Comments from the Editor:

The paper has improved significantly since the last revision, and the reviewers are happy with the result.

- 1. Please clarify with respect to the comments of Reviewer 2.
- We have made relevant changes to each of the points raised by Reviewer 2 (for details see our response below).

2. I would separate Table 1 into 2 tables, one for the data used for configuring the model and the other with the data used for calibrating the model.

An additional table has now been added after Table 2 as follows: "Table 3. Description of data used to calibrate the SWAT model. Data were measured at the Forest Research Institute (FRI) stream–gauge and were considered representative of the downstream/outlet conditions of the Puarenga Stream".

Data	Application	Measurement data details	Source
Stream discharge measurements	Calibration (2004–2008) Validation (1994–1997)	n 15-min stream discharge data were measured at FRI stream-gauge (Fig. 1b) within the catchment and aggregated as daily mean values (1994–1997; 2004– 2008).	
Stream water quality measurements	Calibration (2004–2008) Validation ¹ (1994–1997; 2010–2012)	Monthly grab samples for determination of suspended sediment (SS), total phosphorus (TP) and total nitrogen (TN) concentrations (1994–1997; 2004– 2008), and high–frequency event–based samples for concentrations of SS (nine events), TP and TN (both 14 events) at 1–2 h frequency (2010–2012), were also measured at FRI stream–gauge (Fig. 1b) within the catchment.	BoPRC; Abell et al., 2013

3. I would separate each item (e.g. instead of having stream discharge and water quality, have 2 separate items).

Each item has been separated as "stream discharge measurements" and "stream water quality measurements" as suggested.

¹ Model validation was undertaken using two different datasets. The monthly measurements (1994–1997) were predominantly collected when base flow was the dominant contributor to stream discharge. Data from high–frequency sampling during rain events (2010–2012) were also used to validate model performance during periods when quick flow was high.

4. Also, they should mention where the stream discharge and water quality was measured (e.g. outlet or internal stations), possibly relating each item to the calibration stage where it was used.

The relevant text has been added as follows: "... were measured at FRI stream-gauge (Fig. 1b) within the catchment".

Each item in Table 3 has been linked to text in the section on Model calibration and validation as follows:

"Daily mean discharge was firstly calibrated based on daily mean values of 15-minute measurements (Table 3)",

"Modelled mean daily concentrations were compared with concentrations measured during monthly grab sampling, with monthly measurements assumed equal to daily mean concentrations (Table 3)", and

"In addition, high–frequency (1–2 h) water quality sampling was undertaken at the FRI stream–gauge during 2010–2012 (Table 3) to derive estimates of daily mean contaminant loads during storm events".

5. The table and text should also clarify whether the model was calibrated using only outlet data, or also internal station data.

Please see the caption of Table 3: "...considered representative of the downstream/outlet conditions of the Puarenga Stream".

6. Finally, Figure 1 should list the measurement stations that were actually used for model calibration, possibly cross-referencing with Table 1.

The relevant text has been added in the caption of Figure 1 (but note reference to Table 3) as follows: "... Measurement data (Table 3) used to calibrate the SWAT model were from the Forest Research Institute (FRI) stream–gauge and were considered representative of the downstream/outlet conditions of the Puarenga Stream".

Anonymous Referee #1

After a re-review, I appreciate the authors for accurate responses to all of my comments. I think that quality of the text is very better than the previous version. However, I have also some minor comments before final acceptance:

1. I'm satisfied with the authors' responses on comments# 5, 6, and 8. However, kindly add the statements as the sources of error and/or recommendations for next investigations at the end of the manuscript.

On Page 21, lines 6–11, the relevant text was stated in Conclusions as follows: "Discharge-weighted mean concentrations of TP and SS measured during storm events were greatly underestimated by SWAT, highlighting the potential for uncertainty to be greatly underestimated in catchment model applications that are validated using a sample of contaminant load measurements that is over-represented by measurements made during base flow conditions". This has been now linked with the statement: "Monitoring programmes which collect high-frequency and event-based data should be considered further to support more robust calibration and validation of SWAT model applications".

On Page 21, lines 11–13, the relevant text was also stated in Conclusions as follows: "Accurate simulation of nitrogen concentrations was constrained by the non–steady state of groundwater nitrogen concentrations due to historic variability in anthropogenic nitrogen applications to land". Additional recommendation has been now added as follows: "Improved representation of groundwater processes in the model structure would reduce this aspect of model uncertainty".

2. As the authors known, use of Penman–Monteith (potential/reference evapotranspiration) method leads to increase of uncertainty. Therefore, it is necessary to cite some of the newest (2014 & 2015) successful application of potential/reference evapotranspiration to justify the methodology and indicate the advantages of this approach. For this purpose, cite all of the below papers:
1) Analysis of potential evapotranspiration using 11 modified temperature-based models

2) Evaluation of radiation methods to study potential evapotranspiration of 31 provinces

3) Comparative evaluation of radiation-based methods for estimation of potential evapotranspiration

4) Study of different climatic conditions to assess the role of solar radiation in reference crop evapotranspiration equations

5) Investigation of Valiantzas' evapotranspiration equation in Iran

- 6) Temperature analysis of reference evapotranspiration models
- We have considered each of the papers that the reviewer has provided above.
 However, none of these papers has direct relevance to our study.

Anonymous Referee #2

The authors from the paper entitled: "Effects of hydrologic conditions on SWAT model performance and 1 parameter sensitivity for a small, mixed land use catchment in New Zealand" performed a major revision making the manuscript more clear and better. The organization of the paper is better as well as the description of the model set up, sensitivity analysis and calibration procedures. They attended all of the major points made by the reviewers and editor. There are a few small topics that I wish they take into consideration prior to publication: 1. Line 26, Page 5 : what is LTS? Was it explained before?

LTS has been removed and the text corrected on Page 5, line 27 as follows: "Two cold–water springs (Waipa Spring and Hemo Spring) and one geothermal spring (Fig. 1b) are located in the catchment area".

2. In lines 27-28 of page 6: "The SWAT model version used (SWAT2009_rev488) runs on a daily time step", but then on lines 31-32: "Hourly rainfall estimates were used as hydrologic forcing data" and lines2-4 of page 7 "Therefore, the hourly rainfall/Green & Ampt infiltration/daily routing method (Neitsch et al., 2011) was chosen to simulate upland and in–stream processes." These arguments go against each. The SWAT model runs with hourly data if your input is hourly data and Green & Ampt, but you are using the exported results of daily data, right? Please clarify this on the text saying that the model was run on hourly time step, but daily results were used for the paper;

Thank you for pointing this out. The text has been modified on Page 6, lines 28–29 as follows: "The SWAT model (version SWAT2009_rev488) was run on an hourly time step, but daily mean simulation outputs were used for this study".

3. Also make sure you were running on Green & Ampt with daily routing: as before there was an error on the text material of the SWAT model, and that the Green and Ampt method is used for hourly precipitation and hourly routing as well. As you can see this was clarified in the current version of SWAT in the IEVENT: there is the SCS option and the Subdaily rainfall data/Runoff estimated with Green & Ampt method/Hourly stream routing. I know you used a previous version, but please check if routing was really conducted in daily scale in the model, not only in the model text description. I had some issues running on this format and by checking with the developers this last year, they clarified that it was not running on that format and left only these two options of IEVENT.

Thank you for pointing this out. The text has been modified on Page 7, line 3 as follows: "Therefore, the hourly rainfall/Green & Ampt infiltration/hourly routing method (Neitsch et al., 2011) was chosen to simulate upland and instream processes". Effects of hydrologic conditions on SWAT model performance and
 parameter sensitivity for a small, mixed land use catchment in New Zealand
 3

- 4 W. Me^{1,2}, J. M. Abell^{1,*}, D. P. Hamilton¹
- 5 [1]{Environmental Research Institute, University of Waikato, Private Bag 3105,
- 6 Hamilton 3240, New Zealand}
- 7 [2]{College of Hydrology and Water Resources, Hohai University, Nanjing,
- 8 210098, People's Republic of China}
- 9 [*]{now at: Ecofish Research Ltd., Suite 1220 1175 Douglas Street, Victoria,
- 10 British Columbia, Canada}
- 11 Correspondence to: W. Me (yaowang0418@gmail.com)
- 12

13 Abstract

The Soil Water Assessment Tool (SWAT) was configured for the Puarenga 14 15 Stream catchment (77 km²), Rotorua, New Zealand. The catchment land use is mostly plantation forest, some of which is spray-irrigated with treated wastewater. 16 A Sequential Uncertainty Fitting (SUFI-2) procedure was used to auto-calibrate 17 unknown parameter values in the SWAT model. Model validation was performed 18 using two datasets: 1) monthly instantaneous measurements of suspended 19 sediment (SS), total phosphorus (TP) and total nitrogen (TN) concentrations; and 20 2) high-frequency (1-2 h) data measured during rainfall events. Monthly 21 instantaneous TP and TN concentrations were generally not reproduced well (24% 22 bias for TP, 27% bias for TN, and $R^2 < 0.1$, NSE < 0 for both TP and TN), in 23 contrast to SS concentrations (< 1% bias; R^2 and NSE both > 0.75) during model 24 validation. Comparison of simulated daily mean SS, TP and TN concentrations 25 with daily mean discharge-weighted high-frequency measurements during storm 26 events indicated that model predictions during the high rainfall period 27 considerably underestimated concentrations of SS (44% bias) and TP (70% bias), 28 while TN concentrations were comparable (< 1% bias; R^2 and NSE both ~0.5). 29 This comparison highlighted the potential for model error associated with quick-30

1 flow fluxes in flashy lower-order streams to be underestimated compared with low-frequency (e.g. monthly) measurements derived predominantly from base 2 flow measurements. To address this, we recommend that high-frequency, event-3 4 based monitoring data are used to support calibration and validation. Simulated discharge, SS, TP and TN loads were partitioned into two components (base flow 5 and quick flow) based on hydrograph separation. A manual procedure (one-at a-6 7 time sensitivity analysis) was used to quantify parameter sensitivity for the two hydrologically-separated regimes. Several SWAT parameters were found to have 8 9 different sensitivities between base flow and quick flow. Parameters relating to main channel processes were more sensitive for the base flow estimates, while 10 11 those relating to overland processes were more sensitive for the quick flow estimates. This study has important implications for identifying uncertainties in 12 13 parameter sensitivity and performance of hydrological models applied to catchments with large fluctuations in stream flow, and in cases where models are 14 15 used to examine scenarios that involve substantial changes to the existing flow 16 regime.

17

18 **1 Introduction**

Catchment models are valuable tools for understanding natural processes 19 occurring at basin scales and for simulating the effects of different management 20 21 regimes on soil and water resources (e.g. Cao et al., 2006). Model applications 22 may have uncertainties as a result of errors associated with the forcing variables, measurements used for calibration, and conceptualisation of the model itself 23 (Lindenschmidt et al., 2007). The ability of catchment models to simulate 24 25 hydrological processes and pollutant loads can be assessed through analysis of uncertainty or errors during a calibration process that is specific to the application 26 27 domain (White and Chaubey, 2005).

