Author Response to Reviewer 1:

Important Points:

The approach adopted in the calibration procedure was to continue to run SUFI-2 iterations until a further improvement in one calibration station was not possible without a deterioration in another. The parameter ranges of the last iteration were then used in subsequent model analyses. The current parameter ranges as published in the manuscript were achieved after two iterations. We did perform an additional iteration in an attempt to improve the calibration. However, we found that while an additional iteration resulted in a slight improvement in objective functions for some calibration stations, there was a deterioration in B2H014. This fact, and the fact that statistics for this station were not that good (Table 3) indicate that hydrological processes in this catchment may be fundamentally different from other parts of the catchment and a separate calibration for this part of the catchment may have been more appropriate. This aspect has been addressed in the Conclusion section (Page 18, Line 25 – Page 19, Line 11 – specific reference to B2H007 and B2H014 has been made in this section of the conclusion in the revised version). This section also elaborates on challenges associated with multi-site calibration approaches as adopted in this study. Multi-site calibration results in data from multiple sites across the catchment being used in the calibration process which results in better catchment representation in the calibration process. However, a drawback of this method is that calibration using multiple sites introduces more constraints on the process, which means that fewer iterations will be possible before a calibration iteration will result in an improvement at some sites and a deterioration in other sites.

The poor performance at B2H007 and B2H014 may also be related to the fact that this part of the catchment had a relatively poor weather data record and therefore had to rely on the SWAT weather generator for simulations (this has been discussed in the manuscript – (from Page 11, Line 19). Based on this response I have not performed additional iterations as suggested by the reviewer, however a more detailed explanation of the rationale behind the number of iterations in the calibration procedure has been included in the revised version of the manuscript.

2. The SWAT model does not simulate the effect of management of sewage treatment works (STWs). The only manner in which improved management of STWs can be simulated is simply by reducing the point source loading associated with each of the STWs (with the assumption being that improved management improves the quality of effluent released). In this respect SWAT was indeed used to simulate changes in STWs management, by reducing effluent concentrations from 4 mg/l (estimated current conditions) to 1 mg/l and 0.1 mg/l in the tested scenarios.

3. A warm-up period was included in the model runs. This warm-up period was for the year 1999 to 2001 (i.e. 3 years). Model data was only analysed for the years 2002 to 2010. This explanation has now been included in the text of the manuscript under section 2.3 (Model Calibration).

4. Various aspects in the Conclusion have been elaborated upon in the revised version of the manuscript so as to address this specific comment.

Specific Comments:

Page 1363, Line 9: Naish, replaced with Nash

Page 1363, Line 23: ‘a’ has been deleted.

Page 13630, Line 18: Dryland agriculture is commercial. This has been included in the revised manuscript.

Page 13641, Line 11: The model is likely to be very sensitive to fluctuations in OP effluent concentrations. This can be seen in the large reduction in OP loads when adopting a 1 mg L^{-1} effluent scenario. Improved model performance would be achieved through more rigorous monitoring of effluent concentrations at STWs. However considering the expense of monitoring a compromise needs to be made between increased monitoring or relying on a combination of limited measured data and modelled data (as in this study). The fact that OP loads were well predicted in catchments with and without STWs would indicate that the 4mg L^{-1} effluent concentration was a fairly accurate assessment of STWs throughout the catchment. This point was emphasised in Page 13645, Line 18-22 and has been further emphasised by including the following sentence: *“This further indicates that 4 mg L^{-1} was a reasonable estimate of current effluent concentrations from STWs.”*

Page 13641, Line 21: The author followed the recommendation of the reviewer (“…. to estimate daily weather.”

Page 13461, Line 25: A reference has been included in the revised version of the manuscript (DWA, 2011b).

Page 13642, Line 22: These parameters were obtained from the literature. The sentence has been modified to accurately reflect this: “*For flow calibration, sensitive parameters influencing peak and base flow were obtained from the literature (Neitsch et al., 2002; White and Chaubey, 2005) and used to calibrate flow simulations for the calibration stations.”*

Page 13642, Line 24: See response to “Important Points” above.

Page 13643, Line 2: Naish replaced with Nash

Page 13643, Line 3: R^2 were not in fact used. The author used the incorrect notation for the Nash-Sutcliffe coefficient. The sentence has been corrected as follows: *“The NS coefficients range from infinity to 1. If the NS coefficients values are less than or close to 0, the model simulation is taken as an indication of poor or unacceptable performance.”*

Page 13644, Line 1: A warm-up period was included in the model runs. This warm-up period was for the year 1999 to 2001 (i.e. 3 years). Model data was analysed for the years 2002 to 2010.

