

This is the accepted manuscript version of the contribution published as:

Woodward, S.J.R., Wöhling, T., **Rode, M.**, Stenger, R. (2017):
Predicting nitrate discharge dynamics in mesoscale catchments using the lumped StreamGEM model and Bayesian parameter inference
J. Hydrol. **552**, 684 - 703

The publisher's version is available at:

<http://dx.doi.org/10.1016/j.jhydrol.2017.07.021>

Accepted Manuscript

Research papers

Predicting nitrate discharge dynamics in mesoscale catchments using the lumped StreamGEM model and Bayesian parameter inference

Simon James Roy Woodward, Thomas Wöhling, Michael Rode, Roland Stenger

PII: S0022-1694(17)30472-9

DOI: <http://dx.doi.org/10.1016/j.jhydrol.2017.07.021>

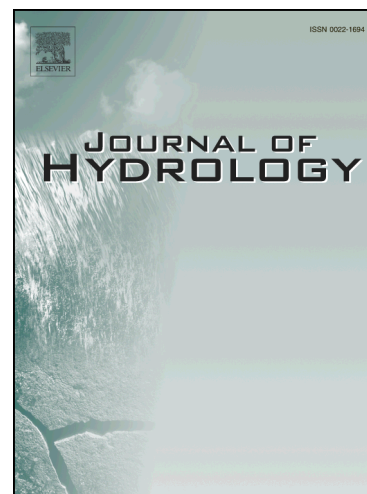
Reference: HYDROL 22123

To appear in: *Journal of Hydrology*

Received Date: 19 August 2016

Revised Date: 21 April 2017

Accepted Date: 11 July 2017



Please cite this article as: Woodward, S.J.R., Wöhling, T., Rode, M., Stenger, R., Predicting nitrate discharge dynamics in mesoscale catchments using the lumped StreamGEM model and Bayesian parameter inference, *Journal of Hydrology* (2017), doi: <http://dx.doi.org/10.1016/j.jhydrol.2017.07.021>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Predicting nitrate discharge dynamics in mesoscale catchments using the lumped StreamGEM model and Bayesian parameter inference

Simon James Roy Woodward ^{a,*}, Thomas Wöhling ^{a,b}, Michael Rode ^c, Roland Stenger ^a

^a Lincoln Agritech Limited, Private Bag 3062, Hamilton 3240, New Zealand,

simon.woodward@lincolnagritech.co.nz, thomas.woehling@tu-dresden.de, michael.rode@ufz.de,

roland.stenger@lincolnagritech.co.nz

^b Department of Hydrology, Technische Universität Dresden, 01069 Dresden, Germany

^c Department of Aquatic Ecosystem Analysis and Management, Helmholtz Centre for Environmental Research-UFZ, 39114 Magdeburg, Germany

* Corresponding author, Lincoln Agritech Limited, Private Bag 3062, Hamilton 3240, New Zealand, telephone +64 7 858 4845, simon.woodward@lincolnagritech.co.nz

SGEMPaper17text_revised3.docx, 12 July 2017, 12:28.

for *Journal of Hydrology*

Abstract

The common practice of infrequent (e.g., monthly) stream water quality sampling for state of the environment monitoring may, when combined with high resolution stream flow data, provide sufficient information to accurately characterise the dominant nutrient transfer pathways and predict annual catchment yields. In the proposed approach, we use the spatially lumped catchment

model StreamGEM to predict daily stream flow and nitrate concentration ($\text{mg L}^{-1} \text{NO}_3\text{-N}$) in four contrasting mesoscale headwater catchments based on four years of daily rainfall, potential evapotranspiration, and stream flow measurements, and monthly or daily nitrate concentrations. Posterior model parameter distributions were estimated using the Markov Chain Monte Carlo sampling code DREAM_{zs} and a log-likelihood function assuming heteroscedastic, t-distributed residuals. Despite high uncertainty in some model parameters, the flow and nitrate calibration data was well reproduced across all catchments (Nash-Sutcliffe efficiency against Log transformed data, NSL, in the range 0.62–0.83 for daily flow and 0.17–0.88 for nitrate concentration). The slight increase in the size of the residuals for a separate validation period was considered acceptable (NSL in the range 0.60–0.89 for daily flow and 0.10–0.74 for nitrate concentration, excluding one data set with limited validation data). Proportions of flow and nitrate discharge attributed to near-surface, fast seasonal groundwater and slow deeper groundwater were consistent with expectations based on catchment geology. The results for the Weida Stream in Thuringia, Germany, using monthly as opposed to daily nitrate data were, for all intents and purposes, identical, suggesting that four years of monthly nitrate sampling provides sufficient information for calibration of the StreamGEM model and prediction of catchment dynamics. This study highlights the remarkable effectiveness of process based, spatially lumped modelling with commonly available monthly stream sample data, to elucidate high resolution catchment function, when appropriate calibration methods are used that correctly handle the inherent uncertainties.

Keywords

Stream flow hydrograph; water quality; direct runoff; groundwater discharge; DREAM; uncertainty

1 Introduction

Catchment nitrate export is a highly topical issue given the importance of both pastoral agriculture and eco-tourism to New Zealand's economy. Nitrate losses from dairy farming, in particular, have been linked with decline in water quality indices in a number of iconic streams and lakes (Wilcock et al., 2006; McDowell et al., 2013; Morgenstern et al., 2015). The ability to characterise and quantify these losses with greater accuracy is highly valuable for developing an improved understanding of nitrate transport and attenuation processes, as well as for identifying opportunities for targeted management and policy interventions.

Nitrogen cycling and transport in mesoscale catchments (10-10000 km²) has predominantly been studied using a spatially-distributed, process-based modelling approach (Breuer et al., 2008; Jiang et al., 2014). In this approach, water percolation and runoff and nitrogen transformations and transport are represented in several surface and subsurface layers at a relatively fine "hydrological response unit" scale, and exports are then aggregated through a stream network model. Spatially-explicit modelling would appear appropriate because the majority of nitrate entering streams in rural areas can be attributed to diffuse land-based sources such as pastoral or arable agriculture (e.g., Bowes et al., 2014). Excess nitrogen applied to pasture as animal excreta or to crops as nitrogen fertiliser rapidly mineralises to nitrate. A small proportion may be carried to streams in direct runoff (i.e., surface runoff, interflow and artificial drainage; USDA, 2004) following rainfall, but being highly soluble, most nitrate leaches below the root zone. Due to the oxidised conditions prevalent in the unsaturated zone, most of this nitrate arrives at the water table and enters the shallow groundwater. The upper, seasonally recharged layer of groundwater tends to be oxidised, and nitrate-bearing; higher water table gradients in winter mean that this nitrate can be rapidly transported laterally to discharge into streams, producing the typically-observed high concentrations of stream nitrate in winter (e.g., Hesser et al., 2010; Jiang et al., 2014; Jomaa et al., 2016). Deeper

groundwater, however, often has much lower nitrate concentrations, if reduced conditions and denitrification are present, or if the water was recharged prior to land use intensification. The longer flow paths and relatively lower discharge rate of this water means that these low concentrations only dominate stream water chemistry following prolonged dry periods, once the shallower, higher nitrate layer has discharged (Woodward et al., 2013). In agreement with this pattern, all 26 stream water monitoring sites in the Waikato region analysed in a regional-scale study featured positive nitrate concentration-discharge relationships (Woodward et al., 2016a). By contrast, where aquifer conditions are not conducive to denitrification and high land use intensity has existed for decades, the deeper groundwater may contain high concentrations of nitrate, and the highest stream nitrate concentrations may be observed during low-flow conditions in summer (e.g., Ruiz et al., 2002; Martin et al., 2004; Oehler et al., 2009; Sackmann, 2010; Li et al., 2010; Wade et al., 2012). Lower nitrate concentrations during high-flow conditions may occur as a result of dilution of contaminated groundwater by surface runoff (that is typically low in nitrate).

Despite their conceptual attractiveness, distributed models are challenging to parameterise. These challenges arise due to the complexity and uncertainty of representing the subsurface processes, the low information content of hydrological data (Jakeman and Hornberger, 1993; Woodward et al., 2016b), and the relative scarcity of hydrogeochemical observations. While parametrisation techniques for complex models continue to evolve, predictive performance so far remains modest (e.g., Jiang et al., 2015). Conversely, alternative approaches are needed for decision making based on readily available data.

Regulatory organisations typically monitor stream water quality on a monthly basis, which provides sufficient information to identify long term trends as these may be related to seasonal, land use or climate changes (e.g., WRC, 2013). This low resolution state of the environment monitoring is not generally considered suitable for the purpose of characterising the mechanisms driving nutrient

export dynamics in response to short term events such as surface runoff, storm flow, or post-storm recession, in order to identify the primary flow paths delivering nutrients within particular catchments and to determine accurate nutrient yields. More powerful analysis of this low resolution water quality data may be possible, however, when it is paired with high resolution (e.g., daily) stream flow information. This leads to an approach which combines the strengths of the purely hydrological and more chemistry-focused methods.

Daily stream flow has been the basis for modelling efforts to characterise catchment flow paths on the basis of hydrological data alone. The low information content of this data, however, means that the standard hydrograph separation or rainfall-runoff modelling approaches are often unable to identify a unique flow path separation on this basis (Jakeman and Hornberger, 1993; Uhlenbrook et al., 1999; Schoups and Vrugt, 2010; Longobardi et al., 2016; Su et al., 2016).

Chemically-based methods of hydrograph separation are also in common use (Christophersen et al. 1990; Klaus and McDonnell, 2013). This “end-member mixing analysis” (EMMA) approach assumes that temporal changes in stream water chemistry reflect changing flow contributions from a known set of water sources (the “end-members”, e.g., rainfall, soil water, groundwater) whose chemical compositions remain constant over time. Because these conditions are most appropriate during short-term, dynamic events, and high resolution sampling is generally required, this method has usually been applied to storm flow separation. Even so, considerable uncertainties may remain due to the difficulty of adequately characterising the source water chemistry (Adams et al., 2009; Delsman et al., 2013).

A combined approach, combining hydrological and water quality information, may resolve these problems. In this approach, the hydrological and hydrogeochemical parameters of a model are estimated through simultaneous calibration to stream flow and water quality data, in principle improving the identifiability of both. This approach was suggested by Ruiz et al. (2002), who

observed that a hydrological model consisting of two flow paths, each discharging water with an assumed fixed nitrate concentration, was sufficient to explain differences in seasonal patterns of nitrate concentration in four small ($0.1\text{--}0.6\text{ km}^2$) catchments in South Western Brittany, France, on the basis of differences in nitrate concentrations in groundwater. Hartmann et al. (2013) similarly used nitrate and sulphate data from a larger (783 km^2) karst system in the Middle East to evaluate alternative model structures of the system.

Following this approach, the three flow path spatially lumped catchment model “StreamGEM” (Bidwell et al., 2008; Woodward et al., 2013) was developed to estimate water and nitrate discharges in a small (15 km^2) dairying catchment in the Waikato region of New Zealand. By calibrating the model to daily stream flow and monthly nitrate data, it was possible to infer daily water and nitrate discharge contributions from near-surface, fast shallow groundwater, and slower deeper groundwater flow paths. The ability to derive this information from monthly water quality data is highly useful, allowing high resolution analysis of catchment dynamics without the need for high resolution water quality data. The objective of the current work, was to extend the StreamGEM approach to quantify nitrate fluxes in larger catchments ($100\text{ to }1000\text{ km}^2$) on the basis of long term daily flow series and associated monthly stream water quality samples. Since policy and management decision are typically based on an annual time step, the calibrated model was also used to calculate annual water and nutrient fluxes for the catchments, including the predictive uncertainty of these flux calculations.

In order to evaluate the effectiveness of the approach, the uncertainties of the calibrated parameters and associated model predictions were quantified using Bayesian uncertainty analysis. Bayesian methods are considered the most rigorous approach to treat uncertain model parameters, errors in the model structure and its inputs, and to quantify the resulting predictive uncertainty. However, the associated numerical methods, such as Markov Chain Monte Carlo (MCMC) sampling,

are numerically costly, and consequently their use is currently limited to relatively fast models with modest numbers of parameters (e.g., Schoups and Vrugt, 2010; Li et al., 2012; Jiang et al., 2015). In fact the StreamGEM model is an ideal candidate for such analysis. A useful overview of the history of MCMC methods is provided by Wu and Zeng (2013), leading up to the development of the Differential Evolution Adaptive Metropolis algorithm, DREAM (Vrugt et al., 2002), a variant of which (DREAM_{zs}) was used in the current study (ter Braak and Vrugt, 2008).

The main objectives of the study, therefore, were:

- to estimate water flow paths and nitrate flux dynamics in contrasting mesoscale catchments based on infrequent stream chemistry data,
- to quantify the uncertainty of model predictions (both forecasts and hindcasts) and identify the main sources of uncertainty,
- to quantify the discrepancy between annual nutrient catchment yields determined from dynamic modelling and monthly balance calculations.

The remainder of the paper is structured as follows. In Section 2 we first describe the characteristics and data sets of the four selected catchments. Next follows a description of the hydrological model and the Bayesian parameter inference scheme used in the study. In Section 3 we present the results of the analysis and discuss their implications, before summarizing our key findings in Section 4.

2 Methods

2.1 Catchment Data Sets

2.1.1 Overview

Four catchments were considered in this study, two of which were simulated under three alternative rainfall data series, and one under two alternative nitrate sampling frequencies, giving a total of nine “data sets”. All were temperate mesoscale catchments (10–10000 km²) with mixed agricultural land use, but varying in their geology, topography and altitude.

The first three catchments, in the Waikato region of New Zealand, were selected from the twenty-six water quality monitoring sites identified by Woodward et al. (2016a), which in turn were a subset of the 114 routinely monitored by the Waikato Regional Council (WRC, 2013). These twenty-six sites have stream flow records available at, or near, the monitoring site. Their geographic locations are shown in Figure 1 of Appendix A (Supplementary Materials). The three catchments were selected to represent mesoscale catchments that contrast in geology, topography, climate and land use. Approximately ten years (2003–2012) of daily stream flow and monthly nitrate data were available at each catchment measurement site, although rainfall and potential evapotranspiration (PET) data had to be obtained from the National Climate Database (cliflo.niwa.co.nz), in some cases from weather stations up to 40 km away.

A fourth catchment, the Weida Stream in Thuringia, Germany, was included to assess the value of high resolution (e.g., daily) nitrate sampling in analysing catchment function. Concentration-discharge relationships for the four catchments are shown in Figure 1, showing the strong positive correlation between nitrate concentration (all concentrations in this paper are mg L⁻¹ NO₃-N) and stream flow in the catchments.

(Figure 1 near here)

2.1.2 Tahunaatara (TA)

The Tahunaatara Stream catchment (TA, 208.1 km², elevation 276-840 m.a.s.l., average 454 m.a.s.l.) is located in the upper Waikato River catchment, within the Lake Taupo volcanic zone. It is dominated by sandy or gravelly, highly porous, pumice soils developed on young volcanic deposits (largely from the Taupo eruption 1.8 ka BP). Podzols occur at greater elevation in forested areas characterised by older volcanic deposits and higher rainfall. The catchment is farmed in the lowlands and includes exotic and native forest in an upland plateau, with 22% dairy, 27% drystock, 29% forestry and 20% native forest overall. Rainfall (average 1258 mm y⁻¹) is measured adjacent to the catchment stream flow gauge at Ohakuri Road (38.337 S, 176.070 E, 276 m.a.s.l., average flow 695 mm y⁻¹) and PET (average 887 mm y⁻¹) is available at the Rotorua Automated Weather Station (AWS) (30.107 S, 176.316 E, 290 m.a.s.l.), 33.4 km away. Nitrate concentrations in Tahunaatara Stream, sampled monthly, ranged from 0.2 to 1.1 mg L⁻¹.