The Soil and Water Assessment Tool (SWAT) model is increasingly used to predict discharge, sediment and nutrient loads on a temporally resolved basis, and to quantify material fluxes from a catchment to the downstream receiving environment such as a lake (e.g. Nielsen et al., 2013). The SWAT model is physically-based and provides distributed descriptions of hydrologic processes at sub-basin scale (Arnold et al., 1998; Neitsch et al., 2011). It has numerous

parameters, some of which can be fixed on the basis of pre-existing catchment 1 data (e.g. soil maps) or knowledge gained in other studies. However, values for 2 other parameters need to be assigned during a calibration process as a result of 3 complex spatial and temporal variations that are not readily captured either 4 through measurements or within the model algorithms themselves (Boyle et al., 5 2000). Such parameter values assigned during calibration are therefore lumped, 6 i.e., they integrate variations in space and/or time and thus provide an 7 approximation for real values which often vary widely within a study catchment. 8 9 Model calibration is an iterative process whereby parameters are adjusted to the 10 system of interest by refining model predictions to fit closely with observations 11 under a given set of conditions (Moriasi et al., 2007). Manual calibration depends on the system used for model application, the experience of the modellers, and 12 13 knowledge of the model algorithms. It tends to be subjective and time-consuming. By contrast, auto-calibration provides a less labour-intensive approach by using 14 optimisation algorithms (Eckhardt and Arnold, 2001). The Sequential Uncertainty 15 Fitting (SUFI-2) procedure has previously been applied to auto-calibrate 16 17 discharge parameters in a SWAT application for the Thur River, Switzerland (Abbaspour et al., 2007), as well as for groundwater recharge, evapotranspiration 18 and soil storage water considerations in West Africa (Schuol et al., 2008). Model 19 validation is subsequently performed using measured data that are independent of 20 those used for calibration (Moriasi et al., 2007). 21

Values for hydrological parameter values in the SWAT model can vary 22 temporally. Cibin et al. (2010) found that the optimum calibrated values for 23 hydrological parameters varied with different flow regimes (low, medium and 24 high), thus suggesting that SWAT model performance can be optimised by 25 assigning parameter values based on hydrological characteristics. Other work has 26 similarly demonstrated benefits from assigning separate parameter values to low, 27 28 medium, and high discharge periods (Yilmaz et al., 2008), or based on whether a catchment is in a dry, drying, wet or wetting state (Choi and Beven, 2007). Such 29 temporal dependence of model parameterisation on hydrologic conditions has 30 31 implications for model performance. Krause et al. (2005) compared different statistical metrics of hydrological model performance separately for base-flow 32 periods and storm events to evaluate the performance. The authors found that the 33 34 logarithmic form of the Nash-Sutcliffe efficiency (NSE) value provided more

information on the sensitivity of model performance for discharge simulations 1 during storm events, while the relative form of NSE was better for base flow 2 periods. Similarly, Guse et al. (2014) investigated temporal dynamics of 3 sensitivity of hydrological parameters and SWAT model performance using 4 Fourier amplitude sensitivity test (Reusser et al., 2011) and cluster analysis 5 (Reusser et al., 2009). The authors found that three groundwater parameters were 6 highly sensitive during quick flow, while one evaporation parameter was most 7 8 sensitive during base flow, and model performance was also found to vary 9 significantly for the two flow regimes. Zhang et al. (2011) calibrated SWAT hydrological parameters for periods separated on the basis of six climatic indexes. 10 Model performance improved when different values were assigned to parameters 11 based on six hydroclimatic periods. Similarly, Pfannerstill et al. (2014) found that 12 13 assessment of model performance was improved by considering an additional performance statistic for very low-flow simulations amongst five hydrologically-14 separated regimes. 15

To date, analysis of temporal dynamics of SWAT parameters has 16 predominantly focussed on simulations of discharge rather than water quality 17 constituents. This partly reflects the paucity of comprehensive water quality data 18 for many catchments; near-continuous discharge data can readily be collected but 19 this is not the case for water quality parameters such as suspended sediment or 20 nutrient concentrations. Data collected in monitoring programmes that involve 21 sampling at regular time intervals (e.g. monthly) are often used to calibrate water 22 23 quality models, but these are unlikely to fully represent the range of hydrologic conditions in a catchment (Bieroza et al., 2014). In particular, water quality data 24 collected during storm-flow periods are rarely available for SWAT calibration, 25 thus prohibiting opportunities to investigate how parameter sensitivity varies 26 27 under conditions which can contribute disproportionately to nutrient or sediment 28 transport, particularly in lower-order catchments (Chiwa et al., 2010; Abell et al., 2013). Failure to fully consider storm-flow processes could therefore result in 29 overestimation of model performance. Thus, further research is required to 30 examine how water quality parameters vary during different flow regimes and to 31 32 understand how model uncertainty may vary under future climatic conditions that affect discharge regimes (Brigode et al., 2013). 33

1 In this study, the SWAT model was configured to a relatively small, mixed land use catchment in New Zealand that has been the subject of an intensive water 2 3 quality sampling programme designed to target a wide range of hydrologic conditions. A catchment-wide set of parameters was calibrated using the SUFI-2 4 procedure which is integrated into the SWAT Calibration and Uncertainty 5 Program (SWAT-CUP). The objectives of this study were to: (1) quantify the 6 7 performance of the model in simulating discharge and fluxes of suspended sediments and nutrients at the catchment outlet; (2) rigorously evaluate model 8 9 performance by comparing daily simulation output with monitoring data collected under a range of hydrologic conditions; and (3) quantify whether parameter 10 11 sensitivity varies between base flow and quick flow conditions.

12

13 **2 Methods**

14 2.1 Study area

The Puarenga Stream is the second-largest surface inflow to Lake Rotorua (Bay 15 of Plenty, New Zealand) and drains a catchment of 77 km². The catchment is 16 situated in the central North Island of New Zealand, which has a warm temperate 17 climate. Annual mean temperature at Rotorua Airport (Fig. 1a) is 15±4 °C and 18 annual mean evapotranspiration is 714 mm yr⁻¹ (1993–2012; National Climatic 19 Data Centre; available at http://cliflo.niwa.co.nz/). Annual mean precipitation at 20 Kaituna rain gauge (Fig. 1a) is 1500 mm yr⁻¹ (1993–2012; Bay of Plenty Regional 21 Council). The catchment is relatively steep (mean slope = 9%; Bay of Plenty 22 Regional Council) with predominantly pumice soils that have high macroporosity, 23 resulting in high infiltration rates and substantial sub-surface lateral flow 24 contributions to stream channels. Two cold-water springs (Waipa Spring and 25 26 Hemo Spring) and one geothermal spring (Fig. 1b) are located in the 27 LTS catchment area. Two cold-water springs have annual mean discharge of ~0.19 m³ s⁻¹ (Rotorua District Council) and one geothermal spring has annual 28 mean discharge of $\sim 0.12 \text{ m}^3 \text{ s}^{-1}$ (White et al., 2004). 29

The predominant land use (47%) is exotic forest (*Pinus radiata*). Approximately 26% is managed pastoral farmland, 11% mixed scrub and 9% indigenous forest. Since 1991, treated wastewater has been pumped from the Rotorua Wastewater Treatment Plant and spray–irrigated over 16 blocks of total

area of 1.93 km² in the Whakarewarewa Forest (Fig. 1a). Following this, it took 1 approximately four years before elevated nitrate concentrations were measured in 2 the receiving waters of the Puarenga Stream (Lowe et al., 2007). Prior to 2002, the 3 irrigation schedule entailed applying wastewater to two blocks per day so that 4 each block was irrigated approximately weekly. Since 2002, 10 to 14 blocks have 5 been irrigated simultaneously at daily frequency. Over the entire period of 6 7 irrigation, nutrient concentrations in the irrigated water have gradually decreased as improvements in treatment of the wastewater have been made (Lowe et al., 8 9 2007).

Measurements from the Forest Research Institute (FRI) stream-gauge (1.7 10 11 km upstream of Lake Rotorua; Fig. 1b) were considered representative of the downstream/outlet conditions of the Puarenga Stream. The FRI stream-gauge was 12 13 closed in mid 1997, then reopened late in 2004 (Environment Bay of Plenty, 2007). Annual mean discharge at this site is 2.0 m³ s⁻¹ (1994–1997 and 2004– 14 15 2008; Bay of Plenty Regional Council). The Puarenga Stream receives a high proportion of flow from groundwater stores and has only moderate seasonality in 16 17 discharge. On average, the lowest mean daily discharge is during summer (December to February; $1.7 \text{ m}^3 \text{ s}^{-1}$) and the highest mean daily discharge is during 18 winter (June to August; 2.4 m³ s⁻¹). Discharge records during 1998–2004 were 19 intermittent and this precluded a detailed comparison of measured and simulated 20 discharge during that period. In July 2010, the gauge was repositioned 720 m 21 downstream to the State Highway 30 (SH 30) bridge (Fig. 1b). 22

23 **2.2 Model configuration**

SWAT input data requirements included a digital elevation model (DEM),
meteorological records, records of springs and water abstraction, soil
characteristics, land use classification, and management schedules for key land
uses (pastoral farming, wastewater irrigation, and timber harvesting). The SWAT
model (version used (SWAT2009_rev488) was runs on an hourlya daily time step,
but daily mean simulation outputs were used for this study.

The DEM was used to delineate boundaries of the whole catchment and individual sub-catchments, with a stream map used to 'burn-in' channel locations to create accurate flow routings. Hourly rainfall estimates were used as hydrologic forcing data. The Penman-Monteith method (Monteith, 1965) was used to

calculate evapotranspiration (ET) and potential ET. The Green and Ampt (1911) 1 method was used to calculate infiltration, rather than the SCS curve number 2 method. Therefore, the hourly rainfall/Green & Ampt infiltration/daily-hourly 3 routing method (Neitsch et al., 2011) was chosen to simulate upland and in-4 stream processes. Ten sub-catchments were represented in the Puarenga Stream 5 catchment, each comprising numerous Hydrologic Response Units (HRUs). Each 6 7 HRU aggregates cells with the same combination of land cover, soil, and slope. A total of 404 HRUs was defined in the model. Runoff and nutrient transport were 8 9 predicted separately within SWAT for each HRU, with predictions summed to 10 obtain the total for each sub-catchment.

Descriptions and sources of the data used to configure the SWAT model are given in Table 1. There were a total of 197 model parameters. Values of SWAT parameters were assigned based on: i) measured data (e.g. some of the soil parameters; Table 1); ii) literature values from published studies of similar catchments (e.g. parameters for dominant land uses; Table 2); or iii) by calibration where parameters were not otherwise prescribed.

17 SWAT simulates loads of 'mineral phosphorus' (MINP) and 'organic 18 phosphorus' (ORGP) of which the sum is total phosphorus (TP). The MINP 19 fraction represents soluble P either in mineral or in organic form, while ORGP 20 refers to particulate P bound either by algae or by sediment (White et al., 2014). 21 Soluble P may be taken up during algae growth, or released from benthic 22 sediment. Either This fraction can be transformed to particulate P contained in 23 algae or sediment.

SWAT simulates loads of nitrate-nitrogen (NO₃-N), ammonium-nitrogen 24 (NH₄–N) and organic nitrogen (ORGN), the sum of which is total nitrogen (TN). 25 Nitrogen parameters were auto-calibrated for each N fraction. The SWAT model 26 27 does not account for the initial nitrate concentration in shallow aquifers, as also 28 noted by Conan et al. (2003). Ekanayake and Davie (2005) indicated that SWAT underestimated N loading from groundwater and suggested a modification by 29 adding a background concentration of nitrate in streamflow to represent 30 groundwater nitrate contributions. Over the period of the first five years of 31 wastewater irrigation, nitrate concentrations in shallow groundwater draining the 32 Waipa Stream sub-catchment were estimated to have increased by c. 0.44 mg L⁻¹ 33 34 (Paku, 2001). SWAT has no capability to dynamically adjust the groundwater

concentration during a simulation run. Therefore we added 0.44 mg N L⁻¹ to all 1 model simulations of TN concentration assuming that groundwater concentrations 2 had equilibrated with the applied wastewater nitrogen. 3

4 2.3 Model calibration and validation

Daily mean discharge was firstly calibrated based on daily mean values of 15-5 minute measurements (Table 3). Water quality variables were then calibrated in 6 the sequence: SS, TP and TN. Modelled mean daily concentrations were 7 8 compared with concentrations measured during monthly grab sampling, with 9 monthly measurements assumed equal to daily mean concentrations (Table 3). 10 One year (1993) was used for model warmup. The calibration period was from 2004 to 2008 and the validation period was from 1994 to 1997. A validation 11 12 period that pre-dated the calibration period was chosen because discharge records were available for two separate periods (1994-1997 and post 2004). In addition, 13 the operational regime for the wastewater irrigation has varied since operations 14 began in 1991, with a marked change occurring in 2002 when operations switched 15 from applying the wastewater load to two blocks (rotated daily for a total of 14 16 17 blocks in a week; i.e., each block irrigated weekly), to 10–14 blocks each irrigated daily. This operational regime continues today and we therefore decided to assign 18 the most recent (post 2002) period (2004–2008) to calibration to ensure that the 19 model was configured to reflect current operations. 20

21

Parameter values that were not derived from measurements or the 22 literature were assigned based on either automated or manual calibration (Table 23 34). Manual calibration was undertaken for 11 parameters related to TP, while a 24 Sequential Uncertainty Fitting (SUFI-2) procedure was applied to auto-calibrate 25 21 parameters for discharge simulations, nine parameters for SS simulations, and 17 parameters related to TN. The SUFI-2 procedure has been integrated into the 26 SWAT Calibration and Uncertainty Program (SWAT-CUP). SUFI-2 is a 27 procedure that efficiently quantifies and constrains parameter uncertainties/ranges 28 from default ranges with the fewest number of iterations (Abbaspour et al., 2004), 29 30 and has been shown to provide optimal results relative to the use of alternative 31 algorithms (Wu and Chen, 2015). SUFI-2 involves Latin hypercube sampling (LHS), which is a method that generates a sample of plausible parameter values 32 from a multidimensional distribution and ensures that samples cover the entire 33

parameter space, therefore ensuring that the optimum solution is not a local
 minimum (Marino et al., 2008).

