Authors' response to Anonymous Referee #2

A. General comments

RC: The manuscript investigates sensitivity of the SWAT model in Wala catchment in Jordan. The main idea of the article is to develop a framework to test the effects of various data sets in hydrological models to support water management and planning in data scarce regions. The results also support to identify gaps that need to be filled by e.g. improved monitoring. For this purpose the authors tested extends of errors in predictions due to the use of different types of input data to SWAT. They developed eighteen hydrological models (using combination of three land use maps, two soil maps, and three climate time series from the local and global sources), and they evaluated the models using measured monthly discharge and constructed sediment yield data at the outlet of the watershed. The authors showed that significant performance gain can be obtained with the proper combination of inputs. They conclude that selection of quality data will reduce uncertainty of hydrological model outputs. The subject falls within the general scope of the journal. The aims of the study are interesting for the readers of the HESS. The obtained results appear encouraging. It is however strongly advisable to extend the following points in the paper:

1. This study tests relative quality of the existing datasets from local/global sources to support the statement in "Page 2, L22-29" on reducing uncertainty in data poor regions when transferring parameters/knowledge from neighboring or geographically similar catchments. While the general idea is very interesting, but I felt the main hypothesis never tested. I think the authors can improve novelty of the work by quantifying how the traditional knowledge (parameter) transfer from the neighboring watersheds (e.g., using calibrated parameters of poor model-scenario in this study) versus parameter transfer from a better model (using quality data model-scenario in the study) help reduce uncertainty in model predictions of the data poor regions. One would expect that this could be done by classical sample test in small portion of the watershed.

AC: The authors would like to express thankfulness to Referee #2 for the thorough review, thoughtful suggestions and raising critical points for discussion. It is a pleasure that this valued review finds the aims of our study interesting and the results encouraging. We have taken all comments and suggestions of Referee #2 on board and prepared the following point-by-point response with applying the required modifications across the manuscript. We hope our response clears up the Referee's concern and strengthens our work.

A modified version of the manuscript is uploaded (a track-changes copy is also provided to facilitate finding edits).

1. We appreciate the Referee's recommendation and apologize for the confusion that made the statement a bit unclear to readers. We would like to clarify that the statement in this paragraph, which states our driving research questions, is in the last two lines: 'which datasets should be employed in modeling and where should investment be targeted to improve data quality?', which we discussed later, for example by recommending improving soil data for their high importance. The lines before that compare between geomorphically similar (e.g. humid) and significantly different (e.g. semi-arid) areas and show that transfer of data/parameters is likely to help in humid areas, but not necessarily applicable in dry lands and this is why we suggest our approach to select suitable data to model dry lands and invest

on improving the most sensitive datasets rather than depending on transfer of parameters which is expected to bring uncertainty.
We do not claim that our approach reduces uncertainty when parameters are transferred from neighbouring areas; rather we encourage testing data available for the area itself and recommend improvement/investment where possible.
We do not currently have parameters from neighboring areas to compare with our poor and good scenarios but this would be a good test to plan for future work. Our main hypothesis is literally stated in the last paragraph of the section.
Please see the modified text; we hope it makes clear our objectives and hypothesis.

A. Specific comments

1. **RC:** Authors may provide more background about previous SWAT applications in the study region (if published in any peer reviewed journals or reports that are available for public); and also strength and limitation of the model in behavior simulation of the major hydrological events in such arid environment with intense, highly intermittent, and often localized storms. Authors may add this in the "model selection" and "discussion" parts. Authors emphasized on the importance of input data uncertainty but never discussed other sources of errors in hydrological modeling: e.g., model structure (process simplifications, which might be case in this work), and parameter estimations.
   AC: Please see the modified section 3.2 (Model selection and structure) for improved background about SWAT and 1 (Introduction) for further details about uncertainty.

