DATA-BASED MECHANISTIC APPROACH TO MODELLING OF DAILY RAINFALL-FLOW RELATIONSHIP: A CASE OF THE UPPER VAAL WATER MANAGEMENT AREA

GM OCHIENG' AND FAO OTIENO

ABSTRACT

Although deterministic models still dominate hydrological modelling, there is a notable paradigm shift in catchment response modelling. An approach to represent the daily rainfall-flow (R-F) relationship using Data-Based Mechanistic (DBM) modelling is presented. DBM modelling is an inductive empirical transfer function (TF) approach relating input to output. The study used secondary data from the Department of Water Affairs and Forestry for the Upper Vaal water management area at station C1H007. The R-F model identification and optimisation was implemented in the CAPTAIN Toolbox in MATLAB. The best estimated R-F model was a 2nd order TF with an input lag of one day and $R_T^2 = 56\%$. In mechanistic interpretation, three parallel flow pathways were discerned; the fast flow, slow flow and the loss component each constituting 49.8%, 24% and 26.2% of the modelled flow respectively. The study demonstrates that the approach adopted herein produces reasonably satisfactory results with a minimum of the readily available catchment data.

Key words: Rainfall, flow, Data-Based Mechanistic Modelling

1. INTRODUCTION

The process of transformation of rainfall into river flow over a catchment is very complex, highly nonlinear, and exhibits both temporal and spatial variability (Rajurkar, Kothyari & Chaube, 2004). A plethora of models differing in their structure, application, data and technical requirements have been developed to simulate this process. These can be categorised as empirical black-box, conceptual, and physically based/deterministic distributed models (Merritt, Letcher & Jakeman, 2003; Rajurkar, Kothyari & Chaube, 2004). The latter category of models are predominantly utilised in hydrological modelling in most parts of the world (Young, 2001a; Hughes, 2004).

Although deterministic models still dominate hydrological modelling, there is a notable paradigm shift in catchment response modelling. This is evident in the growing application of simpler catchment models that perform equally well in representing key identifiable catchment hydrologic responses using readily available catchment data (for instance Young, 2001a, 2001b; Dye & Croke, 2003; Rajurkar, Kothyari & Chaube, 2004). The paradigm shift relates to the fact that studies have shown that:

- Simpler catchment models can perform equally well or may not be substantially out-performed by more complex models (Kokkonen, 2003; Merritt, Letcher & Jakeman, 2003);
- Unlike the complex models, simpler models facilitate systematic uncertainty

assessments (Kokkonen, 2003);

- simpler models are less demanding in data requirements and therefore suitable where only limited types of data are available (Evans & Jakeman, 1998; Kokkonen & Jakeman, 2001); and
- simpler models suffer minimal identifiability and over-parameterisation problems (Ye, Jakeman & Barnes, 1995; Kokkonen & Jakeman, 2001; Perrin, Michel & Andréassian, 2001; 2003; Beven, 2006).

Many situations in practice demand use of simple tools such as the linear system, theoretic models or black-box models. However, these simpler models normally fail to represent the nonlinear dynamics, which are inherent in the environmental processes such as the rainfall-flow transformation (Rajurkar, Kothyari & Chaube, 2004). An approach to represent the daily rainfall-flow relationship using DBM modelling is presented. DBM modelling is an inductive empirical TF approach relating input to output. DBM models differ from conventional black-box models in the sense that the resulting model is only considered acceptable if it can be interpreted in a physically meaningful manner (Young, 2001a, 2001b, 2003).

2. DBM MODELLING

Young and Beven (1994) have suggested a stochastic, Hybrid-Metric-Conceptual modelling approach which they call Data-Based Mechanistic (DBM) modelling. It derives its uniqueness from the following attributes (Young, 2001a, 2003; Pedregal, Taylor & Young, 2004):

- The inductive approach to model synthesis rather than basing the model development on *a priori* assumed conceptual model form. In this case, the model structure is inferred directly from the observed data in relation to a more general class of models. The resulting model is only considered credible if it can be interpreted in physically meaningful terms. This departs from the ordinary black-box models that reveal very little of their internal structure that has any physical meaning. This makes DBM models unique in the sense that unlike ordinary black-box models, the DBM modelling approach considers not only the fit of the data to the resulting model but also the achievement of physically sensible mechanistic interpretation of the resulting model. Hence the term “mechanistic” being used in describing the resulting models.
- The DBM modelling philosophy emphasises the importance of parametrically efficient, low order, dominant mode models. The importance of this dominant mode concept in model identification and estimation is for instance illustrated in Young (2001b) where it is shown how the response of a 26^{th} order hydrological simulation model can be duplicated with exceptional accuracy (0.0001% error by variance) by a much simpler 7^{th} order dominant mode model.
- DBM modelling philosophy also emphasises the development of stochastic methods and the associated statistical analysis required for the identification and estimation of such models.
- The DBM philosophy stresses the importance of explicitly acknowledging the basic uncertainty that is essential to any characterisation of physical, chemical, biological and socio-economic processes. The inherently stochastic nature of the DBM approach differentiates it from alternative

deterministic 'bottom-up' approaches. The inherently stochastic nature implies that the uncertainty in the estimated model is always quantified and this information can then be utilised in various ways, for instance:

(i) It allows for the application of Monte Carlo-based uncertainty and sensitivity analysis. The uncertainty analysis is particularly useful because it is able to evaluate how the covariance properties of the parameter estimates affect the probability distributions of physically meaningful, derived parameters, such as residence times and partition percentages in parallel hydrological pathways for example. This is an important attribute in the process of an informed decision making process as it gives an indication as to what extent to rely on the model predictions/parameters as derived from the stochastic relationship between the input and output variables.

(ii) It allows for the use of the model in statistical forecasting and data assimilation algorithms, such as the Kalman filter.