Page 13644, Line 23: This sentence has been changed as recommended by the reviewer: “*In general, flow simulations at calibration stations compared well to measured flow records and simulated monthly maximum flow volumes were similar to maximum volumes measured at flow gauging stations across the catchment (Fig. 2).*

Page 13645, Line 6: This aspect is dealt with later on in the discussion (Page 13645, Line 25 to Page 13646, Line 2).

Page 13645, Line 11: This sentence has been corrected: *"With the exception of B2H014, NS statistics for the validation stations were slightly lower than for calibration stations across (Table 3)."*

Page 13645, Line 17: This reference has now been included in the revised manuscript. The sentence has been updated as follows: *"The SWAT model has been successfully applied to simulate hydrology in other African countries (Baker and Miller, 2013; Schuol et al., 2008), as well as at the catchment (Andersson et al., 2011; Govender and Everson, 2005) and national scale (Andersson et al., 2012) in South Africa."*

Page 13645, Line 20: This reference has been included in the revised manuscript.

Page 13645, Line 25: abscent corrected to absent.

Page 13646, Line 10: More detail on the manner in which average loads were calculated has been provided: *"In this context, the difference between mean annual measured and predicted loads was considered to be an acceptable form of evaluating model performance. Mean annual loads were calculated by summing monthly loads for a year and determining the annual average for the years 2002 to 2010."*

Page 1346, Line 28: The map has been revised to indicate mean annual OP loads generated per sub-catchment (as opposed to accumulated loads from upstream to downstream). This has the effect of clearly identifying which sub-catchments are responsible for generating high OP loads.

Page 1347, Line 6: This is simply because some STWs are smaller and release lower volumes of effluent (and therefore smaller loads of OP) on a daily basis. While the same concentration of OP in effluent was used for all WWTWs (i.e. 4 mg/L or 1 or 0.1 mg/L for the effluent reduction scenarios), the specific effluent discharge rates for each WWTW were incorporated into the model. This means that OP loads generated by each WWTW will vary according to the discharge rate of the specific WWTW. This was described under section 2.2 (Page 13641, Line 12-13 of the HESSD pdf).

Page 13647, Line 7: The author feels that this paragraph emphasises the point that very few studies have applied SWAT in predicting OP loads and concentrations in rivers and especially reservoirs. This emphasises that the results from this study represent a relatively new contribution to the application of SWAT, within the South African and global context.

Page 13467, Line 25: While the author concedes that there are large discrepancies at the monthly resolution, the PBIAS statistics nevertheless do indicate that the range of simulated OP concentrations fall within an acceptable range of measured concentrations, indicating that the model was able to make realistic simulations of OP concentrations at the monthly time-step. However falling within a range is not the same as an accurate simulation as would be indicated by a good NS statistic (i.e. while the concentrations are within a similar range they do not necessarily correspond well along a monthly time resolution). For this reason the first sentence of this section has been deleted and the following sentence has been modified as follows: *"PBIAS statistics indicated that simulated OP concentrations in the four reservoirs were within an acceptable range of measured concentrations, particularly in the Witbank and Bronkhorstspruit reservoirs (Fig. 5)."*

Page 13649, Line 24: This would include factors such as riparian buffer strips, conservation tillage, fertilizer management etc.. This has been explicitly stated in the revised version of the manuscript.

Page 13651, Line 9: It would not be possible to simulate this with SWAT as currently the model does not simulate the cycling of nutrients within a reservoir system. Thus the study focussed on predicted OP concentrations and associated trophic status associated with the three STW scenarios. These are presented in Figure 8.

Page 13652, Line 3: Under prediction changed to under-prediction.

Page 13652, Line 17: Undoubtedly has been changed to "most likely".

Table 2: Capacity has been corrected.

Table 3:

> The caption has been modified to reflect the fact that these statistics are for monthly simulations.
>
> B1H004 was not spaced differently in the original submission. This has obviously occurred during the type-setting process.
>
> The comment related to the poor statistics for B2H014 has been dealt with in my response to point 1 under "Important Points".

Figure 4: The actual figure and the figure caption have been updated as per the reviewers comments.

