Comments from Editor

The Article is potentially of interest for the SWAT users community and for the broader hydrological audience, but it needs significant revisions. The reviewers offer important suggestions, which I recommend to follow attentively. In addition, I provide some additional points below.

1. A plus of the paper is that it attempts to model transport in addition to flow. The results of transport simulation, however, are not so encouraging. There is a strong difference between calibration and validation performance, with much better model performance during the validation period. I would recommend a split sample approach, where the calibration and validation period are inverted, so to check the consistency of results.

- ➢ The editor stated that "model performance during the validation period showed much better than performance during the calibration period". However, the model performance ratings in Table 5 revealed that simulations of discharge and concentrations of TP and TN during model calibration indicated better model performance than validation performance, represented by statistical indices R^2, NSE, PBIAS.

  We decided not to invert the current calibration (2004–2008) and validation (1994–1997) periods into the counter way, of which the reason has been found in the response to *Reviewer #2, Comment #10*. They were also demonstrated more clearly in the text as follows: "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".

2. Can the bad performance for transport simulation during the calibration period be due to too short warmup period? What is the warmup period, and can it be increased?

- ➢ One year (1993) was used for model warmup. We believe that the length of the warmup period is not related to the poor performance of some aspects of the model. Instead, we believe that inadequate representation of groundwater processes is a key factor that affected nitrogen simulation, as we discuss in our response to *Reviewer #3, Comment #20*. Additional text has been added as follows: "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".

3. The authors seem to compare observed instantaneous concentration data (measured once per month) with modelled monthly averages. They should compared observed averaged with simulated averages, or observed instantaneous values with simulated instantaneous values. Please clarify this aspect and correct the manuscript if necessary.

➢ We compared simulated daily (not monthly) mean concentrations with concentrations measured on respective days. The measured data are 'instantaneous' in that they relate to grab samples that were collected at monthly frequency. In addition, we also compare simulated daily mean concentrations with discharge–weighted mean daily concentrations that were calculated based on samples collected every 1–2 h during high flow events. These measurements are more representative of 'real' daily mean values than single instantaneous samples collected during separate days. Thus, a key focus of our paper is to examine the uncertainties that are associated with using concentration data that are infrequent relative to discharge to calibrate hydrologic models of small catchments; something that is common practice in catchment modelling. This has been clarified as we discuss in our response to *Reviewer #3, Comment #7 (i)*: "Daily mean discharge was firstly calibrated based on daily mean values of 15–minute measurements. 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".

4. The paper structure could be improved. (1) "study area and model configuration" should be 2 separate paragraphs. (2) The model configuration section needs more details. E.g. how many HRUs does the catchment have? How were they defined? (3) How many parameters in total?, etc.

➢ Thanks for the suggestion.

(i) Please see the response to Reviewer #2, Comment #7 that Sections 2.1 'Study area' and 2.2 'Model configuration' have been separated.

(ii) Please see the response to Reviewer #2, Comment #5 that the section Model configuration is now more comprehensive as follows: “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/daily 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”.

(iii) There were a total of 197 parameters involved for the model configuration of this study. This has added (see Model configuration) in the text: “There were a total of 197 model parameters. Values of SWAT parameters were assigned based on…”.

5. Tables 2 and 3: can the parameters corresponding to hydrology, chemistry and sediment transport simulations be clearly separated.

➢ Table 2 shows prior–estimated parameter values for three dominant types of land–cover in the Puarenga Stream catchment.

  Table 3 can be separated for discharge and sediment in more details of their exclusive parameters. Please see a revised version below. Phosphorus and nitrogen parameters have been separated already.

Parameter	**Definition**	**Unit**	**Default range**
Q			
EVRCH.bsn	Reach evaporation adjustment factor		0.5–1
SURLAG.bsn	Surface runoff lag coefficient		0.05–24
ALPHA_BF.gw	Base flow alpha factor (0–1)		0.0071–0.0161
GW_DELAY.gw	Groundwater delay	d	0–500
GW_REVAP.gw	Groundwater “revap” coefficient		0.02–0.2
GW_SPYLD.gw	Special yield of the shallow aquifer	m^3 m^{-3}	0–0.4
GWHT.gw	Initial groundwater height	m	0–25
GWQMN.gw	Threshold depth of water in the shallow aquifer required for return flow to occur	mm	0–5000
RCHRG_DP.gw	Deep aquifer percolation fraction		0–1
REVAPMN.gw	Threshold depth of water in the shallow aquifer required for “revap” to occur	mm	0–500
CANMX.hru	Maximum canopy storage	mm	0–100
EPCO.hru	Plant uptake compensation factor		0–1
ESCO.hru	Soil evaporation compensation factor		0–1
HRU_SLP.hru	Average slope steepness	m m^{-1}	0–0.6
LAT_TTIME.hru	Lateral flow travel time	d	0–180
RSDIN.hru	Initial residue cover	kg ha^{-1}	0–10000
SLSOIL.hru	Slope length for lateral subsurface flow	m	0–150
CH_K2.rte	Effective hydraulic conductivity in the main channel alluvium	mm h^{-1}	0–500
CH_N2.rte	Manning's N value for the main channel		0–0.3
CH_K1.sub	Effective hydraulic conductivity in the tributary channel alluvium	mm h^{-1}	0–300
CH_N1.sub	Manning's N value for the tributary channel		0.01–30
SS			
CH_COV1.rte	Channel erodibility factor		0–0.6
CH_COV2.rte	Channel cover factor		0–1
LAT_SED.hru	Sediment concentration in lateral flow and groundwater flow	mg L^{-1}	0–5000
PRF.bsn	Peak rate adjustment factor for sediment routing in the main channel		0–2

SPCON.bsn	Linear parameter for calculating the maximum amount of sediment that can be re–entrained during channel sediment routing		0.0001–0.01
SPEXP.bsn	Exponent parameter for calculating sediment re–entrained in channel sediment routing		1–1.5
OV_N.hru	Manning's N value for overland flow		0.01–30
SLSUBBSN.hru	Average slope length	m	10–150
USLE_P.mgt	USLE equation support practice factor		0–1