UNIVERSITY OF LIVERPOOL

Ms Esra'a Tarawneh

School of Engineering
The Quadrangle
The University of Liverpool
Brownlow Hill
Liverpool L69 3GH

E erat@liverpool.ac.uk

www.liv.ac.uk

30th August 2016

Dear Editor,

Re. Resubmission of manuscript HESS-2016-242

We are pleased to resubmit for publication the revised version of manuscript HESS-2016-242 "*A pre-calibration approach to select optimum inputs for hydrological models in data-scarce regions.*" We appreciate the positive opinion and constructive criticisms of the Editor and the Referees.

Please find enclosed with this letter a revised manuscript and associated documents (Authors' response to Referees and marker-up manuscript (later in this document) and a supplementary short video of the study area) in respect of our resubmission of the above paper.

We have taken the opportunity to carefully review and revise the manuscript in accordance with the detailed and thoughtful criticisms and suggestions received. We hope that our responses to all the general, specific and technical reviews, both in the point-by-point commentary and on the evidence of the revised manuscript itself, satisfy you and the Referees that the concerns have been addressed resulting in a paper which will be acceptable for publication.

Our justifications for publication in *Hydrology and Earth System Sciences,* assertions of significance and confirmations in respect of authorship and prior publication remain as per our original submission dated 23rd May 2016.

Yours sincerely,

Ms Esra'a Tarawneh

on behalf of co-authors Dr Jonathan Bridge and Dr Neil Macdonald.

Authors' Response to the Review Comments

Editor (Dr Elena Toth) – Received 11ᵗʰ August 2016

1. **EC:** The paper presents a comparison of the rainfall-runoff modelling results obtained with different input and watershed characterization, applied to a data-scarce region, well representative of many parts of the worlds, and therefore may certainly be of interest for a wide hydrologists' audience. The two referees have a positive opinion of the paper and I warmly invite the Authors to submit the revised version they are already working on, taking into account all the comments made by the Referees.
   AC: The Authors would like to thank the Editor for handling the review process and sharing a positive opinion of the paper and constructive comments to improve it. We appreciate the invitation to submit the revised manuscript and hope the improvements we have made by carefully following the review instructions properly address the concerns of the Editor and the Referees.

2. **EC:** I will ask the opinion of Ref 2 on the revised version, while I believe that the Authors have properly addressed the points raised by Ref 1, but the use of monthly instead of daily data (see comments by Ref 1 and the Authors' reply) should be better clarified in the text (it is not explained in the revised version) and consider adding in the paper the results got at daily scale.
   AC: We appreciate the importance of this point and hope our response addresses it. Please see the addition at the end of Section 3 and our revised response to specific comment #5 of Referee #1 below.

3. **EC:** On the comparison of rainfall input: the authors themselves acknowledge in the discussion of the results, the use of a rainfall generator is questionable for filling time-series (and even more so in this case, where the two stations are far from the catchment and their data are not even included in the daily local set). But since the performances are very poor also when excluding the only station where the weather generator is massively used, there are certainly problems in the ability of the remaining 3 raingages (where there almost no missing data, so the weather generator is not used) to represent the precipitation field over the catchment, probably due also to the lack of gauges in the eastern part of the study area; the authors should consider further comments on such issue (some comments are actually already included in the revised version).
   AC: We fully agree with this point and have tried to clarify it in the revised text. Please see the modified sections 3.6 and 4.1.

Anonymous Referee #1 - Received and published 27th June 2016

A. General comments

RC: The authors addressed a major issue of dealing with scarce data having different sources as well temporal and spatial scales for hydrological modelling purposes. The methodology was very clear and the obtained results are of high importance particularly for hydrologists and soil and water conservation specialists working in dry environments. The paper was very well written and illustrated. The bibliography is complete and relatively up to dated.

AC: The authors would like to express appreciation to Referee #1 for the time and effort dedicated to review the manuscript and provide in-depth comments, important suggestions and accurate technical corrections, which have obviously improved the quality of the manuscript. We would also like to thank the Referee for finding the potential of the presented work to assist areas most in need for such approaches and for positively commenting on the results, methodology, structure and technical status of the manuscript. We wish the point-by-point response below addresses the Referee's concerns and improves the manuscript. A modified version of the manuscript will be uploaded

B. Specific comments

1. **RC:** It is recommended to further justify the choose of the SWAT model (P4) with case applications particularly in similar environments:
   AC: Section (3.2 Model selection and structure) is modified to address this important suggestion.

2. **RC:** For the maps (p5): Did you consider that there were no major changes during the simulation period?
   AC: Yes, from the authors' knowledge of the relatively slow land use change in the catchment supported by the available information/maps, it can be assumed that land use change is negligible for this study. Other studies have found that main changes have happened by 1978 (our study starts from 1979). Please see the modified section (3.4.1 Land use).

3. **RC:** A part from the Wala dam, are there other hydraulic structures (soil and water conservation, water harvesting, etc.) in the watershed. If yes, how did you represent them in the model?
   AC: Not up to the authors' knowledge or published work and if any, they can be just small farm ponds/traps. The biggest in the area is the Wala dam with a reservoir surface area of 0.8 Km^2 (compared to ~ 2 000 Km^2 catchment) and it was constructed at the outlet in 2002 (hence, not represented in the model). A note added within Section 3.7 (Scenario comparison) regarding the beginning of impoundment. Please find the supplementary short video of the Wala Catchment for a better display and understanding of the study area.

4. **RC:** Provide a summary of the used soil characteristics for the two soil maps (p5-6) and indicate properly the measured and the estimated ones.
   AC: Comment addressed and section (3.4.2 Soil) is modified accordingly.

5. **RC:** Using average monthly discharge data (p7) in a dry environment needs to be well explained and justified. In fact, in these areas, flood events occur generally in most of the cases in very short periods (some hours). Therefore, even using daily averages may cause some problems with model calibration and validation!

AC: 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 model output time interval. Please see the edits in paragraph 2 of section 3.7 and the text below as inserted at the end of the section:

"Whilst input data (climatic) are based at a daily temporal scale, the model outputs are considered at a monthly timescale for several reasons, i) daily observations of discharge and sediment are unavailable at the Wala station for the whole period of study, with only monthly observations available for model evaluation; ii) A shorter period (1990 – 1996) of daily observations are available at Wala station, but using these yields poor correlations <0.1 between daily model-simulated and observed discharge, iii) with incomplete/low quality measurements, potential for lag within the pairs of daily simulated and observed values (for model statistical evaluation) can present challenges, which can be reduced when using aggregated temporal data; and most importantly, iv) the objective of this study is to determine long-term flux within the catchment, avoiding the complexity presented by ephemeral systems and since the monthly comparison achieves reasonable fit between observed and simulated values, it is considered sufficient for evaluation of the current model with more convenience. However, all calculations of the model occur on a daily time step, which ensures that hydrological events are accounted for separately as they occur each day.

Similar approach is adopted in several studies, particularly using SWAT, in both humid and arid regions. Spruill et al. (2000) evaluate daily and monthly SWAT models simulation for a small watershed in central Kentucky and state that SWAT is an efficient tool for monthly runoff simulation with NSE values of 0.58 – 0.89 compared to - 0.04 – 0.19 for daily runoff simulation during the same period. The reason suggested is that the model poorly detects peak flows and recession rates while it performs better with total monthly values. For reasonable performance of SWAT, Huang and Zhang (2004) select to simulate discharge in a semi-arid catchment in China on a monthly basis, which leads to NSE of 0.88. The difference between daily and monthly simulations is investigated in watersheds of different scales by Heathman and Larose (2007) and the results show that simulating higher discharge rates, which is usually associated with larger watersheds, introduces greater uncertainty in SWAT discharge estimates and the study states that very good model performance is achieved for monthly stream-flow estimation while the outputs of daily simulation are only within acceptable range."

6. **RC:** What do you mean by "Howard ´ Humphreys and Partners (1992) identify a strong log linear relationship . . .. after Tarawneh (2007)" ? (p7). May I understand that the sediment yield was estimated based on this relationship?

   AC: Yes, there are no direct measurements for the study period but the available sediment measurements (for a different period) are linked to discharge by that relationship for the purposes of designing the dam and managing the catchment. Hence, measured discharge is projected to that linear relationship to derive relevant sediment values (as the best available source of data). Equation added and text edited for clarification (section 3.7 Scenario comparisons).

7. **RC:** In Figures 3 (P24) and 4 (P25): are all these classes exist in the study watershed?

   AC: Yes, according to the cited land use maps and general knowledge of the catchment supported by a field visit by the authors in 2013.

C. Technical comments

1. **RC:** P5 L10: Replace 'Luzio et al., 2002' by 'Di Luzio et al., 2002' (it is the same author):
   AC: Done.

2. **RC:** P5 L26: Check if it is Leon, 2007 or Leon, 2013 :
   AC: It is actually 2011 (Version 2). Edited in text and bibliography.

3. **RC:** P6 L23: Replace 'by (Neithsch et al., 2001)' by 'by Neithsch et al. (2001)'
   AC: Done.

4. **RC:** P7 L10: Replace 'see for example Zhang' by 'see for example Zhang'
   AC: Done

5. **RC:** Ageena et al (2014): not found in the text
   AC: EndNote error: Author name was missing and only years appear. Reference added properly.

6. **RC:** Ageena et al. (2013): not found in the text.
   AC: EndNote error: Author name was missing and only years appear. Reference added properly.

7. **RC:** P16 L36: Check if the reference of Montheith is complete?
   AC: Reference completed.

8. **RC:** P17 L22: Check if you need to type twice 2009
   AC: Just once – edited.

9. **RC:** P17 L34: Check if you need to type twice 2008b
   AC: Just once – edited.

10. **RC:** P27: Correct 2000-200 !!
    AC: Corrected to 2003.

11. **RC:** Figure 9 (P30): Check if the scenarios 13, 18, 5 and 3 are included in the graphs!
    AC: Apologies, wrong figure was inserted – corrected now including the scenarios mentioned.

Anonymous Referee #2 - Received and published 7th July 2016

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 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.

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.

B. 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. Please see the modification to section 4.1 for a comment on the representativeness of stations.

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. The Referee is also invited to watch the supplementary video of the Wala catchment for improved display and understanding of the system.

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, SWAT considers daily basis but gives the option to produce monthly or yearly estimates. Please see the link below 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 land use maps in showing two dominant types of vegetation and minor coverage by other land use classes.
- Section 4.2.3 (1st paragraph) states that the performance of the best three scenarios (with the only difference being land use) 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 land use 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 land use types suggested by the global map may provide slightly more accurate representation of the actual land use.

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. Please see the edits in paragraph 2 of section 3.7 and the text below as inserted at the end of the section:

"Whilst input data (climatic) are based at a daily temporal scale, the model outputs are considered at a monthly timescale for several reasons, i) daily observations of discharge and sediment are unavailable at the Wala station for the whole period of study, with only monthly observations available for model evaluation; ii) A shorter period (1990 – 1996) of daily observations are available at Wala station, but using these yields poor correlations <0.1 between daily model-simulated and observed discharge, iii) with incomplete/low quality measurements, potential for lag within the pairs of daily simulated and observed values (for model statistical evaluation) can present challenges, which can be reduced when using aggregated temporal data; and most importantly, iv) the objective of this study is to determine long-term flux within the catchment, avoiding the complexity presented by ephemeral systems and since the monthly comparison achieves reasonable fit between observed and simulated values, it is considered sufficient for evaluation of the current model with more convenience. However, all calculations of the model occur on a daily time step, which ensures that hydrological events are accounted for separately as they occur each day.