Figure 5: For all reservoirs, measured OP concentrations were not well simulated by the model from 2008 onwards, with measured concentrations being considerably lower than simulated predictions. No obvious modelling errors could be detected that could account for this difference. The discrepancy could possibly be related to the inability of the model to simulate nutrient cycling associated with the production of algal blooms in the reservoir. For example the first large algal bloom recorded in Loskop Reservoir was in 2007, after which time blooms have re-occurred on an annual basis (Dabrowski et al., 2013). Subsequent monitoring in the reservoir revealed very low concentrations of OP associated with algal blooms – presumably due to uptake of OP to facilitate growth of algae (Dabrowski et al. 2013). The fact that these blooms occurred near the inflow of the reservoir (in the lacustrine zone) and that monitoring takes place at the dam wall could suggest that algal blooms may intercept OP concentrations and deplete levels before they reach the dam wall. None of the other reservoirs were studied in as sufficient detail to determine whether similar processes occur. As all reservoirs showed very similar declines in OP concentrations, analytical error can also not be ruled out. (This explanation has been included in the revised version of the manuscript).

Dabrowski, J., Oberholster, P. J., Dabrowski, J. M., Le Brasseur, J., and Gieskes, J.: Chemical characteristics and limnology of Loskop Dam on the Olifants River, in light of recent fish and crocodile mortalities, Water SA, 39, 675-686, 2013.

Figure 8: The figure caption has been updated to address the reviewers comment: *"Linear regression plots of simulated loads and mean annual concentrations of orthophosphate associated with three sewage orthophosphate effluent standards (4 mg L^{-1} – ■; 1 mg L^{-1} – ●; 0.1 mg L^{-1} - ▲) in fours reservoirs located in the upper Olifants catchment. Each marker represents a simulation a specific year from the period 2002 to 2010 (dotted lines represent 95 % prediction limits)"*.

Author Response to Reviewer 2:

Specific Comments:

Page 13638, Line 23: The Arnold et al. 2008 reference has been inserted as recommended by the reviewer.

Page 1368, Line 26: The Gassman et al., 2007 reference has been included.

Page 13641, Line 2-3: Given the large area of the catchment it was not possible to get detailed information of fertilizer application rates. There is also no central repository or database where this type of information can be retrieved. We therefore used the FAO data to estimate fertilizer application rates. It is important to note that the FAO data is not a recommended application rate, but is rather an estimate of actual fertilizer use based on a national survey of crops produced in the country.

Page 13641, Line 23-25: A reference for this data source has been included (DWA, 2011b)

Page 13646, Line 24-28: The map has been revised to indicate mean annual OP loads generated per sub-catchment (as opposed to accumulated loads from upstream to downstream). This has the effect of clearly identifying which sub-catchments are responsible for generating high OP loads.

Page 13647, Line 25: The reference for the source of this data has now been included in the Methodology under the 'Model Application' section. Measured data was obtained from Department of Water Affairs monitoring data.

Page 13649, Line 3-4: This sentence has been updated to improve clarity: *"STWs in the catchment area of the Bronkhorstspruit reservoir are therefore responsible for significant loading of OP in this reservoir."*

Page 13649, Line 6-21: Reference to this figure has now been included in the text.

Table 2: This asterisk has been removed.

Technical Corrections:

All technical corrections have been made according to the recommendations of the reviewer.

Author Response to Reviewer 3:

Study Site Description:

- The "Study Area" description has been updated to address comments rasied by the reviewer. In summary:
  - The reservoirs are regulated, with Loskop providing water to the second largest irrigation scheme in South Africa, while the other reservoirs supply water for domestic and industrial use.
  - Dryland agriculture is located primarily in the mid to upper reaches of the catchment, consisting mainly of maize production, which is the primary source of nonpoint source pollution, particularly during summer storm events.
  - STWs contribute a volume of approximately 7 % of mean annual runoff to the catchment (DWAF, 2004). (DWAF (Department of Water Affairs and Forestry): National Water Resource Strategy. Pretoria, South Africa, 2004.).

