Global NEST Journal, Vol 14, No 3, pp 325-334, 2012

USING AN INFORMATIONAL ENTROPY-BASED METRIC AS A DIAGNOSTIC OF FLOW DURATION TO DRIVE MODEL PARAMETER IDENTIFICATION

I.G. PECHLIVANIDIS[1,*]
B.M. JACKSON[1]
H.K. MCMILLAN[2]
H.V. GUPTA[3]

[1] School of Geography, Environment and Earth Sciences
Victoria University of Wellington, Kelburn, Wellington, New Zealand
[2] National Institute for Water and Atmospheric Research (NIWA), Riccarton, Christchurch, New Zealand
[3] Department of Hydrology and Water Resources, University of Arizona Tucson, Arizona, USA

Received: 03/05/12
Accepted: 25/07/12
*to whom all correspondence should be addressed:
e-mail: ilias.pechlivanidis@gmail.com

ABSTRACT
Calibration of rainfall-runoff models is made complicated by uncertainties in data, and by the arbitrary emphasis placed on various magnitudes of the model residuals by most traditional measures of fit. Current research highlights the importance of driving model identification by assimilating information from the data. In this paper, we evaluate the potential use of an entropy-based measure as an objective function or as a model diagnostic in hydrological modelling, with particular interest in providing an appropriate quantitative measure of fit to the flow duration curve (FDC). The proposed Conditioned Entropy Difference (CED) metric is capable of characterising the information in the flow frequency distribution and thereby constrain the model calibration to respect this distributional information. Four years of hourly data from the 46.6 km^2 Mahurangi catchment, NZ, are used to calibrate the 6-parameter Probability Distributed Moisture model. Results are analysed using three measures: the proposed entropy-based measure, the Nash-Sutcliffe (NSE), and the recently proposed Kling-Gupta efficiency (KGE). We also examine a conditioned entropy metric that trades-off and reweights different segments of the FDC to drive model calibration in a way that is based on modelling objectives.

Overall, the entropy-based measure results in good performance in terms of NSE but poor performance in terms of KGE. This entropy measure is strongly sensitive to the shape of the flow distribution and is, from some viewpoints, the single best descriptor of the FDC. By conditioning entropy to respect multiple segments of the FDC, we can reweight entropy to respect those parts of the flow distribution of most interest to the modelling application. This approach constrains the behavioural parameter space so as to better identify parameters that represent both the “fast” and “slow” runoff processes. Use of this importance-weighted, conditioned entropy metric can constrain high flow predictions equally well as the NSE and KGE, while simultaneously providing well-constrained low flow predictions that the NSE or KGE are unable to achieve.

KEYWORDS: calibration, model identification, flow duration curve, performance measures, diagnostics, Kling-Gupta efficiency, conditioned entropy difference.

INTRODUCTION
Hydrological model identification is usually driven by measures of fit, which provide an objective assessment of the agreement between observed and simulated hydrological data (e.g. streamflow) (Beven, 2001). Most traditional measures are a function of the residuals in the modelled and measured quantities, and emphasise different systematic and/or dynamic behaviours within the hydrological system (Pechlivanidis *et al.*, 2011). As a result, a robust assessment of model identification and performance using traditional and/or single measures is difficult (Krause *et al.*, 2005). Recent studies urge the need for robust diagnostic model evaluation, which aims to: 1)

determine the information contained in the data and in the model, 2) examine the extent to which a model can be reconciled with observations, and 3) point towards the aspects of the model that need improvement (Gupta *et al.*, 2008; Schaefli *et al.*, 2011).

Catchment signatures have recently been used to diagnose model inadequacies (Yilmaz *et al.*, 2008; Winsemius *et al.*, 2009). A signature based calibration approach seeks to identify hydrologically meaningful patterns of system behaviour in the data (i.e. pattern of runoff coefficient, monthly runoff variation, recession curves), and to calibrate the model to these signatures (Hingray *et al.*, 2010). The flow duration curve (FDC) is widely considered to be a hydrologically informative signature of catchment behaviour, because it provides a simple, yet comprehensive, representation of the historical distribution of flow variability. FDCs have been used in a wide range of hydrological applications including model calibration (see among others, Westerberg *et al.*, 2010). However, these recently developed FDC calibration methods divide the FDC into segments to represent different flow ranges (i.e. from 30 to 70 percentile flows) and either construct objective measures (i.e. volumetric bias, error in the slope) for each segment (van Werkhoven *et al.*, 2008) or select evaluation points along the FDC and thereby evaluate the deviation between simulated and observed discharge. Therefore, current FDC calibration methods have a limited ability to extract the distributional information contained in the probability density function (pdf) of the flow signal.

