Constraining Dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures.

Freer, J.E. [1], McMillan, H.[2], McDonnell, J.J. [3], and Beven, K.J. [1]

[1] Lancaster University, Department of Environmental Sciences, IENS, Lancaster, Lancaster, LA1 4YQ, United Kingdom

[2] Department of Geography, University of Cambridge, Downing Place, Cambridge, CB2 3EN, UK

[3] Department of Forest Engineering, 015 Peavy Hall, Oregon State University, Corvallis, Oregon 97331-5706, USA

Manuscript for 'Catchment modelling: towards an improved representation of the hydrological processes in real-world model applications', Uhlenbrook, S and Montanari, A. (eds) – **Journal of Hydrology Special Issue** (2002 EGS session) – Post Review Draft

Keywords: Dynamic TOPMODEL, GLUE, Water Table Uncertainty, Parameter Constraining, Fuzzy Rules, multicriteria calibration

DRAFT COPY

*Corresponding Author:
Jim Freer
Department of Environmental Sciences
Lancaster University
Bailrigg
Lancaster
LA1 4YQ
UK

e-mail: j.freer@lancaster.ac.uk
Tel: +44 (0)1524 593563
Fax: +44 (0)1524 593985

Abstract

Dynamic TOPMODEL is applied to the Maimai M8 catchment (3.8 ha), New Zealand using rainfall-runoff and water table information in model calibration. Different parametric representations of hillslope and valley bottom landscape units were used to improve the spatial representation of the model structure. The continuous time series water table information is obtained from tensiometric observations from both near stream (*NS*) and hillslope (*P5*) locations having different responses to rainfall events. For each location, and within an area equivalent to an effective model gridscale (25m^2), a number of tensiometer readings at different depths were available (11 for the *NS* site and 9 for the *P5* site). Using this information a distribution of water table elevations for each time step at each location was calculated. The distribution of water table elevations was used to derive fuzzy estimates of the water table depth for the whole time series that includes the temporal variability of the uncertainty in the observations. These data were used to constrain the spatial representation of the model having previously conditioned the model using the rainfall-runoff data. Model conditioning was assessed using the Generalised Likelihood Uncertainty Estimation procedure.

Results show that many combinations of parameter values for the two landscape units were able to simulate the rainfall-runoff data. Further constraining of the model responses using the fuzzy water table elevations at both locations considerably reduced the number of behavioural parameter sets. An evaluation of the distributions of behavioural parameter sets showed that improvements to the model structure for the two landscape units were required, especially for simulations of the response at the hillslope location.

1 INTRODUCTION

A pragmatic and realistic approach to environmental modelling should recognise that all model structures, regardless of their complexity, are to some extent in error (Beven, 1989; Grayson et al., 1992; Beven, 2002). This can be attributed to two main factors: (1) that our perceptual model is based on imperfect knowledge, and (2) that the formulation of a model necessitates the use of highly simplified mathematical constructs that cannot represent all the details of the many interacting processes within a natural system. Furthermore increasing model complexity, or explanatory depth, increases the possibility that the amount and type of observational data at hand will be inadequate to fully assess model performance. Such data limitations would be especially apparent for semi-distributed or distributed model constructs where the individual spatial components are rarely tested locally.

Model evaluation is made at the catchment scale using stream discharge data. The use of discharge data alone has been shown to have weaknesses in the identification of model structures and parameters (e.g. Freer et al., 1996). This understanding has led to discussions of model identifiability (Sorooshian and Gupta, 1985; Beck and Halfon, 1991) and of the equifinality of model structures and parameters (Beven, 1996; Beven and Freer, 2001b). Increasingly recent papers have shown that being more thoughtful about the specification of objective functions or performance

DRAFT COPY

measures (PM's) and/or the use of multiple objectives ensures that best use is made of limited data in model calibration/evaluation (Gupta et al., 1998; Thiemann et al., 2001; Wagener et al., 2001; Seibert and McDonnell, 2002; Freer et al., 2003). One way of potentially improving the assessment of models has been to introduce multi-response data that describe different characteristics of the system. These measures may improve the identification of model structures and associated parameters without increasing the complexity of the model (Troch et al., 1993). There have now been a number of studies where this has been explored (Kuczera, 1983; de Grosbois et al., 1988; Bloschl et al., 1992; Grayson et al., 1992; Koide and Wheater, 1992; Lamb et al., 1997; Mroczkowski et al., 1997; Franks et al., 1998; Kuczera and Mroczkowski, 1998; Guntner et al., 1999; Motovilov et al., 1999; Vertessy and Elsenbeer, 1999; Anderton et al., 2002a; Aronica et al., 2002; Blazkova et al., 2002; Uhlenbrook and Leibundgut, 2002).

The introduction of data other than discharge into the calibration process has not always produced satisfactory results. *Stephenson and Freeze* (1974) and *Koide and Wheater* (1992), in similar studies using detailed 2D distributed hillslope models calibrated from comprehensively sampled tensiometer and piezometer data, both noted difficulties in the calibration of their models due to numerous data and model simplification/initialisation factors. *Grayson et al.* (1992) found that the "*measurement of catchment response in sufficient detail*" (i.e. limitations imposed by data sparseness) was a limiting factor in the spatial validation of the THALES model. *Hooper et al.* (1988) found that using a combination of rainfall-runoff and geochemical data to identify a model with only six parameters called into question "*the structural validity of more highly parameterised rainfall-runoff models used in water quality prediction*". More recently *Anderton et al.* (2002b) found difficulties in using limited soil moisture and phreatic surface information in the validation of the SHETRAN model due to both the sparseness of the data and the 'mismatch' of the measurement scale to the model gridscale (see detailed discussions on using/interpreting spatial patterns for hydrological modelling in Grayson and Bloschl, 2000).

While the introduction of new data sources (beyond that of discharge) into the assessment of models can increase model identifiability, a number of issues may bias the conclusions:

- The data are uncertain (Sherlock et al., 2000). That is, for many data types there may be an inevitable degradation of quality and/or of the ability of the data to be representative of the system of interest.
- The data may not be appropriate. That is, the phenomenon being represented by the data may not be commensurate with the model formulation, therefore direct comparisons through the specification of simple objective functions may not be realistic
- The observations may be at the wrong scale. That is, observations may be at a different scale to the model scale. For scale discrepancies there might be a range of observed behaviour that is both large and inconsistent over time periods for the effective model gridscale.

DRAFT COPY

As a result of these points, different performance measures may be required to match model assessment with the appropriate level of data quality, representativeness and scale. The error associated with models and data and the limitations of current data technologies directs the practitioner towards an assessment of models that is inherently probabilistic (see for example the use of uncertain saturated area observations in Franks et al., 1998). A probabilistic assessment allows for multiple parameterisations and/or model structures. Nevertheless, rejection is often difficult because of the limitations in the available data or because of our 'imperfect knowledge' of the system under study.

A number of calibration methodologies for this type of approach have been developed, each having to a greater or lesser extent assumptions regarding the nature of the error structure, the sources of error and the complexities of the multidimensional parameter space response surface. This paper introduces multi-response data (discharge and tensiometric information) into the assessment of a hydrological model (Dynamic TOPMODEL) within the uncertainty analysis framework GLUE (Generalised Likelihood Uncertainty Estimation). Both stream discharge and multiple tensiometric readings are used for two topographically distinct sites at the Maimai catchment, New Zealand. The variability in the multiple readings at each site are characterised as a time-variable fuzzy objective function in a way that is more appropriate to the effective model gridscale and the uncertainty within multiple observations. To reflect the differences in these two topographically different sites Dynamic TOPMODEL is configured for two Landscape Units (LU's) one being a Hillslope (HS_{LU}) and the other a Valley Bottom (VB_{LU}), each having independently sampled parameter values. The parameter interactions between the two LU's are assessed and conclusions are drawn as to the usefulness of uncertain (fuzzy) gridscale information in constraining model parameters. Specifically, we address the following questions within the context of this general aim of simulating the discharge and water table (∇_{wt}) responses:

- Can we meet discharge and/or tensiometer criteria for more than one source of information?
- How can fuzzy rules be applied to imperfect and imprecise knowledge when the error structures are time variant?
- How can we constrain model responses and the efficiency of sampling?
- How can we improve the Dynamic TOPMODEL structure and parameter representation?

A recent paper by Seibert and McDonnell (2002)

2 THE STUDY SITE

The Maimai M8 catchment is located in the Tawhai State Forest, North Westland, South Island, New Zealand. It is one of eight small adjoining catchments that have been studied since 1974 as part of a land use change study. The layout of the catchment is shown in Figure 1a-d.

DRAFT COPY

Mean annual gross rainfall in this area is approx. 2600 mm, producing some 1550 mm of runoff from 1950 mm of net rainfall (Rowe, 1979), with little seasonal variation. The Maimai catchments are highly responsive to rainfall, *Pearce and McKercher* (1979) reported that quickflow represents 65% of annual runoff (39% of total rainfall), as defined by *Hewlett and Hibbert's* (1967) separation method. *Sklash* (1990) commented that "*The Maimai catchments are among the most hydrologically responsive forested headwater catchments documented*"

The surficial geology of Maimai catchment is firmly compacted, moderately weathered, early Pleistocene conglomerate, which is known as the Old Man Gravels and has been described as "*effectively impermeable*" by *Mosley* (1979). The relief of the catchment is in the order of 100-150m, with steep (average 34°), short (less than 30m) slopes (see Figure 1a). Soil development has weathered the conglomerate to form (as a broad classification), Blackball Hill soils (Mew et al., 1975). These soils are spatially quite variable in both depth (0.2 - 1.8m) and character, having a thick well developed upper humic horizon (mean 170 mm, *Webster* (1977)). The upper mineral soil has been found to have an average saturated hydraulic conductivity of 250mm hr^{-1} (Webster, 1977). However, using a Guelph permeameter, *McDonnell* (1989) found this value to be highly variable, ranging from <5mm hr^{-1} in poorly drained hollows to the value reported by *Webster* in well drained nose slopes. The average infiltration capacity of the soil surface has been reported by *Webster* (1977) as 6100 mm hr^{-1}.

The vegetation of the catchment is classified as a mixed evergreen forest, the main cover being dominated by southern beech, podocarps and broadleaf hardwoods. The forest is multi-storied, the understorey consists of dense tree fern and shrubs and has a ground cover of ferns and herbs *Pearce et al.* (1986). A more detailed physical description of the Maimai M8 catchment can be found in *Rowe et al.* (1994) and *McGlynn et al.* (2002).

2.1 The tensiometric study sites

The layout of the Maimai M8 catchment is shown in Figure 1, and has been extensively documented by *Pearce et al.* (1986). The intensive monitoring of the 0.3 ha subcatchment and the Near Stream (*NS)* site was undertaken over a number of storm events during September to December 1987 (McDonnell, 1990). The data collected included tensiometer, trough flow, and chemical tracer samples, as well as hydrometric data based on a 10min. time step. Two tensiometer sites were used from this intensive study, these being the *NS* (Figure 1c) and *P5* (*P5* - Figure 1d) sites, both of which have been reported in *McDonnell* (1990) with regard to 3D matric potential (ϕ) responses. Tensiometers were situated away from cracks and voids to ensure they characterised only the changes in the soil matrix (McDonnell, 1990). The topographic position of the two sites differs considerably (see Figure 1a), with the *NS* site having a close proximity to the stream channel (< 4m) and the *P5* site on a steeper upslope section (some 40m from the stream channel).

DRAFT COPY

Consequently the data provide a good test of the possible variation in water table responses in two topographically distinct areas of the catchment. The variations in soil properties between these two sites are given in Table 1 and will be referred to in later sections.

The *P5* site consisted of an electronically multiplexed and logged array of 32 tensiometers (arranged within a grid 6m by 1m), whereas the *NS* site had 24 tensiometers (4m by 0.5m) that were linked via a fluid scanning switch to a single pressure transducer. The *P5* site had continuous logged data, which were recorded at the same time for all tensiometers. Due to the fluid scanning switch at the *NS* site, a single reading was taken in rotation at a maximum resolution of one minute increments, although this increment sometimes increased to 5 min. for short (recession) periods. Details of the tensiometer design and performance are given in *McDonnell* (1993). The tensiometers ranged in their depth below the soil surface from 15-124 cms for the *P5* site and 10-78 cms for the *NS* site.

Readings from the two sites were not available for the complete discharge period (see Figure 2). The data collected for the *P5* site were available from 2/10 at 19:40 to 17/12 at 14:30, and for the *NS* site from 7/10 09:20 to 18/11 at 00:00. Within these limits the data had a considerable number of short and long ‘breaks’ (equipment failure etc.). Most of the longer breaks occurred during recession periods, however some of the smaller disruptions occurred during events, or meant that some of the smaller storm events were not available.