Figure 5:

- For all reservoirs, measured OP concentrations were not well simulated by the model from 2008 onwards, with measured concentrations being considerably lower than simulated predictions. No obvious modelling errors could be detected that could account for this difference. The discrepancy could possibly be related to the inability of the model to simulate nutrient cycling associated with the production of algal blooms in the reservoir. For example the first large algal bloom recorded in Loskop Reservoir was in 2007, after which time blooms have re-occurred on an annual basis (Dabrowski et al., 2013). Subsequent monitoring in the reservoir revealed very low concentrations of OP associated with algal blooms – presumably due to uptake of OP to facilitate growth of algae (Dabrowski et al. 2013). The fact that these blooms occurred near the inflow of the reservoir (in the lacustrine zone) and that monitoring takes place at the dam wall could suggest that algal blooms may intercept OP concentrations and deplete levels before they reach the dam wall. None of the other reservoirs were studied in as sufficient detail to determine whether similar processes occur. As all reservoirs showed very similar declines in OP concentrations, analytical error can also not be ruled out (This explanation has been included in the revised version of the manuscript).

  Dabrowski, J., Oberholster, P. J., Dabrowski, J. M., Le Brasseur, J., and Gieskes, J.: Chemical characteristics and limnology of Loskop Dam on the Olifants River, in light of recent fish and crocodile mortalities, Water SA, 39, 675-686, 2013.

Figure 8.

- This is not a serial correlation. The figures estimate OP concentrations (y) in the reservoirs (dependant variable) as a function of OP loading (x) that occurs upstream in the catchment (independent variable). While concentrations in the reservoirs are dependent on loads, the

SWAT model takes many other factors into account (such as the volume of water in the reservoir, settling out of nutrients, etc.).The graphs do not estimate loads into the reservoir as given from OP concentrations in the reservoir. The graphs estimate concentrations in the reservoir given from OP loads generated in the catchment.

- The Vollenweider equations/models can be used to determine trophic status based on knowledge of P loading (normalised by surface area), reservoir depth and residence time. Based on these input parameters the P concentrations can also be estimated (Nielsen et al. 2013). Under these specific equations, it would be important to use units as specified by the reviewer (mg/m^2/year). The Vollenweider model is a relatively simple approach used to make assessments of P concentrations and trophic status. The Vollenweider models also makes use of benchmark total phosphate concentrations (or guidelines) against which to determine trophic status (e.g. < 10 ug/l = oligotrophic; 10 -20 ug/l = mesotrophic, 20-50 ug/l = eutrophic). While the guideline values we have used here are slightly different (and applicable to ortho-phosphate as opposed to total phosphate), the approach used is essentially the same – this paper simply uses SWAT to estimate OP loading and OP concentrations in the reservoir (which takes a number of other factors into account, such as residence time, reservoir surface area, reservoir volume, nutrient settling etc.) as opposed to using the Vollenweider equations. Considering the study was performed in South Africa, it made sense to use guidelines relevant to South Africa. While this is specific to South Africa, the approach adopted here can be applied throughout the world using appropriate guidelines.

Nielsen, A., Trolle, D., Me, W., Luo, L., Han, B.-P., Liu, Z., 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, 2013.

- The units referred to by the reviewer are very specific to the Vollenweider equations and there are a number of other studies that report P loading in reservoirs as a function of total mass (as opposed to mass per surface area). Furthermore, one of the main messages of this paper is the reduction in loads that would be required to meet a specific trophic status. In terms of catchment management, it is very useful to quantify the total amount of P loading permissible so as to achieve a particular trophic status. For example in Bronkhosrtspruit reservoir it is evident that OP loads would need to be reduced by 4000 kg annually so as to bring OP concentrations within the mesotrophic range. Expressing loads as a unit of weight is the most useful manner in which to communicate this message and provides managers with meaningful targets by which to reduce OP loads at STWs. If loading was expressed as weight per unit surface area of the reservoir, it would in any event be necessary to calculate a total load (expressed as mass only, based on the reservoir surface area) so as to estimate by how much OP loading would need to be reduced further up in the catchment.

- While it would be ideal to have all graphs in the same scale the large differences in OP loading per reservoir would make the graphs for the smaller reservoirs impossible to read.

However, to alert readers to the difference in scale additional text has been added to the caption for the revised version: i.e. *(note the difference in scale on the x axis)*

- The guidelines for hyper-eutrophic conditions have not been indicated as the guideline value is 0.25 ug/l. Therefore adding this trophic level would significantly impair the readability and appearance of the figures. However, knowing the threshold level (0.25 ug/l) it would be possible to calculate the predicted loads for a hyper-eutrophic status by using the regression equations for each reservoir.

Seasonal Variation

OP variation in the reservoirs is predominantly as a result of stream flow changes. The model does not simulate nutrient cycling in reservoirs, which has been mentioned in the revised version of the manuscript. See response to Figure 5 above.