The SUFI-2 procedure analyses relative sensitivities of parameters by 3 4 randomly generating combinations of values for model parameters (Abbaspour et al., 2014). A sample size of 1000 was chosen for each iteration of LHS, resulting 5 in 1000 combinations of parameters and 1000 simulations. Model performance 6 7 was quantified for each simulation based on the Nash–Sutcliffe efficiency (NSE). 8 An objective function was defined as a linear regression of a combination of parameter values generated by each LHS against the NSE value calculated from 9 each simulation. Each compartment was not given weight to formulate the 10 objective function because only one variable was specifically focused on at each 11 time. A parameter sensitivity matrix was then computed based on the changes in 12 the objective function after 1000 simulations. Parameter sensitivity was quantified 13 14 based on the p value from a Student's t-test, which was used to compare the mean 15 of simulated values with the mean value of measurements (Rice, 2006). A parameter was deemed sensitive by if $p \le 0.05$ after 1000 simulations (one 16 iteration). Numerous iterations of LHS were conducted. Values of p from 17 numerous iterations were averaged for each parameter, and the frequency of 18 19 iterations where a parameter was deemed sensitive was summed. Rankings of 20 relative sensitivities of parameters were developed based on how frequently the sensitive parameter was identified and the averaged value of p calculated from 21 22 several iterations. The most sensitive parameter was determined based on the frequency that the parameter was deemed sensitive, and the smallest average p-23 value from all iterations. 24

SUFI-2 considers two criteria to constrain uncertainty in each iteration. 25 One is the P-factor, the percentage of measured data bracketed by 95% prediction 26 uncertainty (95PPU). Another is the R-factor, the average thickness of the 95PPU 27 28 band divided by the standard deviation of measured data. A range was first 29 defined for each parameter based on a synthesis of ranges from similar studies or 30 from the SWAT default range. Parameter ranges were updated after each iteration based on the computation of upper and lower 95% confidence limits. The 95% 31 32 confidence interval and the standard deviation of a parameter value were derived from the diagonal elements of the covariance matrix, which was calculated from 33

the sensitivity matrix and the variance of the objective function. Steps and
 equations used in the SUFI-2 procedure to constrain parameter ranges are
 outlined by Abbaspour et al. (2004).

The total numbers of iterations performed for each simulated variable (Q, 4 SS, MINP, ORGN, NH₄–N and NO₃–N) reflected the numbers required to ensure 5 that > 90% of measured data were bracketed by simulated output and the R-factor 6 7 was close to one. The 'optimal' parameter value was obtained when the Nash-Sutcliffe efficiency (NSE) criterion was satisfied (NSE > 0.5; Moriasi et al., 2007). 8 9 Auto-calibrated parameters for simulations of Q, SS, and TN were changed by absolute values within the given ranges. Some of those given ranges were 10 restricted based on the optimum values calibrated in similar studies. Parameter 11 values for TP simulations were manually-calibrated based on the relative percent 12 deviation from the predetermined values of those auto-calibrated parameters for 13 MINP simulations, given by the objective functions (e.g., NSE). Parameters 14 related to the physical characteristics of the catchment were not changed because 15 their values were considered to be representative of the catchment characteristics. 16

In addition, high-frequency (1-2 h) water quality sampling was 17 undertaken at the FRI stream-gauge during 2010-2012 (Table 3) to derive 18 estimates of daily mean contaminant loads during storm events. Samples were 19 20 analysed for SS (nine events), TP and TN (both 14 events) over sampling periods of 24–73 h. The sampling programme was designed to encompass pre-event base 21 22 flow, storm generated quick flow and post-event base flow (Abell et al., 2013). These data permitted calculation of daily discharge-weighted (Q-weighted) mean 23 24 concentrations to compare with modelled daily mean estimates. We did not use 25 the high-frequency observations to calibrate the model, because of the limited 26 number of high-frequency (1-2 h) samples (nine events for SS and 14 events for 27 TP and TN in 2010–2012). The use of the high–frequency observations for model validation allowed to examine how the model performed during short (1-3 day) 28 high flow periods. The Q-weighted mean concentrations C_{QWM} were calculated as: 29

$$30 \qquad C_{\text{QWM}} = \frac{\sum_{i=1}^{n} C_i Q_i}{\sum_{i=1}^{n} Q_i} \tag{1}$$

31 where n is number of samples, C_i is contaminant concentration measured at time i, 32 and Q_i is discharge measured at time i.

2.4 Hydrograph and contaminant load separation 1

2 The Web-based Hydrograph Analysis Tool (Lim et al., 2005) was applied to partition both measured and simulated discharges into base flow (Q_b) and quick 3 flow (Q_q) . An Eckhardt filter parameter of 0.98 and ratio of base flow to total 4 discharge of 0.8 were assumed (cf. Lim et al., 2005). There were was a total of 60 5 days without quick flow during the calibration period (2004–2008) and 1379 days 6 for which hydrograph separation defined both base flow and quick flow. 7

8 Contaminant (SS, TP and TN) concentrations (C_{sep}) were partitioned into base flow ($C_{b}^{'}$) and quick flow components ($C_{q}^{'}$; cf. Rimmer and Hartmann, 2014) 9 to separately examine the sensitivity of water quality parameters during base flow 10 and quick flow: 11

12
$$C_{\text{sep}} = \frac{Q_{q} \times C_{q} + Q_{b} \times C_{b}}{Q_{q} + Q_{b}}$$
(2)

13

 $C_{\rm b}^{'}$ for each contaminant was estimated as the average concentration for the 60 days with no quick flow. $C_q^{'}$ for each contaminant was calculated by 14 15 rearranging Eq. (2).

To ensure that $C_q^{'}$ is positive, $C_b^{'}$ is constrained to be the minimum of 16 $\overline{C_{sep}}$ and $\overline{C_{sep}}$. Measured and simulated base flow and quick flow contaminant 17 loads were then calculated. 18

A one-at a-time (OAT) routine proposed by Morris (1991) was applied to 19 investigate how parameter sensitivity varied between the two flow regimes (base 20 21 flow and quick flow), based on the ranking of relative sensitivities of parameters that were identified by randomly generating combinations of values for model 22 23 parameters for each individual variable using the SUFI-2 procedure. OAT sensitivity analysis was then employed by varying the parameter of interest 24 25 among ten equidistant values within the default range. The natural logarithm was used by Krause et al. (2005) and therefore the standard deviation (STD) of the ln-26 transformed NSE were was used to indicate parameter sensitivity for the two flow 27 regimes. 28

Parameters were ranked from most to least sensitive on the basis of the 29 sensitivity metric (STD of In-transformed NSE), using a value of 0.2 as a 30 threshold above which parameters were deemed particularly 'sensitive'. The 31

threshold value of "0.2" was chosen in this study, based on the median value
derived from the calculations of the *STD* of ln–transformed NSE. Methods used to
separate the two flow constituents and to quantify parameter sensitivity are
illustrated in Fig. 2.

5 **2.5 Model evaluation**

Model goodness-of-fit was assessed graphically and quantified using coefficient
of determination (R²), Nash-Sutcliffe efficiency (NSE) and percent bias (PBIAS;
Table 4<u>5</u>). R² (range 0 to 1) and NSE (range -∞ to 1) values are commonly used
to evaluate SWAT model performance at daily time step (Gassman et al., 2007).
PBIAS value indicates the average tendency of simulated outputs to be larger or
smaller than observations (Gupta et al., 1999).

Model uncertainty was evaluated by two criteria;-:_R-factor and P-factor (see Section 2.3). They were used to constrain parameter ranges during the calibration using measured Q and loads of SS, MINP, ORGN, NH₄-N and NO₃-N in the SUFI-2 procedure. The R software (<u>R Development Core Team</u>) was used to graphically show the 95% confidence and prediction intervals for measurement data (Neyman, 1937) and model prediction intervals (Seymour, 1993) for Q and concentrations of SS, TP and TN during the calibration period (2004–2008).

19

20 **3 Results**

21 **3.1 Model performance and uncertainty**

Numerous rounds (each comprising 1000 iterations) of LHS were conducted for 22 each simulated variable until the performance criteria were satisfied. The total 23 number of rounds of LHS for each simulated variable was as follows (number in 24 25 parentheses): Q (7), SS (7), MINP (11), ORGN (10), NH₄-N (4) and NO₃-N (4). The parameters that provided the best statistical outcomes (i.e, best match to 26 27 observed data) are given in Table <u>34</u>. Two criteria (R-factor and P-factor) were used to show model uncertainties for simulations of discharge and contaminant 28 29 loads, with values as follows: Q (0.97, 0.43), SS (0.48, 0.19), MINP (2.64, 0.14), ORGN (0.47, 0.17), NH₄-N (1.16, 0.56) and NO₃-N (1.2, 0.29). Model 30 uncertainties for simulations of Q and SS, TP and TN concentrations are shown in 31 Fig. 6. 32

1 Modelled and measured base flow showed high correspondence, although measured daily mean discharge during storm peaks was often underestimated (Fig. 2 3a and 3e). Annual mean percentages of lateral flow recharge, shallow aquifer 3 recharge and deep aquifer recharge to total water yield were predicted by SWAT 4 as 30%, 10%, 58%, respectively. Modelled SS concentrations overestimated 5 measurements of monthly grab samples by an average of 18.3% during calibration 6 7 and 0.32% during validation (Fig. 3b and 3f). Measured TP concentrations in monthly grab samples were underestimated by 23.8% during calibration (Fig. 3c) 8 and 24.5% during validation (Fig. 3g). Similarly, measured TP loads were 9 underestimated by 34.5% and 38.4%, during calibration and validation, 10 11 respectively. Modelled and measured TN concentrations were generally better aligned during base flow (Fig. 3d), apart from a mismatch prior to 1996 when 12 13 monthly measured TN concentrations were substantially lower than model predictions, although the concentrations gradually increased (Fig. 3h) during the 14 15 validation period (1994–1997). The average measured TN load increased from 134 kg N d⁻¹ prior to 1996, to 190 kg N d⁻¹ post 1996. The comparable increase in 16 modelled TN load was 167 kg N d⁻¹ to 205 kg N d⁻¹, respectively. 17

Statistical evaluations of goodness–of–fit are shown in Table $\frac{56}{2}$. The R² 18 values for discharge were 0.77 for calibration and 0.68 for validation, 19 corresponding to model performance ratings (cf. Moriasi et al., 2007) of 'very 20 good' and 'good' (Table 45). Similarly, the NSE values for discharge were 0.73 21 22 (good) for calibration and 0.62 (satisfactory) for validation. Positive PBIAS (7.8% for calibration and 8.8% for validation) indicated a tendency for underestimation 23 of daily mean discharge, however, the low magnitude of PBIAS values 24 corresponded to a performance rating of 'very good'. The R² values for SS were 25 0.42 (unsatisfactory) for calibration and 0.80 for validation (very good). Similarly, 26 the NSE values for SS were -0.08 (unsatisfactory) for calibration and 0.76 (very 27 good) for validation. The model did not simulate trends well for monthly 28 measured TP and TN concentrations. The R^2 values for TP and TN were both < 29 0.1 (unsatisfactory) during calibration and validation and NSE values were both <30 0 (unsatisfactory). Values of PBIAS corresponded to 'good' or 'very good' 31 32 performance ratings for TP and TN.