Recent work has proposed the concept of a diagnostic evaluation approach rooted in information theory (Gupta *et al.*, 2008; Weijs *et al.*, 2010). Information theory provides powerful tools, which make no assumptions about the underlying system dynamics or relationships among the system variables (e.g. they capture any-order correlations among the time series). These tools provide a promising avenue to better identify where information is present and/or conflicting, and to better diagnose model/data/ hypotheses inconsistencies (Weijs *et al.*, 2010). Information theoretic computations ultimately rely on quantities such as entropy, which has drawn the scientific community's attention in a range of problems in hydrology and water resources (see, for example, review by Singh (2000)). However the potential of information entropy measures to serve as objective functions (OFs) and the uses of entropy in conjunction with other measures as diagnostics in hydrological modelling are still unexplored. In this paper, we extend the work of Pechlivanidis *et al.* (2010a) presenting an entropy measure suited to capturing the static (non-dynamical) information contained in streamflow signals as described by the probability distribution (and hence of the FDC).

The paper is organised as follows. Entropy-based statistics are introduced in Section 2, where we present an approach to estimate entropy for streamflow series. In Section 3, the study area and data are introduced. Section 4 describes the rainfall-runoff model and the identification method followed. Section 5 presents results consisting of statistical analysis based on observed and modelled data, which use entropy as an objective function. Finally, Section 6 states the conclusions and discusses on possible ways forward.

USE OF INFORMATIONAL ENTROPY MEASURE AS A MODEL DIAGNOSTIC

Schreiber (2000) stated that information is equivalent to the removal of uncertainty; hence uncertainty and informational entropy are in some senses identical. Entropy (in Greek εντροπία, etymologized from τροπή, i.e. change, turn, drift) has been variably described; examples include "a measure of the amount of chaos" or "of the lack of information about the system" (Koutsoyiannis, 2005).

Informational Shannon entropy

Treating each streamflow observation as a discrete non-negative random variable X, the Shannon entropy can be formulated as (Shannon, 1948):

$$H_X = E[-\log_2 p(X)] = -\sum_{i=1}^{N} p(x_i)\cdot\log_2 p(x_i)$$

where $E[]$ denotes expected value, $p(x_i)$ is the probability of occurrence of outcome x_i such that the probabilities sum to 1, and N is the number of possible outcomes. In this paper the base is 2, in which case entropy estimates the average number of binary digits (bits) needed to optimally encode independent draws of X following a probability distribution $p(x_i)$ (Papoulis, 1991). A low value of entropy indicates a high degree of structure and a low uncertainty. It can be easily shown that with complete information entropy equals 0, otherwise it is greater than 0. If no information is available then entropy will reach its maximum equal to $log_2(N)$. When used as a model diagnostic, we suggest normalisation with respect to the maximum entropy value, where all states are equally probable, i.e., $H_X = log_2(N)$. This normalisation eliminates differences in entropy caused by the number of possible outcomes. Hence, the normalised entropy remains 0 with complete information / maximum order and takes a maximum value of 1 with minimal structure / maximum disorder.

Although a continuous analogue to the Shannon entropy is available, we rarely possess the analytical form of our variable *X*'s probability distribution, and so must generally work with the discrete form presented earlier. Unless *X* is ordinal, a number of discrete bins must be specified with accompanying ranges. In this case, the estimation of the probability distribution and its associated entropy is influenced by the resolution of this data, the number of bins, and the locations of divisions between these bins. The introduction of arbitrary partitions can result in “edge effects”. According to Ruddell and Kumar (2009), with too few/many partitions, “edge effects” become severe and entropy estimates are positively biased. Different approaches can be used to discretise the data set into probability bins. These include function fitting (Knuth *et al.*, 2005), kernel estimation (Nichols, 2006), and binning with fixed mass (e.g. equal probable bins) or fixed width (e.g. linear bins) interval partitions (Ruddell and Kumar, 2009). In this paper, we use a hybrid fixed width-mass interval approach. The hybrid fixed width/fixed mass interval approach is a result of preliminary analysis suggesting the sampling and error characteristics at low flows are most suited to the fixed mass approach, while conversely, medium and high flows are better suited to a fixed width interval approach. The FDC is split into multiple segments; this forces our measures to respect entropy characteristics in each segment rather than merely those parts of the FDC that are most sampled. For most high temporal resolution applications, this effectively reweights the entropy estimation to respect the entire flow range; the more sparsely sampled medium and high flow characteristics as well as the highly sampled low flow characteristics.