McDonnell (1990) detailed results from the *NS* site for tensiometers *T1-9* and from the *P5* site for tensiometers *T1-16* and *T23-25* for the October 29th storm event. There were considerable difficulties in creating a coherent data set for an extended period, these mainly included periods of failed tensiometers. The intention was to incorporate as many tensiometer readings as possible into the ∇_{wt} series, so that a proper account was taken of the variability of the tensiometer response at a scale that was consistent with the model gridscale (see discussion by Bathurst and O'Connell, 1992). Due to problems of equipment failure and extreme electrical noise not all the tensiometers at the two sites were used. Furthermore, shallow tensiometers at the *P5* site, were sensitive to the wetting front propagation down through the soil profile during precipitation events, these sensitivities would not be directly related to the ∇_{wt} formation from the soil-bedrock interface and were also excluded. This resulted in 9 tensiometers at the *P5* site and 11 at the *NS* site that could be used in the following methods. These tensiometers covered areas of 4.5m*lm and 4m*0.5m respectively and are shown as filled circles in Figure 1c and 1d along with their cup depth below the soil surface.

DRAFT COPY

3 METHODS

3.1 Calculation of water table responses at both tensiometer sites

Tensiometer readings have positive matric potential when the porous cup is below the water table surface, negative matric potential when the tensiometer cup is above the water table surface. Variations of matric potential at the *NS* site for all tensiometer readings used in this study are shown in Figure 2 for the whole of the study period. Figure 2 shows that positive (+ve) matric potentials are observed for much of the study period. For the *P5* site +ve potentials were more transient, having steeper recessions (which are reflected in the ∇_{wt} variations shown for both sites in Figure 6).

The relationships between -ve matric potentials and soil water content can be complex and have been well documented (Kosugi and Inoue, 2002; Torres and Alexander, 2002). Soil water retention curves have been determined for many different soil types and generally show hysteresis behaviour between the wetting and drying curves. *Burt & Butcher* (1985; 1986) developed a simple methodology that used average gradient of soil water potentials (from a number of tensiometers at different depths) to predict the depth of the ∇_{wt} at the soil-bedrock interface. Using field calibrations obtained from *Butcher* (1985)] they suggested that the average gradient (at their experimental site at Slapton Wood, UK) under -ve tensions was 1.2 cm soil water potential per cm soil depth (a linear relationship). We used the Butcher [1985] method to develop a relationship between -ve soil water tensions and apparent depth to the ∇_{wt} at Maimai. ∇_{wt} is directly inferred during periods where the deepest tensiometer is below the ∇_{wt} surface (in +ve tension). Matric potentials were linearly adjusted to a ∇_{wt} surface, for both +ve and -ve readings, by correcting readings to the ground surface datum by;

$$\nabla_{wt(t)} = T_{z(t)} - T_{\phi(t)} \qquad [1]$$

where T_z is the depth of the tensiometer (m) and T_ϕ is the matric potential reading of the tensiometer [m H_20] at time *t*. It should be noted that equation [1] is only valid if it is assumed that vertical soil water fluxes are negligible, suggesting that the soil is in equilibrium and total potentials[1] are constant throughout the soil profile. Figure 3a,b show for the recession period of the October 29th storm event the relationship between –ve matric potentials and height above the water table for all tensiometers at the *NS* (Nest 4) and *P5* (Nest 1) sites (see Figure 1) during periods where the deepest tensiometer is in +ve tension (i.e. T10 and T4 respectively). A recession period is chosen to avoid wetting fronts affecting tensiometer readings during precipitation events. Significantly fewer

[1] Total potential, the potential energy of the soil water, is the sum of gravity potential (the product of height above some datum times the density of water times gravitational acceleration) and capillary potential (the amount of capillary rise)

DRAFT COPY

points were available for the *P5* site because the peak response was much more transient so that +ve tensions where not maintained at T4. For the *NS* site a linear relationship provides a good correlation between –ve tension and height above ∇_{wt} (R^2 = 0.94), having a similar slope gradient to that found by *Butcher* (1985), namely 1.16 cm soil water potential per cm soil depth. For the *P5* site the linear relationship does not seem to hold as well (R^2=0.91), the slope gradient is much higher (2.4 cms per cm) and the relationship for the site appears to be only quasi-linear having a lower gradient for smaller –ve tensions. For these hillslope soils a more appropriate relationship is found using a second order polynomial (a = 1.192, b = 0.048, R^2 = 0.97). Such non-linearity may be the result of the -ve matric potential gradients being non-uniform in the unsaturated zone for these soil types (note that the initial slope gradient is again similar to the results of *Butcher* (1985)). Confidence in this relationship increased further after calculating the predicted ∇_{wt} depth for all tensiometers at *P5* (Nest 1) over a much longer recession period (i.e. for a higher –ve potential range) the results of which are shown in Figure 3c. The variability in the ∇_{wt} predictions over the recession period is low but tends to increase with increasing ∇_{wt} depth (to a maximum in this case of 20 cm). However such variability within local tensiometer nests is much less than the ∇_{wt} predictions between the local tensiometers nests (i.e. the variability at the effective model gridscale) for both the *NS* and *P5* sites (see below and section 3.2), including periods where +ve tensions were observed at multiple sites.

The methods described above allowed +ve tension (using eqn. [1]) and –ve tension readings using the linear and polynomial relationships for the *NS* and *P5* sites respectively to be used to predict the ∇_{wt} variations for the whole study period. To summarise the variability of ∇_{wt} predictions throughout the study period Figure 4 shows the variability in the range of ∇_{wt} observations for different depths (classified by the observed mean depth for each timestep), for the minimum to maximum and 25^{th} to 75^{th} inter-quartile range for both the *NS* and *P5* sites. Figure 4 also gives the percentage of time that each depth occurred during the series, this clearly showing the more transient nature of shallow ∇_{wt} observations at the *P5* site with higher frequencies of occurrence being skewed towards deeper ∇_{wt} levels. These plots show that the mean and inter-quartile ranges of ∇_{wt} observations vary with depth, increasing with increasing depth for the *NS* responses and with depths associated with more rapidly changing ∇_{wt} fluctuations during events for the *P5* site (see Figure 6).

3.2 A fuzzy measure of water table responses at the model gridscale

We have identified that the tensiometer responses used in this study are not themselves wholly accurate predictions of the ∇_{wt} as seen in the regression relationships presented in Figure 3. Furthermore significant local variations of the ∇_{wt} are observed at scale that is commensurate with

DRAFT COPY

the model gridscale and that the magnitude and distribution of these variations change with time. What we require therefore is a performance measure that for each timestep and at the model gridscale reflects the noise in the data, the variability in the timings of the ∇_{wt}, and the uncertainty in the information expressed within the regression relationships between –ve tensions and height above the ∇_{wt}.surface. Therefore so as not to unduly bias the assessment of model performance a fuzzy additive definition of the performance measure was used, having the following form of membership function (see Ross, 1995):

$$L\left[M\left(\Theta \middle| Y_T, W_T\right)\right] = \sum_{t=1}^{n} \begin{cases} 0 & z_t \geq {}_{\max}w_t \\ \dfrac{z_t - {}_{75}w_t}{{}_{\max}w_t - {}_{75}w_t} & {}_{75}w_t \leq z_t \leq {}_{\max}w_t \\ 1 \quad ...if... & {}_{25}w_t \leq z_t \leq {}_{75}w_t \\ \dfrac{z_t - {}_{\min}w_t}{{}_{25}w_t - {}_{\min}w_t} & {}_{\min}w_t \leq z_t \leq {}_{25}w_t \\ 0 & z_t \leq {}_{\min}w_t \end{cases} \qquad [2]$$

where $M\left(\Theta \middle| Y_T, W_T\right)$ indicates the i^{th} model simulation run, conditioned on input data Y_T and observations W_T. For each timestep *t* the simulated local ∇_{wt} depth z_t is compared to the distribution of ∇_{wt} observations defined by the *core* (the 25th $\left({}_{25}w_t\right)$ and 75th $\left({}_{75}w_t\right)$ quartiles) and the *support* (the min $\left({}_{\min}w_t\right)$ and max $\left({}_{\max}w_t\right)$ values) of the fuzzy membership function. Essentially equation [2] defines a trapezoidal fuzzy membership set (see Figure 5) for the observed ∇_{wt} responses, the characteristics of which depend on the distribution of the local ∇_{wt} depth at each timestep. The *core* of the set being the range of depths where we believe that the simulated z_t would be a complete and full member of the observations and the *support* being the range of depths either side of the *core* where we have a nonzero membership (i.e. that we become less sure that the simulation is a member the closer this value approaches the *support* limits).

The assignment process that defines the form of the membership function can involve many methods, ranging from intuition (i.e. what is the range of saturated area in this catchment that we believe is possible?) to the use of more formal methods such as inductive reasoning and the use of fuzzy statistics (see Ross, 1995). Membership functions may or may not have a core range and or have much more complex forms (e.g. multi modal, subnormal and nonconvex) depending on the observations that are available. The assignment procedure used here would formally be known as an inference procedure (i.e deductive reasoning from some knowledge of the system). In this case using the 5th and 95th percentiles as the support limits rather than the minimum and maximum values was rejected as there was felt to be no justification for totally rejecting the possibility that the outer ∇_{wt} readings were correct. The trapezoidal measure was chosen as this represented a

DRAFT COPY

compromise between the difficulties of defining what was the 'best' ∇_{wt} observation at each timestep (a membership function without a core range) and the advantage of favouring mid-range ∇_{wt} values that would not be the case using a crisp set (i.e one without *boundaries* – see Figure 5). Figure 6 shows the resultant support limits for the ∇_{wt} membership function (the core is not shown for clarity) and the individual ∇_{wt} observations at both tensiometer sites for all timesteps where data are available. These results clearly show that the amount of uncertainty in the ∇_{wt} surface varies considerably during the study period, that this variation is significant for similar ∇_{wt} levels at different time periods (especially for *P5*) and that each tensiometer study site has different characteristics of variability.

3.3 The Hydrological Model – Dynamic TOPMODEL

The new Dynamic TOPMODEL version is briefly described below to allow the reader to understand the spatial context of the model structure and associated parameters applied to Maimai catchment. For a more detailed treatment of the model application and model theory the reader is referred to the paper by *Peters et al. (2003)* and to the original paper on Dynamic TOPMODEL by *Beven and Freer* (2001a) .

Dynamic TOPMODEL (Beven and Freer, 2001a) is a new version of TOPMODEL that relaxes some of the assumptions of the original form (Beven and Kirkby, 1979) following critiques of TOPMODEL by *Barling et al.* (1994), *Beven* (1997), and *Wigmosta and Lettenmaier* (1999). This new formulation allows for local accounting of hydrological fluxes and storages, relaxing the quasi steady state assumption of a water table parallel to the local surface slope expressed through the derivation of the *ln(a/tanβ)* index of *Kirkby* (1975). Therefore the dynamics of the subsurface saturated zone during wetting and drying event periods (expanding and contracting) can be simulated. Previous field evidence had suggested that the original assumption of an effective upslope contributing area extending to the catchment divide during wetting-up periods was thought to be unrealistic (Barling et al., 1994; Guntner et al., 1999). *Beven* (1997) suggested that the overestimation of the accumulated upslope area '*a*' was being compensated in the results by generally high transmissivity values, this being seen in original TOPMODEL applications. Dynamically varying upslope contributing areas '*a*' are conceptualized in a simple form with the addition of the parameter $S_{\max}$ (the maximum effective deficit of subsurface saturated zone), which in a simple form, as in this example, restricts down slope flow only to areas where the local deficit $s_i \geq S_{\max}$ Areas with shallow regolith depths (small $S_{\max}$) and areas near the catchment divide, would be more likely to 'disconnect' upslope areas during recession periods. *Beven and Freer* (2001a) found the best behavioural simulations of discharge at Slapton Wood catchment in the UK occurred with a dynamically varying upslope contributing area (i.e. when $S_{\max}$ became active).

DRAFT COPY

However, good / acceptable (behavioural) simulations were also obtained for simulations where no change in the upslope contributing areas was predicted.

Without any further information on the spatial variability of hydrological processes a 2D classifier matrix [a, T_0 tanβ] is used as measure of hydrological similarity, where T_0 is the transmissivity measured in the direction of downslope flow, tanβ the local slope angle, and a as before. Using a 2D classification matrix that includes '*a'* ensures the resulting Hydrologically Similar Units (HSU's – now used as the local hydrological accounting units), maintain a general continuity of flow in a downslope direction but whose fluxes are dynamically variable. Topographic analysis allows the calculation of a transition probability matrix for a water drop to move from one class to another (an extension of the multi-flow algorithms of *Quinn et al.* (1991)). In this way, the water balance for each HSU can be solved. Transfers between HSUs are calculated using a kinematic wave approximation, where both the upslope (for inputs) and local (for outputs) storages are required. Flux volumes are a function of the storages and the $T_o tan\beta$ values in each case (Beven and Freer, 2001a). As with the original TOPMODEL, an exponential transmissivity profile and a constant effective storage coefficient are assumed. Experience in a number of catchments in different countries suggests that the transition from hillslope to valley bottom landscape units is often quite marked. These units would be expected to have distinct soil characteristics but the simplifying assumption of homogeneous effective soil parameter values within each of the units is a necessary approximation and will limit the accuracy with which the detailed spatial patterns of response can be predicted. Data from Maimai, Panola and many other sites shows that this transfer from hillslope to riparian boxes can be very threshold-like and non-linear.

