Authors Response to Interactive comments on **"Shallow water table effects on water, sediment and pesticide transport in vegetative filter strips: Part B. model coupling, application, factor importance and uncertainty" by C. Lauvernet an*d* R. Muñoz-Carpena et al.**

RC3- Anonymous (Referee)
Thank you very much for the careful review of the manuscript. Below we address the specific comments raised on the initial submission. Please note that we uploaded the revised manuscript as a supplement to RC1 response comments, with your suggested changes also there. [RC3-#: Reviewer 3 comment #; AR-#: Authors response to comment #).

RC3-0. The authors provide a good level of discussion providing necessary background and references concerning VFS simulation, the novel infiltration simulation model, study site data, and the GSA methods used to assess the model sensitivity and uncertainty As with Part A of this manuscript, the methods and assumption are clearly described, and the results sufficiently support many of the conclusions; however, the validation of the model with observed field data is limited.

> AR-0. Thank you for your positive comments. Regarding the model validation, the main objective of paper B is to build on the numerical validation obtained in the first manuscript, to couple SWINGO with the existing VFSMOD model and analyze the potentially important effects that WT can have in VFS performance. This manuscript does not intend to provide a thorough experimental testing of the coupled model. Indeed, this is an important next step that our team undertook through laboratory controlled experiments and it is the subject of a follow up manuscript (under review elsewhere and working on minor revisions --we can provide a copy upon request). In Paper B, our approach was to select two markedly different (soils, slope, climate, crops) experimental sites where filters are already in use and with seasonal WT present, to assess the impacts of VFS performance under realistic conditions. We did that in the hope to lend additional credibility to the sensitivity analysis part of the paper, in contrast with synthetic conditions that we could have selected arbitrarily. The field testing of a detailed numerical model like VFSMOD requires intensive field measurements, in particular detailed hydrographs and pollutographs that are not often available unless the monitoring sites were initially designed with model testing in mind. Moreover, to our knowledge there is no field VFS experiment with a shallow water table published before.

> This is a limitation of the experimental sites used. In the interest of conciseness, we omitted and simplified important details related to the data selected for the application and this has led to the excellent questions raised by the reviewer. Indeed, the description provided was incomplete and led to confusion. Thank you for identifying this weakness. We are expanding the description here and in the revised manuscript. While both Morcille and Jaillière provide sufficient details for application of the model (field parameters, initial and boundary conditions), VFS outflow was only available for Morcille. In particular, earlier studies at Jaillière by Patty et al. (1997) monitored VFS efficiency in the same site but in the absence of a shallow water table. Although they provide some of the model inputs they are not directly applicable for this WT model application. Later, working on the same watershed Branger et al. (2009) and Fontaine (2010) studied the shallow water table effects on runoff at the edge-of-the-field and a receiving drainage ditch, but did not monitor the efficiency of the VFS. We selected one average event (dynamics and volume) in the middle of the high-water season based on Fontaine (2010) for our model application. In Morcille, Lacas (2005) and Lacas et al. (2012) monitored the effectiveness of the VFS, but because of the high permeability of the soil and deeper shallow water conditions, only 5 out of the 24 natural rainfall events recorded generated outflow from the VFS. From these 5, the one closer to the average for the high water table season was selected for application of the model.

Because of these limitations, we do not state field testing as a main objective (title, abstract, conclusions) rather (title): "coupling, application, factor importance and uncertainty". Instead, we identify this as a limitation of the study and a future research need (that we address in another experimental manuscript under review). With this in mind, in addition to the revised text summarizing the paragraphs above, we attempt to add more information about the limited comparison as suggested by the reviewer in the comments below.

RC3-1. Visual comparison of the simulated versus modeled VFS out flow is provided in Figure 4a for 1 event. Could the authors provide some goodness of fit statistics, such as Nash-Sutcliffe efficiency or others in addition to the visual interpretation?

AR-1. Yes, we now provide Nash-Sutcliffe efficiency (NSE) and root mean square error (RMSE) ranges for the model uncertainty bounds in Fig. 4a, with median NSE = 0.610 and 95CI [0.448 - 0.943], and RMSE= 4.284e-05 [1.179e-05 - 7.472e-05] m^3/s. The performance was assessed based on FitEval software (Ritter and Muñoz-Carpena, 2013), with data uncertainty included using the modification of the NSE based on the probable error range (PER) method (Harmel et al, 2007). FitEval evaluation files are included in the revised Supp. Materials file.

RC3-2. Why are observations at the Jailliere not shown for comparison with the model simulation in Figure 4b? Even if detailed time series are not available, an event total volume could be compared between model and observations.

AR-2. Please see detailed explanation about this in AR-0 above.

