Spatial variability of hydrological processes and model structure diagnostics in a 50 km² catchment

Journal:	Hydrological Processes	
Manuscript ID:	HYP-12-0783.R1	
Wiley - Manuscript type:	Research Article	
Date Submitted by the Author:	n/a	
Complete List of Authors:	McMillan, Hilary; National Institute of Water and Atmospheric Research Ltd, Hydrological Processes Gueguen, Myriam; Montpellier SupAgro, - Grimon, Elisabeth; ETH Zürich, Department of Civil, Environmental and Geomatic Engineering Woods, Ross; NIWA, Hydrology and Catchment Processes Clark, Martyn; NCAR, Hydrometeorological Applications Program; Rupp, David; Oregon State University, Oregon Climate Change Research Institute	
Keywords:	Hydrology, Process, Variability, Model Structure, Diagnostic, Signatures	



1		
2		
3		
4		
5	1	Spatial variability of hydrological processes and model structure
6 7	2	diagnostics in a 50 km ² catchment
/ Q		
9	3	Hilary McMillan ^{*1}
10	Δ	Myriam Gueguen ²
11	5	Elicabeth Grimon ³
12	5	
13	6	Ross woods
14	7	Martyn Clark
15	8	David E. Rupp ^⁵
10	9	
18	10	*h.mcmillan@niwa.co.nz
19	11	¹ National Institute of Water and Atmospheric Research, Christchurch, NZ
20	12	² Montpellier SupAgro, France
21	13	³ Department of Civil Environmental and Geomatic Engineering ETH Zürich Switzerland
22	1/	⁴ Department of Civil Engineering University of Bristol UK
23	14	⁵ National Contar for Atmospheric Descent, Devider, Colorado, UCA
24	15	National Center for Atmospheric Research, Boulder, Colorado, USA
20	16	^o Oregon Climate Change Research Institute, College of Earth, Ocean, and Atmospheric Sciences,
20 27	17	Oregon State University, USA
28	10	
29	18	
30	10	Revision submitted to Hydrological Processes
31	15	Revision submitted to right orgical rocesses
32	20	14 th June 2012
33		
34 35	21	
36		
37		
38	22	Abstract

In this paper we develop diagnostic methods to assess spatial variability in hydrological processes, particularly those relevant to catchment modelling. We target a range of catchment responses, including runoff volume, runoff timing, storage-discharge relationships and threshold responses to rainfall and soil moisture. The diagnostics allow us to map the scales and patterns of process variability, to test whether climate or physical catchment characteristics can be used to predict patterns in processes, and to explore the implications for appropriate spatial variability in hydrological model structures or parameters.

We apply the diagnostic tests to the mid-sized (50 km²) Mahurangi catchment in Northland, New Zealand, combining data from 28 flow gauges, 13 rain gauges and 18 soil moisture measurement sites to build a comprehensive description of spatial variation in catchment response. The results show a complex picture: different diagnostics reveal different patterns of hydrological processes, and large variations in processes occur, even over the short length scales involved (~10 km). Catchment and climate characteristics almost all show the same pattern, i.e. that subcatchments in the far North and South of the Mahurangi are similar to each other (higher elevation, steep,

forested), but contrast with central subcatchments (lower elevation, shallower slopes, pasture).
Surprisingly, this pattern is not reflected in the patterns of diagnostic indices, demonstrating the
difficulty of defining realistic a-priori estimates of spatial variability in processes. We discuss how
process variations correspond to the design of components of a simple lumped conceptual model. In
the Mahurangi catchment, we find that spatial variations in multiple aspects of the hydrological
response imply a need for spatial variation in both model structures and parameters.

7 1 Introduction

Conceptual hydrological models are simplifications of the flow pathways in a catchment. Creating a hydrological model involves learning about water partitioning and runoff generation processes, and describing these mathematically (Beven, 2001; Gupta et al., 2012). Because each catchment is unique (Beven, 2000), recent work has called for hydrological model structure to be tailored to the catchment (Kirchner, 2006; Gupta et al., 2008; Savenije, 2009; Fenicia et al., 2011; Kavetski and Fenicia, 2011). The contrasting view seeks a unified hydrological theory, where catchments reflect common organising principles such as minimisation of entropy generation or flow resistance (Sivapalan, 2005; Troch et al., 2008). Both viewpoints emphasise diagnostic evaluation of process heterogeneity, whether to learn about specific catchments or to generalise and extrapolate (McDonnell et al., 2007).

Diagnostic signatures provide targeted analyses of catchment response data, which are used to build
understanding of hydrological processes (Li *et al.*, 2012; Tian *et al.*, 2012) and to choose appropriate
model structures (Clark *et al.*, 2011; Kavetski and Fenicia, 2011; McMillan *et al.*, 2011, Euser *et al.*,
2013). The concept follows that of diagnostic tests to choose model parameter values (Gupta *et al.*,
2008; Yilmaz *et al.*, 2008; Pokhrel *et al.*, 2012). Diagnostic signatures can help us test hypotheses
about catchment function, which is a useful framework for hydrological learning (Beven, 2001;
Beven, 2008; Wagener *et al.*, 2010; Clark *et al.*, 2011).

In this study, we use diagnostic signatures to examine the extent to which hydrological behaviour, and hence recommended model structure and parameters, vary within the mid-size (50 km²) Mahurangi catchment in New Zealand. Our long-term goal is to develop generalisable strategies for model structure selection. In this paper, we view alternative model structures in the context of simple, lumped, conceptual models, as recently popularised by the FUSE (Clark et al., 2008) and FLEX (Fenicia et al., 2011; Kavetski and Fenicia, 2011) multi-model frameworks; and used in the FARM model evaluation framework (Euser et al., 2013). We recognise that complex, physically based models may be able to represent a wide variety of processes within a single model structure. In these cases, there may be multiple, equally plausible model structures, and findings of spatial variability in processes would map to spatially variable model parameters rather than structures.

We use distributed rainfall, flow and soil moisture data from the 50 km² Mahurangi catchment in New Zealand. The Mahurangi provides an unusually rich data set of hydrological measurements, including rainfall, flow and soil moisture data. Previous authors have demonstrated the value of auxillary data sets (in addition to rainfall and flow) to improve the range of diagnostic tools available (Seibert and McDonnell, 2002; Son and Sivapalan, 2007; Fenicia *et al.*, 2008; Blume *et al.*, 2009). This paper builds on diagnostic analysis of model structure in the 0.25 km² 'Satellite Right' subcatchment

Hydrological Processes

of the Mahurangi (Clark *et al.*, 2011; McMillan *et al.*, 2011). By extending the approach to all subcatchments, we can use inter-site comparisons to investigate controls and mechanisms of hydrological behaviour (e.g. Jones, 2005; Carrillo *et al.*, 2011). The Mahurangi varies in slope, soil texture and land-use, but is relatively small, with a relatively narrow range of annual precipitation. The Mahurangi is therefore a valuable location to test hydrological controls without the influence of a dominant climate gradient (Sawicz *et al.*, 2011).

If different catchments are found to require different model structures, a related question is to ask whether these structures can be predicted *a priori*, using readily available data on characteristics to hydrological data in the catchment. Finding relationships between process descriptions and catchment characteristics would allow us to preselect model structures for predictions on a regional or national scale. This aim is analogous to parameter regionalisation (e.g. Merz and Bloschl, 2004; Wagener and Wheater, 2006), and has many similarities to catchment classification which aims to characterise the drivers of hydrological function (McDonnell and Woods, 2004; Wagener et al., 2007). Relationships between process and catchment indices can be tested using correlations (e.g. Zecharias and Brutsaert, 1988; Krakauer and Temimi, 2011), or informally by visual comparison of maps (e.g. Sivapalan et al., 2011).

The aim of this paper is therefore to test the null hypothesis that a single process description, and hence conceptual model structure, would be suitable for all subcatchments of the Mahurangi. Our alternative hypothesis is that processes and hence appropriate structures vary over small scales (1 -10 km), driven by physical characteristics. These characteristics may be linked to commonly-available data such as land-use, terrain or soil type; or they may reflect more complex processes such as co-evolution of flora, soil structure and connectivity, which we are not yet able to quantify. We will achieve the aim as follows: [1] Select a range of diagnostic signatures which evaluate different aspects of catchment response. [2] Evaluate the diagnostics for each subcatchment of the Mahurangi to help determine important differences in processes. [3] Translate the process descriptions into recommendations for spatial variations in conceptual model structure. [4] Identify any spatial patterns in the signatures or model conceptualisations, the scale over which process descriptions vary, and whether patterns of different signatures are related. [5] Determine whether patterns of diagnostics can be attributed to identifiable patterns of physical catchment characteristics.

The paper is structured as follows. Section 2 describes how we selected the diagnostic analyses, reviewing previous literature. Section 3 describes the Mahurangi catchment and data, and current understanding of the catchment processes. Section 4 presents the methods we used to calculate and interpret the diagnostic signatures. Section 5 gives the results of the diagnostic signature calculations. Section 6 combines the signatures to assess spatial variability in processes and model conceptualisations, the significance for model building, and relationships with physical characteristics. In Section 7 we discuss our results in the context of previous work. We conclude in Section 8.