The model also allows for the spatial organisation and connectivity of different HSU's, each having potentially different functional forms of hydrological (and/or other) responses. Including different functional forms requires some knowledge of the spatial variability of hydrological response, which may often be limited (especially within the subsurface) at a scale pertinent to catchment scale responses. *Peters et al.* (2001) conceptualised Dynamic TOPMODEL to include the spatial variability of distinct LU's, primarily though the distribution of regolith depths. These LU's were assumed to have different hydrological / physical characteristics that were controlling hydrological response and therefore required the definition of different parameter ranges and/or model structure. In this study the catchment has been separated into 2 LU's, a HS_{LU} and a VB_{LU} component (see Figure 1), the general break in slope between the VB_{LU} and HS_{LU} areas defined the spatial extent of these units. Two LU's were identified primarily to study the interactions between parameter sets when simulating the discharge and the *NS* and *P5* ∇_{wt} responses. The spatial variability of hydrological response due to additional topographic features at Maimai (i.e. ridges and hollows), is characterised explicitly in the model using the classifer matrix of hydrological similarity (described above).

DRAFT COPY

A previous application using the original form of TOPMODEL by *Freer* (1998) showed that no parameter sets (using homogeneous parameter values) could be identified that satisfactorily simulated the ∇_{wt} responses at both sites, although when treated individually behavioural parameter sets could be identified for each site. The functional differences in the LU's are here expressed by the differences in the parameter ranges for each unit (see Table 2), i.e., the same functional <u>form</u> is retained for each LU, including the assumption of an exponential decline in transmissivity with depth. The 2-LU model has 14 parameters (*CHV* and SR_{init} are sampled once for each simulation, then assumed constant for all LU's). The catchment was divided using digital terrain analyses (5 m^2 DEM) into 41 Hydrologically Similar Units (HSUs) for the model simulations.

3.4 GLUE Simulations of Discharge and Water Table Information

Freer (1998) used the original form of TOPMODEL applied to Maimai for both discharge and ∇_{wt} simulations (at the *NS* and *P5* sites) as reported here but using a weighted Nash efficiency measure (weights calculated from the uncertainty in the ∇_{wt} depth) within the GLUE procedure. As noted above no homogeneous catchment parameter sets could be found that simulated the ∇_{wt} responses at both tensiometer sites. Recently *Beven and Freer* (2001b), also using the original form of TOPMODEL, analysed multiple years of discharge data at Maimai and found that once uniform prior distributions had been constrained using 1 year of data, subsequent years did little to constrain parameter estimates further. This paper extends these analyses by assessing multi-objective variations in model performance for dynamic TOPMODEL within the GLUE methodology using fuzzy performance measures. The multi-objective data in this case are the discharge and the ∇_{wt} information at both tensiometer sites. For the initial simulation runs all parameters listed in Table 2 were randomly assigned a value appropriate to the ranges specified for each LU (where appropriate). For initial simulations a uniform sampling strategy of the parameter ranges was deployed to express the lack of knowledge of the expected distribution and covariance of the parameter values. The model streamflow and ∇_{wt} predictions for the study period were compared to the observed data using one of the 3 Performance Measures and rejection criteria defined in Table 3. For each tensiometer site the midpoint position of the tensiometer nests was used to georeference this data with the DTM coverage. The time series of the simulated ∇_{wt} predictions for both corresponding HSU increments and the catchment outfall discharge predictions were retained for post analysis along with the parameter values for the model run. Differences among behavioural parameter sets were evaluated for each performance measure.

The GLUE simulations were conducted on a parallel LINUX PC system at Lancaster University. The system consists of 47 nodes having a combination of AMD 800MHz, 1500MHz and 2600MHz processors. The topology used was a simple master slave combination via 100Mbps Ethernet using

DRAFT COPY

basic batch processing scripts for job submissions (one job per slave unit). The initial 5,600,000 simulations took 2 days to complete (on 6 fast nodes) for the 1987 study period.

4 RESULTS AND DISCUSSION

In total 6.8million runs of the model were generated. The initial 5.6million runs described above are known as run_1. To see how much the efficiency of sampling could be improved from run_1 a further 1.2million more runs of the model (run_2) with reduced parameter ranges (where these could be determined from behavioural simulations that resulted in constrained parameter ranges from run_1, see Table 4) were generated. This run also employed a uniform sampling strategy. The results presented in the following result sections are initially from run_1, but the final dotty plots and confidence limits presented in Figures 8, 9 and 10 are calculated from run_2, having a total number of behavioural parameters sets shown in the second part of Table 5.

4.1 Simulating the discharge and ∇_{wt} responses separately

Figure 7 shows the distribution of behavioural parameter values (from run_1) for both LU's over the sampled ranges listed in Table 2. Each column is associated with parameter ranges that meet one or more behavioural criteria using the multiple objectives identified in Table 3. Table 5 lists the number of behavioural simulations associated with each criteria.

Simulations meeting the behavioural criteria for discharge (Figure 7 - column 1) show limited parameter sensitivity for the ranges sampled, primarily *SZM* and $\ln(T_0)$ from the HS_{LU} show any sensitivity, with only the $\ln(T_0)$ parameter constrained to its lower range from the initial sampling limits listed in Table 2. Table 5 also lists the large number (41% of the initial sample of 5.6million runs) of simulations that meet the behavioural threshold for the discharge criteria. Surprisingly almost no sensitivity is seen in the VB_{LU} parameters for simulations of discharge. *Freer et al.* (2003) reported a similar effect for Dynamic TOPMODEL simulations at PMRW where 3 LU's were identified. In that study, parameters for the VB_{LU} showed little sensitivity to discharge simulations. At PMRW this was attributed to a greater sensitivity of the HS_{LU} dynamics to wetting and drying cycles needed to capture the high seasonality in observed discharge. In both cases the insensitivity of the VB_{LU} could well be attributed to the relatively small areal extent of the LU (9% for PMRW and 12% for Maimai) as well as the product of landscape position and model conceptualisation. This suggests that for discharge simulations a simpler conceptual form for the VB_{LU} could be identified, potentially resulting in fewer, more easily identifiable parameters. Finally S_{max}, proven to be an important parameter for other applications of Dynamic TOPMODEL (i.e. Beven and Freer, 2001a) appears redundant here, perhaps reflecting the climatic and physical conditions found at Maimai (i.e. steep slopes, rapid transmissivities, wet conditions). The need for a dynamic subsurface saturated zone that primarily controls wetting and drying cycles is not required for behavioural simulations. For

DRAFT COPY

simulations meeting the fuzzy criteria for both tensiometer sites independently (see Table 3) there are different but somewhat consistent results with the discharge simulations (Figure 7 columns 2 and 3 for the *P5* and *NS* sites respectfully).

For the *P5* site the HS_{LU} parameters *SZM* and ln(T_0) show similar distributions to the discharge simulations however $\Delta\theta_1$ is now highly sensitive to its lower range and SR_{max} also shows some sensitivity. The high sensitivity of $\Delta\theta_1$ should be expected, this parameter is one of the primary controls of the mean depth of the predicted ∇_{wt} within the model (the difference in the water content between saturation and field capacity), effectively a simple scaling of the local moisture deficit. For this criteria the number of behavioural simulations is much reduced (see Table 5), and can be attributed to the high sensitivity of $\Delta\theta_1$ reported reducing the efficiency of the uniform sampling employed.

Comparing the number of behavioural simulations for the *NS* site with those for the *P5* site the latter produces a considerably greater number. This is directly reflected in the broader range of $\Delta\theta_1$ for the VB_{LU} that partly results from wider fuzzy limits in the observed ∇_{wt} series (especially at depth) shown in Figure 4. What is surprising about the simulations meeting the *NS* behavioural criteria is how little sensitivity is observed within the HS_{LU}, especially given the proximity of this LU to the *NS* site (see Figure 1.b).

4.2 Meeting discharge and/or tensiometer criteria for more than one source of information

Parameter distributions from simulations that are behavioural for a combination of two PM criteria from Table 3 are shown in Figure 7 columns 4-6 and for a combination of all PM in Figure 8. To highlight the combined effect of the PM's to the parameter sensitivity the dotty plots shown in Figure 7 are a multiplicative combination of discharge and ∇_{wt} PM's and an additive combination of the combined *NS* and *P5* ∇_{wt} PM's. These resultant sensitivities would be similar to those that would be shown though the more general application of Bayes equation in the standard GLUE procedure. Due to the insensitivity of the VB_{LU} for discharge, coupled with the similarity of the behavioural distributions for discharge and *P5* PM for the HS_{LU}, the combined behavioural PM distributions almost always reflect the PM sensitivity for the individual ∇_{wt} distributions previously shown in Figure 7 (columns 2 and 3). Combining discharge with the *NS* and *P5* PM's further reduces the number of behavioural parameter sets (Table 5). However only 3.8% of parameter sets are retained for a combined *NS* and *P5* PM from the maximum possible number of behavioural parameter sets for either of these two sites. This incompatibility of parameter distributions is the result of the general insensitivity of each LU's parameters to simulating the other LU's ∇_{wt} information. In combination the void space throughout the parametric hyperspace (i.e. the area of the parameter space where no behavioural simulations are found) increases rapidly due to the constraining of parameter ranges in

DRAFT COPY

both LU's, thus reducing sampling efficiencies (i.e. a reduction in the percentage of the total number of simulations that are behavioural).

Figure 8 shows the marginal posterior likelihood weighted distributions of individual parameters as histograms, and the interaction of parameters both within and between LU's for the final behavioural parameter sets constrained using all 3 PM's from run_2. Parameter sensitivities are similar to those shown in Figure 7 column 6 for the combined *NS* and *P5* PM's (note parameter ranges in Figure 8 are consistent with the ranges listed in Table 4 for run_2). Although a number of parameters are sensitive across their individual ranges, the bi-variate plots of parameter interactions show that few correlation structures are clearly identified, especially for parameter interactions between LU's. This point is confirmed by the strength of the correlation co-efficients, where only $\Delta\theta_1$ and its relationship to *SZM* and $\ln(T_0)$ for both LU's have co-efficients above +/-0.25. However more complex, non-linear and multi-dimensional structures may well exist, but this still results in parameter distributions that are equifinal. In part the poor sampling efficiencies suggest a complexity of structure within the parametric hyperspace.

The behavioural simulations for all PM's identified in Table 5 and Figure 8 were then used to determine the upper and lower possibility limits for the discharge, *NS* ∇_{wt} levels and *P5* ∇_{wt} levels, these results are presented in Figure 9 (note that discharge is also plotted in log units in Figure 9 B.). The results show that although the range of simulations generally envelope (or are within the range of) the different observations, this is not the case for all time steps, and for some periods there are significant departures. For discharge the results are encouraging, even when shown as log transformed flows (Figure 9 B.). Periods of rapid fluctuations from a general recession form are likely to reflect observed data uncertainties not yet accounted for.

The *P5* simulations are within the range of the ∇_{wt} uncertainty limits for most of the study period. The exception to this is the period before the 29th October storm event where the distribution of simulated ∇_{wt} levels are deeper than those of the observed. This period is preceded by a considerable recession period (for Maimai), that may suggest even moderate wetting up sequences are not well represented in the model dynamics. Non-linearity in catchment response can be highlighted by the relationship between peak discharges and the maximum ∇_{wt} rise at the *P5* site. A consistent pattern is not apparent, where considerable differences in discharge peaks produce similar rises in observed ∇_{wt} levels, in some cases smaller discharge peaks result in the highest ∇_{wt} rises.

The *NS* simulations show the most extensive departures from the range of ∇_{wt} observations., primarily during the recession period previously mentioned above. This rapid decline in the observed ∇_{wt} levels (that seems to begin to be replicated at the end of the *NS* observations) may be systematic of local phenomenon such as non-linearity in the storage-discharge relationship with the

DRAFT COPY

different soil horizons. However this could also be the result of a breakdown in the relationship between the –ve matric potentials and height above ∇_{wt} at the *NS* site. Certainly *McDonnell* (1990) reported a bedrock depth of 0.5m at the *NS* pit face, however this local depth is highly variable (as noted by McDonnell, 1990) as identified by the 0.78m tensiometer placement in Nest 1 (see Figure 1(c)). Departures occur in the *NS* simulations during periods where +ve matric potential readings are observed, which suggest the rapidly declining ∇_{wt} levels have some validity.

Finally the characteristics in the *NS* and *P5* simulations relate well to the variability in the parameter distributions for these LU's. For the HS_{LU} lower *SZM* and higher $\Delta\theta_1$ parameter distributions reflect the steeper and deeper *P5* ∇_{wt} recession characteristics. Previously *Freer* (1998) using the original form of TOPMODEL identified these controlling parameters as the reason why the model was unable to simulate the ∇_{wt} responses at both sites using homogeneously applied catchment scale parameters.