2.1.3 Puniu (PB, PN, PA)

The Puniu River catchment (519.1 km², elevation 29-960 m.a.s.l., average 220 m.a.s.l.) is located within the Waipa River catchment, and has headwaters in the northern Pureora Forest Park. The deep, loamy allophanic soils dominating the catchment have developed on a range of volcanic ashes older than the Taupo deposits (1.8 ka BP). Podzols occur in the headwater area under forest vegetation. The catchment is characterised by farmed lowlands, with 41% dairy, 34% drystock, 2% forestry and 18% native forest overall. Rainfall is measured at Barton's Corner on Pokuru Road (PB, 38.030 S, 175.293 E, 29 m.a.s.l., average rainfall 1186 mm y⁻¹), adjacent to the catchment stream gauge (average flow 915 mm y⁻¹) and also at the settlement of Ngaroma in the hilly uplands (PN, 38.329 S, 175.544 E, 599 m.a.s.l., average rainfall 2297 mm y⁻¹), 39.8 km away. The average of these two rainfall data series was used as an alternative model input (PA). PET (average 827 mm y⁻¹) is

available at Te Kuiti Electronic Weather Station (EWS) (38.334 S, 175.153 E, 80 m.a.s.l.), 35.9 km away. Nitrate concentrations in the Puniu River, sampled monthly, ranged from 0.2 to 2.0 mg L⁻¹.

2.1.4 Mangatangi (MV, MD, MA)

The Mangatangi River catchment (194.5 km², elevation 14-680 m.a.s.l., average 148 m.a.s.l.) is in the lower Waikato River catchment, and includes flows released from a municipal water supply dam in the forested Hunua Ranges (greywacke). Soils in the agriculturally dominated lower parts of the catchment comprise clayey ultic soils developed from strongly weathered greywacke and clayey granular soils derived from strongly weathered volcanic rocks or ash. Land use includes 23% dairy, 35% drystock, 6% forestry and 35% native forest (mainly in the dam catchment). Rainfall is measured at the dam (MD, 37.113 S, 175.208 E, 199 m.a.s.l., average rainfall 1764 mm y⁻¹), which is 15.0 km from the catchment stream gauge at State Highway 2 (37.247 S, 175.185 E, 14 m.a.s.l., average flow 428 mm y⁻¹). Rainfall estimates are also available as a Virtual Climate Station site (VCS, Tait et al., 2012) (MV, 37.275 S, 175.175 E, 42 m.a.s.l., average rainfall 1167 mm y⁻¹), 3.2 km from the catchment outlet. The average of these two rainfall data series was used as an alternative model input (MA). PET (average 884 mm y⁻¹) is available at Pukekohe EWS (37.206 S, 174.863 E, 94 m.a.s.l.), 28.8 km away. Nitrate concentrations, sampled monthly, ranged from 0.0 to 1.5 mg L⁻¹.

Outflow from the dam is controlled by Auckland Council Watercare Services Limited to maintain “environmental flows” throughout the year and is measured at a weir 1 km downstream of the dam (37.172 S, 175.224 E, average flow 289 mm y⁻¹), which represents a subcatchment of area 42.9 km². Because the dam releases result in an artificial flow in the dam subcatchment, only the 151.6 km² area below the dam weir was considered in this study; the catchment outlet flow and nitrate data were adjusted accordingly (assuming negligible nitrate in the dam release water, since it drains an area of native forest). The main effect of this adjustment was to significantly reduce the stream flow during summer, since dam release is relatively high at this time (at least 0.3 m³ s⁻¹). The

error in the flow data, especially at low flows, is consequently much higher, which could result in a poor model fit. To avoid excessive errors resulting from subtracting flows of similar magnitude, dam release flows which were above 40% of the total catchment outlet flow were truncated, prior to being subtracted from the catchment outlet flow data.

2.1.5 Weida (WD, WM)

Finally, the Weida Stream catchment (99.5 km², elevation 357-552 m.a.s.l.) was included as an example of a site with high resolution stream nitrate data and high nitrate loss. The Weida Stream is located in the lower mountain range in the state of Thuringia, Germany, and is monitored by the Thuringian Remote Water Supply. Land use is 40% crops, 29% forest, 26% grassland and 5% urban (Hesser et al., 2010; Jomaa et al., 2016). Stream flow is measured at Lăwitz (50.631 N, 11.919 E, 361 m.a.s.l., average flow 228 mm y⁻¹), including daily stream nitrate data from 31 October 1997 to 31 December 2003 (Hesser et al., 2010; Jomaa et al., 2016), with a range of 0.7 to 18.9 mg L⁻¹. The catchment has been the subject of numerous research studies (e.g., Hesser et al., 2010; Shrestha et al., 2013; Jomaa et al., 2016). Rainfall data for this period was available from five stations; the daily arithmetic mean of these was used in the current paper (average rainfall 640 mm y⁻¹). Potential evapotranspiration (average 567 mm y⁻¹) was also available for a weather station near the catchment outlet.

In order to assess the value of this high resolution (daily) nitrate data (WD), an additional data set was constructed using a monthly sample from the daily time series on the 14th of each month (WM). The range of nitrate concentrations in this subsample was 3.5 to 18.9 mg L⁻¹.

2.1.6 Calibration Data

The main types of calibration data in each data set were daily sampled flow and monthly sampled nitrate (or daily sampled nitrate in the case of WD). In order to maintain an accurate water

balance, monthly total stream flow was added as a third variate (Wöhling et al., 2013), calculated as the sum of the daily sampled values.

Because daily average rainfall and PET rates are used to represent storm events in this study, interpretation of stream water chemistry samples must be done carefully. Storm events typically occur on a time frame shorter than 24 hours, and stream responses to such events also occur rapidly (Brocca et al., 2009; Massari et al., 2014). Daily average rainfall data does not include information about storm timing and intensity, which may be critical for correct interpretation of stream chemistry sample data associated with storm events. We do not know whether the sample was taken before the storm (in which case the sample will not reflect storm effects) or after (in which case it may, although the degree of impact will depend on the exact timing of the sampling relative to the storm signal). Furthermore, the concentration response is likely to be delayed relative to the flow response, which in turn is delayed relative to the storm itself (Ferrant et al., 2013; Shrestha et al., 2013). For this reason, we assumed that stream concentration sample data represent stream water chemistry at the beginning of the sampling day, and hence are unaffected by storm events on the day of sampling.

The data sets are summarised in Table 1.

(Table 1 near here)

2.2 Hydrological Model

2.2.1 Model Concept

The lumped catchment model used in this paper is an advanced version of the three flow path model “StreamGEM” described by Woodward et al. (2013). The three flow paths represented in the model are: near-surface (*N*), representing both event and pre-event water entering the stream in response to rainfall events as overland flow and/or interflow (and artificial drainage if applicable);

fast-response shallow groundwater (F), representing saturated discharge that responds rapidly to recharge; and slow-response deeper groundwater (S), representing flow paths with longer contact times with the subsurface materials. Following Ruiz et al. (2002), the nitrate concentrations of water discharged from each flow path were assumed to be quasi-stationary (the end-member mixing assumption; Christophersen et al., 1990). Flow model parameters and end-member concentrations are estimated by calibration simultaneously to daily flow and monthly (or daily) nitrate data. Figure 2 illustrates the water reservoirs and flow paths represented in the model, and the model parameters, along with their assumed physically appropriate ranges, are listed in Table 2.

(Figure 2 near here)

(Table 2 near here)

Two modifications were made to the original model of Woodward et al. (2013) for the purpose of applying it to larger catchments with potentially more diverse land use, topography, weather and hydrology. First, a parameter was included to allow scaling of point rainfall data to whole catchment rainfall, and second, a more realistic treatment of surface runoff was introduced.

2.2.2 Catchment Rainfall and Potential Evapotranspiration

Because available rainfall data (R) may be available only at a single station, from a site that may not be representative of the catchment as a whole, a calibration parameter (f_R , “rain factor”) was added to the model to adjust for possible bias in the rainfall data, following Vrugt et al. (2008) and Jiang et al. (2014). The importance of suitable data to represent catchment rainfall will be discussed later; for a more comprehensive approach see Li et al. (2012). Catchment average daily rainfall rate (R') was therefore calculated as,

$$R' = f_R R$$

Since potential evapotranspiration (P , typically calculated using the Penman-Monteith method) is much less variable than rainfall in space and time, and random and systematic errors in PET have relatively minor impact on catchment hydrology (Oudin et al., 2006), adjustments for spatial and crop variability effects were considered negligible.

2.2.3 Catchment Direct Runoff

The soil water component of the original StreamGEM model (Woodward et al., 2013) was based on a point scale soil water balance model developed by Woodward et al. (2001). In that model, both infiltration excess (Horton) and saturation excess (Dunne) runoff were represented using thresholds. This approach fails to take account of spatial heterogeneity, which is important for both of these processes at larger scales. Inclusion of spatial heterogeneity requires not only an intensity threshold, but also an estimate of the increasing source area as intensity increases (Sivapalan et al., 1996; Liang and Xie, 2001; Srinivasan and McDowell, 2009).

The success of the well known “SCS Runoff Curve Number Method” (USDA, 1986), which estimates direct runoff for a given rain event, can be attributed to its implicit handling of both threshold and source area within a single empirical equation (Steenhuis et al., 1995; Dahlke et al., 2012). Treating each day’s rainfall as a separate event (Hawkins et al., 2010), the curve number method predicts a daily average direct runoff rate of,

$$R_N = \frac{(R' - \lambda Z)^2}{R' + (1 - \lambda)Z}$$

where R' is the daily average rainfall rate, λZ is the rainfall threshold above which runoff occurs and $(1 + \lambda)Z$ is the maximum infiltration rate. The parameter λ is usually assumed to be 0.2, but following Hawkins et al. (2010), a value of 0.05 was assumed here.

This model can be improved by redefining the parameters, so that,

$$R_N = \frac{(R' - \lambda Z)^2}{R' + (1 - 2\lambda)Z}$$

where R' and λZ are interpreted as before, but the maximum infiltration rate is now Z . The advantage of this formulation is that setting $\lambda = 1$ makes this equivalent to a threshold response. Based on this approach, daily average infiltration rate is $R' - R_N$, which is a curvilinear function with an initial slope of 1, and an asymptote at Z .

Dahlke et al. (2012) showed that both Z and λZ depend on antecedent soil moisture conditions. Analysis of our data suggested reducing Z by a fraction of the soil moisture content on the previous day (W), so that,

$$R_N = \frac{(R' - \lambda Z')^2}{R' + (1 - 2\lambda)Z'}$$

where $Z' = \max(0, Z - k W)$ and the constant $k \text{ d}^{-1}$ is greater than zero. A similar relationship was derived experimentally by Brocca et al. (2009).

The temporal dynamics of near-surface (N) flow are modelled as first order reservoir discharging directly into the stream network with a relative discharge rate $\alpha_N \text{ (d}^{-1}\text{)}$:

$$\frac{dN}{dt} = R_N - D_N$$

$$D_N = \alpha_N N$$

This approach allows some carryover of near-surface flow from one day to the next. As the mean response times for this reservoirs is (α_N^{-1}) is expected to be less than a day (Massari et al., 2014), near-surface discharge (D_N) to the stream is considered as an instantaneous rate rather than a daily average rate, to be compared directly to instantaneous stream flow data.

2.2.4 Catchment Soil Water Drainage

The remaining water percolating through the soil zone was partitioned between bypass flow (Thorburn and Rose, 1990; Radulovich et al., 1992; Phillips, 2010) and soil wetting (McDonnell, 2014). Only soil water is subject to evapotranspiration, as conceptualized in the “two water worlds” hypothesis of McDonnell (2014). Soil water content in the catchment is represented by two storages (W and W_s), based on the tipping bucket model of Woodward et al. (2001). Both storages are expressed in mm, with W representing the total water storage, and W_s representing storage in a rapidly-wetted surface zone with a maximum storage of 25 mm. Inclusion of the latter storage provides improved simulation following rainfall in the dry season, compared with single store models (Scotter et al., 1979; Woodward et al., 2001; Rushton et al., 2006; Horne and Scotter, 2016).

Recharge to soil water (R_w mm d⁻¹), then, is equal to adjusted rainfall (R') less near-surface (R_N) and bypass (R_B) flows,

$$R_B = b(R' - R_N)$$

$$R_w = (1 - b)(R' - R_N)$$

where b is the “bypass” fraction of infiltration, that bypasses the soil matrix and contributes to the soil drainage directly. The rate of actual evapotranspiration from W and W_s (E and E_s , respectively) is determined using the model of Woodward et al. (2001)

Daily drainage rate from the soil zone into the vadose (D_w) is calculated as ,

$$D_w = \max(0, DSW - Y) \times 1 \text{ d}^{-1}$$

where drainable soil water (DSW) is,

$$DSW = W + (R_w - E) \times 1 \text{ d}$$

The available water holding capacity (Y) was determined by calibration. The rate of change of soil water storage is then,

$$\frac{dW}{dt} = R_W - E - D_W$$

with a similar equation for W_s , and the rate of water entering the vadose zone is,

$$R_V = R_B + D_W$$

The release of water from the vadose zone (V) to groundwater recharge (D_V) is modelled as a first order reservoir with relative discharge rate α_V (d^{-1}),

$$\frac{dV}{dt} = R_V - D_V$$

$$D_V = \alpha_V V$$

Since the mean response time for this first order reservoir (α_V^{-1}) may be less than a day (Thorburn and Rose, 1990; Radulovich et al., 1992; Woodward et al., 2016b), vadose zone drainage is modelled as an instantaneous rate. Indeed, at the catchment scale, effective response time is likely to be small, due to the importance of critical source areas (Srinivasan and McDowell, 2009). In order to preserve internal mass balance, daily recharge to groundwater is calculated as the daily average of D_V .

2.2.5 Groundwater Reservoirs

StreamGEM simulates two serially-connected groundwater reservoirs (F and S), which are modelled using eigenvalue-eigenfunction solutions to a linearised Boussinesq equation (Sahuquillo, 1983; Sloan, 2000; Pulido-Velazquez et al., 2005; Bidwell et al., 2008; Woodward et al., 2013). In the solution, each spatially distributed groundwater reservoir is characterised by a single response

parameter (α) represented by a set of parallel first order reservoirs ($i = 1 \dots n$). The groundwater dynamics equations during each stress period are then,

$$\frac{dF_i}{dt} = \beta_{Fi} D_V - \alpha_{Fi} F_i$$

$$\frac{dS_i}{dt} = \beta_{Si} f_S D_F - \alpha_{Si} S_i$$

where F_i and S_i are the eigenmodel storage components, β_{Fi} and β_{Si} are the gain coefficients, and α_{Fi} and α_{Si} are the discharge coefficients, as described in Bidwell et al. (2008). Instantaneous discharge from the fast reservoir (D_F) is then calculated as,

$$D_F = \sum \alpha_{Fi} F_i$$

and is partitioned between discharge to the slow groundwater reservoir, $f_S D_F$, and discharge to the stream network, $(1-f_S) D_F$, as indicated in Figure 2. Instantaneous drainage from the slow groundwater reservoir (D_S) discharges directly into the stream network, and is calculated as:

$$D_S = \sum \alpha_{Si} S_i$$

2.2.6 Stream Discharge

Discharge to the stream (D) is the sum of discharges from the near-surface, fast groundwater and slow groundwater reservoirs:

$$D = D_N + (1 - f_S) D_F + D_S$$

It is assumed that water discharged to the stream network travels rapidly to the catchment outlet. Discharge to the stream can then be compared directly with measured stream flow at the catchment outlet.