2 Selection of diagnostic analyses

Our objective in selecting the diagnostic analyses was to evaluate as wide a range as possible of aspects of catchment behaviour, and hence model structure decisions, given the data available. The selection of diagnostics was guided by McMillan et al. (2011), who considered the structure decisions required in a typical conceptual model, and which data sources influenced those decisions. We also follow advice from Sawicz et al. (2011) that diagnostic signatures are most useful for catchment classification when they have an interpretable link to catchment function. We chose four themes of catchment response: [1] Water balance characteristics, [2] Hydrograph characteristics, [3] Recession characteristics, [4] Hydrological thresholds. Each of these themes will be discussed in more detail below.

2.1 Water balance characteristics: Runoff ratio.

A fundamental descriptor of catchment function is the water balance: partitioning of water by the landscape into evaporation, runoff and recharge. Inter-catchment variability of mean annual water balance can be used to study regional patterns and relationships with catchment similarity indices (Sivapalan et al., 2011; Norbiato et al., 2009; Merz and Bloschl, 2009). The water balance can also be explored via 'catchment elasticity', the sensitivity of annual streamflow to precipitation (Harman et al., 2011). The diagnostic that we selected for this study was runoff ratio, i.e. the proportion of rainfall that becomes runoff, which is a useful characteristic of the water balance. We calculated runoff ratio for both storm events ('event runoff ratio'), and over continuous rainfall and runoff series ('total runoff ratio'). The two calculations give insights into different components of hydrological behaviour. Total runoff ratio is controlled by water that bypasses the gauge, i.e. evaporation and groundwater fluxes, whereas event runoff ratio quantifies the split between fast and slow runoff processes.

2.2 Hydrograph characteristics: Runoff timing.

The timing of events can give insight into catchment processes. Clark et al. (2011) suggest that when assembling a hydrological model, a strong control on runoff timing is partitioning between surface/near-surface runoff and baseflow, with higher surface runoff volumes corresponding to faster response times. Runoff timing therefore helped to determine the preferred model relationship between soil moisture and drainage, which influenced this partitioning. Li and Sivapalan (2011) noted that, theoretically, characteristic overland flow response time can increase under wet conditions, as areas distant to the channel become saturated. It has also been widely noted that fast runoff responses may result from pipe flow mechanisms (Beven, 2001). To characterise event timing, we used the metric suggested by Clark et al. (2011), that is, the length of time between 50% of event rainfall depth occurring and 50% of event discharge depth occurring. This metric distinguishes between flashy responses and damped responses. We are interested in speed of the dynamic response rather than travel time of water particles.

2.3 Recession characteristics: Timescale, nonlinearity and seasonality.

The storage-discharge relationship is a key element of catchment function, and one of the fundamental building blocks of most conceptual models. This relationship describes catchment behaviour after the immediate response to rainfall has passed, when slow-flow and evaporation processes dominate. At this time, runoff rate is controlled by the quantity and distribution of water in the catchment. In a conceptual model, the storage-discharge relationship is controlled mostly strongly by the number of lower zone reservoirs, their release characteristics, and distribution of

Hydrological Processes

water volume between them (Clark *et al.*, 2011). There are also more minor influences from the
unsaturated zone representation, particularly in summer (Rupp *et al.*, 2009; Staudinger *et al.*, 2011),
and from return flow (Wang, 2011).

In this study, we use recession analysis as a diagnostic to identify storage-discharge relationships
(Hall, 1968; Tallaksen, 1995). An established recession analysis method is to study the relationship
between flow and its time-derivative. In the theoretical case of a conceptual model with a single
reservoir, where flow Q is a power function of storage S,

$$Q = cS^d$$

9 And assuming that flow is equal to change in storage (i.e. negligible evaporation), this leads to the
 10 recession relationship

Eq 1

Eq 2

$$11 \qquad dQ/dt = aQ^b$$

where *c* and *d* can be expressed in terms of *a* and *b*. The same relationship (Eq 2) was derived by Brutsaert and Nieber (1977) to describe groundwater outflow in an idealised aquifer. It is common to plot -dQ/dt against Q on logarithmic axes, to find *a* and *b*. Values of b typically range between 1 and 3. A value of 1 implies a simple linear reservoir, with b > 1 implying greater nonlinearity. However, b > 1 can also be interpreted as a recession controlled by multiple water stores draining at different rates. As the intercept *a* depends on flow magnitude, flows Q can be scaled by the median flow Q₀ and written as

19
$$d\hat{Q}/dt = -\hat{Q}^b/T_0$$

20 Where $\hat{Q} = Q/Q_0$ and T_0 is a characteristic recession time, at the median flow.

Clark et al. (2009) showed that in the Panola catchment (an experimental catchment in Georgia, USA), recessions from subcatchments of sizes 0.1 ha, 10 ha and 41 ha resulted in approximate b-values of 1, 2 and 3 respectively. They suggested that increasing b reflected increasingly complex processes, with spatial variability and riparian controls on flow becoming more important with increasing catchment size. Wang (2011) noted that at Panola, reservoirs could function in series as well as in parallel due to bedrock leakage to the aquifer. Harman et al. (2009) showed that in the general case, the exponent b increases with increasing heterogeneity in catchment hydraulic properties.

To characterise catchment recessions, we use parameters T_0 (timescale), *b* (nonlinearity) and interannual variation in T_0 (seasonality). Refer to Section 4.1.3 for details of the implementation.

31 2.4 Hydrological thresholds

Thresholds exist in many flow pathways, for example field capacity and saturation point of soils, and are a key contributor to hydrological complexity (Blöschl and Zehe, 2005; Ali *et al.*, 2013). For the relationship between soil moisture and vertical drainage, popular conceptual models exist with (Leavesley *et al.*, 1983; Quick, 1995) and without (Wood *et al.*, 1992) a hard threshold at field capacity, and this modelling decision can be a strong control on predictions. Therefore, diagnostics which test the strength of threshold behaviours can provide key clues to suitable model
 conceptualisations.

The ability of a catchment to transport water often has a threshold response to catchment wetness. The threshold may occur in lateral hillslope flow response to rainfall depth (Tromp-van Meerveld and McDonnell, 2006; Tromp-van Meerveld and McDonnell, 2006; Lehmann *et al.*, 2007; Graham *et al.*, 2010), runoff ratio response to antecedent soil moisture (Woods *et al.*, 2001; Penna *et al.*, 2011), or runoff ratio response to maximum soil moisture or water table elevation (Peters *et al.*, 2003). The threshold can be interpreted as showing a causal relationship, or purely co-variation, as discussed by Tromp-can Meerveld and McDonnell (2005) and Western *et al.* (2005).

For this study, we chose to test for threshold behaviour in two relationships. Firstly, we tested for
thresholds between antecedent soil moisture and event runoff ratio. This analysis was chosen as soil
moisture is known to be a strong control on runoff in Satellite subcatchment (Woods *et al.*, 2001).
Secondly, we tested for threshold responses between rainfall depth and flow. This type of threshold
was noted in Satellite subcatchment by McMillan et al. (2011).

3 Case study: Mahurangi catchment

To test the ability of diagnostic signatures to identify and differentiate aspects of catchment response, we used a case study in the Mahurangi catchment in the North island of New Zealand. As we will consider the spatial variability of catchment function, it is appropriate to start by reviewing the spatially-variable characteristics of the Mahurangi, and to what extent these are represented by measured data. The Mahurangi has a warm, humid climate; annual rainfall is approximately 1600 mm, annual pan evaporation is approximately 1300 mm. Catchment land use is a mixture of native and exotic forestry, and pasture. The soils are typically clay or clay loams, less than 1 m deep. Maps of elevation and land use for the Mahurangi are shown in Figure 1.

24 3.1 Instrumentation

Data were collected from 1997 – 2001 during the Mahurangi River Variability Experiment (Woods et al., 2001). Figure 2 shows the locations of rain gauges, flow gauges and soil moisture measurement sites. The 28 flow gauges used compound v-notch weirs, taking measurements every two minutes. Tipping bucket rain gauges were used, also recording every two minutes. Soil moisture measurements were taken every 30 minutes, at 6 sites, each site having Campbell Scientific CS615 sensors at three hillslope locations and 2 depths, a total of 36 sensors. The shallow sensors were placed in the top 300 mm of soil and the deep sensors over a 250 mm depth at the base of the column (generally around 500-800 mm). An extensive effort was undertaken to calibrate each sensor, include comparisons with Neutron Moisture Meter measurements (Western and Seyfried, 2005). For this study, we aggregated all measurements to 1 hr intervals.