4.3 Constraining model responses and the efficiency of sampling

For the different behavioural parameter sets identified in Table 5, Figure 10 shows the distribution of a number of summary model responses calculated from each simulation run. Figure 10 shows that the range of model behaviour can vary considerably between the different behavioural parameter sets (i.e. peak discharge). In nearly all cases (apart from Sum Discharge Figure 10a where limits for this measure are generally consistent for all PM's) simulations conditioned using all the PM's show the smallest range of model behaviours. Treated individually, the *P5* PM constrains the model responses most; that this also occurs for the range of Peak Discharge responses is somewhat surprising. Perhaps this is indicative of the discriminatory power of the R^2 measure, the strength of which has been questioned in a number of recent studies (i.e. Gupta et al., 1998; Legates and McCabe Jr., 1999; Freer et al., 2003). The average ∇_{wt} depth ranges for the *P5* and *NS* sites (Figure 10e and f) identify why only a small proportion of simulations that are behavioural for one site are also behavioural for the other. The distributions of these average ∇_{wt} depths have very little overlap and must reflect a general inability to simulate the ∇_{wt} observations to an acceptable level.

Distributions of model responses for the *NS* PM, coupled with the lack of sensitivity in parameter distributions for the same PM shown in Figure 8, suggest this PM has the least explanatory power. This leads to output model responses that seem uncharacteristic of catchment behaviour (i.e. the high peak discharges and maximum saturated areas shown in Figure 10b and c). Partly this is a product of the information content in the fuzzy *NS* ∇_{wt} observations, i.e. generally wider limits and lower amplitudes of responses compared to the *P5* data, but also this reflects the general insensitivity of this LU described in section 4.1.

DRAFT COPY

4.4 Can we improve the model structure and parameter representation?

The simulation results thus far presented have resulted in good simulations of the Maimai catchment discharge and ∇_{wt} responses. Where this has not been the case (i.e. the deeper ∇_{wt} recessions at the *NS* site) further data collection would be required to confirm the potential for increased observational errors. What is not clear from these results is whether additional data sets (i.e. more ∇_{wt} sites or the use of tracer data) would still maintain a compatible set of parameter estimates or lead to the rejection of all model simulations. Would each new information require a new set of parameter distributions and/or changes to the basic model structure, similar to that reported by *Lamb et al.* (1998)? An important question for modellers in this regard is how approximate can a model be and still retain an element of realism in predicting quantities and fluxes of interest. Even if the general structure of dynamic TOPMODEL is a reasonable approximation for the hydrological response at Maimai, model parameters are more heterogeneous in space than our definition of 2 LU's have characterised. The use of internal state data is desirable, but should we expect such information to have overlapping joint probability distributions of behaviour with the model dynamics without biasing the results unduly? Certainly the information pertaining to the characterisation of hydrological responses at Maimai used in this study are still limited. This is important as the effective gridscale uncertainties in the ∇_{wt} responses for both the *NS* and *P5* sites may well be greater than those currently identified. We may still be biasing our range of simulated behaviour due to poorly defined observational uncertainties. Perhaps what is more likely if additional observations were available is that each new site that is added to the constraining information (in this case ∇_{wt} information) will have characteristics that are in some way unique (Beven, 2000). Small to potentially large variations in local parameter distributions may be required to simulate such information. Observations from the *NS* and *P5* sites clearly show that the subsurface dynamics are different, the sites are clearly drawn from topographically distinct regions of the catchment, and that these differences have been reflected in the behavioural parameter distributions for the two LU's.

The general hydrological regime at Maimai lends itself to the primary assumptions embedded in the dynamic TOPMODEL framework. However the perceptual model of the subsurface flow processes at Maimai includes mechanisms that are not explicitly accounted for in the model structure, i.e. horizontal preferential macropore flowpaths, vertical bypassing to depth, variable porosity values in the organic and mineral soil horizons (see Mosley, 1979; McDonnell, 1990; McGlynn et al., 2002). With this in mind our modelling results are surprisingly good for the ∇_{wt} dynamics. That parameter estimates and model responses seem to make physical sense with observational data from Maimai is also encouraging. For example *Pearce et al.* (1986) suggest maximum saturated areas at Maimai are in the region of 4-7%, comparing well with the results presented in Figure 10c for the simulations constrained by all the PM's.

DRAFT COPY

However insensitivity in many parameters and a lack of interaction between parameters suggests the conceptual framework of the model for Maimai catchment could be improved, if only to reduce the redundancy of certain parameters. Would it be possible with increased information to identify for the local place (i.e. a LU) a subset of parameters that characterises the uniqueness of place? In this case a combination of *SZM*, ln(T_0) and $\Delta\theta_1$ would seem appropriate for characterising the ∇_{wt} dynamics. Or would new subsets of parameters and model function be required for the inclusion of each new place? What data are certain enough for local places to ensure that inverse reasoning, namely that local observations can be effectively used to estimate an appropriate distribution of parameters (e.g. Jordan, 1994; Seibert et al., 1997; Lamb et al., 1998) and model function, can be applied effectively?

4.5 The use of fuzzy rules applied to imperfect and imprecise knowledge

The results discussed so far have assumed the rejection of all models that did not meet the behaviourability criteria. However these criteria are not absolute; they were identified using knowledge of both standard hydrological practice in uncertainty acceptance and specific understanding of the study sites and measurement techniques employed. It has been highlighted in the preceding section that the inclusion of further observations which the model is required to replicate may necessitate increasing complexity in the model: therefore if no such additional complexity is allowed, uncertainty limits may have to be relaxed to take into account the local deviations inherent in the catchment.

The use of the fuzzy measures demonstrated in particular the difficulties of using imprecise knowledge of the catchment behaviour in a meaningful way. Originally a multiplicative form of the fuzzy measure was considered, where multiplication rather than summation was used to combine the fuzzy scores for each timestep: however this lead to all models being rejected as each had at least one point outside the observed tensiometer limits, thus setting the score to zero. This may reflect inadequacies in the model, but equally may reflect the incomplete knowledge of ∇_{wt} at gridcell scale. Given that ∇_{wt} was sampled at 9 and 11 locations in the *P5* and *NS* sites respectively, demonstrating significant variation across the gridcell, it would be reasonable to suppose that this sample reflects only part of the range actually present across the gridcell. A case therefore could be made for widening the fuzzy limits beyond the observed extreme points. The extent of this widening would have to be determined based on observed range and gradient of tensiometer readings, together with coverage of gridcell. At the Maimai study site, tensiometers only covered areas of 4.5m*1m and 4m*0.5m at the *P5* and *NS* sites, therefore capturing variability in only 0.18 and 0.08 of the 25m^2 gridcell area, we should not forget the limitations of our observations in relation to the scale at which our model simulations are being applied.

DRAFT COPY

A further source of uncertainty occurs in the determination of the ∇_{wt} position. The tensiometer reading itself is constrained by the scale at which the measurement is taken: variable values may therefore be unsuitable for use at different scales and cannot be held to represent the full small-scale complexity of the system. Further to this, Section 3.1 noted that the transformation from tensiometer reading to ∇_{wt} was subject to uncertainty which increases with the magnitude of the negative potential recorded. It would therefore be preferable to formulate fuzzy limits to take into account this changing uncertainty in our transformation equations.

The variability of the tensiometer readings in time and space are clearly not the only source of observed data uncertainty driving our model simulations. Hydrologist have tended to treat our main input forcing errors (in rainfall and evapotranspiration) and our observed outputs (discharge) as either unimportant, implicitly in the relaxation of the acceptability criteria (i.e. this paper and Beven and Binley, 1992) or through the derivation of likelihood measures with assumed error structures (i.e. Sorooshian, 1981). However, there is clearly a need to more explicitly account for the potential for errors (having known and unknown error structures) in all the observed data series we use to drive our model simulations, especially for rainfall inputs and for stage-discharge relationships for rated channel sections. Recently papers have begun to confront these issues, an analysis of the effect of rainfall forcing errors (using multipliers on the rainfall totals) has been undertaken by Kavatski et al. ????. Perhaps we need to consider that all of our observations are neither deterministic or have a known and stationary error structure, that they are in fact ‘grey’ in quantity, and that the level of greyness is likely to be variable in both space and time. The challenge will be to develop methods that are both realistic and flexible about the nature of such errors but still maintain a sound scientific justification and/or evaluation of the error terms. We agree with the recent comments of Seibert and McDonnell (2002) that the use of fuzzy membership functions is one method that lends itself to this type of error analysis approach but for all observed data series.

To sum up, the use of fuzzy performance measures is a powerful and flexible tool in situations where there is no or incomplete knowledge of the error structure and local variability of the phenomenon. The exact form of the measure can be designed to reflect uncertainties particular to the modelling situation. Equally, this very adaptability means that consistent, global rules for function definition cannot be specified; instead the user must be clear as to the motivation that underlies the chosen measure, as was the aim in this paper.

5 CONCLUSIONS

This paper presents an approach to assessing the internal accuracy of dynamic TOPMODEL, recognising that internal state data available to the modeller are inherently uncertain. The model was applied to the Maimai M8 catchment in New Zealand, and was refined by using two

DRAFT COPY

topographically-distinct landscape units ('Hillslope' and 'Valley Bottom') with separate parameterisations. For each location, a nest of tensiometers located within an area commensurate with the model gridscale provided a distribution of matric potentials which were then converted to water table depth. These depths were used together with rainfall-runoff data to constrain the model using the Generalised Likelihood Uncertainty Estimation methodology.

The use of localised data to assess model performance presents particular problems to the modeller. Unlike aggregated components such as river discharge, water table levels more strongly reflect localised and smaller-scale characteristics of catchment processes, and this is clearly demonstrated in the variability shown in the tensiometer readings within the area of one gridcell. Although tensiometers were placed such as to avoid cracks and voids in the soils, the effects of heterogenity of soil characteristics and flow pathways, such as macropores and soil structure, cannot be avoided. When these locally conditioned data are used to constrain the model, the model structure and parameterisation may then be biased towards these local structures which may not be representative of the catchment as a whole, or indeed at a scale comparable to the model gridscale, the smallest spatial scale of hydrological process representation in the model. As increasing numbers of these local criteria are enforced, the model is unable to incorporate the complexity of local observations and the danger is that all simulations are rejected as non-behavioural.

In an attempt to respond to these problems, fuzzy performance measures rather than a more formal deterministic evaluation were used. These allow the modeller to include knowledge of errors in the internal state data presented and are not constrained by the need for a particular error structure. In this study a trapezoidal form of fuzzy measure was used to incorporate knowledge of the distribution of water table levels at the two test sites. However, despite the use of fuzzy measures to relax the strictness of the criteria, the retention rate for parameter sets picked using the more efficient constrained sampling ranges in run$_2$ dropped from 84.69% (discharge only) to 0.26% when using all three performance measures. This sparseness of behavioural parameter sets suggests both a complex structure within the parameter space, and individuality of water table levels internally to the catchment. Intuition suggests that the *NS* water table data should be less location dependent and more representative of the overall ∇_{wt} dynamics of the catchment than the *P5* data. The *NS* site integrates a greater catchment drainage area and therefore proportionally this site is more likely to be representative of the overall catchment dynamics that characterise streamflow response. This was borne out by the higher sampling efficiency when using internal data only from the *NS* site as oppose to only the *P5* site.

This study has demonstrated that when using dynamic TOPMODEL to make predictions about internal catchment dynamics, it is not sufficient to condition the model using aggregate performance data such as discharge. The uniqueness of place demonstrated at and within each gridcell area is not reflected in such integrated measures; and therefore internal state data are required to enable

DRAFT COPY

model calibration if the model is to provide an accurate representation of the catchment processes. Important questions have been raised as to the feasibility of introducing multi-criteria performance measures, these will become ever more pertinent as internal state data become more readily available through improved measurement and remote-sensing techniques. Fuzzy measures are becoming more widely accepted as an appropriate method for dealing with uncertain calibration data (e.g. Seibert and McDonnell, 2002), and have been shown here to present a flexible structure within which the modeller can combine observational data and site-specific knowledge on within-gridcell variability.

6 ACKNOWLEDGEMENTS

We thank Stefan Uhlenbrook, Rodger Grayson, Ralf Ludwig and one anonymous reviewer for their valuable comments. This research was partly funded by the ????

Thanks to NZ people ????

DRAFT COPY

7 REFERENCES

Anderton, S., Latron, M. and Gallart, F., 2002a. Sensitivity analysis and multi-response, multi-criteria evaluation of a physically based distributed model. Hydrological Processes, 16(2): 333-353.

Anderton, S.P. et al., 2002b. Internal evaluation of a physically-based distributed model using data from a Mediterranean mountain catchment. Hydrology and Earth System Sciences, 6(1): 67-83.

Aronica, G., Bates, P.D. and Horritt, M.S., 2002. Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE. Hydrological Processes, 16(10): 2001-2016.

Barling, R.D., Moore, I.D. and Grayson, R.B., 1994. A Quasi-Dynamic Wetness Index for Characterizing the Spatial- Distribution of Zones of Surface Saturation and Soil-Water Content. Water Resources Research, 30(4): 1029-1044.

Bathurst, J.C. and O'Connell, P.E., 1992. Future of Distributed Modelling - the Systeme-Hydrologique- Europeen. Hydrological Processes, 6(3): 265-277.