RC3-3. The description of the study sites indicates that monitoring was conducted over a multi-year period at each site, and that around 20 events per site were analyzed. However, observations from only 1 event at 1 site were included for review and discussion in the manuscript. This leaves the reader wondering how the model performs for all the other storm events. It is important for readers to see how the model performs over a broader range of events, as 1 single event at one site is not enough to assess the validity of the model. The authors should provide the model simulation comparisons with observed data for the complete set of events available at each site which is necessary to better validate the new approach and any improvement it has over the approach that does not include the shallow water table.

AR-3. Yes, the description provided was incomplete and leads to confusion. Thank you for identifying this weakness. For the reasons presented in AR-0, there is no data available for systematic validation of the model and this was left outside the scope of the paper. The single event at Morcille with sufficient data indicates that the model responds in the same range as the measured field data. Since to our knowledge there is no field VFS experiment with a shallow water table that has ever been published, this motivated us to conduct our follow up lab experimental work (in a separate manuscript).

RC3-4. The model performance compared to observations at the Morcille site was fair, and the authors deemed the model performance as satisfactory given that the model was not calibrated. If the authors were to provide an example of the performance of the model when calibrated, readers would better understand the potential improvement in a calibrated model and be better able to make a judgement regarding the validity of this new modeling approach.

AR-4. Please notice that uncalibrated or "cold" testing of the model is the most stringent test a model could be subject to. As a way of calibration, the 95% confidence interval obtained by varying only K_s within measured values (Table 2) was presented (grey area in Fig. 4a), with NSE = 0.855

[0.455 - 0.982]. Within those uncertainty bounds, the model is classified as 'unacceptable' to 'very good' based on the FitEval methodology (Ritter and Muñoz-Carpena, 2013) discussed above in AR-1. Again, the intend of this paper was not to experimentally test the model, but to evaluate the effects of the presence of the shallow water table on VFS performance through a realistic application to 2 contrasting field conditions that experience seasonally high water table and use VFS for water quality protection.

RC3-5. Figure 4a, 4b: In both figures, the legends are incomplete. Symbology for all time series shown in the figures should be included.

AR-5. Yes, we revised the legends to reflect this.

RC3-6. Figure 5: The marker symbols that are included as part of several of the lines should be larger. The line symbology should be more distinct for the dashed lines if possible.

AR-6. Yes, we revised the figure to reflect this.

RC3-7. Figure 6: The labeling for the majority of the points on these plots in not legible. The labels that cannot be read should be removed from each plot. Labels should only be included for points of greatest significance or ones that are referred to in the text, and must be legible if included.

AR-7. Yes, we revised the figure to reflect this.

RC3-8. Figure 7: The x-axis labels are pretty tough to read because there are so many parameters shown. This could be potentially fixed by not showing some of the parameter results in the plot.

AR-8. We considered this but the importance of the factors changes within each graph, so that there is not a common subset important to all. Selecting different inputs in each of the figures would lead to confusion and instead we opted for leaving all of them so they could be compared across all cases. Notice that this is now in Supplementary Materials following Reviewers 1 and 2 comments.

RC3-9. Figure 8: The symbology is a little difficult to match between the legend and the figures. Part of the reason for this is that "dE" is concentrated up against the 100% level of the x-axis (accept for Figure 8b). One solution to improve this would be to also label each of the curves with an "callout" or arrow.

AR-9. We added the callout arrows.

References

Branger, F., Tournebize, J., Carluer, N., Kao, C., Braud, I., Vauclin, M., 2009. A simplified modelling approach for pesticide transport in a tile-drained field: The PESTDRAIN model. Agricultural Water Management 96, 415–428. doi:10.1016/j.agwat.2008.09.005

Fontaine, A., 2010. Optimizing the size of grassed buffer strips to limit pesticides transfer from land to surface water in overland flow (MSc THESIS). Cranfield University, UK.

Harmel, R. D., and P. K. Smith. 2007. Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling. *J. Hydrol.* 337(3-4): 326-336.

Lacas, J.-G., 2005. Processus de dissipation des produits phytosanitaires dans les zones tampons enherbées. Etude expérimentale et modélisation en vue de limiter la contamination des eaux de surface. Unversité Montpellier II. Sciences et techniques du Languedoc.

Lacas, J.-G., Carluer, N., Voltz, M.: Efficiency of a Grass Buffer Strip for Limiting Diuron Losses from an Uphill Vineyard Towards Surface and Subsurface Waters. Pedosphere, 22, 580-592, doi:10.1016/S1002-0160(12)60043-5, 2012.

Patty A., B. Réal and J.J. Gril. The use of grassed buffer strips to remove pesticides, nitrate and soluble phosphorus compounds from runoff water. Pestic. Sci. 49, 243-251, 1997.

Ritter, A. and R. Muñoz-Carpena. 2013. Predictive ability of hydrological models: objective assessment of goodness-of-fit with statistical significance. J. of Hydrology 480(1):33-45. doi:10.1016/j.jhydrol.2012.12.004