1 Observed Q-weighted daily mean concentrations derived from hourly 2 measurements and simulated daily mean concentrations of SS, TP and TN during an example two-day storm event are shown in Fig. 4a-4c. The simulations of SS 3 4 and TN concentrations was were somewhat better than for TP. Comparisons of 5 Q-weighted daily mean concentrations (C_{OWM}) during storm events from 2010 to 6 2012 are shown in Fig. 4d–4f for SS (nine events), TP and TN (both 14 events). The C_{QWM} of TP exceeded the simulated daily mean by between 0.02 and 0.2 mg 7 8 P L^{-1} , and on average, the model underestimated measurements by 69.4% (Fig. 4e). Although R^2 and NSE values for C_{QWM} of TN were unsatisfactory (Table <u>56</u>), 9 they were both close to the threshold for satisfactory performance (0.5). For C_{QWM} 10 of SS and TP, R² and NSE values indicated that the model performance was 11 unsatisfactory. The PBIAS value of -0.87 for $C_{\rm QWM}$ of TN corresponded to model 12 performance ratings of 'very good', while the PBIAS values for C_{OWM} of SS and 13 TP were 43.9 and 69.4, respectively, indicating satisfactory model performance. 14

15 Measured and simulated discharge and contaminant loads separated for the two flow regimes (base flow and quick flow) are shown in Fig. 5. Model 16 17 performance statistics differed between the two flow regimes (Table 67). Simulations of discharge and constituent loads under quick flow were more 18 closely related to the measurements (i.e., higher values of R² and NSE) than 19 simulations under base flow. Base flow TN load simulations during the validation 20 period showed better model performance than simulations under quick flow. 21 Additionally, measurements under quick flow were better reproduced by the 22 23 model than the measurements for the whole simulation period. Simulations of 24 contaminant loads matched measurements much better than for contaminant concentrations, as indicated by statistical values for model performance given in 25 Table 5-6 and 67. 26

27 **3.2 Separated parameter sensitivity**

Based on the ranking of relative sensitivities of hydrological and water quality parameters derived from the SUFI-2 procedure (see Table 78), the OAT sensitivity analysis undertaken separately for base flow and quick flow identified three parameters that most influenced the quick flow estimates, and five parameters that most influenced the base flow estimates (parameters above the dashed line in Fig. 7a). Channel hydraulic conductivity (CH_K2) is used to

estimate the peak runoff rate (Lane, 1983). Lateral flow slope length (SLSOIL) 1 and lateral flow travel time (LAT_TIME) have an important controlling effect on 2 the amount of lateral flow entering the stream reach during quick flow. Both slope 3 (HRU_SLP) and soil available water content (SOL_AWC) were particularly 4 sensitive for the base flow simulation because they affect lateral flow within the 5 kinematic storage model in SWAT (Sloan and Moore, 1984). The aquifer 6 7 percolation coefficient (RCHRG_DP) and the base flow alpha factor (ALPHA_BF) strongly influenced base flow calculations (Sangrey et al., 1984), 8 9 as did the channel Manning's N value (CH N2) which is used to estimate channel flow (Chow, 2008). 10

For SS loads, 12 and four parameters, respectively, were identified as 11 sensitive in relation to the simulations of base flow and quick flow (parameters 12 13 above the dashed line in Fig. 7b). Parameters that control main channel processes (e.g. CH_K2 and CH_N2) and subsurface water transport processes (e.g. 14 LAT TIME and SLSOIL) were found to be much more sensitive for base flow SS 15 load estimations. Exclusive parameters for SS estimations, such as SPCON (linear 16 17 parameter), PRF (peak rate adjustment factor), SPEXP (exponent parameter), CH_COV1 (channel erodibility factor), and CH_COV2 (channel cover factor) 18 were found to be much more sensitive in base flow SS load, while LAT_SED (SS 19 concentration in lateral flow and groundwater flow) was more sensitive in quick 20 flow SS load. Parameters that control overland processes, e.g. CN2 (the curve 21 number), OV N (overland flow Manning's N value) and SLSUBBSN (sub-basin 22 slope length), were found to be much more sensitive for quick flow SS load 23 estimations. 24

Of the sensitive parameters, BC4 (ORGP mineralization rate) was 25 particularly sensitive for the simulation of base flow MINP load (Fig. 7c). RCN 26 (nitrogen concentration in rainfall) related specifically to the dynamics of the base 27 28 flow NO₃–N load and NPERCO (nitrogen percolation coefficient) significantly affected quick flow NO₃-N load (Fig. 7d). Parameter CH_ONCO (channel ORGN 29 concentration) similarly affected both flow components of ORGN load (Fig. 7e) 30 and SOL_CBN (organic carbon content) was most sensitive for the simulations of 31 quick flow ORGN and NH₄–N loads. Parameter BC1 (nitrification rate in reach) 32 was particularly sensitive for the simulation of base flow NH₄–N load (Fig. 7f). 33

1 4 Discussion

This study examined temporal dynamics of model performance and parameter 2 sensitivity in a SWAT model application that was configured for a small, 3 relatively steep and lower order stream catchment in New Zealand. This country 4 faces increasing pressures on freshwater resources (Parliamentary Commissioner 5 for the Environment, 2013) and models such as SWAT potentially offer valuable 6 7 tools to inform management of water resources although, to date, the SWAT model has received limited consideration in New Zealand (Cao et al., 2006). 8 9 Model evaluation on the basis of the data collected during an extended monitoring programme enabled a detailed examination of how model performance varied 10 11 during different flow regimes. It also permitted error in daily mean estimates of 12 contaminant loads to be quantified with relative precision, allowing assessment of 13 the ability of the SWAT model to simulate contaminant loads during storm events when lower-order streams typically exhibit considerable sub-daily variability in 14 15 both discharge and contaminant concentrations (Zhang et al., 2010). Separating discharge and loads of sediments and nutrients into those associated with base 16 17 flow and quick flow for separate OAT sensitivity analyses provided important insights into the varying dependency of parameter sensitivity on hydrologic 18 conditions. 19

20 4.1 Temporal dynamics of model performance

21 The modelled estimates of deep aquifer recharge (58%) and combined lateral flow 22 and shallow aquifer recharge (40%) were comparable with estimates derived by Rutherford et al. (2011), who used an alternative catchment model to derive 23 respective estimates of 30% and 70% for these two fluxes. Our decision to 24 deliberately select a validation period (1994–1997) during which the boundary 25 conditions of the system (specifically anthropogenic nutrient loading) differed 26 considerably from the calibration period allowed us to rigorously assess the 27 capability of SWAT to accurately predict water quality under an altered 28 management scenario (i.e. the purpose of most SWAT applications). 29

30 Overestimation of TN concentrations prior to 1996 reflects higher NO₃–N 31 concentrations in groundwater during the calibration period (2004–2008) due to 32 the wastewater irrigation operation. Nitrate concentrations appeared to reach a 33 new quasi–steady state as wastewater loads and in–stream attenuation came into balance. SWAT may not adequately represent the dynamics of groundwater
nutrient concentrations (Bain et al., 2012) particularly in the presence of changes
in catchment inputs (e.g., with start-up of wastewater irrigation). The
groundwater delay parameter was set to five years (cf. Rotorua District Council,
2006), but this did not appear to capture adequately the lag in response to
increases in stream nitrate concentrations following wastewater irrigation from
1991.

The poor fit between simulated daily mean TP concentrations and monthly 8 9 instantaneous measurements may partly reflect a mismatch between the dominant processes affecting phosphorus cycling in the stream and those represented in 10 11 SWAT. The ORGP fraction that is simulated in SWAT includes both organic and inorganic forms of particulate phosphorus, however, the representation of 12 13 particulate phosphorus cycling only focusses on organic phosphorus cycling, with limited consideration of interactions between inorganic streambed sediments and 14 15 dissolved reactive phosphorus in the overlying water (White et al., 2014). This contrasts with phosphorus cycling in the study stream where it has been shown 16 17 that dynamic sorption processes between the dissolved and particulate inorganic phosphorus pools exert major control on phosphorus cycling (Abell and Hamilton, 18 2013). 19

Our finding that measured Q-weighted mean concentrations (C_{OWM}) of TP 20 and SS during storm events (2010-2012) were greatly underestimated relative to 21 simulated daily mean TP and SS concentrations has important implications for 22 studies that examine effects of altered flow regimes on contaminant transport. For 23 example, studies which simulate scenarios comprising more frequent large rainfall 24 events (associated with climate change predictions for many regions; IPCC, 2013) 25 may considerably underestimate projected future loads of SS and associated 26 27 particulate nutrients if only base flow water quality measurements (i.e. those 28 predominantly collected during 'state of environment' monitoring) are used for 29 calibration/validation (see Radcliffe et al., 2009 for a discussion of this issue in relation to phosphorus). This is also reflected by the two model performance 30 31 statistics relating to validation of modelled SS concentrations using monthly grab samples (predominantly base flow; 'very good') and COWM estimated during 32 storm sampling ('unsatisfactory') based on R² and NSE values. 33

1 4.2 Key uncertainties

Model uncertainty in this study may arise from four main factors: 1) model 2 parameters; 2) forcing data; 3) in measurements used for evaluation of model fit, 3 and; 4) model structure or algorithms (Lindenschmidt et al., 2007). The values of 4 most parameters assigned for model calibration, although specific to different soil 5 types (e.g. soil parameters), were lumped across land uses and slopes in this study. 6 7 They integrated spatial and temporal variations, thus neglecting any variability throughout the study catchment. In terms of forcing data, the assumption of 8 9 constant values of spring discharge rate and nutrient concentrations may inadequately reflect the temporal variability and therefore increase model 10 11 uncertainty, although this should contribute little to the model error term. Most water quality data used for model calibration comprised monthly instantaneous 12 13 samples taken during base flow conditions. The use of those measurements for model calibration would likely lead to considerable underestimation of constituent 14 15 concentrations (notably SS and TP) due to failure to account for short-term high flow events. Inadequate representation of groundwater processes in the model 16 17 structure is another key factor that is likely to affect model uncertainty, particularly for nitrogen simulations. The analysis of model performance based on 18 datasets separated into base flow and quick flow constituents enabled 19 uncertainties in the structure of hydrological models to be identified, denoted by 20 different model performance between these two flow constituents. Furthermore, 21 the disparity in goodness-of-fit statistics between discharge (typically 'good' or 22 'very good') and nutrient variables (often 'unsatisfactory') highlights the potential 23 for catchment models which inadequately represent contaminant cycling 24 processes (manifest in unsatisfactory concentration estimates) to nevertheless 25 produce satisfactorily load predictions (e.g., compare model performance statistics 26 27 for prediction of nutrient concentrations in Table $\frac{5-6}{6}$ with statistics for prediction 28 of loads in Table 67). This highlights the potential for model uncertainty to be underestimated in studies which aim to predict the effects of scenarios associated 29 30 with changes in contaminant cycling, such as increases in fertiliser application 31 rates.

1 4.3 Temporal dynamics of parameter sensitivity

To date, studies of temporal variability of parameters have focused on 2 hydrological parameters, rather than on water quality parameters. The 3 characteristics of concentration-discharge relationships for SS and TP are 4 different to that for TN (Abell et al., 2013). In quick flow, there is a positive 5 relationship between Q and concentrations of SS and TP, reflecting mobilisation 6 7 of sediments and associated particulate P. Total nitrogen concentrations declined slightly in quick flow, reflecting the dilution of nitrate from groundwater. 8 9 Defining separate contaminant concentrations in base flow and quick flow enabled us to examine how the sensitivity of water quality parameters varied 10 11 depending on hydrologic conditions.