Beck, M.B. and Halfon, E., 1991. Uncertainty, Identifiability and the Propagation of Prediction Errors - a Case-Study of Lake-Ontario. Journal of Forecasting, 10(1-2): 135-161.

Beven, K.J., 1989. Changing Ideas in Hydrology - the Case of Physically-Based Models. Journal of Hydrology, 105(1-2): 157-172.

Beven, K.J., 1996. Equifinality and uncertainty in geomorphological modelling., The Scientific Nature of Geomorphology, Proceedings of the 27th Binghampton Symposium in Geomorphology held 27-29 September 1996,, Binghampton.

Beven, K.J., 1997. TOPMODEL: A critique. Hydrological Processes, 11(9): 1069-1085.

Beven, K.J., 2000. Uniqueness of place and process representations in hydrological modelling. Hydrology and Earth System Sciences, 4(2): 203-213.

Beven, K.J., 2002. Towards a coherent philosphy for modelling the environment. Proceedings of the Royal Society London A, 458: 2465-2484.

Beven, K.J. and Binley, A., 1992. The future of distributed models: Model calibration & uncertainty prediction. Hydrological Processes, 6: 279-298.

Beven, K.J. and Freer, J., 2001a. A dynamic TOPMODEL. Hydrological Processes, 15(10): 1993-2011.

Beven, K.J. and Freer, J., 2001b. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology, 249(1-4): 11-29.

Beven, K.J. and Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. Hydrological Sciences Bulletin-Bulletin Des Sciences Hydrologiques, 24(1): 43-69.

Blazkova, S., Beven, K.J. and Kulasova, A., 2002. On constraining TOPMODEL hydrograph simulations using partial saturated area information. Hydrological Processes, 16(2): 441-458.

Bloschl, G., Gutknecht, D. and Kirnbauer, R., 1992. On the evaluation of distributed hydrologic models., Advances in Distributed Hydrology (workshop), Seriate (Bergamo), Italy, 25-26th June.

Burt, T.P. and Butcher, D.P., 1985. Topographic controls of soil moisture distributions. Journal of Soil Science, 36: 469-486.

Burt, T.P. and Butcher, D.P., 1986. Development of topographic indices for use in semi-distributed hillslope runoff models. Z. Geomorphological, 58: 1-19.

Butcher, D.P., 1985. Field verification of topographic indicies for use in a hillslope runoff model. Unpublished Ph. D Thesis, Huddersfield Polytechnic, Huddersfield.

de Grosbois, E., Hooper, R.P. and Christophersen, N., 1988. A Multisignal Automatic Calibration Methodology for Hydrochemical Models - a Case-Study of the Birkenes Model. Water Resources Research, 24(8): 1299-1307.

Franks, S.W., Gineste, P., Beven, K.J. and Merot, P., 1998. On constraining the predictions of a distributed moder: The incorporation of fuzzy estimates of saturated areas into the calibration process. Water Resources Research, 34(4): 787-797.

Freer, J., Beven, K. and Ambroise, B., 1996. Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach. Water Resources Research, 32(7): 2161-2173.

Freer, J.E., 1998. Uncertainty and Calibration of Conceptual Rainfall Runoff Models. Unpublished PhD Thesis, Lancaster University, Lancaster, 282 pp.

DRAFT COPY

Freer, J.E., Beven, K.J. and Peters, N.E., 2003. Multivariate seasonal period model rejection within the generalised likelihood uncertainty estimation procedure. In: Q. Duan, H. Gupta, S. Sorooshian, A.N. Rousseau and R. Turcotte (Editors), Calibration of watershed models. AGU, Water Science and Application Series, Washington, pp. 346.

Grayson, R. and Bloschl, G., 2000. Spatial patterns in catchment hydrology: Observations and modelling. Cambridge University Press, Cambridge, 404 pp.

Grayson, R.B., Moore, I.D. and McMahon, T.A., 1992. Physically Based Hydrologic Modelling .2. Is the Concept Realistic. Water Resources Research, 28(10): 2659-2666.

Guntner, A., Uhlenbrook, S., Seibert, J. and Leibundgut, C., 1999. Multi-criterial validation of TOPMODEL in a mountainous catchment. Hydrological Processes, 13(11): 1603-1620.

Gupta, H.V., Sorooshian, S. and Yapo, P.O., 1998. Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. Water Resources Research, 34(4): 751-763.

Hewlett, J.D. and Hibbert, A.R., 1967. Factors affecting the response of small watersheds to precipitation in humid areas., International Symposium on Forest Hydrology.

Hooper, R.P., Stone, A., Christophersen, N., de Grosbois, E. and Seip, H.M., 1988. Assessing the Birkenes Model of Stream Acidification Using a Multisignal Calibration Methodology. Water Resources Research, 24(8): 1308-1316.

Jordan, J.P., 1994. Spatial and Temporal Variability of Stormflow Generation Processes on a Swiss Catchment. Journal of Hydrology, 153(1-4): 357-382.

Kirkby, M., 1975. Hydrograph Modelling Strategies. In: R. Peel, M. Chisholm and P. Haggett (Editors), Processes in Physical and Human Geography. Heinemann, London, pp. 69-90.

Koide, S. and Wheater, H.S., 1992. Subsurface Flow Simulation of a Small Plot at Loch-Chon, Scotland. Hydrological Processes, 6(3): 299-326.

Kosugi, K. and Inoue, M., 2002. Estimation of hydraulic properties of vertically heterogeneous forest soil from transient matric pressure data. Water Resources Research, 38(12): art. no.-1322.

Kuczera, G., 1983. Improved Parameter Inference in Catchment Models .2. Combining Different Kinds of Hydrologic Data and Testing Their Compatibility. Water Resources Research, 19(5): 1163-1172.

Kuczera, G. and Mroczkowski, M., 1998. Assessment of hydrologic parameter uncertainty and the worth of multiresponse data. Water Resources Research, 34(6): 1481-1489.

Lamb, R., Beven, K.J. and Myrabo, S., 1997. Discharge and water table predictions using a generalized TOPMODEL formulation. Hydrological Processes, 11(9): 1145-1167.

Lamb, R., Beven, K.J. and Myrabo, S., 1998. Use of spatially distributed water table observations to constrain uncertainty in a rainfall-runoff model. Advances in Water Resources, 22(4): 305-317.

Legates, D.R. and McCabe Jr., G.J., 1999. Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation. Water Resources Research, 35(1): 233-241.

McDonnell, J.J., 1989. The age, origin and pathway of subsuirface stormflow in a steep, humid headwater catchment. Unpublished PhD Thesis, University of Canterbury, Christchurch, N.Z., 270 pp.

McDonnell, J.J., 1990. A Rationale for Old Water Discharge through Macropores in a Steep, Humid Catchment. Water Resources Research, 26(11): 2821-2832.

McDonnell, J.J., 1993. Technical notes: Electronic versus fluid multiplexing in recording tensiometer systems. ASAE Publication, 36(2): 459-462.

McGlynn, B.L., McDonnell, J.J. and Brammer, D.D., 2002. A review of the evolving perceptual model of hillslope flowpaths at the Maimai catchments, New Zealand. Journal of Hydrology, 257(1-4): 1-26.

Mew, G., Webb, T.H., Ross, C.W. and Adams, J., 1975. Soils of Ignangahua depression, South Island, New Zealand. 17, N.Z. Soil Survey, Christchurch.

Mosley, M.P., 1979. Streamflow generation in a forested watershed, New Zealand. Water Resources Research, 15(4): 795-806.

Motovilov, Y., Gottschalk, L., Engeland, K. and Rodhe, A., 1999. Validation of a distributed hydrological model against spatial observations. Agricultural and Forest Meteorology, 98-99: 257-277.

Mroczkowski, M., Raper, G.P. and Kuczera, G., 1997. The quest for more powerful validation of conceptual catchment models. Water Resources Research, 33(10): 2325-2335.

Pearce, A.J. and McKercher, A.I., 1979. Upstream generation of storm runoff. In: D.L.A. Murray, P. (Editor), Physical Hydrology - New Zealand Experience. New Zealand Hydrological Society, Wellington, pp. 165-192.

Pearce, A.J., Stewart, M.K. and Sklash, M.G., 1986. Storm Runoff Generation in Humid Headwater Catchments .1. Where Does the Water Come From. Water Resources Research, 22(8): 1263-1272.

DRAFT COPY

Peters, N.E., Freer, J. and Beven, K.J., 2003. Modeling hydrologic responses in a small forested catchment (Panola Mountain, Georgia, USA) - a comparison of the original and a new dynamic TOPMODEL. Hydrological Processes, 17(3): 345-362.

Peters, N.E., Freer, J.E. and Beven, K.J., 2001. Modeling hydrologic responses in a small forested watershed by a new dynamic TOPMODEL (Panola Mountain, Georgia, USA). In: S. Uhlenbrook, C. Leibundgut and J.J. McDonnell (Editors), Runoff Generation and Implications for River Basin Modeling. Freiburger Schriften zur Hydrologie, Band 13, Frieburg, pp. 318-325.

Quinn, P.F., Beven, K.J., Chevallier, P. and Planchon, O., 1991. The prediction of hillslope flow patterns for distributed hydrological modelling using digital terrain models. Hydrological Processes, 5: 59-59.

Ross, T.J., 1995. Fuzzy logic with engineering applications. McGraw-Hill, New York, 600 pp.

Rowe, L.K., 1979. Rainfall interception by a Beech- Podocarp- Hardwood forest near Reefton, North Westland, New Zealand. Journal of Hydrology (NZ), 18(2): 63-87.

Rowe, L.K., Pearce, A.J. and O'Loughlin, C.L., 1994. Hydrology and Related Changes after Harvesting Native Forest Catchments and Establishing Pinus-Radiata Plantations .1. Introduction to Study. Hydrological Processes, 8(3): 263-279.

Seibert, J., Bishop, K.H. and Nyberg, L., 1997. A test of TOPMODEL's ability to predict spatially distributed groundwater levels. Hydrological Processes, 11(9): 1131-1144.

Seibert, J. and McDonnell, J.J., 2002. On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. Water Resources Research, 38(11): art. no.-1241.

Sherlock, M.D., Chappell, N.A. and McDonnell, J.J., 2000. Effects of experimental uncertainty on the calculation of hillslope flow paths. Hydrological Processes, 14(14): 2457-2471.

Sklash, M.G., 1990. Environmental isotope studies of storm & snowmelt runoff generation. In: M.G.B. Anderson, T.P. (Editor), Process Studies in Hillslope Hydrology. John Wiley, New York, pp. 401-435.

Sorooshian, S., 1981. Parameter-Estimation of Rainfall - Runoff Models with Heteroscedastic Streamflow Errors - the Non-Informative Data Case. Journal of Hydrology, 52(1-2): 127-138.

Sorooshian, S. and Gupta, V.K., 1985. The Analysis of Structural Identifiability - Theory and Application to Conceptual Rainfall-Runoff Models. Water Resources Research, 21(4): 487-495.

Stephenson, G.R. and Freeze, R.A., 1974. Mathematical simulation of subsurface flow contributions to snowmelt runoff, Reynolds Creek watershed, Idaho. Water Resources Research, 10(2): 284-298.

Thiemann, M., Trosset, M., Gupta, H. and Sorooshian, S., 2001. Bayesian recursive parameter estimation for hydrologic models. Water Resources Research, 37(10): 2521-2535.

Torres, R. and Alexander, L.J., 2002. Intensity-duration effects on drainage: Column experiments at near-zero pressure head. Water Resources Research, 38(11): art. no.-1240.

Troch, P.A., Mancini, M., Paniconi, C. and Wood, E.F., 1993. Evaluation of a Distributed Catchment Scale Water-Balance Model. Water Resources Research, 29(6): 1805-1817.

Uhlenbrook, S. and Leibundgut, C., 2002. Process-oriented catchment modelling and multiple-response validation. Hydrological Processes, 16(2): 423-440.

Vertessy, R.A. and Elsenbeer, H., 1999. Distributed modelling of storm flow generation in an Amazonian rain forest catchment: Effects of model parameterization. Water Resources Research, 35(7): 2173-2187.

Wagener, T. et al., 2001. A framework for development and application of hydrological models. Hydrology and Earth System Sciences, 5(1): 13-26.

Webster, J., 1977. The hydrological properties of the forest floor under Beech-Podocarp-Hardwood forest, North Westland. Unpublished Ph. D Thesis Thesis, Christchurch University, Canterbury, New Zealand.

Wigmosta, M.S. and Lettenmaier, D.P., 1999. A comparison of simplified methods for routing topographically driven subsurface flow. Water Resources Research, 35(1): 255-264.

DRAFT COPY

	DTA Results		Observed Field Data			
Site	**ln(a/tanβ)**	**Acc. Area**	**Soil Depth (m)**	**Slope (°)**	**Total Porosity (%)**	**Saturated Conductivity (m/hr)**
Near Stream*	4.03	183.0	0.5	15	52	-
Pit 5*	3.04	101.9	1.5	34	68	-
Catchment#	-	-	0.6	-	45**	0.01 – 0.3***

*Observed field data from *McDonnell, J.J.* (pers comm.)