The DBM approach to modelling is widely applicable. It has been applied successfully to the characterisation of numerous environmental systems. Several examples as quoted in Young (2001a) include the development of *Aggregated Dead Zone* (ADZ) model for pollution transport and dispersion in rivers (Wallis, Young & Beven*,* 1989; Young, 1992); rainfall-flow modelling and forecasting (Young, 2001b; Bogner, Hingray & Musy, 2002); adaptive flood warning (Lees *et al.,* 1994; Young & Tomlin, 2000); and modelling of ecological and biological systems (Jarvis *et al*., 1999). The DBM approach has also been applied for control designs examples of which include the modelling and control of climate in glasshouses (Lees *et al*., 1996); forced ventilation in agricultural buildings (Price *et al*., 1999) and inter-urban road traffic systems (Taylor *et al*., 1998). They have also been applied in macro-economic modelling (Young & Pedregal, 1999).

2.1 Generic DBM model

The methodologies underpinning the DBM modelling approach are well documented (for instance Young & Beven, 1994; Young, 1998; Beven, 2001; Young, 2002; Young *et al.*, 2004). The general DBM model in TF terms for a single input single output takes the form:

$$y_t = \frac{B(z^{-1})}{A(z^{-1})} u_{t-\delta} + \xi_t \quad \text{...(1)}$$

where: is the measured output is the measured input. is a pure time delay, measured in sampling intervals, which is introduced to allow for any temporal delay that may occur between the incidence of a change in and its first effect on . The TF polynomials are defined as:

$$A(z^{-1}) = 1 + a_1 z^{-1} + a_2 z^{-2} + \ldots\ldots + a_n z^{-n}$$
$$B(z^{-1}) = b_0 + b_1 z^{-1} + b_2 z^{-2} + \ldots\ldots + b_m z^{-m} \quad \text{..(2)}$$

where: is the backshift operator (i.e.), and is the stochastic residual representing uncertainty in the relationship arising from a combination of measurement noise, the effect of other unmeasured inputs and modelling error. It is defined as follows:

$$\xi_t = y_t - \frac{\hat{B}(z^{-1})}{\hat{A}(z^{-1})} u_t = \hat{e}_t$$

and

$$\hat{e}_t \sim N(0, \hat{\sigma}_e^2) \qquad (3)$$

where: $\hat{A}(z^{-1})$ & $\hat{B}(z^{-1})$ are the estimates in Equation 2 and $\hat{\sigma}_e^2$ is the variance of the model residuals. The model order is defined by $\{n, m\}$ while the triad $\{n, m, \delta\}$ defines the structure of the model.

2.2 Generic transfer function rainfall-flow model

The generic model (Equation 1) applies directly for linear systems. However, as stated earlier, environmental processes such as transformation of rainfall into river flow are often complex and nonlinear. In DBM modelling, if the system is found to be nonlinear where changes in the parameters are functions of the state or input variables, then either the time variable parameter (TVP) or the more robust state dependent parameter (SDP) modelling is applied within the DBM modelling tenets to represent the system. An extension of these principles to R-F modelling over time has led to the development of a generic R-F model structure. Figure 1 presents such a structure emanating from studies by Jakeman and Hornberger (1993) based on earlier aspects of hydrologic research as well.

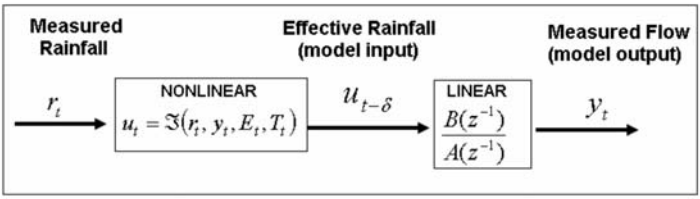

FIGURE 1: Block diagram of the generic transfer function rainfall-flow model (Young, 2001b)

Where r_t is the measured rainfall at time t, u_t is the effective rainfall, is the effective rainfall, $\Im(.)$ denotes an unknown nonlinear functional relationship (rainfall filter) defining the unobserved catchment storage state S_t considered as a function of potentially important variables that may affect or be related to catchment storage (Young, 2001b). Such variables may be rainfall (r_t) temperature (T_t), potential evapotranspiration (ET_t) and stream discharge (y_t) all of which could help to define the changes in soil moisture and storage if they are available. δ is any associated time delay in the system and $\frac{B(z^{-1})}{A(z^{-1})}$ is the transfer function relating the effective rainfall input to the measured flow. The SDP modelling technique is used to define $\Im(.)$ whereby the nonlinearity is located and characterised using both nonparametric and parametric tools. More information on SDP modelling can be obtained from Young (2000, 2001b) and Pedregal, Taylor and Young, (2004) and the references therein, for instance.

In this study, rainfall (r_t) evaporation (E_t) and stream flow (y_t) data were used to define the effective rainfall input (u_t) In this case, the flow data was used as an objectively identified surrogate for the catchment moisture storage. This was motivated by previous research findings (for instance Young, 2003) that have shown that there is a similarity in the pattern of temporal changes of catchment storage and streamflow with the later as a function of the former. The evaporation data was used as a descriptor of seasonality effect as well as an index of evapotranspiration in the catchment. The motivation to use the evaporation data was from the fact that apart from being a readily available and directly measurable variable, it is a direct contributor to primary nonlinearity between the occurrence of rainfall in the catchment and the subsequent increase of flow in the river. Further, from previous similar studies (for instance Young 2001b), temperature has been predominantly used as an index of evapotranspiration or a descriptor of the seasonality effect. As a means to add to the already existing body of knowledge in this field, an alternative hydrometeorological variable was thus opted for. The advantage of using the measured evaporation data is that it is used in the model as a directly measured variable rather than as a latent variable. The results obtained therefore are directly related to the system rather than inferred as the case is with using measured streamflow as a surrogate catchment moisture storage function.

3. METHODOLOGY

3.1 Catchment description and data

This paper presents the results of the DBM modelling approach to R-F modelling of station number C1H007 located at Vaal River at Goedgeluk. This station has a catchment area of 4686 km^2 and forms part of a larger Upper Vaal water management area in drainage region C (Department of Water Affairs and Forestry, 2004).

