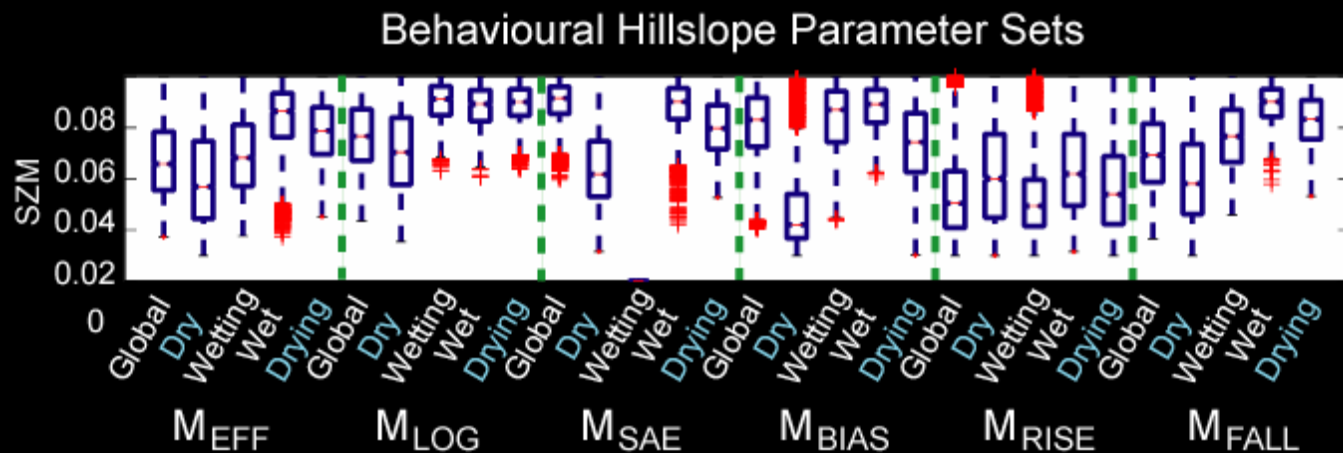
More robust analysis - Rejecting Models?

1993 Simulations – **Dry**, **Wetting**, **Wet** and **Drying** Seasonal Periods and 6 Performance Measures

Initial **Global** M_{EFF} Behavioural Parameter sets **10,850**

After Seasonal multi-objective Analysis **0**

Why 0?



Conclusion: We need to be more thoughtful about assessing the performance of our models – also more ‘expert analysis’ approaches of Boyle et al.

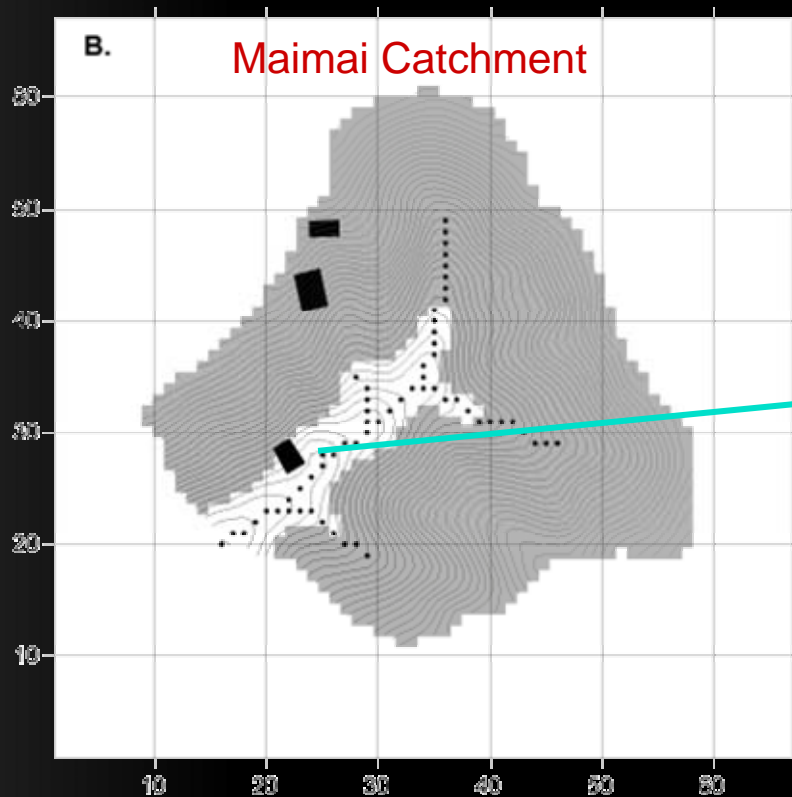
Freer, J.E., Beven, K.J. and Peters, N.E., 2003

Is Rainfall-Runoff data ever enough?