**Top 0.17m organic horizon 86% total porosity (39% macroporosity)

***Soil Infiltration rate 6.1m/hr

#Data taken from *McGlynn et al.*, 2002

Table 1: Local DTA values, soil and topographic characteristics for both the Near Stream and Pit 5 sites as well as average data for the Maimai Catchment

DRAFT COPY

Parameter	**Units**	**Lower Limits***	**Upper Limits***	**Description**
SZM	[m]	0.001 {*0.005*}	0.012 {*0.017*}	Form of the exponential decline in conductivity
ln(T0)	[m^2 hr^{-1}]	7.0 {-7.0}	3.0 {3.0}	Effective lateral saturated transmissivity
SR_{max}	[m]	0.005 {0.005}	0.08 {0.08}	Maximum soil root zone deficit
SR_{ini}	[m]	0.00 {0.00}	0.01 {0.01}	Initial root zone deficit
CHV	[m hr^{-1}]	250 {250}	1500 {1500}	Channel routing velocity
T_d	[hr]	0.10 {0.10}	40.0 {40.0}	Unsaturated zone time delay
$\Delta\theta$		0.05 {0.01}	0.60 {0.30}	Effective porosity
S_{max}	[m]	0.60 {0.60}	2.00 {2.00}	Maximum effective deficit of the subsurface storage zone

*Parameter upper and lower ranges for both the valley bottom and hillslope

Table 2: Parameter ranges for the VB_{LU} and (in {}'s) the HS_{LU} for the Monte-Carlo simulations

DRAFT COPY

Performance Measure	Equation	Acceptability Criteria
R^2 Discharge	* $L\left[M\left(\Theta \middle| Y_T, W_T\right)\right] = \left(1 - \sigma_\varepsilon^2 / \sigma_o^2\right)^N$	0.6
Near Stream Fuzzy Additive	Equation 2 (in text)	1000 (maximum possible 2464)
P5 Fuzzy Additive	Equation 2 (in text)	2000 (maximum possible 4149)

* Where σ_ε^2 is the error variance; σ_o^2 is the variance of the observations and N = 1

Table 3: Discharge and ∇_{wt} Performance Measures and their acceptability criteria evaluated for the Dynamic TOPMODEL GLUE simulations

DRAFT COPY

	run_1 behavioural simulations*		run_2 behavioural simulations**		run_1 and run_2
Acceptability Criteria	**Total Number**	**Sampling Efficiency (%)**	**Total Number**	**Sampling Efficiency (%)**	**Sampling Efficiency Increase**
Discharge only	2,327,664	41.56	1,016,325	84.69	2.0
NS ∇_{wt} only	196,591	3.51	118,519	9.87	2.8
P5 ∇_{wt} only	16,195	0.28	39,128	3.26	11.5
Discharge and NS ∇_{wt}	84,636	1.51	98,218	8.18	5.4
Discharge and P5 ∇_{wt}	11,987	0.21	34,205	2.80	13.3
NS ∇_{wt} and P5 ∇_{wt}	614	0.011	3,692	0.31	28.2
Discharge, NS and P5 ∇_{wt}	419	0.007	3,184	0.26	37.1

* Total number of all simulations was 5,600,000

** Total number of all simulations was 1,200,000

Table 4: Behavioural simulations for individual and combined acceptance criteria for the performance measures identified in Table 3 from both run_1 and run_2.

DRAFT COPY

Figure 1: Maimai M8 catchment: (A) The spatial variability of the ln(a/tanβ) index and (B) the spatial distribution of the VB_{LU} and HS_{LU} LU's. Details of the study area showing the position of the tensiometer instrumentation at (C) the Near Stream and (D) the Pit 5 sites.

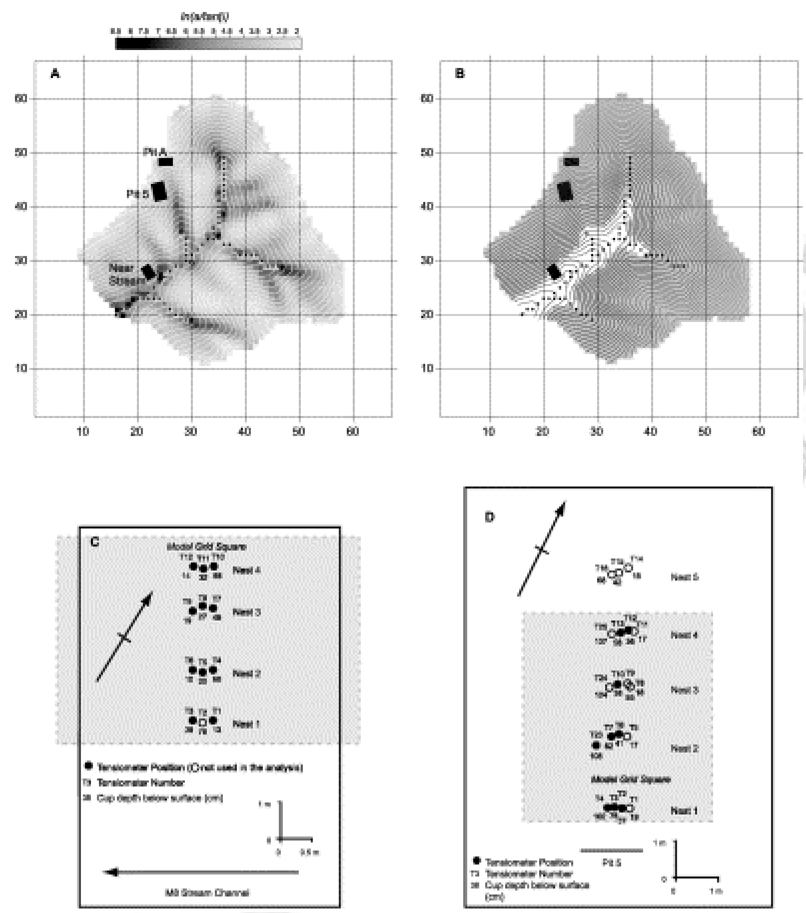


Figure 2: Near Stream site shallow and deep tensiometer readings adjusted to matric potentials. All available tensiometer responses are plotted against catchment discharge for the whole of the observation record used in this study.

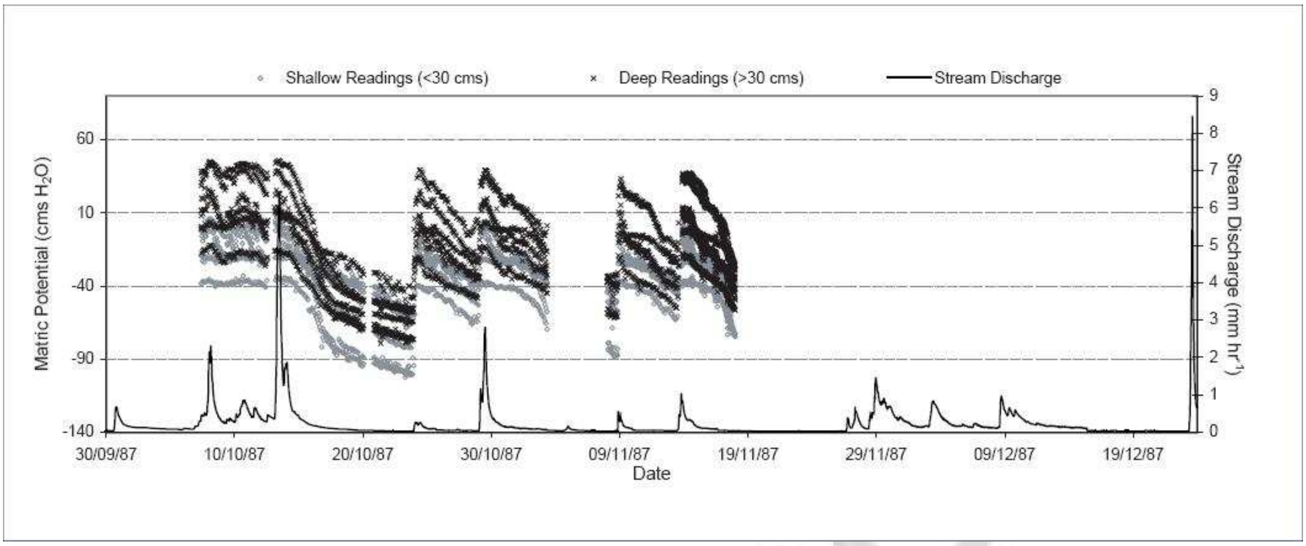


Figure 3: The relationships between observed –ve matric potentials and heights above a know ∇_{wt} (at least one tensiometer in +ve tension) for (A) the Near Stream site, Nest 4 and (B) the Pit 5 site, Nest 1. The plots show the regression curves used to describe these relationships for both cases. For the Pit 5 site, Nest 1 (C) shows for each tensiometer the depth to the water table for an extended recession period calculated using the regression relationship shown in (B).

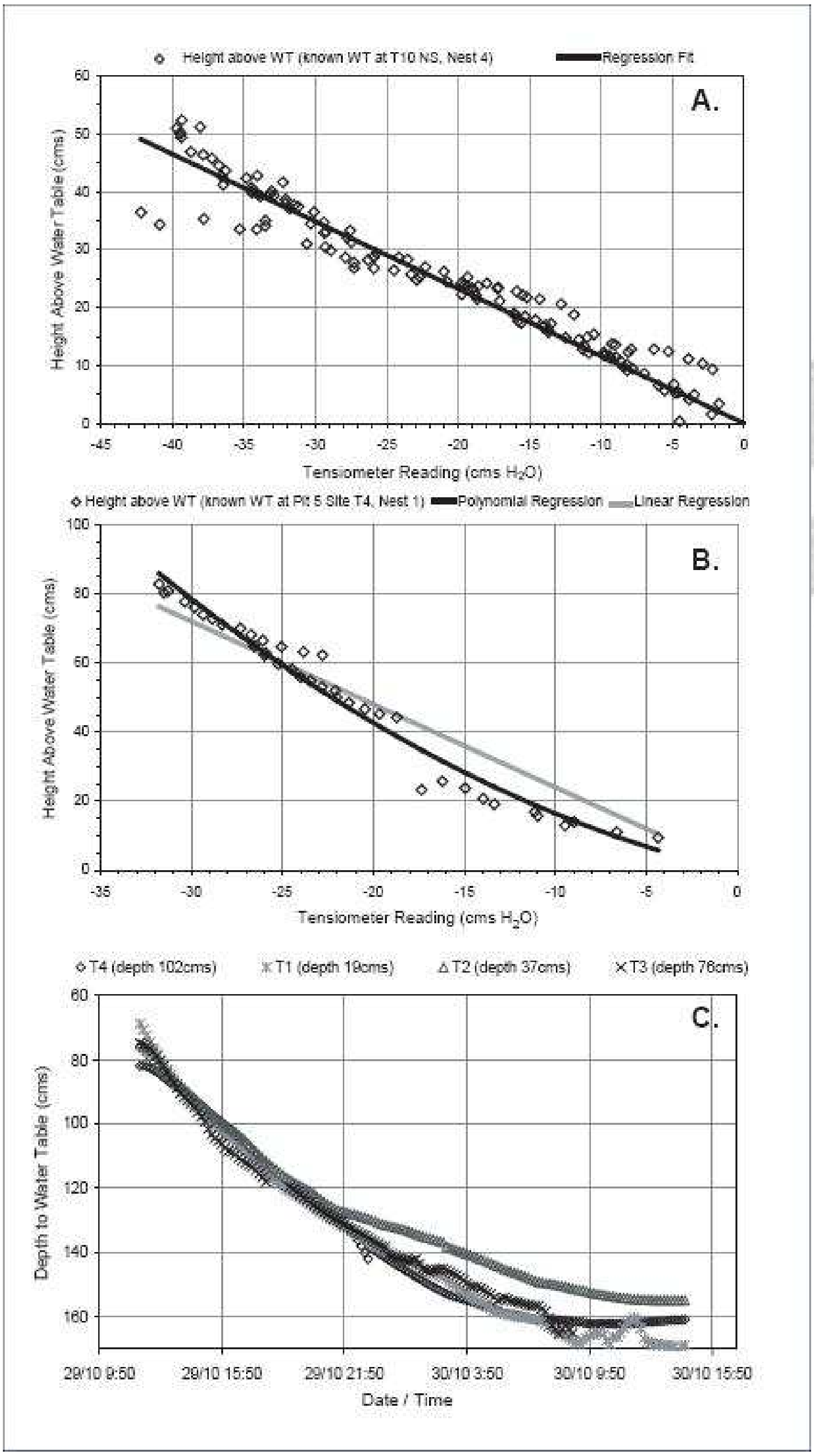


Figure 4: The variability in the range of ∇_{wt} levels for the two distribution limits used to define the fuzzy numbers for each timestep (i.e. the min-max for the *support* and the 25th and 75th for the *core* values of the fuzzy number) summarised for the whole of the observed data series by categorising the readings at each time step by the mean ∇_{wt} level. The range and mean ∇_{wt} levels are determined separately for each tensiometer site from the variability in all tensiometer observations adjusted to depth using the regression relationships shown in Figure 3. Results for the Near Stream site are shown for the *support* limits in (A) and the *core* limits in (B), with the same limits shown for the Pit 5 site in (C) and (D) respectfully. For all plots the frequency that each mean ∇_{wt} category is sampled for the whole data series is also shown.