Similar approach is adopted in several studies, particularly using SWAT, in both humid and arid regions. Spruill et al. (2000) evaluate daily and monthly SWAT models simulation for a small watershed in central Kentucky and state that SWAT is an efficient tool for monthly runoff simulation with NSE values of 0.58 – 0.89 compared to - 0.04 – 0.19 for daily runoff simulation during the same period. The reason suggested is that the model poorly detects peak flows and recession rates while it performs better with total monthly values. For reasonable performance of SWAT, Huang and Zhang (2004) select to simulate discharge in a semi-arid catchment in China on a monthly basis, which leads to NSE of 0.88. The difference between daily and monthly simulations is investigated in watersheds of different scales by Heathman and Larose (2007) and the results show that simulating higher discharge rates, which is usually associated with larger watersheds, introduces greater uncertainty in SWAT discharge estimates and the study states that very good

model performance is achieved for monthly stream-flow estimation while the outputs of daily simulation are only within acceptable range."

(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). However, if that still causes confusion, would you please consider the revised Figures 9 and 10 with all selected scenarios displayed with a consistent 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.

A pre-calibration approach to select optimum inputs for hydrological models in data-scarce regions

Esraa Tarawneh[1], Jonathan Bridge[1] and Neil Macdonald[2]

[1]Centre for Engineering Sustainability, School of Engineering, University of Liverpool, Liverpool, L69 3GH, UK

[2] School of Environmental Sciences, University of Liverpool, Liverpool, L69 7ZT, UK

Correspondence to: Esraa Tarawneh (e.tarawneh@liverpool.ac.uk)

Abstract. This study uses the Soil and Water Assessment Tool (SWAT) model to quantitatively compare available input datasets in a data-poor dryland environment (Wala catchment, Jordan; 1 743 km^2). Eighteen scenarios combining best available land-use, soil and weather datasets (1979 - 2002) are considered to construct SWAT models. Data include local observations and global reanalysis data products. Uncalibrated model outputs assess the variability in model performance derived from input data sources only. Model performance against discharge and sediment load data are compared using r^2, ~~Nash Sutcliffe Efficiency (~~NSE~~)~~, RSR and PBIAS. NSE statistic varies from 0.56 to -12 and 0.79 to -85 for best and poorest-performing scenarios against observed discharge and sediment data respectively. Global weather inputs yield considerable improvements on discontinuous local datasets, whilst local soil inputs perform considerably better than global-scale mapping. The methodology provides a rapid, transparent and transferable approach to aid selection of the most robust suite of input data.

Keywords: soil; land-use; weather; sediment; sensitivity; SWAT

1 INTRODUCTION

Arid and semi-arid regions of the world suffer from water scarcity exacerbated by growing populations, increasing per capita water consumption and agricultural intensification. Depletion of surface water and over-abstraction of non-renewable groundwater adversely impact ecosystems and human quality of life (Wheater et al., 2008b). Effective water management is crucial and relevant decision making can be assisted by approximating the complex hydrologic systems of arid and semi-arid regions through modelling. This enables scenario-testing and forecasting to inform decision-making in water and land management (Tessema, 2011; Wheater et al., 2008a).

The ability of a model to successfully predict catchment behaviour relies on the reliability and representativeness of the data against which it is calibrated, the quality of the processes and parameters assumed internally within the model and the accuracy of the input datasets used to define the catchment (Griensven and Meixner, 2006). Unfortunately in many cases, the regions most in need of reliable hydrological models are those with limited economic resources and fragmented environmental monitoring infrastructure (Ragab and Prudhomme, 2002). Data available to underpin models may, therefore, vary significantly, both in quality and quantity (Pilgrim et al., 1988). This encourages the use of modelling 'rules of thumb' or estimations based on spatially or temporally aggregated data for the modelled area, or data obtained from comparable better-studied regions (Gee and Hillel, 1988; Nyong et al., 2007; Tingsanchali and Gautam, 2000).

~~Imperfect fit of mo~~del results to measured data is called modelling uncertainty while predictive uncertainty measures predictability of the model when used for future scenarios ~~Predictive and modelling uncertainty (the range of error in the calibrated model output about the observed calibration dataset) can both be reduced by effective calibration~~ (Krupnick et al., 2006). Structure of hydrologic models can lead to uncertainty issues, particularly when assumptions are inherent within the model design~~,~~. ~~h~~However, choices available to semi-arid regions are still limited and reducing associated uncertainty requires intensive research to improve incorporated mathematical models and their ability of representing physical processes and extract information from available data. Uncertainty related to measurements used for model assessment can only be reduced

by improving observation techniques and networks (Griensven and Meixner, 2006). This study focuses on modelling uncertainty and particularly on inputs as one of its main sources in models of data-sparse semi-arid regions. Minimizing inputs uncertainty is an important aspect of the planning process for modelling projects, it ensures that input data and parameters are more accurate and suitable, reduces predictive uncertainty and assists decision-making (Liu and Gupta, 2007; USEPA, 2002). ~HoweverFurthermore, in the absence of high quality 'ground truth' data for soils, land use and weather inputs, powerful automated calibration algorithms can alter model parameters to produce a structurally biased model (Kalantari et al., 2015) which provides a good fit to specified calibration data, but may diverge significantly from true catchment behaviour under other conditions (Beven, 2011).

The relationship between model inputs and performance are investigated at a range of scales in different hydrologic settings (Beeson et al., 2014; Chaplot, 2013; Lobligeois et al., 2013; Lobligeois et al., 2014; Müller Schmied et al., 2014). For example, Legesse et al. (2003) use distributed Precipitation-Runoff Modelling (PRM) to investigate the impact of climatic and land-use variations on hydrologic response in data-scarce Tropical Africa. Di Luzio et al. (2005) determine that Digital Elevation Model (DEM) construction is critical to stream flow and sediment predictions of a SWAT (Arnold et al., 1998) model for a 21.3 km^2 watershed in the Mississippi, with a significant effect of land-use and limited influence of soil data. Liong et al. (2013) present SWAT model results for a catchment in Southeast Asia and conclude that the highest uncertainty results from applying global climate models for regional and localized applications, recommending the use of higher spatial resolution regional data. Recently, Faramarzi et al. (2015) show in a SWAT analysis of Alberta, Canada that choice of optimal input datasets significantly affects the overall model performance by reducing unnecessary and arbitrary adjustment of parameters to compensate for structural errors in the model. Crucially, better model performance is not necessarily correlated with accumulation of a mass of data; rather it depends on data reliability and relevance (Tessema, 2011).

In settings where input and output datasets are robust and comprehensive, such as in humid areas, this issue may present rarely or be mitigated by transfer of knowledge from neighbouring or geomorphically similar catchments. By contrast, in semi-arid regions, for example, where data coverage and quality are historically poor (Edmunds et al., 2013) and hydrological systems operate under significantly different conditions from those in well-monitored temperate environments (Chehbouni et al., 2008), such transferability of parameters may not be the best option and can be is itself a source of uncertainty (Wheater et al., 2008a). Therefore, developing methodologies that target optimizing data and parameters of the area itself, in addition to This can only be reduced by improving observation techniques and networks are more efficient ways to reduce uncertainty (Griensven and Meixner, 2006). When the relative integrity of available datasets is unknown and research resources are limited, the questions arise: which dataset(s) should be employed in modelling, and where should investment be targeted to improve data quality?

This study explores a methodology for differentiating between various input datasets of unknown relative quality for a hydrological model of a typical semi-arid catchment in Jordan. We start with the proposition that the combination of input datasets which produce the best fit to observed output data prior to full model calibration will yield a model that is less computationally intensive and which minimises the potential for structural errors arising from systematic biases introduced during calibration. Our objective is to test the specific hypothesis that different combinations of a small number of available datasets will result in a significant variation in pre-calibration model performance, allowing rapid estimation of relative input data quality. The aim is to develop a simple, resource-efficient and transferable method for use in the design and specification of catchment models to support water resource management where data is of uncertain quality and/or quantitylity and decisions on where to invest efforts to improve them are limited by available resources.

2 STUDY AREA

Jordan is one of the poorest countries globally in terms of water resources and availability, with ~~extremely scarce renewable water resources and~~ less than 200 mm annual rainfall across 91% of its area (Abdulla and Al-Assa'd, 2006). Hence, severe water stress and the ongoing unsustainable drawdown of fossil groundwater reserves in Jordan make pilot schemes for increasing the capture of seasonal storm flows of considerable strategic importance. The Wala basin forms the northern 2 100 km^2 of the Mujib basin in central Jordan (Fig. 1). Its main drainage stream is Wadi Wala, which flows from an elevation 750 m to -100 m (a.s.l), where it joins Wadi Mujib and their confluence flows to the Dead Sea (Cordova, 2008). It has a Mediterranean climate characterized by hot dry summers and cold wet winters (Al-Bakri and Al-Jahmany, 2013). Maximum precipitation occurs in December and January while the rainy season extends between October and May. Average annual rainfall decreases in a northwest-southeast gradient from 500 mm a^{-1} to less than 100 mm a^{-1}, with an average of 181 mm a^{-1} (Margane et al., 2009).

The area of this study, the Wala catchment, occupies 1 743 km^2 upstream of the Wala Dam (Fig. 1). The Wala catchment and the aquifer beneath it form an important hydrologic system in Jordan, with the Wala Dam (31.56 ºN, 35.80 ºE) constructed between 1999 and 2002 to artificially recharge groundwater storage. This recharge supports agricultural activities in downstream cultivated areas as well as supplementing the potable water needs of the capital city Amman via abstraction wells approximately 9 km downstream of the dam at Al-Heidan (Ta'any, 2011). Wadi Wala had a permanent discharge before intense pumping started in the 1990s (Cordova, 2008). The main agricultural activity within the catchment is sheep and goat grazing, and its land cover is characterized by scrub vegetation, minor tree cover and some irrigated and non-irrigated crops. Since plans are in place for an expansion in the number of artificial recharge schemes, funded by UN and other international aid monies (JNFP, 2012; Margane et al., 2009), Wala provides a critical and influential case study in the development and management of catchment water resources in Jordan and the wider region.

3 METHODS

3.1 Approach

All available weather, soil, land-use and topographic datasets for the catchment are collated and characterised as described below. The general form and boundary conditions of the catchment are implemented in the SWAT (ArcSWAT 2012, http://swat.tamu.edu/software/arcswat/) model framework and used as input factorial combinations of the available datasets, yielding 18 different model representations of the Wala catchment. These models are run prior to internal calibration in order to elucidate the range of input influences on model performance, using observed discharge and empirical sediment data as benchmarks for comparison of model outputs. Visual and statistical assessment criteria are applied to check goodness of fit of scenarios outputs and observations both visually and quantitatively.

3.2 Model selection and structure

SWAT is selected for this work as it enables a continuous real-time model to simulate hydrology, land management and sedimentation processes on a basin-wide scale (Arnold et al., 1998; Srinivasan et al., 1998). SWAT also uses physical data for topography, weather, soil properties and land-use to directly simulate physical processes, rather than depending on regression formulas to determine input-output relationships (Arnold et al., 2013). Model parameters, input variables and methods that pertain to each type of the main inputs discussed in this study are detailed by Neitsch et al. (2011). There is extensive literature on SWAT and its applications in general (Arnold and Fohrer, 2005; Arnold et al., 2012) and in dryland in particular (see for example Adem et al. (2016); Adham et al. (2016); Havrylenko et al. (2016); Özcan et al. (2016)). The applicability of SWAT in arid environments is assessed by Wu et al. (2016) and the results obtained encourage using SWAT

in regions of similar characteristics. Zhang et al. (2016) show by comparing the performance of a conceptual model and SWAT, as a physically-based model, in simulating arid regions in China that both models perform reasonably well; however, data constraints and deficiencies, if not addressed properly, can limit SWAT's performance. In another comparison of modelling tools capability of simulating arid regions undertaken by Liu et al. (2016), SWAT shows strengths in different hydrological processes important to study arid regions such as lateral flow, while it performs relatively poorer in other processes, which are not main interest in this study such as snow, hence the outcome supports our selection. Marek et al. (2016) investigate evapotranspiration in the semi-arid Texas High Plains and state that SWAT is a suitable tool to simulate it. Shrestha et al. (2016) demonstrate the capability of SWAT to perform realistically in areas of extreme conditions like the semi-arid Onkaparinga catchment, South Australia and stress on the value of improving data sources for more realistic performance and robust simulation using SWAT. Several hydrological modelling studies undertaken in arid and semi-arid regions in Tunisia indicate that SWAT simulates various hydrological processes of these areas with reasonable accuracy and reliability (Ouessar et al., 2009; Ouessar et al., 2008). In Jordan, and adjacent to the Wala catchment, SWAT is employed successfully by Ijam and Al-Mahamid (2012) to estimate sediment inflow to the Mujib Dam reservoir and identify patterns of soil erosion across the Mujib basin. However, their study strongly recommends improving field measurements of sedimentation in the area for more confidence in the proposed model in simulating sediments as the data utilized for model calibration are constructed based on previous studies to cover the shortage in observed data. Ijam and Tarawneh (2012) also predict water and sediment yields from the Wala catchment using SWAT and conclude that it satisfactorily simulates hydrological processes and sedimentation in the area but again stress on the obstacles posed by the lack of data, especially those required for model calibration, the case in which reducing input uncertainty may partially account for potential errors. The key features of the SWAT model are briefly described below.

