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.**

RC1- M Vanclooster (Referee)

Thank you very much for the careful review and edits to the initial submission. Below we address the comments raised on the initial submission and we have also revised the manuscript accordingly to accommodate these. Please note that we uploaded the revised manuscript as a supplement to these response comments. [RC1-#: Reviewer 1 comment #; AR-#: Authors response to comment #).

RC1-1: […]The paper enters in the scope of HESS, is well written and it follows a logical approach. It is therefore a significant contribution to hydrological science.

Yet, the added value of 2 sensitivity analysis approaches is not clear. Authors implement the Morris and eFast approaches that clearly are consistent and coherent. There seems to be little added value of implementing 2 sensitivity analysis approaches. The paper therefore loose some focus by complicating this analysis. It could be suggested to eliminate the Morris analysis which does not add new information as compared to the eFast analysis.

> AR-1: Yes, our rationale when including both GSA approaches was to ensure the robustness of the results. Admittedly, under common conditions, both types of analysis provide comparable results. However, methods have not been compared often in studies with complex environmental models where non-linearities can be high, although doing this allows for testing the robustness of the sensitivity indices (Pianosi et al. 2016). We should not ignore that initially Morris is a "qualitative" method since it is based on a sparse sampling (in our case with r=10 and k=18 (no WT) and 20 (with WT), N=r(k+1) =190 and 210 samples for each scenario) that could lead to inaccurate results when the model is highly non-linear or discontinuous in some region of the input factor space, compared to variance-decomposition methods like eFAST based on dense sampling (M=497, N=Mk=8946 and 9940). To increase the reliability of Morris, improved sampling techniques have been developed (e.g. Khare, Muñoz-Carpena et al., 2015) that intend to increase the robustness of the method and approximate more quantitative, comparable to those of variance methods results. This opens important opportunities for application to large models where only Morris might be feasible (e.g. see Srivastava, Graham, Muñoz-Carpena et al., 2014). In addition, we believe that the Morris plots provide an intuitive and clear way to assess the importance of the input factors and their interactions. On the other hand, the dense variance-based sampling allows for a follow up quantitative uncertainty analysis. Thus, the inclusion of the two methods in this work and the results obtained further corroborates the Morris efficiency for complex models and confirms the sensitivity of the input factors of the model.
>
> In spite of this rationale, we agree that the inclusion of both methods shifts the main focus away from the main objective of the paper related to analysis of WT effects in the coupled processes that occur in a vegetative filter strip. We now focus on the Morris in the revised manuscript and moved the eFAST figure to Supplementary Materials. For the interested reader, we leave a brief comment in the GSA results section (with reference to Supp. Mat.) on the robustness and insights that eFAST results lend to the work.

RC1-2: Further, the manuscript suffers from some editorals that should be considered in a minor revision of the manuscript before it can be accepted for publication.

> AR-2: We revised the manuscript to include all the suggested editorial comments.

References

Y.P. Khare, R. Muñoz-Carpena, R.W. Rooney, C.J. Martinez, A multi-criteria trajectory-based parameter sampling strategy for the screening method of elementary effects, Environmental Modelling & Software, Volume 64, February 2015, Pages 230-239, ISSN 1364-8152, https://doi.org/10.1016/j.envsoft.2014.11.013.

Pianosi, F.; Beven, K.; Freer, J.; Hall, J. W.; Rougier, J.; Stephenson, D. B. & Wagener, T. Sensitivity analysis of environmental models: A systematic review with practical workflow, In Environmental Modelling & Software, Volume 79, 2016, Pages 214-232, ISSN 1364-8152, https://doi.org/10.1016/j.envsoft.2016.02.008.

Srivastava, V., W.D. Graham, R. Muñoz-Carpena and R. Maxwell., Insights on geologic and vegetative controls over hydrologic behavior of a large complex basin – Global Sensitivity Analysis of an integrated parallel hydrologic model, Journal of Hydrology, Volume 519, Part B, 27 November 2014, Pages 2238-2257, ISSN 0022-1694, https://doi.org/10.1016/j.jhydrol.2014.10.020.

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.**

RC2- S. Reichenberger (Referee)

Thank you very much for the careful review and edits to the initial submission. Below we address the main 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. [RC2-#: Reviewer 2 comment #; AR-#: Authors response to comment #).

RC2-1. I think this paper is of good quality and that it complements its companion paper (HESS-2017-405) well. However, I have a few comments (see below).