2.2.7 Stream Nitrate

Distributed models for catchment nitrate export usually simulate deposition, transformation and transport processes explicitly (e.g., Hesser et al., 2010; Jiang et al., 2014, 2015; Fovet et al., 2015). Despite the considerable data and computational requirements of such models, predictions of in-stream nitrate concentration are often mediocre. Furthermore, the uncertainty of these predictions is usually unable to be assessed. Ruiz et al. (2002), however, showed that the nitrate concentrations representing different flow paths (C_N , C_F and C_S in the current model; Figure 2) may often be considered to be quasi-stationary. While nitrate concentrations at the bottom of the root zone will vary substantially in space and time, various “mixing” processes will result in much more stable concentrations arriving through the different transfer pathways at the stream (Hrachowitz et al., 2016). The degree of dampening will increase with increasing transit time. Accordingly, discharges from short, near-surface pathways are likely to have the most dynamic nitrate concentrations, while discharges from the slow, deep groundwater pathway generally have the most stable concentrations. These features are reflected in the concentration-discharge relationships (Figure 1). Stream sampling has shown that the groundwater sustaining baseflow had mean transfer times ranging from 6 years in the Mangatangi catchment to 16 years at Puniu and 34 years at Tahunaatara (WRC, 2015). These values indicate that assuming quasi-constant nitrate concentrations is an acceptable simplification in the catchments concerned.

It is not required for our approach that concentrations differ between N , F , and S in a pre-determined particular order. However, the more distinct the concentration differences are, the more reliable can pathway contributions be estimated. The approach also requires that the end-member concentrations remain constant over time. This is more likely to be valid using a short calibration period relative to land use or climate changes. In the current study, the calibration period is short (4 years; Section 2.3.4), and nitrate trends have been shown to be negligible (Hesser et al.,

2010; WRC, 2013). Furthermore the validity of this and other model assumptions are validated against data from a separate time period, and the residuals are evaluated visually (Section 3.3).

In-stream concentrations of nitrate (C) at the catchment outlet, therefore, are calculated using a mixing assumption, dividing the total yield (i.e., the sum of discharge concentration multiplied by discharge rate) by stream flow (D):

$$C = \frac{C_N D_N + C_F (1 - f_S) D_F + C_S D_S}{D}$$

Attenuation of nitrate in the hyporheic zone or in-stream cannot be explicitly modelled on the basis of the data used, but is implicitly included in the concentrations assigned to each flow path.

2.3 Model Calibration

2.3.1 Sources of Uncertainty

The Bayesian methods used in this study provide tools for handling the various sources of uncertainty inherent in the modelling process. These include errors in the model input/forcing data (e.g., rainfall, PET), errors in the model structure (e.g., functional forms, fixed parameter values), and errors in the calibration data (e.g., stream flow, concentration). These errors are unknown; but modelling proceeds under the assumption that they do not preclude estimation of the desired catchment properties.

Errors in the model input data are frequently a large source of uncertainty (Huard and Mailhot, 2006; Li et al., 2012). Precipitation data is particularly problematic since it is the main forcing function for catchment hydrology and spatial distributions or averages over relatively large areas have to be derived from point measurements, which are themselves uncertain. This situation is exacerbated by sparse collection networks in rural areas of interest, and inherent nonlinearities in catchment function (and catchment models), which mean that model predictions may be

significantly biased due to the forcing data used, let alone other sources of error. The importance of input data errors is examined in the current study by comparing results derived under three different rainfall inputs for two of the catchments (Puniu and Mangatangi), and by including a calibration parameter to adjust for rainfall bias (f_R).

Errors in the model structure (Beven, 2005) may be examined directly by comparing alternative models (e.g., Hartmann et al., 2013; Ruiz-Pérez et al., 2016). This approach was not attempted in the current study, however. Instead, an attempt was made to minimise structural error by ensuring that the model was (1) consistent with hydrological process knowledge, (2) parsimonious with the catchment data, and (3) flexible in its parameterisation. The success of this approach was then assessed by post-calibration analysis of the model residuals.

Errors in the calibration data are usually estimated on the basis of measurement methodology. Measurement errors for stream flow and concentration data were reviewed by McMillan et al. (2012), who suggested errors for rating-curve-estimated streamflow of up to 20% at mid-high flow ranges, and up to 100% at low flow ranges, and of around 5% for inorganic nitrogen concentrations circa 1 mg L^{-1} . It was not possible to use these estimates directly in our study, as the impact of model input and structural errors means that the calibration residuals are much larger than the measurement errors alone. McMillan et al. (2012) noted, for example, that point rainfall data could have standard errors as large as 65% when compared with the mean from gauges distributed over an area of 135 km^2 .

The final source of model error is the uncertainty of the model parameters (including the initialisation values of the model state variables); this is inferred during the Bayesian calibration process.

2.3.2 Bayesian Parameter Inference

Bayesian parameter inference methods provide the means to estimate a vector of model parameters (\mathbf{x}) on the basis of a given vector of real world observations (\mathbf{Y}). In particular, Bayes' theorem states that the conditional probability density of \mathbf{x} , $p(\mathbf{x}|\mathbf{Y})$, called the “posterior”, is proportional to the conditional probability density of \mathbf{Y} , $p(\mathbf{Y}|\mathbf{x})$, called the “likelihood”.

Mathematically,

$$p(\mathbf{x}|\mathbf{Y}) = \frac{p(\mathbf{Y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{Y})}$$

where $p(\mathbf{x})$ is called the “prior”, and represents any prior knowledge about the parameters, and $p(\mathbf{Y})$ is called the “evidence”, which is constant over \mathbf{x} (Vrugt, 2016). In application, the analyst provides the prior parameter distribution $p(\mathbf{x})$ and defines the likelihood function $p(\mathbf{Y}|\mathbf{x})$, and a numerical method is then used to search the parameter space in order to infer the posterior parameter distribution $p(\mathbf{x}|\mathbf{Y})$. In the current study, the priors were assumed to be uniform.

2.3.3 Likelihood Function

The misfit between the model simulations and the calibration data (i.e., the set of “model residuals”) is typically assumed to be due to observational errors in the calibration data alone. This misfit is quantified in a likelihood function, $p(\mathbf{Y}|\mathbf{x})$, which represents the probability of a given set of residuals occurring by chance, given a certain model, inputs and parameterization. The most probable values of the model parameters (i.e., the posterior distributions of the parameters) are then derived on this basis.

In reality, as described above, the residuals are affected by errors in the forcing data and in the model structure, and cannot be characterised simply as measurements errors in the calibration data. The form of the likelihood function is therefore necessarily subjective. In addition, the residuals from hydrological model calibration are highly unlikely to be statistically well-behaved (i.e.,

independent, identically distributed, Gaussian). Since the likelihood function is therefore unavoidably incorrect, bias in the posterior parameter distributions and other predictions is also possible (Beven, 2005; He et al., 2010; Schoups and Vrugt, 2010; Li et al., 2012).

As far as possible, the likelihood function should consider the probability density of each residual (e.g., standard error, functional form, heteroscedasticity), as well as correlations between the residuals (e.g., autocorrelation within data sets, correlation between data sets) (Schoups and Vrugt, 2010; He et al., 2010). Various forms of likelihood functions can be constructed. Sorooshian and Dracup (1980) and Schoups and Vrugt (2010) suggest that adequate representation of the actual statistical distribution of the residual errors will yield tighter predictive error bands and more accurate parameter uncertainty estimates that are less sensitive to the calibration data used (i.e., less prone to overfitting). In the current study a likelihood function (specified as a log-likelihood, LL) was constructed assuming uncorrelated, non-Gaussian, heteroscedastic errors. The log-likelihoods of the three calibration data types (daily flow, monthly or daily nitrate, monthly flow) were combined by addition (method of Bayes' multiplication, He et al., 2010), and these were scaled to adjust for the different numbers of residuals of each type. The log-likelihood to be maximised was therefore the sum of the log-likelihoods of the daily discharge residuals (LL_D), concentration residuals (LL_C), and monthly flow residuals (LL_M), each scaled by the number of data points relative to the number of monthly flow residuals (n_M):

$$LL = \frac{n_M}{n_D} LL_D + \frac{n_M}{n_C} LL_C + LL_M$$

$$LL_D = \sum_{j=1}^{n_D} \ln p \left(\frac{\hat{D}_j - D_j}{\sigma_{D_j}}, v \right)$$

$$LL_C = \sum_{j=1}^{n_C} \ln p \left(\frac{\hat{C}_j - C_j}{\sigma_{C_j}}, v \right)$$

$$LL_M = \sum_{j=1}^{n_M} \ln p\left(\frac{\hat{M}_j - M_j}{\sigma_{M_j}}, \nu\right)$$

$$p(t, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi} \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

where discharge and concentration observations are indicated by a hat symbol ($\hat{\cdot}$), σ_{D_j} is the assumed standard error (for model prediction D_j , for example), and $p(t, \nu)$ is the probability density of a Student's t -distribution with ν degrees of freedom. Although the assumption of Gaussian residuals has performed well in many case studies, and results in efficient convergence of search algorithms (e.g., He et al., 2010), it may produce biased estimates in heavy-tailed data sets (Schoups and Vrugt, 2010) or if outliers exist (Motulsky and Brown, 2006). The Student's t -distribution, on the other hand, with a small degrees of freedom (ν), is less sensitive to large residuals (this is illustrated in Figure 2 of Appendix A: Supplementary Materials). A value of $\nu = 7$ was used for all variables in the current study, based on Jeffreys' arguments regarding the measurement error distributions in real data generally (Jeffreys, 1938), and inspection of the residuals in this study, as discussed in Section 3.2.

The assumed standard error was 20% for all data types. By assuming this error model we are, in a way, hypothesising that the model represents changes in the observed variables equally well across all scales. For comparison, Schoups and Vrugt (2010) estimated standard errors of daily flow of approximately 10% and 33%, respectively, for two rivers in the United States.

Correlation between the residuals is more difficult to account for. Corrections for autocorrelation are possible (Schoups and Vrugt, 2010) but do not work well in practice for hydrological time series, where the correlation is structural rather than random. In the current study the daily flow residuals (and daily nitrate residuals in WD) are approximately 30 times more numerous than the monthly observation types, but are highly autocorrelated. Scaling the log-

likelihoods by the number of observations compensates somewhat for this structure. Essentially, a single “average” probability is calculated for each month, and these are assumed to be independent between months.

2.3.4 Computation

Based on this log-likelihood function, the posterior (most likely) distributions of the parameter values were estimated for each catchment using a Matlab implementation of the DREAM_{zs} Markov Chain Monte Carlo algorithm of ter Braak and Vrugt (2008). DREAM_{zs} is a refinement of the DREAM differential evolution adaptive Metropolis sampler described in Vrugt et al. (2008) which stores and exploits information from past samples to greatly improve computational efficiency, particularly by reducing the number of sampling chains.

The first 90 days of data were considered as a warm-up period (c.f., Li et al., 2012, who used a 30 day warm-up period), and the following 1461 days (4 years) were used for model calibration (Table 1). Ninety days was considered a sufficient warm-up period for the model, as most state variables change much more rapidly than this. The exceptions are total soil water (W) and slow groundwater storage (S). The initial values of W and the slowest component of S (S_1) were therefore fitted as additional model parameters (not reported here).

A four-year calibration period proved sufficient (and was available) for all the data sets, and matched that used in an earlier study (Woodward et al., 2013). The remaining data was used for model validation (Table 1).

DREAM_{zs} convergence was checked by examining the Gelman-Rubin statistic (Vrugt, 2016) as well as the 3 chains for each parameter. A run length of 50000 generations was sufficient for apparent convergence in all cases. The last 10000 samples were taken to represent the posterior parameter distributions.

3 Results and Discussion

3.1 Posterior Log-Likelihoods

In order to avoid parameter sets whose log-likelihood was extremely low, only posterior parameter sets in top 95% of log-likelihoods were shown in the plots of the posterior analysis. The distribution of log-likelihoods of the remaining posterior parameter sets are shown in Figure 3. The box and whisker plots therefore show the median, quartiles, and 95% range. The results show narrow ranges for all data sets, consistent with convergence of the MCMC procedure. Notably, the maximum log-likelihood parameter sets (i.e., maximum likelihood parameter sets, indicated by crosses on the plots) perform well for all data types.

Differences in log-likelihood between the Puniu River and Mangatangi River data sets shows the effect of input error on model calibration: at both sites, the best fits to the catchment data (indicated by relatively high LL values) were obtained by using the average of the catchment outlet and catchment headwater rain gauges. Furthermore, in both of these catchments, a better fit was achieved using the headwater rainfall data (PN and MD) compared with the catchment outlet rainfall data (PB and MV) respectively. This appears to be due to an orographic effect; lowland rain gauges may underrepresent (or even miss) certain storms compared with upland gauges (Corradini, 1985).

Differences in log-likelihood between the Weida Stream data sets cannot be interpreted directly, as they are calculated using different subsets of the available nitrate data. Nevertheless the calibrations to WD and WM are for all intents and purpose identical; StreamGEM reproduced the daily nitrate data set (WD) just as well as the monthly data set (WM), indicating that minimal overfitting occurred in the WM calibration. This result is discussed in more detail in Sections 3.5 and 3.6.

(Figure 3 near here)

3.2 Residual Analysis

The likelihood function describes the goodness of fit between data and model simulations resulting from a single candidate parameter set (i.e., the residuals) and summarises it into a single log-likelihood value. It is also instructive to examine the structure of the residuals visually (Xu, 2001; Schoups and Vrugt, 2010; Li et al., 2012; Gelman and Shalizi, 2013; Jiang et al., 2014), allowing us to ask “questions such as whether simulations from the fitted model resemble the original data, whether the fitted model is consistent with other data not used in the fitting of the model, and whether variables that the model says are noise (“error terms”) in fact display readily-detectable patterns” (Gelman and Shalizi, 2013).

Figure 4 illustrates the distribution of flow and nitrate model residuals during the calibration and validation periods for a single parameter set (the maximum likelihood parameter set from the Puniu River with rainfall from Barton’s Corner data set (PB) calibration). Since a heteroscedastic error model was assumed, the residuals were scaled by the standard error and expressed as z-values, as in Section 2.3.3. Figure 4 includes both the hydrograph and chemograph fits during the calibration and validation periods (note the log-scale), model-data plots for the three calibration variables (daily flow, monthly nitrate, monthly total flow), scatter plots of model residuals (as z-values) against the model values (which is useful to assess dependence of the residuals on the model outputs), histograms of the model residuals (which is useful to assess the suitability of the assumed log-likelihood function), the autocorrelation of the daily flow residuals (note that the autocorrelation declines to zero at a lag of approximately 30 days) and the data and model concentration-discharge plots, superimposed on one another (which highlight certain salient features of the data as explained below).

This simulation produces well-behaved residuals overall, in both the calibration and validation periods, although some remaining structure can be observed in the daily flow residuals. In

particular we notice the following three features, which are typical in most of our calibrations: large negative residuals for large simulated flows, large positive residuals at low-medium simulated flows, and one-tailed, positively skewed residuals at low simulated flows. Similar patterns can be observed in the standard least squares fits of Xu (2001), Schoups and Vrugt (2010) and Jiang et al. (2014).

The first two features can be traced back to differences in timing between rainfall events and flow events in the data. In cases where daily storm rainfall is not reflected in a concomitant stream flow increase on the same day (e.g., because of rainfall overestimation, or unmodelled delays in the surface or subsurface flow), a large negative residual occurs at high model flow. Conversely, where a stream flow increase occurs but the associated rainfall event is not recorded or is underestimated, a large positive residual occurs, at low-medium model flow. These discrepancies can therefore be attributed to the shortcomings of using daily, lumped data and modelling as the basis for simulating the highly dynamic hydrological processes in these small-to-medium scale catchments. Much higher resolution data and modelling would be required to resolve these issues, which is beyond the scope of the current work, whose focus is on utilising commonly available information with relatively poor resolution.