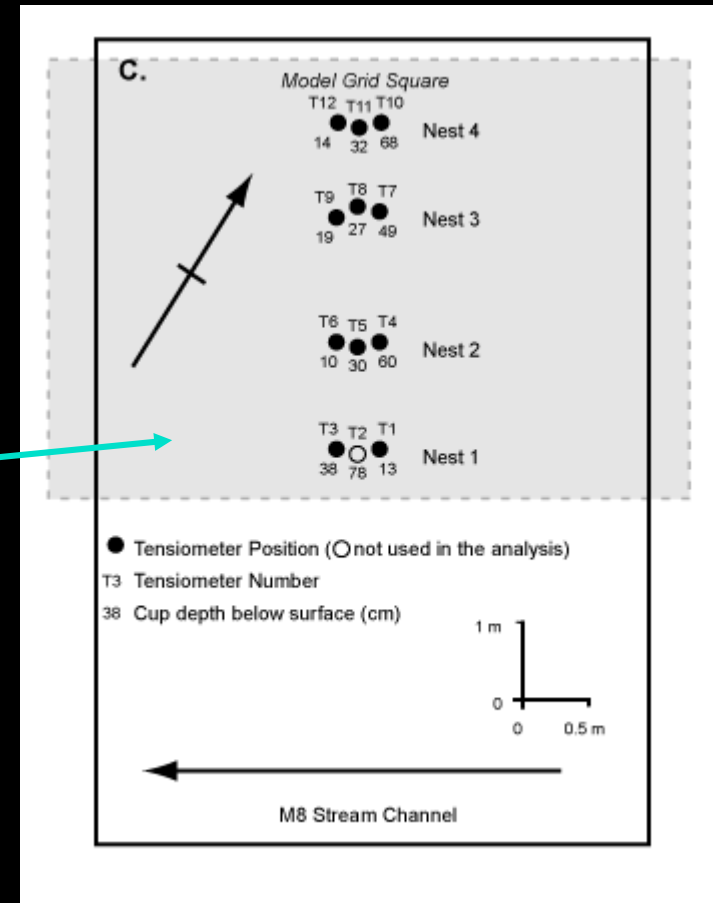
- Can / should we accept / reject models on rainfall - runoff data alone?
- When should we use 'multi-response' data - Should we expect simplified models to represent local processes?
- How do we weight this data - how representative is it (point measurements)?
- What about the modelling paradox?

Maimai Example - Multi-response Analysis

Main Objectives - Making use of diverse data to improve model identifiability and ensuring spatial data 'uncertainties' were properly accounted for at the effective model scale

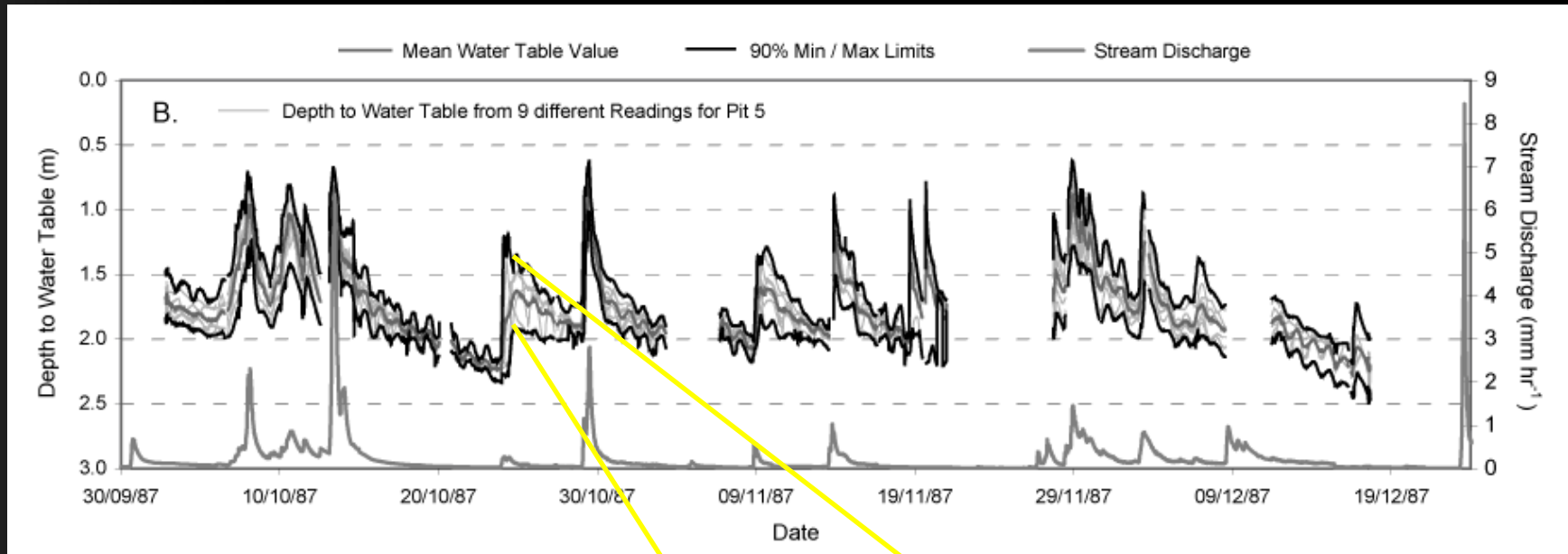


Rainfall-Runoff and Tensiometer Data



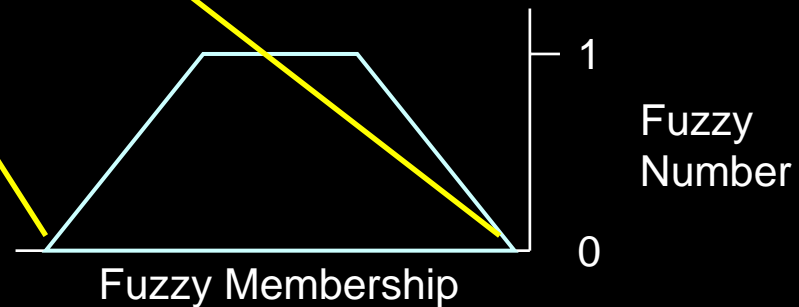
Maimai Example - Multi-criteria Analysis