Definition of proposed entropy difference metric

Although Shannon entropy is a quantification of the distribution of values within a dataset, its static probabilistic nature cannot characterise the temporal structure of information. It therefore shows no sensitivity to differences in timing. In addition, this measure is not usually discretised to depend on the range of the data, so mass balance errors can be introduced. In this study we propose an estimate of entropy suitable for hydrological applications, based on trading off the unscaled and scaled Shannon entropy difference, SUSE, defined as:

$$SUSE = \max[\,abs(H^U_{sim} - H^U_{obs}),\, abs(H^S_{sim} - H^S_{obs})]$$

where H^U is the un-scaled entropy using different bin ranges for simulated and observed data based on their individual specific maximum range (this measure respects shape conservation irrespective of mass/scaling), and H^S is the scaled entropy using identical bins for both simulated and observed data (it attempts to conserve mass and shape).

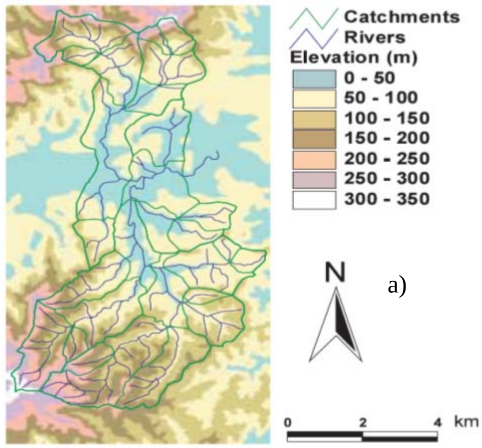
To reweight different segments of the FDC so as to better characterise the information in the FDC we use an importance-weighted (conditioned) entropy metric (CED; Conditioned Entropy Difference metric), wherein we subjectively partition the curve into different segments (Figure 1b): i) high (<2% probability of exceedance) flow segment characterising the catchment response to large precipitation events, ii) medium and intermediate (2-20% and 20-70% respectively) flow segments characterising the catchment response to moderate size precipitation events and also relating to the intermediate-term primary and secondary baseflow relaxation response of the catchment, and iii) low (>70%) flow segments relating to the long-term sustainability of flow; note that these partitions can be changed to accord with the specific requirements of an application. SUSE is estimated for each segment and eventually CED is defined as the maximum entropy difference present in the different FDC segments, given by:

$$CED = \max[(SUSE_{0-sg_1}),\ldots,(SUSE_{sg_m-100})]$$

where *m* is the number of the FDC segments, and *sg* is the probability of exceedance for each segment of the FDC. Linear binning (in this case 150 bins) was used to characterise the information in the high, medium and intermediate flow segments, whereas equally-probable binning (in this case 60 bins) was used to characterise information in the low flow segments. The number of bins was selected using the algorithm by Knuth (2005).

STUDY SITE AND DATA DESCRIPTION

The analysis is based on observed data from the experimental Mahurangi River (Figure 1a) in northern New Zealand, which drains 46.6 km^2 of steep hills and gently rolling lowlands. The Mahurangi River Variability Experiment, MARVEX, ran from 1997-2001, and investigated the space-time variability of the catchment water balance. A network of 28 flow gauges and 13 rain gauges has been installed, collecting records at 15 minutes intervals as part of the MARVEX project (Woods, 2004).



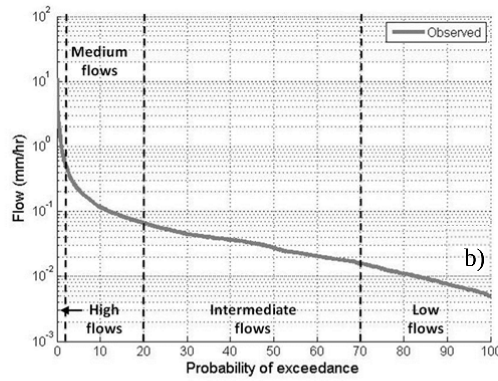


Figure 1. a) The Mahurangi River catchment, and b) the catchment's flow duration curve (Figure 1a reproduced from Woods (2004))

The catchment experiences a warm humid climate (frosts are rare and snow and ice are unknown), with mean annual rainfall and evaporation of 1,600 and 1,310 mm respectively. The catchment elevation ranges from sea level to 300 m. Most of the soils in the catchment are clay loams, no more than a metre deep, while much of the lowland area is used for grazing. Plantation forestry occupies most of the hills in the south, and a mixture of native forest, scrub and grazing occurs on the hills in the north. Further details are given in Woods (2004). Historical rainfall, streamflow and potential evapotranspiration data at hourly time steps were provided by the National Institute of Water and Atmospheric Research, New Zealand, for the period 1998-2001. The arithmetic average of the 13 rain gauge records was used as the mean areal precipitation and was distributed uniformly over the catchment. Only the flow gauge at the outlet of the catchment was considered in the present study. Its flow duration curve is presented in Figure 1b.