The third feature of the daily flow residuals, the one-tailed, positive skew of residuals at low flows, can be attributed to the nature of the recession portion of the hydrograph. Recession curves in real catchments are more complex than can be represented in any lumped model, being the superposition of drainage contributions from multiple land areas with different recession characteristics and which may have received different amounts of recharge. This means that each event can have different recession characteristics, and a lumped model with a small number of subsurface reservoirs will be unlikely to fit every event well. Calibration will tend to suggest a model that is somewhat “average”, one consequence of which is that it will tend to overpredict flow during the periods of lowest flow (e.g., Wöhling et al., 2013). This results in a clustering of negative

residuals at low flows, which are small in magnitude (reflecting the relatively small variation in model and observed flows during the recession portion of the hydrograph). Any variation then tends to occur only in the positive direction, due to stream responses which are not captured by the model.

The observed biases in the daily flow residuals may therefore be unavoidable in any daily, spatially-lumped modelling. Since our focus is on estimating the relative magnitudes of water and nitrate fluxes discharging along separate flow paths, this bias in the fine detail of the flow predictions will be accepted as an inevitable price of simplification.

In contrast to the daily flow residuals, the monthly nitrate and monthly flow residuals generally have a much more ideal structure, appearing independent, with constant variance, and small bias and variance.

(Figure 4 near here)

3.3 Goodness of Fit Check

As well as visual inspecting sets of residuals, the posterior parameter distributions derived from MCMC calibration were checked by making model predictions for the same catchment for a different time period (split-sample validation). Comparison of the distribution of residuals for the calibration and validation data sets provided an indication of the predictive performance of the calibrated model (Gelman and Shalizi, 2013). If the validation residuals are much worse than the calibration residuals, this indicates overfitting of the model to the calibration data, or that the system is not stationary (i.e., its parameters change over time).

While it is theoretically possible to carry out the analysis in Figure 4 for every posterior parameter set, it is more practical to choose suitable summary statistics for this purpose.

Unfortunately, almost all commonly used goodness-of-fit statistics assume constant error variance

(Moriassi et al., 2007), and so are unsuitable for use here. However, because relative errors were assumed for all variables in this study, fits may be assessed in log space. The commonly-used Nash-Sutcliffe model efficiency (NSE ; Nash and Sutcliffe, 1970),

$$NSE_D^{Calib} = 1 - \frac{\sum_{j=1}^{n_D} (\hat{D}_j - D_j)^2}{\sum_{j=1}^{n_D} (\hat{D}_j - \bar{D})^2}$$

was therefore calculated in log space (NSL ; Krause et al., 2005; Wöhling et al., 2013),

$$NSL_D^{Calib} = 1 - \frac{\sum_{j=1}^{n_D} (\ln(\hat{D}_j) - \ln(D_j))^2}{\sum_{j=1}^{n_D} (\ln(\hat{D}_j) - \ln(\bar{D}))^2}$$

to assess the model fit for each data type during the calibration and validation periods (Table 1). NSL therefore represents the proportion of variation explained by the model in the second row of charts in Figure 4. As with NSE , the range of the NSL is between $-\infty$ and 1, and its optimal value is 1.

Figure 5 shows the calculated NSL values. The NSL for the calibration period were generally consistent and well constrained across all data sets. Flow calibrations were particularly good, with calibration NSL in the range 0.62–0.83 for daily flow and 0.68–0.93 for monthly flow. NSL for nitrate concentration was somewhat lower, in the range 0.17–0.88; the lower values reflect the narrow range of concentration data at some sites (e.g., Tahunaatara, Puniu) rather than increased model prediction error.

The NSL for the validation period assesses the model's ability to predict catchment behaviour over a separate time period. Ideally the residuals should have a similar magnitude in the validation period compared with the calibration period. In Figure 5 we see that the NSL for the validation period was also generally good. Again, flow fits were the best, with NSL in the range 0.60–0.89 for daily flow and 0.64–0.95 for monthly flow. The similarity of these values with the calibration fits suggests that the flow model was optimally fitted; overfitting in the calibration period would

have been reflected in relatively poorer predictions in the validation period. NSL for nitrate was again lower, in the range -0.39–0.74. The lower values for WM are due to the variation and small number of data points in the validation period (20 samples; Table 1); WD had similar model parameter values (Section 3.4) but achieved much better validation NSL due to the larger validation data set (566 samples).

These fits compare favourably with those from other studies. The calibrated best fit (maximum likelihood) NSL values in the current study were in the range 0.69–0.81 for daily flow and 0.63–0.84 for monthly or daily nitrate. These can be compared with NSE values from studies based on a homoscedastic (constant variance) error model. Jiang et al. (2015), for example, using a similar approach in the Selke catchment in Germany, reported best fit NSE values in the range 0.86–0.90 for daily flow and -0.42–0.70 for biweekly inorganic nitrogen; this suggests that a better trade-off between flow and nitrate fit was achieved in the current study. Hesser et al. (2010) similarly reported relatively low NSE of 0.36 during calibration and 0.22 during validation for nitrate concentration in the Weida catchment. It might be expected that the complex, distributed models used in these two studies would allow excellent calibration to the observed data. In practice, the use of highly parameterised, complex models precludes a thorough search of the parameter space, resulting in poorer results than those that can be achieved using a parsimonious approach such as the one used in the current study.

(Figure 5 near here)

3.4 Parameter Uncertainty

The output of the Bayesian calibration with DREAM₂₅ is the joint posterior distribution of the model parameters. Following common convenience, we base our analysis of parameter uncertainty on the marginal parameter distributions which are shown in Figure 6.

Different model parameters have different uncertainty ranges (indicated by the length of the box and whiskers plot), and in some cases these ranges differ between data sets. Due to the interdependent structure of the model, the posterior distributions of any two parameters may also be correlated. In this case, the uncertainties in Figure 6 are somewhat overstated, as fixing or redefining the parameters can give the same results with less parameter variation. Although the DREAM_{zS} code does provide correlation analysis of the posterior parameters sets, analysing these correlations is complex, and only the uncertainties of individual parameters will be presented here.

As expected, the rainfall scaling parameter f_R varied between data sets, with values between 0.56 and 1.60. Values for each data set, however, were well defined. This parameter corrects for rainfall location and other catchment water balance errors, so it is expected to be accurately determined.

Surface runoff parameters (Z , k , α_N) were relatively uncertain, reflecting the difficulty in accurately capturing storm dynamics in a simple model based on daily rainfall and stream flow data. Near-surface flow rate (α_N) was constrained *a priori* to be between 0.7 and 2.1 d⁻¹ to ensure rapid discharge of NS flow. Without this constraint, values of α_N were extremely variable and other parameters were variable also. The effect of antecedent moisture conditions on near-surface runoff (k) was more pronounced in the steeper catchments with less subsurface storage (Mangatangi, Weida). Maximum infiltration (Z) was highly correlated with k at most sites (analysis not shown). The greater uncertainty in k and Z in catchments with relatively high baseflow (Tahunaatara, Puniu) probably reflects the relatively rarity of significant direct runoff in these catchments: however the maximum likelihood values of these parameters appear to be relatively consistent with the values for the steeper, low baseflow catchments.

The uncertainty of unsaturated zone parameters (b , Y , α_v) differed between catchments with high baseflow (Tahunaatara and Puniu) and those with low baseflow (Mangatangi and Weida).

The proportion of bypass flow (b) was uncertain for Tahunaatara and Puniu, but consistently low at the other sites. This parameter is important for accurate representation of streamflow response at low flow ranges when there is little direct runoff. Soil water holding capacity (Y) was also relatively poorly identified in these two catchments, although values were within the ranges observed in field scale studies (Horne and Scotter, 2016). This parameter controls the timing and degree of groundwater recharge, which is not directly observed in the calibration data. Schoups and Vrugt (2010) similarly found the soil water holding capacity parameter to be poorly identified (unrealistically high) for one of the rivers in their study. The low values of Y estimated for Mangatangi and Weida may reflect the relatively high importance of steeper land in these catchments. The vadose zone discharge parameter (α_v) was completely insensitive. Although this parameter was retained in the model for physical reasons, vadose zone passage appears to have little effect on the stream hydrograph.

Partitioning of fast groundwater to deep percolation versus stream discharge (f_s) varied from catchment to catchment. This parameter is often positively correlated with fast groundwater nitrate-N concentration (since it reduces the amount of fast groundwater discharge, C_F increases to compensate) and slow groundwater discharge rate. The highest values of f_s are at Mangatangi, for example, and are associated with high and uncertain values of C_F at that site.

Fast and slow groundwater discharge parameters (α_F and α_S), on the other hand, were quite well defined for all data sets. Values for α_F ranged from 0.010 to 0.052 d⁻¹ (corresponding to hydrodynamic response times of 6.4-33.3 days). Values for α_S ranged from 0.000097 to 0.0040 d⁻¹ for the Waikato catchments (corresponding to hydrodynamic response times of 83-3400 days), but were much slower for Weida, ranging from 0.000013 to 0.00048 (700 to 25000 days). The Waikato results are similar to the response times of 3.2-21 d and 21-417 d for fast and slow flow paths across a range of catchments derived by Gupta et al. (2009), 32 d and 6600 d derived by Ruiz et al. (2002),

and 40 d and 127 d by Fovet et al. (2015), using linear models. The low values and large uncertainty of α_s for Weida (Figure 6) may reflect the low hydraulic conductivity of subsurface materials (Hesser et al., 2010) and small contribution of slow groundwater to stream flow (Section 3.5) in this catchment.

As expected, nitrate concentrations in fast and slow groundwater discharge (C_F and C_S) were well defined for most sites, being closely related to the maximum and minimum values of stream nitrate-N concentration. The high values and uncertainty in C_F at Mangatangi are related to the low predicted discharge of F in this catchment (Section 3.5). Near-surface nitrate concentration (C_N), which was intermediate in magnitude for these catchments, was well defined at most sites. Nitrate concentrations from the Weida catchment (which are scaled by a factor of 10 in Figure 6), ranged from 4.6-13.9 for C_N , 12.0-17.1 for C_F , and 3.7-6.8 for C_S for the WD data set. These values correspond well with estimates derived by Hesser et al. (2010) of 10-25 for interflow and fast baseflow and 3-6.5 for slow baseflow (which had a measured range of 4.1-8.6 mg L⁻¹, the highest values being observed in spring and the lowest in summer). These results contrast with those of Woodward et al. (2013) for the small (15 km²) Toenepi Stream dairy catchment, where the highest nitrate concentrations were observed in storm flow, resulting in high values of C_N , possibly reflecting a relative importance of high nitrate interflow and/or artificial drainage in that catchment, compared with the catchments in the current study where surface runoff could be more important (with more dilute nitrate concentrations).

(Figure 6 near here)

3.5 Predictive Uncertainty

There are several types of predictions of interest (in this context, predictions are primarily hindcasts, rather than forecasts). The first type are the values of physically meaningful parameters, such as the end member concentrations. These have already been discussed in Section 3.4.

The second type of predictions of interest are the high resolution hydrograph and chemograph predictions, examples of which are presented in Figure 7 for the Weida Stream catchment. Although hydrographs are typically presented on a linear scale (e.g., Schoups and Vrugt, 2010; Jiang et al., 2015), presenting the data on a log scale gives considerably more insight into behaviour across the entire flow range (e.g., Woodward et al., 2013; Ruiz-Pérez et al., 2016). Discharge modelled with StreamGEM generally followed the data very well in this and the other three catchments; minor discrepancies may be due to rainfall errors or unmodelled spatial variability.

The high resolution chemograph, on the other hand, offers insights into the dynamics of in-stream nitrate, and the representativeness of the monthly data. It is clear from the daily nitrate data in Figure 7b that nitrate may be much more dynamic than is captured by the monthly samples in Figure 7a. Wade et al. (2012) similarly showed that low resolution nitrate samples do not capture many features that only become apparent in high resolution sampling. The StreamGEM model, however, even when calibrated to the low resolution data as in Figure 7a, predicted similar dynamics to those observed in the high resolution data (Figure 7b). The implication is that the calibrated model may represent stream nitrate dynamics much more accurately than apparent from monthly sampling data, and that model calibration using monthly sampling data may in fact be adequate for the kind of dynamic catchment flux analysis proposed in this paper.

(Figure 7 near here)

The third kind of predictions of interest are the flow and nitrate fluxes attributable to the three flow paths: near-surface discharge, fast groundwater discharge and slow groundwater discharge. Typical dynamics of modelled streamflow contributions are illustrated in Figure 8; slow groundwater discharge is relatively stable through the year, fast groundwater discharge is seasonal, being dominant in the winter, and near-surface flow is associated with storm events. The dynamics of nitrate yield look somewhat similar to the flow dynamics, as in StreamGEM, the nitrate concentrations associated with each flow path are constant with time.

Averaging the flow and nitrate discharge from each flow path over time then allows us to estimate annual average proportions and totals. The resulting predictions of catchment fluxes are shown in Figure 9, with the box and whisker plots representing the posterior parameter uncertainty. As expected, the total annual flow and nitrate yield predictions in Figure 9 are tightly constrained for all data sets, being largely determined by the calibration data. Average annual flow and nitrate yield estimates were also calculated directly from the data (the latter using the Beale Ratio Estimator method, Beale, 1962) and are shown also in Figure 9 for comparison. The range of model predictions are consistent with these empirical estimates.

The contributions of flow and nitrate from each flow path however are less certain, as the calibration data contains this information only indirectly (e.g., in the curvature of post-storm recession or the ratio of concentration and flow). The predicted proportions of slow groundwater discharge in Figure 9a broadly match expectations based on the catchment geology; catchments with relatively deep, porous subsurface materials discharged considerably more slow groundwater (e.g., Tahunaatara 49-81%) than catchments with relatively shallow, impermeable materials (i.e., Weida 17-28%), with Puniu (24-43%) and Mangatangi (29-46%) being intermediate in both respects (both these catchments include significant lowlands as well as hilly uplands). In all cases slow groundwater contribution was greater than the 9-18% estimated for the small Toenepi Stream

catchment in previous work (Woodward et al., 2013), as expected, due to the very low effective storage in that catchment. Modelled slow flow for Weida is somewhat higher than the <16% estimated by Hesser et al. (2010); the discrepancy may be due to the different calibration criteria used (i.e., constant vs relative standard errors), which would be expected to result in more accurate modelling of low flows in the current study.

Discharge of shallower, faster groundwater generally showed the opposite pattern, being lower in lowland catchments (e.g., Tahunaatara 10-43%) and higher in steeper catchments (Weida 40-73%). However, relatively low seasonal (fast groundwater) flow was inferred for the Mangatangi catchment (6-26%), where near-surface flow was predicted to be more important on a percentage basis (37-59%). This could be due to near-surface flows being encouraged both in the steep, forested upland parts of the catchment as well as in the largely clay-textured lowland parts. The dominance of near-surface flow paths in this catchment is consistent with tritium-based mean residence time (MRT) estimates for baseflow reported by WRC (2015), being only 6 years for Mangatangi, compared with 16 years for Puniu and 34 years for Tahunaatara.

Nitrate yield from slow groundwater was generally low across all catchments (Figure 9b), reflecting low nitrate concentrations (C_s , Figure 6). The larger slow groundwater nitrate contributions predicted for Tahunaatara (18-63%) reflect the relatively high discharge of nitrate-bearing groundwater in that catchment (although concentrations are still low by international standards). The bulk of nitrate flux in all catchments was attributed to fast groundwater discharge (31-97%), as this flowpath has both high concentration and high water flux. This confirmed the result derived by Hesser et al. (2010) for the Weida Stream. Near-surface runoff was a major contributor only at Mangatangi (24-63%), being generally low in the other catchments (0-38%).