2. **RC:** Page 6, L23-24: How the two weather stations (Qatraneh and Errabbah) represent climate conditions in the study area? If the stations are not representing actual conditions, the generated data (to fill the gaps in the recorded time series) will be subjective, and as a result poor hydrological performance will be obtained (it is seen also from the results).
   AC: It is common in hydrological modelling (and specifically using SWAT) to use data from nearby stations (Fuka et al., 2014) if they are reasonably close to the catchment because SWAT uses geographical weightage/interpolation when using data from nearby stations and each subbasin is linked to the station closest to its centroid. However, we agree with the Referee that this can be a source of uncertainty but in such data-challenging environment, these two stations were the closest with complete datasets that the model requires to run. This potential uncertainty emphasises the need to improve data monitoring (and accessibility) in these areas for better studies. The text is edited to point to this important note.

3. **RC:** Page 6, Section 3.5: How the HRUs were defined? Please indicate if you used dominant or multiple (/threshold?).
   AC: Section 3.5 is edited and threshold criteria defined.

4. **RC:** Page 7, Section 3.7: In this study the sediment yields are not measured but estimated using the streamflow data. This cause a subjective comparison results when testing different scenarios in this study: (i) It is obvious from the results that model scenarios that perform better in the simulation of streamflow, present higher performance in modeling sediment too. Therefore, any judgment on the performance of the input data in model simulation will be subjective. (ii) The inherent errors in the estimated sediment yields may be compromised/offset by the model prediction errors due to less quality input data, resulting in a wrong conclusion in scenario selection. The authors may provide a background on how the

sediment data were estimated and how the above mentioned points may be justified when evaluating model-scenarios in this work.
AC Please see the edits in section 3.7 for details of how sediment data were constructed. The equation provided was accredited and used to design the dam and manage the catchment. We hope it represents the conditions in the area (furthermore, it is the best available source for sediment data). However, we agree that there is dependency of the constructed sediment data on discharge data they correspond to, this may justify the correlation between model performance in simulating discharge and sediment.
The inputs used to simulate sediments are completely independent from the observation-constructed sediment data; therefore, we think that the goodness of fit of sediment simulation is properly assessed by comparing two independent series of data (sim & obs). Quantitatively, the better statistics of sediment simulation could be a result of the nature and magnitude of sediment events compared to those of discharge and not because of the explained dependency. Given the complication of sediment simulation and data scarcity, we hope this reveals the Referee's concern and sheds light on the challenges facing data-poor areas.

5. **RC:** Page 7, L 29-31: The strong relationship between Q and T should be explained. The provided reference is not enough and not available for readers.
   AC: Appreciating the importance of this note, the equation used to construct sediment data is added (please see section 3.7). Howard Humphreys and Partners (1992) is a well-recognised consultant study undertaken for the purposes of designing the Wala dam and we are afraid it is beyond the scope of the current study to present the finer details of that study.

6. **RC:** There are many studies suggesting multi gauge evaluation (rather than single outlet comparison) in spatially explicit hydrological modeling, to prevent the spatial errors in the upstream catchments that may offset and not observed in the outlet. If the gauge data are not available in other tributaries of this study to conduct multi-gauge evaluation, at least the importance of this effort should be highlighted to also support the statement on Page 3, L3-5, on the investment efforts to improve the limited data.
   AC: We totally agree with this and recommend improving field measurements to provide trustworthy observed data. Unfortunately, we could hardly obtain observed data for the outlet (which is of high importance for being a dam location) and could not get hold of any further data (if any), otherwise multi-gauge evaluation would have been undertaken. This key recommendation is added to the conclusion (last two lines).

7. **RC:** Authors may add three other layers of data into Fig. 5: (i) river network; (ii) stream- flow station (Wadi Wala) where the models were evaluated; (iii) Wala dam. With the geographic coordinates only, it is hard to understand the system.
   AC: Please see Fig. 2 for better illustration of the system and magnified location of the dam.

8. **RC:** Page 7, L27-28: The authors evaluated their model-scenarios using the discharge data of Wadi Wala station that is located downstream of the dam location. How the operation of dam and the effects on hydrological regime in downstream station was considered? Authors may explain if the operation of dam was simulated in the model.
   AC: True, but the dam was put in operation and impoundment started after 2002 and the measurements used were before that. Currently, there are discharge measurements

upstream the dam and managed water release through the dam tunnels to downstream areas (these will be used for future uses of the calibrated model). Please see the edited text for clarity.

