Responses to Reviewer #4

Point #1

The present work assesses the impact of (i) reservoir management and (ii) agriculturalwater withdrawn on the hydrologic behavior of the Yakima River Basin (YRB). The Authors provide a convincing and well documented motivation supporting the present study, due to the (not so well explored) role that reservoir management and agricultural activities could have on the hydrological dynamics. I think that the paper is worth for publication after some minor revisions.

Response: Thank you for the valuable comments. We addressed each point of your suggestions and qauality of the work has been improved significantly

Point #2

Comment 1 The Authors consider 5 different scenario: R0 – the SWAT model is used to simulate the basins dynamics and neither the reservoirs management operations or the agricultural activities are included; R1 - the SWAT model is extended to consider reser- voirs management operations; R2 – the reservoirs management practice are modelled with RiverWare (which provide the flowrate downstream the each simulated reservoir according to with a set of management rules) which is then combined with SWAT; R2S2 – leverage on R2 including agricultural activities under the hypothesis that all the demand water comes from reservoirs and streams; R2S1 – is the counterpart of R2S2 in which the agricultural demand is satisfied by superficial aquifers. For each scenario (I would rather say modelling scenario/choice) there is a calibration period and a validation period. Model performance metrics are the correlation coefficient r (note that in the figure is referred as R, please change it) and the Sutcliffe efficiency coefficient ENS. The Authors found a better model performance in the following order R2S1 > R2S2 > R2 > R1 > R0, leading to the conclusion that using RiverWare to model the reservoirs management and modelling the agricultural demands as satisfied by surface bodies is the best option. How does the Authors ensure that ensuing models ordering is not influenced by the fact that a better fit between model results and data (which is at the base of r and ENS metrics) is not an artefact of model flexibility/complexity (i.e., more parameters) rather than being a 'true' more realistic model? See e.g., "A primer for model selection: the decisive role of model complexity" by Hoge et al., 2018 (https://doi.org/10.1002/2017WR021902)

Response: We think the reviewer pointed out a critical question around how to determine the robustness of a hydrologic/watershed model. Frankly speaking, we do not have the silver bullet, instead we are trying to use commonly accepted statistical metrics to measure the performance of different models, and rank those models scenarios/choices based on values of the metrics. We really appreciate that the reviewer point us to the recent publication "A primer for model selection: the decisive role of model complexity" by Hoge et al., 2018, which we believe is a milestone paper reviewing existing methods/criteria in model selection and laying out the fundamentals about the philosophy and critical aspect to be considered in model selection. We strongly agree that model complexity/flexibility should be a factor in model selection. In our case, we do not have reliable prior information about a model structure and associated parameters, making it difficult to give quantitatively factor in model complexity in model evaluation process. However, we think we should follow the conclusion drawn by Hoge et al. (2018) that "Regardless of which explicit or implicit approach is suitable and used for model selection, we want to emphasize that one should consider and report how the particular method interprets complexity and what this means for the model which is selected".

Accordingly, we added the following discussion in the manuscript to highlight the potential limitation of simply relying on one or multiple statistics to determine model performance. The model selection process deserve more robust methods or at least provide information about how model complexity is considered.

"4.4 Caveats in model selection

Among the multiple modeling scenarios, we found that linking RiverWare reservoir model with SWAT achieved better performance than those model structures that reply on simplified reservoir operations, as evidenced by relatively higher correlation coefficient and E_{ns}. However, it is worth noting that these statistical metrics are calculated based on a limited set of hydrological variables (e.g. streamflow), but cannot guarantee other hydrological processes are well represented (Zhang et al. 2013). Therefore, we further used MODIS estimated ET and reported irrigation water demand data to justify the favorable performance of the combined SWAT-RiverWare watershed model configuration.

Our model evaluation process follows the widely accepted procedures for model calibration and evaluation (Moriasi et al. 2007; Arnold et al. 2012). We also would like to point out that the complexity difference between the SWAT-RiverWare and other watershed model configurations was not explicitly considered in model evaluation. Previous research note that model complexity is an important factor in selecting the most robust model configuration that can fulfill a specific purpose. For example, Höge et al. (2018) reviewed existing methods and laid the foundation for a comprehensive framework for understanding the critical role of model complexity in model selection. The lack of reliable prior knowledge of the model structure and associated model parameters makes it difficult to directly consider model complexity here. However, the framework laid out by Höge et al. (2018) deserve further exploration in comparing the performance of different watershed model configurations in the future."