3.2 Dominant processes in the the Mahurangi catchment

Field and model-based work contribute to our knowledge of dominant flow pathways in the Mahurangi. The most intensively studied area is the Satellite subcatchment. Detailed mapping showed that shallow soil moisture is controlled by soil texture and macroporosity at small scales, topography being less important (Wilson *et al.*, 2003; Western *et al.*, 2004). The lack of topographic

influence on soil moisture, even in wet conditions, provides evidence against shallow downslope flow through the soil matrix (Woods et al., 2001). Findings of perennial discharge areas at the base of hillslopes, together with a gradational soil profile, suggested the presence of deeper lateral flow paths (Western et al., 2004). An unpublished study by Bowden (2009, pers. comm.), based on a hillslope in Satellite Left subcatchment, rejected earlier hypotheses of fast lateral flow, concluding instead that vertical, preferential flow dominates. The study applied bromide tracer to the upper hillslope, and chloride and deuterated water to the lower hillslope. However, no tracer response was detected in the stream at the base of the hillslope over a two-month period after tracer application, and tracers often bypassed samplers in the soil matrix. These observations suggest that hillslope precipitation percolates downwards via preferential flow paths to the saturated zone, and led to the conclusion that despite the small size of Satellite catchment, groundwater processes are involved in the runoff response. This theory is supported by modelling studies (Chirico et al., 2003; Clark et al., 2011; McMillan et al., 2011) which found that nonlinear and seasonally varying slow flow processes required models with multiple nonlinear storages.

In the Mahurangi catchment, variability in runoff generation is a recurring theme. Atkinson et al. (2003a; 2003b) echoed results from Satellite, finding that conceptual models needed multiple storage buckets for good streamflow predictions. This was most important in summer when early rainfall contributed to wetting up the catchment. The interpretation was that these buckets (with storage capacities chosen at equally-spaced quantiles along a one-parameter probability distribution) mimicked variable source areas for saturation excess flow. Although variable saturation excess flux can also be represented using a single soil moisture state (for example Moore, 2007), the multiple-bucket representation also allows for variable water volume distribution across the buckets. Atkinson et al. (2003b) also tested other changes which increased structural complexity, such as allowing spatially variable inputs or parameters. Explicit representation of rainfall variability also improved predictions, but the multiple bucket representation was the single most effective change. McMillan (2012) provided a possible explanation for the need for multiple storage buckets, by showing that seasonality-changing variability in soil moisture controls the nonlinearity of emergent catchment-scale drainage behaviour.

29 4 Methods

The methods section is divided into two subsections. In Section 4.1, we explain how each diagnostic signature from Section 2 is applied to our case study catchment. In Section 4.2 we describe how the signatures are used to interpret spatial variation in processes and model conceptualisations.

33 4.1 Calculation of diagnostic signatures

34 4.1.1 Runoff Ratio

We calculated runoff ratios for each subcatchment of the Mahurangi (28 flow gauges). This required areal mean rainfall estimates for each subcatchment. We used inverse distance weighting to interpolate the hourly rainfall data, collected at 13 locations within the Mahurangi catchment, onto the centroid of each flow gauge catchment. Total runoff ratios were calculated by dividing total runoff depth (mm) by total rainfall depth (mm), over the whole measurement period (1997 – 2001).

We also calculated runoff ratios on a per-event basis. Storm events were identified from the interpolated rainfall series (see previous paragraph). We defined events when more than 2 mm/hour or 10 mm/day of precipitation fell. Events were considered distinct if they were separated by at least 12 dry hours. Events were deemed to end 5 days after the last rainfall, or when rainfall greater than 0.2 mm/hr signifies the start of a new event. To reduce subjective choices, no baseflow separation was implemented. A minimum of 158 (Satellite Right) and a maximum of 171 (Marine Road) events were identified. Event precipitation depths ranged from 2.56 mm to 250.46 mm. The runoff ratio for each event was defined as event runoff (mm) divided by event rainfall (mm). Event runoff ratio was calculated for each subcatchment as the mean over all events.

10 4.1.2 Runoff timing

We calculated runoff timing for each subcatchment, and for each event, as the length of time between 50% of event rainfall depth occurring and 50% of event discharge depth occurring. We used rainfall series and events as defined in Section 4.1.1. We took the average over all events to give mean runoff timing for each subcatchment.

15 4.1.3 Recession characteristics

16 Criteria for recession periods

17 Recessions were defined as periods of at least 12 hr with no rainfall greater than 0.2 mm/hour. The 18 rainfall series were described in Section 4.1.1. A delay was imposed after rainfall, to eliminate 19 quickflow; previous studies used 1 day for the whole catchment (Atkinson *et al.*, 2002), or 1 hour for 20 the small Satellite subcatchment (Chirico *et al.*, 2003). We used a sliding scale according to 21 catchment size, with delays of 4 - 16 hours.

It is difficult to obtain accurate values of dQ/dt during low flows. We used the accumulated volume method of Rupp and Selker (2006) to increase the period over which dQ/dt was calculated for low values of Q. Despite this, the flow often displayed steps due to limited measurement precision, or small fluctuations due to a diurnal cycle (assumed to be a response to evaporation in the riparian zone). Therefore, prior to recession period selection, a 24-hour moving average filter was applied to all non-storm periods (defined as less than twice the median flow). Visual inspection showed that this method provided a good fit to the measured flow series.

29 Calculation of recession parameters

We used flow data from each subcatchment to plot -dQ/dt against Q on logarithmic axes. A linear fit gives b as the slope. We used Total Least Squares regression that allows for errors in both log(Q) and log(-dQ/dt) (Brutsaert and Lopez, 1998). Shaw and Riha (2012) show that seasonal variation in the Q - dQ/dt relationship can change the derived b (nonlinearity) and T_0 (timescale) values depending whether recessions are fitted (1) using all data, (2) by month or season, (3) on individual events. We used all three methods, and compared the results. Seasonal variation is caused by changing distributions of water within the catchment. For example, in summer, shallow stores may be depleted, with a higher proportion of water lying in deeper stores. Catchment conceptualisations with a single slow-flow store cannot produce seasonally varying Q - dQ/dt relationships. To quantify the degree of seasonality in recessions, we calculated the interquartile range of T_0 for each catchment. Interquartile range is robust to outliers, which can easily occur where few recessions are available in a particular month. T_0 was used in preference to b as variation in recession shapes

commonly takes the form of translations, which do not alter the *b* value (Biswal and Marani, 2010; McMillan et al. 2011; Shaw and Riha, 2012). conditions. catchment scale variability. We use the findings of spatial variability in diagnostic signatures to test our null hypothesis that the whole Mahurangi can be represented with a single process description or model structure. We do this by considering which process differences could be represented using varying parameter values, versus those which might require a different model. In some ways this separation is arbitrary, as many model structure decisions could be redefined as parameter choices. Here we draw on previous

work in Satellite subcatchment by Clark et al. (2011) using the FUSE multi-model framework which allows modular combinations of popular hydrological model components. That work explored dependencies between model and parameter choices and corresponding diagnostic signatures.

4.1.4 Hydrological thresholds

Soil moisture thresholds. We analysed threshold response of event runoff ratio to antecedent soil moisture at each of the 18 soil moisture sensor locations. In each case, we used only the lower sensor at depth c. 600 mm, for simplicity (results from the upper sensors were similar; not shown). For each event, we took soil moisture at the time rainfall started, and event runoff ratio as defined in Section 4.1.1. We plotted these points on a graph as a visual guide to threshold behaviour, and also colour coded points by rainfall depth, to identify large events where rainfall might override the initial

We define strong threshold behaviour as (1) Runoff ratio is constant below the threshold and (2) Runoff ratio increases rapidly with soil moisture above the threshold. To quantify this, we fitted each data set with two intersecting lines (a 'broken stick' fit), using a least-squares measure to optimise the slopes and intersect. We tested two null hypotheses which relate to the two definitions of strong threshold behaviour above: (1) The slope of the first line is zero (2) The two lines have equal slopes. These tests return z-statistics which quantify the strength of evidence for each hypothesis: where the absolute value exceeds 1.96, the null hypothesis can be rejected at the 5% level.

Rainfall thresholds. We also analysed threshold response of flow to rainfall depth, at each flow gauge catchment. We plotted event rainfall against runoff depths, as calculated in Section 4.1.1. We also colour-coded events by season, as seasonally-varying thresholds were previously found in Satellite subcatchment by McMillan et al (2011). Using the same line-fitting technique as the previous section, a threshold relationship was fitted at each site, for all storms combined and for summer and winter seasons separately. We also noted the threshold location (i.e. precipitation depth at threshold) as an indicator of catchment rainfall storage capacity.

4.2 Spatial variability in diagnostic signatures and conceptualisations

4.2.1 Spatial variability of process descriptions

Having calculated values of diagnostic signatures across a range of response types, we then combine the analyses to build a picture of spatial variability of signatures, and interpret these in terms of catchment function. We ask: What are the spatial patterns in inferred process descriptions? Are the patterns of different signatures related? Where possible, we compare the range of signatures values in the Mahurangi with national or literature ranges, to determine the relative importance of

4.2.2 Spatial variability in conceptual model structure

Using these results, we report on the scales over which recommended model structures change in
 the Mahurangi.