This catchment has several meteorological stations providing rainfall and evaporation data. The meteorological stations chosen for this study were: Driehoek at Ermelo (C1E002), Standerton (C1E003), Bethel (C1E004), Nooitgedacht at Ermelo (C1E006) and Riversdale at Grootdraai dam (C1E007). The selection of these stations was based on the availability of sufficient rainfall and evaporation data as well as on whether the available data was commensurate in date of records. Figure 2 presents a section of the Upper Vaal water management area showing the gauging station as well as the meteorological stations from which data used in the model development was obtained.

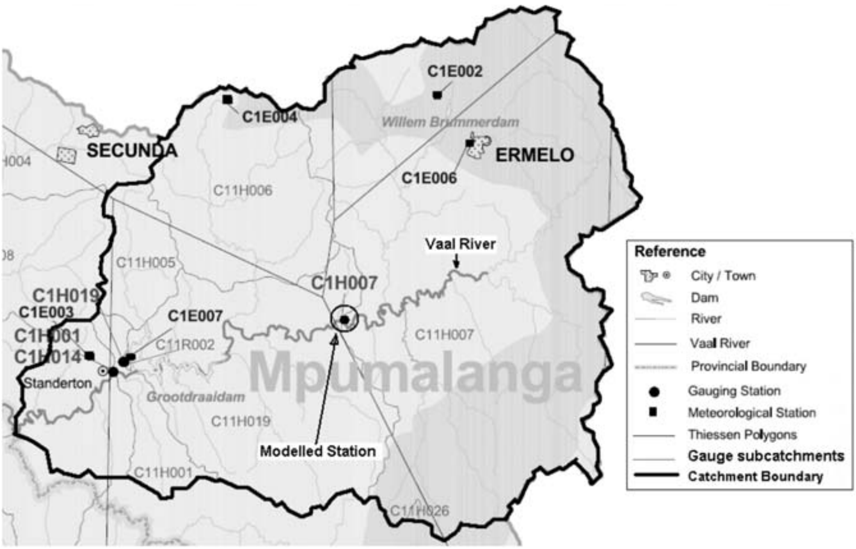

FIGURE 2: A section of the Upper Vaal water management area showing the gauging and meteorological stations used in this study

Note: Stations with C and E denote the meteorological stations in drainage region C for instance C1E003. Stations with C and H denote the gauging stations in drainage region C for instance C1H007

The hydrometeorological data (rainfall, evaporation and stream flow) used in this study was obtained from the daily recorded data courtesy of the Department of Water Affairs and Forestry (DWAF). Further, daily rainfall data was also obtained from the Raster Database developed by Lynch (Lynch, 2004) on behalf of the Water Research Commission (WRC) of South Africa. The data from all the stations was compared on a spreadsheet and matched for dates of observation to check on stations with missing data and those that were regularly monitored. This was important since a daily time scale model was used for the rainfall-flow model hence for any meaningful inferences to be made on the results, the data had to be of the same date and period. The Thiessen polygon method was applied to calculate the average areal daily rainfall while the arithmetic mean method was used to calculate the average daily evaporation data. Figure 3 presents the data for station C1H007 used in this study.

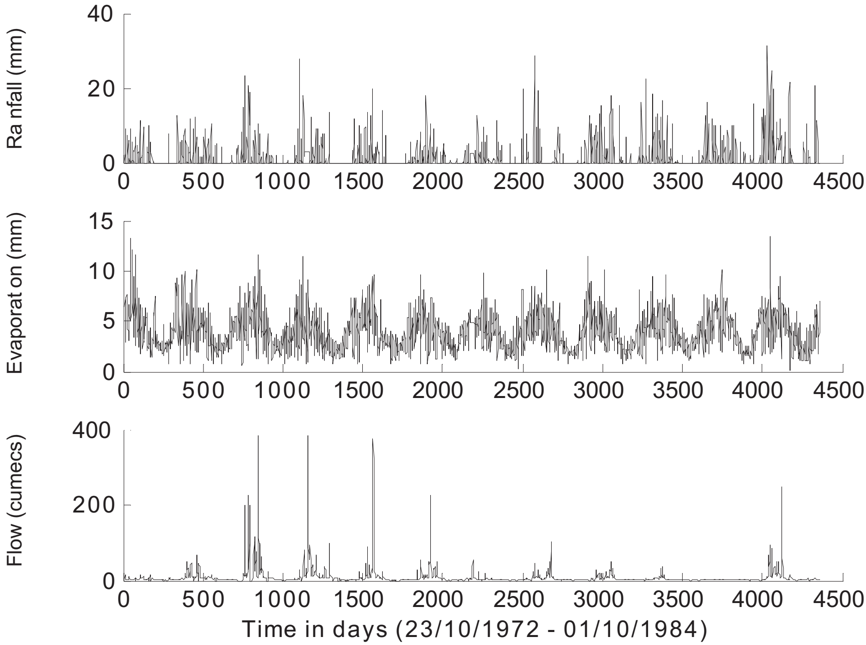

FIGURE 3: Time series data for station C1H007

3.2 Model identification and parameter estimation

Modelling in this study was done in a MATLAB® (MathWorks, 2006) environment. More particularly, the use of CAPTAIN Toolbox was made for the model identification and parameter estimation. CAPTAIN (Pedregal, Taylor & Young , 2004) is a MATLAB® compatible toolbox for non-stationary time series analysis, system identification, signal processing and forecasting. For more information on CAPTAIN, reference can be made to Pedregal, Taylor and Young (2004). One model that has received special attention in this toolbox is the multiple-input, single output TF model. This is the generic class of model that was applied in this study. CAPTAIN includes functions for robust unbiased identification and estimation (for example Refined Instrumental Variable – RIV and Simplified Refined Instrumental Variable – SRIV algorithms) of discrete-time and continuous-time TF models. One advantage of the TF model is its simplicity and ability to characterise the dominant modal behaviour of a dynamic system.