Uncertainty in Water Table Position for the Effective Model Grid Square

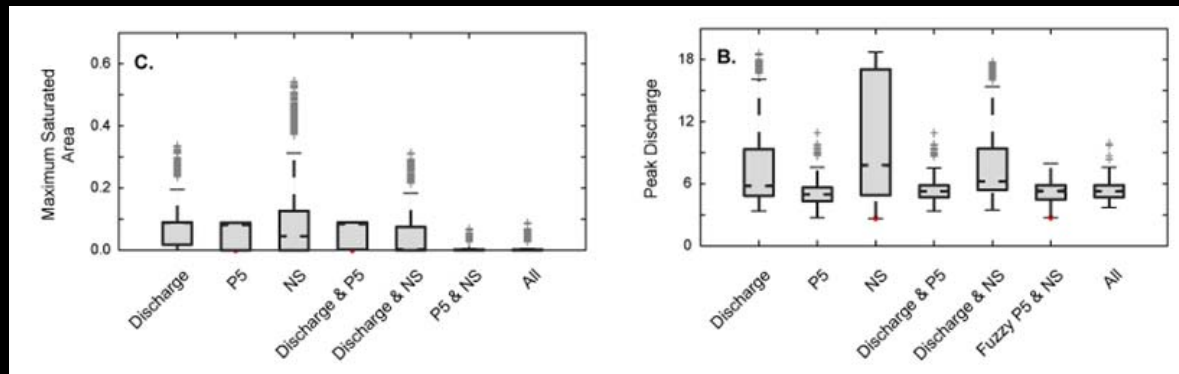
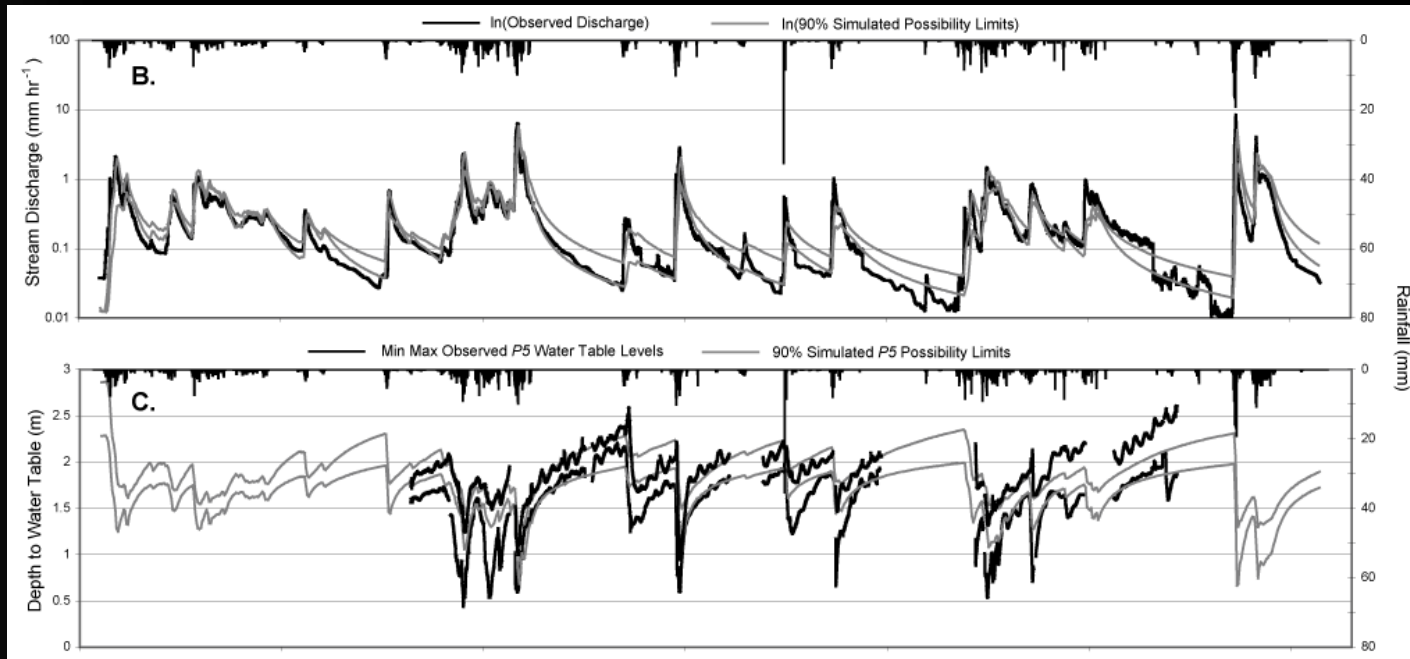


Convert to a fuzzy membership function for each time step

Use this to evaluate model Performance (additive)



Maimai Example - GLUE Results



Conclusion: Important to include uncertainty estimates and use diverse data

Quantifying Measurement Uncertainty

- Observations have a certain measurement error due to the instrumentation - analytical error

but this is not the whole story

- There are commensurability issues - the integration scale of the observations compared to the model variable
- There are commensurability issues - how to measure parameters at the effective gridscale of models?