(Figure 8 near here)

(Figure 9 near here)

3.6 Analysis of Model Structure and Modelling Approach

The success of the StreamGEM model in reproducing the flow and concentration dynamics in a variety of catchments at this scale (100-520 km²), including catchments with significant baseflow (Tahunaatara) and catchments with minimal subsurface flow (Mangatangi, Weida), is striking. While smaller catchments might be expected to behave in relatively homogenous way (e.g., Woodward et al., 2013), the spatial and temporal variation inevitable at larger scales might be expected to result in much less well-defined patterns of response, which would be more difficult to fit with a non-spatially explicit model. Many authors have therefore used spatially explicit models at this scale (e.g., Hesser et al., 2010; Li et al., 2012; Bieger et al., 2012; Shrestha et al., 2013; Jiang et al., 2014; Ruiz-Pérez et al., 2016). However, rainfall-runoff responses appear to still be sufficiently clear to allow calibration of a simple lumped model. Furthermore, the derived parameter values and model predictions are generally consistent with known catchment geology and land use features.

Use of the eigenmodel representation of groundwater discharge also appears successful, although it was not compared against the linear store model used by other authors (e.g., Ruiz et al., 2002; Gupta et al., 2009; Schoups and Vrugt, 2010; Hartmann et al., 2013). More general forms of the eigenmodel are possible, that include bedrock slope and hillslope shape (Pauwels et al., 2002; Huyck et al., 2005), but it is unlikely that the data would support additional parameterisation, making these less attractive than the single-parameter version used here.

The model modifications appear to have been beneficial: the inclusion of a rainfall scaling factor (f_R) appears to be crucial, particularly when a single rainfall gauge is used. The importance of using representative rainfall data was also highlighted, as lowland rain gauges may provide a poor estimate of catchment rainfall.

The increased detail in the direct-runoff model (Section 2.2.3) similarly allowed more realistic near-surface flow estimates compared with the previous model (Woodward et al., 2013),

despite the fact that the near-surface parameters were frequently rather uncertain (Figure 6). The decision to ignore in-stream uptake of nitrate and delays due to surface water routing also appears to be reasonable for the purpose of modelling fitting. In reality in-stream nitrate uptake is likely to be most significant during periods of low flow and warm temperatures (Rode et al., 2016); this implies that the nitrate concentration in slow groundwater discharge might be higher than the calibrated C_s parameter in our model.

Bayesian inference offered several significant advantages over traditional (e.g., nonlinear least squares) model calibration approaches. Because traditional approaches are not embedded in a rigorous statistical framework, they yield only a single “best” parameter set, and lack a basis for characterising the uncertainty of the solution. By contrast, Bayes’ theorem incorporates prior information about the parameter values as well as uncertainty in the observations; this allows the full joint probability distribution for the parameters to be characterised, elegantly handling and highlighting model parameters that are poorly identified or correlated with other parameters. Within the Bayesian framework, Markov Chain Monte Carlo algorithms were developed to provide a computationally efficient means for exploring the parameter space. MCMC calibration nevertheless requires a large number of model runs, and so can only be feasibly applied for fast models with modest numbers of parameters, which generally precludes its use with distributed models. It is ideal, however, for application to lumped models such as StreamGEM.

Use of relative errors to assess the flow and nitrate fits seemed to work well, and allowed the same error model to be applied across a wide range of catchments. Weighting between the flow and nitrate data sets was done in an ad hoc way, however, which may be able to be improved in future work. In particular, the high resolution nitrate data at Weida may have been too heavily down-weighted, as the additional information provided by this data did not reduce the uncertainty of the model predictions.

Compared with other modelling approaches, the predictions derived in this study appear to have very wide uncertainty ranges. Even then, the ranges shown in the box and whisker plots reflect only the posterior parameter uncertainty (omitting the 5% of parameter sets with the lowest log-likelihood values). Model input and model structural uncertainty remain, as well as possible bias due to the non-ideal error propagation and structure in the model and data; these uncertainties are illustrated here only in Figure 7. The estimation and reporting of uncertainty should not be considered a weakness, however. As well as tempering the temptation to overstate the certainty of our modelling results, uncertainty analysis provides insights into the models, data and catchments of interest that are not available with traditional approaches. Given that the tools are increasingly available to the practitioner (Refsgaard et al., 2007; Matott et al., 2009; Wu and Zeng, 2013), it can only benefit hydrological science to apply them more widely and consistently (Juston et al., 2013; Andréassian et al., 2013).

4 Conclusions

The catchment modelling approach described and applied in this paper shows that mechanisms of flow and nitrate discharge can be inferred from daily stream flow and monthly nitrate data. A simple lumped process model, assuming constant flow path nitrate concentrations in near-surface, fast groundwater and slow groundwater flow paths, allows water and nitrate flux dynamics to be predicted, with robust estimates of uncertainty based on Bayesian inference.

The analysis highlighted the bias inherent in using a single rainfall gauge to represent catchment rainfall, and showed that this can be corrected by including an adjustable rainfall factor (Vrugt et al., 2008) and/or by providing spatially averaged rainfall data (Li et al., 2012). The availability of daily stream nitrate data at one site appeared to provide no advantage over the monthly stream samples such as are commonly available from routine environmental monitoring

programmes, although this may have been due to its low weighting in the composite log-likelihood function.

The approach allowed estimation of nitrate concentrations of the near-surface, fast groundwater and slow groundwater flow path end members. Fast groundwater had the highest nitrate concentrations in all of the catchments modelled here, and slow groundwater the lowest. Near-surface flow concentration were intermediate; this meant that summer storms in these catchments tended to result in increased stream nitrate concentration, whereas winter storms tended to result in decreased concentrations, a dynamic that has recently been highlighted by Dupas et al. (2016).

The predicted water and nitrate flux components were consistent with knowledge of catchment geology, yielding relatively greater groundwater flux in catchments with greater subsurface storage, and conversely, greater near-surface runoff in steeper catchments with lower storage. Annual nitrate yields calculated using the model were consistent with those calculated using a common ratio estimator method. The proposed approach was therefore demonstrated to be highly valuable for practitioners as it is a relatively simple, computationally efficient method for predicting highly dynamic nutrient fluxes in mesoscale catchments which also provides accurate estimates of catchment nutrient yields without the need to increase sampling frequency.

Acknowledgements

This work was funded under the New Zealand Ministry for Business, Innovation and Employment “Groundwater Assimilative Capacity” (C03X1001) and “Transfer Pathways Programme” (LVLX1502) contracts. Special thanks to Jasper Vrugt (University of California, Irvine) for sharing his DREAM codes, and to Maria Elena Orduña Alegría (Department of Hydrology and Geohydrology, Universität Stuttgart) who carried out early work on this project as part of her Master of Science dissertation research. Thanks to Waikato Regional Council, NIWA, Mighty River Power, Auckland

Council Watercare Services Limited, and Thuringian Remote Water Supply for permission to use their data.

Appendix A: Supplementary Materials

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jhydrol.2017.xx.xxx>. These include a map of the Waikato study sites, and a figure showing the effect of using a Student's t-distribution, instead of a Gaussian distribution, on the weighting of extreme residuals.

References

- Adams, G.A., Cornish, P.S., Croke, B.F.W., Hart, M.R., Hughes, C.E., Jakeman, A.J., 2009. A new look at uncertainty in end member mixing models for streamflow partitioning. Proceedings of 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, 13-17 July 2009, Cairns, Australia.
http://www.mssanz.org.au/modsim09/I1/adams_ga.pdf
- Andréassian, V., Lerat, J., Loumagne, C., Mathevet, T., Michel, C., Oudin, L., Perrin, C., 2007. What is really undermining hydrologic science today? *Hydrological Processes* 21, 2819-2822, doi:10.1002/hyp.6854.
- Beale, E.M.L., 1962. Some uses of computers in operational research. *Industrielle Organisation* 31, 51-52.
- Beven, K., 2005. On the concept of model structural error. *Water Science and Technology* 52(6), 167-175.
- Bidwell, V.J., Stenger, R., Barkle, G.F., 2008. Dynamic analysis of groundwater discharge and partial-area contribution to Pukemanga Stream, New Zealand. *Hydrology and Earth System*

Sciences 12, 975-987.

Bieger, K., Hörmann, G., Fohrer, N., 2012. Using residual analysis, auto- and cross-correlations to identify key processes for the calibration of the SWAT model in a data scarce region. *Advances in Geosciences* 31, 23-30, doi:10.5194/adgeo-31-23-2012.

Bowes, M. J., Jarvie, H. P., Naden, P. S., Old, G. H., Scarlett, P. M., Roberts, C., Armstrong, L. K., Harman, S. A., Wickham, H. D., Collins, A. L., 2014. Identifying patterns for nutrient mitigation using river concentration-flow relationships: The Thames basin, U.K. *Journal of Hydrology* 517, 1-12, doi:10.1016/j.jhydrol.2014.03.063.

Breuer, L., Vache, K.B., Julich, S., Frede, H., 2008. Current concepts in nitrogen dynamics for mesoscale catchments. *Hydrological Sciences Journal* 53, 1059-1074, doi:10.1623/hysj.53.5.1059.

Brocca, L., Melone, F., Moramarco, T., Singh, V.P., 2009. Assimilation of observed soil moisture data in storm rainfall-runoff modeling. *Journal of Hydrologic Engineering* 14(2), 153-165, doi:10.1061/(ASCE)1084-0699.

Christophersen, N., Neal, C., Hooper, R.P., Vogt, R.D., Andersen, S., 1990. Modelling streamwater chemistry as a mixture of soilwater end-members—a step towards second-generation acidification models. *Journal of Hydrology* 116, 307-320, doi:10.1016/0022-1694(90)90130-P.

Corradini, C., 1985. Analysis of the effects of orography on surface rainfall by a parameterized numerical model. *Journal of Hydrology* 77, 19-30, doi:10.1016/0022-1694(85)90195-7.

Dahlke, H.E., Easton, Z.M., Walter, M.T., Steenhuis, T.S., 2012. Field test of the variable source area interpretation of the curve number rainfall-runoff equation. *Journal of Irrigation and*

Drainage Engineering 138(3), 235-244, doi:10.1061/(ASCE)IR.1943-4774.0000380.

Delsman, J.R., Oude Essink, G.H.P., Beven, K.J., Stuyfzand, P.J., 2013. Uncertainty estimation of end-member mixing using generalized likelihood uncertainty estimation (GLUE), applied in a lowland catchment. *Water Resources Research* 49, 4792-4806, doi:10.1002/wrcr.20341.

Dupas, R., Jomaa, S., Musolff, A., Borchardt, D., Rode, M., 2016. Disentangling the influence of hydroclimate patterns and agricultural management on river nitrate dynamics from infra-hourly to decadal time scales. *Science of the Total Environment* 7, 53, doi:10.1016/j.scitotenv.2016.07.053.

Ferrant, S., Laplanche, C., Durbe, G., Probst, A., Dugast, P., Durand, P., Sanchez-Perez, J.M., Probst, J.L., 2013. Continuous measurement of nitrate concentration in a highly event-responsive agricultural catchment in south-west of France: is the gain of information useful? *Hydrological Processes* 27, 1751-1763, doi:10.1002/hyp.9324.

Fovet, O., Ruiz, L., Faucheux, M., Molénat, J., Sekhar, M., Vertès, F., Aquilina, L., Gascuel-Oudou, C., Durand, P., 2015. Using long time series of agricultural-derived nitrates for estimating catchment transit times. *Journal of Hydrology* 522, 603-617, doi:10.1016/j.jhydrol.2015.01.030.

Gelman, A., Shalizi, C.R., 2013. Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology* 66, 8-38, doi:10.1111/j.2044-8317.2011.02037.x.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *Journal of Hydrology* 377, 80-91, doi:10.1016/j.jhydrol.2009.08.003.

- Hartmann, A., Wagener, T., Rimmer, A., Lange, J., Brielmann, H., Weiler, M., 2013. Testing the realism of model structures to identify karst system processes using water quality and quantity signatures. *Water Resources Research* 49, 3345-3358, doi:10.1002/wrcr.20229.
- Hawkins, R.H., Ward, T.J., Woodward, D.E., van Mullem, J.A., 2010. Continuing evolution of rainfall-runoff and the curve number precedent. Proceedings of the Second Joint Federal Interagency Conference, Las Vegas, Nevada, 27 June–1 July 2010.
<http://acwi.gov/sos/pubs/2ndJFIC/>
- He, J.Q., Jones, J.W., Graham, W.D., Dukes, M.D., 2010. Influence of likelihood function choice for estimating crop model parameters using the generalised likelihood uncertainty estimation method. *Agricultural Systems* 103, 256-264, doi:10.1016/j.agsy.2010.01.006.
- Hesser, F.B., Franko, U., Rode, M., 2010. Spatially distributed lateral nitrate transport at the catchment scale. *Journal of Environmental Quality* 39, 193-203, doi:10.2134/jeq2009.0031.
- Horne, D.J., Scotter, D.R., 2016. The available water holding capacity of soils under pasture. *Agricultural Water Management* 1777, 165-171, doi:10.1016/j.agwat.2016.07.012.
- Hrachowitz, M., Benettin, P., van Breukelen, B.M., Fovet, O., Howden, N.J.K, Ruiz, L., van der Velde, Y., Wade, A.J., 2016. Transit times—the link between hydrology and water quality at the catchment scale. *WIREs Water* 3, 629-657, doi:10.1002/wat2.1155.
- Huard, D., Mailhot, A., 2006. A Bayesian perspective on input uncertainty in model calibration: application to hydrological model “abc”. *Water Resources Research* 42, W07416, doi:10.1029/2005WR004661.
- Huyck, A.A.O., Pauwels, V.R.N., Verhoest, N.E.C., 2005. A base flow separation algorithm based on the linearized Boussinesq equation for complex hillslopes. *Water Resources Research*

41, W08415, doi:10.1029/2004WR003789.

Jakeman, A.J., Hornberger, G.M., 1993. How much complexity is warranted in a rainfall-runoff model? *Water Resources Research* 29(8), 2637-2649, doi:10.1029/93WR00877.

Jeffreys, H., 1938. The law of error and the combination of observations. *Philosophical Transactions of the Royal Society (London) Series A* 237, 231-271, doi:10.1098/rsta.1938.0008.

Jiang, S., Jomaa, S., Rode, M., 2014. Modelling inorganic nitrogen leaching in nested mesoscale catchments in central Germany. *Ecohydrology* 7, 1345-1362, doi:10.1002/eco.1462.

Jiang, S., Jomaa, S., Büttner, O., Meon, G., Rode, M., 2015. Multi-site identification of a distributed hydrological nitrogen model using Bayesian uncertainty analysis. *Journal of Hydrology* 529, 940-950, doi:10.1016/j.jhydrol.2015.09.009.

Jomaa, S., Jiang, S., Thraen, D., Rode, M., 2016. Modelling the effect of different agricultural practices on stream nitrogen load in central Germany. *Energy, Sustainability and Society* 6(11), 1-16, doi: 10.1186/s13705-016-0077-9.

Juston, J.M., Kauffeldt, A., Quesada Montano, B., Seibert, J., Beven, K.J., Westerberg, I.K., 2013. Smiling in the rain: seven reasons to be positive about uncertainty in hydrological modelling. *Hydrological Processes* 27, 1117-1122, doi:10.1002/hyp.9625.

Klaus, J., McDonnell, J.J., 2013. Hydrograph separation using stable isotopesP: Review and evaluation. *Journal of Hydrology* 505, 47-64, doi:10.1016/j.jhydrol.2013.09.006.