Höge, M., Wöhling, T. and Nowak, W., 2018. A primer for model selection: The decisive role of model complexity. *Water Resources Research*, *54*(3), pp.1688-1715.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D. and Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, *50*(3), pp.885-900.

Arnold, J.G., Moriasi, D.N., Gassman, P.W., Abbaspour, K.C., White, M.J., Srinivasan, R., Santhi, C., Harmel, R.D., Van Griensven, A., Van Liew, M.W. and Kannan, N., 2012. SWAT: Model use, calibration, and validation. *Transactions of the ASABE*, *55*(4), pp.1491-1508.

Zhang, X., Beeson, P., Link, R., Manowitz, D., Izaurralde, R.C., Sadeghi, A., Thomson, A.M., Sahajpal, R., Srinivasan, R. and Arnold, J.G., 2013. Efficient multi-objective calibration of a computationally intensive hydrologic model with parallel computing software in Python. *Environmental modelling & software*, *46*, pp.208-218.

Point #3

Comment 2 On top of the GSA results the Authors conclude that (line 187-189) : "In general, selected parameters demonstrated similar sensitives among all scenarios, particularly for the ten most sensitive parameters, indicating that the five scenarios captured critical processes regulating water cycling in the basin". I would disagree with the second part of the sentence. The fact that in all the 5 scenarios the top 10 most influential parameters are quite the same suggests a similarity in the behaviour/dynamics/functioning of the 5 models investigated (since for each of them the key parameters are the same), which does not ensure the fidelity with the true-world dynamics!

Response: Thank you for the comments. We agree with the reviewer that same key parameters among the scenarios does not necessarily mean that the simulations accurately represent water cycling in the real world. We improved this sentence as follows:

"In general, selected parameters demonstrated similar sensitives among all scenarios, particularly for the ten most sensitive parameters, indicating snow melting (SMFMX, SFTMP and SMTMP), soil water dynamics (CN2, SOL_k, and SOL_Z), and water routing (CH_N2 and SLSUBBSN) are critical for water cycling in the basin."

Point #4

Furthermore, this result seems to contradict in part the relevance of having different modelling strategies for the reservoirs management and/or the agricultural activities, since the 5 model results are mainly ruled by the adopted representation of common processes, like the moisture conditions (parameter CN2 is always the most important one) or snow melt process/parameters (see Line 190, with 5 parameters occupying the 2nd -6th places). As a matter of fact, the joint inspection of Fig. 3-6 does not reveal a dramatic change in the r and Ens values. It is also true that starting from the 7th position parameters associated with reservoir management and agricultural activities alternate as importance. I would appreciate a more detailed description of the GSA methodology (how it is possible to vary a parameter, and so test its sensitivity on the model outputs, related with the reservoirs, e.g., the RES_K, if the reservoirs are not included in R0? The same for agricultural activities related parameters).

Response: We agree with the reviewer that although the reservoir and irrigation related parameters play more important roles in scenarios simulating management activities, soil moisture and snow melting related parameters may still play dominant roles. We add following sentences to section:

"Note that although inclusion of management activities altered the sensitivity of reservoir and irrigation related parameter, snow melting and soil water dynamics may still play the fundamental role in water cycling, as evidenced by the high sensitivity of CN2 and SFTMP."

We added following description of the sensitivity analysis to the supplementary material:

"According to Abbaspour et al., (2017), SWAT sensitivity analysis is to find the influential parameters in the model. In the above sensitivity analysis, we held all other parameters unchanged while modifying one specific parameter to identify how this parameter affect streamflow simulation. In the sensitivity analysis, model simulations were run hundreds of times to quantify how model output is affected by changes in each parameter. Specifically, the sensitivity analysis uses a multiple regression approach to quantify sensitivity of each parameter:

$$\mathrm{Ob} = \alpha + \sum \beta_i b_i$$

In the above equation Ob is the objective function value (r-squared, E*ns* etc.), α is the regression constant, and β is the coefficient of parameters, and b refers to a specific parameter (Abbaspour et al., 2017). In the analysis, the t-test is employed to identify the relative significance of each parameter. "

Abbaspour, K., Vaghefi, S. and Srinivasan, R.: A Guideline for Successful Calibration and Uncertainty Analysis for Soil and Water Assessment: A Review of Papers from the 2016 International SWAT Conference, Water, 10(1), 6, doi:10.3390/w10010006, 2017.