3.2.1 R-F modelling

Based on the given data and in following the generic TF model given in Figure 1, the SDP modelling was used to define the effective rainfall input. Several discernable nonlinear functions defining the effective rainfall were tried. These were: polynomial, power law, rational, and exponential functions. Objective assessment of these functions in terms of their goodness of fit coefficients as well as graphical observations was done. The results showed that the effective rainfall in this case was best defined by a rational function of 2nd order numerator and 3rd order denominator.

The resulting effective rainfall time series together with the stream flow data was then used to develop a discrete time linear TF model. The SRIV estimation as applied in CAPTAIN toolbox was used to identify a family of discrete time linear TF models. The SRIV is practically useful in this situation since it does not require concurrent estimation of a noise model hence is robust to the assumption that the system noise $\hat{e}_t \sim N(0,\hat{\sigma}_e^2)$

4. RESULTS

Following the procedure in Section 3.2.1, the best selected model was of the structure {2 2 1} in the $\{n, m, \delta\}$ nomenclature. This implies a 2nd order TF with a 1 day input lag. The model has $R_T^2 = 56.3\%$ implying that based on the variance of the model errors; it explains approximately 56% of the flow data. In the $\{n, m, \delta\}$ nomenclature, the model is interpreted as follows:

$$\hat{y}_t = \frac{0.4919 - 0.4735z^{-1}}{1 - 1.4634z^{-1} + 0.4735z^{-2}} u_{t-1} + \xi_t$$

$$\xi_t = \frac{1}{1 - 1.1037z^{-1} + 0.5815z^{-2} - 0.3475z^{-3} + 0.1843z^{-4} - 0.056z^{-5}} \hat{e}_t \quad \text{.................(4)}$$

where:

$$u_t = y_t . \left(\frac{p_1 E_t^2 + p_2 E_t + p_3}{E_t^3 + q_1 E_t^2 + q_2 E_t + q_3} \right) . r_t$$

$$p_1 = 3.822(0.095); p_2 = -26.75(0.64); p_3 = 64.31(1.3)$$

$$q_1 = -4.241(0.282); q_2 = -12.15(1.92); q_3 = 187.2(4.9)$$

The variable u_t is the effective rainfall. The variable y_t is the measured flow while E_t and are measured evaporation and rainfall respectively. The values in brackets indicate the standard error bounds defined at 95% confidence interval. The estimated model error is $\hat{e}_t = (y_t - \hat{y}_t)$ where $\hat{y}_t$ the estimated flow, is the SRIV identified TF in Equation 4. The standard errors of the parameters of the TF model (Equation 4) are given in Table 1. Figure 4 compares the output of the deterministic part of the model with the observed flow data. Figure 5 presents the model stochastic residuals.

TABLE 1: Standard errors of the R-F model parameters

Parameter	a_1	a_2	b_0	b_1
Estimate	-1.4634	0.4735	0.4919	-0.4735
Standard error	0.0161	0.0154	0.0113	0.0109

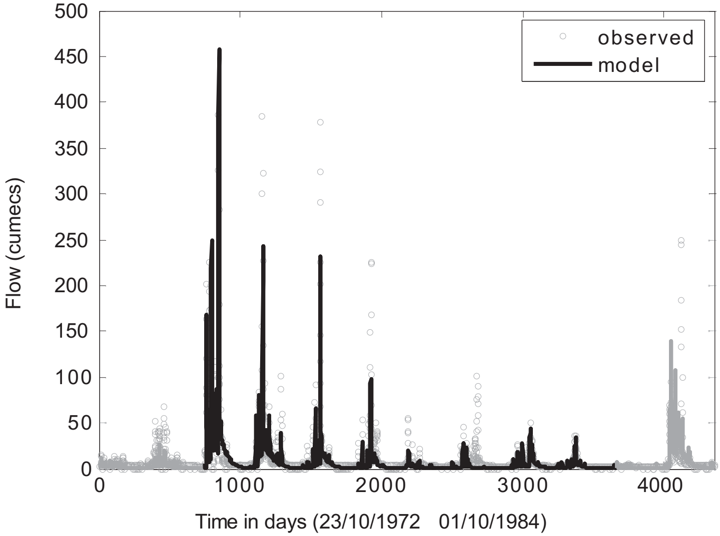

FIGURE 4: Comparison of R-F model transfer function output with the observed flow

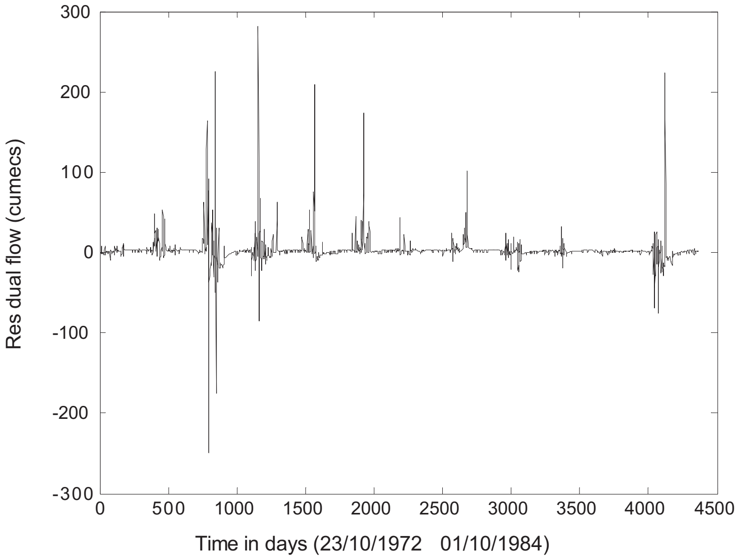

FIGURE 5: R-F model stochastic residuals

4.1 Mechanistic interpretation of R-F model

As noted in Beven (2001) and Young (2001a, 2001b, 2003), an important attribute of the DBM modelling is that the resulting models are only considered credible if they can be interpreted in physically meaningful terms. This attribute is an important point of departure from the ordinary "black-box" models in that model analysis considers not only the fit of the data to the resulting model but also the achievement of physically sensible mechanistic interpretation of the model in line with the subject paradigms.

