Effects of hydrologic conditions on SWAT model performance and parameter sensitivity for a small, mixed land use catchment in New Zealand

W. Me[1,2], J. M. Abell[1,*], D. P. Hamilton[1]

[1]{Environmental Research Institute, University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand}

[2]{College of Hydrology and Water Resources, Hohai University, Nanjing, 210098, People's Republic of China}

[*]{now at: Ecofish Research Ltd., Suite 1220 – 1175 Douglas Street, Victoria, British Columbia, Canada}

Correspondence to: W. Me (yaowang0418@gmail.com)

Abstract

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

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

1 Introduction

Catchment models are valuable tools for understanding natural processes occurring at basin scales and for simulating the effects of different management regimes on soil and water resources (e.g. Cao et al., 2006). Model applications may have uncertainties as a result of errors associated with the forcing variables, measurements used for calibration, and conceptualisation of the model itself (Lindenschmidt et al., 2007). The ability of catchment models to simulate hydrological processes and pollutant loads can be assessed through analysis of uncertainty or errors during a calibration process that is specific to the application 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 data (e.g. soil maps) or knowledge gained in other studies. However, values for other parameters need to be assigned during a calibration process as a result of complex spatial and temporal variations that are not readily captured either through measurements or within the model algorithms themselves (Boyle et al., 2000). Such parameter values assigned during calibration are therefore lumped, i.e., they integrate variations in space and/or time and thus provide an approximation for real values which often vary widely within a study catchment. Model calibration is an iterative process whereby parameters are adjusted to the system of interest by refining model predictions to fit closely with observations 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 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 optimisation algorithms (Eckhardt and Arnold, 2001). The Sequential Uncertainty Fitting (SUFI–2) procedure has previously been applied to auto–calibrate discharge parameters in a SWAT application for the Thur River, Switzerland (Abbaspour et al., 2007), as well as for groundwater recharge, evapotranspiration and soil storage water considerations in West Africa (Schuol et al., 2008). Model validation is subsequently performed using measured data that are independent of those used for calibration (Moriasi et al., 2007).

Values for hydrological parameter values in the SWAT model can vary temporally. Cibin et al. (2010) found that the optimum calibrated values for hydrological parameters varied with different flow regimes (low, medium and high), thus suggesting that SWAT model performance can be optimised by assigning parameter values based on hydrological characteristics. Other work has similarly demonstrated benefits from assigning separate parameter values to low, 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 temporal dependence of model parameterisation on hydrologic conditions has implications for model performance. Krause et al. (2005) compared different statistical metrics of hydrological model performance separately for base–flow periods and storm events to evaluate the performance. The authors found that the logarithmic form of the Nash–Sutcliffe efficiency (NSE) value provided more information on the sensitivity of model performance for discharge simulations during storm events, while the relative form of NSE was better for base flow periods. Similarly, Guse et al. (2014) investigated temporal dynamics of sensitivity of hydrological parameters and SWAT model performance using Fourier amplitude sensitivity test (Reusser et al., 2011) and cluster analysis (Reusser et al., 2009). The authors found that three groundwater parameters were highly sensitive during quick flow, while one evaporation parameter was most sensitive during base flow, and model performance was also found to vary significantly for the two flow regimes. Zhang et al. (2011) calibrated SWAT hydrological parameters for periods separated on the basis of six climatic indexes. Model performance improved when different values were assigned to parameters based on six hydroclimatic periods. Similarly, Pfannerstill et al. (2014) found that assessment of model performance was improved by considering an additional performance statistic for very low–flow simulations amongst five hydrologically–separated regimes.

To date, analysis of temporal dynamics of SWAT parameters has predominantly focussed on simulations of discharge rather than water quality constituents. This partly reflects the paucity of comprehensive water quality data for many catchments; near–continuous discharge data can readily be collected but this is not the case for water quality parameters such as suspended sediment or nutrient concentrations. Data collected in monitoring programmes that involve sampling at regular time intervals (e.g. monthly) are often used to calibrate water 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 collected during storm–flow periods are rarely available for SWAT calibration, thus prohibiting opportunities to investigate how parameter sensitivity varies under conditions which can contribute disproportionately to nutrient or sediment 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 overestimation of model performance. Thus, further research is required to examine how water quality parameters vary during different flow regimes and to understand how model uncertainty may vary under future climatic conditions that affect discharge regimes (Brigode et al., 2013).

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 quality sampling programme designed to target a wide range of hydrologic conditions. A catchment–wide set of parameters was calibrated using the SUFI–2 procedure which is integrated into the SWAT Calibration and Uncertainty Program (SWAT–CUP). The objectives of this study were to: (1) quantify the performance of the model in simulating discharge and fluxes of suspended sediments and nutrients at the catchment outlet; (2) rigorously evaluate model performance by comparing daily simulation output with monitoring data collected under a range of hydrologic conditions; and (3) quantify whether parameter sensitivity varies between base flow and quick flow conditions.

2 Methods

2.1 Study area

The Puarenga Stream is the second–largest surface inflow to Lake Rotorua (Bay of Plenty, New Zealand) and drains a catchment of 77 km^2. The catchment is situated in the central North Island of New Zealand, which has a warm temperate climate. Annual mean temperature at Rotorua Airport (Fig. 1a) is 15±4 °C and annual mean evapotranspiration is 714 mm yr^{-1} (1993–2012; National Climatic Data Centre; available at http://cliflo.niwa.co.nz/). Annual mean precipitation at Kaituna rain gauge (Fig. 1a) is 1500 mm yr^{-1} (1993–2012; Bay of Plenty Regional Council). The catchment is relatively steep (mean slope = 9%; Bay of Plenty Regional Council) with predominantly pumice soils that have high macroporosity, resulting in high infiltration rates and substantial sub–surface lateral flow contributions to stream channels. Two cold–water springs (Waipa Spring and Hemo Spring) and one geothermal spring (Fig. 1b) are located in the catchment area. Two cold–water springs have annual mean discharge of ~0.19 m^3 s^{-1} (Rotorua District Council) and one geothermal spring has annual mean discharge of ~0.12 m^3 s^{-1} (White et al., 2004).

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

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

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 SWAT2009_rev488) was run on an hourly 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) method was used to calculate infiltration, rather than the SCS curve number method. Therefore, the hourly rainfall/Green & Ampt infiltration/hourly routing method (Neitsch et al., 2011) was chosen to simulate upland and in–stream processes. Ten sub–catchments were represented in the Puarenga Stream catchment, each comprising numerous Hydrologic Response Units (HRUs). Each 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 predicted separately within SWAT for each HRU, with predictions summed to 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.

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

SWAT simulates loads of nitrate–nitrogen (NO_3–N), ammonium–nitrogen (NH_4–N) and organic nitrogen (ORGN), the sum of which is total nitrogen (TN). Nitrogen parameters were auto–calibrated for each N fraction. The SWAT model does not account for the initial nitrate concentration in shallow aquifers, as also noted by Conan et al. (2003). Ekanayake and Davie (2005) indicated that SWAT underestimated N loading from groundwater and suggested a modification by adding a background concentration of nitrate in streamflow to represent groundwater nitrate contributions. Over the period of the first five years of wastewater irrigation, nitrate concentrations in shallow groundwater draining the Waipa Stream sub–catchment were estimated to have increased by c. 0.44 mg L^{-1} (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 model simulations of TN concentration assuming that groundwater concentrations had equilibrated with the applied wastewater nitrogen.

2.3 Model calibration and validation

Daily mean discharge was firstly calibrated based on daily mean values of 15–minute measurements (Table 3). Water quality variables were then calibrated in the sequence: SS, TP and TN. Modelled mean daily concentrations were compared with concentrations measured during monthly grab sampling, with monthly measurements assumed equal to daily mean concentrations (Table 3). 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 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, the operational regime for the wastewater irrigation has varied since operations began in 1991, with a marked change occurring in 2002 when operations switched from applying the wastewater load to two blocks (rotated daily for a total of 14 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 the most recent (post 2002) period (2004–2008) to calibration to ensure that the model was configured to reflect current operations.