9. **RC:** AC: Page 7, L32: The "parameterised" SWAT model is not clear. Does it mean that authors make change in model parameters (for any calibration purposes?). If so, it contradicts with several statements in the text that that the evaluations were performed prior to calibration.
   AC: Apologies, the term "parameterized" is removed to clear up confusion. We mean the model with its default parameters which are extracted from its original inputs.

10. **RC:** Page 7, L32: SWAT model operates on "daily" not monthly or yearly or seasonal. Please rephrase the sentence. Authors may aggregated data from daily to monthly etc.
    AC : True, we have averaged daily data to produce monthly records as SWAT gives the option of simulating loadings on a daily, monthly or yearly basis. Please see the link for definitions of SWAT inputs (http://swat.tamu.edu/media/69392/ch31_input_meas.pdf).

11. **RC:** Page 9, L7-8: "NSE drops ... and all cases". These are interesting results. Authors may explain why CFSR data performed better than local observations. Are the findings consistent with other studies around the world that applied CFSR in hydrological models?
    AC: Yes, similar results were found by several studies. Please see the added references recommending using the CFSR over local records (modified sections 4.1 and 4.2.2).

12. **RC:** Page 9, L11-13: usually a better performance is expected when using locally produced high resolution maps (e.g., the one from Al Bakri). Authors may explain the possible reasons why the course resolution map of global source performed better in this study.
    AC:
    - Section 3.4.1 describes briefly the similarity between the three landuse maps in showing two dominant types of vegetation and minor coverage by other landuse classes.
    - Section 4.2.3 (1st paragraph) states that the performance of the best three scenarios (with the only difference being landuse) is almost equal (just slight differences found).
    - Section 4.2.3 (2nd paragraph) explains the relatively little variation in spatial distribution and range of physical characteristics among the three landuse maps and the close range of CN values (ranging from 80 to 84). In addition, it is suggested that "the method of HRU definition within SWAT selects the major land-use types in each HRU, thus potentially nullifying the gains of higher-resolution land-use maps with numerous smaller land-use classes", which means that the gain of having more minor details in Al-Bakri's map is not obvious in this specific case and the landuse types suggested by the global map may provide slightly more accurate representation of the actual landuse.

13. **RC:** Page 10, section Weather data: Please refer to my previous comment in this section (comment #2).
    AC: Please see response to comment #2 and the edited text in sections 4.2.2 and 4.2.4.

14. **RC:** Page 10, L12-16: (i) When looking at the top left graph in Fig. 9, the high performing statistics are usually representing low flows, while high flows are almost completely underestimated (or not predicted at all). How the authors will consider both low flows (important for drought) and high flows (major events and important for flooding and water

saving) explain in their model evaluation and scenario selection? (ii) As indicated in several places in the text, the Wala basin is located in semi-arid area that is prone to intense (sub-daily) rainfall events. My concern is that how model evaluation at the monthly scale will ensure representation of locally important short-term events? (iii) As mentioned previously, the data are compared at downstream station that might be affected by operation of a dam.
AC: i) We agree with the Referee's concern but as the statistics are unable to differentiate between low and high flows, the observed good relationship for scenario 16 with the observed at high flows is encouraging, whilst recognising that the simulated at times struggles to determine smaller events. Currently, we hope this fulfils the aim of this study and the next step will be an in-depth calibration to deal with low/high flows and preferentially alter their simulated values to match observed data. We noted an error in Figure 9 and have presented the correct figure. (ii) We appreciate that ideally the use of daily information as both input and output would be preferable. Unfortunately, attempting to run the analysis at the daily temporal scale failed to achieve satisfactory results, this we feel could be the result of several factors (e.g. catchment lag, partial precipitation coverage of the catchment, data quality issues). The use of the monthly provided a more satisfactory output, we feel that this is the result of the temporally short (sub)daily precipitation being averaged over the month, removing the ephemeral nature of precipitation and subsequent river flows and sediment production. The original discharge measurements were daily and their monthly averages were calculated to match the simulation time interval (iii) Please see our response to comment (8): the dam was put into operation and impoundment started after 2002 and the measurements used were before that.