In this context, if the resulting TF identifying the model is greater than first order and characterised by real eigen values, the roots of the $A(z^{-1})$.

polynomial, then it can be decomposed into different pathways either parallel or series depending on the resulting model and the subject being modelled. The various pathways are identified as first order equations in each pathway. From the resulting decompositions, it is then possible to calculate the following: (i) the residence times (time constants) and (ii) the percentage contribution of each pathway to the overall system.

The deterministic part of Equation 4 can be rewritten as follows:

$$\hat{y}_t = \frac{0.4919 - 0.4735z^{-1}}{1 - 1.4634z^{-1} + 0.4735z^{-2}} u_{t-1} \text{...} (5)$$

where $\hat{y}_t$ is the modelled flow due to effective rainfall. To expand the TF in Equation 5 into its first order components, use was made of the partial fraction expansion algorithms of the Signal Processing Toolbox in MATLAB® environment. The results obtained are given in Equation 6.

$$\hat{y}_t = \left[\frac{0.982}{1 - 0.483z^{-1}} + \frac{0.0179}{1 - 0.9804z^{-1}} - 0.1 \right] u_{t-1} \text{...} (6)$$

The model identified in this case is one with three stores in parallel. The two obvious stores are the quick flow and slow flow components. The third store that is uniquely identified here is the loss component. These stores combine additively to yield the modelled streamflow $\hat{y}_t$. Equation 6 can thus be rewritten as follows:

$$\hat{y}_t = x_t^q + x_t^s + x_t^l$$

$$\Rightarrow x_t^q = \left[\frac{0.982}{1 - 0.483z^{-1}} \right] u_{t-1}; x_t^s = \left[\frac{0.0179}{1 - 0.9804z^{-1}} \right] u_{t-1}; x_t^l = -0.1u_{t-1} \text{..........................} (7)$$

where x_t^q is the quick flow component, x_t^s is the slow flow component and x_t^l is the loss component. The details of the individual first order transfer functions in the decomposition are shown in Table 2. Figure 6 shows the modelled streamflow while Figure 7 shows the flow components generated by the parallel decomposition of the modelled streamflow.

TABLE 2: Decomposed transfer function components of the R-F model

Decomposed TF Components														
TF(A) – Fast Flow					TF(B) – Slow Flow					TF(C) – Loss Component				
RT	SSG	SSG[a]	TC	% $\hat{y}_t$	RT	SSG	SSG[a]	TC	% $\hat{y}_t$	RT	SSG	SSG[a]	TC	% $\hat{y}_t$
0.48	1.9	0.25	1.93	49.8	0.98	0.91	0.12	51.02	24	0	-1.0	-0.13	0	26.2

TF = Transfer Function; RT = Root or Eigen value of the decomposed TF; SSG = Steady State Gain; SSG[a] = Adjusted SSG; TC = Residence time; % $\hat{y}_t$ = Percent of $\hat{y}_t$, the modelled flow due to effective rainfall = $\frac{SSG^a}{\sum SSG^a} * 100$

Note

The adjustment of the SSG was done so that the resulting model is physically meaningful in hydrological terms. This was done by multiplying the SSG by a constant φ so that the total effective rainfall over the observation interval is equal to the total flow. Thus:

$$\varphi = \frac{\sum_{t=1}^{t=N} u_t}{\sum_{t=1}^{t=N} y_t}$$ and

$$SSG^a = \varphi . SSG \qquad (8)$$

Initial attempts were made to let this constant (φ) enter the model through the effective rainfall but the models that resulted were ill defined with imaginary roots for the decomposed TF components. Entering through the SSG resulted in models that could be interpreted meaningfully in hydrological terms. In this case φ was determined as $\varphi = 0.13$.

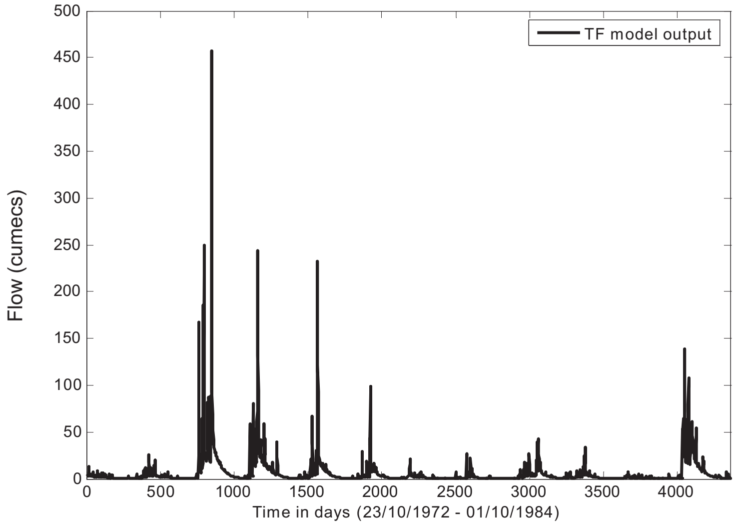

FIGURE 6: Modelled streamflow (transfer function model output)