Parameter values that were not derived from measurements or the literature were assigned based on either automated or manual calibration (Table 4). Manual calibration was undertaken for 11 parameters related to TP, while a Sequential Uncertainty Fitting (SUFI–2) procedure was applied to auto–calibrate 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 SWAT Calibration and Uncertainty Program (SWAT–CUP). SUFI–2 is a procedure that efficiently quantifies and constrains parameter uncertainties/ranges from default ranges with the fewest number of iterations (Abbaspour et al., 2004), and has been shown to provide optimal results relative to the use of alternative algorithms (Wu and Chen, 2015). SUFI–2 involves Latin hypercube sampling (LHS), which is a method that generates a sample of plausible parameter values from a multidimensional distribution and ensures that samples cover the entire 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 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 in 1000 combinations of parameters and 1000 simulations. Model performance was quantified for each simulation based on the Nash–Sutcliffe efficiency (NSE). 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 each simulation. Each compartment was not given weight to formulate the objective function because only one variable was specifically focused on at each time. A parameter sensitivity matrix was then computed based on the changes in the objective function after 1000 simulations. Parameter sensitivity was quantified based on the p value from a Student's t–test, which was used to compare the mean of simulated values with the mean value of measurements (Rice, 2006). A parameter was deemed sensitive if $p \leqslant 0.05$ after 1000 simulations (one iteration). Numerous iterations of LHS were conducted. Values of p from numerous iterations were averaged for each parameter, and the frequency of iterations where a parameter was deemed sensitive was summed. Rankings of relative sensitivities of parameters were developed based on how frequently the sensitive parameter was identified and the averaged value of p calculated from several iterations. The most sensitive parameter was determined based on the frequency that the parameter was deemed sensitive, and the smallest average p–value from all iterations.

SUFI–2 considers two criteria to constrain uncertainty in each iteration. One is the P–factor, the percentage of measured data bracketed by 95% prediction uncertainty (95PPU). Another is the R–factor, the average thickness of the 95PPU band divided by the standard deviation of measured data. A range was first defined for each parameter based on a synthesis of ranges from similar studies or 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% confidence interval and the standard deviation of a parameter value were derived from the diagonal elements of the covariance matrix, which was calculated from 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, SS, MINP, ORGN, NH_4–N and NO_3–N) reflected the numbers required to ensure that > 90% of measured data were bracketed by simulated output and the R–factor 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). 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 restricted based on the optimum values calibrated in similar studies. Parameter values for TP simulations were manually–calibrated based on the relative percent deviation from the predetermined values of those auto–calibrated parameters for MINP simulations, given by the objective functions (e.g., NSE). Parameters related to the physical characteristics of the catchment were not changed because their values were considered to be representative of the catchment characteristics.

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. Samples were 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 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 concentrations to compare with modelled daily mean estimates. We did not use the high–frequency observations to calibrate the model, because of the limited number of high–frequency (1–2 h) samples (nine events for SS and 14 events for 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) high flow periods. The Q–weighted mean concentrations C_{QWM} were calculated as:

$$C_{\mathrm{QWM}} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} C_{\mathrm{i}} Q_{\mathrm{i}}}{\sum_{\mathrm{i}=1}^{\mathrm{n}} Q_{\mathrm{i}}} \tag{1}$$

where n is number of samples, C_{i} is contaminant concentration measured at time i, and Q_{i} is discharge measured at time i.

2.4 Hydrograph and contaminant load separation

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 flow (Q_q). An Eckhardt filter parameter of 0.98 and ratio of base flow to total discharge of 0.8 were assumed (cf. Lim et al., 2005). There was a total of 60 days without quick flow during the calibration period (2004–2008) and 1379 days for which hydrograph separation defined both base flow and quick flow.

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

$$C_\mathrm{sep} = \frac{Q_\mathrm{q} \times C_\mathrm{q}^{'} + Q_\mathrm{b} \times C_\mathrm{b}^{'}}{Q_\mathrm{q} + Q_\mathrm{b}} \qquad (2)$$

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

To ensure that $C_\mathrm{q}^{'}$ is positive, $C_\mathrm{b}^{'}$ is constrained to be the minimum of $\overline{C_\mathrm{sep}}$ and C_sep. Measured and simulated base flow and quick flow contaminant loads were then calculated.

A one–at a–time (OAT) routine proposed by Morris (1991) was applied to investigate how parameter sensitivity varied between the two flow regimes (base flow and quick flow), based on the ranking of relative sensitivities of parameters that were identified by randomly generating combinations of values for model parameters for each individual variable using the SUFI–2 procedure. OAT sensitivity analysis was then employed by varying the parameter of interest 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–transformed NSE was used to indicate parameter sensitivity for the two flow regimes.

Parameters were ranked from most to least sensitive on the basis of the sensitivity metric (STD of ln–transformed NSE), using a value of 0.2 as a threshold above which parameters were deemed particularly ‘sensitive’. The 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.

2.5 Model evaluation

Model goodness–of–fit was assessed graphically and quantified using coefficient of determination (R^2), Nash–Sutcliffe efficiency (NSE) and percent bias (PBIAS; Table 5). R^2 (range 0 to 1) and NSE (range $-\infty$ to 1) values are commonly used to evaluate SWAT model performance (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, NH_4–N and NO_3–N in the SUFI–2 procedure. The R software (R Development Core Team) 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).

3 Results

3.1 Model performance and uncertainty

Numerous rounds (each comprising 1000 iterations) of LHS were conducted for each simulated variable until the performance criteria were satisfied. The total number of rounds of LHS for each simulated variable was as follows (number in parentheses): Q (7), SS (7), MINP (11), ORGN (10), NH_4–N (4) and NO_3–N (4). The parameters that provided the best statistical outcomes (i.e, best match to observed data) are given in Table 4. Two criteria (R–factor and P–factor) were used to show model uncertainties for simulations of discharge and contaminant 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), NH_4–N (1.16, 0.56) and NO_3–N (1.2, 0.29). Model uncertainties for simulations of Q and SS, TP and TN concentrations are shown in Fig. 6.

Modelled and measured base flow showed high correspondence, although measured daily mean discharge during storm peaks was often underestimated (Fig. 3a and 3e). Annual mean percentages of lateral flow recharge, shallow aquifer recharge and deep aquifer recharge to total water yield were predicted by SWAT as 30%, 10%, 58%, respectively. Modelled SS concentrations overestimated measurements of monthly grab samples by an average of 18.3% during calibration 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) and 24.5% during validation (Fig. 3g). Similarly, measured TP loads were underestimated by 34.5% and 38.4%, during calibration and validation, respectively. Modelled and measured TN concentrations were generally better aligned during base flow (Fig. 3d), apart from a mismatch prior to 1996 when monthly measured TN concentrations were substantially lower than model predictions, although the concentrations gradually increased (Fig. 3h) during the 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 modelled TN load was 167 kg N d^{-1} to 205 kg N d^{-1}, respectively.

Statistical evaluations of goodness–of–fit are shown in Table 6. The R^2 values for discharge were 0.77 for calibration and 0.68 for validation, corresponding to model performance ratings (cf. Moriasi et al., 2007) of ‘very good’ and ‘good’ (Table 5). Similarly, the NSE values for discharge were 0.73 (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 of daily mean discharge, however, the low magnitude of PBIAS values corresponded to a performance rating of ‘very good’. The R^2 values for SS were 0.42 (unsatisfactory) for calibration and 0.80 for validation (very good). Similarly, the NSE values for SS were -0.08 (unsatisfactory) for calibration and 0.76 (very good) for validation. The model did not simulate trends well for monthly measured TP and TN concentrations. The R^2 values for TP and TN were both $<$ 0.1 (unsatisfactory) during calibration and validation and NSE values were both $<$ 0 (unsatisfactory). Values of PBIAS corresponded to ‘good’ or ‘very good’ performance ratings for TP and TN.

Observed Q–weighted daily mean concentrations derived from hourly 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 and TN concentrations were somewhat better than for TP. Comparisons of Q–weighted daily mean concentrations (C_{QWM}) during storm events from 2010 to 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 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 6), they were both close to the threshold for satisfactory performance (0.5). For C_{QWM} of SS and TP, R^2 and NSE values indicated that the model performance was unsatisfactory. The PBIAS value of -0.87 for C_{QWM} of TN corresponded to model performance ratings of 'very good', while the PBIAS values for C_{QWM} of SS and TP were 43.9 and 69.4, respectively, indicating satisfactory model performance.