In a study of a lowland catchment (481 km²), Guse et al. (2014) found that 12 three groundwater parameters, RCHRG_DP (aquifer percolation coefficient), 13 GW_DELAY (groundwater delay) and ALPHA_BF (base flow alpha factor) were 14 15 highly sensitive in relation to simulating discharge during quick flow, while ESCO (soil evaporation compensation factor) was most sensitive during base flow. 16 17 This is counter to the findings of this study for which the base-flow discharge simulation was sensitive to RCHRG_DP and ALPHA_BF. This result may reflect 18 that, relative to our study catchment, the catchment studied by Guse et al. (2014) 19 had moderate precipitation (884 mm y⁻¹) with less forest cover and flatter 20 topography. Although the GW_DELAY parameter reflects the time lag that it 21 takes water in the soil water to enter the shallow aquifers, its lack of sensitivity 22 under both base flow and quick flow conditions in this study is a reflection of 23 higher water infiltration rates and steeper slopes. The ESCO parameter controls 24 the upwards movement of water from lower soil layers to meet evaporative 25 26 demand (Neitsch et al., 2011). Its lack of sensitivity in our study may reflect relatively high and seasonally-consistent rainfall (1500 mm y⁻¹), in addition to 27 extensive forest cover in the Puarenga Stream catchment, which reduces soil 28 evaporative demand by shading. Soil texture is also likely a contributor to this 29 result. The predominant soil horizon type in the Puarenga Stream catchment was 30 A, indicating high macroporosity which promotes high water infiltration rate and 31 inhibits upward transport of water by capillary action (Neitsch et al., 2011). The 32 variability in the sensitivity of the parameter SURLAG (surface runoff lag 33 34 coefficient) between this study (relatively insensitive) and that of Cibin et al.

(2010; relatively sensitive) likely reflects differences in catchment size. The 1 Puarenga Stream catchment (77 km²) is much smaller than the study catchment 2 (St Joseph River; 2800 km²) of Cibin et al. (2010) and, consequently, distances to 3 the main channel are much shorter, with less potential for attenuation of surface 4 runoff in off-channel storage sites. The curve number (CN2) parameter was found 5 to be insensitive in both this study and Shen et al. (2012), because surface runoff 6 7 was simulated based on the Green and Ampt method (1911) requiring the hourly rainfall inputs, rather than the curve number equation which is an empirical model. 8 9 By contrast, the most sensitive parameters in our study are those that determine the extent of lateral flow, an important contributor to streamflow in the catchment, 10 due to a general lack of ground cover under plantation trees and formation of 11 gully networks on steep terrain. 12

13 Parameters that control surface water transport processes (e.g. LAT_TIME and SLSOIL) were found to be much more sensitive for base flow SS load 14 estimation than parameters that control groundwater processes (e.g. ALPHA BF 15 and RCHRG_DP), reflecting the importance of surface flow processes for 16 17 sediment transport. Sensitive parameters for quick flow SS load estimation related to overland flow processes (e.g. OV_N and SLSUBBSN), thus reflecting the fact 18 that sediment transport is largely dependent on rainfall-driven processes, as is 19 typical of steep and lower-order catchments. Modelled base flow NO₃-N loads 20 were most sensitive to the nitrogen concentration in rainfall (RCN) because of 21 rainfall as a predominant contributor to recharging base flow. The nitrogen 22 percolation coefficient (NPERCO) was more influential for quick flow NO₃-N 23 load estimation, probably indicating that the quick flow NO₃-N load is more 24 influenced by the mobilisation of concentrated nitrogen sources associated with 25 26 agriculture or treated wastewater distribution. High sensitivity of the organic 27 carbon content (SOL_CBN) for quick flow ORGN load estimates likely reflects 28 mobilisation of N associated with organic material following rainfall. The finding that base flow NH₄–N load was more sensitive to nitrification rate in reach (BC1) 29 30 likely reflects that base flow provides more favourable conditions to complete this 31 oxidation reaction, as NH₄–N is less readily leached and transported. Similarly, 32 the ORGP mineralization rate (BC4) strongly influenced base flow MINP load estimation, reflecting that base flow phosphorus transport is relatively more 33 34 influenced by cycling from channel bed stores, whereas quick flow phosphorus

transport predominantly reflects the transport of phosphorus that originated from
 sources distant from the channel.

3

4 5 Conclusions

The performance of a SWAT model was quantified for different hydrologic 5 6 conditions in a small catchment with mixed land use. Discharge-weighted mean concentrations of TP and SS measured during storm events were greatly 7 8 underestimated by SWAT, highlighting the potential for uncertainty to be greatly 9 underestimated in catchment model applications that are validated using a sample 10 of contaminant load measurements that is over-represented by measurements made during base flow conditions. Monitoring programmes which collect high-11 12 frequency and event-based data have an important roleshould be considered further to in-supporting themore robust calibration and validation of SWAT model 13 applications. Accurate simulation of nitrogen concentrations was constrained by 14 the non-steady state of groundwater nitrogen concentrations due to historic 15 anthropogenic nitrogen applications 16 variability in to land. Improved 17 representation of groundwater processes in the model structure would reduce this aspect of model uncertainty. The sensitivity of many parameters varied depending 18 on the relative dominance of base flow and quick flow, while curve number, soil 19 evaporation compensation factor, surface runoff lag coefficient, and groundwater 20 delay were largely invariant to the two flow regimes. Parameters relating to main 21 22 channel processes were more sensitive when estimating variables (particularly Q 23 and SS) during base flow, while those relating to overland processes were more sensitive for simulating variables associated with quick flow. Temporal dynamics 24 25 of both parameter sensitivity and model performance due to dependence on hydrologic conditions should be considered in further model applications. 26 Monitoring programmes which collect high-frequency and event-based data have 27 an important role in supporting the robust calibration and validation of SWAT 28 model applications. This study has important implications for modelling studies of 29 similar catchments that exhibit short-term temporal fluctuations in stream flow. In 30 31 particular these include small catchments with relatively steep terrain and lower order streams with moderate to high rainfall. 32

1 Acknowledgements

2 This study was funded by the Bay of Plenty Regional Council and the Ministry of Business, Innovation and Employment (Outcome Based Investment in Lake 3 Biodiversity Restoration UOWX0505). We thank the Bay of Plenty Regional 4 Council (BoPRC), Rotorua District Council (RDC) and Timberlands Limited for 5 6 assistance with data collection. In particular, we thank Alison Lowe (RDC), 7 Alastair MacCormick (BoPRC), Craig Putt (BoPRC) and Ian Hinton (Timberlands Limited). Theodore Alfred Kpodonu (University of Waikato) is 8 thanked for assisting with manuscript preparation. 9

1 References

- Abbaspour, K.C.: Swat–Cup4: SWAT Calibration and Uncertainty Programs
 Manual Version 4, Department of Systems Analysis, Integrated
 Assessment and Modelling (SIAM), Eawag, Swiss Federal Institute of
 Aquatic Science and Technology, Duebendorf, Switzerland, pp 106, 2014.
- Abbaspour, K.C., Johnson, C.A., and van Genuchten, M.T.H.: Estimating
 uncertain flow and transport parameters using a sequential uncertainty
 fitting procedure, Vadose Zone J., 3, 1340–1352, doi:
 10.2136/vzj2004.1340, 2004.
- 10 Abbaspour, K.C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J.,
- Zobrist, J., and Srinivasan, R.: Modelling hydrology and water quality in
 the pre–alpine/alpine Thur watershed using SWAT, J. Hydrol., 333, 413–
 430, doi: 10.1016/j.jhydrol.2006.09.014, 2007.
- Abell, J.M. and Hamilton, D.P.: Bioavailability of phosphorus transported during
 storm flow to a eutrophic polymictic lake, New Zeal. J. Mar. Fresh., 47,
 481–489, doi: 10.1080/00288330.2013.792851, 2013.
- Abell, J.M., Hamilton, D.P., and Rutherford, J.C.: Quantifying temporal and
 spatial variations in sediment, nitrogen and phosphorus transport in stream
 inflows to a large eutrophic lake, Environ. Sci.: Processes Impacts, 15,
 1137–1152, doi: 10.1039/c3em00083d, 2013.
- Arnold, J.G., Srinivasan, R., Muttiah, R.S., and Williams, J.R.: Large area
 hydrologic modeling and assessment Part I: Model development, J. Am.
 Water Resour. As., 34, 73–89, doi: 10.1111/j.1752-1688.1998.tb05961.x,
 1998.
- Bain, D.J., Green, M.B., Campbell, J.L., Chamblee, J.F., Chaoka, S., Fraterrigo,
 J.M., Kaushal, S.S., Martin, S.L., Jordan, T.E., and Parolari, A.J.: Legacy
 effects in material flux: structural catchment changes predate long-term
 studies, BioScience, 62, 575–584, doi: 10.1525/bio.2012.62.6.8, 2012.
- Bi, H.Q., Long, Y.S., Turner, J., Lei, Y.C., Snowdon, P., Li, Y., Harper, R.,
 Zerihun, A., and Ximenes, F.: Additive prediction of aboveground
 biomass for *Pinus radiata* (D. Don) plantations, Forest Ecol. Manag., 259,
 2301–2314, doi: 10.1016/j.foreco.2010.03.003, 2010.

1	Bieroza, M.Z., Heathwaite, A.L., Mullinger, N.J., and Keenan, P.O.:
2	Understanding nutrient biogeochemistry in agricultural catchments: the
3	challenge of appropriate monitoring frequencies, Environ. Sci.: Processes
4	Impacts, 16, 1676–1691, doi: 10.1039/c4em00100a, 2014.
5	Boyle, D.P., Gupta, H.V., and Sorooshian, S.: Toward improved calibration of
6	hydrologic models: Combining the strengths of manual and automatic
7	methods, Water Resour. Res., 36, 3663-3674,
8	doi: 10.1029/2000WR900207, 2000.
9	Brigode, P., Oudin, L., and Perrin, C.: Hydrological model parameter instability:
10	A source of additional uncertainty in estimating the hydrological impacts
11	of climate change?, J. Hydrol., 476, 410-425, doi:
12	10.1016/j.jhydrol.2012.11.012, 2013.
13	Cao, W., Bowden, W.B., Davie, T., and Fenemor, A.: Multi-variable and multi-
14	site calibration and validation of SWAT in a large mountainous catchment
15	with high spatial variability, Hydrol. Process., 20, 1057–1073,
16	doi: 10.1002/hyp.5933, 2006.
17	Chiwa, M., Ide, J., Maruno, R., Higashi, N., and Otsuki, K.: Effects of storm flow
18	samplings on the evaluation of inorganic nitrogen and sulfate budgets in a
19	small forested watershed, Hydrol. Process., 24, 631-640, doi:
20	10.1002/hyp.7557, 2010.
21	Choi, H.T. and Beven, K.J.: Multi-period and multi-criteria model conditioning
22	to reduce prediction uncertainty in an application of TOPMODEL within
23	the GLUE framework, J. Hydrol. (NZ), 332, 316-336, doi:
24	10.1016/j.jhydrol.2006.07.012, 2007.
25	Chow, V.T.: Open-channel hydraulics, Blackburn Press, Caldwell, New Jersey,
26	2008.
27	Cibin, R., Sudheer, K.P., and Chaubey, I.: Sensitivity and identifiability of stream
28	flow generation parameters of the SWAT model, Hydrol. Process., 24,
29	1133–1148, doi: 10.1002/hyp.7568, 2010.
30	Conan, C., Bouraoui, F., Turpin, N., de Marsily, G., and Bidglio, G.: Modelling
31	flow and nitrate fate at catchment scale in Britany (France), J. Environ.
32	Qual., 32, 2026–2032, doi:10.2134/jeq2003.2026, 2003.Dairying Research
33	Corporation, AgResearch, Fert Research: Fertilizer use on New Zealand
34	Dairy Farms, In New Zealand Fertiliser Manufacturers' Research

1	Association, Roberts, A.H.C. and Morton, J.D. (eds), Auckland, New
2	Zealand, 36, 1999.
3	Eckhardt, K. and Arnold, J.G.: Automatic calibration of a distributed catchment
4	model, J. Hydrol., 251, 103–109, 2001.
5	Ekanayake, J. and Davie, T.: The SWAT model applied to simulating nitrogen
6	fluxes in the Motueka River catchment, Landcare Research ICM Report
7	2004–05/04, Landcare Research, Lincoln, New Zealand, 18, 2005.
8	Environment Bay of Plenty: Historical data summary, Report prepared for Bay of
9	Plenty Regional Council, New Zealand, 522. 2007.
10	Fert Research: Fertilizer Use on New Zealand Sheep and Beef Farms, In New
11	Zealand Fertiliser Manufacturers' Research Association, Balance, J.M.
12	and Ravensdown, A.R. (eds), Newmarket, Auckland, New Zealand, 52,
13	2009.
14	Gassman, P.W., Reyes, M.R., Green, C.H., and Arnold, J.G.: The Soil and Water
15	Assessment Tool: Historical development, applications, and future
16	research directions, T. ASABE, 50, 1211–1250, 2007.
17	Glover, R.B.: Rotorua Chemical Monitoring to June 1993, GNS Client Report
18	prepared for Bay of Plenty Regional Council, #722305.14, Bay of Plenty
19	Regional Council, New Zealand, 38, 1993.
20	Green, W.H. and Ampt, G.A.: Studies on soil physics, part I – the flow of air and
21	water through soils, J. Agr. Sci., 4, 1–24, doi:
22	10.1017/S0021859600001441, 1911.
23	Gupta, H.V., Sorooshian, S., and Yapo, P.O.: Status of automatic calibration for
24	hydrologic models: Comparison with multilevel expert calibration, J.
25	Hydrol. Eng., 4, 135–143, doi: 10.1061/(ASCE)1084–0699(1999)4:2(135)
26	1999.
27	Guse, B., Reusser, D.E., and Fohrer, N.: How to improve the representation of
28	hydrological processes in SWAT for a lowland catchment-temporal
29	analysis of parameter sensitivity and model performance, Hydrol. Process.,
30	28, 2651–2670, doi: 10.1002/hyp.9777, 2014.
31	Hall, G.M.J., Wiser, S.K., Allen, R.B., Beets, P.N., and Goulding, C.J.: Strategies
32	to estimate national forest carbon stocks from inventory data: the 1990
33	New Zealand baseline, Glob. Change Biol., 7, 389-403,
34	doi: 10.1046/j.1365–2486.2001.00419.x, 2001.