~~The key features of the model are briefly described below.~~

SWAT applies two levels of physical discretization: i) watershed into subbasins; and, ii) subbasins into hydrologic response units (HRUs), which are regions of unique soil, slope and land-use combinations (Arnold et al., 1998; Srinivasan et al., 1998). SWAT employs input weather information along with water budget techniques to quantitatively describe interrelated watershed hydrology components on a daily basis (Betrie et al., 2011). ~~Basin characteristics are directly represented in the model by applying~~ The model applies the US Soil Conservation Service Curve Number (CN) method to transform daily rainfall to surface runoff (USDA, 1972) while the CN varies according to basin characteristics including the type of land-use, soil group and antecedent moisture content (USDA, 1986). These parameters are extracted/calculated from the soil and land-use data. A set of widely tested sub-models are incorporated into SWAT to simulate key hydrological functions:

the Rational method (Chow et al., 1988) to predict peak discharge rate depending on daily precipitation, calculated surface runoff and topographic parameters derived from the DEM;

crack-discharge model combined with storage routing techniques and direct soil parameters such as hydraulic conductivity, percent clay content, available water capacity and bulk density to estimate percolation (Arnold et al., 1995);

the Penman-Monteith method (Monteith, 1964) for evapotranspiration, which depends on daily wind speed; maximum/minimum temperature, evaporative demand of soil and characteristics of land-cover leaves. These are all provided to the model through weather, soil and land-use data;

the variable storage coefficient to compute channel discharge routing (Williams, 1969), for which length, slope and Manning's value of channels are important information derived from the DEM and soil data (Neitsch et al., 2011);

the Williams and Singh (1995) formulas introduced for the Universal Soil Loss Equation (USLE) and its modification (MUSLE) to predict gross soil erosion and sediment yield at HRU-level respectively. Key parameters required for

4

these formulas are soil erodibility factor, percent rock in soil, cover/ management and practice factors taken from land-use data, topographic factor and parameters of surface runoff and peak discharge as detailed clearly in (Neitsch et al., 2011).

The model considers channel degradation as a result of stream energy and sediment deposition in channels, according to particle fall velocity, to investigate sediment transport (Williams, 1980). The input datasets required to represent the physical characteristics of the area and provide the model parameters are described in the following sections. Running the SWAT model yields a range of outputs for different model components, including the watershed, subbasins, HRUs and channel system (see Arnold et al. (2013) for full description)

3.3 Catchment configuration

The catchment area is defined by reference to the 30 m-resolution DEM obtained from ASTER-GDEM version 2 (Tachikawa et al., 2011). The DEM is used to derive topographic parameters for the catchment area, such as overland slope and slope length, and define stream pattern according to customized threshold of the area contributing to each branch (Di Luzio et al., 2002). An optimal subbasin area threshold of 3 % (5000 ha for Wala) of watershed area (Jha et al., 2004) resulted in 23 model subbasins, for which main streams and outlets are defined (Fig. 2). The main streams of subbasins 15 and 19 form the arms of the Wala dam reservoir, and their confluence is the main stream of subbasin 16, representing the catchment outlet and dam location (Tarawneh, 2007).

Land slope is derived from the 30-m DEM described above and shows relatively flat topography over the upper catchment compared to steep canyons near the outlet. Relatively steep slopes characterize the western and northern parts of the watershed. Average and maximum slopes within the area are 5.42° and 54.4°, respectively.

3.4 Input data

3.4.1 Land use

Land-use maps from three different sources are used: i) a 30 m-resolution raster grid reprocessed from the detailed map developed and presented by Tarawneh (2007) based on the 1:250 000-scale map of the National Soil Map and Land-use Project of Jordan (Ministry of Agriculture, 1994); ii) the land use/cover map of Jordan produced by Al-Bakri et al. (2013) and also reprocessed to a 30 m-resolution grid; iii) the Europe/Asia land-use grid (WaterBase, 2012) constructed from the Global Land Cover Characterization (GLCC) database with a 1:2 000 000 scale and 1 km spatial resolution. Figure 3 illustrates that the three maps all show two dominant types of vegetation over the area, with minor coverage by other land-use classes. Considering the importance of land use to modelling and planning of drylands, it is essential to take into consideration any major land use differences over time (Wolffc, 2011). By investigating land use variation in several sites in the Badia zone in Jordan over the period 1953 – 1992, Al-Bakri et al. (2001) conclude that land use changes are from rangeland to cultivated areas and urban settlements in addition to the appearance of some irrigation fields after 1978. For the Wala catchment, land use maps and information from different sources and periods tend to exhibit only minor differences in the major land use types and this supports our assumption that land use changes in the study area can be neglected for the time frame of the current study (1979 – 2002), particularly that the model considers the major land use type (greater than specified threshold) in each hydrologic response unit.

3.4.2 Soil

The physical soil characteristics required by the model include soil hydrologic group (Wood and Blackburn, 1984), depth of soil layers, moist bulk density, available water capacity, saturated hydraulic conductivity, organic carbon content, erodibility

5

factor, moist soil albedo, rock fragment content and percentages of silt, sand and clay (Arnold et al., 2011). Two soil datasets are compared (Fig. 4): i) the Europe/Asia soil grid (WaterBase, 2012) produced from the Food and Agriculture Organization (FAO) map with 1:25 000 000-scale and a coarse spatial resolution of 10 km (Leon, 2011), showing only three types of two-layer soils over the Wala catchment; ii) the map produced by Tarawneh (2007) and processed to 30 m-resolution based on the 1:250 000-scale soil map and analysis released by the Jordanian government (Ministry of Agriculture, 1994). In the latter case, the catchment is divided into 17 three-layer soil units, each linked to a soil properties database based on thorough sampling undertaken by the national project to study soil profile, composition and spatial distribution. The Tarawneh (2007) map provides higher resolution and level of detail and more importantly, measured ground-truth based data including silt, sand and clay percentages, percent organic carbon and rock content, which are used to define soil texture and estimate or calculate the remaining characteristics using pre-developed models, equations or graphs.

3.4.3 Topography

Land slope is derived from the 30-m DEM described above and shows relatively flat topography over the upper catchment compared to steep canyons near the outlet. Relatively steep slopes characterize the western and northern parts of the watershed. Average and maximum slopes within the area are 5.42° and 54.4°, respectively.

3.5 Hydrologic response units (HRUs)

Soil, land-use and slope combinations define HRUs, over which water and sediment loadings are estimated. Hence, each set of input data defines a unique set of HRUs and this provides a key structural characteristic that governs sensitivity of the model to changes in these fundamental input datasets. Different combinations of the input datasets specified above result in significant differences in the number and physical characteristics of HRUs and consequently water and sediment loading simulation at both HRU and subbasin levels. For instance, combining the WaterBase (2012) soil data (Fig. 4a) with each of the three land-use maps shown in Fig. 3 results in 47, 67 and 68 HRUs respectively, based on a multiple threshold criteria of 20, 30 and 30 % applied on land use, soil and slope respectively. The number of HRUs generated in each of the different scenarios is displayed alongside modelling results presented later in this paper.

3.6 Input weather datasets

SWAT requires daily series of climatic data as model input. Where incomplete climate records exist, SWAT uses a built-in weather generator algorithm to statistically process monthly data taken from representative weather stations to produce full daily series or fill any missing records in the available measured data (Arnold et al., 2013). The SWAT generator uses a first-order Markov chain model to predict wet/dry days depending on monthly wet/dry probabilities provided by the user. Daily precipitation is estimated for wet days using a skewed distribution while a normal distribution is used to generate missing maximum/minimum temperature and solar radiation in conjunction with a continuity equation. These values are adjusted depending on wet/dry conditions so that the monthly average of generated daily values agrees with the averages provided by the user. Details of the SWAT weather generator are provided by Neitsch et al. (2011). Average monthly climatic parameter data from the Qatraneh (31.24 ºN, 36.04 ºE) and Errabbah (31.27 ºN, 35.74ºE) weather stations (Fig. 5) over ten years are processed to provide two weather generator files. The mentioned stations are located outside the watershed but it is a common practice in watershed modelling to use weather data monitored outside the study area, though some potential complications may arise regarding validity and representativeness of these stations (Fuka et al., 2014).

Daily precipitation records from 26 gauges in and around the study area obtained from the Ministry of Water and Irrigation of Jordan are used. Record lengths vary, with the earliest record starting in 1938. However, detailed analysis of these datasets revealed poor quality and gaps in most of the series, leaving only three gauges of sufficient quality to provide

6

continuous daily records between January 1971 and September 2002. Figure 5 shows the rain gauges used in this study (noting their respective missing record percentage over 31 years). The stations are preferentially distributed to the west of the catchment, as such the representativeness of these stations may be poorer to the east of the catchment. A considerable portion of the Madaba gauge (31.71 ºN, 35.79 ºE) record is missing. The Madaba gauge is used in this study to demonstrate model sensitivity to gaps in rainfall information in the semi-arid region characterised by intense, highly intermittent storms.

Temperature is important for key processes in the hydrologic cycle such as evapotranspiration and vegetation growth (Sandholt et al., 2002). The weather stations used in this study, Qatraneh and Errabbah (Fig. 5) hold records of daily maximum/minimum temperature for the period 1971 - 2002, with infrequent gaps. By reviewing the temperature variation, we found it followed a smoother pattern than that of precipitation, hence easier to estimate or forecast to fill the gaps. Figure 6 shows an illustrative subset of daily precipitation and temperature (maximum and minimum) for the two stations used in the model in the representative period 2000 - 2003. Recent research suggests significant increasing trends in daily maximum and minimum temperatures in the Middle East and north Africa over the last 50 - 100 years (Ageena et al., 2014, 2013) which are directly proportional to increasing aridity (see for example Zhang et al. (2005), Trondalen (2009)). However, since the current model is run for shorter periods, long-term climate change is not considered significant in this study.

Global atmospheric reanalyses such as the Climate Forecast System Reanalysis (CFSR) are routinely used to provide catchment-scale hydrological simulations with the required climatic data, particularly in locations for which measured data are scarce (Wang et al., 2011). The CFSR is designed by the National Centre for Environmental Prediction (NCEP) to provide continuous weather data for grid points across the globe for the period 1979 – 2010 (Saha et al., 2010). In an area such as Wala, characterised by intermittent, intense, often localised rainstorms, it is pertinent to query whether such a global reanalyses can adequately capture the local drivers of hydrological activity. Four of the CFSR data points are in or close to the study area (Fig. 5) and therefore their data (daily precipitation, temperature, solar radiation, relative humidity and wind speed) are used as an additional input dataset to compare with the local weather station data.