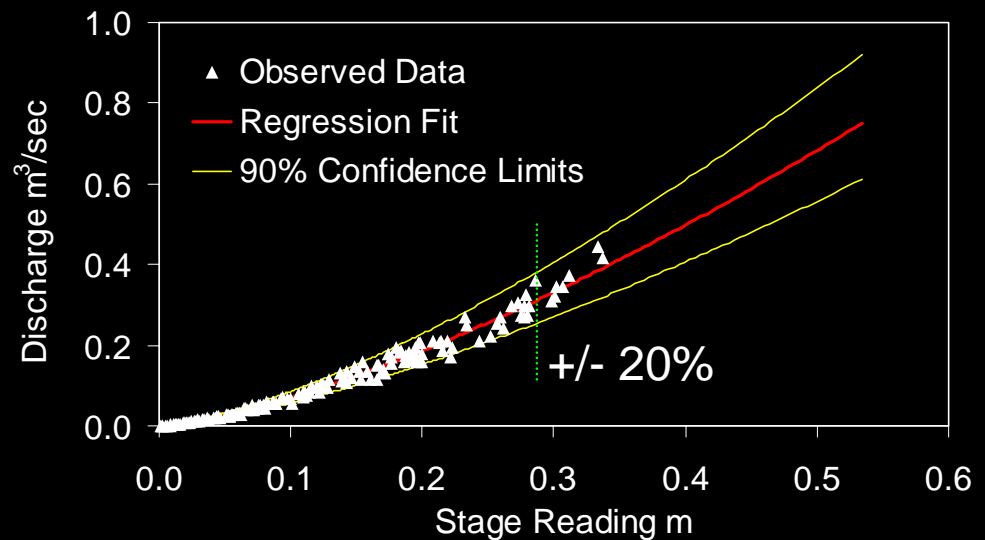
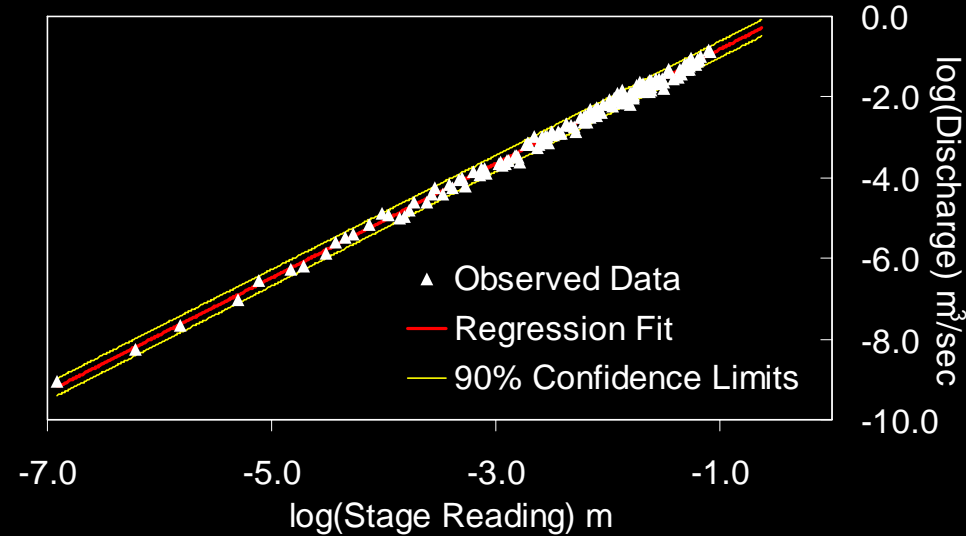
Stage-Discharge Errors - example