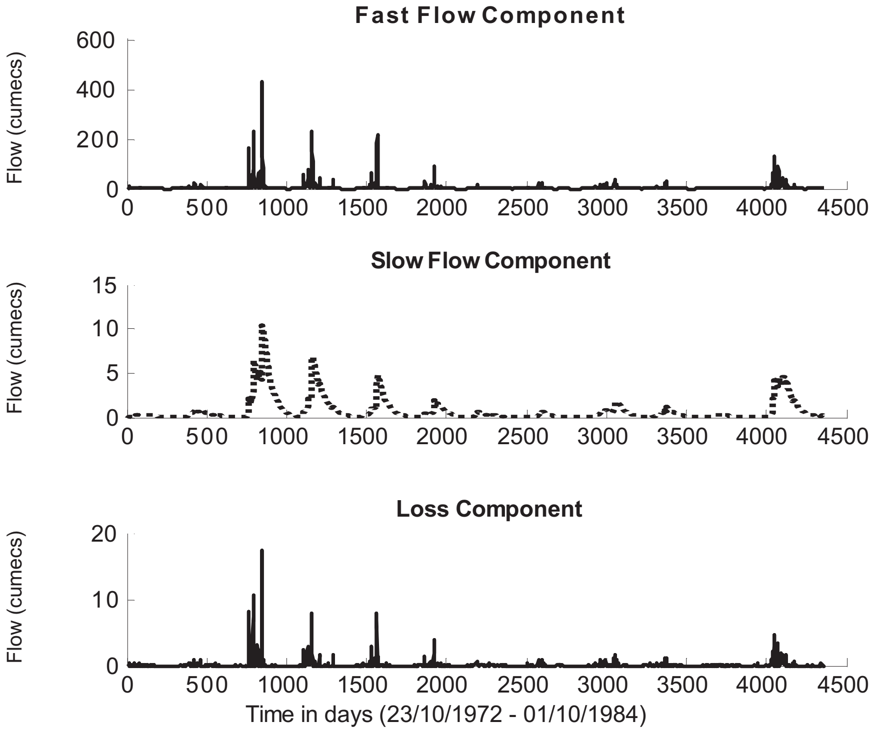

FIGURE 7: Flow components generated by the parallel decomposition of the R-F transfer function model

In block diagram terms, the model components in their decomposed parallel pathway form are presented in Figure 8.

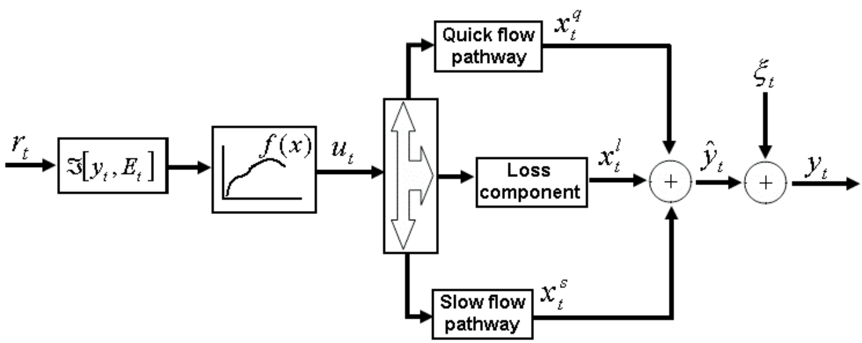

FIGURE 8: Block diagram of the R-F model showing the decomposed parallel flow components

4.1.1 Discussion on the mechanistic interpretation of the R-F model

The first interpretation of the model in Equation 6 is that there is a lag of one day (that is $\delta = 1$) on the effect of measured rainfall on the output. The one day time delay represents the advective delay between the occurrence of rainfall and its first effect on flow. This implies that the rainfall event recorded in the catchment one day back results in an effective rainfall which contributes to a flow component whose effect is manifested at the gauging station a day after the recorded rainfall event that generated it.

The hydrological relevance derived from the decomposed TF in Equations 6 and 7 and the information in Table 2 could be interpreted as follows. Using the adjusted steady state gain values, 25% of the generated effective rainfall measured the previous day (1 day lag) constantly enters the river as a "quick flow" component with a residence time of 1.93 days and constitutes 49.8% of the modelled flow $\hat{y}_t$. The residence time represents the longevity of the effect. This implies that from the first day the quick flow component reaches the gauging station; its effect is realised continuously for a period of 1.93 days before it fizzles out. This "quick flow" component could probably be associated with the surface and sub-surface processes in the catchment.

Concurrently, using the adjusted steady state gain values, it could be interpreted that 12% of the generated effective rainfall measured the previous day constantly enters the river as a "slow flow" component with a residence time of 51.02 days and constitutes 24% of the modelled flow $\hat{y}_t$. The 51.02 days residence time for the "slow flow" component seem longer than would practically be expected. A possible explanation for this very long residence time could be attributed to the DBM modelling philosophy and model identifiability criteria as follows. The tenets underpinning the DBM modelling approach are based on unobserved component models. These models collectively define the system in a combination of linear and nonlinear dynamic relationships. The identification of these components is based on dominant mode concepts which further rely on the information content of the data. This implies that if the available time series data is not sufficiently informative to allow for the estimation of a uniquely identifiable model form from an otherwise identifiable model structure then, the estimated models are limited to a limited number of dominant modes of the system that are excited to any significant extent. By extension of this, it could be argued that the 51.02 days residence time here is considered a lumped time constant constituting all the possible "slow flow" components, parts of which could not be uniquely identified due to possible unidentified limitations in the information content of the available data.

The third pathway is a component herein referred to as "loss component" given by transfer functions C in Table 2. Again using the adjusted steady state gain values, it could be interpreted that 13% of the generated effective rainfall measured the previous day enters the river but is lost from the system on the same day of entry and constitutes 26.2% of the modelled flow $\hat{y}_t$. This pathway is unique in this study in the sense that it is contrary to the normally expected positive flow. The flow here is obtained as negative, occurring within one day. A possible and very significant explanation of this scenario could be related to surface evaporation and or channel transmission losses that are quite often inadequately quantified and understood in the conventional models. As noted by Hughes (2004, S.a.), lack of quantitative understanding of the process of channel transmission losses as well as accurate surface evaporation measurement at various scales is a main limitation to further development of existing models currently in use in the arid and semi arid lands (ASALs) such as those found in parts of South Africa. The channel transmission losses could be attributed to the flow contribution to the river bank storage. It is however not easy at this stage to discern the proportions possibly associated with either of the processes.