3.7 Scenario comparison

Three land use maps and two soil maps are combined factorially with three sets of weather data obtained from: i) CFSR; ii) local stations including Madaba; iii) local stations excluding Madaba, yielding 18 different model scenarios (Tables 1 and 2). Running the SWAT model yields a range of outputs for different model components, including the watershed, subbasins, HRUs and channel system (see Arnold et al. (2013) for full description). We compare the average monthly stream outflow (discharge, m3s-1) and the monthly sediment transported out of reaches (t) with observations of discharge and observation-derived sediment yield, respectively. The observed discharge data comprise average monthly discharge (m3s-1) obtained from daily measurements at the Wadi Wala flow station CD0038 (31.55 ºN, 35.77 ºE), located 5 km downstream of the current dam location (Margane et al., 2009) for the period January 1971 to September 2002 (before impoundment started), available from the Ministry of Water and Irrigation of Jordan. Howard Humphreys and Partners (1992) provide a sediment-rating curve identifying a strong log linear relationship between log sediment yield (kg s-1) and log discharge (m3s-1) for the Wala gauging station during the design studies of the Wala Dam. Howard Humphreys and Partners (1992) provide a sediment-rating curve identifying a strong log linear relationship between log sediment yield (kg s^{-1}) and log discharge (m^3 s^{-1}) for the Wala gauging station (CD0038) during the design studies of the Wala Dam. This relationship is presented by Tarawneh (2007) in Eq 1 below and used to develop values of sediment yield (t) corresponding to the available discharge values at station CD0038, which are used in this study. Observed sediment accumulation in Wala Dam since construction (2002-2007) closely relates to that modelling over the coeval period (Tarawneh, 2007; Wala Dam Management, 2013).

$$\log Discharge = 0.5833 \log Sediment\ yield + 0.0168 \qquad\qquad\qquad\qquad\qquad (1)$$

~~Howard Humphreys and Partners (1992) identify a strong log linear relationship between sediment yield (kg s-1) and log discharge (m3s-1) for the Wala catchment and this relationship is used Tarawneh (2007) to develop the observed sediment yield (t) for the catchment at station CD0038 after Tarawneh (2007).~~

The ~~parameterised~~ SWAT model ~~is run~~ in this study is setup to produce monthly output by averaging daily estimates ~~on a monthly time interval~~ to simulate seasonal variation of discharge and sediment in the period January 1979 through to January 2003. Model performance under each scenario (combination of input datasets) is evaluated by quantitative comparison with the observed discharge and observation-derived sediment load data series using standard graphical and statistical techniques for watershed modelling (Moriasi et al., 2007). Hydrograph comparison (Yen, 1995) of simulated and observed discharge and sediment yield are combined with quantitative measures.

A suite of four standard statistical instruments are employed to compare input scenarios on the basis of pre-calibration modelled vs observed catchment outputs: coefficient of determination r^2 (Eq.~~1~~2) (Goodwin and Leech, 2006):

$$r^2 = \left(\frac{\sum_{i=1}^{n}(O_i-\bar{O})(P_i-\bar{P})}{\sqrt{\sum_{i=1}^{n}(O_i-\bar{O})^2 \sum_{i=1}^{n}(P_i-\bar{P})^2}} \right)^2 \tag{$\cancel{1}$2}$$

where O is observed and P is predicted values; Nash-Sutcliffe Efficiency (NSE) (Eq. ~~2~~3) developed by Nash and Sutcliffe (1970):

$$NSE = 1 - \frac{\sum_{i=1}^{n}(O_i-P_i)^2}{\sum_{i=1}^{n}(O_i-\bar{O})^2} \tag{$\cancel{2}$3}$$

root mean square error standard deviation ratio (RSR) (Eq. ~~3~~4):

$$RSR = \frac{\sqrt{\sum_{i=1}^{n}(O_i-P_i)^2}}{\sqrt{\sum_{i=1}^{n}(O_i-\bar{O})^2}} \tag{$\cancel{3}$4}$$

and percent bias (PBIAS) (Eq. ~~4~~5) (Moriasi et al., 2007):

$$PBIAS = \frac{\sum_{i=1}^{n}(O_i-P_i)*100}{\sum_{i=1}^{n}O_i} \tag{5}$$

Whilst input data (climatic) are based at a daily temporal scale, the model outputs are considered at a monthly timescale for several reasons, i) daily observations of discharge and sediment are unavailable at the Wala station for the whole period of study, with only monthly observations available for model evaluation; ii) A shorter period (1990 – 1996) of daily observations are available at Wala station, but using these yields poor correlations <0.1 between daily model-simulated and observed discharge, iii) with incomplete/low quality measurements, potential for lag within the pairs of daily simulated and observed values (for model statistical evaluation) can present challenges, which can be reduced when using aggregated temporal data; and most importantly, iv) the objective of this study is to determine long-term flux within the catchment, avoiding the complexity presented by ephemeral systems and since the monthly comparison achieves reasonable fit between observed and simulated values, it is considered sufficient for evaluation of the current model with more convenience. However, all calculations of the model occur on a daily time step, which ensures that hydrological events are accounted for separately as they occur each day.

A similar approach is adopted in several comparable studies, particularly using SWAT, in both humid and arid regions. Spruill et al. (2000) evaluate daily and monthly SWAT models simulation for a small watershed in central Kentucky and state that SWAT is an efficient tool for monthly runoff simulation with NSE values of 0.58 – 0.89 compared to - 0.04 – 0.19 for daily runoff simulation during the same period. The reason suggested is that the model poorly detects peak flows and recession rates

while it performs better with total monthly values. For reasonable performance of SWAT, Huang and Zhang (2004) select to simulate discharge in a semi-arid catchment in China on a monthly basis, which leads to NSE of 0.88. The difference between daily and monthly simulations is investigated in watersheds of different scales by Heathman and Larose (2007) and the results show that simulating higher discharge rates, which is usually associated with larger watersheds, introduces greater uncertainty in SWAT discharge estimates and the study states that very good model performance is achieved for monthly stream-flow estimation while the outputs of daily simulation are only within acceptable range.

4 RESULTS AND DISCUSSION

4.1 Comparison of statistical measures

All four statistics exhibit significant variability in model behaviour among the eighteen input scenarios. Figures 7 and 8 show the r^2 and PBIAS statistics, respectively, for each scenario. For discharge prediction, highest r^2 is obtained from scenarios 16, 10 and 4 (group 1); and a far lower r^2 is associated with scenarios 2, 8 and 14 (group 3). Values of r^2 for the remaining twelve scenarios (group 2) are located between these two groups, with slight or no difference between successive scenarios. r^2 values for sediment prediction show higher correlation than that of predicted discharge (Fig. 7b). It should be noted that r^2 quantifies only the dispersion; therefore in some cases very good r^2 values may be obtained when the model is over/under-predicting all the time regardless of the accuracy. For PBIAS, Fig. 8a shows that scenarios 16, 10 and 4 (which use the CFSR data) tend to underestimate discharge, while all remaining scenarios show overestimation. Figure 8b shows that scenarios 16, 10 and 4 have least tendency to over/under-predict sediment yield with PBIAS values of 31, -17 and -33, respectively, while all other scenarios significantly overestimate sediment yield. Both indicators consistently identify the input scenarios that most closely represent the observed discharge and sediment data prior to calibration, but yield little further information with which to differentiate between scenarios.

NSE and RSR enable a finer distinction between scenarios, revealing clear trends arising from the influence of the different input datasets (Table 1 and Table 2). Table 1 shows the 18 scenarios arranged according to NSE and RSR, with similar descending order for both criteria and a clear structure evident in the importance of different inputs. Scenarios are divided into two distinct groups: i) those using the CFSR dataset, and ii) those applying a combination of generated and locally measured series of weather variables. NSE drops and RSR increases significantly for equivalent scenarios when only the weather varies (e.g. scenario pairs 16/18 and 1/2), with improved performance for the CFSR in all cases. Several studies worldwide lead to similar results and show that CFSR data out-perform local records. Potential causes for this focus on the data incompleteness, data quality, representativeness of instrumental site data to wider regions, data management and instrument maintenance being common factors; these present a number of challenges, particularly in areas and regions with challenging climates ((Wheater et al., 2008b)). With special reference to SWAT, Fuka et al. (2014) states that providing SWAT models with CFSR data substantially improves model performance over forcing the model to use data acquired from local weather stations (please see Saleh et al. (2000) for a case study leading to a similar statement). A clear improvement is obtained using the Tarawneh (2007) produced soil map in preference to the global Waterbase map (e.g. scenario pairs 16/13 and 5/2). As with other statistical measures, inclusion of the Madaba rainfall gauge cause a sharp drop in NSE and increase in RSR for otherwise identical scenarios (e.g. scenario pairs 18/17 and 15/14). Consideration of the three land-use classes shows a clear variation in uncalibrated model performance, as the global land-use layer out-performs the two locally processed maps.

Similarly, Table 2 shows the order of NSE and RSR calculated to assess sediment yield prediction. The scenario rank order differs from that in Table 1 due to the different sensitivity of discharge and sediment simulation to various types of inputs. The global soil map produces considerably poorer model performance than the Tarawneh (2007) map, (e.g. scenario pairs 16/13, 18/15 and 5/2). The comparisons lead to initial classification of scenarios into two groups defined by the

specification of the soil input dataset. Within each group the ranking order of land-use/weather combinations is similar. This confirms the high sensitivity of the sediment simulation to input soil data. Across the rest of scenarios, using the CFSR data results in significantly higher NSE and lower RSR, with a wide gap between them and the successive values (scenarios 18, 12 and downward). This is consistent with the results of the discharge assessment (Table 1). The importance of land-use in determining sediment yield is clear by the priority it takes over the weather in the ranking of scenarios 18, 12, 17, 11, 6, and 5. In all cases, excluding the Madaba rain gauge always yields a closer correlation to observations between scenarios of similar conditions.

4.2 Case study results

4.2.1 Soil data

The choice of soil dataset is a strong control on model behaviour by all measures (Table 1, Table 2). The pre-calibration performance of the model against both discharge and sediment data is better using the more detailed local soil map (Tarawneh, 2007 and Ministry of Agriculture, 1994) and the CFSR dataset, in combination with the global land-use map. Conversely, the weakest uncalibrated performance against observations results from applying the global soil map (Waterbase, 2012) and local weather data including the Madaba rain gauge (i.e. the combination of measured and SWAT-generated weather data). It is clear from Fig. 4 that there is a significant difference in the granularity of data between the two input soil maps. The additional detail embodied in the Tarawneh (2007) dataset yields significantly more range in soil class and key SWAT parameters, such as permeability, which presumably directly influences model calculation of both discharge and sediment loading.

4.2.2 Weather data

In contrast to the soil datasets (where more granular, sampled-derived data yield best model performance), the global reanalysis (CFSR) weather data consistently yield better pre-calibration model performance (scenarios 16, 10, 4, 13, 7, 1) than scenarios using locally-recorded weather data (Table 1 and 2). This difference is further exacerbated when the data from the Madaba recording station is included in the locally-recorded input dataset. A qualitative inspection of rainfall data series shows high values of rainfall recorded at Madaba compared to the global CFSR for similar periods. This in turn influences the extensive infill values generated by the SWAT weather generator for this dataset. Fuka et al. (2014) suggests that using CFSR data provides a remedy to the potential uncertainty linked to using local weather records, which are seldom complete and may not realistically represent the watershed and provide point rainfall measurements neglecting the effects of hydro-climatic gradients (Ciach, 2003). To understand how prediction of discharge differs between scenarios and for visual comparison between observed and simulated discharges, six scenarios are selected (Fig. 9) to visually assess model performance. Figure 9 clearly illustrates that better fit is associated with scenarios of lower RSR (closer to zero) and higher NSE and r^2 values, with over-prediction resulting from using local weather data regardless of the inclusion of the strongly discontinuous Madaba dataset. Graphical comparison of four sets of observed and simulated sediment yield is displayed in Fig. 10 to demonstrate the tendency of the poorly performing scenarios to significantly overestimate sediment yield. This is consistent with records containing anomalously high rainfall readings.