Measured and simulated discharge and contaminant loads separated for the two flow regimes (base flow and quick flow) are shown in Fig. 5. Model performance statistics differed between the two flow regimes (Table 7). Simulations of discharge and constituent loads under quick flow were more closely related to the measurements (i.e., higher values of R^2 and NSE) than simulations under base flow. Base flow TN load simulations during the validation period showed better model performance than simulations under quick flow. Additionally, measurements under quick flow were better reproduced by the model than the measurements for the whole simulation period. Simulations of contaminant loads matched measurements much better than for contaminant concentrations, as indicated by statistical values for model performance given in Table 6 and 7.

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 8), 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) and lateral flow travel time (LAT_TIME) have an important controlling effect on the amount of lateral flow entering the stream reach during quick flow. Both slope (HRU_SLP) and soil available water content (SOL_AWC) were particularly sensitive for the base flow simulation because they affect lateral flow within the kinematic storage model in SWAT (Sloan and Moore, 1984). The aquifer percolation coefficient (RCHRG_DP) and the base flow alpha factor (ALPHA_BF) strongly influenced base flow calculations (Sangrey et al., 1984), as did the channel Manning's N value (CH_N2) which is used to estimate channel flow (Chow, 2008).

For SS loads, 12 and four parameters, respectively, were identified as sensitive in relation to the simulations of base flow and quick flow (parameters 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. LAT_TIME and SLSOIL) were found to be much more sensitive for base flow SS load estimations. Exclusive parameters for SS estimations, such as SPCON (linear parameter), PRF (peak rate adjustment factor), SPEXP (exponent parameter), CH_COV1 (channel erodibility factor), and CH_COV2 (channel cover factor) were found to be much more sensitive in base flow SS load, while LAT_SED (SS concentration in lateral flow and groundwater flow) was more sensitive in quick flow SS load. Parameters that control overland processes, e.g. CN2 (the curve number), OV_N (overland flow Manning's N value) and SLSUBBSN (sub–basin slope length), were found to be much more sensitive for quick flow SS load estimations.

Of the sensitive parameters, BC4 (ORGP mineralization rate) was particularly sensitive for the simulation of base flow MINP load (Fig. 7c). RCN (nitrogen concentration in rainfall) related specifically to the dynamics of the base flow NO_3–N load and NPERCO (nitrogen percolation coefficient) significantly affected quick flow NO_3–N load (Fig. 7d). Parameter CH_ONCO (channel ORGN concentration) similarly affected both flow components of ORGN load (Fig. 7e) and SOL_CBN (organic carbon content) was most sensitive for the simulations of quick flow ORGN and NH_4–N loads. Parameter BC1 (nitrification rate in reach) was particularly sensitive for the simulation of base flow NH_4–N load (Fig. 7f).

4 Discussion

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

4.1 Temporal dynamics of model performance

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

Overestimation of TN concentrations prior to 1996 reflects higher NO_3–N concentrations in groundwater during the calibration period (2004–2008) due to the wastewater irrigation operation. Nitrate concentrations appeared to reach a 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 instantaneous measurements may partly reflect a mismatch between the dominant processes affecting phosphorus cycling in the stream and those represented in SWAT. The ORGP fraction that is simulated in SWAT includes both organic and inorganic forms of particulate phosphorus, however, the representation of particulate phosphorus cycling only focusses on organic phosphorus cycling, with limited consideration of interactions between inorganic streambed sediments and 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 that dynamic sorption processes between the dissolved and particulate inorganic phosphorus pools exert major control on phosphorus cycling (Abell and Hamilton, 2013).

Our finding that measured Q–weighted mean concentrations (C_{QWM}) of TP and SS during storm events (2010–2012) were greatly underestimated relative to simulated daily mean TP and SS concentrations has important implications for studies that examine effects of altered flow regimes on contaminant transport. For example, studies which simulate scenarios comprising more frequent large rainfall events (associated with climate change predictions for many regions; IPCC, 2013) may considerably underestimate projected future loads of SS and associated particulate nutrients if only base flow water quality measurements (i.e. those predominantly collected during 'state of environment' monitoring) are used for 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 statistics relating to validation of modelled SS concentrations using monthly grab samples (predominantly base flow; 'very good') and C_{QWM} estimated during storm sampling ('unsatisfactory') based on R^2 and NSE values.

4.2 Key uncertainties

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

4.3 Temporal dynamics of parameter sensitivity

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

In a study of a lowland catchment (481 km^2), Guse et al. (2014) found that three groundwater parameters, RCHRG_DP (aquifer percolation coefficient), GW_DELAY (groundwater delay) and ALPHA_BF (base flow alpha factor) were highly sensitive in relation to simulating discharge during quick flow, while ESCO (soil evaporation compensation factor) was most sensitive during base flow. 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 that, relative to our study catchment, the catchment studied by Guse et al. (2014) had moderate precipitation (884 mm y^{-1}) with less forest cover and flatter topography. Although the GW_DELAY parameter reflects the time lag that it takes water in the soil water to enter the shallow aquifers, its lack of sensitivity under both base flow and quick flow conditions in this study is a reflection of higher water infiltration rates and steeper slopes. The ESCO parameter controls the upwards movement of water from lower soil layers to meet evaporative 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 extensive forest cover in the Puarenga Stream catchment, which reduces soil evaporative demand by shading. Soil texture is also likely a contributor to this result. The predominant soil horizon type in the Puarenga Stream catchment was A, indicating high macroporosity which promotes high water infiltration rate and inhibits upward transport of water by capillary action (Neitsch et al., 2011). The variability in the sensitivity of the parameter SURLAG (surface runoff lag coefficient) between this study (relatively insensitive) and that of Cibin et al.

(2010; relatively sensitive) likely reflects differences in catchment size. The Puarenga Stream catchment (77 km^2) is much smaller than the study catchment (St Joseph River; 2800 km^2) of Cibin et al. (2010) and, consequently, distances to the main channel are much shorter, with less potential for attenuation of surface runoff in off–channel storage sites. The curve number (CN2) parameter was found to be insensitive in both this study and Shen et al. (2012), because surface runoff 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. 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, due to a general lack of ground cover under plantation trees and formation of gully networks on steep terrain.

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 estimation than parameters that control groundwater processes (e.g. ALPHA_BF and RCHRG_DP), reflecting the importance of surface flow processes for 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 that sediment transport is largely dependent on rainfall–driven processes, as is typical of steep and lower–order catchments. Modelled base flow NO_3–N loads were most sensitive to the nitrogen concentration in rainfall (RCN) because of rainfall as a predominant contributor to recharging base flow. The nitrogen percolation coefficient (NPERCO) was more influential for quick flow NO_3–N load estimation, probably indicating that the quick flow NO_3–N load is more influenced by the mobilisation of concentrated nitrogen sources associated with agriculture or treated wastewater distribution. High sensitivity of the organic carbon content (SOL_CBN) for quick flow ORGN load estimates likely reflects mobilisation of N associated with organic material following rainfall. The finding that base flow NH_4–N load was more sensitive to nitrification rate in reach (BC1) likely reflects that base flow provides more favourable conditions to complete this oxidation reaction, as NH_4–N is less readily leached and transported. Similarly, the ORGP mineralization rate (BC4) strongly influenced base flow MINP load estimation, reflecting that base flow phosphorus transport is relatively more 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.

5 Conclusions

The performance of a SWAT model was quantified for different hydrologic conditions in a small catchment with mixed land use. 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. 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. 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. Improved representation of groundwater processes in the model structure would reduce this aspect of model uncertainty. The sensitivity of many parameters varied depending on the relative dominance of base flow and quick flow, while curve number, soil evaporation compensation factor, surface runoff lag coefficient, and groundwater delay were largely invariant to the two flow regimes. Parameters relating to main channel processes were more sensitive when estimating variables (particularly Q and SS) during base flow, while those relating to overland processes were more sensitive for simulating variables associated with quick flow. Temporal dynamics of both parameter sensitivity and model performance due to dependence on hydrologic conditions should be considered in further model applications. This study has important implications for modelling studies of similar catchments that exhibit short–term temporal fluctuations in stream flow. In particular these include small catchments with relatively steep terrain and lower order streams with moderate to high rainfall.

Acknowledgements

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

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.

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.

Bieroza, M.Z., Heathwaite, A.L., Mullinger, N.J., and Keenan, P.O.: Understanding nutrient biogeochemistry in agricultural catchments: the challenge of appropriate monitoring frequencies, Environ. Sci.: Processes Impacts, 16, 1676–1691, doi: 10.1039/c4em00100a, 2014.

Boyle, D.P., Gupta, H.V., and Sorooshian, S.: Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods, Water Resour. Res., 36, 3663–3674, doi: 10.1029/2000WR900207, 2000.