In summing up the percentage of generated effective rainfall, it is noted that some 50% of the generated effective rainfall is unaccounted for from the model output. A possible explanation for this could be that some of the effective rainfall could be entering the stream through some other pathways not identified at this stage. This could be due to a possible lack of measurable dependent states or further presence of unobserved components within the system. At this stage all the unidentified pathways enter the model via the stochastic residual model ξ_t.

Also worth noting from the ensuing interpretation is the absence of "instantaneous flow" component. This could be attributed to the earlier observations made with regard to the identified model and its representation of the high flows as depicted in the model stochastic residuals given in Figure 5. The absence of the instantaneous flow component at this stage could again be interpreted possibly to imply that the high flows might not be generated via the natural hydrological processes in the catchment but through possible uncontrolled flow regulation in the catchment. Alternatively it could be an exposition of one of the possible weaknesses of TF models such as DBM models in representing high flow conditions. Further, the presence of spikes in the stochastic residuals of the model could imply presence of outliers during high flows when one might expect measurement errors and extremes of behaviour to be present. This could imply that the instantaneous flow component is improperly defined as it would be represented within the high peak sections of the flow data.

However, the absence of the instantaneous component of flow seems to be in line with results from other studies where metric-conceptual models have been applied to model the rainfall-runoff in a catchment. For instance Evans and Jakeman (1998) note that from the application of IHACRES to many catchments it has been found that the best configuration is generally two stores in parallel. The two stores have been identified as the quick flow and slow flow that combine additively to yield streamflow. IHACRES, as earlier stated, is a metric-conceptual rainfall-runoff model. The model undertakes **i**dentification of **h**ydrographs **a**nd **c**omponent flows purely from **r**ainfall, t**e**mperature and **s**treamflow data. From the foregoing, it could be deduced that the performance of the DBM model identified here is in line with other tried metric-conceptual models. The difference here is that evaporation and not temperature data was used to capture the seasonality effect in the model.

5. CONCLUSIONS

The results and discussions have demonstrated the ability of the stochastic DBM approach to model the catchment fairly well in a simple manner and to obtain a model that is interpretable in meaningful hydrological terms. In terms of the model fit to the data, the transfer function model explains approximately 56% of the data with the parameter standard error bounds as shown in Table 1 defined at 95% confidence interval showing fairly narrow interval widths. Hence the parameters could be considered fairly reliable. Even though the parameters obtained can be considered lumped and catchment specific, the model could be considered to provide a fair representation of the catchment's rainfall-flow dynamics.

The DBM modelling approach has further demonstrated its ability to discern very distinct flow paths. Of particular significance in these results was the channel transmission loss flow path. This pathway is seldom identified and quantified in conventional models. This is considered a major contribution in the field of arid and semi arid hydrological modelling. This is so because as noted by Hughes (2004), one of the major pitfalls of the conventional models currently applied in arid and semi arid hydrological modelling is their lack of identification and quantification of the channel transmission losses. Further, it has been demonstrated that even with the few readily available sets of data at the catchment scale, simple models can still be developed to effectively represent the catchment hydrological dynamics.

6. RECOMMENDATIONS

There is room for possible improvement on the model fit. This could be achieved for instance by a further analysis of the streamflow data to identify and eliminate possible outliers especially in the high flow periods as errors are more likely to be encountered in such periods. The physical interpretation of the resulting TF models could further be enhanced by increasing the amount of information available to identify the model parameters. This could be done for instance by using additional output variables to constrain the parameter space yet not compromising the DBM modelling philosophy.

7. REFERENCES

BEVEN, K. 2001. Rainfall-runoff modelling the Primer. England: John Wiley & Sons.

BEVEN, K. 2006. A manifesto for the equifinality thesis. Journal of Hydrology, 320(1-2):18-36.

BOGNER, K., HINGRAY, B. & MUSY, A. 2002. Data-based mechanistic modelling of rainfall-runoff processes and its application in a complex hydrological context. In: Rizzoli, A.E. & Jakeman, A.J., (eds.). Integrated assessment and decision support: proceedings of the First Biennial Meeting of the International Environmental Modelling and Software Society, iEMSs, 2002, 1:381-386.

DEPARTMENT of Water Affairs and Forestry see SOUTH AFRICA. Department of Water Affairs and Forestry.

DYE, P.J. & CROKE, B.F.W. 2003. Evaluation of streamflow predictions by the IHACRES rainfall-runoff model in two South African catchments. Journal of environmental modelling & software, 18(8-9), Oct-Nov.:705-712.

EVANS, J. & JAKEMAN, A. 1998. Development of a simple, catchment-scale rainfall-evapotranspiration-runoff model. Journal of environmental modelling and software, 13(3-4):385-393.

HUGHES, D.A. 2004. Three decades of hydrological modelling research in South Africa: special Rhodes University centenary edition. South African Journal of Science, 100:638-642.

HUGHES, D.A. S.a. Modelling semi arid and arid hydrology and water resources: the Southern African experience [Online]. Grahamstown: Institute for Water Research, Rhodes University. Available from: http://www.gwadi.org/shortcourses/chapters/hughes_L5.pdf [Accessed: 05/07/2006].

JAKEMAN, A.J. & HORNBERGER, G.M. 1993. How much complexity is warranted in a rainfall-runoff model? Water Resources Research, 29 (8): 2637-2649.

JARVIS, A.J, YOUNG, P.C, TAYLOR, C.J. & DAVIES, W.J. 1999. An analysis of the dynamic response of stomatal conductance to a reduction in humidity over leaves of Cadrella odorata. Plant cell and environment journal, 22:913-924.

KOKKONEN, T. 2003. Rainfall-runoff modelling: comparison of modelling strategies with a focus on ungauged predictions and model integration. D.Sc.Tech. dissertation, Helsinki University of Technology: Helsinki University of Technology Water Resources.

KOKKONEN, T. & JAKEMAN, A.J. 2001. A comparison of metric and conceptual approaches in rainfall-runoff modelling and its implication. Water Resources Research, 37(9):2345-2352.