1	Hopmans, P. and Elms, S.R.: Changes in total carbon and nutrients in soil profiles
2	and accumulation in biomass after a 30-year rotation of Pinus radiata on
3	podzolized sands: Impacts of intensive harvesting on soil resources, Forest
4	Ecol. Manag., 258, 2183–2193, doi: 10.1016/j.foreco.2009.02.010, 2009.
5	IPCC: Climate Change 2013: The Physical Science Basis. Contribution of
6	Working Group I to the Fifth Assessment Report of the Intergovernmental
7	Panel on Climate Change. Stocker, T.F., Qin, D., Plattner, G.K., Tignor,
8	M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley,
9	P.M. (eds), Cambridge University Press, Cambridge, United Kingdom and
10	New York, NY, USA, 1535, 2013.
11	Jowett, I.: Instream habitat and minimum flow requirements for the Waipa Stream
12	Ian Jowett Consulting Client report: IJ0703, Report prepared for Rotorua
13	District Council, Rotorua, New Zealand, 31, 2008.
14	Kirschbaum, M.U.F. and Watt, M.S.: Use of a process-based model to describe
15	spatial variation in Pinus radiata productivity in New Zealand, Forest Ecol
16	Manag., 262, 1008–1019, doi: 10.1016/j.foreco.2011.05.036, 2011.
17	Krause, P., Boyle, D.P., and Bäse, F.: Comparison of different efficiency criteria
18	for hydrological model assessment, Advances in Geosciences, 5, 89-97,
19	2005.
20	Kusabs, I. and Shaw, W.: An ecological overview of the Puarenga Stream with
21	particular emphasis on cultural values: prepared for Rotorua District
22	Council and Environment Bay of Plenty, Rotorua, New Zealand, 42, 2008.
23	Lane, L.J.: Chapter 19: Transmission Losses, In Soil Conservation Service,
24	National engineering handbook, section 4: hydrology, U.S. Government
25	Printing Office, Washington, D.C., 19-1–19-21, 1983.
26	Ledgard, S. and Thorrold, B.: Nitrogen Fertilizer Use on Waikato Dairy Farms,
27	AgResearch and Dexcel, New Zealand, 5, 1998.
28	Lim, K.J., Engel, B.A., Tang, Z., Choi, J., Kim, K., Muthukrishnan, S., and
29	Tripathy, D.: Automated Web GIS-based Hydrograph Analysis Tool,
30	WHAT, J. Am. Water Resour. As., 41, 1407-1416, doi: 10.1111/j.1752-
31	1688.2005.tb03808.x, 2005.
32	Lindenschmidt, K., Fleischbein, K., and Baborowski, M.: Structural uncertainty in
33	a river water quality modelling system, Ecol. Model., 204, 289-300,
34	doi: 10.1016/j.ecolmodel.2007.01.004, 2007.

1	Lowe, A., Gielen, G., Bainbridge, A., and Jones, K.: The Rotorua Land Treatment
2	Systems after 16 years, In New Zealand Land Treatment Collective-
3	Proceedings for 2007 Annual Conference, Rotorua, 14-16 March 2007,
4	66–73, 2007.
5	Mahon, W.A.J.: The Rotorua geothermal field: technical report of the Geothermal
6	Monitoring Programme, 1982-1985, Ministry of Energy, Oil and Gas
7	Division, Wellington, New Zealand, 1985.
8	Marino, S., Hogue, I.B., Ray, C.J., and Kirschner, D.E.: A methodology for
9	performing global uncertainty and sensitivity analysis in systems biology,
10	J. Theor. Biol., 254, 178–196, doi: 10.1016/j.jtbi.2008.04.011, 2008.
11	McKenzie, B.A., Kemp, P.D., Moot, D.J., Matthew, C., and Lucas, R.J.:
12	Environmental effects on plant growth and development, In New Zealand
13	Pasture and Crop Science, White, J.G.H. and Hodgson, J. (eds), Oxford
14	University Press: Auckland, New Zealand, 29–44, 1999.
15	Monteith, J.L.: Evaporation and the environment. In the state and movement of
16	water in living organisms, 19th Symposia of the Society for Experimental
17	Biology, Cambridge Univ. Press, London, U.K., 1965.
18	Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., and
19	Veith, T.L.: Model evaluation guidelines for systematic quantification of
20	accuracy in watershed simulations, T. ASAE, 50, 885–900, 2007.
21	Morris, M.D.: Factorial sampling plans for preliminary computational
22	experiments, Technometrics, 33, 161–174, 1991.
23	Neitsch, S.L., Arnold, J.G., Kiniry, J.R., and Williams, J.R.: Soil and Water
24	Assessment Tool Theoretical Documentation Version 2009, Texas Water
25	Resources Institute Technical Report No. 406, Texas A&M University
26	System, College Station, Texas, 647, 2011.
27	Neyman, J.: Outline of a Theory of Statistical Estimation Based on the Classical
28	Theory of Probability, Phil. Trans. R. Soc. A, 236, 333-380, doi:
29	10.1098/rsta.1937.0005, 1937.
30	Nielsen, A., Trolle, D., Me, W., Luo, L.C., Han, B.P., Liu, Z.W., Olesen, J.E., and
31	Jeppesen, E.: Assessing ways to combat eutrophication in a Chinese
32	drinking water reservoir using SWAT, Mar. Freshwater Res., 64, 475–492,
33	doi: 10.1071/MF12106, 2013.

1	Paku, L.K.: The use of carbon–13 to trace the migration of treated wastewater and
2	the chemical composition in a forest environment, Master Thesis, Science
3	in Chemistry, the University of Waikato, Hamilton, New Zealand, 92,
4	2001.
5	Parliamentary Commissioner for the Environment: Water Quality in New Zealand
6	Land Use and Nutrient Pollution, New Zealand, 82, 2013.
7	Pfannerstill, M., Guse, B., and Fohrer, N.: Smart low flow signature metrics for an
8	improved overall performace evaluation of hydrological models, J.
9	Hydrol., 510, 447–458, 2014.
10	Radcliffe, D.E., Lin, Z., Risse, L.M., Romeis, J.J., and Jackson, C.R.: Modeling
11	phosphorus in the Lake Allatoona watershed using SWAT: I. Developing
12	phosphorus parameter values, J. Environ. Qual., 38, 111-120,
13	doi:10.2134/jeq2007.0110, 2009.
14	Reusser, D.E., Blume, T., Schaefli, B., and Zehe, E.: Analysing the temporal
15	dynamics of model performance for hydrological models, Hydrol. Earth.
16	Syst. Sc., 13, 999–1018, doi:10.5194/hess-13-999-2009, 2009.
17	Reusser, D.E. and Zehe, E.: Inferring model structural deficits by analysing
18	temporal dynamics of model performance and parameter sensitivity, Water
19	Resour. Res., 47, W07550, 15pp, doi: 10.1029/2010WR009946, 2011.
20	Rice, J.A.: Mathematical statistics and data analysis, Boston, MA, Cengage
21	Learning, 2006.
22	Rimmer, A. and Hartmann, A.: Optimal hydrograph separation filter to evaluate
23	transport routines of hydrological models, J. Hydrol., 514, 249-257, doi:
24	10.1016/j.jhydrol.2014.04.033, 2014.
25	Rotorua District Council, Rotorua Wastewater Treatment Plant, Rotorua, New
26	Zealand, 22, 2006.
27	Rutherford, K., Palliser, C., Wadhwa, S.: Prediction of nitrogen loads to Lake
28	Rotorua using the ROTAN model, Report prepared for Bay of Plenty
29	Regional Council, New Zealand, 183. 2011.
30	Sangrey, D.A., Harrop-Williams, K.O., and Klaiber, J.A.: Predicting ground-
31	water response to precipitation, J. Geotech. Eng., 110, 957-975, doi:
32	10.1061/(ASCE)0733-9410(1984)110:7(957), 1984.

Schuol, J., Abbaspour, K.C., Yang, H., and Srinivasan, R.: Modeling blue and 1 2 green water availability in Africa, Water Resour. Res., 44, W07406, 18 pp, doi: 10.1029/2007WR006609, 2008. 3 4 Seymour, G.: Predictive Inference: An Introduction, Chapman & Hall, New York, 5 pp 280, 1993. Shen, Z.Y., Chen, L., and Chen, T.: Analysis of parameter uncertainty in 6 7 hydrological and sediment modeling using GLUE method: a case study of SWAT model applied to Three Gorges Reservoir Region, China, Hydrol. 8 9 Earth. Syst. Sc., 16, 121–132, doi: 10.5194/hess-16-121-2012, 2012. Sloan, P.G. and Moore, I.D.: Modelling subsurface stormflow on steeply sloping 10 forested watersheds, Water Resour. Res., 20, 1815–1822, 11 doi: 10.1029/WR020i012p01815, 1984. 12 13 Statistics New Zealand: Fertiliser use in New Zealand, Statistics New Zealand, New Zealand, 13, 2006. 14 15 van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., and Srinivasan, R.: A global sensitivity analysis tool for the parameters of 16 17 multi-variable catchment models, J. Hydrol., 324, 10-23, doi: 10.1016/j.jhydrol.2005.09.008, 2006. 18 Watt, M.S., Clinton, P.W., Coker, G., Davis, M.R., Simcock, R., Parfitt, R.L., and 19 Dando, J.: Modelling the influence of environment and stand 20 characteristics on basic density and modulus of elasticity for young Pinus 21 radiata and Cupressus lusitanica, Forest Ecol. Manag., 255, 1023-1033, 22 doi: 10.1016/j.foreco.2007.09.086, 2008. 23 White, K.L. and Chaubey, I.: Sensitivity analysis, calibration, and validations for 24 a multisite and multivariable SWAT model, J. Am. Water Resour. As., 41, 25 26 1077–1089, doi: 10.1111/j.1752-1688.2005.tb03786.x, 2005. White, M.J., Storm, D.E., Mittelstet, A., Busteed, P.R., Haggard, B.E., and Rossi, 27 28 C.: Development and testing of an in-stream phosphorus cycling model for the Soil and Water Assessment Tool, J. Environ. Qual., 43, 215-223, 29 30 doi: 10.2134/jeq2011.0348, 2014. White, P.A., Cameron, S.G., Kilgour, G., Mroczek, E., Bignall, G., Daughney, C., 31 and Reeves, R.R.: Review of groundwater in Lake Rotorua catchment, 32 Prepared for Environment Bay of Plenty, Institute of Geological & 33

- Nuclear Sciences Client Report 2004/130, Whakatane, New Zealand, 245,
 2004.
- Whitehead, D., Kelliher, F.M., Lane, P.M., and Pollock, D.S.: Seasonal
 partitioning of evaporation between trees and understorey in a widely
 spaced *Pinus radiata* stand, J. Appl. Ecol., 31, 528–542, 1994.
- Wu, H., Chen, B. 2015. Evaluating uncertainty estimates in distributed
 hydrological modeling for the Wenjing River watershed in China by
 GLUE, SUFI-2, and ParaSol methods. Ecological Engineering 76: 110–
 121.
- Ximenes, F.A., Gardner, W.D., and Kathuria, A.: Proportion of above–ground
 biomass in commercial logs and residues following the harvest of five
 commercial forest species in Australia, Forest Ecol. Manag., 256, 335–346,
 doi: 10.1016/j.foreco.2008.04.037, 2008.
- Yilmaz, K.K., Gupta, H.V., and Wagener, T.: A process–based diagnostic
 approach to model evaluation: Application to the NWS distributed
 hydrologic model, Water Resour. Res., 44, W09417, 18 pp, doi:
 10.1029/2007WR006716, 2008.
- Zhang, H., Huang, G.H., Wang, D.L., and Zhang, X.D.: Multi-period calibration
 of a semi-distributed hydrological model based on hydroclimatic
 clustering, Adv. Water Resour., 34, 1292–1303, 2011.
- Zhang, Z., Tao, F., Shi, P., Xu, W., Sun, Y., Fukushima, T., and Onda, Y.:
 Characterizing the flush of stream chemical runoff from forested
 watersheds, Hydrol. Process., 24, 2960–2970, doi: 10.1002/hyp.7717,
 2010.