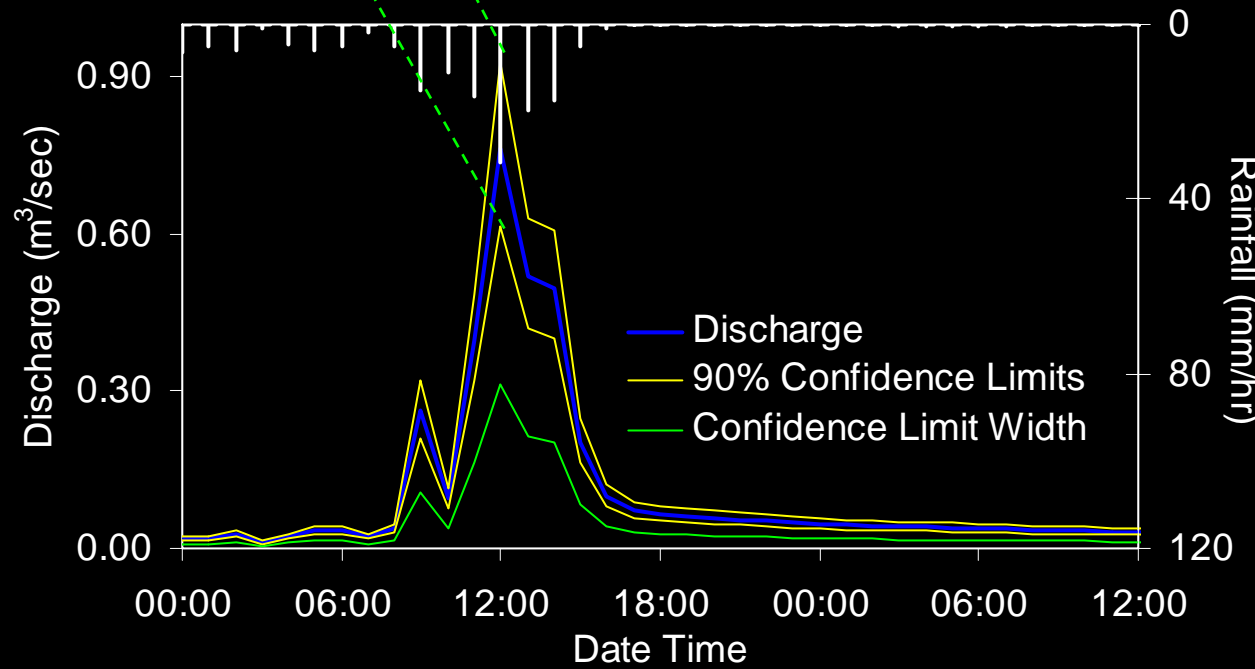
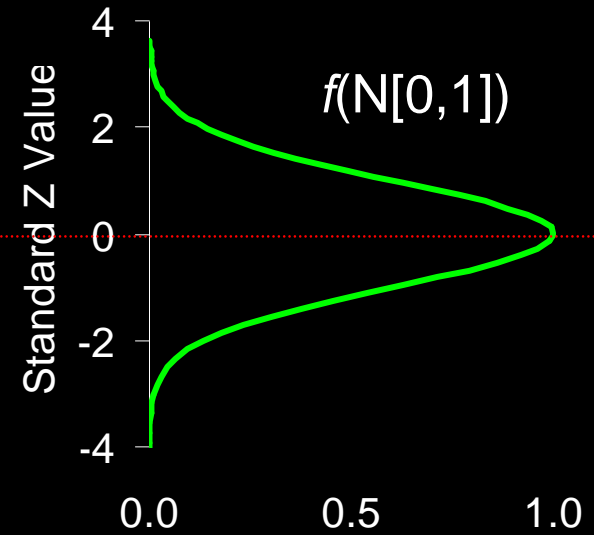
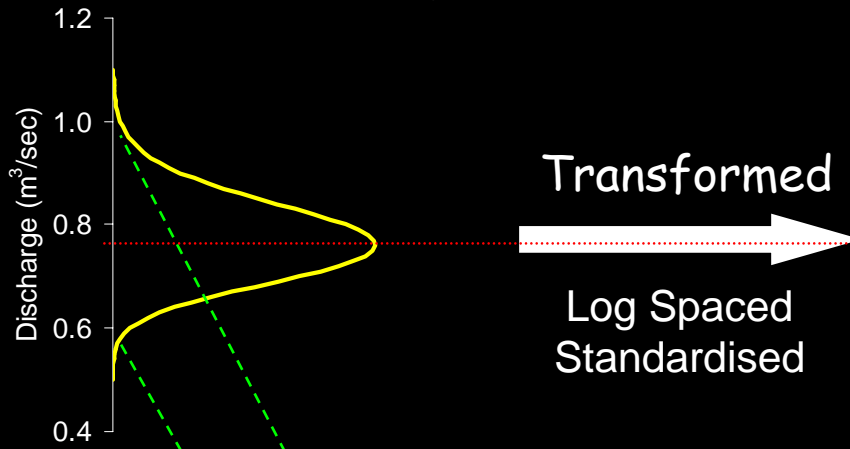
Bukmoongol - Concrete Dam has three sharp-crested rectangle Weirs - 2 only active at peak discharges

Errors were normally distributed in log with this model

But how far can we extrapolate this relationship?, should we?