Point #5

Furthermore, is the employed GSA able to provide a quantification of the global sensitivity of each parameter and not just their ranking (e.g., could be that CN2 is dominating the process and the others parameters induce just very small variations in the output)

Response: Thank you for the valuable suggestions. We added T and P values to the selected parameters in the sensitivity analysis to the supplementary material

Table S2 Parameter sensitivity under various scenarios.

Parameters	**Description**	***T and P* values for parameter sensitivity^{1,2}**				
		R0	R1	R2	R2S1	R2S2

SFTMP	Snowfall temperature (°C)	-8.06 (0.00)	-15.08 (0.00)	-12.18 (0.00)	-1.03 (0.31)	-9.32 (0.00)
CN2	Initial SCS runoff curve number for moisture condition	-10.75 (0.00)	-17.76 (0.00)	-15.25 (0.00)	-22.50 (0.00)	-11.22 (0.00)
SMFMX	Maximum melt rate for snow during year (occurs on summer solstice) (mm H_2O/°C/day)	-4.89 (0.00)	-6.72 (0.00)	-2.54 (0.01)	-0.32 (0.75)	-3.47 (0.00)
SMTMP	Snow melt base temperature (°C)	4.28 (0.00)	11.45 (0.00)	7.91 (0.00)	0.76 (0.45)	4.43 (0.00)
CH_N2	Manning's "n" value for the main channel	3.22 (0.00)	1.69 (0.09)	2.70 (0.01)	-0.73 (0.47)	1.16 (0.25)
SMFMN	Minimum melt rate for snow during the year (occurs on winter solstice) (mm H_2O/°C/day)	-1.19 (0.23)	-1.97 (0.05)	-0.10 (0.92)	-0.87 (0.38)	-0.77 (0.44)
SLSUBBSN	Average slope length (m)	5.04 (0.00)	5.54 (0.00)	3.32 (0.00)	8.65 (0.00)	4.85 (0.00)
CH_N1	Manning's "n" value for the tributary channels	-0.18 (0.86)	0.44 (0.66)	0.72 (0.47)	-0.45 (0.65)	-0.06 (0.95)
SOL_K	Saturated hydraulic conductivity (mm/hr)	-2.63 (0.01)	-1.98 (0.05)	-2.49 (0.01)	-8.57 (0.00)	-2.57 (0.01)
GW_REVAP	Groundwater "revap" coefficient	-1.21 (0.23)	-1.19 (0.23)	1.34 (0.18)	1.34 (0.18)	-0.80 (0.43)
CANMX	Maximum canopy storage (mm H_2O)	0.05 (0.96)	-0.31 (0.75)	0.69 (0.49)	-0.06 (0.95)	-0.01 (0.99)
HRU_SLP	Average slope steepness (m/m)	-0.87 (0.38)	-2.17 (0.03)	-0.25 (0.80)	-4.60 (0.00)	-0.46 (0.65)
RES_K	Hydraulic conductivity of the reservoir bottom (mm/hr)	-1.46 (0.14)	2.14 (0.03)	0.11 (0.91)	5.33 (0.00)	0.15 (0.88)
GW_DELAY	Groundwater delay (days)	-1.45 (0.15)	-0.51 (0.61)	0.71 (0.47)	-0.25 (0.81)	-1.60 (0.11)
EVRSV	Lake evaporation coefficient	-0.60 (0.55)	3.36 (0.00)	0.66 (0.51)	1.37 (0.17)	-0.49 (0.63)
TIMP	Snow pack temperature lag factor	-0.02 (0.98)	-0.04 (0.97)	-0.73 (0.46)	-0.06 (0.95)	-0.10 (0.92)
ESCO	Soil evaporation compensation coefficient	-0.11 (0.91)	-1.82 (0.07)	-0.13 (0.90)	-0.95 (0.34)	0.11 (0.91)
GWQMN	Threshold water level in the shallow aquifer for the base flow (mm)	0.26 (0.79)	0.47 (0.64)	-0.82 (0.41)	-0.89 (0.38)	0.02 (0.99)