Brigode, P., Oudin, L., and Perrin, C.: Hydrological model parameter instability: A source of additional uncertainty in estimating the hydrological impacts of climate change?, J. Hydrol., 476, 410–425, doi: 10.1016/j.jhydrol.2012.11.012, 2013.

Cao, W., Bowden, W.B., Davie, T., and Fenemor, A.: Multi–variable and multi–site calibration and validation of SWAT in a large mountainous catchment with high spatial variability, Hydrol. Process., 20, 1057–1073, doi: 10.1002/hyp.5933, 2006.

Chiwa, M., Ide, J., Maruno, R., Higashi, N., and Otsuki, K.: Effects of storm flow samplings on the evaluation of inorganic nitrogen and sulfate budgets in a small forested watershed, Hydrol. Process., 24, 631–640, doi: 10.1002/hyp.7557, 2010.

Choi, H.T. and Beven, K.J.: Multi–period and multi–criteria model conditioning to reduce prediction uncertainty in an application of TOPMODEL within the GLUE framework, J. Hydrol. (NZ), 332, 316–336, doi: 10.1016/j.jhydrol.2006.07.012, 2007.

Chow, V.T.: Open–channel hydraulics, Blackburn Press, Caldwell, New Jersey, 2008.

Cibin, R., Sudheer, K.P., and Chaubey, I.: Sensitivity and identifiability of stream flow generation parameters of the SWAT model, Hydrol. Process., 24, 1133–1148, doi: 10.1002/hyp.7568, 2010.

Conan, C., Bouraoui, F., Turpin, N., de Marsily, G., and Bidglio, G.: Modelling flow and nitrate fate at catchment scale in Britany (France), J. Environ. Qual., 32, 2026–2032, doi:10.2134/jeq2003.2026, 2003.Dairying Research Corporation, AgResearch, Fert Research: Fertilizer use on New Zealand Dairy Farms, In New Zealand Fertiliser Manufacturers’ Research

Association, Roberts, A.H.C. and Morton, J.D. (eds), Auckland, New Zealand, 36, 1999.

Eckhardt, K. and Arnold, J.G.: Automatic calibration of a distributed catchment model, J. Hydrol., 251, 103–109, 2001.

Ekanayake, J. and Davie, T.: The SWAT model applied to simulating nitrogen fluxes in the Motueka River catchment, Landcare Research ICM Report 2004–05/04, Landcare Research, Lincoln, New Zealand, 18, 2005.

Environment Bay of Plenty: Historical data summary, Report prepared for Bay of Plenty Regional Council, New Zealand, 522. 2007.

Fert Research: Fertilizer Use on New Zealand Sheep and Beef Farms, In New Zealand Fertiliser Manufacturers' Research Association, Balance, J.M. and Ravensdown, A.R. (eds), Newmarket, Auckland, New Zealand, 52, 2009.

Gassman, P.W., Reyes, M.R., Green, C.H., and Arnold, J.G.: The Soil and Water Assessment Tool: Historical development, applications, and future research directions, T. ASABE, 50, 1211–1250, 2007.

Glover, R.B.: Rotorua Chemical Monitoring to June 1993, GNS Client Report prepared for Bay of Plenty Regional Council, #722305.14, Bay of Plenty Regional Council, New Zealand, 38, 1993.

Green, W.H. and Ampt, G.A.: Studies on soil physics, part I – the flow of air and water through soils, J. Agr. Sci., 4, 1–24, doi: 10.1017/S0021859600001441, 1911.

Gupta, H.V., Sorooshian, S., and Yapo, P.O.: Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration, J. Hydrol. Eng., 4, 135–143, doi: 10.1061/(ASCE)1084–0699(1999)4:2(135), 1999.

Guse, B., Reusser, D.E., and Fohrer, N.: How to improve the representation of hydrological processes in SWAT for a lowland catchment–temporal analysis of parameter sensitivity and model performance, Hydrol. Process., 28, 2651–2670, doi: 10.1002/hyp.9777, 2014.

Hall, G.M.J., Wiser, S.K., Allen, R.B., Beets, P.N., and Goulding, C.J.: Strategies to estimate national forest carbon stocks from inventory data: the 1990 New Zealand baseline, Glob. Change Biol., 7, 389–403, doi: 10.1046/j.1365–2486.2001.00419.x, 2001.

Hopmans, P. and Elms, S.R.: Changes in total carbon and nutrients in soil profiles and accumulation in biomass after a 30–year rotation of *Pinus radiata* on podzolized sands: Impacts of intensive harvesting on soil resources, Forest Ecol. Manag., 258, 2183–2193, doi: 10.1016/j.foreco.2009.02.010, 2009.

IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Stocker, T.F., Qin, D., Plattner, G.K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P.M. (eds), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535, 2013.

Jowett, I.: Instream habitat and minimum flow requirements for the Waipa Stream, Ian Jowett Consulting Client report: IJ0703, Report prepared for Rotorua District Council, Rotorua, New Zealand, 31, 2008.

Kirschbaum, M.U.F. and Watt, M.S.: Use of a process–based model to describe spatial variation in *Pinus radiata* productivity in New Zealand, Forest Ecol. Manag., 262, 1008–1019, doi: 10.1016/j.foreco.2011.05.036, 2011.

Krause, P., Boyle, D.P., and Bäse, F.: Comparison of different efficiency criteria for hydrological model assessment, Advances in Geosciences, 5, 89–97, 2005.

Kusabs, I. and Shaw, W.: An ecological overview of the Puarenga Stream with particular emphasis on cultural values: prepared for Rotorua District Council and Environment Bay of Plenty, Rotorua, New Zealand, 42, 2008.

Lane, L.J.: Chapter 19: Transmission Losses, In Soil Conservation Service, National engineering handbook, section 4: hydrology, U.S. Government Printing Office, Washington, D.C., 19-1–19-21, 1983.

Ledgard, S. and Thorrold, B.: Nitrogen Fertilizer Use on Waikato Dairy Farms, AgResearch and Dexcel, New Zealand, 5, 1998.

Lim, K.J., Engel, B.A., Tang, Z., Choi, J., Kim, K., Muthukrishnan, S., and Tripathy, D.: Automated Web GIS–based Hydrograph Analysis Tool, WHAT, J. Am. Water Resour. As., 41, 1407–1416, doi: 10.1111/j.1752-1688.2005.tb03808.x, 2005.

Lindenschmidt, K., Fleischbein, K., and Baborowski, M.: Structural uncertainty in a river water quality modelling system, Ecol. Model., 204, 289–300, doi: 10.1016/j.ecolmodel.2007.01.004, 2007.

Lowe, A., Gielen, G., Bainbridge, A., and Jones, K.: The Rotorua Land Treatment Systems after 16 years, In New Zealand Land Treatment Collective–Proceedings for 2007 Annual Conference, Rotorua, 14–16 March 2007, 66–73, 2007.

Mahon, W.A.J.: The Rotorua geothermal field: technical report of the Geothermal Monitoring Programme, 1982–1985, Ministry of Energy, Oil and Gas Division, Wellington, New Zealand, 1985.

Marino, S., Hogue, I.B., Ray, C.J., and Kirschner, D.E.: A methodology for performing global uncertainty and sensitivity analysis in systems biology, J. Theor. Biol., 254, 178–196, doi: 10.1016/j.jtbi.2008.04.011, 2008.

McKenzie, B.A., Kemp, P.D., Moot, D.J., Matthew, C., and Lucas, R.J.: Environmental effects on plant growth and development, In New Zealand Pasture and Crop Science, White, J.G.H. and Hodgson, J. (eds), Oxford University Press: Auckland, New Zealand, 29–44, 1999.

Monteith, J.L.: Evaporation and the environment. In the state and movement of water in living organisms, 19th Symposia of the Society for Experimental Biology, Cambridge Univ. Press, London, U.K., 1965.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., and Veith, T.L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, T. ASAE, 50, 885–900, 2007.

Morris, M.D.: Factorial sampling plans for preliminary computational experiments, Technometrics, 33, 161–174, 1991.

Neitsch, S.L., Arnold, J.G., Kiniry, J.R., and Williams, J.R.: Soil and Water Assessment Tool Theoretical Documentation Version 2009, Texas Water Resources Institute Technical Report No. 406, Texas A&M University System, College Station, Texas, 647, 2011.