Discharge Fuzzy Measure



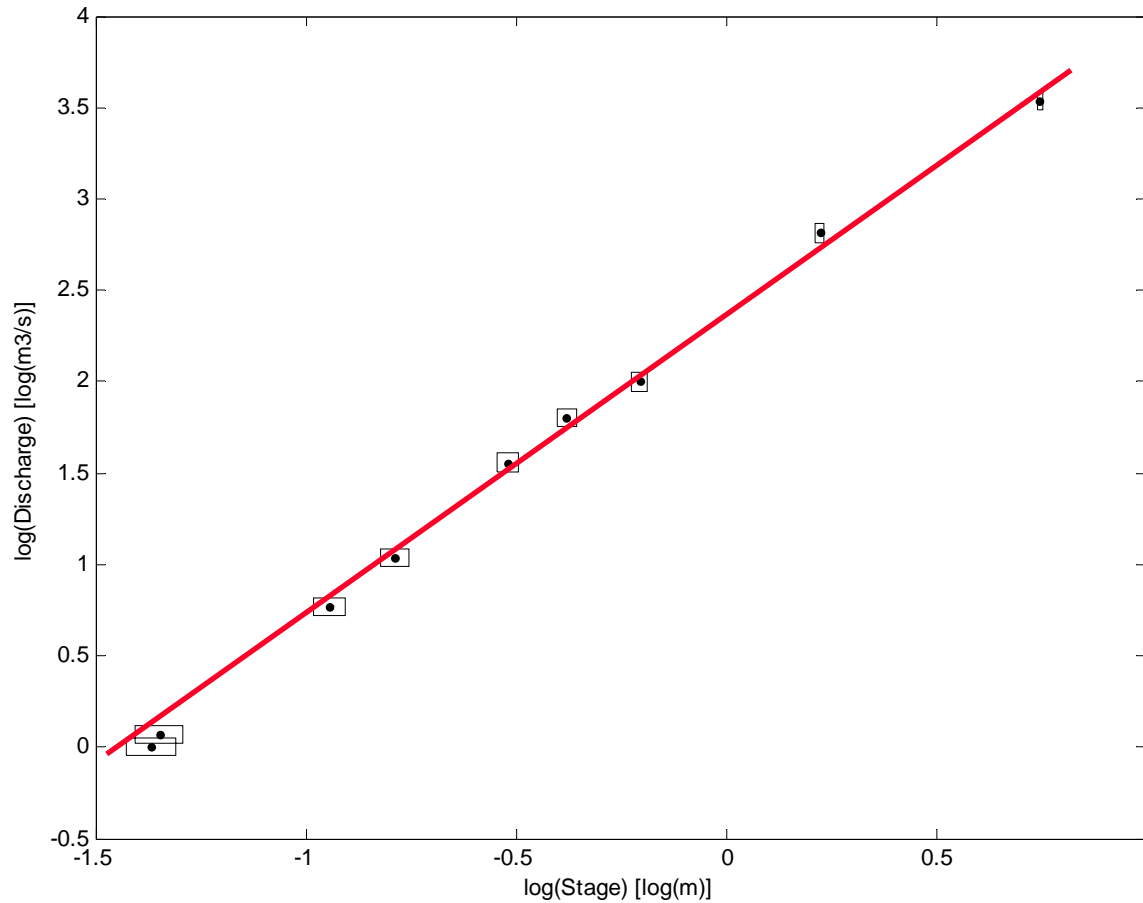
Fuzzy Membership Function at each Time step

An Alternative - set theoretic approach

- Determine your potential errors for both stage and 'field measured' discharge
- Associate an appropriate error window to each measured point
- Fit a series of models - accept all models that pass through every error window, this is your range of stage-discharge relationships

Paul Smith's Current PhD Research Work

Set Theoretic Approach - an example



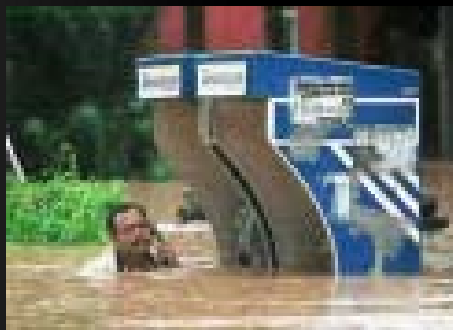
Errors in Different Discharge Structures



Even for well calibrated structures we have potential error structure thresholds out the 'V'
Systematic bias - Gauge Underflow
Regional Subsurface Linkages (input/output)