MODEL IDENTIFICATION

Model description

The Probability Distributed Moisture (PDM) model is a conceptual model, which uses a distribution of soil moisture storage capacities for soil moisture accounting and, in this application and most others, two linear reservoirs in parallel for the routing component (Moore, 2007; Pechlivanidis *et al.*, 2010b) (Figure 2).

The soil moisture storage capacity, *C* (mm), is assumed to be described by a Pareto distribution having the following function:

$$F(C) = 1 - (1 - C / Cmax)^b$$

where *C* is the storage capacity in the catchment, *Cmax* is the maximum capacity at any point in the catchment, and the parameter *b* (-) controls the spatial variability of storage capacity over the catchment. Within each time step, the soil moisture storage is depleted by evaporation as a linear function of the potential rate and the volume in storage, and augmented by rainfall. Effective rainfall is then equal to the soil moisture excess.

The effective rainfall is split into "quick" and "slow" pathways, which are routed via parallel storage components. The parameter *q* defines the proportion of total effective rainfall going to the fast response reservoir. The simulated streamflow is determined by the combination of the two pathways. This model component has three parameters: a residence time for each reservoir, *Kq* and *Ks* (hours) and *q* (-). The total streamflow is finally delayed by a parameter *T* (hours) to adjust the time to peak response.

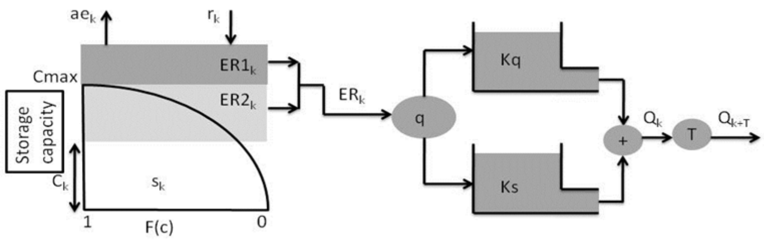


Figure 2. Structure of the Probability Distributed Moisture model

Selection of the measures of fit

A Monte Carlo uniform random search was used to explore the feasible parameter space (Table 1) and to investigate parameter identifiability (50,000 samples). The first year (1998) was used as a model warm-up period, the next two years for model calibration (1999-2000) and the final year for independent performance evaluation (2001). The PDM was calibrated using streamflow data at the catchment outlet using three OFs: the SUSE, the Nash-Sutcliffe Efficiency, (NSE), and the Kling-Gupta Efficiency (KGE). KGE has recently been proposed by Gupta *et al.* (2009) to reduce calibration problems that arise when using the widely applied NSE (Nash and Sutcliffe, 1970) or the un-normalized mean square error criterion. NSE and KGE are defined as:

$$NSE = 1 - \frac{\sum_{i=1}^{n} (Qobs_i - Qsim_i)^2}{\sum_{i=1}^{n} (Qobs_i - Q\overline{obs})^2}$$

$$KGE = 1 - \sqrt{(cc-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2}$$

where *Qsim* is the calculated flow, *Qobs* is the observed flow, *n* is the length of the time series, *cc* is the linear cross-correlation coefficient between *Qobs* and *Qsim*, α is a measure of variability in the data values (equal to the standard deviation of *Qsim* over the standard deviation of *Qobs)*, and β is equal to the mean of *Qsim* over the mean of *Qobs*. In optimisation NSE and KGE are subject to maximisation with an ideal value at unity. Another reason for our choice of KGE is that this measure sees the calibration problem from a multi-objective perspective, by focusing on the correlation, variability error and bias error as separate criteria to be optimised (see Gupta *et al.* (2009) for further details of the KGE and its components).

As explained earlier, the entropy measure is insensitive to timing errors and hence the corresponding simulated runoff is not sensitive to the final routing delay parameter. To overcome this, the routing parameter *T* was individually adjusted through manual calibration (*T* is equal to 2 hours).