Krause, P., Boyle, D.P., Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences* 5, 89-97, doi:10.5194/adgeo-5-89-2005.

Li, H., Sivapalan, M., Tian, F., Liu, D., 2010. Water and nutrient balance in a large tile-drained

agricultural catchment: A distributed modeling study. *Hydrology and Earth System Sciences* 14, 2259-2275, doi:10.5194/hess-14-2259-2010.

Li, M., Yang, D., Chen, J., Hubbard, S., 2012. Calibration of a distributed flood forecasting model with input uncertainty using a Bayesian framework. *Water Resources Research* 48, W08510, doi:10.1029/2010WR010062.

Liang, X., Xie, Z., 2001. A new surface runoff parameterization with subgrid-scale soil heterogeneity for land surface models. *Advances in Water Resources* 24, 1773-1193.

Longobardi, A., Villani, P., Guida, D., Cuomo, A., 2016. Hydro-geo-chemical streamflow analysis as a support for digital hydrograph filtering in a small, rainfall dominated, sandstone watershed. *Journal of Hydrology* 539, 177-187, doi:10.1016/j.jhydrol.2016.05.028.

Martin, C., Aquilina, L., Gascuel-Oudou, C., Molénat, J., Fauchoux, M., Ruiz, L., 2004. Seasonal and interannual variations of nitrate and chloride in stream waters related to spatial and temporal patterns of groundwater concentrations in agricultural catchments. *Hydrological Processes* 18, 1237-1254, doi:10.1002/hyp.1395.

Massari, C., Brocca, L., Barbetta, S., Papathanasiou, C., Mimikou, M., Moramarco, T., 2014. Using globally available soil moisture indicators for flood modelling in Mediterranean catchments. *Hydrology and Earth System Sciences* 18, 839-853, doi:10.5914/hess-18-839-2014.

Matott, L.S., Babendreier, J.E., Purucker, S.T., 2009. Evaluating uncertainty in integrated environmental models: a review of concepts and tools. *Water Resources Research* 45, W06421, doi:10.1029/2008WR007301.

McDonnell, J.J., 2014. The two water worlds hypothesis: ecohydrological separation of water

between streams and trees. WIREs Water 2014, doi:10.1002/wat2.1027.

McDowell, R.W., Snelder, T.H., Cox, N., Booker, D.J., Wilcock, R.J., 2013. Establishment of reference or baseline conditions of chemical indicators in New Zealand streams and rivers relative to present conditions. Marine and Freshwater Research 64, 387–400, doi:10.1071/MF12153.

McMillan, H., Krueger, T., Freer, J., 2012. Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. Hydrological Processes 26, 4078–4111, doi:10.1002/hyp.9384.

Morgenstern, U., Daughney, C.J., Leonard, G., Gordon, D., Donath, F.M., Reeves, R., 2015. Using groundwater age and hydrochemistry to understand sources and dynamics of nutrient contamination through the catchment into Lake Rotorua, New Zealand. Hydrology and Earth System Sciences 19, 803–822, doi:10.5194/hess-19-803-822.

Moriasi, D.N., Arnold, J.G., van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the American Society of Agricultural and Biological Engineers 50(3), 885–900, doi:10.13031/2013.23153.

Motulsky, H.J., Brown, R.E., 2006. Detecting outliers when fitting data with nonlinear regression—a new method based on robust nonlinear regression and the false discovery rate. BMC Bioinformatics 7, 123, doi:10.1186/1471-2105-7-123.

Nash, J.E., Sutcliffe, J.V., 1970, River flow forecasting through conceptual models: part 1 a discussion of principles. Journal of Hydrology 10(3), 282–290, doi:10.1016/0022-1694(70)90255-6.

- Oehler, F., Durand, P., Bordenave, P., Saadi, Z., Salmon-Monviola, J., 2009. Modelling denitrification at the catchment scale. *Science of the Total Environment* 407, 1726-1737, doi: 10.1016/j.scitotenv.2008.10.069.
- Oudin, L., Perrin, C., Mathevet, T., Andréassian, V., Michel, C., 2006. Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models. *Journal of Hydrology* 320, 62-83, doi:10.1016/j.jhydrol.2005.07.016.
- Pauwels, V.R.N., Verhoest, N.E.C., De Troch, F.P., 2002. A metahillslope model based on an analytical solution to a linearised Boussinesq equation for temporally variable recharge rates. *Water Resources Research* 38(12), 1297, doi:10.1029/2001WR000714.
- Phillips, F.M., 2010. Soil-water bypass. *Nature Geoscience* 3, 77-78.
- Pulido-Velazquez, M.A., Sahuquillo-Herraz, A., Ochoa-Rivera, J.C., Pulido-Velazquez, D., 2005. Modeling of stream-aquifer interaction: the embedded multireservoir model. *Journal of Hydrology* 313, 166-181, doi:10.1016/j.jhydrol.2005.02.026.
- Radulovic, R., Sollins, P., Baveye, P., Solórzano, E., 1992. Bypass water flow through unsaturated microaggregated tropical soils. *Soil Science Society of America Journal* 56(3), 721-726.
- Refsgaard, J.C., van der Sluijs, J.P., Højberg, A.L., Vanrolleghem, P.A., 2007. Uncertainty in the environmental modelling process: a framework and guidance. *Environmental Modelling and Software* 22, 1543-1556, doi:10.1016/j.envsoft.2007.02.004.
- Rode, M., Halbedel, S., Anis, M.R., Borchardt, D., Weitere, M., 2016. Continuous in-stream assimilatory nitrate uptake from high-frequency sensor measurements. *Environmental Science and Technology* 50, 5685-5694, doi:10.1021/acs.est.6b00943.
- Ruiz, L., Abiven, S., Martin, C., Durand, P., Beaujouan, V., Molénat, J., 2002. Effect on nitrate

concentration in stream water of agricultural practices in small catchments in Brittany: II. temporal variations and mixing processes. *Hydrology and Earth System Sciences* 6(3), 507-513, doi:10.5194/hess-6-507-2002.

Ruiz-Pérez, G., Medici, C., Latron, J., Llorens, P., Gallart, F., Francés, F., 2016. Investigating the behaviour of a small Mediterranean catchment using three different hydrological models as hypotheses. *Hydrological Processes* 30, 2050-2062, doi:10.1002/hyp.10738.

Rushton, K.R., Eilers, V.H.M., Carter, R.C., 2006. Improved soil moisture balance methodology for recharge estimation. *Journal of Hydrology* 318, 379-399, doi:10.1016/j.jhydrol.2005.06.022.

Sackmann, B.S., 2011. Deschutes River continuous nitrate monitoring. Washington State Department of Ecology Publication 11-03-030.
<https://fortress.wa.gov/ecy/publications/documents/1103030.pdf>.

Sahuquillo, A., 1983. An eigenvalue numerical technique for solving unsteady linear groundwater models continuously in time. *Water Resources Research* 19, 87-93, doi: 10.1029/WR019i001p00087.

Schoups, G., Vrugt, J.A., 2010. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated heteroscedastic, and non-Gaussian errors. *Water Resources Research* 46, W10531, doi:10.1029/2009WR008933.

Scotter, D.R., Clothier, B.E., Turner, M.A., 1979. The soil water balance in a fragiaqualf and its effect on pasture growth in central New Zealand. *Australian Journal of Soil Research* 17, 455-465.

Shrestha, R.R., Osenbrück, K., Rode, M., 2013. Assessment of catchment response and

calibration of a hydrological model using high-frequency discharge-nitrate concentration data. *Hydrology Research* 44(6), 995-1012, doi:10.2166/nh.2013.087.

Sivapalan, M., Ruprecht, J.K., Viney, N.R., 1996. Water and salt balance modelling to predict the effects of land-use changes in forested catchments: 1. Small catchment water balance model. *Hydrological Processes* 10, 393-411, doi:10.1002/(SICI)1099-1085.

Sloan, W.T., 2000. A physics-based function for modeling transient groundwater discharge at the watershed scale. *Water Resources Research* 36, 225-241, doi:10.1029/1999WR900221.

Sorooshian, S., Dracup, J.A., 1980. Stochastic parameter estimation procedure for hydrologic rainfall-runoff models: correlated and heteroscedastic error cases. *Water Resources Research* 16(2), 430-442, doi:10.1029/WR016i002p00430.

Srinivasan, M.S., McDowell, R.W., 2009. Identifying critical source areas for water quality: 1. mapping and validating transport areas in three headwater catchments in Otago, New Zealand. *Journal of Hydrology* 379, 54-67, doi:10.1016/j.jhydrol.2009.09.044.

Steenhuis, T.S., Winchell, M., Rossing, J., Zollweg, J.A., Walter, M.F., 1995. SCS runoff equation revisited for variable-source runoff areas. *Journal of Irrigation and Drainage Engineering* 121(3), 234-238, doi:10.1061/(ASCE)0733-9437.

Su, C.H., Peterson, T.J., Costelloe, J.F., Western, A.W., 2016. A synthetic study to evaluate the utility of hydrological signatures for calibrating a base flow separation filter. *Water Resources Research* 52, doi:10.1002/2015WR018177.

Tait, A., Sturman, J., Clark, M., 2012. An assessment of the accuracy of interpolated daily rainfall for New Zealand. *Journal of Hydrology (NZ)* 51, 25-44.

- ter Braak, C.J.F., Vrugt, J.A., 2008. Differential evolution Markov Chain with snooker updater and fewer chains. *Statistical Computing* 18, 435-446, doi:10.1007/s11222-008-9104-9.
- Thorburn, P.J., Rose, C.W., 1990. Interpretation of solute profile dynamics in irrigated soils: III. a simple model of bypass flow in soils. *Irrigation Science* 11(4), 219-225, doi:10.1007/BF00190536.
- Uhlenbrook, S., Siebert, J., Leibundgut, Ch., Rodhe, A., 1999. Predictive uncertainty of a rain-flow conceptual model due to the identification of parameters and model structure. *Hydrological Sciences Journal* 44, 779-797, doi:10.1080/02626669909492273.
- USDA, 1986. Urban hydrology for small watersheds. Technical Release 55, Natural Resources Conservation Service, United States Department of Agriculture.
www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1044171.pdf.
- USDA, 2004. Estimation of direct runoff from storm rainfall. Chapter 10, National Engineering Handbook, Part 630, Natural Resources Conservation Service, United States Department of Agriculture. <https://directives.sc.egov.usda.gov/OpenNonWebContent.aspx?content=17752.wba>.
- Vrugt, J.A., 2016. Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling and Software* 75, 273-316, doi:10.1016/j.envsoft.2015.08.013.
- Vrugt, J.A., ter Braak, C.J.F., Diks, C.G.H., Robinson, B.A., Hyman, J.M., Higdon, D., 2002. Accelerating Markov Chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation* 10(3), 271-288.
- Vrugt, J.A., ter Braak, C.J.F., Clark, M.P., Hyman, J.M., Robinson, B.A., 2008. Treatment of input

uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resources Research* 44, W00B09, doi:10.1029/2007WR006720.

Wade, A.J., Palmer-Felgate, E.J., Halliday, S.J., Skeffington, R.A., Loewenthal, M., Jarvie, H.P., Bowes, M.J., Greenway, G.M., Haswell, S.J., Bell, I.M., Joly, E., Fallatah, A., Neal, C., Williams, R.J., Gozzard, E., Newman, J.R., 2012. Hydrochemical processes in lowland rivers: insights from in situ, high-resolution monitoring. *Hydrology and Earth System Sciences* 16, 4323-4342, doi:10.5194/hess-16-4323-2012.

Wilcock, R.J., Monaghan, R.M., Quinn, J.M., Campbell, A.M., Thorrold, B.S., Duncan, M.J., McGowan, A.W., Betteridge, K., 2006. Land-use impacts and water quality targets in the intensive dairying catchment of the Toenepi Stream, New Zealand. *New Zealand Journal of Marine and Freshwater Research* 40, 123-140, doi:10.1080/00288330.2006.9517407.

Wöhling, T., Samaniego, L., Kumar, R., 2013. Evaluating multiple performance criteria to calibrate the distributed hydrological model of the upper Neckar catchment. *Environmental Earth Sciences* 69(2), 1–16, doi:10.1007/s12665-013-2306-2.

Woodward, S.J.R., Barker, D.J., Zyskowski, R.F., 2001. A practical model for predicting soil water deficit in New Zealand pastures. *New Zealand Journal of Agricultural Research* 44, 91-109, doi:10.1080/00288233.2001.9513464.

Woodward, S.J.R., Stenger, R., Bidwell, V.J., 2013. Dynamic analysis of stream flow and water chemistry to infer subsurface water and nitrate fluxes in a lowland dairying catchment. *Journal of Hydrology* 505, 299-311, doi:10.1016/j.jhydrol.2013.07.044.

Woodward, S.J.R., Stenger, R., Hill, R.B., 2016a. Flow stratification of river water quality data to elucidate nutrient transfer pathways in mesoscale catchments. *Transactions of the American Society of Agricultural and Biological Engineers* 59(2), 545-551,

doi:10.13031/trans.59.11145.

Woodward, S.J.R., Stenger, R., Wöhling, Th., 2016b. Uncertainty in the modelling of spatial and temporal patterns of shallow groundwater flow paths: the role of geological and hydrological site information. *Journal of Hydrology* 534, 680-694, doi: 10.1016/j.jhydrol.2016.01.045.

WRC, 2013. Trends in river water quality in the Waikato Region, 1993-2012. Waikato Regional Council Technical Report 2013/20. Available at <http://www.waikatoregion.govt.nz/tr201320/>

WRC, 2015. Groundwater field investigations over the 2014-2015 summer in support of the Healthy Rivers Project. Waikato Regional Council Internal Series Report 2015/38. Available at <http://www.waikatoregion.govt.nz/>

Wu, J.C, Zeng, X.K., 2013. Review of the uncertainty analysis of groundwater numerical simulation. *Chinese Science Bulletin* 58(25), 3044-3052, doi:10.1007/s11434-013-5950-8.

Xu, C.Y., 2001. Statistical analysis of parameters and residuals of a conceptual water balance model: methodology and case study. *Water Resources Management* 15, 75-92, doi: 10.1023/A:1012559608269.

Table Captions

Table 1: Summary of catchment data sets, including average annual rainfall, PET and stream flow totals, ranges of observed nitrate concentration, and number of daily discharge (n_D), monthly or daily concentration (n_C) and monthly flow (n_M) data points in the calibration and validation periods..

Table 2: Model parameters ranges, including fixed parameters.

Figure Captions

Figure 1: Concentration-discharge relationships for nitrate ($\text{mg L}^{-1} \text{NO}_3\text{-N}$) for the four catchments in this study. The monthly and daily Weida Stream data are shown separately.

Figure 2: Catchment water reservoirs and flow paths simulated in the revised StreamGEM model described in this paper.

Figure 3: Distribution of posterior log-likelihoods for each calibration variable, and for the weighted total. Plots show 95% ranges and quartiles (boxes and whiskers), and maximum likelihood values (crosses). Data set codes are given in Table 1.

Figure 4: Example analysis of residuals for one likely parameter realisation for the Puniu River with rainfall from Barton's Corner (PB) data set.