4.2.3 Land-use data and sensitivity analysis

The only difference between the three scenarios (16, 10, 4) achieving best performance for both discharge and sediment prediction is the land-use data source. These scenarios show good correlation between simulated and observed variables, with lowest r^2 and highest NSE and r^2. A similar order of the three land-use scenarios is identified for both discharge and sediment, but the performance of all three scenarios is almost equal. A standard SWAT model 32-parameter global sensitivity analysis (Dechmi et al., 2012; Van Griensven, 2005) is applied using the best-performing scenario 16 to identify quantitatively which

internal parameters are the most sensitive for the Wala catchment SWAT model. Table 3 shows the results of this sensitivity analysis using observed discharge and observation-derived sediment load values at the Wala flow station during the simulation period. After discounting parameters which score low sensitivities, it is clear that the seven highest-ranked parameters are closely related to the properties defined by the soils and land use data inputs.

The most sensitive parameter is the SCS Curve Number (CN), derived directly from land-use data (Neitsch et al., 2011). Nevertheless, our pre-calibration results show that selection among the land-use datasets available in this study yields least influence on model performance. This apparent contradiction can be resolved by inspection of Fig. 3 which shows, in contrast to the soils datasets shown in Fig. 4, that there is relatively little variation both in spatial distribution and range of physical characteristics among the three available land use maps. Reviewing the CN values for the dominant land-use classes in the three land-use maps, we find them to be close (ranging from 80 to 84) due to the similarity of properties defined for each dataset. Furthermore, the method of HRU definition within SWAT selects the major land-use in each HRU, thus potentially nullifying the gains of higher-resolution land-use maps with numerous smaller land-use classes. While the sensitivity analysis emphasizes the general importance of land-use definition in SWAT catchment models, this case study shows the value of quantitative interrogation of the available datasets for any specific application.

4.2.4 Global reanalysis vs locally-derived datasets

For prediction of discharge (Table 1, Fig. 9), the scenario analysis strongly confirms that the most sensitive constituent is the input weather data. It is obvious to say that precipitation is a fundamental driver of runoff and discharge time-series. However, considerably higher model performance is achieved by the reanalysed CFSR versus local weather datasets, regardless of the other input datasets. We suggest the reason for this is the continuity and consistency of the CFSR dataset, which is provided by the NCEP reanalysis climate data derived from global satellite imagery for a grid of statistically interpolated points (Saha et al., 2010). Although the local dataset might be expected to capture average daily events more precisely, this relies on well-calibrated, well-maintained instrumentation and proper representativeness of measurement stations within the study area.

In the Wala catchment, as in many locations world-wide, poor data continuity and reliability necessitate generation of infill data points by the SWAT weather generator. The potential of individual recording stations as a source of error in model output is further demonstrated by the observation that for otherwise similar scenarios, incorporating the Madaba rain gauge (which depends on the SWAT weather simulator to generate 32.75% of its daily records) significantly reduces the performance of the model. This does not fully negate the potential inappropriateness of the weather stations utilized to construct the weather generator, however, by reviewing the statistics generated using these stations, they do not seem to vary significantly from measured weather parameters within the area. We suggest that basic weaknesses in the recording of data are compounded by the challenges posed to the rainfall generator algorithm by strong daily, monthly, and interannual variability in an arid-climate rainfall regime. One possible area for investigation in this respect is the use in the generator of the Markov chain model, which does not account for the interannual variability in the daily weather causing clear inconsistencies with measurements (Jiang et al., 2011). This is crucial in semi-arid and arid regions where precipitation is much more variable on all timescales than in temperate and humid regions.

The control of the choice of soil dataset on model performance is substantial in our analysis, which corresponds with the sensitivity of the model to its internal soil characteristic parameters (Table 3). The results show clear improvement in pre-calibration model performance using higher resolution maps built using field sampling rather than the global map, which is of lower classification quality and resolution, relying heavily on satellite remote sensing. The local soil map yields better pre-calibration performance than the global map, even with different weather data (with/without Madaba gauge), emphasizing the primary importance of soil data in this model of the Wala catchment. Sediment simulation is highly influenced by changes in soil definition; this is expected because soil parameters are directly needed by the USLE (Wischmeier and Smith, 1965) and

MUSLE to predict soil erosion and sediment yield and are also required to simulate discharge, which is important for sediment yield prediction.

4.3 Effect on calibrated model performance

While it is clear that use of SWAT pre-calibration enables rapid, quantitative comparison of different input datasets, we wished to confirm that optimization of the model in this way yields improved performance after calibration also. To test the functionality of the presented pre-calibration approach in enhancing subsequent calibrated model performance, automatic internal parameter calibration (Abbaspour et al., 2007) is performed for the best (16) and poorest (2) pre-calibration scenarios as described above. Standard SWAT calibration is undertaken by applying consistent conditions and criteria for each scenario separately for discharge simulation. The calibration targets the set of parameters defined in the sensitivity analysis as being the strongest controls on model performance (Table 3). NSE is selected as an objective function and 1000 iterations run.

Table 4 displays a comparison between uncalibrated and calibrated scenarios. Calibration improves the NSE for discharge simulation from 0.56 to 0.64 and from -12 to -11.29 for scenario 16 and scenario 2, respectively. This represents a 14 % performance gain for scenario 16 and a 6 % improvement in scenario 2. It is clear that our pre-calibration methodology accurately reflects the fully-calibrated performance of models based on different input data combinations, yet with a fraction of the computational effort and time. These findings emphasize the value of reducing model uncertainty by undertaking preliminary screening of input datasets and selecting the best available conditions to construct models that achieve the best possible calibrated performance.

5 CONCLUSIONS

Previous use of a SWAT model to simulate discharge and sediment yield across the Wala catchment led to a detailed understanding of the hydrological system of the area and the interaction between its components and processes (Ijam and Tarawneh, 2012). In this paper we have developed a discrete methodology (Fig. 11) for using a SWAT model framework, comprising an analytical stage prior to full model calibration, to support decision-making in the selection and application of input datasets for use with catchment hydrological models. This should be of value in the specification and design of catchment modelling in many semi-arid, arid and data-poor regions, since the factorial scenario-testing approach allows rapid, quantitative comparison among a range of datasets of uncertain quality. Model sensitivity to various types and resolutions of data is clear and demonstrates the significant influence of input selection on model performance prior to the calibration step and hence the potential to minimise the computational effort and possible systematic biases inherent in the calibration process (Beven, 2011; Wheater et al, 2008a). In summary, we find:

continuity and quality of record are critical factors in selecting weather data, over and above use of local measurements. In our case study, inclusion/exclusion of the poor quality, incomplete Madaba dataset results in significant variability in model performance and leads us to recommend the preferential use of global reanalysis data where there is any doubt about local data quality, even in rainfall regimes which are characterised by infrequent, irregular, intense storm events;

high resolution, high quality soil data (likely to be available only through detailed local survey) yields significant improvements in pre-calibration model performance over globally-available datasets obtained from e.g. remote sensing;

land-use definition in this specific case at the Wala shows the least impact of the three inputs assessed. We propose this is due to broad similarities between available land use datasets, which is likely to be representative of conditions across much of central Jordan and surrounding arid and semi-arid regions. Our results suggest that only significant

and spatially-extensive deviation in actual land use from global freely-available datasets such as GLCC – either as a result of rapid (or predicted) land development or land degradation – will significantly impact on overall model function and performance.

The key benefit of this work in the context of the Wala Dam and the management of water resources in Jordan is an improvement in the confidence with which catchment data and models can be used in decision making. This applies both to management of existing artificial recharge catchments, such as Wadi Wala, and to the options assessment and selection of new schemes which are critical to securing a more sustainable water resource for the country (JNFP, 2012). The potential utility of SWAT in this context has been previously demonstrated (Ijam and Tarawneh, 2011); this current work provides a rational basis for supporting the selection and use of available input datasets, and targeting of field resources to improve the reliability and coverage of these data.

A general observation is that globally-available weather and land-use datasets tended to perform equal to or better than local data as inputs to the catchment model over a range of dataset combinations, suggesting that these may be preferable sources of inputs where local data is sparse or unreliable. However, we found obtaining a high-quality, ground-truthed soil dataset offers substantial improvements in pre-calibration model performance over regional or global soil datasets. We therefore recommend detailed soil mapping as a priority for targeting desk and field resources to support studies in settings comparable to that studied here. It is highly recommended also to quantitatively and qualitatively improve field measurements to provide trustworthy observations which can be used in model assessment and calibration, for example by increasing the number of gauges within the area.

AUTHOR CONTRIBUTION

Tarawneh undertook the SWAT analysis and led the paper writing. Bridge and Macdonald as supervisors contributed to structure and research design and supported the writing of the paper and analysis.

DATA AVAILABILITY

This work forms part of a thesis at the University of Liverpool which will be submitted in early 2017. At that time all digital data products specifically associated with the work will be made publically available via the University of Liverpool data catalogue (datacat.liverpool.ac.uk). Before then, all enquiries for data can be made to the corresponding author.

ACKNOWLEDGMENT

REFERENCES

Abbaspour, K., Vejdani, M., and Haghighat, S.: SWAT-CUP calibration and uncertainty programs for SWAT, Modelling and Simulation Society of Australia and New Zealand,2007.

Abdulla, F. and Al-Assa'd, T.: Modeling of groundwater flow for Mujib aquifer, Jordan, J Earth Syst Sci, 115, 289-297, 2006.

Adem, A. A., Tilahun, S. A., Ayana, E. K., Worqlul, A. W., Assefa, T. T., Dessu, S. B., and Melesse, A. M.: Climate Change Impact on Sediment Yield in the Upper Gilgel Abay Catchment, Blue Nile Basin, Ethiopia. In: Landscape Dynamics, Soils and Hydrological Processes in Varied Climates, Springer, 2016.

Adham, A., Riksen, M., Ouessar, M., and Ritsema, C. J.: A Methodology to Assess and Evaluate Rainwater Harvesting Techniques in (Semi-) Arid Regions, Water, 8, 198, 2016.

Ageena, I., Macdonald, N., and Morse, A.: Temporal and spatial variation of maximum and climatic temperature across Libya (1945-2009), Theoretical and Applied Climatology, 2014. 549-563, 2014.

Ageena, I., Macdonald, N., and Morse, A.: Variability of minimum temperature across Libya (1945–2009), International Journal of Climatology, 33, 641-653, 2013.

Al-Bakri, J. T. and Al-Jahmany, Y. Y.: Application of GIS and Remote Sensing to Groundwater Exploration in Al-Wala Basin in Jordan, Journal of Water Resource and Protection, 5, 962-971, 2013.

Al-Bakri, J. T., Salahat, M., Suleiman, A., Suifan, M., Hamdan, M. R., Khresat, S., and Kandakji, T.: Impact of Climate and Land Use Changes on Water and Food Security in Jordan: Implications for Transcending "The Tragedy of the Commons", Sustainability, 5, 724-748, 2013.

Al-Bakri, J. T., Taylor, J. C., and Brewer, T. R.: Monitoring land use change in the Badia transition zone in Jordan using aerial photography and satellite imagery, Geographical Journal, 167, 248-262, 2001.

Arnold, J. and Fohrer, N.: SWAT2000: current capabilities and research opportunities in applied watershed modelling, Hydrological processes, 19, 563-572, 2005.

Arnold, J., Kiniry, J., Srinivasan, R., Williams, J., Haney, E., and Neitsch, S.: Input/Output Documentation - Version 2012. Texas Water Resources Institute Texas, 2013.