Table 1. Prior range of PDM model parameters

	Parameter	Range
Cmax	Maximum storage capacity (mm)	0 – 1000
b	Shape of Pareto distribution (-)	0 – 2
Kq	Time constant for quick flow reservoir (hours)	0 – 300
Ks	Time constant for slow flow reservoir (hours)	3000 – 7000
q	Fraction of flow through quick flow reservoir (-)	0 – 1
T	Time delay of channel routing (hours)	0 – 10

Identification of behavioural model parameter sets, using the conditioned entropy difference measure (CED), was based on simultaneous satisfaction of a criterion for each segment of the FDC. A threshold value was used to condition and identify these sets.

Method for model evaluation

We first investigate the potential of each objective function to enable a successful calibration of the PDM model. Both the NSE and KGE are used as benchmarks; the former has been widely applied in hydrology as a benchmark measure of fit, while KGE has recently been proposed to reduce NSE-related calibration problems. We next investigate the potential of the entropy-based metric to select model calibrations which more accurately reproduce the observed streamflow time series. Model evaluation is also based on model's ability to reproduce FDC related signature measures which according to Yilmaz *et al.* (2008) are related to the "fast" and "slow" runoff processes. We examined biases towards specific properties of the FDC using the following diagnostic signature measures: the percent bias in FDC high-segment volume (High volume bias (%)), the percent bias in FDC intermediate-segment slope (Intermediate slope bias (%)), and the percent bias in FDC low-segment volume (Low volume bias (%)). In addition, we apply a signature measure using the median log flow as an index of intermediate flow behaviour (Midflow bias (%)).

We next investigate the parameter sensitivity on the performance measure and how the parameter space is constrained using the CED; at this stage multiple thresholds for the four segments of the FDC are used to reweight the importance of each segment (entropy differences are less than 0.15, 0.11, 0.11, and 0.11 for the low, intermediate, medium and high flow segment of the FDC respectively). To achieve this, we present the solution space using NSE only, and then how the parameter ranges are reduced when using CED. Finally, we investigate the improvement in our model simulations based on their ability to reproduce the observed streamflow using the reduced parameter ranges.

RESULTS

Benchmark comparison of entropy-based metric with NSE and KGE

The first step is to establish the potential of each objective function to achieve a successful calibration of the model (Table 2a). The first column indicates the criterion used for model calibration. The next sets of columns indicate the calibration and evaluation period values achieved for each of the three performance measures by the calibrated parameter set so obtained. The second step is to evaluate the model simulations (streamflow is simulated based on the behavioural model parameter sets) based on their ability to: 1) reproduce the observed streamflow time series, and 2) minimise the biases in different ranges of the FDC (Table 2b).

Table 2. a) Model performance using the three objective functions, and b) absolute biases for each segment of the FDC

a)	NSE		KGE		SUSE	
	Cal.	Eval.	Cal.	Eval.	Cal.	Eval.
NSE	0.86	0.80	0.90	0.89	0.03	0.06
KGE	0.84	0.80	0.92	0.90	0.02	0.07
SUSE	0.79	0.72	0.67	0.68	0.00	0.04

b)		NSE	KGE	SUSE
High volume bias (%)	Mean	14.97	12.13	20.55
	Min	0.06	0.06	0.02
	Max	41.60	40.43	97.68
Intermediate slope bias (%)	Mean	42.52	40.13	27.25
	Min	1.02	1.21	0.18
	Max	133.60	175.21	85.93
Low volume bias (%)	Mean	30.59	29.71	22.79
	Min	0.25	0.25	0.03
	Max	60.37	60.37	118.14
Midflow bias (%)	Mean	13.02	15.50	9.49
	Min	0.06	0.06	0.01
	Max	82.68	66.29	32.01