Neyman, J.: Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability, Phil. Trans. R. Soc. A, 236, 333–380, doi: 10.1098/rsta.1937.0005, 1937.

Nielsen, A., Trolle, D., Me, W., Luo, L.C., Han, B.P., Liu, Z.W., Olesen, J.E., and Jeppesen, E.: Assessing ways to combat eutrophication in a Chinese drinking water reservoir using SWAT, Mar. Freshwater Res., 64, 475–492, doi: 10.1071/MF12106, 2013.

Paku, L.K.: The use of carbon–13 to trace the migration of treated wastewater and the chemical composition in a forest environment, Master Thesis, Science in Chemistry, the University of Waikato, Hamilton, New Zealand, 92, 2001.

Parliamentary Commissioner for the Environment: Water Quality in New Zealand: Land Use and Nutrient Pollution, New Zealand, 82, 2013.

Pfannerstill, M., Guse, B., and Fohrer, N.: Smart low flow signature metrics for an improved overall performace evaluation of hydrological models, J. Hydrol., 510, 447–458, 2014.

Radcliffe, D.E., Lin, Z., Risse, L.M., Romeis, J.J., and Jackson, C.R.: Modeling phosphorus in the Lake Allatoona watershed using SWAT: I. Developing phosphorus parameter values, J. Environ. Qual., 38, 111–120, doi:10.2134/jeq2007.0110, 2009.

Reusser, D.E., Blume, T., Schaefli, B., and Zehe, E.: Analysing the temporal dynamics of model performance for hydrological models, Hydrol. Earth. Syst. Sc., 13, 999–1018, doi:10.5194/hess-13-999-2009, 2009.

Reusser, D.E. and Zehe, E.: Inferring model structural deficits by analysing temporal dynamics of model performance and parameter sensitivity, Water Resour. Res., 47, W07550, 15pp, doi: 10.1029/2010WR009946, 2011.

Rice, J.A.: Mathematical statistics and data analysis, Boston, MA, Cengage Learning, 2006.

Rimmer, A. and Hartmann, A.: Optimal hydrograph separation filter to evaluate transport routines of hydrological models, J. Hydrol., 514, 249–257, doi: 10.1016/j.jhydrol.2014.04.033, 2014.

Rotorua District Council, Rotorua Wastewater Treatment Plant, Rotorua, New Zealand, 22, 2006.

Rutherford, K., Palliser, C., Wadhwa, S.: Prediction of nitrogen loads to Lake Rotorua using the ROTAN model, Report prepared for Bay of Plenty Regional Council, New Zealand, 183. 2011.

Sangrey, D.A., Harrop–Williams, K.O., and Klaiber, J.A.: Predicting ground–water response to precipitation, J. Geotech. Eng., 110, 957–975, doi: 10.1061/(ASCE)0733–9410(1984)110:7(957), 1984.

Schuol, J., Abbaspour, K.C., Yang, H., and Srinivasan, R.: Modeling blue and green water availability in Africa, Water Resour. Res., 44, W07406, 18 pp, doi: 10.1029/2007WR006609, 2008.

Seymour, G.: Predictive Inference: An Introduction, Chapman & Hall, New York, pp 280, 1993.

Shen, Z.Y., Chen, L., and Chen, T.: Analysis of parameter uncertainty in hydrological and sediment modeling using GLUE method: a case study of SWAT model applied to Three Gorges Reservoir Region, China, Hydrol. 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 forested watersheds, Water Resour. Res., 20, 1815–1822, doi: 10.1029/WR020i012p01815, 1984.

Statistics New Zealand: Fertiliser use in New Zealand, Statistics New Zealand, New Zealand, 13, 2006.

van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., and Srinivasan, R.: A global sensitivity analysis tool for the parameters of multi–variable catchment models, J. Hydrol., 324, 10–23, doi: 10.1016/j.jhydrol.2005.09.008, 2006.

Watt, M.S., Clinton, P.W., Coker, G., Davis, M.R., Simcock, R., Parfitt, R.L., and Dando, J.: Modelling the influence of environment and stand characteristics on basic density and modulus of elasticity for young *Pinus radiata* and *Cupressus lusitanica*, Forest Ecol. Manag., 255, 1023–1033, doi: 10.1016/j.foreco.2007.09.086, 2008.

White, K.L. and Chaubey, I.: Sensitivity analysis, calibration, and validations for a multisite and multivariable SWAT model, J. Am. Water Resour. As., 41, 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, C.: Development and testing of an in–stream phosphorus cycling model for the Soil and Water Assessment Tool, J. Environ. Qual., 43, 215–223, doi: 10.2134/jeq2011.0348, 2014.

White, P.A., Cameron, S.G., Kilgour, G., Mroczek, E., Bignall, G., Daughney, C., and Reeves, R.R.: Review of groundwater in Lake Rotorua catchment, Prepared for Environment Bay of Plenty, Institute of Geological &

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.

Table 1. Description of data used to configure 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)
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 analysis to estimate properties for unmeasured functional horizons.	New Zealand Land Resource Inventory & digital soil map (available at http://smap.landcareresearch.co.nz)
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 practices	Agricultural management schedules	Stock density	Statistics New Zealand, 2006; Ledgard and Thorrold, 1998
		Applications of urea and di–ammonium phosphate	Statistics New Zealand, 2006; Fert Research, 2009
		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.)

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

Land–cover type	Parameter	Definition	Value	Source
PINE (*Pinus radiata*)	HVSTI	Percentage of biomass harvested	0.65	(Ximenes et al., 2008)
	T_OPT (℃)	Optimal temperature for plant growth	15	(Kirschbaum and Watt 2011)
	T_BASE (℃)	Minimum temperature for plant growth	4	(Kirschbaum and Watt 2011)
	MAT_YRS	Number of years to reach full development	30	(Kirschbaum and Watt 2011)
	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)
FRSE (Evergreen forest)	HVSTI	Percentage of biomass harvested	0	–
	BMX_TREES (tonnes ha^{-1})	Maximum biomass for a forest	372	(Hall et al., 2001)
	MAT_YRS (years)	Number of years for tree to reach full development	100	–
PAST (Pastoral farm)	T_OPT (℃)	Optimal temperature for plant growth	25	(McKenzie et al., 1999)
	T_BASE (℃)	Minimum temperature for plant growth	5	(McKenzie et al., 1999)

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)	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).	BoPRC; Abell et al., 2013
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

[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.

Table 4. Summary of calibrated SWAT parameters. Discharge (Q), suspended sediment (SS) and total nitrogen (TN) parameter values were assigned using auto–calibration, while total phosphorus (TP) parameters were manually calibrated. SWAT default ranges 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
TP				
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 ℃	d^{-1}	0.01–0.7	0.3
RS2.swq	Benthic (sediment) source rate for dissolved phosphorus in the reach at 20 ℃	mg m^{-2} d^{-1}	0.001–0.1	0.02
RS5.swq	Organic phosphorus settling rate in the reach at 20 ℃	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	mg N L^{-1}	0–200	0.055
SHALLST_N.gw	Nitrate–nitrogen concentration in the shallow aquifer	mg N L^{-1}	0–1000	1
ERORGN.hru	Organic nitrogen enrichment ratio		0–5	3
CH_ONCO.rte	Organic nitrogen concentration in the channel	mg N L^{-1}	0–100	0.01
BC1.swq	Rate constant for biological oxidation of ammonium–nitrogen to nitrite–nitrogen in the reach at 20 ℃	d^{-1}	0.1–1	1
BC2.swq	Rate constant for biological oxidation of nitrite–nitrogen to nitrate–nitrogen in the reach at 20 ℃	d^{-1}	0.2–2	0.7
BC3.swq	Rate constant for hydrolysis of organic nitrogen to ammonium–nitrogen in the reach at 20 ℃	d^{-1}	0.2–0.4	0.4
RS3.swq	Benthic (sediment) source rate for ammonium–nitrogen in the reach at 20 ℃	mg m^{-2} d^{-1}	0–1	0.2
RS4.swq	Rate coefficient for organic nitrogen settling in the reach at 20 ℃	d^{-1}	0.001–0.1	0.05

Table 5. Criteria for model performance. Note: o_{n} is the n^{th} observed datum, s_{n} is the n^{th} simulated datum, $\bar{\mathrm{o}}$ is the observed mean value, $\bar{\mathrm{s}}$ is the simulated daily mean value, and N is the total number of observed data. Performance rating criteria are based on Moriasi et al. (2007) for Q: discharge, SS: suspended sediment, TP: total phosphorus and TN: total nitrogen. Moriasi et al. (2007) derived these criteria based on extensive literature review and analysing the reported performance ratings for recommended model evaluation statistics.