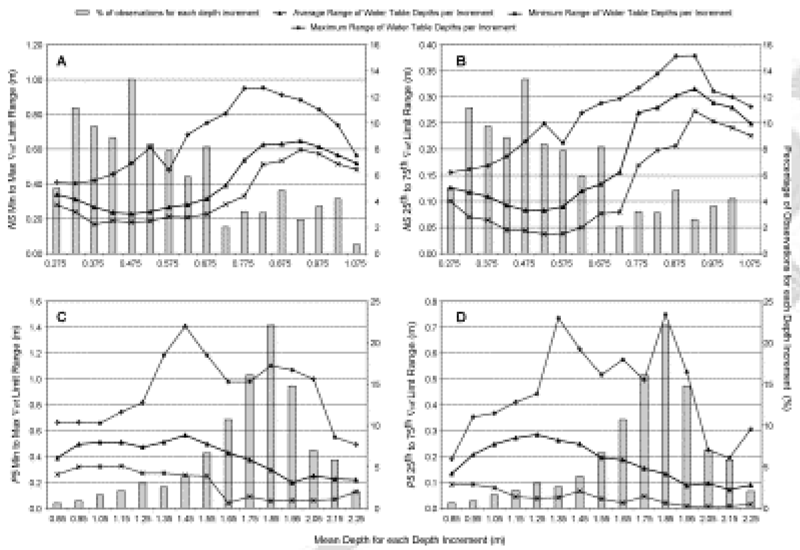
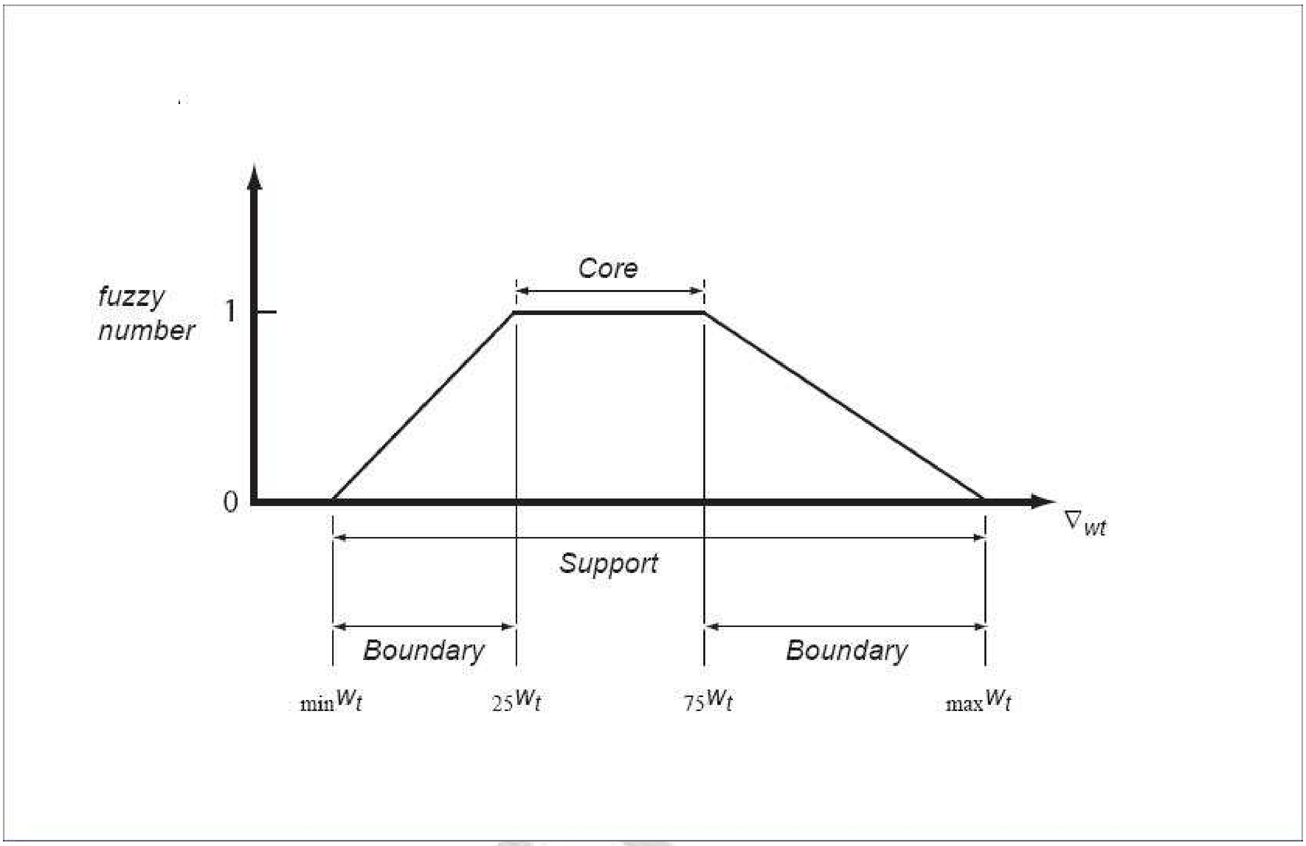


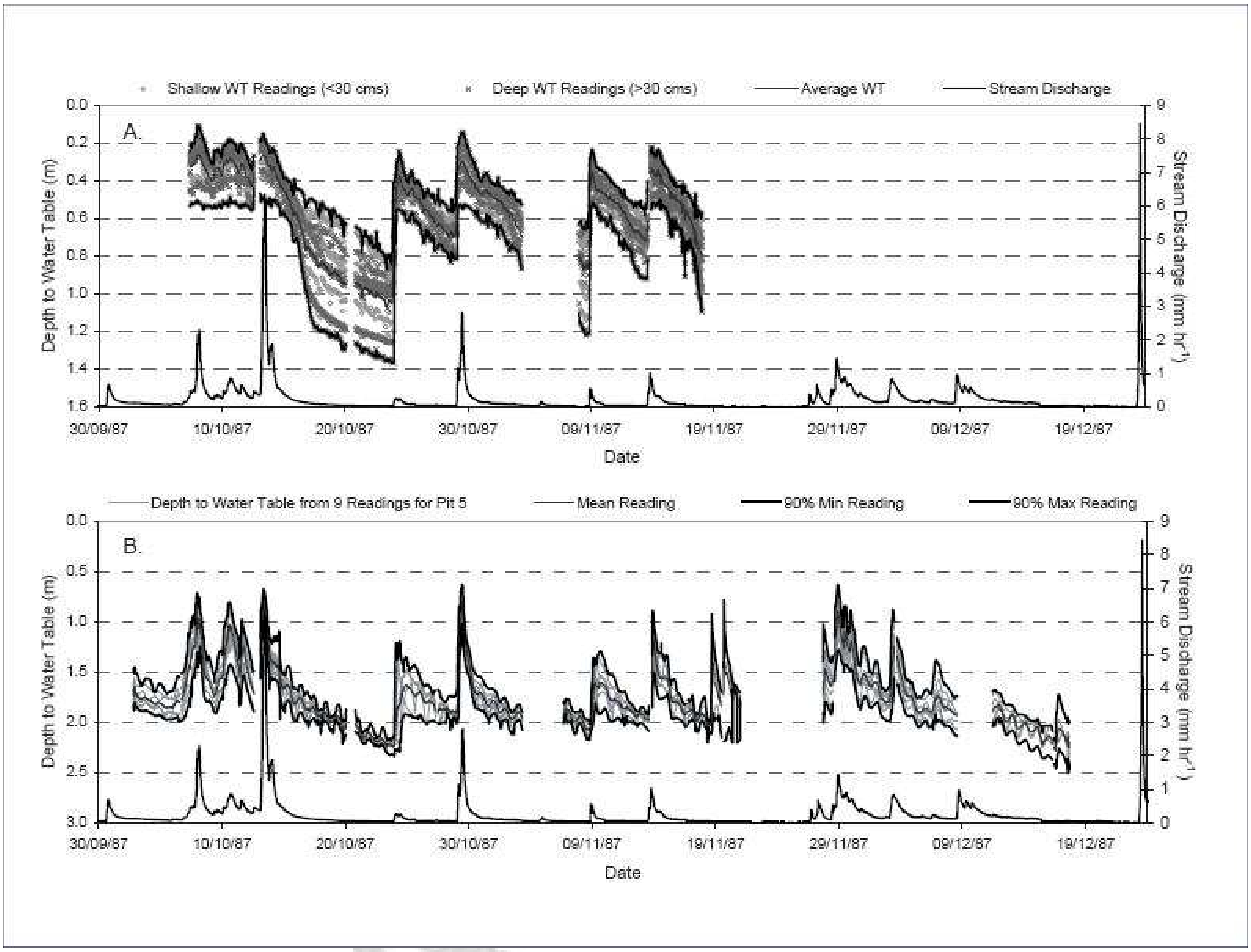
Figure 5: An example of the construction and terminology of a fuzzy number used in this study. The limits ($_{min}w_t$, $_{25}w_{t,}$ $_{75}w_t$ and $_{max}w_t$) are determined using equation 2

867



DRAFT

870 **Figure 6**: Observed water table responses calculated from the tensiometer data for both (A) Near Stream and (B) Pit 5 tensiometer sites. The plot shows the resultant upper and lower min and max limits for the water table responses defining the model gridscale variability of the observations.



873

DRAFT

Figure 7: Dotty plots of behavioural parameter distributions for both the VB_{LU} (rows1-7) and the HS_{LU} (rows 8-14) for the different performance measures (or combinations of measures) listed in Table 3. Each column distinguishes between the different performance measures or combinations of measures (plots show a random sample of up to 1,000 points from the total number of behavioural parameter sets listed in Table 4)

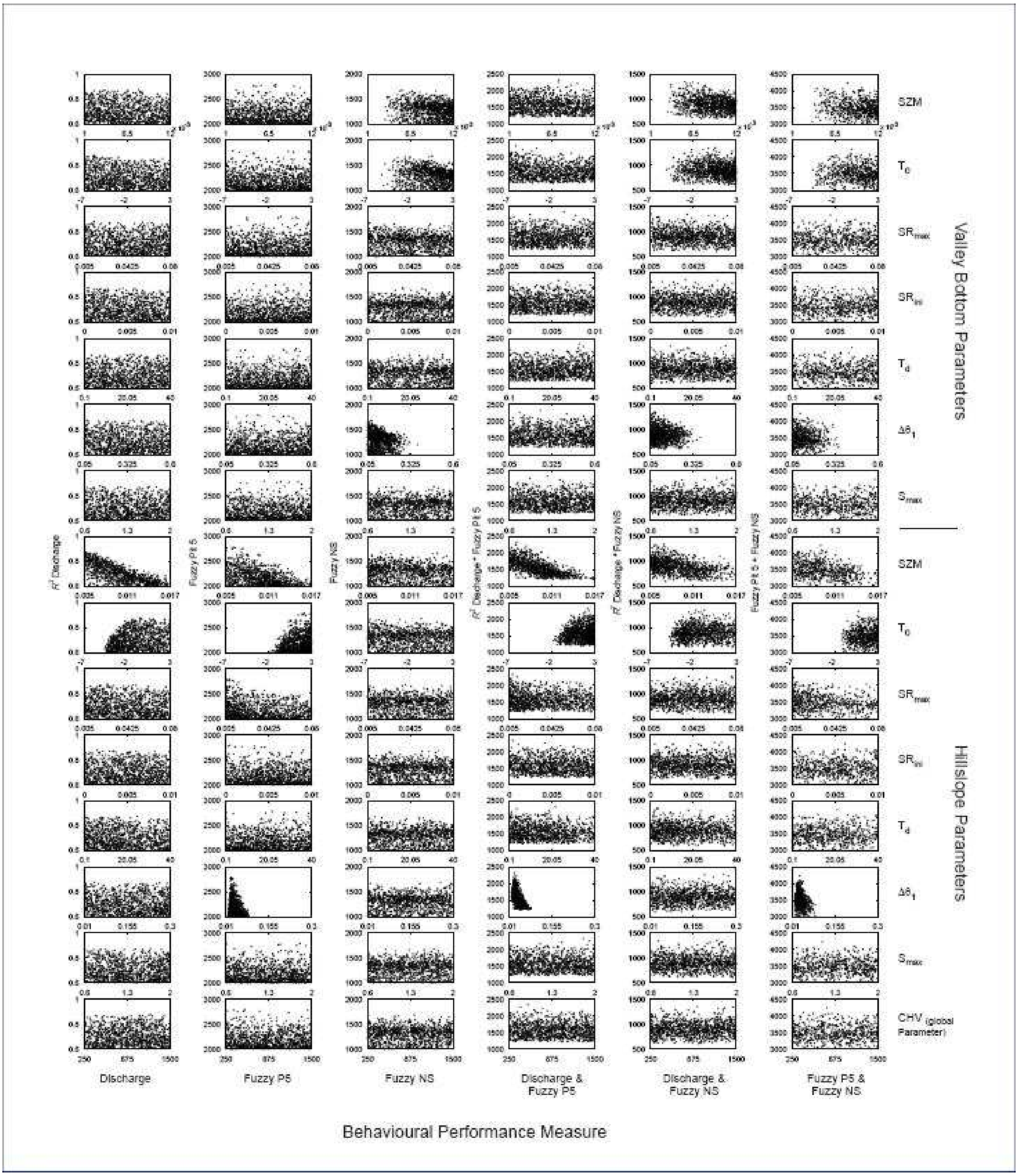


Figure 8: Dotty plots and histograms of behavioural parameter distributions from run_2 for both the VB_{LU} and the HS_{LU} for parameter sets that were classed as behavioural for all three performance measures listed in Table 3. The main matrix of dotty plots shows the correlation between pairs of parameters within the same LU and between the HS_{LU} and VB_{LU} LU's (the greyed area). Each histogram shows the distribution of behavioural parameters within each parameter range (note the range limits are shown for the run_1 limits).