15. **RC:** Page 11, L13-22: Another reason is the use of statistics generated from inappropriate weather stations: please refer to my comment in this section (comment#2)
AC: This is a possibility and we are surely taking it into consideration. To check that, we tried to review the statistics generated for these stations and found them within expected ranges for the area. Please see response to comment 2 and the modified text in sections 4.2.2 and 4.2.4.

16. **RC:** Page 11, section 4.3: It would be interesting to apply the calibrated parameters of the model-scenarios to support the main idea of the study on how appropriate model setup and quality data help managers of data scarce watersheds when transferring knowledge (e.g., parameters) from neighboring watersheds. Please refer to my general comment.
AC: Please see our response to the General Comment. Apologies again for the confusion, the main idea is to reduce input uncertainty by using the best available datasets and invest on improving the sensitive ones (if required). However, the suggestion is interesting and it is already planned to use the calibrated model to support decision making in the area, for example the feasibility of raising the Wala dam which is being under investigation by the Jordanian Government currently. The calibrated model will be also used to suggest land management scenarios in the area but this is beyond the scope of this paper. Please also see the last two paragraphs of the conclusion which support the above.

C. Technical corrections

1. **RC:** Authors may keep the acronyms consistent: either provide full definition for all of the NSE, RSR, and PBIANS; with the acronyms in the bracket; or provide acronyms only.

AC: 'Nash-Sutcliffe Efficiency' removed in the abstract and added in the text. All consistent now (acronyms in the abstract and full definitions in section 3.7).

2. **RC:** Page 3, L3: "..quality and/or quality" : "..quality and/or quantity".
   AC: Corrected.

3. **RC:** Page 2, L6: "..with extremely . . . and" is repetition of previous statement. Either remove this sentence or rephrase.
   AC: Repeated statement removed. (Please note, the statement was in P3 not P2)

4. **RC:** Page 4, L16: "basin characteristics" are NOT represented by the USDA CN method, but the "surface runoff' is simulated.
   AC: Sentence modified.

5. **RC:** Page 24, Fig. 5: for the ease of visual comparison, I suggest using similar coloring pattern for the same LU classes in the three maps.
   AC: New figure with similar coloring pattern inserted. (it is Fig. 3)

6. **RC:** Page 6, section 3.4.3: the authors may move this short paragraph to section 3.3, where you first discuss the DEM in the study area.
   AC: Paragraph moved as suggested.

7. **RC:** Page 7, L23: "Running SWAT . . .for full description": statement does not fit in this section. I suggest to move it to the section 3.2, where the authors introduce SWAT model, or to remove it.
   AC: Statement moved as suggested.

8. **RC:** Page 30, Fig. 9: Titles and legends are too small. Only scenario 16, and scenario 2 are presented in the figure, while the caption indicates scenarios 13, 18, 5, and 3 as well as 16, and 2. Please carefully check the figures, captions, and discussion in the text to match these three parts.
   AC: Apologies, a wrong figure was inserted, which is corrected now. It matches the text and captions.

9. **RC:** Page 30, Fig. 9: the explanation of the sediment yield graphs are missing from the caption and text.
   AC: Figure is corrected. Only discharge is meant to be displayed in Fig. 9.

10. **RC:** Page 31, Fig. 10: please provide high quality graphs with consistent size and formatting, and consistent scale in the vertical axis to help comparison of the graphs. With the current format it is hard to read and compare the graphs.
    AC: We apologize for the inconvenience but this was felt to be the best way of presenting the images as a previous attempt to standardise the y-axis for all four scenarios made the figure difficult to read (and split into two pages), hence we kept the x-axis consistent and the background lines at a consistent interval in all figures (0.2) and we kindly ask the readers to consider the different scale of scenario 2 from the first 3 scenarios, which are presented on

the same y-axis scale). We hope the comments of Referee #1 on the quality and illustrations support our response.

References

Fuka, D. R., Walter, M. T., MacAlister, C., Degaetano, A. T., Steenhuis, T. S., and Easton, Z. M.: Using the Climate Forecast System Reanalysis as weather input data for watershed models, Hydrological Processes, 28, 5613-5623, 2014.
Howard Humphreys and Partners: Dams on Wadi Wala and Mujib, Jordan Valley Authority of Jordan, AmmanFinal report, 1992.