Arnold, J., Kiniry, J., Srinivasan, R., Williams, J., Haney, E., and Neitsch, S.: Soil and Water Assessment Tool input/output file documentation: Version 2009, Texas Water Resources Institute Technical Report, 365, 2011.

Arnold, J., Moriasi, D., Gassman, P., Abbaspour, K., White, M., Srinivasan, R., Santhi, C., Harmel, R., van Griensven, A., M. Van Liew, Kannan, N., and Jha, M.: SWAT: Model use, calibration, and validation, Transactions of the American Society of Agricultural and Biological Engineers, 55, 1494-1508, 2012.

Arnold, J., Williams, J., and Maidment, D.: Continuous-time water and sediment-routing model for large basins, Journal of Hydraulic Engineering, 121, 171-183, 1995.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R.: Large area hydrologic modeling and assessment part I: Model development1. Wiley Online Library, 1998.

Beeson, P., Sadeghi, A., Lang, M., Tomer, M., and Daughtry, C.: Sediment delivery estimates in water quality models altered by resolution and source of topographic data, Journal of Environmental Quality, 43, 26-36, 2014.

Betrie, G., Mohamed, Y., Van Griensven, A., and Srinivasan, R.: Sediment management modelling in the Blue Nile Basin using SWAT model, Hydrology and Earth system sciences, 15, 807-818, 2011.

Beven, K.: Rainfall-runoff modelling: the primer, John Wiley & Sons, 2011.

Chaplot, V.: Impact of spatial input data resolution on hydrological and erosion modeling: Recommendations from a global assessment, Physics and Chemistry of the Earth, Parts A/B/C, 2013. 2013.

Chehbouni, A., Escadafal, R., Duchemin, B., Boulet, G., Simonneaux, V., Dedieu, G., Mougenot, B., Khabba, S., Kharrou, H., Maisongrande, P., Merlin, O., Chaponnière, A., Ezzahar, J., Er‐Raki, S., Hoedjes, J., Hadria, R., Abourida, A., Cheggour, A., Raibi, F., Boudhar, A., Benhadj, I., Hanich, L., Benkaddour, A., Guemouria, N., Chehbouni, A. H., Lahrouni, A., Olioso, A., Jacob, F., Williams, D. G., and Sobrino, J. A.: An integrated modelling and remote sensing approach for hydrological study in arid and semi‐arid regions: the SUDMED Programme, International Journal of Remote Sensing, 29, 5161-5181, 2008.

Chow, V., Maidment, D., and Mays, L.: Applied hydrology, McGraw-Hill, New York, 1988.

Ciach, G. J.: Local random errors in tipping-bucket rain gauge measurements, Journal of Atmospheric and Oceanic Technology, 20, 752-759, 2003.

Cordova, C.: Floodplain degradation and settlement history in Wadi al-Wala and Wadi ash-Shallalah, Jordan, Geomorphology, 101, 443-457, 2008.

Dechmi, F., Burguete, J., and Skhiri, A.: SWAT application in intensive irrigation systems: Model modification, calibration and validation, Journal of Hydrology, 470 471, 227-238, 2012.

Di Luzio, M., Arnold, J., and Srinivasan, R.: Effect of GIS data quality on small watershed stream flow and sediment simulations, Hydrological Processes, 19, 629-650, 2005.

Di Luzio, M., Srinivasan, R., and Arnold, J.: Integration of watershed tools and SWAT model into basins, JAWRA Journal of the American Water Resources Association, 38, 1127-1141, 2002.

Edmunds, W. M., Jayakumar, R., Mishra, A., Salih, A., Sorooshian, S., Wheater, H. S., and Logan, W.: G-WADI—the first decade, Sciences in Cold and Arid Regions, 5, 0001-0005, 2013.

Faramarzi, M., Srinivasan, R., Iravani, M., Bladon, K. D., Abbaspour, K. C., Zehnder, A. J., and Goss, G. G.: Setting up a hydrological model of Alberta: Data discrimination analyses prior to calibration, Environ Modell Softw, 74, 48-65, 2015.

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.

Gee, G. W. and Hillel, D.: Groundwater recharge in arid regions: review and critique of estimation methods, Hydrological Processes, 2, 255-266, 1988.

Goodwin, L. and Leech, N.: Understanding Correlation: Factors That Affect the Size of r, Journal of Experimental Education, 74, 249-266, 2006.

Griensven, A. v. and Meixner, T.: Methods to quantify and identify the sources of uncertainty for river basin water quality models, Water Science & Technology, 53, 51-59, 2006.

Havrylenko, S. B., Bodoque, J. M., Srinivasan, R., Zucarelli, G. V., and Mercuri, P.: Assessment of the soil water content in the Pampas region using SWAT, CATENA, 137, 298-309, 2016.

Heathman, G. and Larose, M.: Influence of scale on SWAT model calibration for streamflow, 2007, 2747-2753.

Howard Humphreys and Partners: Dams on Wadi Wala and Mujib, Jordan Valley Authority of Jordan, AmmanFinal report, 1992.

Huang, Q.-h. and Zhang, W.-c.: Improvement and application of GIS-based distributed SWAT hydrological modeling on high altitude, cold, semi-arid catchment of Heihe River Basin, China, 2, 2004.

Ijam, A. and Tarawneh, E.: Assessment of sediment yeild for Wala dam catchemnt area in Jordan, European Water, 38, 43-58, 2012.

Ijam, A. Z. and Al-Mahamid, M. H.: Predicting sedimentation at Mujib dam reservoir in Jordan, Jordan Journal of Civil Engineering, 6, 448-463, 2012.

Jha, M., Gassman, P., Secchi, S., Gu, R., and Arnold, J.: Effect of watershed subdivision on SWAT flow, sediment, and nutrient predictions. In: Journal of the American Water Resources Association, 3, Wiley-Blackwell, 2004.

Jiang, Z., Ding, Y., Zheng, C., and Chen, W.: An improved, downscaled, fine model for simulation of daily weather states, Advances in Atmospheric Sciences, 28, 1357-1366, 2011.

JNFP: Badia Restoration Program (BRP)- Community action plan 2011-2025, Amman, Report, 2012.

Kalantari, Z., Lyon, S. W., Jansson, P.-E., Stolte, J., French, H. K., Folkeson, L., and Sassner, M.: Modeller subjectivity and calibration impacts on hydrological model applications: An event-based comparison for a road-adjacent catchment in south-east Norway, Science of The Total Environment, 502, 315-329, 2015.

Krupnick, A., Morgenstern, R., Batz, M., Nelson, P., Burtraw, D., Shih, J.-S., and McWilliams, M.: Not a sure thing: Making regulatory choices under uncertainty, Resources for the Future Washington, DC, 2006.

Legesse, D., Vallet-Coulomb, C., and Gasse, F.: Hydrological response of a catchment to climate and land use changes in Tropical Africa: case study South Central Ethiopia, Journal of Hydrology, 275, 67-85, 2003.

Leon, L. F.: http://www.waterbase.org, last access: 25 November 2013.

Liong, S., Raghavan, S., and Vu, M.: Climate Change and Its Impacts on Streamflow: WRF and SCE-Optimized SWAT Models. In: Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. II), Springer, 2013.

Liu, J., Liu, T., Bao, A., De Maeyer, P., Feng, X., Miller, S. N., and Chen, X.: Assessment of Different Modelling Studies on the Spatial Hydrological Processes in an Arid Alpine Catchment, Water Resources Management, 30, 1757-1770, 2016.

Liu, Y. and Gupta, H. V.: Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, Water Resources Research, 43, 2007.

Lobligeois, F., Andréassian, V., Perrin, C., and Loumagne, C.: Can we improve streamflow simulation by using higher resolution rainfall information?, 2013, 11078.

Lobligeois, F., Andréassian, V., Perrin, C., Tabary, P., and Loumagne, C.: When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events, Hydrol. Earth Syst. Sci., 18, 575-594, 2014.

Marek, G. W., Gowda, P. H., Evett, S. R., Baumhardt, R. L., Brauer, D. K., Howell, T. A., Marek, T. H., and Srinivasan, R.: Calibration and validation of the SWAT model for predicting daily ET over irrigated crops in the Texas High Plains using lysimetric data, 2016. 2016.

Margane, A., Borgstedt, A., Hamdan, I., Subah, A., and Hajali, Z.: Delineation of surface water protection zones for the Wala Dam, Ministry of Water and Irrigation (MWI) and Federal Institute for Geosciences (BGR), Amman, Technical report Report No.12, 126 pp., 2009.

Ministry of Agriculture: National soil map and land-use project - Level 1. Amman. , 1994.

Monteith, J.: Evaporation and environment, Symp. Soc. Exp. Biol, 19, 205-234, 1964.

Moriasi, D., Arnold, J., Van Liew, M., Bingner, R., Harmel, R., and Veith, T.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, 2007. 2007.

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., and Döll, P.: Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration, Hydrol. Earth Syst. Sci., 18, 3511-3538, 2014.

Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I — A discussion of principles, Journal of hydrology, 10, 282-290, 1970.

Neitsch, S., Arnold, J., Kiniry, J., Williams, J., and King, K.: Soil and Water Assessment Tool theoretical documentation version 2009. Grassland soil and water research laboratory and blackland research center., Texas, 2011.

Nyong, A., Adesina, F., and Elasha, B. O.: The value of indigenous knowledge in climate change mitigation and adaptation strategies in the African Sahel, Mitigation and Adaptation Strategies for Global Change, 12, 787-797, 2007.

Ouessar, M., Bruggeman, A., Abdelli, F., Mohtar, R., Gabriels, D., and Cornelis, W.: Modelling water-harvesting systems in the arid south of Tunisia using SWAT, Hydrology and Earth System Sciences, 13, 2003-2021, 2009.

Ouessar, M., Bruggeman, A., Mohtar, R., Ouerchefani, D., Abdelli, F., and Boufelgha, M.: Future of Drylands – An Overview of Evaluation and Impact Assessment Tools for Water Harvesting. In: The Future of Drylands: International Scientific Conference on Desertification and Drylands Research Tunis, Tunisia, 19-21 June 2006, Lee, C. and Schaaf, T. (Eds.), Springer Netherlands, Dordrecht, 2008.

Özcan, Z., Kentel, E., and Alp, E.: Determination of unit nutrient loads for different land uses in wet periods through modelling and optimization for a semi-arid region, Journal of Hydrology, 540, 40-49, 2016.

Pilgrim, D., Chapman, T., and Doran, D.: Problems of rainfall-runoff modelling in arid and semiarid regions, Hydrological Sciences Journal, 33, 379-400, 1988.

Ragab, R. and Prudhomme, C.: SW—Soil and Water: Climate Change and Water Resources Management in Arid and Semi-arid Regions: Prospective and Challenges for the 21st Century, Biosystems Engineering, 81, 3-34, 2002.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., and Behringer, D.: The NCEP climate forecast system reanalysis, Bulletin of the American Meteorological Society, 91, 2010.

Saleh, A., Arnold, J., Gassman, P. W. a., Hauck, L., Rosenthal, W., Williams, J., and McFarland, A.: Application of SWAT for the upper North Bosque River watershed, Transactions of the ASAE, 43, 1077, 2000.

Sandholt, I., Rasmussen, K., and Andersen, J.: A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status, Remote Sensing of Environment, 79, 213-224, 2002.

Shrestha, M. K., Recknagel, F., Frizenschaf, J., and Meyer, W.: Assessing SWAT models based on single and multi-site calibration for the simulation of flow and nutrient loads in the semi-arid Onkaparinga catchment in South Australia, Agricultural Water Management, doi: http://dx.doi.org/10.1016/j.agwat.2016.02.009, 2016. 2016.

Spruill, C., Workman, S., and Taraba, J.: Simulation of daily and monthly stream discharge from small watersheds using the SWAT model, Transactions of the ASAE, 43, 1431, 2000.

Srinivasan, R., Arnold, J., and Jones, C.: Hydrologic modelling of the United States with the soil and water assessment tool, International Journal of Water Resources Development, 14, 315-325, 1998.