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

R^2: coefficient of determination

NSE: Nash–Sutcliffe efficiency

PBIAS: percent bias

Table 6. 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
Calibration with instantaneous measurements (2004–2008)		n = 1439	n = 43	n = 45	n = 39
	R^2	0.77 (Very good)	0.42 (Unsatisfactory)	0.02 (Unsatisfactory)	0.08 (Unsatisfactory)
	NSE	0.73 (Good)	-0.08 (Unsatisfactory)	-1.31 (Unsatisfactory)	-0.30 (Unsatisfactory)
	±PBIAS%	7.8 (Very good)	-18.3 (Very good)	23.8 (Very good)	-0.05 (Very good)
Validation with instantaneous measurements (1994–1997)		n = 1294	n = 37	n = 37	n = 36
	R^2	0.68 (Good)	0.80 (Very good)	0.01 (Unsatisfactory)	0.01 (Unsatisfactory)
	NSE	0.62 (Satisfactory)	0.76 (Very good)	-0.97 (Unsatisfactory)	-2.67 (Unsatisfactory)
	±PBIAS%	8.8 (Very good)	-0.32 (Very good)	24.5 (Very good)	-26.7 (Good)
Validation with Q–weighted mean concentrations (2010–2012)		–	n = 12	n = 18	n = 18
	R^2	–	0.38 (Unsatisfactory)	0.06 (Unsatisfactory)	0.46 (Unsatisfactory)
	NSE	–	-0.03 (Unsatisfactory)	-4.88 (Unsatisfactory)	0.42 (Unsatisfactory)
	±PBIAS%	–	43.9 (Satisfactory)	69.4 (Satisfactory)	-0.87 (Very good)

Table 7. 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.

Model performance	Statistics	Q			SS			TP			TN		
		Q_b	Q_q	Q_{all}	L_b	L_q	L_{all}	L_b	L_q	L_{all}	L_b	L_q	L_{all}
Calibration (2004–2008)	R^2	0.84	0.84	0.77	0.66	0.68	0.61	0.24	0.65	0.39	0.72	0.97	0.95
	NSE	0.6	0.71	0.73	0.33	0.33	0.27	-6.2	0.09	-0.17	0.5	0.89	0.85
	±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
Validation (1994–1997)	R^2	0.87	0.81	0.68	0.36	0.98	0.95	0.27	0.27	0.06	0.79	0.33	0.58
	NSE	0.56	0.62	0.62	-0.03	0.43	0.85	-1.9	0.04	-0.64	0.58	-0.07	0.33
	±PBIAS%	11.3	-1.2	8.8	34.5	-79.7	11.1	45.8	-9.3	37	-7.6	14.3	-2.5

R^2: coefficient of determination; NSE: Nash–Sutcliffe efficiency; PBIAS: percent bias

Table 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), NH_4–N (ammonium–nitrogen), and NO_3–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 ($p \leqslant 0.05$). 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 4.

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					

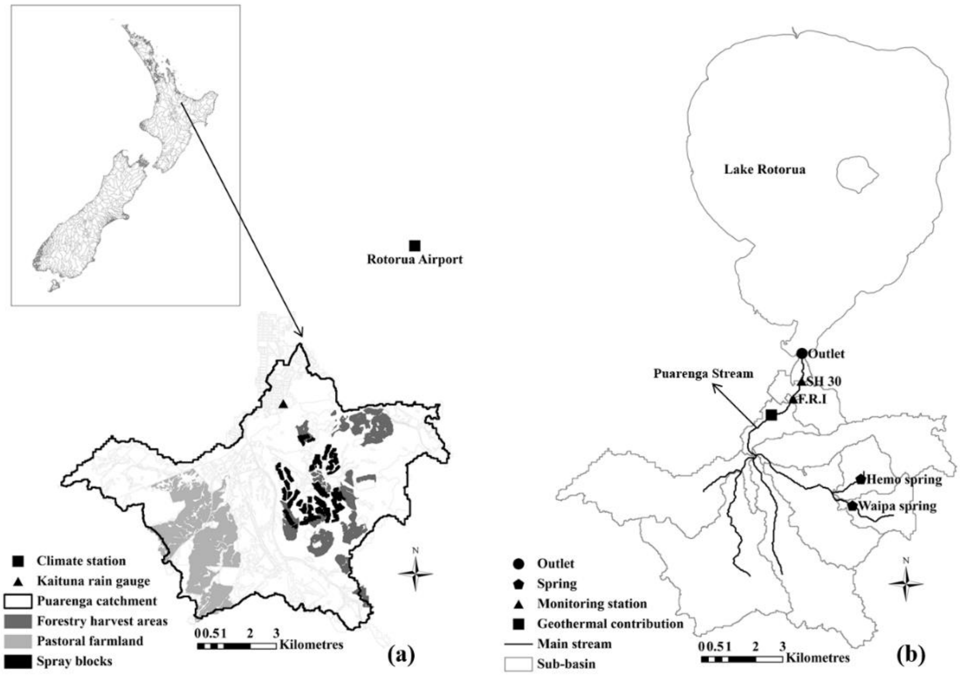

Figure 1. (a) Location of Puarenga Stream surface catchment in New Zealand, Kaituna rain gauge, climate station and managed land areas for which management schedules were prescribed in SWAT, and (b) location of the Puarenga Stream, major tributaries, monitoring stream–gauges, two cold–water springs and the Whakarewarewa geothermal contribution. 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.

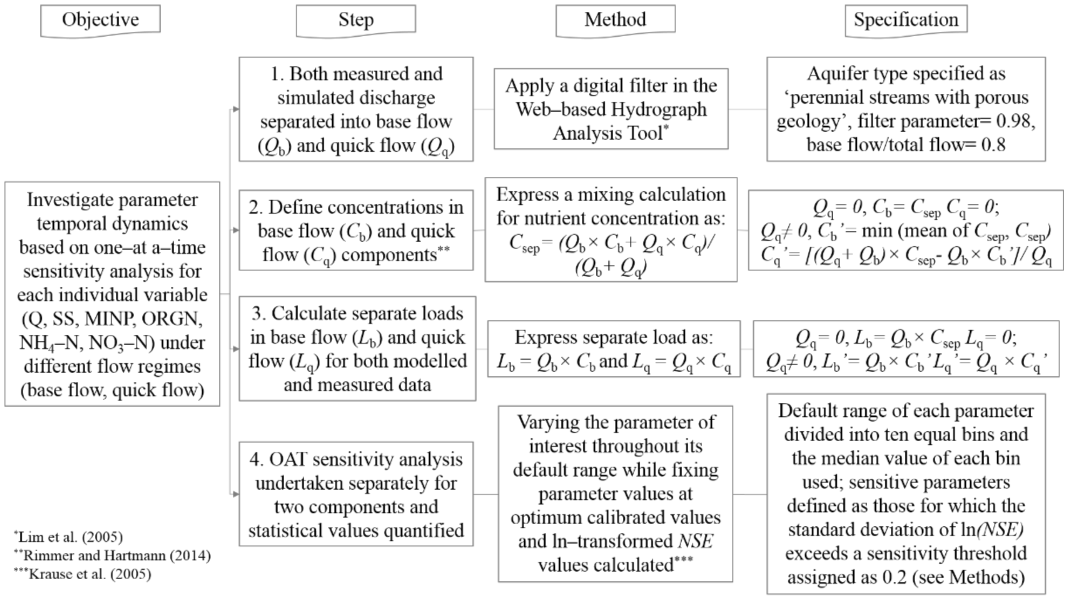

Figure 2. Flow chart of methods used to separate hydrograph and contaminant loads and to quantify parameter sensitivities for: Q (discharge), SS (suspended sediment), MINP (mineral phosphorus), ORGN (organic nitrogen), NH_4–N (ammonium–nitrogen), and NO_3–N (nitrate–nitrogen). *NSE*: Nash–Sutcliffe efficiency.