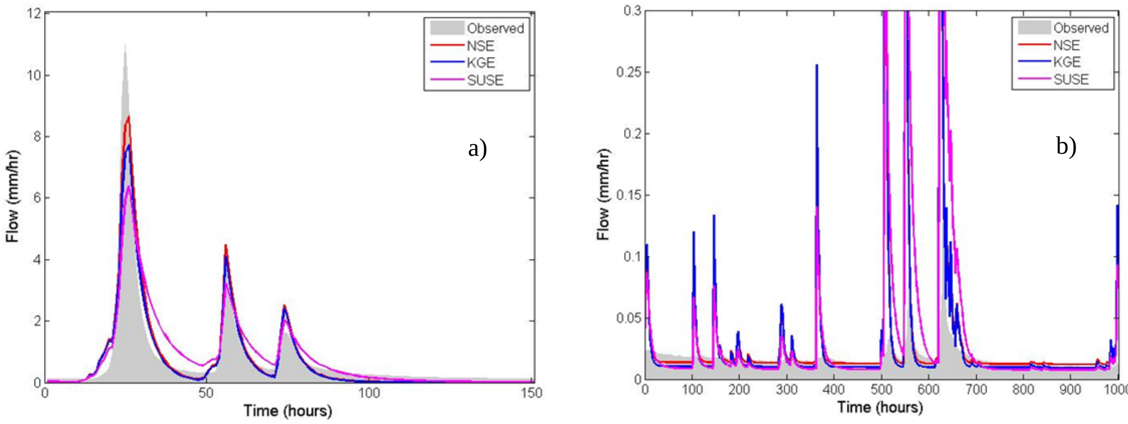


Figure 3. Simulated streamflow using the three OFs during: a) high, and b) low flow periods

Overall, good performance of the model in terms of NSE and KGE was achieved for both calibration and evaluation periods (see Table 2a). Calibration using entropy provided acceptable values for NSE (higher than 0.7 for both calibration and validation periods; NSE values greater than 0.7 are generally considered acceptable in many hydrological applications (Freer *et al.*, 2004)), but not for KGE (0.67 and 0.68 for calibration and evaluation periods respectively), highlighting the limitation of SUSE as a stand-alone objective function. Table 2b shows that calibrations using NSE and KGE are able to represent adequately the high flow range of the FDC, indicating their suitability for flood prediction applications; however they introduce significant bias in the other segments of the FDC. While SUSE distributes its weight equally towards all aspects of the FDC, it is unable to perform as well as the other two OFs in matching the high flow segment. Similar conclusions can be drawn from Figure 3, which shows the model fit using the three OFs during a high and low flow period. Both NSE and KGE tend to better fit the highest flow event (10.7 mm hr^{-1}) than SUSE; however, they overestimate the other two peaks. In contrast, fitting using NSE and KGE during low flow periods is poor; while fitting baseflow is improved using SUSE (see also low flow volume bias in Table 2b).

Parameter identifiability to performance measure

We further investigate how the parameter space is constrained using CED compared to the single NSE objective. To achieve this, we present the solution space using NSE only, and then how the parameter ranges are “reduced” when using CED. Figure 4 presents the solution space using the conditioned entropy metric (CED; identified parameter sets have SUSE value less than 0.11 for each segment of the FDC) against the behavioural sets in terms of NSE (NSE>0.7; identifying 811 behavioural model parameter sets) for each model parameter.

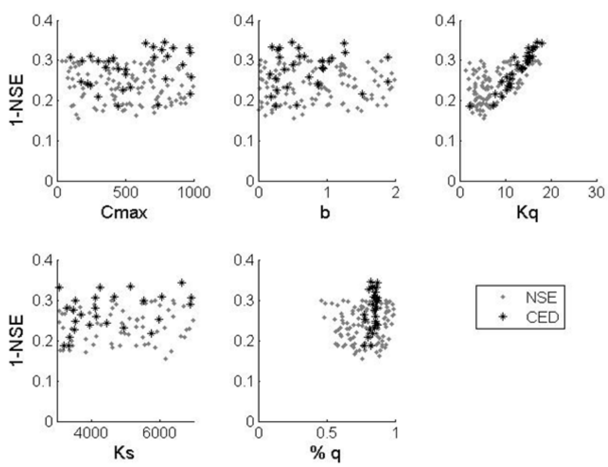


Figure 4. The effects of conditioning entropy on parameter identification

A narrower parameter space is achieved when CED is used. *Cmax* and *b* parameters are poorly identifiable using the conditioned metrics (and also relatively insensitive to many other objective functions not reported in this study). High identifiability is observed for *Kq* and *q*. Conditioning the solution space using the CED and threshold value equal to 0.11 for each segment, the average NSE and KGE values are 0.66 and 0.67 respectively (varying between 0.52 - 0.81 and 0.52 - 0.82 respectively).