3 4.2.3 Relationships between process descriptions and catchment characteristics

We test relationships between process and catchment indices by using visual comparisons and Spearman's rank scores for nonparametric correlation. A range of catchment indices are calculated, guided by previous studies in the Mahurangi (Atkinson et al., 2003b; Woods, 2004) and elsewhere (Post and Jakeman, 1996; Berger and Entekhabi, 2001; Pena-Arancibia et al., 2010; Krakauer and Temimi, 2011; Price, 2011). The indices are: percentage of 1st order streams, percentage forest, percentage unweathered sandstone geology, percentage soils classified as high saturated conductivity, area, percentage North aspect, mean slope, standard deviation of slope, drainage density, and elongation ratio. Example maps of a selection of indices are given in Figure 3.

5 Results: Diagnostic analyses

13 5.1 Runoff Ratio

We calculated event runoff ratios for each subcatchment, as described in Section 2.1, and took the mean over all events; total runoff ratios were also calculated (Figure 4). Subcatchments are coloured by runoff ratio, with smaller nested subcatchments overlying the parent catchments. Spatial organisation is similar for event and total runoff ratios. Total runoff ratios are generally lower than 0.5 (24/28 catchments). The range of total runoff ratios (0.25 - 0.64) suggest differences in evapotranspiration characteristics and/or losses or redistribution by groundwater which bypasses the flow gauges. Values for event runoff ratios are lower than for total runoff ratios by 1-16% (mean 7%) of annual rainfall depth, which demonstrates the importance of slow runoff processes throughout the Mahurangi, as a significant proportional of rainfall reaches the stream outside the 5-day event window.

The highest total runoff ratios occur in steeper, forested catchments in the North. The lowest occur
in the South-East, and particularly in the two small catchments of Marine Road East and Grimmers.
Small catchments thus have more variability in runoff ratios, and therefore more variability in
evapotranspiration (ET) and groundwater fluxes as a proportion of rainfall, while larger catchments
take mid-range values.

29 5.2 Hydrograph Timing

The mean event timings show a moderate range, from 14 to 21 hours (Figure 5). The intercatchment differences are minor when compared to the intra-catchment standard deviations, which have a mean of 15 hours. The wide range in timings is partly caused by extended, intermittent rainfall events. All the timings are relatively long given the small catchment sizes (e.g. see McGlynn et al., 2004 for a comparison), suggesting the existence of slower, deeper flow pathways consistent with low event runoff ratios. Although some of the smallest catchments (e.g. in the North-East and East) have the fastest median runoff timing, there are also small catchments (e.g. in the South West) that have relatively slow median runoff timing.

38 5.3 Recession characteristics

Figure 6 shows b values for each catchment, using all data, or as median and quartiles when calculated per month or per recession. The three methods can produce different b values. The values calculated per month usually lie between those calculated with all the data or by individual recession, showing that seasonality explains a part of the variability. The b values were plotted as a map (Figure 7). The median monthly values were used as these reflect seasonality without the high variability of individual recession values. The b values follow a trend, increasing from North to South. Exceptions occur in the two small catchments of Waterfall Right and Grimmers. The Mahurangi displays a wide range of b values (1.8 - 3.6) compared to typical ranges quoted elsewhere of 1 - 3(e.g. Brutsaert and Nieber, 1977; Figure 9 in Harman et al. 2009).

Recession seasonality is characterised by the interquartile range of T_0 (Figure 7c). Seasonal variation in T_0 typically took the form of higher T_0 (i.e. longer recession timescale) during winter months (not shown). Figure 7 shows similar trends in both T_0 and T_0 interquartile range as in b. In other words, catchments with more nonlinear recessions also have shallower recessions at median flow (higher T_0), and greater seasonal variation. The latter result is consistent with catchments which have greater internal variation in hillslope response time, and therefore have greater potential for both seasonal change and change of dominant store within a single recession. Rupp et al. (2009) noted that where T_0 characteristic times were orders of magnitude greater than inter-storm times, accuracy of b and T_0 estimates was reduced. In the Mahurangi, this is not a concern as T_0 values are of the same order of magnitude as inter-storm times, which have a mean of 5.5 days.

20 5.4 Hydrological thresholds

21 5.4.1 Soil Moisture Thresholds

Figure 8 shows threshold responses between soil moisture and runoff ratio, demonstrating variation in threshold strength between sites. In particular, Marine Road sites appear to have a weaker threshold. The z-statistics for the tests [1] constant runoff ratio below the threshold and [2] no change of slope at the threshold, are shown on each figure as zstat1 and zstat12 respectively. In most subcatchments both null hypotheses are rejected, i.e. runoff ratio increases with soil moisture even at low values, and a distinct change of slope (i.e. threshold) exists. However, for 7 of the 18 sites there is a constant runoff ratio below the threshold. Only the two Lower Marine Road sites have weak thresholds, i.e. the difference in slope of the two lines is not statistically significant. Thresholds are less clear (i.e. lower absolute z statistic) in all the Marine Road sites. Satellite Station sites have the strongest thresholds (highest absolute z statistic). The difference may be due to land cover (forested at Marine Road) or other co-occurring differences such as steeper topography. We return to these explanations in Section 6.

Event rainfall depth is an additional control on runoff ratio. At the Claydens sites, events with rainfall greater than 50 mm typically lie well above the two fitted lines, and could be fitted with a single line with intercept close to zero. The same is true to a lesser extent at Carrans, but the runoff ratios at other sites are less sensitive to rainfall depth. This analysis is consistent with independent data from the soil moisture probe installation, which recorded thinner soils and higher clay contents at the two Northern soil moisture sites (Carrans and Claydens).

5.4.2 Rainfall Thresholds

Figure 9 shows threshold responses between rainfall depth and flow. The gradient of the fitted line
 segments is consistently lower for summer than for winter, giving stronger thresholds in summer

when runoff ratios for small rainfall events are low. The threshold location (i.e. precipitation depth threshold) is usually consistent between seasons. In fact, the threshold location is consistent throughout the Mahurangi, varying only between 51 and 67 mm, with the exception of Grimmers which is an outlier at 99mm (Figure 10). Grimmers has the lowest runoff ratio, giving a less distinct threshold, which may account for the anomaly. It also displays an unusually low recession *b* value, so we conjecture that a lack of fast flow pathways in this small, forested catchment gives a simpler, slow-flow dominated response.

8 6 Results: Spatial variability in conceptualisations

9 In Section 6.1, we combine the diagnostic analyses to build a picture of spatial variability in 10 processes. In 6.2 we consider the implications of spatially variable processes for model structure and 11 parameter choices, and test the hypothesis that 'a single process description, and hence conceptual 12 model structure, would be suitable for use over all subcatchments of the Mahurangi'. In 6.3, we test 13 whether process conceptualisations are related to physical catchment characteristics.

6.1 Spatial variability of process descriptions

 Total and event runoff ratios were higher in the Northern subcatchments, and lower in the South East. As previously noted, the range of total runoff ratios (0.25 - 0.64) suggests high variability in water partitioning and possible redistribution by groundwater. Runoff timing is not correlated to runoff ratio, with faster timing in both the South East (low runoff ratio) and the North East (high runoff ratio). The variation in timing is small, from 14 - 21.5 hours between midpoints of rainfall and flow, and is minor compared to the large within-subcatchment variation. These results suggest that throughout the Mahurangi, vertical drainage is an important flow pathway, driving slower runoff timing and lower runoff ratios than might be expected for small catchments. However, the North East area shows the highest runoff ratios and the fastest runoff timing, suggesting that there a higher proportion of the flow follows shallower, more rapid pathways.

Recession characteristics follow a different pattern, with a North West – South East gradient. Moving South East, recessions become more nonlinear (higher b value), and show a slower recession at median flow (higher T_0) and greater seasonality (larger range of T_0). These diagnostics point to more complex behaviour in the South East, which may represent higher within-subcatchment spatial variability in water storage and release. The interquartile range of b values in the Mahurangi, 1.8 -3.5 for the monthly means, is large given the catchment size. For comparison, the interquartile range of b values for all major rivers in New Zealand is 1.75 - 3 (R. Woods, pers comm). This result shows that large variability in recession shapes, and hence storage-discharge relationships, can occur even within a small domain.

The threshold precipitation depth (50 - 60 mm) for significant runoff generation was the only diagnostic which showed consistent behaviour over the Mahurangi. It was also consistent across seasons, showing a link between temporal and spatial variability emphasised by Sivapalan *et al.* (2011). The threshold response of runoff ratio to antecedent soil moisture showed greater variation in value and strength. The threshold was strongest at Satellite catchment (South East), weakest at Marine Road (South West), moderate at Carrans and Claydens in the North. This pattern is similar to that of runoff timing, so we can empirically link a strong threshold response to antecedent soil moisture to faster runoff timing. The findings from all diagnostic analyses are summarised in Figure
11.

3 6.2 Modelling implications of process variability

We considered the extent to which spatial variation in hydrological processes would require differentiation of model structures and parameters. Our findings are given in Table 1, which lists the spatial variation in each diagnostic and gives a commentary on implications for model building. Based on Table 1 and Figure 11, we reject the hypothesis that a single process description could be used across the whole catchment; instead, key aspects of hydrological response would most appropriately be treated as spatially variable.