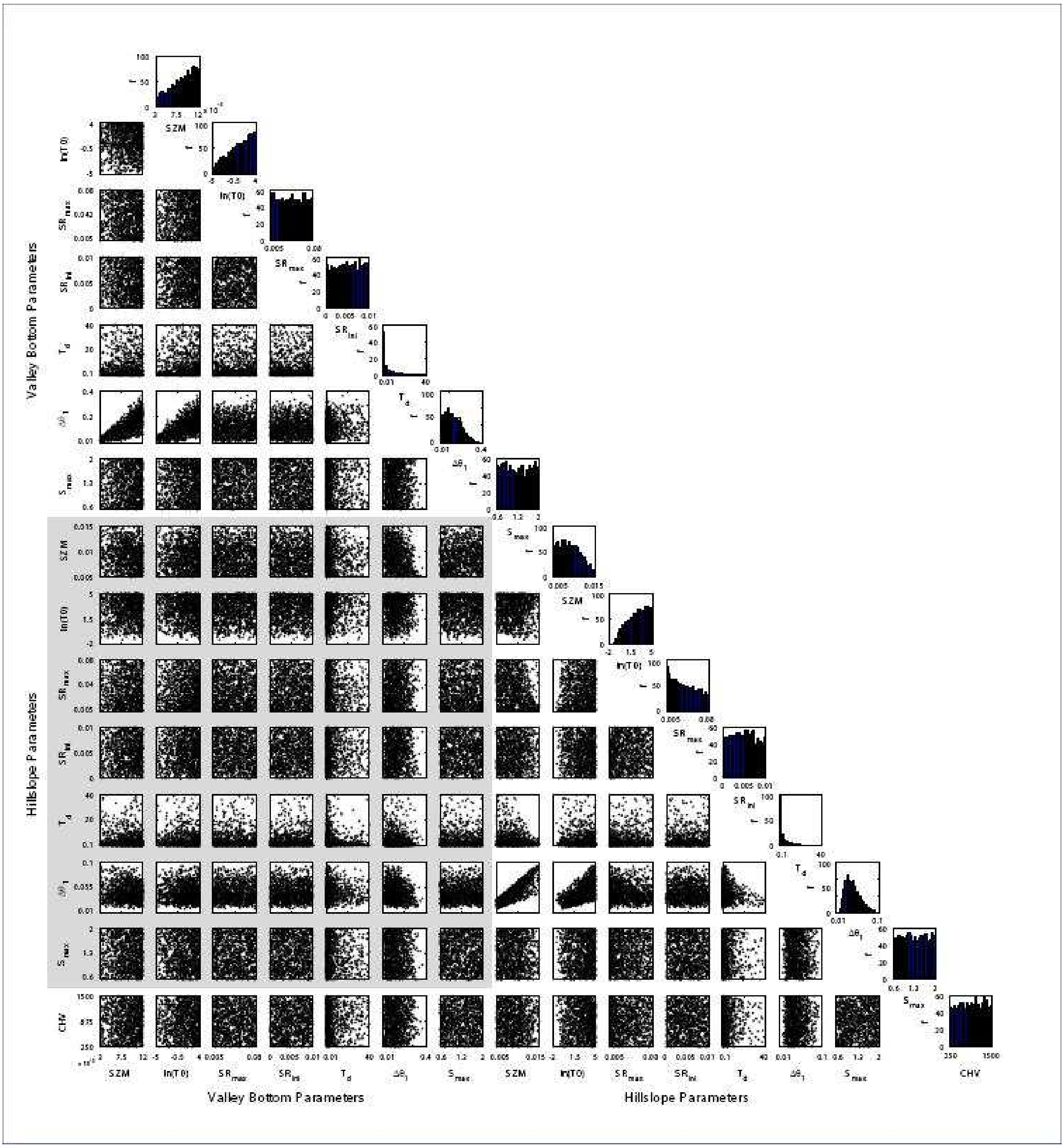


Figure 9: GLUE Discharge, NS ∇_{wt} and P5 ∇_{wt} updated behavioural possibility bounds for a) Discharge, b) ln(Discharge), C) P5 ∇_{wt} and D) NS ∇_{wt} simulations.

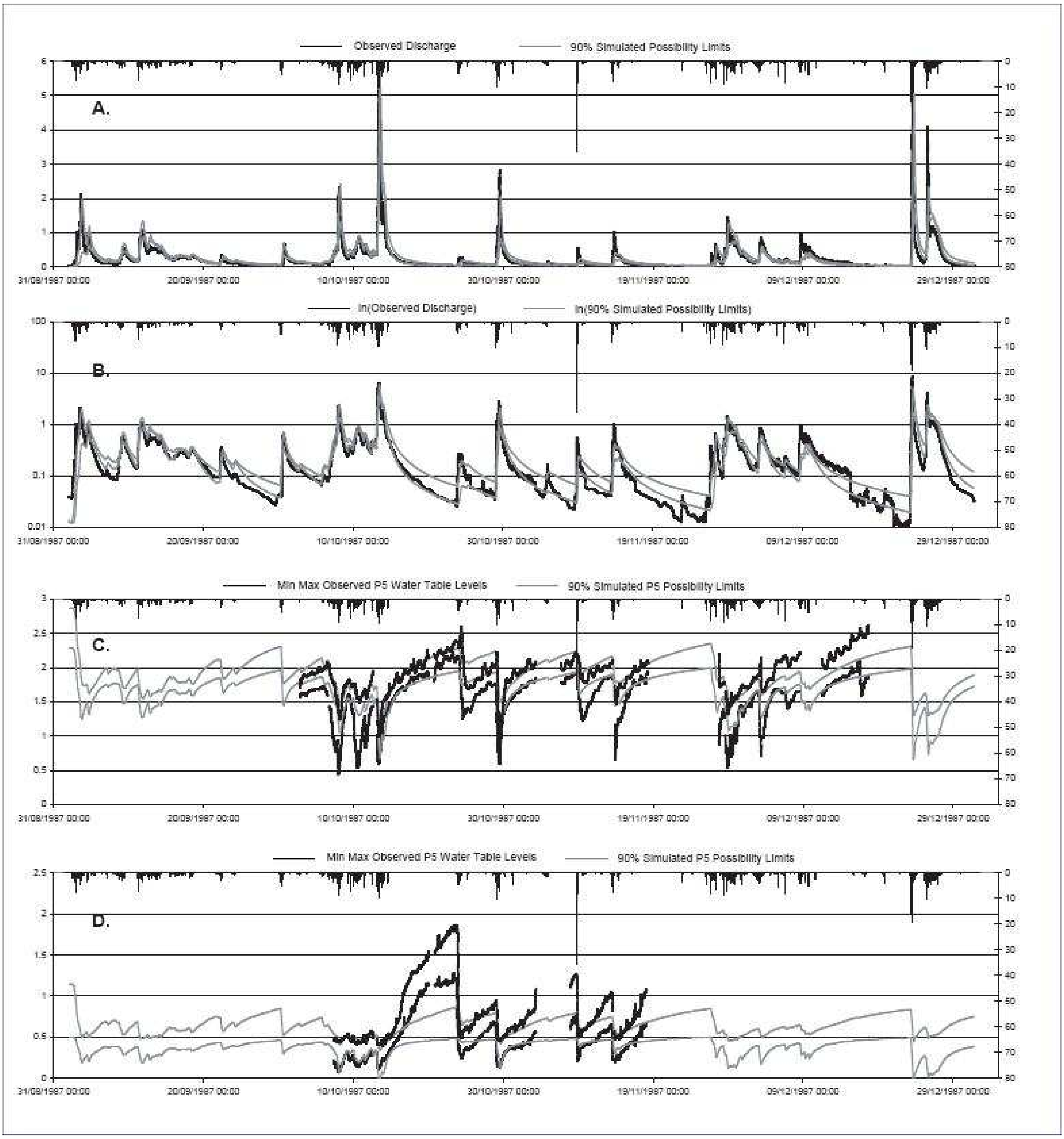
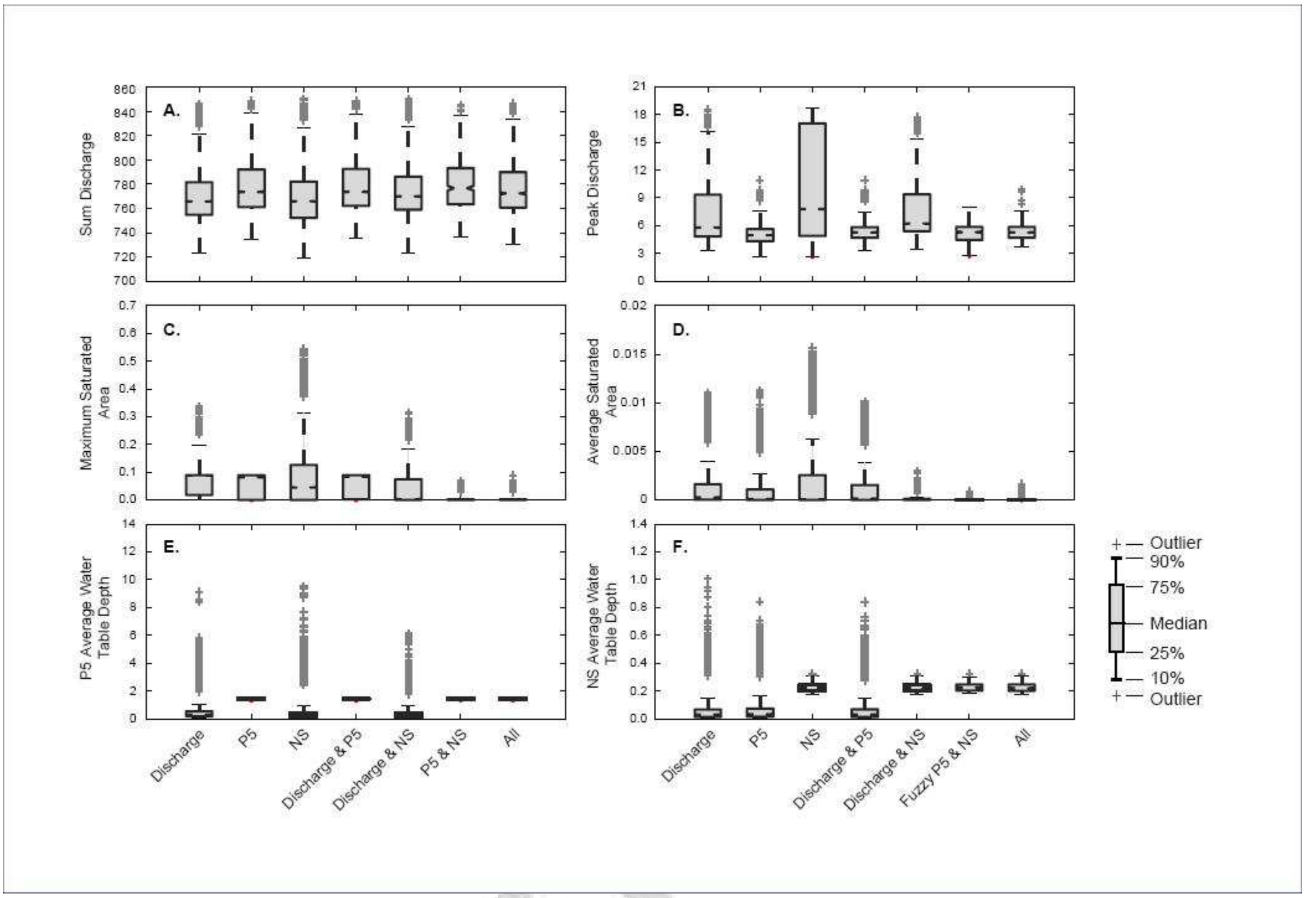


Figure 10: Distributions of summary model responses for behavioural simulations using different PM's or combinations of PM's listed in Table 4



DRAFT