LEES, M, J., TAYLOR, J., CHOTAI, A., YOUNG, P.C. & CHALABI, Z. S. 1996. Design and implementation of a Proportional-Integral-Plus (PIP) control system for temperature, humidity and carbon dioxide in a glasshouse. Acta Horticulture, 406:115-123.

LEES, M., YOUNG, P.C., BEVEN, K.J, FERGUSON, S. & BURNS, J. 1994. An adaptive flood warning system for the River Nith at Dumfries. In: WHITE, W.R. & WATTS, J. (eds). 2nd International Conference on River Flood Hydraulics organised by Institute of Hydrology, Wallingford. Chichester: Wiley: 65-77.

LYNCH, S.D. 2004. Development of a Raster Database of annual, monthly and daily rainfall for Southern Africa. Pretoria: Water Research Commission. (Report, 1156/1/04).
MATHWORKS. 2006. MATLAB, the language of technical computing. Version 7.0.0. Release 14. [CD-ROM]. Mathworks Inc.

MERRITT, W.S., LETCHER, R.A. & JAKEMAN, A.J. 2003. A review of erosion and sediment transport models. Environmental modelling and software, 18:761-799.

PEDREGAL, D.J., TAYLOR, C.J. & YOUNG, P.C. 2004. System identification, time series analysis and forecasting: the Captain toolbox. Handbook v 1.1. Lancaster: Centre for Research on Environmental Systems and Statistics, Lancaster University.

PERRIN, C., MICHEL, C. & AndrÉAssian, V. 2001. Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. Journal of Hydrology, 242:275-301.

PERRIN, C., MICHEL, C. & AndrÉAssian, V. 2003. Improvement of a parsimonious model for streamflow simulation. Journal of Hydrology, 279:275-289.

PRICE, L., YOUNG, P.C., BERKMANS, D., JANSSENS, K. & TAYLOR, J. 1999. Data-based mechanistic modelling and control of mass and energy transfer in agricultural buildings. Annual Reviews in Control, 23:71-82.

RAJURKAR, M.P., KOTHYARI, U.C. & CHAUBE, U.C. 2004. Modelling of the daily rainfall-runoff relationship with artificial neural network. Journal of Hydrology, 285:96-113.

SOUTH AFRICA. Department of Water Affairs and Forestry. 2004. National Water Resource Strategy. Pretoria: Government Printer.

TAYLOR, C.J., YOUNG, P.C., CHOTAI, A. & WHITTAKER, J. 1998. Non-minimal state space approach to multi-variable ramp metering control of motorway bottlenecks. IEE Proc. Control Theory Application, 145:568-574.

WALLIS, S.G., YOUNG, P.C. & BEVEN, K.J. 1989. Experimental investigation of the aggregated dead zone (ADZ) model for longitudinal solute transport in stream channels. Proc. Inst. of Civil Engineers, 87(2):1-22.

YE, W., JAKEMAN, A.J. & BARNES, C.J. 1995. A parametrically efficient model for prediction of streamflow in an Australian benchmark catchment with complex storage dynamics. Environment International, 21(5):539-544.

YOUNG, P.C. 1992. Parallel processes in hydrology and water quality: a unified time series approach. Journal, Institute of Water and Environmental Management, 6:598-612.

YOUNG, P.C. 1998. Data-based mechanistic modelling of environmental, ecological, economic, and engineering systems. Environmental Modelling and Software, 13:105-122.

YOUNG, P.C. 2000. Stochastic, dynamic modelling and signal processing: time variable and state dependent parameter estimation. In: Fitzgerald, W.J., Walden, A. & Smith, R. (eds.). Nonstationary and Nonlinear Signal Processing. Cambridge: Cambridge University Press:74-114.

YOUNG, P.C. 2001a. Data-based mechanistic modelling of environmental systems. (Keynote paper presented at the International Federation on Automatic Control (IFAC) Workshop on Environmental Systems, August 21, Tokyo, Japan). Unpublished.

YOUNG, P.C. 2001b. Data-based mechanistic modelling and validation of rainfall-flow processes. In: Anderson, M.G. & Bates, P.D. (eds.). Model Validation: perspectives in hydrological science. Chichester: Wiley:117-161.

YOUNG, P.C. 2002. Data-based mechanistic and top-down modelling. In: Rizzoli, A.E. & Jakeman, A.J. (eds.). Integrated Assessment and Decision Support: proceedings of the First Biennial Meeting of the International Environmental Modelling and Software Society, iEMSs. S.I.: iEMSs:1:363-374.

YOUNG, P.C. 2003. Top-down and data-based mechanistic modelling of rainfall-flow dynamics at the catchment scale. Hydrological Processes, 17:2195-2217.

YOUNG, P.C. & BEVEN, K.J. 1994. Data-based mechanistic modelling and the rainfall-flow nonlinearity. Environmetrics, 5:335-363.

YOUNG, P.C. & PEDREGAL, D.J. 1999. Macroeconomic relativity: government spending, private investment and unemployment in the USA 1948-1998. Journal of Structural Change and Economic Dynamics, 10:359-380.

YOUNG, P.C. & TOMLIN, C.M. 2000. Data-based mechanistic modelling and adaptive flow forecasting. In: Lees, M.J. & Walsh, P. (eds.). Flood Forecasting: What Does Current Research Offer the Practitioner? (BHS Occasional paper No. 12, produced by the Centre for Ecology and Hydrology on behalf of the British Hydrological Society). S.I.: s.n.:26-40.

YOUNG, P.C., TAYLOR, C.J., TYCH, W., PEDREGAL, D.J. & McKENNA, P.G. 2004. Captain toolbox. Lancaster: Centre for Research on Environmental Systems and Statistics (CRES), Lancaster University.

Acknowledgements

This paper is based on partial work undertaken towards a doctorate study. We would like to thank Tshwane University of Technology for awarding a bursary to the principal author to undertake his studies. The data provided by the Department of Water Affairs and Forestry (DWAF), and Council for Scientific and Industrial Research (CSIR), South Africa are also greatly appreciated.