In some cases, process variation can be directly linked to the components of a simple conceptual model, e.g. recession behaviour is predominantly controlled by lower zone reservoirs. In other cases, the correspondence between processes and model component is more nuanced, for example runoff ratio is affected most obviously by model schemes for evapotranspiration and deep groundwater losses, but is also affected by any model components which change the soil water dynamics (refer to Table 1). In such cases, the implementation of inter-catchment variability will necessarily be model-specific, and will also depend on information obtained from other diagnostics to constrain interacting processes and model components. We provide a graphical summary in Figure 12 which depicts how model structures and parameters might vary over the Mahurangi catchment.

We therefore tentatively also reject the null hypothesis that the whole Mahurangi could be represented using a single model structure of the simple lumped conceptual type. Our diagnostic analyses suggest strong process variations over the landscape, which would require careful model treatment. Model implementations may take several forms, including spatially variable model structures, spatially variable members of model ensembles, or careful choices of suitable, more complex models with the potential for tuning multiple, interacting components and the flexibility to represent variability in signatures through parameter variation. The development of single- or multi-structure modelling approaches which are able to accommodate the wide heterogeneity in processes over small catchment areas presents a significant challenge. We would welcome future collaboration with others in the modeling community who would like to test the behaviour of their model(s) against multiple spatially variable signatures of hydrologic behaviour in the Mahurangi subcatchments, and evaluate the extent to which a single model structure can represent the spatial variability in processes that we document here.

The spatial scales of process variation in the Mahurangi are of the order of 10 km; that is, the patterns could typically be characterised as variation along gradients rather than scatter (Figures 4,5,7), but we would not expect variation to increase significantly for larger catchments as the Mahurangi already shows a spread of diagnostic values approaching the national range (as discussed in Section 5). This length scale was relatively consistent across diagnostic types in the Mahurangi, however we recognise that in other landscapes, abrupt landscape changes (e.g. in geology), may result in discontinuities in process patterns.

The finding of short scales of process variation has implications for model building; especially in moderate or large size catchments. Where multiple flow gauges or other hydrological data sources are available within a catchment, we would suggest that these data should be used to select

appropriate model structures as well as model parameters. We have also shown that different modelling decisions (i.e. different parts of the model) may have different spatial patterns, and therefore should be considered individually. In our case where diagnostic values tended to vary smoothly over scales of the order of 10 km, this variation would need to be reconciled with discrete changes from one model structure to another. Smooth variation might be achieved by using a combination of structure and parameter changes, e.g. changes to the number of lower zone reservoirs and their storage discharge behaviours, or potentially by using an ensemble of model structures where the selection of structures in the ensemble is gradually changed.

9 6.3 Relationship of process descriptions to catchment characteristics

The typical pattern of physical characteristics in the Mahurangi is for the North and South extremities to be similar to each other, but to contrast with the central catchment. This is true for mean slope, slope standard deviation, percent forest, and mean annual precipitation (Figure 3) which all tend to increase with distance from the outlet (located centrally). Such correlations in predictor variables are common (Krakauer and Temimi, 2011). The geology varies little over the Mahurangi: all subcatchments lie on Waitemata Sandstones (typically having alternating layers of sandstones and mudstones) overlying a basement of greywacke. There is some variation in the extent of sandstone weathering; this also tends to follow the pattern described above. Variables that do not conform to the pattern include percent north aspect (higher for Southern subcatchments) and elongation ratio (higher for North and East catchments). Given the propensity for physical characteristics to follow the first pattern, it is unexpected that none of the diagnostic analyses are similarly distributed.

The North-South gradient of runoff ratios is most similar to patterns of percent North aspect or percent soil with high hydraulic conductivity (although the latter is somewhat subjective due to high variation of soil types over small scales; Western and Seyfried, 2005). Both variables could affect runoff ratios, as south-facing slopes have lower potential evaporation, and highly conductive soils drain quickly, both of which act to reduce actual evapotranspiration. Runoff timing, which is lower in the North and East, is similar in pattern to both percent forest and elongation ratio, with faster timing corresponding to low forest cover and more compact catchments. Recession parameters have a North-West to South-East gradient, which is not similar to any physical gradient, except perhaps the percent of high hydraulic conductivity soil, with higher conductivities linked loosely to recessions with lower b and lower seasonality. Soil moisture threshold strength is highest for Eastern Satellite catchments and lowest for South-West Marine Road catchments, a similar pattern to mean slope, with lower slopes indicating sharper thresholds.

Following the visual identification of relationships between diagnostic indices and catchment characteristics, we calculated the Spearman's rank correlation between each pair. For each diagnostic index, the two characteristics with strongest correlation are shown in Table 2. These results largely back up the visual results. The correlations are weak: only runoff timing has predictor variables (% forest, % unweathered sandstone geology) with correlation greater than 0.5.

It is important to note that uncertainties in the diagnostic indices make the fitting of relationships to catchment characteristics more questionable, especially given the weak correlations found. Uncertainty is present both in the observed data (McMillan *et al.*, 2012) and in the subjective or methodological decisions needed to implement the diagnostics (e.g. the fitting method for recession

Hydrological Processes

b; Stoelzle *et al.*, 2012). For example, we plotted recession analysis *b* values against catchment area (Figure 13). Since the *b* values were fitted to each month separately, and the median taken, we estimate uncertainty magnitude by displaying the interquartile range as an error bar. This example has a Spearman's rank correlation of 0.41, typical of the correlations found. A linear regression is fitted as an example. It is clear that the low correlation and high uncertainties mean that the fitted relationship should be treated with caution.

7 Discussion

It is worthwhile to compare process descriptions derived in this study with previous work in the Mahurangi. Low runoff ratios in Satellite subcatchment are consistent with previous work (McMillan et al., 2011), and suggest that a significant proportion of rainfall becomes recharge to deeper stores. Complex recession characteristics (b = 2.6 at the catchment outlet) are consistent with Atkinson et al. (2003a; 2003b) who found that multiple buckets were required to represent lower zone storage. Findings by the same authors that explicit representation of rainfall uncertainty improved model predictions also hint at the importance of spatially varied processes in the Mahurangi. Seasonal variation in slow flow processes, which we analysed using the annual range of the recession timescale parameter T_{0} , was found by Chirico (2003) to be important in Satellite catchment. Our results showed that while seasonality in T_0 was observed at Satellite, it was even more pronounced in the South West of the Mahurangi. Spatial differences in the strength of the threshold response of runoff to antecedent soil moisture, add to findings of temporal differences in the emergent threshold behaviour (McMillan, 2012) and demonstrate links between temporal and spatial variability. The consistency with previous findings increases confidence in our diagnostic results, despite uncertainties in the data or diagnostic methods.

We found that diagnostic descriptors of process variability are not strongly correlated to physical catchment characteristics (the best predictor variables had Spearman's rank correlations of 0.4 -0.63). Although most of the diagnostics have not previously been calculated over extended areas in order to assess model structure, the recession parameter *b* has been the subject of many previous studies, so provides an opportunity for comparison. Reviews of recession analysis by Hall (1986) and Tallaksen (1995) both found little success in comparisons between catchment characteristics and recession parameters.

Some successes have been reported: for example Tague and Grant (2004) related recession and timing characteristics to the percentage of highly permeable young volcanic bedrock, in an area with two strongly contrasting geological types. Peña-Arancibia et al. (2010) found reasonable correlations of T_0 with climate indices (annual precipitation, aridity) in a dataset of tropical catchments spanning the globe. In an indication of spatial organisation outside of the physical characteristics used in the studies, both Peña-Arancibia et al. (2010) and van Dijk (2010) found spatially correlated residuals, with van Dijk (2010) attributing correlations at scales of 100-150 km to "substrate characteristics not captured by the available soil and geology data". It remains a challenge for hydrologists to develop physical catchment descriptors which characterise the substrate and soil structures, organisation and variability which control runoff generation. For example, Harman et al. (2009) suggest that the recession b parameter may represent the heterogeneity of catchment hydraulic conductivities, but we do not yet have a method to quantify this property.

1 8 Conclusion

2 The contribution of this paper has been to develop a suite of diagnostic tests, which facilitate 3 detailed analysis of the spatial variability of hydrologic processes and help hydrologists to identify 4 model structures which are consistent with dominant processes in catchments where data is 5 available.

We tested the method by applying it in the 50 km² Mahurangi catchment, using flow data from 28 nested small- to meso-scale catchments, alongside 13 rain gauges and 18 soil moisture measurement sites. We used a range of diagnostic signatures to evaluate runoff ratio, runoff timing, storage-discharge relationships and threshold responses to rainfall and soil moisture. Our results showed that there is tremendous heterogeneity in hydrologic signatures over this small geographical area; for example the range of recession shapes was similar to that for all New Zealand. The signatures showed a range of spatial patterns, which varied between diagnostic types, suggesting a high number of degrees of freedom in process variation. Diagnostic indices tended to vary smoothly across the Mahurangi: spatial scales of process changes were in the order of 10 km. We used visual comparison of maps and Spearmans rank correlation to test the predictive power of physical catchment characteristics to explain the spatial variability in diagnostic indices, and found only weak relationships.