PLAPS	Precipitation lapse rate (mm H_2O/km)	-0.33 (0.74)	-5.01 (0.00)	-2.70 (0.01)	-2.54 (0.01)	-0.89 (0.38)
OV_N	Manning's "n" value for overland flow	-2.51 (0.01)	0.42 (0.67)	0.44 (0.66)	1.53 (0.13)	-2.11 (0.04)
REVAPMN	Threshold depth of water in the shallow aquifer for "revap" to occur (mm)	0.08 (0.94)	-0.23 (0.81)	0.57 (0.57)	0.56 (0.58)	-0.03 (0.98)
SOL_AWC	Available water capacity of the soil layer (mm H_2O/mm soil)	0.00 (1.00)	-1.89 (0.06)	-0.10 (0.92)	-0.35 (0.72)	0.63 (0.53)
NDTARGR	Number of days to reach target storage from current reservoir storage	-1.27 (0.21)	0.44 (0.66)	1.48 (0.14)	2.48 (0.01)	-0.46 (0.65)
ALPHA_BF	Baseflow alpha factor (1/day)	0.37 (0.71)	-0.47 (0.64)	1.10 (0.27)	1.70 (0.09)	0.59 (0.55)
SOL_Z	Depth from soil surface to the bottom of the layer (mm)	3.89 (0.00)	2.43 (0.01)	2.40 (0.02)	4.75 (0.00)	3.87 (0.00)
TLAPS	Temperature lapse rate (°C/km)	-0.44 (0.66)	8.91 (0.00)	1.25 (0.21)	3.02 (0.00)	0.21 (0.83)
SURLAG	Surface runoff lag coefficient	-0.53 (0.60)	-0.03 (0.98)	0.11 (0.91)	0.18 (0.85)	-1.35 (0.18)
EPCO	Plant uptake compensation factor	1.56 (0.12)	1.34 (0.18)	-2.29 (0.02)	0.66 (0.51)	1.14 (0.25)

[1], format the T and P values is T (P)

[2], For *P* values less than 0.01, we use “0.00” in the above table.

Point #6

Comment 3 Caption of Fig. 2, note that should be ‘ baseline simulation does NOT consider management activities’.

Response: We corrected this mistake.

Point #7

Comment 4 Line 279. “Management schemes developed and evaluated in this study will be transferable and applicable to future SWAT and other watershed models applications for investigating water cycling that is influenced by reservoir operations and water withdrawal for irrigation across broader spatial scales” Which are the developed schemes? It was my understanding that the Authors employed SWAT and RiverWare without any modification to them.

Response: Although we did not develop new algorithms, this study combined the widely used watershed model (SWAT) with the reservoir management model (RiverWare) in investigating management impacts on streamflow simulations, and set an example for future application of the two models to better investigate hydrology in managed watershed. We agree with the reviewer that it is necessary to make the statement more specific and accurate:

"Management schemes **employed** and evaluated in this study will be transferable and applicable to future SWAT and other watershed models applications for investigating water cycling that is influenced by reservoir operations and water withdrawal for irrigation across broader spatial scales."

Point #8

Comment 5 Eq. (1): Vnet seems to me like the water volume after one day, being Vstored the water stored at the beginning of the day. Please clarify that Vflowout is the focus of the diverse management schemes. Eq. (4): Vflowout has been already used to indicate a volume in Eq. (1), please modify the notation in order to avoid confusion.

Response: Thank you for the valuable suggestions. We use

$V_{swat\_flowout}$ and $V_{RiverWare\_flowout}$ in equations 4 and 5 to avoid confusion with variables in equation 1

Point #9

Comment 6 Line 207: delete water after release.

Response: We deleted the second 'water'.

Point #10

Comment 7 Line 321-323: "The irrigation operation scheme that used surface water as the single source may have introduced uncertainties to streamflow simulations, since ground water is also an important water source for irrigation, particularly in dry years in the YRB". The choice of using surface water as the single source for the irrigation surely introduce (increasing, by e.g., lack of knowledge on some parameter, or reducing, e.g., by inserting salient dynamics in the model) uncertainties to streamflow, but I think that in the context of the sentence this choice has to be seen as a 'bias' to the streamflow simulations, i.e., streamflow are biased by having choice to consider only surface water bodies.

Response: Thank you for the valuable suggestions and we improved this sentence as follows:

"Streamflow simulations are biased by considering only surface water bodies since groundwater is also an important water source for irrigation, particularly in dry years in the YRB"

Point #11

Caption of Fig. 6: is it R2S1 based on SWAT or RiverWare?

Response: It is from RiverWare. We corrected this mistake as follows:

"Figure 6. Calibration and validation results under the R2S1 scenario (RiverWare for reservoir operation and surface water as the water source for irrigation)"