General comments: As already pointed out by Marnik Vanclooster in his comment, the Morris sensitivity screening and eFAST are not really complementary methods: Both methods produce indices with similar interpretation, as you state yourselves: "Interestingly, the Morris indices (μ, σ) have been found to provide a good approximation to the eFAST indices (STi, STi-Si) at a much lower computational cost (Saltelli et al. 2004, Campolongo et al. 2007) making it ideal for large and computationally expensive models." The main advantage of the Morris sensitivity screening is that is computationally much more efficient than eFAST, especially if one has a large number of input factors. Usually, Morris and a true variance-based method such as eFAST are used in sequence: First, a Morris sensitivity screening is performed to identify and eliminate non-sensitive input factors. In the second step, eFAST is run with a reduced set of input factors (i.e. without the ones identified with Morris as non-sensitive). This two-step approach makes sense because most modellers do not have the computational resources available for a "brute force" approach. If, however, after the Morris sensitivity screening one runs eFAST with the same number of input factors as before, the Morris sensitivity screening becomes obsolete. As a justification of using the same sets of input factors for both Morris and eFAST, you wrote "In this study, both methods were run with the full set of inputs as a check for the consistency of the GSA results." If you want to keep both sensitivity analyses in the paper, maybe you should expand this further. However, this would draw the focus from the new VFSMOD version with shallow water table. Maybe an in-depth comparison of Morris and eFAST results should be the subject of another paper?*

AR-1. Yes, Morris is generally used as a screening step in large model evaluations. However, as explained in our comments to reviewer one, our rationale was other. Our rationale when including both GSA approaches was to ensure the robustness of the results. Admittedly, under common conditions, both types of analysis provide comparable results. However, methods have not been compared often in studies with complex environmental models where non-linearities can be high, although doing this allows for testing the robustness of the sensitivity indices (Pianosi et al. 2016). We should not ignore that initially Morris is a "qualitative" method since it is based on a sparse sampling (in our case with r=10 and k=18 (no WT) and 20 (with WT), N=r(k+1) =190 and 210 samples for each scenario) that could lead to inaccurate results when the model is highly non-linear or discontinuous in some region of the input factor space, compared to variance-decomposition methods like eFAST based on dense sampling (M=497, N=Mk=8946 and 9940). To increase the reliability of Morris, improved sampling techniques have been developed (e.g. Khare, Muñoz-Carpena et al., 2015) that intend to increase the robustness of the method and approximate more quantitative, comparable to those of variance methods results. This opens important opportunities for application to large models where only Morris might be feasible (e.g. see Srivastava, Graham, Muñoz-Carpena et al., 2014). In addition, we believe that the Morris plots provide an intuitive and clear way to assess the importance of the input factors and their interactions. On the other hand, the dense variance-based sampling allows for a follow up quantitative uncertainty analysis. Thus, the inclusion of the two methods in this work and the results obtained further corroborates the Morris efficiency for complex models and confirms the sensitivity of the input factors of the model.

In spite of this rationale, we agree that the inclusion of both methods shifts the main focus away from the main objective of the paper related to analysis of WT effects in the coupled processes that occur in a vegetative filter strip. We now focus on the Morris in the revised manuscript and moved the eFAST figure to Supplementary Materials. For the interested reader, we leave a brief comment in the GSA results section (with reference to Supp. Mat.) on the robustness and insights that eFAST results lend to the work.

RC2-2. p. 5, l. 130: equation 4: deltaRO is defined as the "change of cumulative excess rainfall" within a given time step. Maybe this quantity would be easier understandable if referred to as "surface runoff volume within a given time step"? In general, I think that "surface runoff" would be easier to understand than "excess rainfall" (which is a sort of ambiguous term).

AR-2. Notice that in strict hydrological terms these are not exactly the same. "Excess rainfall" is a "point value" in the landscape, where "runoff" refers to the aggregation of the excess rainfall along the slope into surface overland flow. However, we changed it to "surface runoff" as suggested for clarity.

RC2-3. p. 5, l. 140: equation 8: As pointed out by Marnik Vanclooster in his comments on the companion paper (part A), the infiltration rate f cannot be a function of the vertical coordinate z.

AR-3. Fixed in this revision.

RC2-4. p. 6, l. 149: "phytosanitary products": This term is very French. In English one would say "crop protection products".

AR-4. Changed as suggested.

RC2-5. p. 6, l. 151: Please mention the soil type according to WRB or FAO, if available.

AR-5. Table 1 provides the widely used USDA Soil Taxonomy classification.

RC2-6. p. 6, l. 154: Feel free to mention that the Jaillière site is also the basis of the regulatory scenario FOCUSsw D5 (La Jaillière)

AR-6. Good point, thanks! We added this.

RC2-7. p. 6, l. 163: "on Morcille", "on Jaillière" It should be "at", not "on"

AR-7. Corrected.

RC2-8. p. 6, l. 1746: "complementary": In what sense are the two methods complementary? (cf. above)

AR-8. See response above on AR-1 and action taken.

RC2-9. p. 6, l. 177: "a variance-based method is computed": either "variance-based measures are computed" or "a variance-based method is run/applied"

AR-9. Changed to "a variance-based method is applied"

RC2-10. p. 6, l. 190 f.: You mention only three groups. Would a fourth group (low mu and high sigma) be mathematically possible? (cf. comment by Marnik Vanclooster)*

AR-10. Agreed, fixed.