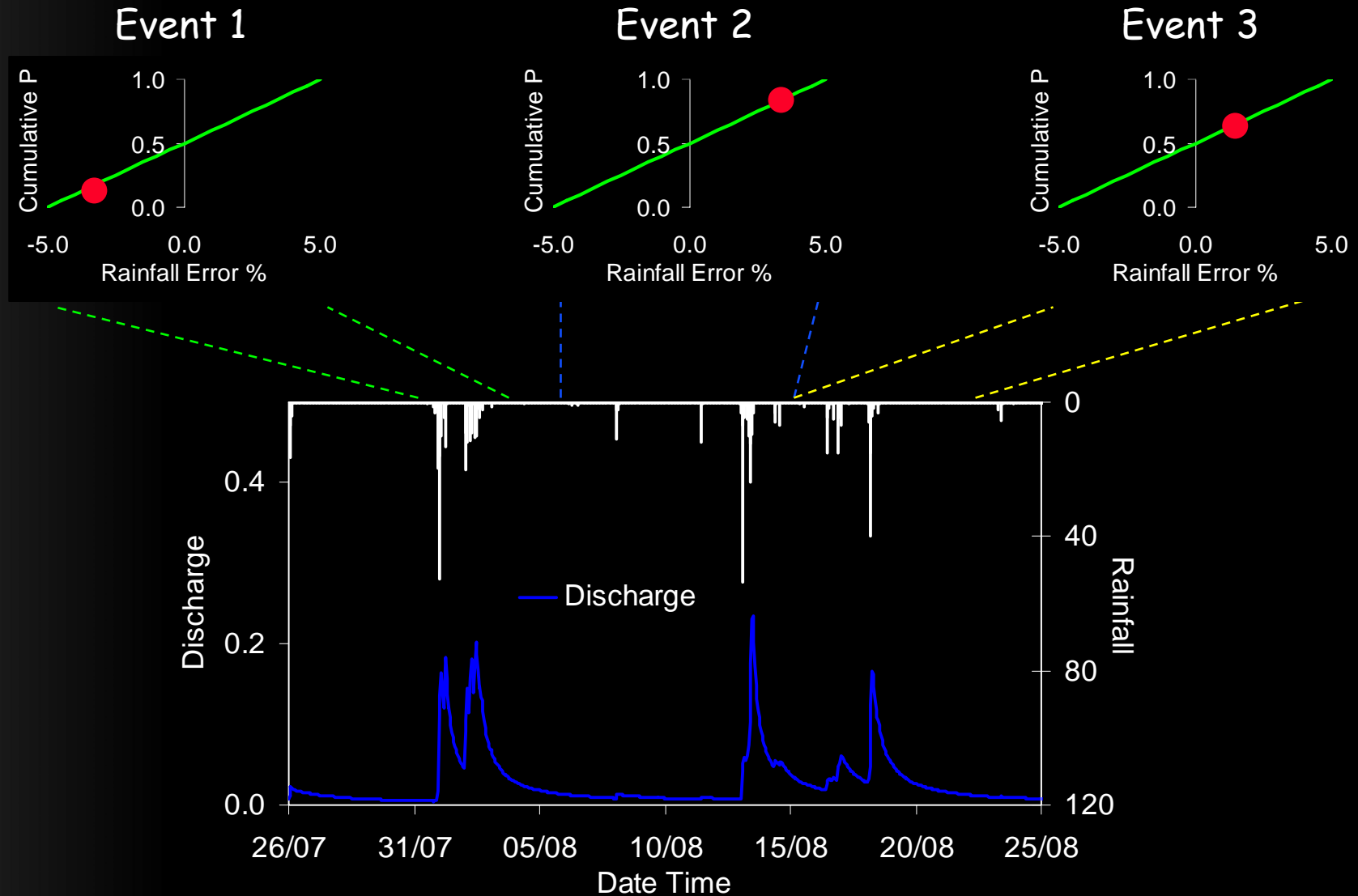


Rated Sections
Point velocity measurements
Interpolation of the velocity field
Changing stage base – how often is this checked?

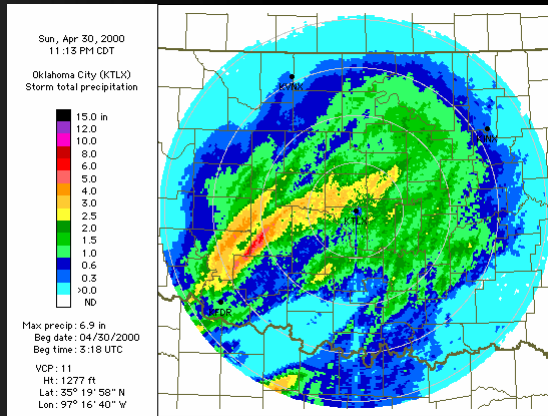


What are the error structures associated with out of bank peak flood events ? – Do we need to phone for help !!
Should we be constraining models to flood peaks?

Rainfall Magnitude Errors (i.e. Dmitri Kavetski)



Rainfall Errors - Larger Scales



When does the spatial variability of rainfall inputs become important as an error term?

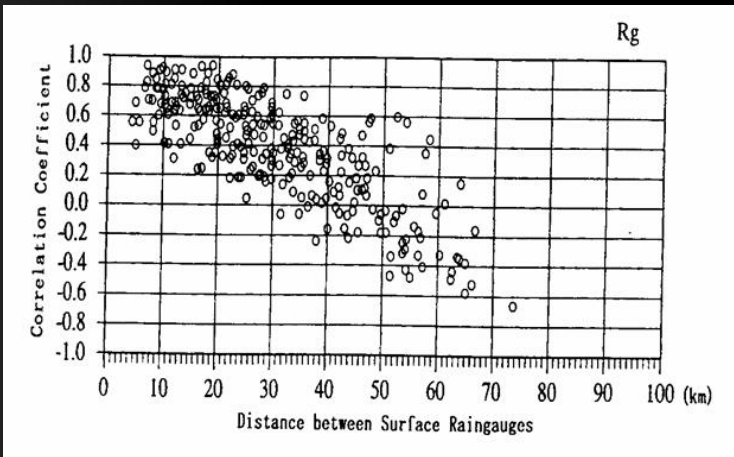
How can this be quantified, won't uncertainty vary due to:

Size of the catchment

Storm Type (frontal system or convective cell)

Geomorphological properties

Similar Issues with Evaporation ?