1 Table 1. Description of data used to configure and calibrate the SWAT model.

Data	Application	Data description and configuration details	Source
Digital elevation model (DEM) & digitized stream network	Sub–basin delineation (Fig. 1b)	25 m resolution. Used to define five slope classes: $0-4\%$, $4-10\%$, $10-17\%$, $17-26\%$ and $>26\%$.	Bay of Plenty Regional Council (BoPRC)
Stream discharge and water quality measurements	Calibration (2004–2008) and validation ² (1994–1997; 2010–2012)	FRI: 15-min stream discharge data were aggregated as daily mean values (1994–1997; 2004–2008), monthly grab samples for determination of suspended sediment (SS), total phosphorus (TP) and total nitrogen (TN) concentrations (1994–1997; 2004–2008), high-frequency event-based samples for concentrations of SS (nine events), TP and TN (both 14 events) at 1–2 h frequency (2010–2012).	BoPRC; Abell et al., 2013
Spring discharge and nutrient loads	Point source (Fig. 1b)	Constant daily discharge and nutrient concentrations assigned to two cold–water springs (Waipa Spring and Hemo Spring) and one geothermal spring.	White et al., 2004; Proffit, 2009 (Unpublished Site Visit Report); Paku, 2001; Mahon, 1985; Glover, 1993; Rotorua District Council (pers. comm.)
Water abstraction volumes	Water use	Monthly water abstraction assigned to two cold–water springs.	Kusabs and Shaw, 2008; Jowett, 2008
Land use	HRU definition	25 m resolution, 10 basic land–cover categories. Some particular land–cover parameters were prior–estimated (Table 2).	New Zealand Land Cover Database Version 2; BoPRC
Soil characteristics	HRU definition	22 soil types. Properties were quantified based on measurements (if available) or estimated using regression	New Zealand Land Resource Inventory & digital soil map

^a-Model validation was undertaken using two different datasets. The monthly measurements (1994–1997) were predominantly collected when base flow was the dominant contributor to stream discharge. Data from high–frequency sampling during rain events (2010–2012) were also used to validate model performance during periods when quick flow was high.

	-	analysis to estimate properties for unmeasured functional horizons.	(available at http://smap.landcareresearch.co.n z)
Meteorological data	Meteorological forcing	Daily maximum and minimum temperature, daily mean relative humidity, daily global solar radiation, daily (9 am) surface wind speed and hourly precipitation.	Rotorua Airport Automatic Weather Station, National Climate Database (available at http://cliflo.niwa.co.nz/); Kaituna rain gauge (Fig. 1a)
Agricultural management		Stock density	Statistics New Zealand, 2006; Ledgard and Thorrold, 1998
	Agricultural management	Applications of urea and di-ammonium phosphate	Statistics New Zealand, 2006; Fert Research, 2009
practices	schedules	Applications of manure-associated nutrients	Dairying Research Corporation, 1999
Nutrient loading by wastewater application	Nonpoint– source from land treatment irrigation	Wastewater application rates and effluent composition (TN and TP concentration) for 16 spray blocks from 1996–2012. Each spray block was assigned an individual management schedule specifying daily application rates.	Rotorua District Council, 2006
Forest stand map and harvest dates	Forestry planting and harvesting operations	Planting and harvesting data for 472 ha forestry stands. Prior to 2007 we assumed stands were cleared one-year prior to the establishment year. Post 2007, harvesting date was assigned to the first day of harvesting month.	Timberlands Limited, Rotorua, New Zealand (pers. comm.)

- 1 Table 2. Prior–estimated parameter values for three dominant types of land–cover in the Puarenga Stream catchment. Values of other
- 2 land use parameters were based on the default values in the SWAT database.

Land-cover type	Parameter	Definition	Value	Source
	HVSTI	Percentage of biomass harvested	0.65	(Ximenes et al., 2008)
	T_OPT (°C)	Optimal temperature for plant growth	15	(Kirschbaum and Watt 2011)
	T_BASE (°C)	Minimum temperature for plant growth	4	(Kirschbaum and Watt 2011)
PINE	MAT_YRS	Number of years to reach full development	30	(Kirschbaum and Watt 2011)
(Pinus radiata)	BMX_TREES (tonnes ha ⁻¹)	Maximum biomass for a forest	400	(Bi et al., 2010)
	$GSI(m s^{-1})$	Maximum stomatal conductance	0.00198	(Whitehead et al., 1994)
	$BLAI (m^2 m^{-2})$	Maximum leaf area index	5.2	(Watt et al., 2008)
	BP3	Proportion of P in biomass at maturity	0.000163	(Hopmans and Elms 2009)
	BN3	Proportion of N in biomass at maturity	0.00139	(Hopmans and Elms 2009)
	HVSTI	Percentage of biomass harvested	0	-
FRSE	BMX_TREES (tonnes ha ⁻¹)	Maximum biomass for a forest	372	(Hall et al., 2001)
(Evergreen forest)	MAT_YRS (years)	Number of years for tree to reach full development	100	_
PAST		Optimal temperature for plant growth	25	(McKenzie et al., 1999)
(Pastoral farm)	T_BASE (°C)	Minimum temperature for plant growth	5	(McKenzie et al., 1999)

1 Table 3. Description of data used to calibrate the SWAT model. Data were measured at the Forest Research Institute (FRI) stream-

2 gauge and were considered representative of the downstream/outlet conditions of the Puarenga Stream.

Data	Application	Measurement data details	Source
Stream discharge measurements	<u>Calibration</u> (2004–2008) <u>Validation</u> (1994–1997)	<u>15-min stream discharge data were measured at FRI stream-</u> gauge (Fig. 1b) within the catchment and aggregated as daily mean values (1994–1997; 2004–2008).	BoPRC; Abell et al., 2013
<u>Stream water quality</u> <u>measurements</u>	<u>Calibration</u> (2004–2008) <u>Validation³</u> (1994–1997; 2010–2012)	Monthly grab samples for determination of suspended sediment (SS), total phosphorus (TP) and total nitrogen (TN) concentrations (1994–1997; 2004–2008), and high– frequency event–based samples for concentrations of SS (nine events), TP and TN (both 14 events) at 1–2 h frequency (2010–2012), were also measured at FRI stream–gauge (Fig. 1b) within the catchment.	BoPRC; Abell et al., 2013

³ Model validation was undertaken using two different datasets. The monthly measurements (1994–1997) were predominantly collected when base flow was the dominant contributor to stream discharge. Data from high–frequency sampling during rain events (2010–2012) were also used to validate model performance during periods when quick flow was high.

1 Table <u>34</u>. Summary of calibrated SWAT parameters. Discharge (Q), suspended sediment (SS) and total nitrogen (TN) parameter

2 values were assigned using auto-calibration, while total phosphorus (TP) parameters were manually calibrated. SWAT default ranges

3 and input file extensions are shown for each parameter.

Parameter	Definition	Unit	Default range	Calibrated value
Q				
EVRCH.bsn	Reach evaporation adjustment factor		0.5–1	0.9
SURLAG.bsn	Surface runoff lag coefficient		0.05–24	15
ALPHA_BF.gw	Base flow alpha factor (0–1)		0.0071– 0.0161	0.01
GW_DELAY.gw	Groundwater delay	d	0–500	500
GW_REVAP.gw	Groundwater "revap" coefficient		0.02-0.2	0.08
GW_SPYLD.gw	Special yield of the shallow aquifer	$m^{3} m^{-3}$	0-0.4	0.13
GWHT.gw	Initial groundwater height	m	0–25	14
GWQMN.gw	Threshold depth of water in the shallow aquifer required for return flow to occur	mm	0–5000	372
RCHRG_DP.gw	Deep aquifer percolation fraction		0–1	0.87
REVAPMN.gw	Threshold depth of water in the shallow aquifer required for "revap" to occur	mm	0–500	260
CANMX.hru	Maximum canopy storage	mm	0–100	0.6
EPCO.hru	Plant uptake compensation factor		0–1	0.34
ESCO.hru	Soil evaporation compensation factor		0–1	0.9
HRU_SLP.hru	Average slope steepness	$m m^{-1}$	0–0.6	0.5
LAT_TTIME.hru	Lateral flow travel time	d	0–180	3
RSDIN.hru	Initial residue cover	kg ha ⁻¹	0-10000	1
SLSOIL.hru	Slope length for lateral subsurface flow	m	0–150	40
CH_K2.rte	Effective hydraulic conductivity in the main channel alluvium	mm h ⁻¹	0–500	20
CH_N2.rte	Manning's N value for the main channel		0–0.3	0.16

CH_K1.sub	Effective hydraulic conductivity in the tributary channel alluvium	mm h ⁻¹	0–300	100
CH_N1.sub	Manning's N value for the tributary channel		0.01–30	20
SS				
USLE_P.mgt	USLE equation support practice factor		0–1	0.5
PRF.bsn	Peak rate adjustment factor for sediment routing in the main channel		0–2	1.9
SPCON.bsn	Linear parameter for calculating the maximum amount of sediment that can be re-entrained during channel sediment routing		0.0001-0.01	0.001
SPEXP.bsn	Exponent parameter for calculating sediment re–entrained in channel sediment routing		1–1.5	1.26
LAT_SED.hru	Sediment concentration in lateral flow and groundwater flow	mg L ⁻¹	0–5000	5.7
OV_N.hru	Manning's N value for overland flow		0.01–30	28
SLSUBBSN.hru	Average slope length	m	10-150	92
CH_COV1.rte	Channel erodibility factor		0–0.6	0.17
CH_COV2.rte	Channel cover factor		0–1	0.6
ТР				
P_UPDIS.bsn	Phosphorus uptake distribution parameter		0–100	0.5
PHOSKD.bsn	Phosphorus soil partitioning coefficient		100-200	174
PPERCO.bsn	Phosphorus percolation coefficient		10-17.5	14
PSP.bsn	Phosphorus sorption coefficient		0.01–0.7	0.5
GWSOLP.gw	Soluble phosphorus concentration in groundwater loading	mg P L ⁻¹	0-1000	0.063
LAT_ORGP.gw	Organic phosphorus in the base flow	mg P L ⁻¹	0–200	0.01
ERORGP.hru	Organic phosphorus enrichment ratio		0–5	2.5
CH_OPCO.rte	Organic phosphorus concentration in the channel	mg P L ⁻¹	0–100	0.02
BC4.swq	Rate constant for mineralization of organic phosphorus to dissolved phosphorus in the reach at 20 °C	d ⁻¹	0.01–0.7	0.3
RS2.swq	Benthic (sediment) source rate for dissolved phosphorus in the reach at 20 °C	mg m ⁻² d ⁻¹	0.001–0.1	0.02
RS5.swq	Organic phosphorus settling rate in the reach at 20 °C	d ⁻¹	0.001-0.1	0.05