Ta'any, R.: Impact of Wala Dam on groundwater enhancement of Wadi Wala catchment area in Jordan, Indian Journal of Agricultural Research, 45, 255-265, 2011.

Tachikawa, T., Hato, M., Kaku, M., and Iwasaki, A.: Characteristics of ASTER GDEM version 2, 2011, 3657-3660.

Tarawneh, E.: Water and sediment yield for Wala Dam catchment area, MSc Thesis, Civil and environmental engineering, Mutah University, 106 pp., 2007.

Tessema, S.: Hydrological modeling as a tool for sustainable water resources management: a case study of the Awash River Basin, 2011. 2011.

Tingsanchali, T. and Gautam, M. R.: Application of tank, NAM, ARMA and neural network models to flood forecasting, Hydrological Processes, 14, 2473-2487, 2000.

Trondalen, J. M.: Climate changes, water security and possible remedies for the Middle East. In: United Nations Educational, Scientific and Cultural Organization, 2009.

USDA, S.: National Engineering Handbook, Section 4: Hydrology, Washington, DC, 1972. 1972.

USDA, S.: Urban hydrology for small watersheds, Technical release, 55, 2-6, 1986.

USEPA: Guidance for quality assurance project plans for modeling., USEPA, Office of Environmental Information, Washington, D.C. Report EPA/240/R-02/007, 2002.

Van Griensven, A.: Sensitivity, auto-calibration, uncertainty and model evaluation in SWAT2005, Unpublished report, 2005. 2005.

Wala Dam Management: Field visit to the Wala catchemment,. Bridge, J. and Tarawneh, E. (Eds.), 2013.

Wang, W., Xie, P., Yoo, S., Xue, Y., Kumar, A., and Wu, X.: An assessment of the surface climate in the NCEP climate forecast system reanalysis, Climate dynamics, 37, 1601-1620, 2011.

WaterBase: http://www.waterbase.org/home.html, last access: 20 November 2013.

Wheater, H., Sorooshian, S., and Sharma, K.: Hydrological modelling in arid and semi-arid areas, Cambridge University Press, 2008a.

Wheater, H., Sorooshian, S., and Sharma, K.: Modelling hydrological processes in arid and semi-arid areas: an introduction to the workshop. In: Hydrological Modelling in Arid and Semi-Arid Areas.(Cambridge University Press) Cambridge, 2008b.

Williams, J.: Flood routing with variable travel time or variable storage coefficients, Trans. ASAE, 12, 100-103, 1969.

Williams, J.: SPNM, a model for predicting sediment, hosphorous and nitrogen yields from agricultural basins Journal of the American Water Resources Association, 16, 843-848, 1980.

Williams, J. R. and Singh, V.: The EPIC model: Computer models of watershed hydrology., Water Resources Publications, USA, 1995.

Wischmeier, W. and Smith, D.: USDA Agriculture Handbook 282. Washington DC, US Department of Agriculture, 1965.

Wolffc, I. K.: Land-use change scenarios for the Jordan River region, Int. J. of Sustainable Water and Environmental Systems, 3, 25-31, 2011.

Wood, M. K. and Blackburn, W. H.: AN EVALUATION OF THE HYDROLOGIC SOIL GROUPS AS USED IN THE SCS RUNOFF METHOD ON RANGELANDS1. Wiley Online Library, 1984.

Wu, Y., Li, C., Zhang, C., Shi, X., Bourque, C. P.-A., and Zhao, S.: Evaluation of the applicability of the SWAT model in an arid piedmont plain oasis, Water Science and Technology, 73, 1341-1348, 2016.

Yen, B.: Criteria for Evaluation of Watershed Models, Journal of Irrigation and Drainage Engineering, 121, 130-131, 1995.

Zhang, L., Jin, X., He, C., Zhang, B., Zhang, X., Li, J., Zhao, C., Tian, J., and DeMarchi, C.: Comparison of SWAT and DLBRM for Hydrological Modeling of a Mountainous Watershed in Arid Northwest China, Journal of Hydrologic Engineering, 21, 04016007, 2016.

Zhang, X., Aguilar, E., Sensoy, S., Melkonyan, H., Tagiyeva, U., Ahmed, N., Kutaladze, N., Rahimzadeh, F., Taghipour, A., Hantosh, T. H., Albert, P., Semawi, M., Karam Ali, M., Said Al-Shabibi, M. H., Al-Oulan, Z., Zatari, T., Al Dean Khelet, I., Hamoud, S., Sagir, R., Demircan, M., Eken, M., Adiguzel, M., Alexander, L., Peterson, T. C., and Wallis, T.: Trends in Middle East climate extreme indices from 1950 to 2003, Journal of Geophysical Research: Atmospheres, 110, D22104, 2005.

TABLES

Scenario No.	Land-use map[1]	Soil map[2]	Weather data	Madaba station	No. of HRUs	NSE	RSR
16	c	b	CFSR	–	47	0.56	0.66
10	b	b	CFSR	–	67	0.56	0.67
4	a	b	CFSR	–	68	0.55	0.67
13	c	a	CFSR	–	48	-0.32	1.15
7	b	a	CFSR	–	63	-0.36	1.17
1	a	a	CFSR	–	60	-0.36	1.17
18	c	b	Local	Excluded	47	-0.36	1.17
12	b	b	Local	Excluded	67	-0.43	1.19
6	a	b	Local	Excluded	68	-0.69	1.30
17	c	b	Local	Included	47	-2.90	1.97
11	b	b	Local	Included	67	-3.16	2.04
5	a	b	Local	Included	68	-3.56	2.13
15	c	a	Local	Excluded	48	-4.69	2.38
9	b	a	Local	Excluded	63	-4.84	2.42
3	a	a	Local	Excluded	60	-5.39	2.53
14	c	a	Local	Included	48	-11.25	3.50
8	b	a	Local	Included	63	-11.42	3.52
2	a	a	Local	Included	60	-12.00	3.61

Table 1. Number of HRUs and values of NSE and RSR calculated for 18 scenarios for comparison of observed and simulated average monthly discharge (m³s⁻¹) at the Wala catchment outlet. 1 Land-use maps: a) Tarawneh (2007); b) Al-Bakri et al. (2013); c) WaterBase (2012). 2 Soil maps: a) WaterBase (2012); b) Tarawneh (2007).

Scenario No.	Land-use map[1]	Soil map[2]	Weather data	Madaba station	No. of HRUs	NSE	RSR
16	c	b	CFSR	_	47	0.79	0.46
10	b		CFSR	_	67	0.66	0.58
4	a				68	0.60	0.64
18	c		Local	Excluded	47	-0.11	1.06
12	b				67	-0.11	1.06
17	c		Local	Included	47	-1.67	1.63
11	b				67	-1.81	1.68
6	a		Local	Excluded	68	-2.97	1.99
5	a		Local	Included	68	-7.21	2.86
13	c	a	CFSR	_	48	-12.74	3.71
7	b		CFSR	_	63	-16.47	4.18
1	a				60	-22.70	4.87
15	c		Local	Excluded	48	-26.72	5.26
9	b				63	-36.01	6.08
14	c		Local	Included	48	-42.16	6.57
8	b				63	-48.98	7.07
3	a		Local	Excluded	60	-59.72	7.79
2	a		Local	Included	60	-85.06	9.28

Table 2. Number of HRUs and values of, NSE and RSR calculated for 18 scenarios for comparison of observed and simulated average monthly sediment yield (t/month) at the Wala catchment outlet. 1 Land-use maps: a) Tarawneh (2007); b) Al-Bakri et al. (2013); c) WaterBase (2012). 2 Soil maps: a) WaterBase (2012); b) Tarawneh (2007).

Name	Description[1]	Rank
CN2	Initial SCS CN II value (Curve Number)	1
SOL_AWC	Available water capacity (mm H_2O/mm soil)	2
SOL_Z	Soil depth (mm)	3
SURLAG	Surface runoff lag time (days)	4
ESCO	Soil evaporation compensation factor	5
CH_N	Manning's n value for main channel	6
ALPHA_BF	Baseflow alpha factor [days]	7

[1]   (Van Griensven, 2005)

Table 3. Results of parameters sensitivity analysis of scenario 16

Scenario No.	NSE (uncalibrated)	NSE (calibrated)
16	0.56	0.64
2	-12.00	-11.29

Table 4. Values of NSE calculated for uncalibrated and calibrated best and poorest-performing scenarios (16 and 2, respectively) for discharge simulation.