<http://www.unesco.org.uy/phi/libros/radar>

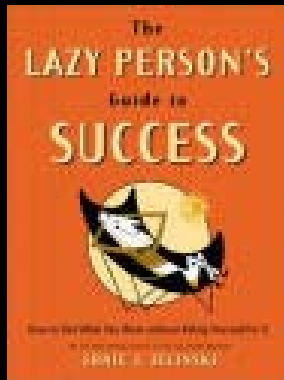
K. Takemura et al

25 raingauges – 10min readings

Possible 'Future' Directions

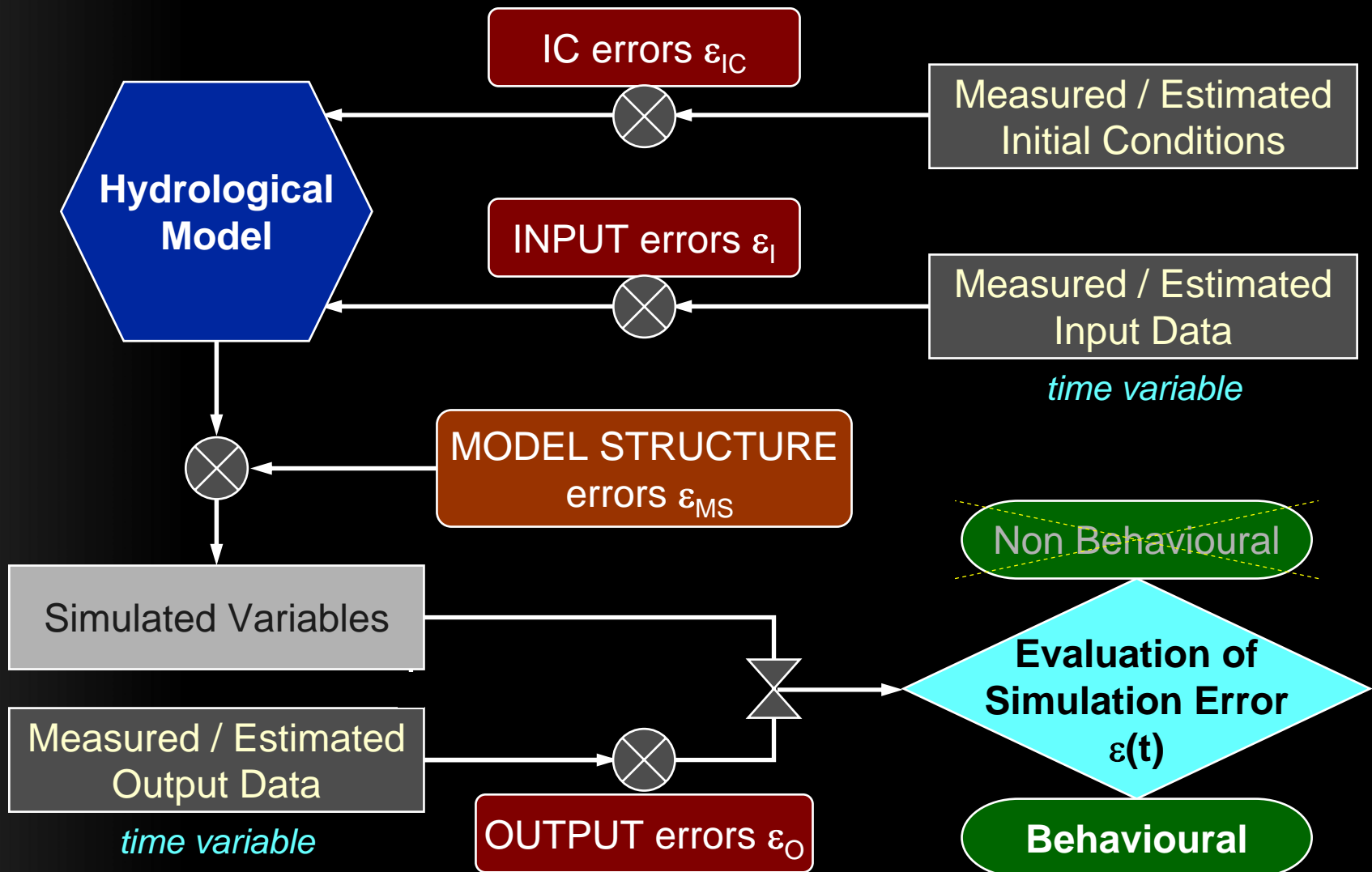
- The use of qualitative data in a soft information form

But how can we limit what an experimentalist believes are unreasonable simulations? - he might be:



- Visualisation of Models using virtual techniques

Extended GLUE - A rejectionist framework



Closing Remarks

- We can always reject all our models if we try hard enough - we learn most when we are wrong
- What is wrong with 'explicit subjectivity' (Bayesian Priors!)? - Is this not a response to imperfect knowledge?
- Are the potential sources of error in our measurements the same as that we assume in our PM?
- Lets make sure we are critical of our results and methods and data

Closing Remarks (continued).....

- Can we even consider at this stage the concept of observational uncertainty benchmarks - how could we move this concept forwards
- Lets think laterally and 'out the box' How much can we use non-quantified PM, soft methods, intuitive visualisation of models to improve our reasoning and rejectionist framework
- This is a PUB-IAHS workshop - Some of our focus should be directed at the goals of PUB

Finally an Announcement!!

Please consider taking part in AGU Fall Meeting
2004 San Francisco

Session Title:

Incorporating Observational Uncertainties
into the Evaluation of Hydrological
Models

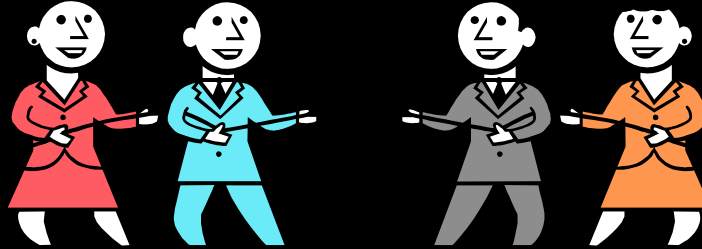
Conveners:

Jim Freer and Jake Peters

j.freer@lancaster.ac.uk

nepeters@usgs.gov

Which side are you on (not mutually exclusive)?



We know our error structures!

A model is never verified

Our model structure is in error

We need some benchmarks

Spatial rainfall fields aren't important

I'm a rejectionist

Equifinality is generic

Uncertainty is everywhere and is uncertain – help!

No we don't !

Of course it is, it works!

Ahhh, it's not that bad

It's impossible to agree them

Well you know they might be

I'm an optimist

But Models are simply ill-conditioned

We can handle it

Enjoy yourselves!!