Model ability to reproduce streamflow series: CED using multiple thresholds

To investigate the potential of the CED metric, streamflow is simulated based on the behavioural model parameter sets selected using this measure, and compared to those selected using NSE and KGE. Assigning a different threshold value to each segment of the FDC allows reweighting of the entropy measure based on the modelling objectives. For instance, in this study, the model is not capable of representing every segment of the FDC to similar precision; a trade off exists between fitting high and low flow values. To demonstrate the importance of choosing threshold values appropriately, Figure 5 presents the simulated runoff for high and low flow periods using 0.15, 0.1, 0.1, and 0.1 as thresholds for the flow segments (low - high). The range of simulated runoff during the high flow period is similar when using the NSE, KGE and conditioned parameter sets. There is only a slight overestimation of the peaks when the KGE is used. Both NSE and KGE are relatively insensitive to the low flows, with considerable overestimates and underestimates in some cases (Figure 5b). The potential of the conditioned entropy difference measure to capture detail of the "slow" runoff processes is illustrated in Figure 5b, since the envelope of simulated runoff is very close to the baseflow.

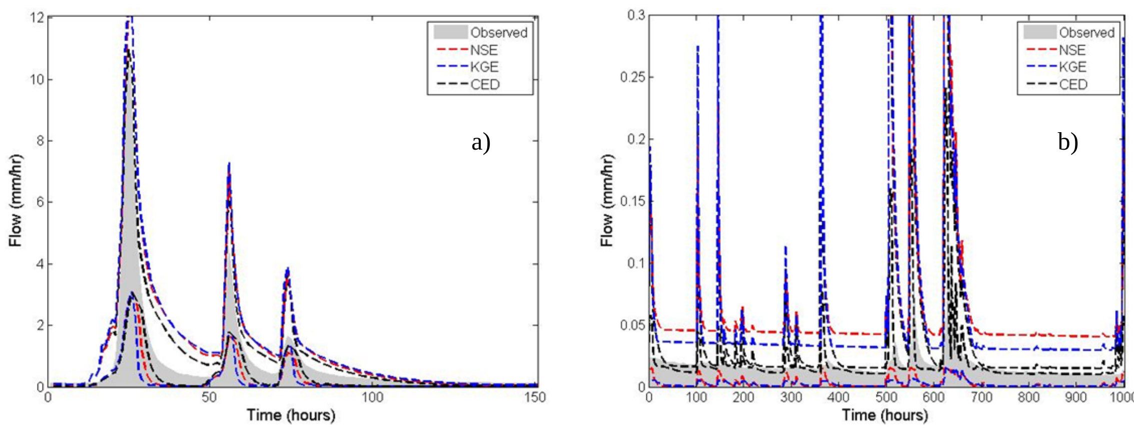


Figure 5. Simulated streamflow during high (a) and low (b) flow periods using the behavioural parameter sets based on the NSE, KGE and CED

CONCLUSIONS

In this paper, we have explored the potential use of informational entropy-based measures in hydrological modelling with particular interest in extracting the distributional structure of the flow time series, as described by the flow duration curve (FDC). The PDM rainfall-runoff model was calibrated using the NSE and KGE objective functions, and our new proposed Shannon entropy-based measure. Overall, results support our theoretical observations that the probabilistic structure of the Shannon entropy measure is strongly related to the FDC, while our proposed estimation of entropy is capable of characterising information of interest in the probability distribution of flow. This metric uses equally probable bins at the low segment of the FDC and linear bins at the intermediate, medium and high segments. In this first application in the Mahurangi catchment, our metric outperforms the NSE and KGE at medium, intermediate and low flows; as might be expected however, both NSE and KGE achieve better performance at the high flow segment of the FDC. This is likely due, at least in part, to entropy's statistical nature; fundamentally the highest flow values will be under-sampled and hence not statistically robust; entropy considers this as possessing negligible information.

Bias towards different aspects of the FDC can be overcome, to a certain extent, by using an importance-weighted, conditioned entropy measure. This overcomes the tendency of entropy to emphasise information within the low (most-sampled) flows by estimating the entropy over multiple segments of the FDC and setting criteria for each individual metric value to identify acceptable parameter sets. These criteria could be a single threshold or a set of entropy thresholds for each segment of the FDC. The latter approach seems more generally applicable in applications where data or model structural errors lead to a trade-off between, for example, the high and low FDC segment, and where the user is more concerned with specific regions of the FDC rather than overall performance.

REFERENCES

Beven K. (2001), *Rainfall-Runoff modelling. The Primer*, 1-360 pp., John Wiley and Sons, Chichester, UK.

Freer J.E., McMillan H., McDonnell J.J. and Beven K.J. (2004), Constraining dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures, *Journal of Hydrology*, **291**(3-4), 254-277.

Gupta H., Wagener T. and Liu Y. (2008), Reconciling theory with observations: elements of a diagnostic approach to model evaluation, *Hydrological Processes*, **22**(18), 3802-3813.