We used the variation in diagnostic signatures to recommend how the structures and parameters of a range of model components could change across the subcatchments of the Mahurangi. Diverse diagnostic patterns led to diverse model structure/parameter patterns, and led us to reject the initial hypothesis of a single process description and single model structure for the Mahurangi. In particular, spatial variations in total runoff ratio and recession seasonality were linked to model structural changes in the context of simple lumped conceptual models. Designing generalisable model-building methods which represent variability in multiple interacting processes, with appropriate levels of complexity, remains an ongoing challenge for the hydrological community.

27 Acknowledgements

28 H. McMillan was funded by NZ Ministry for Business, Innovation and Employment, Grants C01X1006 and

- 29 C01X0812.

2		
3	1	References:
4		
5	2	Ali G, Oswald CJ, Spence C, Cammeraat ELH, McGuire KJ, Meixner T, Reaney SM. 2013. Towards a unified threshold-based
6	3	hydrological theory: necessary components and recurring challenges. Hydrol. Process., 27: 313–318. doi:
/	4	10.1002/hyp.9560 Atkinson SE, Siyanalan M, Vinov NP, Woods PA, 2002a, Dradicting chase time variability of hourly streamflow and the role
8	6	of climate seasonality: Mahurangi Catchment, New Zealand, Hydrological Processes, 17 : 2171-2193, DOI:
9	7	10.1002/hyp.1327.
10	8	Atkinson SE, Sivapalan M, Woods RA, Viney NR. 2003b. Dominant physical controls on hourly flow predictions and the role
10	9	of spatial variability: Mahurangi catchment, New Zealand. Advances in Water Resources, 26: 219-235. DOI:
12	10	10.1016/s0309-1708(02)00183-5.
1/	11	Atkinson SE, Woods RA, Sivapalan M. 2002. Climate and landscape controls on water balance model complexity over
14	13	Berger KP. Entekhabi D. 2001. Basin hydrologic response relations to distributed physiographic descriptors and climate
16	14	Journal of Hydrology, 247 : 169-182. DOI: 10.1016/s0022-1694(01)00383-3.
17	15	Beven K. 2001. On hypothesis testing in hydrology. Hydrological Processes, 15: 1655-1657. DOI: 10.1002/hyp.436.
18	16	Beven K. 2001. Rainfall-Runoff Modelling: The Primer. Wiley.
19	17	Beven K. 2008. On doing better hydrological science. Hydrological Processes, 22 : 3549-3553. DOI: 10.1002/hyp.7108.
20	18 10	Beven KJ. 2000. Uniqueness of place and process representations in hydrological modelling. Hydrol. Earth Syst. Sci., 4: 203-
21	20	Biswal B. Marani M. 2010. Geomorphological origin of recession curves. Geophys. Res. Lett. 37 (24403)
22	21	Blöschl G, Zehe E. 2005. Invited Commentary - on hydrological predictability. Hydrol. Processes, 19, 3923–3929.
23	22	Blume T, Zehe E, Bronstert A. 2009. Use of soil moisture dynamics and patterns at different spatio-temporal scales for the
24	23	investigation of subsurface flow processes. Hydrol. Earth Syst. Sci., 13 : 1215-1233.
25	24	Brutsaert W, Nieber JL. 1977. Regionalized drought flow hydrographs from a mature glaciated plateau. Water Resources
26	25	Research, 13 : 637-643. Brutspert W. Lonez J. B. 1008. Basin-scale geobydrologic drought flow features of riparian aquifers in the southern Great
27	20	Plains. Water Resour. Res. 34, 233–240.
28	28	Carrillo G, Troch PA, Sivapalan M, Wagener T, Harman C, Sawicz K. 2011. Catchment classification: hydrological analysis of
29	29	catchment behavior through process-based modeling along a climate gradient. Hydrol. Earth Syst. Sci., 15: 3411-
30	30	3430. DOI: 10.5194/hess-15-3411-2011.
31	31	Chirico GB, Grayson RB, Western AW. 2003. A downward approach to identifying the structure and parameters of a
32	32	process-based model for a small experimental catchment. Hydrological Processes, 17: 2239-2258. DUI:
33	34	Clark MP. Kavetski D. Fenicia F. 2011. Pursuing the method of multiple working hypotheses for hydrological modeling.
34	35	Water Resources Research, 47 : W09301. DOI: 10.1029/2010wr009827.
35	36	Clark MP, McMillan H, Collins D, Kavetski D, Woods R. 2011. Hydrological field data from a modeller's perspective: Part 2:
30	37	process-based evaluation of model hypotheses. Hydrological Processes, 25 : 523-543.
37	38	Clark MP, Rupp DE, Woods RA, Tromp-van Meerveld HJ, Peters NE, Freer JE. 2009. Consistency between hydrological models and field observations: linking processes at the billslape scale to by declarical responses at the waterched
30 20	40	scale. Hydrological Processes 23 : 311-319, DOI: 10.1002/hyp.7154
39 40	41	Clark MP, Slater AG, Rupp DE, Woods RA, Vrugt JA, Gupta HV, Wagener T, Hay LE. 2008. Framework for Understanding
40 //1	42	Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. Water
41	43	Resources Research 44:W00B02.
43	44	Euser T, Winsemius HC, Hrachowitz M, Fenicia F, Uhlenbrook S, Savenije HHG. 2013. A framework to assess the realism of
40	45	model structures using hydrological signatures, Hydrol. Earth Syst. Sci., 17, 1893-1912, doi:10.5194/hess-1/-
45	40	1895-2015. Fenicia E Kavetski D Savenije HHG 2011 Elements of a flexible approach for concentual hydrological modeling [.] 1
46	48	Motivation and theoretical development. Water Resources Research, 47 : W11510
47	49	DOI: 10.1029/2010wr010174.
48	50	Fenicia F, McDonnell JJ, Savenije HHG. 2008. Learning from model improvement: On the contribution of complementary
49	51	data to process understanding. Water Resources Research, 44 . DOI: 10.1029/2007wr006386.
50	52	Graham CB, Woods RA, McDonnell JJ. 2010. Hillslope threshold response to rainfall: (1) A field based forensic approach.
51	54	Gunta BY Clark MP, Vrugt IA, Abramowitz G, Ye M, 2012, Towards a comprehensive assessment of model structural
52	55	adequacy. Water Resources Research, 48 : W08301.
53	56	Gupta HV, Wagener T, Liu YQ. 2008. Reconciling theory with observations: elements of a diagnostic approach to model
54	57	evaluation. Hydrological Processes, 22: 3802-3813. DOI: 10.1002/hyp.6989.
55	58	Hall FR. 1968. Base flow recessions - a review. Water Resources Research, 4 : 973-983.
56	59 60	Harman CJ, Sivapaian M, Kumar P. 2009. Power law catchment-scale recessions arising from heterogeneous linear small-
57	00	SCARE UYNANNICS. WALER RESOULCES RESEATCH, 45 . DUI: 10.1029/2008WF00/392.
58		
59		
60		17

1
2
3
4
5
5
0
1
8
9
10
11
12
13
14
15
10
10
17
18
19
20
21
22
23
20
24
25
26
27
28
29
30
31
22
32
33
34
35
36
37
38
39
40
40
41
42
43
44
45
46
47
48
49
50
50
51
52
53
54
55
56
57
58
50
ວອ