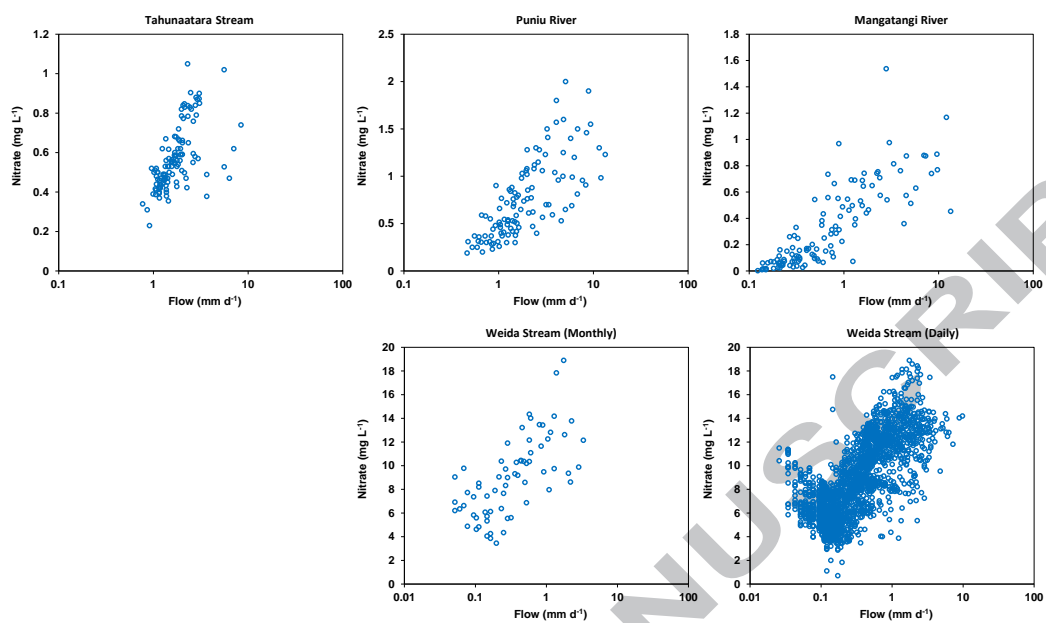
Figure 5: Calibration and validation performance (NSL, Nash-Sutcliffe Log, calculated from posterior parameter distributions), showing 95% ranges and quartiles (box and whisker plot) and maximum likelihood values (crosses). Data set codes are given in Table 1.

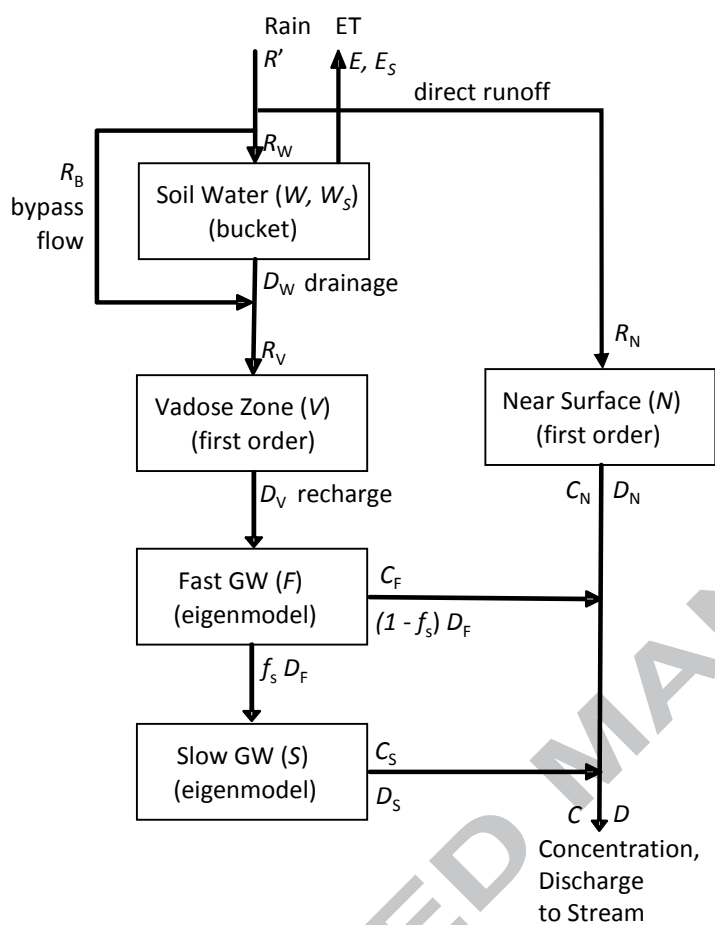
Figure 6: Calibrated parameter values (posterior distributions), showing prior bounds (dotted lines), 95% ranges, and quartiles (box and whisker plot) and maximum likelihood values (crosses). Concentration parameters for the WM and WD data sets have been divided by 10. Data set codes are given in Table 1.

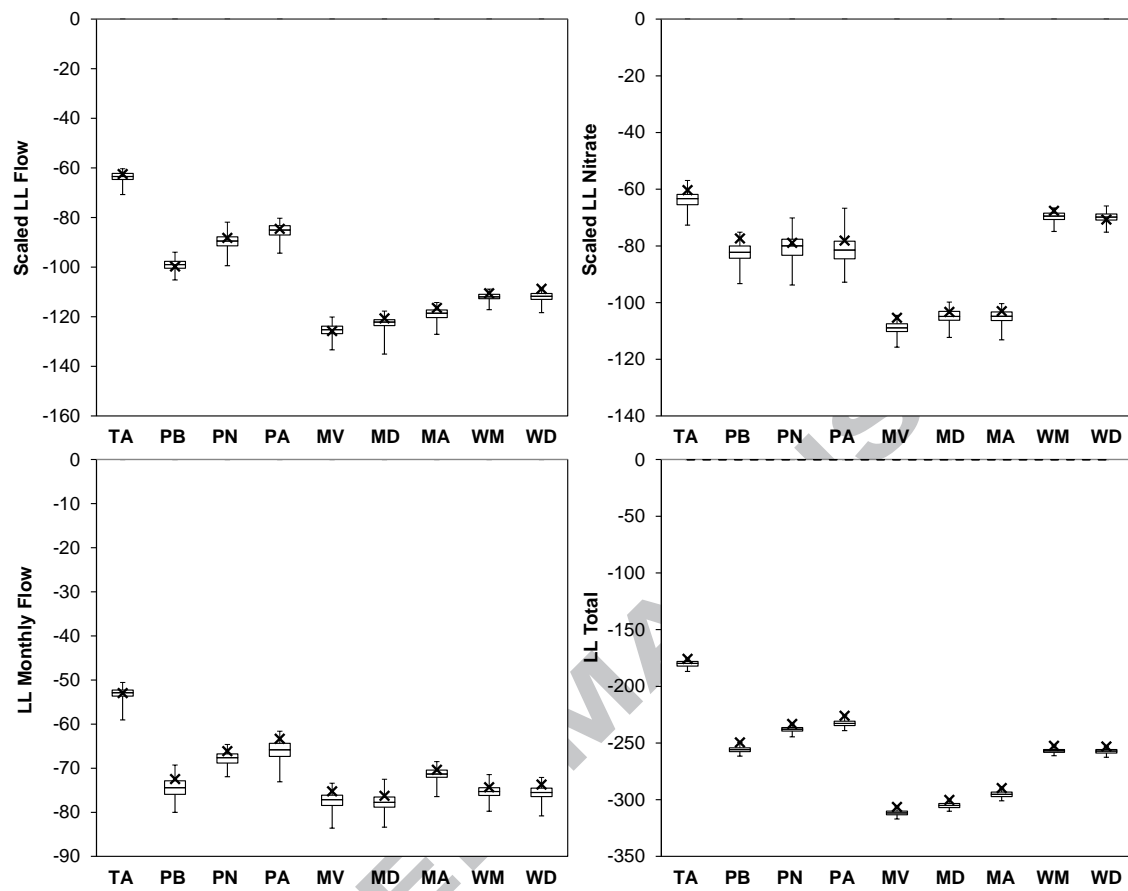
Figure 7: High resolution hydrograph and chemograph predictions taken from the calibration period for (a) the Weida Stream with Monthly nitrate data set (WM), and (b) the Weida Stream with Daily nitrate data set (WD). Data are indicated by blue circles. Model prediction parameter uncertainty bounds are indicated by dark shading, and total 95% uncertainty bounds by light shading.

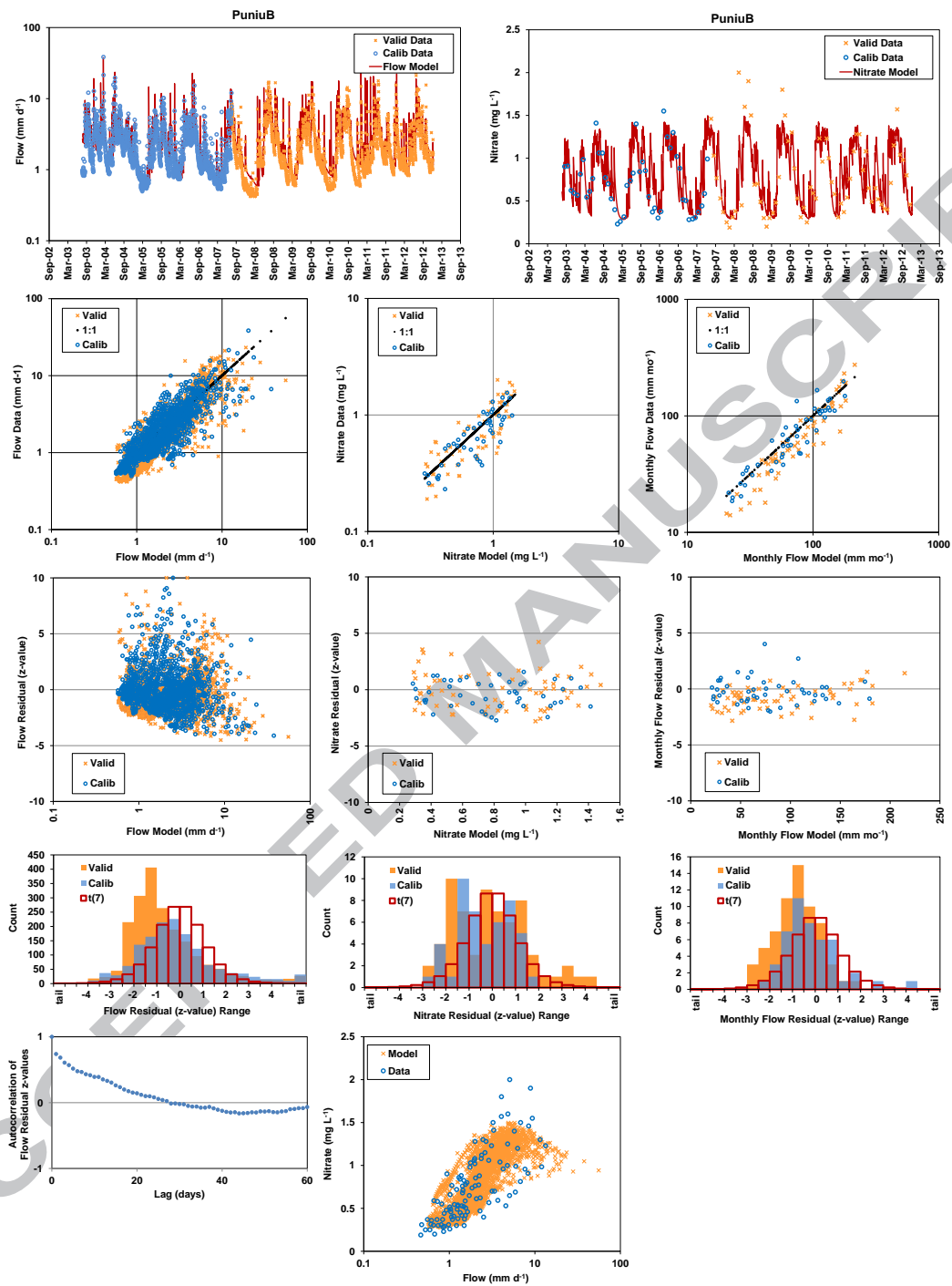
Figure 8: Example of dynamic stream flow contributions, for one likely parameter realisation for the Puniu River with rainfall from Barton's Corner (PB) data set, corresponding to Figure 4.

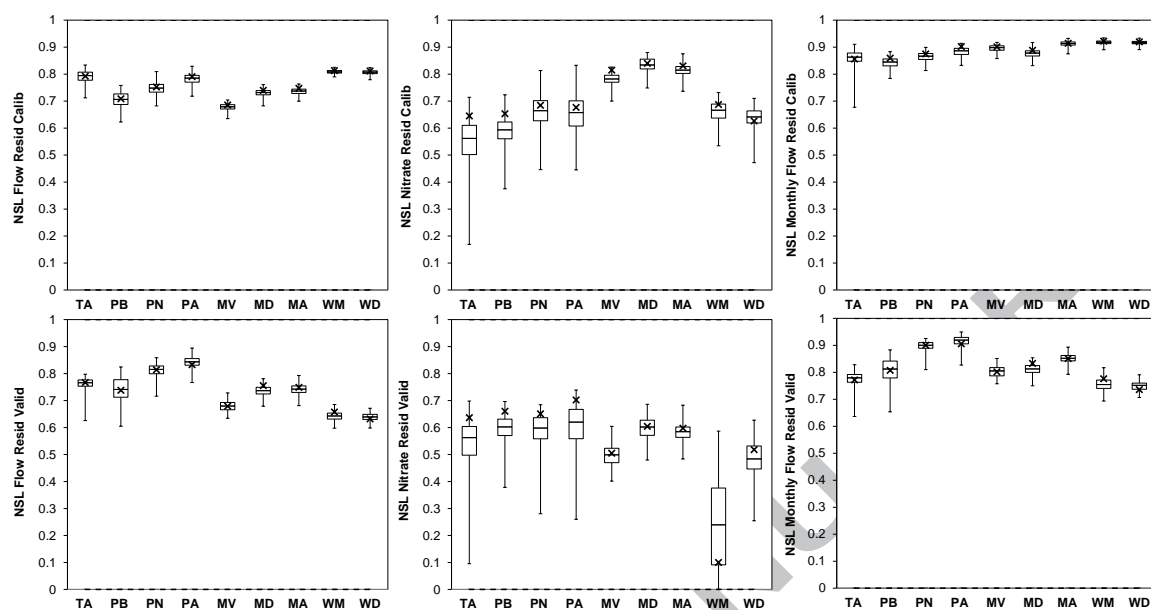
Figure 9: Predicted catchment fluxes (calculated from posterior parameter distributions), showing 95% ranges and quartiles (box and whisker plot) and maximum likelihood values (crosses); (a) average annual stream flows expressed in mm y^{-1} , and (b) average annual stream nitrate yields in $\text{kg ha}^{-1} \text{y}^{-1}$. Yields for the WM and WD data sets have been divided by 10. Circles in (a) indicate average measured annual flow, and in (b), average annual nitrate yield calculated using the Beale Ratio Estimator approach. Data set codes are given in Table 1.