Gupta H.V., Kling H., Yilmaz K.K. and Martinez G.F. (2009), Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *Journal of Hydrology,* **377**(1-2), 80-91.

Hingray, B., Schaefli, B., Mezghani, A. and Hamdi, Y. (2010), Signature-based model calibration for hydrological prediction in mesoscale Alpine catchments, *Hydrological Sciences Journal*, **55**(6), 1002-1016.

Knuth K.H. (2005), Lattice duality: The origin of probability and entropy, *Neurocomputing*, **67**, 245-274.

Knuth K.H., Gotera A., Curry C.T., Huyser K.A., Wheeler K.R., Rossow W.B. (2005), Revealing relationships among relevant climate variables with information theory. In Proceedings of the *Earth-Sun System Technology Conference* (ESTC 2005), July, Adelphi, MD.

Koutsoyiannis D. (2005), Uncertainty, entropy, scaling and hydrological statistics. 1. Marginal distributional properties of hydrological processes and state scaling, *Hydrological Sciences Journal*, **50**(3), 381-404.

Krause P., Boyle D.P. and Base F. (2005), Comparison of different efficiency criteria for hydrological model assessment, *Advances in Geosciences*, **5**, 89-97.

Moore R.J. (2007), The PDM rainfall-runoff model, *Hydrol. Earth Syst. Sci.*, **11**(1), 483-499.

Nash J.E. and Sutcliffe J.V. (1970), River flow forecasting through conceptual models, *Journal of Hydrology*, **10**, 282-290.

Nichols J.M. (2006), Examining structural dynamics using information flow, *Probabilistic Engineering Mechanics*, **21**(4), 420-433.

Papoulis, A. (1991), *Probability, Random Variables and Stochastic Processes*, 3rd edition, McGraw Hill.

Pechlivanidis, I.G., Jackson, B. and McMillan, H. (2010a), The use of entropy as a model diagnostic in rainfall-runoff modelling, *iEMSs 2010:* International Congress on Environmental Modelling and Software, 5-8 July, Ottawa, Canada, **2**, 1780-1787.

Pechlivanidis I.G., Jackson B., McIntyre N. and Wheater H.S. (2011), Catchment scale hydrological modelling: A review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications. *Global NEST Journal*, **13**(3)193-214.

Pechlivanidis I.G., McIntyre N.R. and Wheater H.S. (2010b), Calibration of the semi-distributed PDM rainfall-runoff model in the Upper Lee catchment, UK, *Journal of Hydrology*, **386**(1-4), 198-209.

Ruddell, B.L. and Kumar, P. (2009), Ecohydrologic process networks: 1. Identification, *Water Resources Research*, **45**, W03419.

Schaefli B., Harman C.J., Sivapalan M. and Schymanski S.J. (2011), HESS Opinions: Hydrologic predictions in a changing environment: behavioral modeling, *Hydrology and Earth System Sciences*, **15**, 635-646.

Schreiber T. (2000), Measuring Information Transfer, *Phys. Review Letters*, **85**(2), 461-464.

Shannon C.E. (1948), A mathematical theory of communication, *Bell System Technical Journal*, **27**, 379-423 (Part I) and 623-656 (Part II).

Singh V. (2000), The entropy theory as a tool for modelling and decision-making in environmental and water resources, *Water S.A.*, **26**(1), 1-12.

van Werkhoven K., Wagener T., Reed P. and Tang T. (2008), Characterization of watershed model behavior across a hydroclimatic gradient, *Water Resources Research*, **44**(W01429), 1-16.

Weijs, S.V., Schoups, G. and van de Giesen, N. (2010), Why hydrological predictions should be evaluated using information theory, *Hydrol.Earth Syst. Sci.*, **14**(12), 2545-2558.

Westerberg I.K., Guerrero J.-L., Younger P.M., Beven K.J., Seibert J., Halldin S., Freer J.E. and Xu C.-Y. (2010), Calibration of hydrological models using flow-duration curves, *Hydrology and Earth System Sciences*, **7**(6), 9467-9522.

Winsemius H.C., Schaefli B., Montanari A. and Savenije H.H.G. (2009), On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information, *Water Resources Research*, **45**(12), W12422.

Woods R. (2004), The impact of spatial scale on spatial variability in hydrologic response: experiments and ideas, *Scales in Hydrology and Water Management*, **287**, 1-15.

Yilmaz K.K., Gupta H.V. and Wagener T. (2008), A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrological model, *Water Resources Research*, **44**(W09417), 1-18.