1	Harman CJ, Troch PA, Sivapalan M. 2011. Functional model of water balance variability at the catchment scale: 2. Elasticity
2	of fast and slow runoff components to precipitation change in the continental United States. Water Resources
3	Research, 47 : W02523 DOI: 10.1029/2010wr009656.
4	Jones J. 2005. Intersite Comparisons of Rainfall-runoff Processes. In: Encyclopedia of Hydrological Sciences, Anderson MG
5	(ed.) Wiley, pp: 1839-1854.
6	Kavetski D, Fenicia F. 2011. Elements of a flexible approach for conceptual hydrological modeling: 2. Application and
0	experimental insights. Water Resources Research, 47 : W11511 DUI: 10.1029/2011Wr010748.
o Q	All Andrewski and All All All All All All All All All Al
10	Krakauer NV. Temimi M. 2011. Stream recession curves and storage variability in small watersheds. Hydrol. Farth Syst. Sci.
11	15 : 2377-2389 DOI: 10.5194/hess-15-2377-2011
12	Leaveslev GH. Lichty RW. Troutman BM. Saindon LG. 1983. Precipitation-Runoff Modelling System: User's Manual. In: U.S.
13	Geol. Surv. Water Invest. Rep. , pp: 207.
14	Lehmann P, Hinz C, McGrath G, Tromp-van Meerveld HJ, McDonnell JJ. 2007. Rainfall threshold for hillslope outflow: an
15	emergent property of flow pathway connectivity. Hydrol. Earth Syst. Sci., 11 : 1047-1063.
16	Li HY, Sivapalan M. 2011. Effect of spatial heterogeneity of runoff generation mechanisms on the scaling behavior of event
17	runoff responses in a natural river basin. Water Resources Research, 47: W00h08 DOI: 10.1029/2010wr009712.
18	Li HY, Sivapalan M, Tian FQ. 2012. Comparative diagnostic analysis of runoff generation processes in Oklahoma DMIP2
19	basins: The Blue River and the Illinois River. Journal of Hydrology, 418 : 90-109. DOI:
20	10.1016/j.jhydrol.2010.08.005.
21	McDonnell JJ, Sivapalan M, Vache K, Dunn S, Grant G, Haggerty R, Hinz C, Hooper R, Kirchner J, Roderick ML, Selker J,
22	Weiler M. 2007. Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology.
23	Water Resources Research, 43 : W0/301 DOI: 10.1029/2006wr005467.
24	MicDonnell JJ, Woods KA. 2004. On the need for catchment classification. Journal of Hydrology, 299: 2-3. DUI:
25	10.1010/J.JIIYUI 0I.2004.09.003. McClypp RL McDoppell II. Seibert I. Kendall C. 2004. Scale effects on beadwater catchment runoff timing, flow sources
20	and groundwater-streamflow relations. Water Resources Research 40 : W07504 DOI: 10.1029/2003wr002494
28	McMillan H. 2012. Effect of spatial variability and seasonality in soil moisture on drainage thresholds and fluxes in a
29	conceptual hydrological model. Hydrological Processes. 26 : 2838-2844.
30	McMillan H, Clark M, Bowden WB, Duncan MJ, Woods R. 2011. Hydrological field data from a modeller's perspective. Part
31	1: Diagnostic tests for model structure. Hydrological Processes, 25 : 511-522.
32	McMillan H, Krueger T, Freer J. 2012. Benchmarking observational uncertainties for hydrology: Rainfall, river discharge and
33	water quality. Hydrological Processes, 26(26): 4078-4111. DOI 10.1002/hyp.9384.
34	Merz R, Bloeschl G. 2009. A regional analysis of event runoff coefficients with respect to climate and catchment
35	characteristics in Austria. Water Resources Research, 45 : W01405 DOI: 10.1029/2008wr007163.
36	Merz R, Bloschl G. 2004. Regionalisation of catchment model parameters. Journal of Hydrology, 287: 95-123. DOI:
37	10.1016/j.jhydrol.2003.09.028.
38	Moore, R. J. 2007 The PDM rainfall-runoff model. Hydrology and Earth System Sciences, 11, 483-499.
39	Norbiato D, Borga M, Merz R, Bloeschl G, Carton A. 2009. Controls on event runoff coefficients in the eastern Italian Alps.
40	Journal of Hydrology, 375 : 312-325. DOI: 10.1016/J.Jhydrol.2009.06.044.
41 12	Pend-Arancibid JL, Van Dijk Aljiki, Mulligan M, Bruijnzeel LA. 2010. The fole of climatic and terrain attributes in estimating
42 13	2010
43 44	Penna D. Tromp-van Meerveld HI. Gobhi A. Borga M. Dalla Fontana G. 2011. The influence of soil moisture on threshold
45	runoff generation processes in an alpine headwater catchment. Hydrol. Earth Syst. Sci., 15: 689-702. DOI:
46	10.5194/hess-15-689-2011.
47	Peters NE, Freer J, Aulenbach BT. 2003. Hydrological Panola Mountain Research Watershed, Georgia. Ground Water, 41:
48	973-988. DOI: 10.1111/j.1745-6584.2003.tb02439.x.
49	Pokhrel P, Yilmaz KK, Gupta HV. 2012. Multiple-criteria calibration of a distributed watershed model using spatial
50	regularization and response signatures. Journal of Hydrology, 418: 49-60. DOI: 10.1016/j.jhydrol.2008.12.004.
51	Post DA, Jakeman AJ. 1996. Relationships between catchment attributes and hydrological response characteristics in small
52	Australian mountain ash catchments. Hydrological Processes, 10 : 877-892. DOI: 10.1002/(sici)1099-
53	1085(199606)10:6<877::aid-hyp377>3.0.co;2-t.
54	Price K. 2011. Effects of watershed topography, soils, land use, and climate on baseflow hydrology in humid regions: A
55	review. Prog. Phys. Geogr., 35 : 465-492. DOI: 10.1177/0309133311402714.
50	QUICK IVIC. 1995. The UBC Watershed Model. In: Computer Models of Watershed Hydrology, Singh VP (ed.) Water
5/	Resources Publications, pp: 233-280.
50	times and landscape influences in a pested mesoscale catchment. Hudrol. Earth Suct. Sci. 0 : 120-155
60	Runn DE Schmidt I Woods RA Bidwell VI 2009 Analytical assessment and parameter estimation of a low-dimensional
61	groundwater model, Journal of Hydrology. 377: 143-154. DOI: 10.1016/i ihydrol 2009 08 018
62	Rupp DE, Selker JS, 2006, Information, artifacts, and noise in dO/dt-O recession analysis. Advances in Water Resources 29 .
63	154-160. DOI: 10.1016/j.advwatres.2005.03.019.

1		
2		
3	1	Savenije HHG. 2009. HESS Opinions "The art of hydrology". Hydrol. Earth Syst. Sci., 13 : 157-161.
4	2	Sawicz K, Wagener T, Sivapaian M, Troch PA, Carrillo G. 2011. Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA. Hydrol Earth Syst. Sci. 15 : 2895-2911. DOI:
5	4	10.5194/hess-15-2895-2011.
6	5	Seibert J, McDonnell JJ. 2002. On the dialog between experimentalist and modeler in catchment hydrology: Use of soft
/	6	data for multicriteria model calibration. Water Resources Research, 38 : 1241 DOI: 10.1029/2001wr000978.
8	7	Shaw SB, Riha SJ. 2012. Examining individual recession events instead of a data cloud: Using a modified interpretation of
9 10	8 Q	dQ/dt-Q streamflow recession in glaciated watersheds to better inform models of low flow. Journal of Hydrology,
10	10	Sivapalan M. 2005. Pattern. Process and Function: Elements of a Unified Theory of Hydrology at the Catchment Scale. In:
12	11	Encyclopedia of Hydrological Sciences, Anderson MG (ed.) Wiley, pp: 194-219.
12	12	Sivapalan M, Yaeger MA, Harman CJ, Xu XY, Troch PA. 2011. Functional model of water balance variability at the catchment
14	13	scale: 1. Evidence of hydrologic similarity and space-time symmetry. Water Resources Research, 47: W02522
15	14	DOI: 10.1029/2010wr009568.
16	15 16	Son K, Sivapalan M. 2007. Improving model structure and reducing parameter uncertainty in conceptual water balance
17	10	Sneed M Tetzlaff D. Soulshy C. Hrachowitz M. Waldron S. 2010. Isotonic and geochemical tracers reveal similarities in
18	18	transit times in contrasting mesoscale catchments. Hydrological Processes, 24 : 1211-1224. DOI:
19	19	10.1002/hyp.7593.
20	20	Staudinger M, Stahl K, Seibert J, Clark MP, Tallaksen LM. 2011. Comparison of hydrological model structures based on
21	21	recession and low flow simulations. Hydrol. Earth Syst. Sci., 15 : 3447-3459. DOI: 10.5194/hess-15-3447-2011.
22	22	Stoeizie M, Stani K, Weller M. 2012. Are streamflow recession characteristics really characteristic?, Hydroi. Earth Syst. Sci.
23	23	Tague C & Grant GE (2004) A geological framework for interpreting the low-flow regimes of Cascade streams. Willamette
24	25	River Basin, Oregon. Water Resources Research, 40, W04303, doi:10.1029/2003WR002629.
25	26	Tallaksen LM. 1995. A review of baseflow recession analysis. Journal of Hydrology, 165: 349-370.
26	27	Tian FQ, Li HY, Sivapalan M. 2012. Model diagnostic analysis of seasonal switching of runoff generation mechanisms in the
27	28	Blue River basin, Oklahoma. Journal of Hydrology, 418 : 136-149. DOI: 10.1016/j.jhydrol.2010.03.011.
28	29	I roch PA, Carrillo GA, Heldbuchel I, Rajagopal S, Switanek M, Volkmann THM, Yaeger M. 2008. Dealing with Catchment
29	31	Compass. 3 : 375–392.
30	32	Tromp-van Meerveld HJ, McDonnell JJ. 2005. Comment to "Spatial correlation of soil moisture in small catchments and its
31	33	relationship to dominant spatial hydrological processes, Journal of Hydrology 286 : 113-134". Journal of
32	34	Hydrology, 303 : 307-312. DOI: 10.1016/j.jhydrol.2004.09.002.
33	35	Tromp-van Meerveld HJ, McDonnell JJ. 2006. Threshold relations in subsurface stormflow: 1. A 147-storm analysis of the
34	30	Tromp-van Meerveld HT McDonnell II. 2006. Threshold relations in subsurface stormflow: 2. The fill and snill hypothesis
35	38	Water Resources Research. 42. DOI: 10.1029/2004wr003800.
30 27	39	Van Dijk A (2010) Climate and terrain factors explaining streamflow response and recession in Australian catchments,
31 20	40	Hydrology and Earth System Sciences, 14, 159-169.
30	41	Wagener T, Sivapalan M, Troch PA, McGlynn BL, Harman CJ, Gupta HV, Kumar P, Rao PSC, Basu NB, Wilson JS. 2010. The
40	42	tuture of hydrology: An evolving science for a changing world. Water Resources Research, 46 : W05301 DDI:
40 41	44	Wagener T. Siyapalan M. Troch PA. Woods R. 2007. Catchment Classification and Hydrologic Similarity. Geography
42	45	Compass, 1 : 901-931.
43	46	Wagener T, Wheater HS. 2006. Parameter estimation and regionalization for continuous rainfall-runoff models including
44	47	uncertainty. Journal of Hydrology, 320 : 132-154. DOI: 10.1016/j.jhydrol.2005.07.015.
45	48	Wang D. 2011. On the base flow recession at the Panola Mountain Research Watershed, Georgia, United States. Water
46	49 50	Western AW Sevfried MS 2005 A calibration and temperature correction procedure for the water-content reflectometer
47	51	Hydrological Processes, 19 : 3785-3793. DOI: 10.1002/hyp.6069.
48	52	Western AW, Zhou SL, Grayson RB, McMahon TA, Bloschl G, Wilson DJ. 2004. Spatial correlation of soil moisture in small
49	53	catchments and its relationship to dominant spatial hydrological processes. Journal of Hydrology, 286: 113-134.
50	54	DOI: 10.1016/j.jhydrol.2003.09.014.
51	55 56	western Aw, Zhou SL, Grayson KB, Micivianon TA, Bioschi G, Wilson DJ. 2005. Reply to comment by Tromp van Meerveld and McDonpell on Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial
52	57	hydrological processes. Journal of Hydrology. 303 : 313-315. DOI: 10.1016/i.ihydrol.2004.09.001.
53	58	Wilson DJ, Western AW, Grayson RB, Berg AA, Lear MS, Rodell M, Famiglietti JS, Woods RA, McMahon TA. 2003. Spatial
54	59	distribution of soil moisture over 6 and 30 cm depth, Mahurangi river catchment, New Zealand. Journal of
55	60	Hydrology, 276 : 254-274. DOI: 10.1016/s0022-1694(03)00060-x.
56	61 62	wood EF, Lettenmaler DP, Zartarian VG. 1992. A land-surface hydrology parameterization with subgrid variability for
5/	02	general-circulation mouels. Journal of Geophysical Research-Atmospheres, 97: 2717-2728.
50 50		
59 59		40
00		19