RC2-11 p. 9, l. 245: "end-vertical boundary condition": you mean bottom boundary condition?

AR-11. It is an "end" as it only applies after t>tw. Added "bottom" the first time it appears to "end vertical bottom boundary condition, hereon referred to as vertical boundary condition"

RC2-12 p. 10, l. 279: "average soluble properties": better "average sorption properties". In practice, the distribution between water and soil or water and sediment is usually governed only by the adsorption parameters (Kf and nf for Freundlich adsorption, Kd for linear sorption) and not by water solubility, since concentrations in solution usually stay way below solubility limits.

AR-12. Agreed. Changed as suggested.

RC2-13 p. 10, l. 291: "sub-saturates": What do you mean by that?

AR-13. Changed to "saturates from the bottom"

RC2-14 p. 10, l. 294: "This is exacerbated with WT": There is something missing here before WT.

AR-14. Changed to "With WT, the infiltration is limited even further in these fine soils, where…"

RC2-15 p. 11, l. 326: What is a "formal" uncertainty analysis? Can you explain briefly how you did the uncertainty analysis (probably better in the methods section)?

AR-15. "Formal" means that the variance method provides a large number of results that lend themselves to the construction of output probability distribution curves and calculation of statistics of uncertainty like median, percentiles, 95CI (Fig. 8 and Table S3). We changed "formal" to "quantitative".

RC2-16 p. 11, l. 335: "Diflufenicanil" –> diflufenican

AR-16. Corrected

RC2-17 p. 11, l. 336: "Reduction . . . higher than the other two pesticides": There is something missing here. Maybe: "higher than for the other two pesticides" or "higher than reduction of the other two pesticides"

AR-17. "higher than reduction of the other two pesticides"

RC2-18 p. 13, l. 369: "rivers drainage networks": "river drainage networks"

AR-18. Changed as suggested.

RC2-19 p. 13, l. 375 ff.: Just for information: The GERDA software package which was developed for the German EPA (UBA) as a future regulatory tool for surface water does include VFSMOD simulations with a shallow water table where present. However, while the final report of the GERDA project has been published recently (https://www.umweltbundesamt.de/publikationen/bewertung- des-eintrags-von-pflanzenschutzmitteln-in), the GERDA tool itself is not publicly available yet.

AR-19. Good point, thanks! We added this.

RC2-20 p. 18, l. 542: "UIPAC": This reference seems to be wrong. First, it should be IUPAC. Second, the link http://sitem.herts.ac.uk/aeru/iupac/index.htm is just a mirror of the original PPDB homepage (http://sitem.herts.ac.uk/aeru/ppdb/en/index.htm), and both pages and the PPDB are maintained by AERU, University of Hertfordshire, UK

AR-20. Yes, corrected.

RC-21 Figure 7: The legend needs clarification. I presume that the grey bars denote Si, the black bars (STi – Si) and the sum of both bars STi?

AR-21. Yes, we corrected the caption to reflect this. Notice that this is now Fig. S1 in Supplementary Materials following Reviewers 1 and your comments.

RC-22 Figure 8: "Uncertainty analysis results": Some further explanations needed. It seems that these are probability density functions?

AR-16. This means that the variance method provides a large number of results that lend themselves to the construction of output probability distribution curves and calculation of statistics of uncertainty like median, percentiles, 95CI (Fig. 8 and Table S3). We changed the caption to "Probability density functions from the uncertainty analysis of eFAST simulations on output variables […]"

References

Y.P. Khare, R. Muñoz-Carpena, R.W. Rooney, C.J. Martinez, A multi-criteria trajectory-based parameter sampling strategy for the screening method of elementary effects, Environmental Modelling & Software, Volume 64, February 2015, Pages 230-239, ISSN 1364-8152, https://doi.org/10.1016/j.envsoft.2014.11.013.

Pianosi, F.; Beven, K.; Freer, J.; Hall, J. W.; Rougier, J.; Stephenson, D. B. & Wagener, T. Sensitivity analysis of environmental models: A systematic review with practical workflow, In Environmental Modelling & Software, Volume 79, 2016, Pages 214-232, ISSN 1364-8152, https://doi.org/10.1016/j.envsoft.2016.02.008.

Srivastava, V., W.D. Graham, R. Muñoz-Carpena and R. Maxwell., Insights on geologic and vegetative controls over hydrologic behavior of a large complex basin – Global Sensitivity Analysis of an integrated parallel hydrologic model, Journal of Hydrology, Volume 519, Part B, 27 November 2014, Pages 2238-2257, ISSN 0022-1694, https://doi.org/10.1016/j.jhydrol.2014.10.020.

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), which has just been accepted for publication (Fox et al., 2017). We updated the conclusions to reflect this and the need for additional field work.

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 = 0.610 [0.448 - 0.943]. 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.

Fox, G., R. Muñoz-Carpena and R. Purvis. 2017. Controlled laboratory experiments and modeling of vegetative filter strips with shallow water tables. *J. of Hydrology* (in press).

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