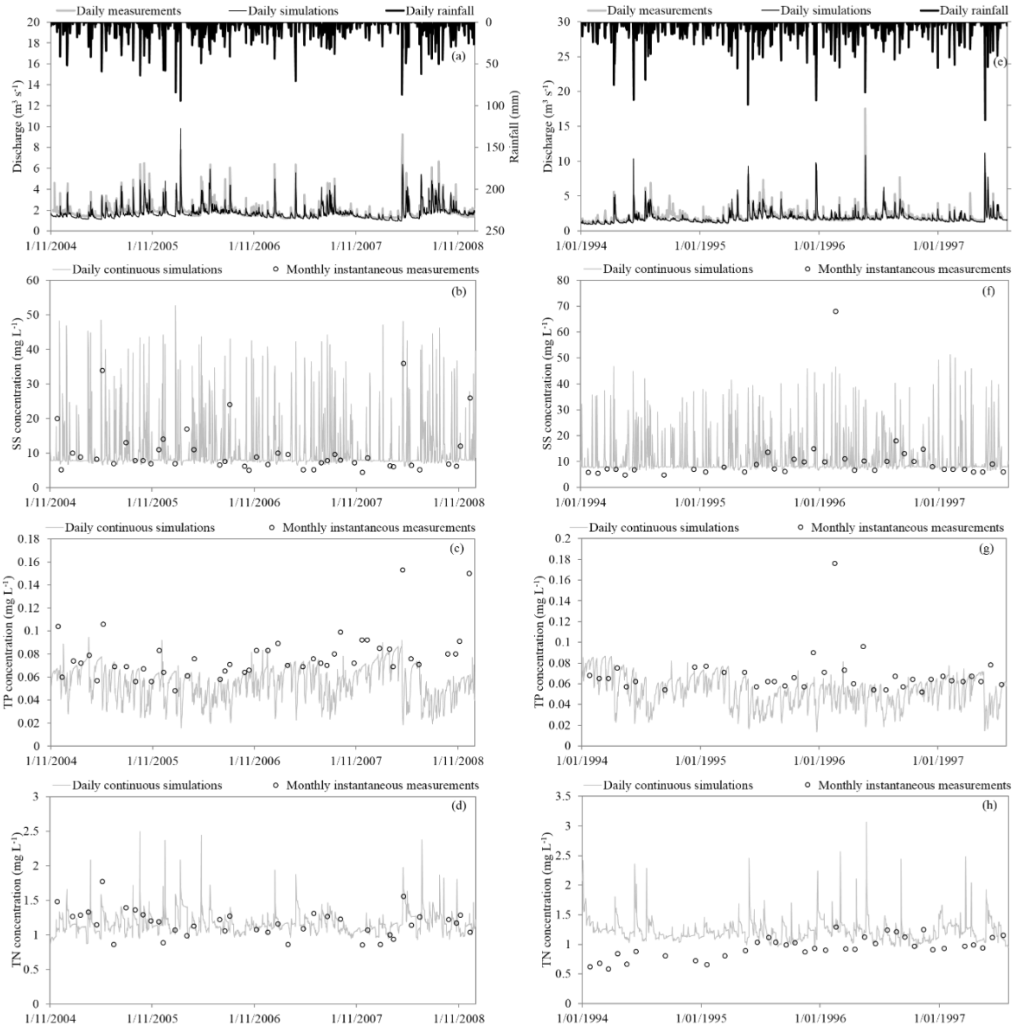

Figure 3. Measurements and daily mean simulated values of discharge, suspended sediment (SS), total phosphorus (TP) and total nitrogen (TN) during calibration (a–d) and validation (e–h). Measured daily mean discharge was calculated from 15–min observations and measured concentrations of SS, TP and TN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

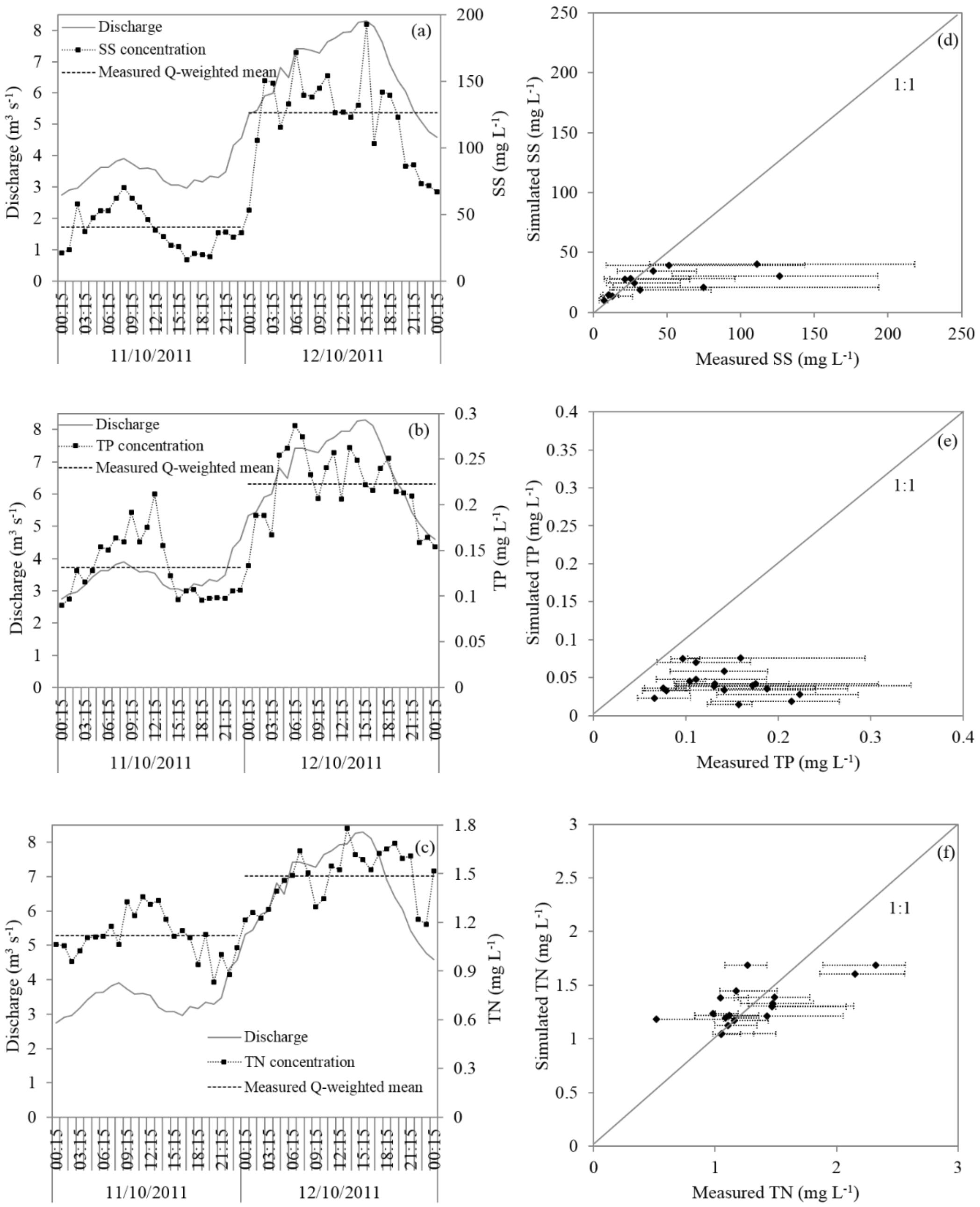

Figure 4. Example of a storm event showing derivation of discharge (Q)–weighted daily mean concentrations (dashed horizontal line) based on hourly measured concentrations (black dots) of suspended sediment (SS), total phosphorus (TP) and total nitrogen (TN) over two days (a–c). Comparisons of Q–weighted 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.