2	
3	
4	
4	
5	
6	
7	
1	
8	
9	
10	
10	
11	
12	
13	
11	
14	
15	
16	
17	
10	
10	
19	
20	
21	
∠ I 00	
22	
23	
24	
25	
25	
26	
27	
28	
20	
29	
30	
31	
22	
32	
33	
34	
35	
30	
36	
37	
38	
00	
39	
40	
41	
12	
42	
43	
44	
45	
40	
40	
47	
48	
10	
49	
50	
51	
52	
52	
53	
54	
55	
56	
50	
57	
58	

Woods	RA. 2004. The impact of spatial scale on spatial variability in hydrologic response: experiments and ideas. Scales in
	Hydrology and Water Management/Echelles en hydrologie et gestion de l'eau, 287 : 153-167.
Woods	RA, Grayson RB, Western AW, Duncan MJ, Wilson DJ, Young RI, Ibbitt RP, Henderson RD, McMahon TA. 2001.
	Experimental Design and Initial Results from the Mahurangi River Variability Experiment: MARVEX. In:
	Observations and Modelling of Land Surface Hydrological Processes, Lakshmi V, Albertson JD, Schaake J (eds.)
	American Geophysical Union, pp: 201-213.
Yilmaz	KK, Gupta HV, Wagener T. 2008. A process-based diagnostic approach to model evaluation: Application to the NWS
	distributed hydrologic model. Water Resources Research, 44 . DOI: 10.1029/2007wr006716.
Zechar	as YB, Brutsaert W. 1988. Recession characteristics of groundwater outflow and base flow from mountainous
	watersheds. Water Resources Research, 24: 1651-1658.

Diagnostic Analysis	Spatial Variability	Spatially varying model component	Spatially varying parameters	Commentary
Total Runoff Ratio	High (0.24 –0.64)	ET Scheme (+ infiltration, drainage schemes)		Total Runoff Ratio is controlled by water losses to ET and deep groundwater. In a lumped, conceptual model, simulated ET depends on the ET scheme (e.g. lower runoff ratios where ET demand is preferentially satisfied from wetter surface soils; Clark et al, 2011) and on soil water dynamics which depend on infiltration and drainage schemes, soil depth and field capacity. Losses to deep groundwater are commonly assumed to be negligible.
				Some of these components may be determined independently by other diagnostics (e.g. soil depth by model thresholds; see below). Depending on these other factors, spatially variable or more complex model components for ET and deep groundwater may be needed to simulate high variations in total runoff ratio.
Runoff Timing	Low (14 – 21.5 hrs between the points of rainfall and flow). Not correlated to catchment size or runoff ratio	-	Surface Runoff	Runoff timing is sensitive to the split between surface and subsurface flow (Cla et al, 2011). Variation could be accommodated by changing parameters which control the surface runoff quantity.
Recession nonlinearity (<i>b</i> value)	Very high (1.8 – 3.5 interquartile range); comparable with national variability	-	Storage- Discharge relationship	Recession characteristics are controlled by the number and type of lower zone reservoirs (Clark et al, 2011). Different <i>b</i> values can be fitted by modifying the storage-discharge exponent in a single reservoir (refer to Section 2.3).
Recession <i>T</i> ₀ seasonality	High (5 – 50 days) Correlated with recession b value	Number of lower zone reservoirs	-	High seasonality in T_0 requires multiple lower zone reservoirs which are unnecessary in catchments with weakly seasonal T_0 .
Threshold response to precipitation	Low: consistent threshold precipitation depth (51 – 67 mm)	-	-	Threshold response is controlled by model soil depth, and is consistent across the catchment.
Threshold response to antecedent soil moisture	Moderate: threshold strength varies from significant to not significant. Pattern is similar to that of runoff timing	-	Drainage exponent	Threshold response is controlled by model relationship between soil moisture and drainage. Previous work showed that parameterisation of drainage as a power function of soil moisture was suitable for the Mahurangi (Clark et al, 2011; McMillan, 2012). The exponent could be modified to change the thresho strength.

Table 1: Summary of diagnostic variability and implications for model structure in the Mahurangi

Diagnostic Index	Characteristic with highest Spearman	Spearman Rho	Characteristic with second highest	Spearman Rho
	Rho		Spearman Rho	
b	% High K Soil	-0.45	Area	0.41
T ₀	% Forest	0.45	% High K Soil	0.40
T_0 range	% High K Soil	-0.45	Slope SD	-0.42
Total Runoff Ratio	% N aspect	-0.40	Area	0.31
Event Runoff Ratio	% N aspect	-0.41	Slope SD	0.29
Runoff Timing	% Forest	0.63	% Sandstone	0.51

2 Table 2: Diagnostic Indices with the two physical characteristics that have the highest Spearman's

3 rank correlation coefficient







Figure 2: Instrumentation of the Mahurangi catchment: (A) Rain and Flow gauges (B) Soil moisture probes



Figure 3: Examples of catchment characteristic indices used in this study, (A) Mean Slope (°) (B) % Forest (C) % High Hydraulic Conductivity Soil (D) % North Aspect (E) Annual precipitation (mm)



Figure 4: Runoff ratios of the Mahurangi catchment, using (A) all data and (B) mean runoff ratio calculated for individual storm events.

Page 27 of 35





Figure 5: Mean runoff timing for each subcatchment of the Mahurangi



Figure 6: b values for each catchment, ordered by median monthly b



Figure 7: Maps of (A) Recession nonlinearity parameter: b (B) Recession characteristic time at median flow: T_0 (days) (C) Recession seasonality parameter: T_0 interquartile range (days)



Figure 8: Event runoff ratio as a function of antecedent soil moisture at each soil moisture measurement site. Events are colour coded by rainfall depth. Two line segments are fitted to the storms with rainfall depth less than 50mm.



Figure 9: Relationships between event precipitation depth and event runoff depth, for each gauged subcatchment of the Mahurangi, during summer (Nov-Apr; filled circles) and winter (May-Oct; open circles). A threshold relationship (two line segments) is fitted to the points by season and combined.



Figure 10: Threshold precipitation depth by subcatchment, calculated using combined (summer and winter) storm events.



Figure 11: Simplified graphic of process variability over the Mahurangi





Figure 12. Simplified graphics of suggested model structure and parameter variability over the Mahurangi

3.5

3

2.5

2

10⁰

b value

Monthly b: median and quartiles

Linear Regression Fit

10¹

log(a) L

٠

Area (km²)

Figure 13: Plot of b value as a function of log(area), with linear regression line





60

http://mc.manuscriptcentral.com/hyp