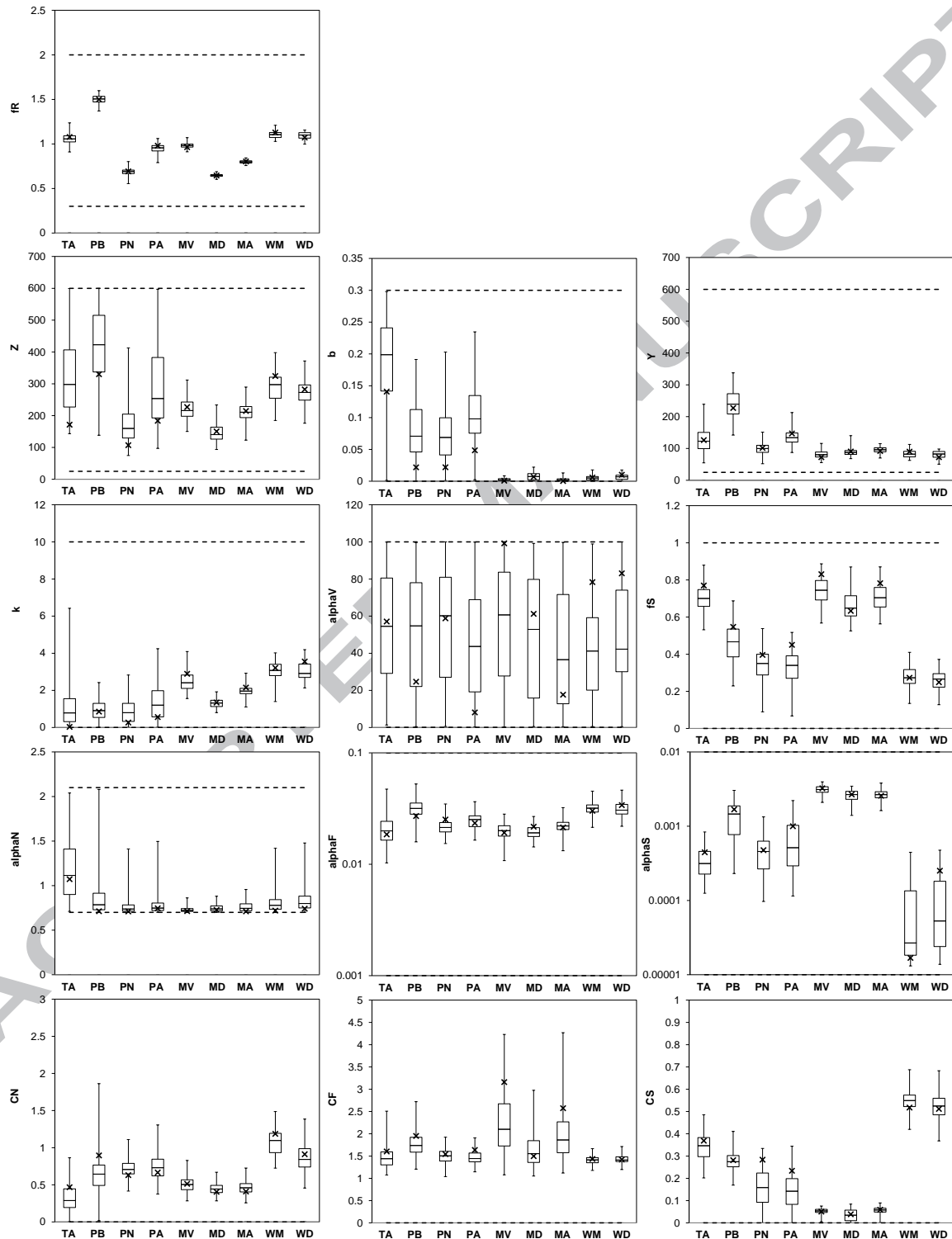


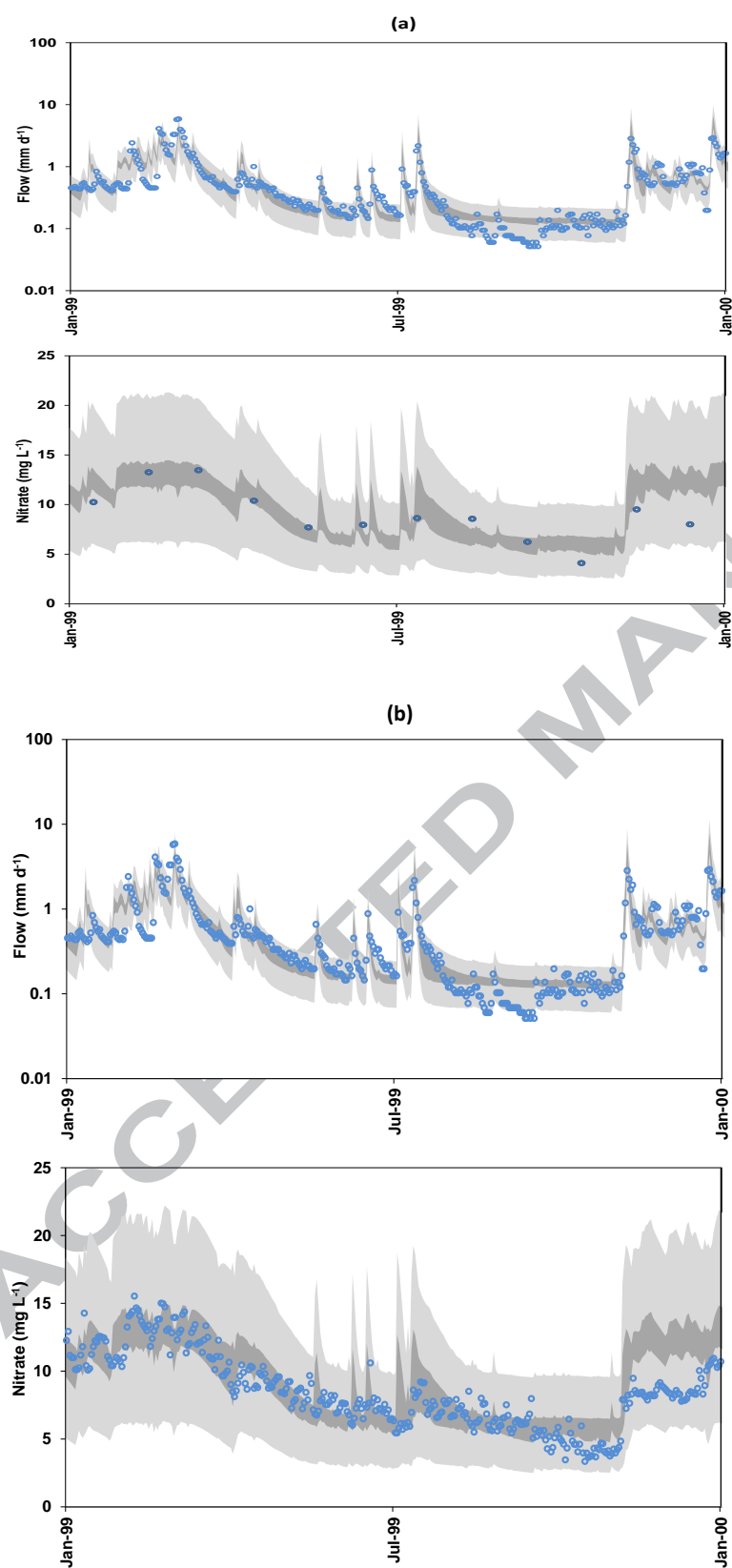


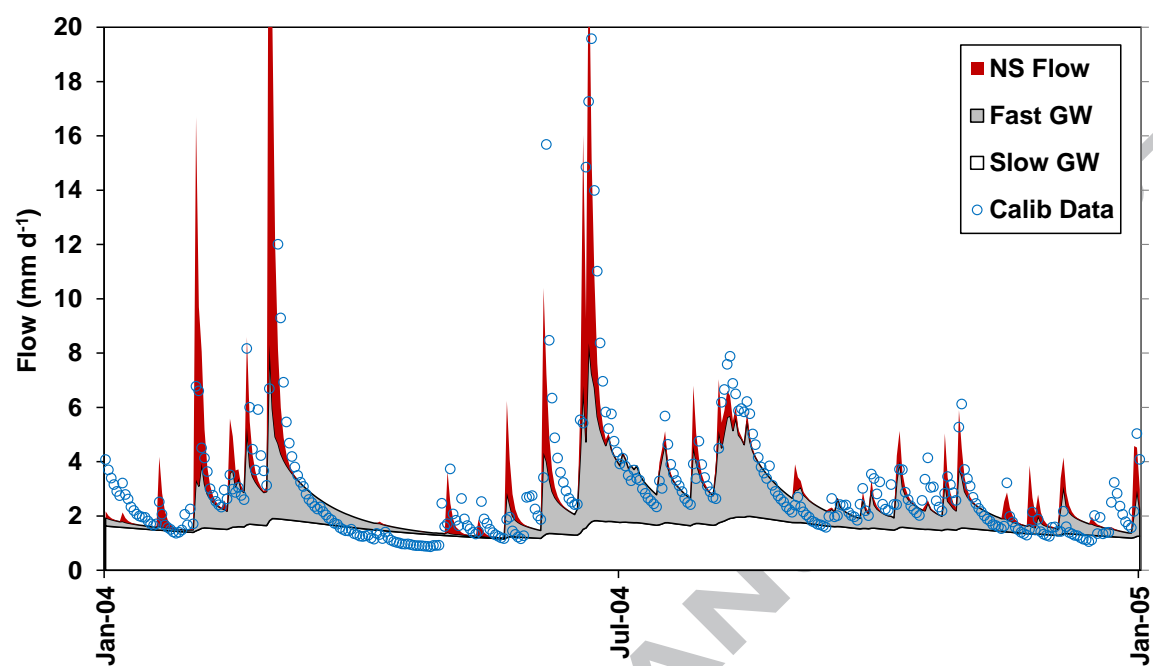


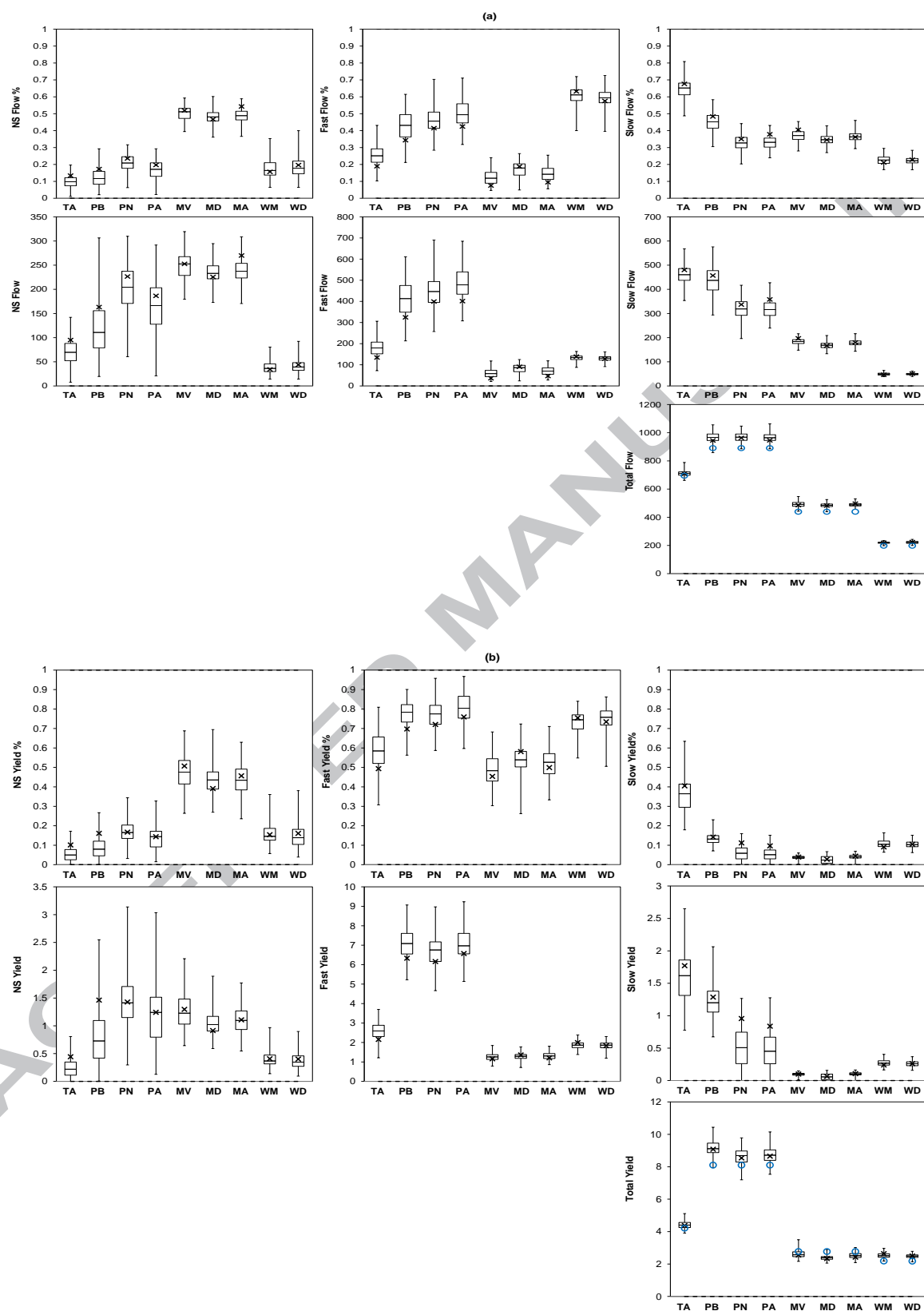












Dat a Set	Catchmen t (Rainfall Site)	Area km ²	Altitud e m.a.s.l	Rainfa ll mm y ⁻¹	PE T m m y ⁻¹	Flo w mm y ⁻¹	Nitrat e mg L ⁻¹	Samplin g	Calibration (n _D ,n _C ,n _M)	Validation (n _D ,n _C ,n _M)
TA	Tahunaatar a Stream (Ohakuri Road)	208. 1	276- 840	1258	887	695	0.2- 1.1	Monthly	1/4/03- 31/3/07 (1461,48,48)	1/4/07- 31/12/12 (2102,69,6 9)
PB	Puniu River (Barton's Corner)	519. 1	29-960	1186	827	915	0.2- 2.0	Monthly	7/8/03-6/8/07 (1461,47,47)	7/8/07- 31/12/12 (1974,65,6 4)
PN	Puniu River (Ngaroma)			2297						
PA	Puniu River (Average)			1742						
MV	Mangatang i River (Virtual Climate)	194. 5	14-680	1167	884	428	0.0- 1.5	Monthly	1/4/03- 31/3/07 (1461,48,48)	1/4/07- 31/12/12 (2098,69,6 8)
MD	Mangatang i River (Mangatan			1764						

gi Dam)

MA Mangatang 1466

i River

(Average)

WM	Weida	99.5	357-	640	567	228	3.5-	Monthly	30/10/97-	30/10/01-
	Stream		552				18.9		29/10/01	31/12/03
	(Average)								(1461,48,47)	(793,20,26)
WD	Weida						0.7-	Daily	30/10/97-	30/10/01-
	Stream						18.9		29/10/01	31/12/03
	(Average)								(1461,1431,4	(793,566,2
									7)	6)

	Symbol	Prior Range	Units
Model Parameters			
Rainfall data scaling factor	f_R	0.3-2.0	-
Maximum infiltration to soil water	Z	25-600	mm d ⁻¹
Near surface runoff threshold fraction	λ	0.05	-
Antecedent moisture effect parameter	k	0-10	d ⁻¹
Soil water bypass fraction	b	0-0.3	-
Maximum soil water storage	Y	25-600	mm
Drainage threshold fraction	μ	1.0	-
Near-surface discharge rate	α_N	0.7-2.1	d ⁻¹
Vadose zone discharge rate	α_V	0.01-100	d ⁻¹
Fast groundwater parameter	α_F	0.001-0.1	d ⁻¹
Slow groundwater parameter	α_S	0.00001-0.01	d ⁻¹
Fraction recharge to slow groundwater	f_S	0-1	-
Near-surface nitrate-N	C_N	0-20	mg L ⁻¹
Fast groundwater nitrate-N	C_F	0-20	mg L ⁻¹
Slow groundwater nitrate-N	C_S	0-20	mg L ⁻¹

Highlights

- The StreamGEM model was calibrated using monthly nitrate data from four catchments
- Posterior parameter distributions were derived by Bayesian analysis with DREAMZS
- Predicted flow path water and nitrate fluxes were consistent with catchment geology
- Use of daily nitrate data in one catchment did not improve model predictions
- Four years of monthly samples was sufficient to characterize catchment dynamics