The Comments on the manuscript entitled ”Modelling water, sediment and nutrient fluxes from a mixed land–use catchment in New Zealand: Effects of hydrologic conditions on SWAT model performance” submitted for possible publication in the Hydrology and Earth System Sciences:

I appreciate the Editor to give me a chance to review an interesting paper on application of the SWAT model for assessment of hydrological situations in a case study. In my opinion, there are some merits in this paper and it has a potential to publish in HESS. However, I have also some concerns on the different parts of the text. If the authors address carefully to all of my comments, I'll recommend publication of the manuscript:

1. Don't use “we” and “they” in the manuscript, “the authors” has a suitable replacement for these. Revise whole of the text with this correction.
2. Add some of the most important quantitative results to the Abstract.
3. Page 4317, line 12: Change “spatial and temporal” to “spatiotemporal”. Apply this for whole of the manuscript.
4. There are many useful and more new papers on auto-calibration in the different fields of hydrology which can increase reliability aspect of the methodology. Therefore, cite all of the below papers for this purpose:

Critical Areas of Iran for Agriculture Water Management According to the Annual Rainfall
Monthly Inflow Forecasting using Autoregressive Artificial Neural Network
Long-term runoff study using SARIMA and ARIMA models in the United States
Simulation of open- and closed-end border irrigation systems using SIRMOD
Analysis of potential evapotranspiration using 11 modified temperature-based models
A comprehensive study on irrigation management in Asia and Oceania
Future of agricultural water management in Africa

5. In this study, the authors measured the discharge every 15 minutes. In this case, why did the authors use daily scale instead hourly scale?
6. The length of the calibration period is 5 years while, the length of the validation period is 4 years. This leads to increase of uncertainty, because two-third of the data is commonly applied for calibration period.
7. The data used for validation period (1994-1997) occurred before calibration data (2004-2008)! How do the authors justify this abnormal selection?
8. Why is there a gap between calibration and validation periods (1998-2003)? Is this due to lack of measuring? Why?
9. In Table 4, what is the criterion to this classification? For instance, why did the values of R-square more than 0.7 indicate a very good correlation?
10. Figure 3 underline poor performance of the model in peak and low points. I suggest to the authors to use a separate index for evaluation of the error of peak points as follows:

$$PVC = \frac{\sqrt[4]{\sum_{i=1}^{N_p}\left(X_i - Y_i\right)^2 \times \left(X_i\right)^2}}{\sqrt{\sum_{i=1}^{N_p}\left(X_i\right)^2}}$$

$$LVC = \frac{\sqrt[4]{\sum_{i=1}^{N_l}\left(X_i - Y_i\right)^2 \times \left(X_i\right)^2}}{\sqrt{\sum_{i=1}^{N_l}\left(X_i\right)^2}}$$

Where, Xi and Yi are the ith observed and estimated values, respectively; $\overline{X}$ and $\overline{Y}$ are the average of Xi and Yi, Np is number of peak parameter greater than one-third of the mean peak parameter observed, Nl is

number of low parameter lower than one-third of the mean low parameter observed and n is the total numbers of data.

11. A temporal evaluation of error indices could be useful for better understanding of performance the SWAT model. The authors can read and cite the below papers:

Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir

Parameters Estimate of Autoregressive Moving Average and Autoregressive Integrated Moving Average Models and Compare Their Ability for Inflow Forecasting

12. How did the authors calculate evapotraspiration as an input parameters for the SWAT model?

13. In the Conclusion, discuss on the most important factors which are effective on variations of the base and quick flow in the study area.