HUGE Case Study 1 – A GLUE Application

P Smith, J Freer and K J Beven

Lancaster University, Department of Environmental Sciences, IENS, Lancaster, UK (p.j.smith@lancs.ac.uk)

Introduction

This poster outlines the initial GLUE analysis of case 1 of the HUGE experiment where a model and 'real life' data is provided. Initially though some time will be taken to demonstrate the conditions under which GLUE can be thought of as a strict Bayesian methodology. In the analysis of the experiment some investigation is carried out into the effect of assuming conditional independence between the data and the relationship between the number of behavioural simulations and the data characteristics.

The GLUE Posterior: A true distribution?

GLUE proposes for data Γ and parameters $\theta \in \Theta$ the posterior measure of the parameters $PM(\theta|\Gamma)$ is related to the measure of the data $M(\Gamma|\theta)$ by Bayes rule.

$$PM(\theta|\Gamma) = \frac{M(\Gamma|\theta)P(\theta)}{\int_{\Theta} M(\Gamma|\theta)P(\theta)\partial\theta}$$

The measure if defined as having two properties

1. $M(\Gamma|\theta) \geq 0$
2. The measure increases with increasing performance

For $PM(\theta|\Gamma)$ to be a proper distribution the following conditions must be met

1. $PM(A|\Gamma) = \int_A PM(\theta|\Gamma)\partial\theta \geq 0 \quad \forall A \subseteq \Theta$
   - True if integral converges since $M(\Gamma|\theta) \geq 0$ and prior a proper distribution
2. $\int_{\Theta} PM(\theta|\Gamma)\partial\theta = 1$
   - True from the specification if the integral converges to a finite real number
3. $PM(A \cup B|\Gamma) = PM(A|\Gamma) + PM(B|\Gamma) \quad A \subseteq \Theta \; B \subseteq \Theta \; A \cap B = \emptyset$
   - Simple to show by the rules of integration if the integral required for 1 converges

$PM(\theta|\Gamma)$ a formal distribution if

$$\int_A PM(\theta|\Gamma)\partial\theta$$

converges to a real value for all $A \subseteq \Theta$

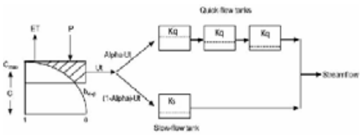
Simple to show if $M(\Gamma|\theta)$ has an upper bounded and the prior defines a bounded space

- e.g. A Nash-Sutcliffe measure and a Uniform prior as commonly used

HUGE Case 1 – Initial Results

Method

•HyMod parameters $\theta = (\text{Cmax, b, Alpha, Ks, Kq})$



•Priors – Independently distributed

- Cmax ~ U[1, 2000]
- b ~ U[0, 4]
- Alpha ~ U[0, 1]
- Ks ~ U[0.0002, 0.3]
- Kq ~ U[0.3, 0.99]

•To methods of incorporating additional information

- 'Independent' - With conditional independence between years

$$PM(\theta|\Gamma,\Gamma_{new}) \propto M(\Gamma_{new}|\theta)M(\Gamma|\theta)P(\theta)$$

- 'Block' - Without conditional independence between years

$$PM(\theta|\Gamma,\Gamma_{new}) \propto M(\Gamma,\Gamma_{new}|\theta)P(\theta)$$

•Two Performance measures related computed flow *c* to observed flow *o*

•Nash-Sutcliffe efficiency $M(\theta|\Gamma) = \max\left(1 - \frac{\sum_{i=1}^{n}(c_i - o_i)^2}{\sum_{i=1}^{n}(o_i - \bar{o})^2}, 0\right)$

•Bias $M(\theta|\Gamma) = \max\left(1 - \left|\frac{\sum_{i=1}^{n} c_i}{\sum_{i=1}^{n} o_i} - 1\right|, 0\right)$

•Random Sampling on prior distribution

•Behavioural model if

•Nash-Sutcliffe efficiency > 0.5

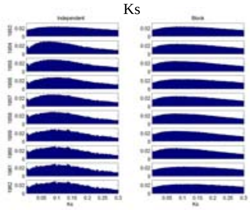
and

•Bias > 0.9

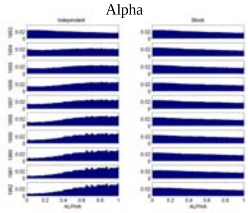
Posterior Probability Density Functions

- Constructed using Nash-Sutcliffe efficiency measure
- Show distribution formed by data up to and including the labelled year

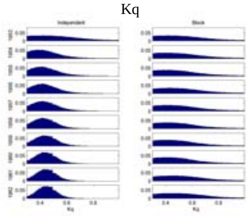
Ks



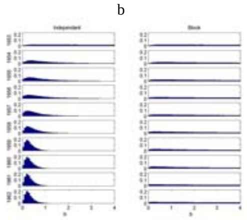
Alpha



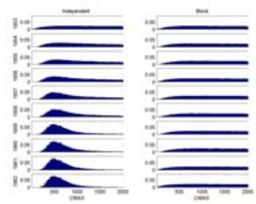
Kq



b



Cmax



- Nash-Sutcliffe efficiency measure perceives more information in the data if water years are treated as independent
- Note the location of the modes of the distributions - Why the discrepancy in Cmax & Alpha?

A Behavioural view point

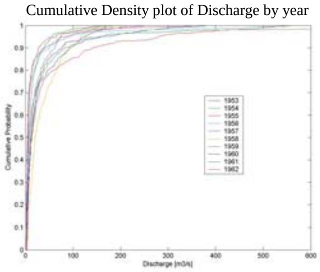
Summary of the number of Behavioural Simulations

Water Year	Behavioural for year (% of simulations)	Behavioural for all water years to date independently (% of simulations)	Behavioural for all data to end of water year as block (% of simulations)
1953	3.26	3.26	3.26
1954	0.49	0.10	1.67
1955	0.12	0	0.76
1956	1.64	0	0.92
1957	0.23	0	0.58
1958	0.79	0	0.57
1959	0.05	0	0.49
1960	1.33	0	0.50
1961	4.41	0	1.23
1962	20.00	0	2.01

Summary of precipitation record

Water Year	Days with Precipitation	Days with Precipitation greater then ... [mm/d]			
		20	50	80	110
1953	143	27	5	0	0
1954	143	15	2	0	0
1955	133	18	3	0	0
1956	145	18	1	1	0
1957	181	15	6	2	0
1958	156	28	2	0	0
1959	156	12	0	0	0
1960	169	28	2	0	0
1961	172	26	6	1	0
1962	158	23	9	2	1

Cumulative Density plot of Discharge by year



- A greater number of behavioural simulations are found for the years where there is a heavier upper tail to the flow distribution.
- The relationship between the number of behaviour models and the rainfall intensity is less clear.
- No Single parameter set sampled is behavioural for all water years (when treated independently)

Conclusions

1. GLUE posterior distributions have all the properties of a probability distribution providing
   - The integral, over part or all of the parameter sample space, of the GLUE measure multiplied by the prior can be shown to converge to a real number
   - The measure is always greater or equal to zero
   - The measure increases in value as the model performance improves
2. The assumption of independence between periods of the data record allows common GLUE measures to perceive more information in the record.
3. The formulation of a well defined posterior distribution does not imply the model fits all sections of the data adequately.

Acknowledgements

Paul Smith is supported by the NERC long term research programme on Uncertainty in Environmental Modelling (NER/L/2001/00658).