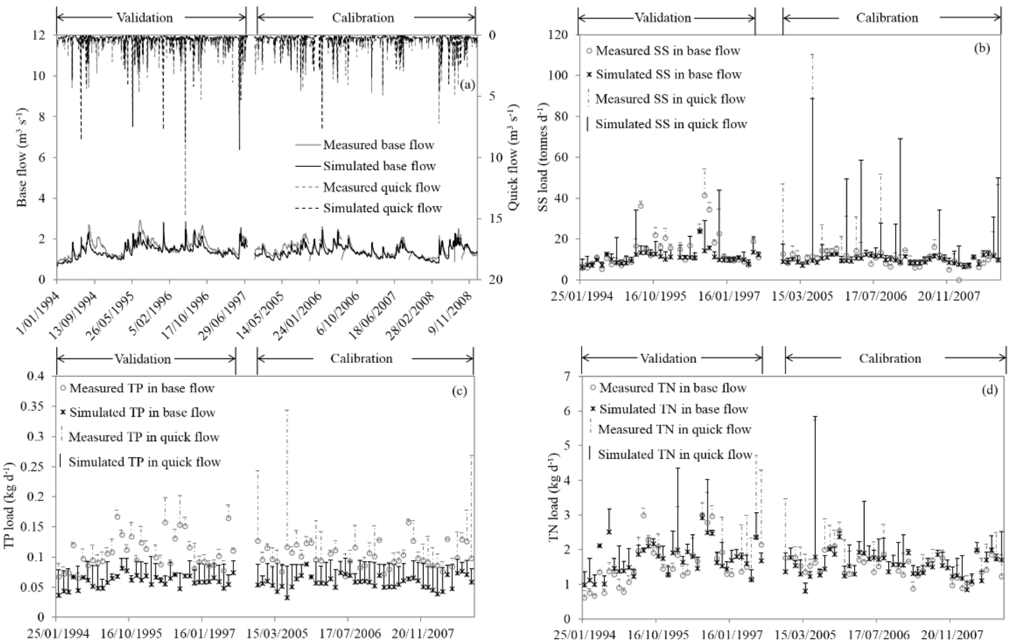

Figure 5. Measurements and simulations derived using the calibrated set of parameter values. Data are shown separately for base flow and quick flow. (a) Daily mean base flow and quick flow; (b) suspended sediment (SS) load; (c) total phosphorus (TP) load; (d) total nitrogen (TN) load. Vertical lines 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.

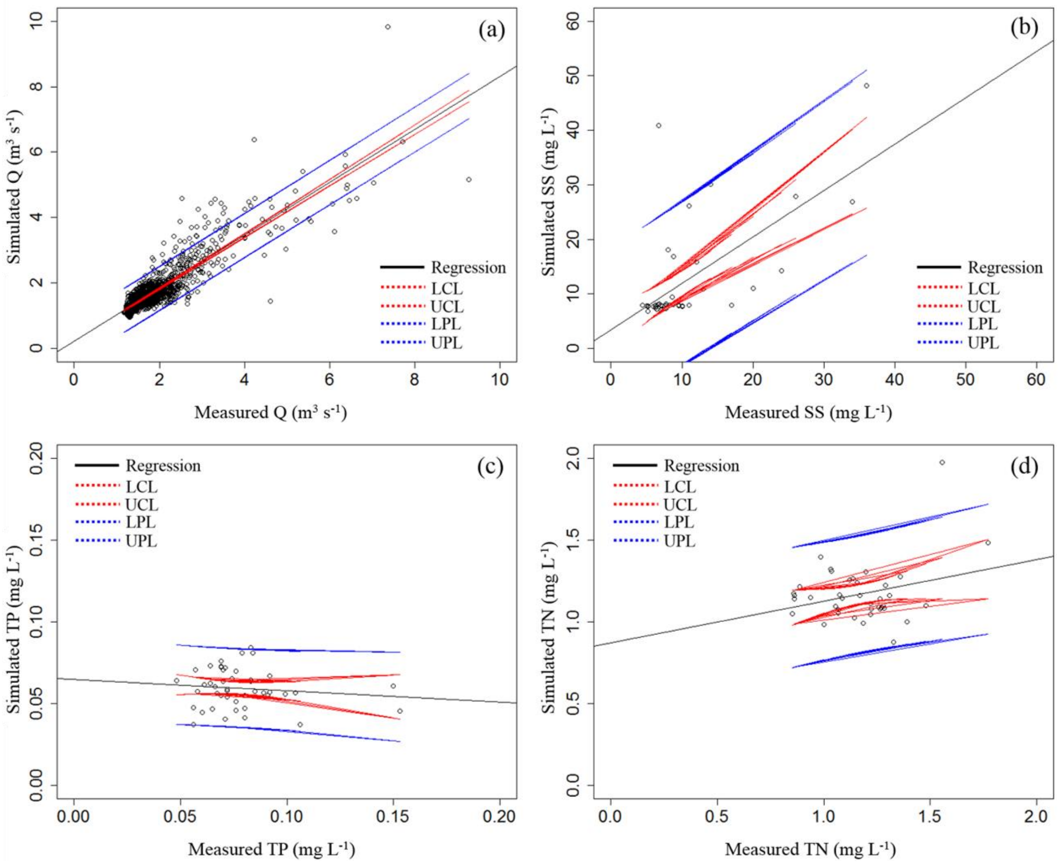

Figure 6. Regression of measured and simulated (a) discharge (Q), concentrations of (b) suspended sediment (SS), (c) total phosphorus (TP), and (d) total nitrogen (TN) including lower and upper 95% confidence limits (LCL and UCL) and lower and upper 95% prediction limits (LPL and UPL). Note that the "choppy" shape of confidence limits shown in figures b–d resulted from the few data points (< 50) in the regressions of measured and simulated SS, TP and TN concentrations.

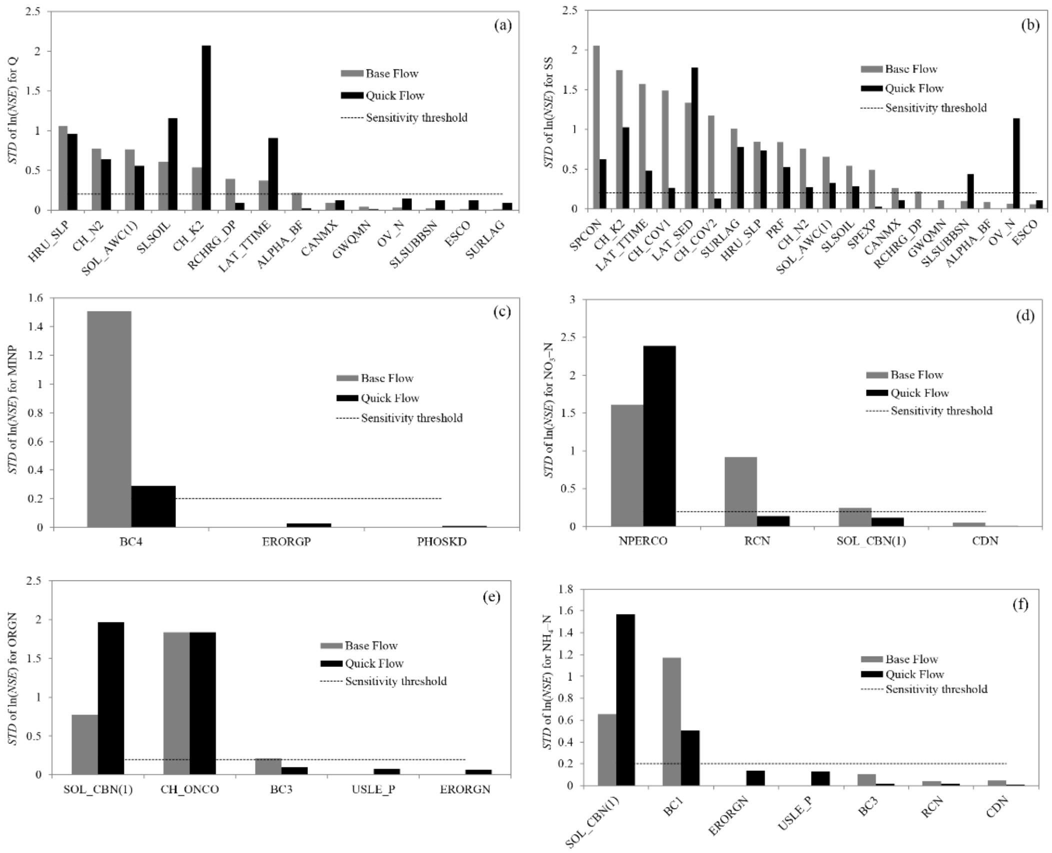

Figure 7. The standard deviation (STD) of the ln–transformed Nash–Sutcliffe efficiency (NSE) used to indicate parameter sensitivity based on one–at a–time (OAT) sensitivity analysis for separate base and quick flow components: (a) Q (discharge); (b) SS (suspended sediment); (c) MINP (mineral phosphorus); (d) NO_3–N (nitrate–nitrogen); (e) ORGN (organic nitrogen); (f) NH_4–N (ammonium–nitrogen). A median value (0.2) derived from the STD of ln–transformed NSE was chosen as a threshold above which parameters were deemed to be ‘sensitive’. Definitions of each parameter are shown in Table 4.