TN				
RSDCO.bsn	Residue decomposition coefficient	0.02-0.1	0.09	
CDN.bsn	Denitrification exponential rate coefficient	0–3	0.3	
CMN.bsn	Rate factor for humus mineralization of active organic nitrogen	0.001-0.003	0.002	
N_UPDIS.bsn	Nitrogen uptake distribution parameter	0-100	0.5	
NPERCO.bsn	Nitrogen percolation coefficient	0–1	0.0003	
RCN.bsn	Concentration of nitrogen in rainfall mg N L ⁻¹		0–15	0.34
SDNCO.bsn	Denitrification threshold water content		0–1	0.02
HLIFE_NGW.gw	Half–life of nitrate–nitrogen in the shallow aquifer d		0–200	195
LAT_ORGN.gw	Organic nitrogen in the base flow n		0–200	0.055
SHALLST_N.gw	Nitrate–nitrogen concentration in the shallow aquifer mg N L ⁻		0-1000	1
ERORGN.hru	Organic nitrogen enrichment ratio		0–5	3
CH_ONCO.rte	Organic nitrogen concentration in the channel n		0–100	0.01
BC1.swq	Rate constant for biological oxidation of ammonium–nitrogen to nitrite–nitrogen in the reach at 20 °C		0.1–1	1
BC2.swq	Rate constant for biological oxidation of nitrite–nitrogen to nitrate– nitrogen in the reach at 20 °C	d ⁻¹	0.2–2	0.7
BC3.swq	Rate constant for hydrolysis of organic nitrogen to ammonium- nitrogen in the reach at 20 °C	d ⁻¹	0.2–0.4	0.4
RS3.swq	Benthic (sediment) source rate for ammonium–nitrogen in the reach $mg d^{-1}$		0–1	0.2
RS4.swq	Rate coefficient for organic nitrogen settling in the reach at 20 °C	d ⁻¹	0.001-0.1	0.05

1 Table 4<u>5</u>. Criteria for model performance. Note: o_n is the n^{th} observed datum, s_n is the n^{th} simulated datum, \overline{o} is the observed mean

2 value, \bar{s} is the simulated daily mean value, and N is the total number of observed data. Performance rating criteria are based on

3 Moriasi et al. (2007) for Q: discharge, SS: suspended sediment, TP: total phosphorus and TN: total nitrogen. Moriasi et al. (2007)

4 derived these criteria based on extensive literature review and analysing the reported performance ratings for recommended model

5 evaluation statistics.

Statistic equation	Constituent -	Performance ratings			
Statistic equation		Unsatisfactory	Satisfactory	Good	Very good
$R^{2} = \frac{\{\sum_{n=1}^{N} [(s_{n} - \bar{s})(o_{n} - \bar{o})]\}^{2}}{\sum_{n=1}^{N} (o_{n} - \bar{o})^{2} \times \sum_{n=1}^{N} (s_{n} - \bar{s})^{2}} $ (3)	All	< 0.5	0.5 - 0.6	0.6 - 0.7	0.7 – 1
NSE = $1 - \frac{\sum_{n=1}^{N} (o_n - s_n)^i}{\sum_{n=1}^{N} (o_n - \overline{o})^i}$ $i = 2$ (4)	All	< 0.5	0.5 - 0.65	0.65 - 0.75	0.75 – 1
	Q	> 25	15 - 25	10 - 15	< 10
$\pm PBIAS\% = \frac{\sum_{n=1}^{N} (o_n - s_n)}{\sum_{n=1}^{N} o_n} \times 100 (5)$	SS	> 55	30 - 55	15 - 30	< 15
$\Sigma_{n=1} \sigma_n$	TP, TN	> 70	40 - 70	25 - 40	< 25

6 R²: coefficient of determination

7 NSE: Nash–Sutcliffe efficiency

8 PBIAS: percent bias

Table <u>56</u>. Model performance ratings for simulations of discharge (Q), concentrations of suspended sediment (SS), total phosphorus
 (TP) and total nitrogen (TN). n indicates the number of measurements. Q-weighted mean concentrations were calculated using Eq.
 (1).

Model performance	Statistics	Q	SS	TP	TN
		n = 1439	n = 43	n = 45	n = 39
-	n ²	0.77	0.42	0.02	0.08
Calibration with	K-	(Very good)	(Unsatisfactory)	(Unsatisfactory)	(Unsatisfactory)
instantaneous measurements	NCE	0.73	-0.08	-1.31	-0.30
(2004–2008)	INSE	(Good)	(Unsatisfactory)	(Unsatisfactory)	(Unsatisfactory)
		7.8	-18.3	23.8	-0.05
	\pm PDIA5%	(Very good)	(Very good)	(Very good)	(Very good)
		n = 1294	n = 37	n = 37	n = 36
	ם 2	0.68	0.80	0.01	0.01
Validation with	Γ	(Good)	(Very good)	(Unsatisfactory)	(Unsatisfactory)
instantaneous measurements	NSE	0.62	0.76	-0.97	-2.67
(1994–1997)		(Satisfactory)	(Very good)	(Unsatisfactory)	(Unsatisfactory)
		8.8	-0.32	24.5	-26.7
	<u> </u>	(Very good)	(Very good)	(Very good)	(Good)
		_	n = 12	n = 18	n = 18
	D ²	_	0.38	0.06	0.46
Validation with	K-		(Unsatisfactory)	(Unsatisfactory)	(Unsatisfactory)
Q-weighted mean concentrations	NCE	ISE –	-0.03	-4.88	0.42
(2010–2012)	INSE		(Unsatisfactory)	(Unsatisfactory)	(Unsatisfactory)
	±PBIAS%	_	43.9	69.4	-0.87
			(Satisfactory)	(Satisfactory)	(Very good)

Table 67. Model performance statistics for simulations of discharge (Q), and loads of suspended sediment (SS), total phosphorus (TP) and total nitrogen (TN). Statistics were calculated for both overall and separated simulations. Q_{all} and L_{all} indicate the overall simulations; Q_b and L_b indicate the base flow simulations; Q_q and L_q indicate the quick flow simulations.

> SS TP TN Q Model performance Statistics Q_{q} $Q_{
> m all}$ L_{all} L_{b} L_q L_q Q_{b} $L_{\rm b}$ L_q Lall $L_{\rm b}$ L_{all} \mathbb{R}^2 0.84 0.77 0.97 0.95 0.84 0.66 0.68 0.61 0.24 0.65 0.39 0.72 0.71 0.73 0.33 0.33 0.27 -6.2 0.89 0.85 Calibration (2004-2008) NSE 0.6 0.09 -0.17 0.5 \pm PBIAS% 7.5 8.7 7.8 7.57 -23.4 -3.6 45.4 40.1 43.6 0.8 6.6 2.7 R² 0.87 0.81 0.68 0.36 0.98 0.95 0.27 0.27 0.33 0.58 0.06 0.79 0.33 Validation (1994–1997) NSE 0.56 0.62 0.62 -0.03 0.43 0.85 -1.9 0.04 -0.64 0.58 -0.07 ±PBIAS% 11.3 -1.2 8.8 34.5 -79.7 45.8 -7.6 14.3 -2.5 11.1 -9.3 37

4 R²: coefficient of determination; NSE: Nash–Sutcliffe efficiency; PBIAS: percent bias

Table 7-8 Rankings of relative sensitivities of parameters (from most to least) for variables (header row) of Q (discharge), SS (suspended sediment), MINP (mineral phosphorus), ORGN (organic nitrogen), NH₄-N (ammonium-nitrogen), and NO₃-N (nitrate-nitrogen). Relative sensitivities were identified by randomly generating combinations of values for model parameters and comparing modelled and measured data with a Student's t test ($p \le 0.05$). Bold text denotes that a parameter was deemed sensitive relative to more than one simulated variable. Shaded text denotes that parameter deemed insensitive to any of the two flow components (base and quick flow; see Figure 7) using one-at a-time sensitivity analysis. Definitions and units for each parameter are shown in Table 34.

Q	SS	MINP	ORGN	NH ₄ –N	NO ₃ –N
SLSOIL	LAT_SED	CH_OPCO	CH_ONCO	CH_ONCO	NPERCO
CH_K2	CH_N2	BC4	BC3	BC1	CDN
HRU_SLP	SLSUBBSN	RS5	SOL_CBN(1)	CDN	ERORGN
LAT_TTIME	SPCON	ERORGP	RS4	RS3	CMN
SOL_AWC(1)	ESCO	PPERCO	RCN	RCN	RCN
RCHRG_DP	OV_N	RS2	N_UPDIS		RSDCO
GWQMN	SLSOIL	PHOSKD	USLE_P		
GW_REVAP	LAT_TTIME	GWSOLP	SDNCO		
GW_DELAY	SOL_AWC(1)	LAT_ORGP	SOL_NO3(1)		
CH_COV1	EPCO		CMN		
CH_COV2	CANMX		HLIFE_NGW		
EPCO	CH_K2		RSDCO		
SPEXP	GW_DELAY		USLE_K(1)		
CANMX	ALPHA_BF				
CH_N1	GW_REVAP				
PRF	CH_COV1				
SURLAG					



1

2 Figure 1. (a) Location of Puarenga Stream surface catchment in New Zealand, 3 Kaituna rain gauge, climate station and managed land areas for which 4 management schedules were prescribed in SWAT, and (b) location of the Puarenga Stream, major tributaries, monitoring stream-gauges, two cold-water 5 springs and the Whakarewarewa geothermal contribution. Measurement data 6 7 (Table 3) used to calibrate the SWAT model were from the Forest Research Institute (FRI) stream-gauge and were considered representative of the 8 downstream/outlet conditions of the Puarenga Stream. 9



Figure 2. Flow chart of methods used to separate hydrograph and contaminant
loads and to quantify parameter sensitivities for: Q (discharge), SS (suspended
sediment), MINP (mineral phosphorus), ORGN (organic nitrogen), NH₄–N
(ammonium–nitrogen), and NO₃–N (nitrate–nitrogen). *NSE*: Nash–Sutcliffe
efficiency.



Figure 3. Measurements and daily mean simulated values of discharge, suspended
sediment (SS), total phosphorus (TP) and total nitrogen (TN) during calibration
(a–d) and validation (e–h). Measured daily mean discharge was calculated from
15–min observations and measured concentrations of SS, TP and TN correspond
to monthly grab samples.



Figure 4. Example of a storm event showing derivation of discharge (Q)–weighted daily mean concentrations (dashed horizontal line) based on hourly measured concentrations (black dots) of suspended sediment (SS), total phosphorus (TP) and total nitrogen (TN) over two days (a–c). Comparisons of Q–weighted daily mean concentrations with simulated daily mean estimates of SS, TP and TN (scatter plot, d–f). The horizontal bars show the ranges in hourly measurements during each storm event in 2010–2012.



Figure 5. Measurements and simulations derived using the calibrated set of parameter values. Data are shown separately for base flow and quick flow. (a) Daily mean base flow and quick flow; (b) suspended sediment (SS) load; (c) total phosphorus (TP) load; (d) total nitrogen (TN) load. Vertical lines in b–d show the contaminant load in quick flow. Time series relate to calibration (2004–2008) and validation (1994–1997) periods (note time discontinuity). Measured instantaneous loads of SS, TP, and TN correspond to monthly grab samples.



of (b) suspended sediment (SS), (c) total phosphorus (TP), and (d) total nitrogen (TN) including lower and upper 95% confidence limits (LCL and UCL) and lower and upper 95% prediction limits (LPL and UPL). Note that the "choppy" shape of confidence limits shown in figures b–d were-resulted from the few data points (< 50) in the regressions of measured and simulated SS, TP and TN concentrations.



1

Figure 7. The standard deviation (STD) of the ln-transformed Nash-Sutcliffe 2 efficiency (NSE) used to indicate parameter sensitivity based on one-at a-time 3 4 (OAT) sensitivity analysis for separate base and quick flow components: (a) Q (discharge); (b) SS (suspended sediment); (c) MINP (mineral phosphorus); (d) 5 6 NO₃-N (nitrate-nitrogen); (e) ORGN (organic nitrogen); (f) NH₄-N (ammonium-7 nitrogen). A median value (0.2) derived from the STD of In-transformed NSE was chosen as a threshold above which parameters were deemed to be 'sensitive'. 8 9 Definitions of each parameter are shown in Table 34.