FIGURES
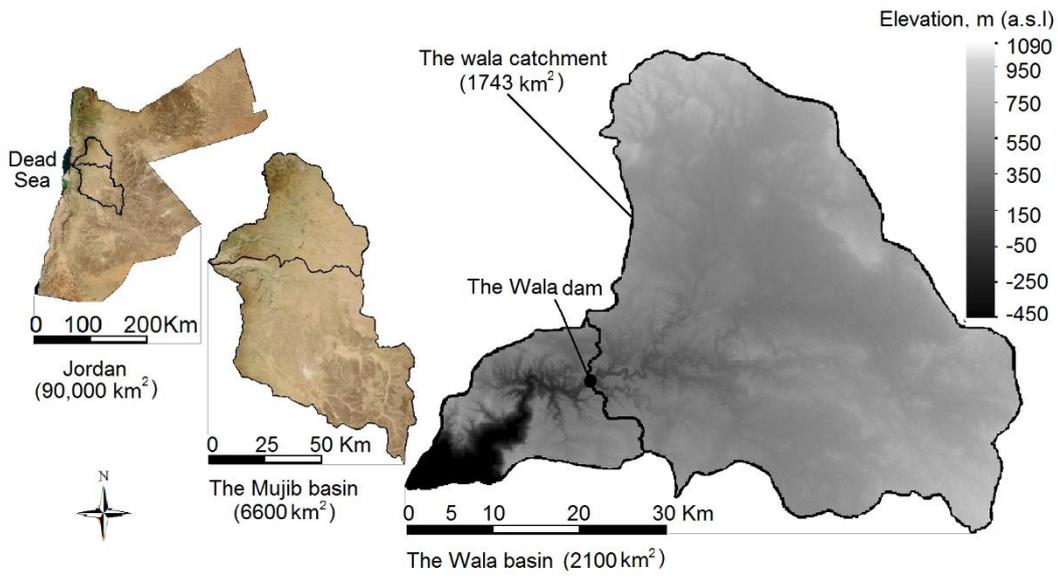
Figure 1. Location of the Wala Catchment.
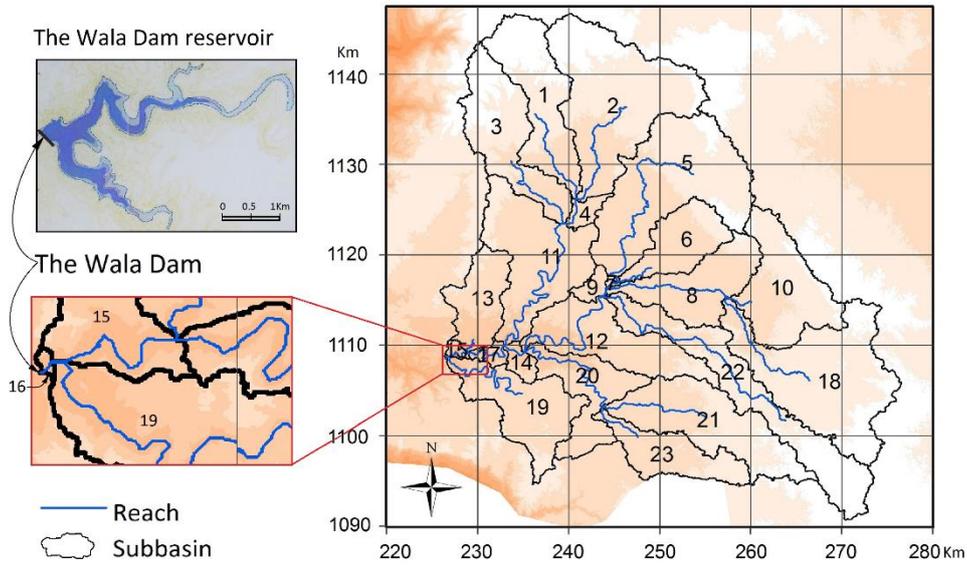
Figure 2. The Wala catchment delineation into subbasins, stream pattern and the catchment outlet.
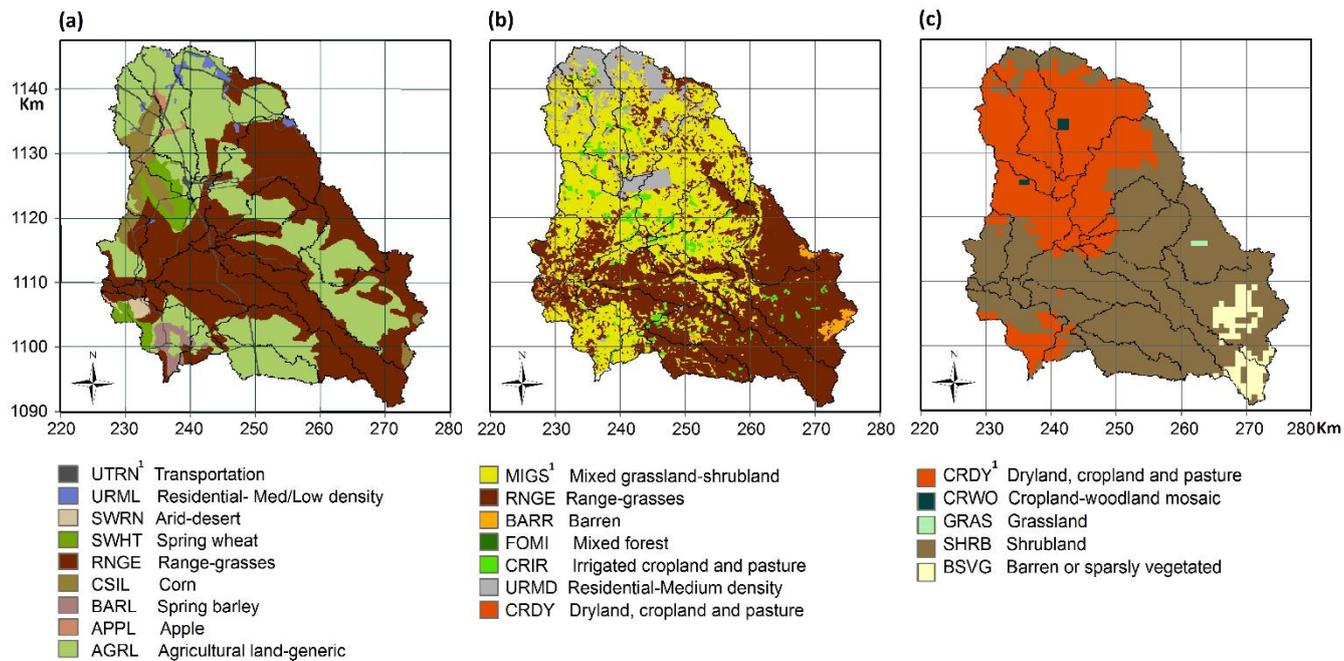
Figure 3. Land-use classification over the Wala catchment: a) Tarawneh (2007); b) Al-Bakri et al. (2013) ; c) WaterBase (2012). 1 SWAT land-use codes (Arnold et al., 2013).
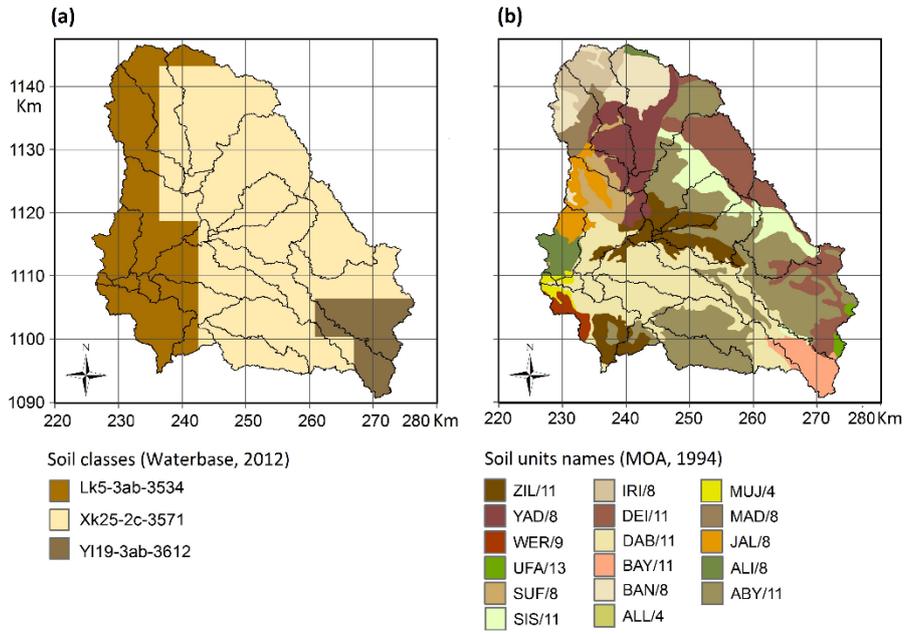
Figure 4. Soil classification of the Wala catchment: a) WaterBase (2012); b) Tarawneh (2007).
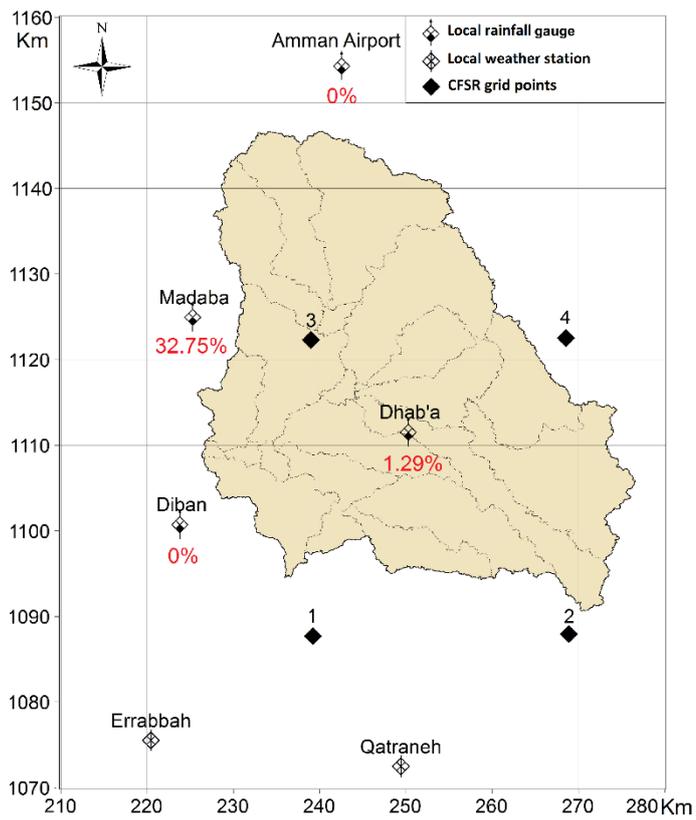
Figure 5. Local weather stations, CFSR grid points; and local rainfall gauges with their percentage of missing data over the period 1971-2002.
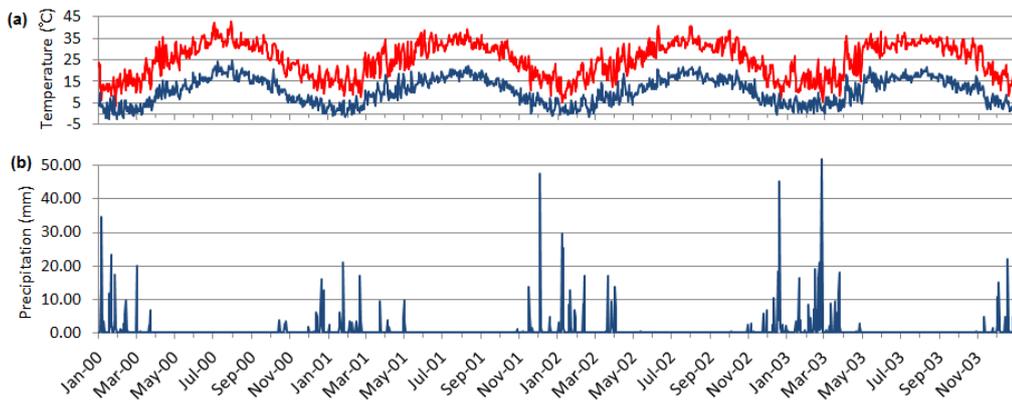
Figure 6. (a) Daily minimum and maximum temperature of Qatraneh station; (b) Daily precipitation of Amman Airport gauge, for the period 2000 - 2003.
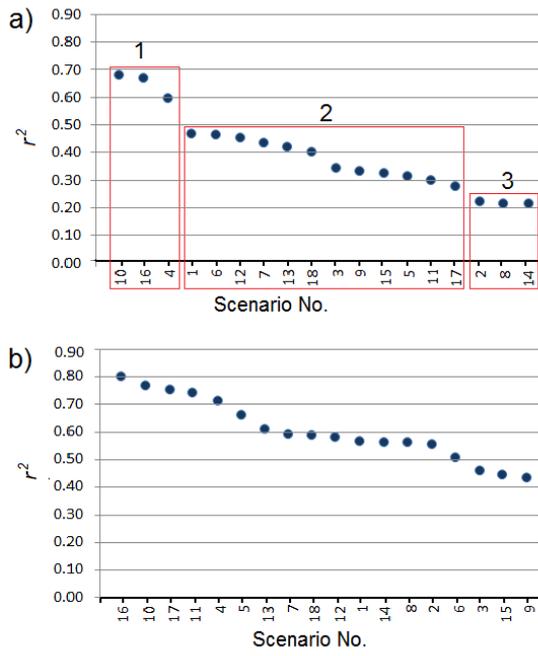
Figure 7. Values of r^2 calculated for prediction of a) discharge; b) sediment yield, from the 18 scenarios.
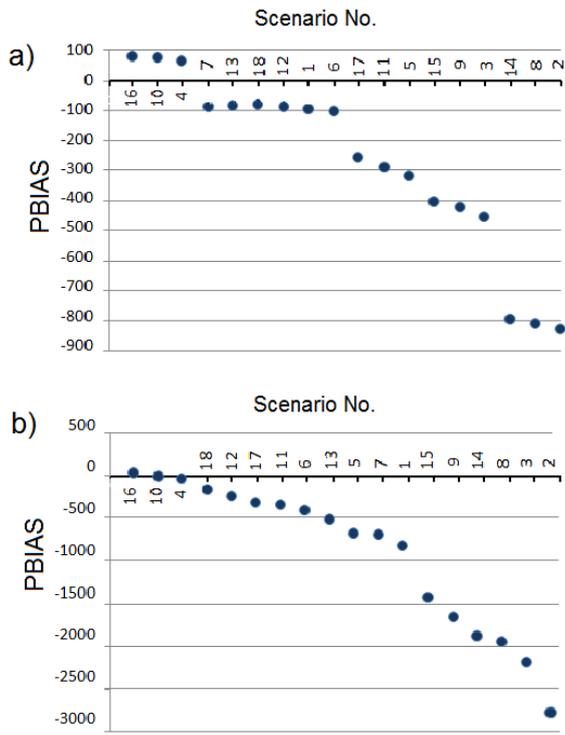
Figure 8. Values of PBIAS calculated for prediction of a) discharge; b) sediment yield, from the 18 scenarios.
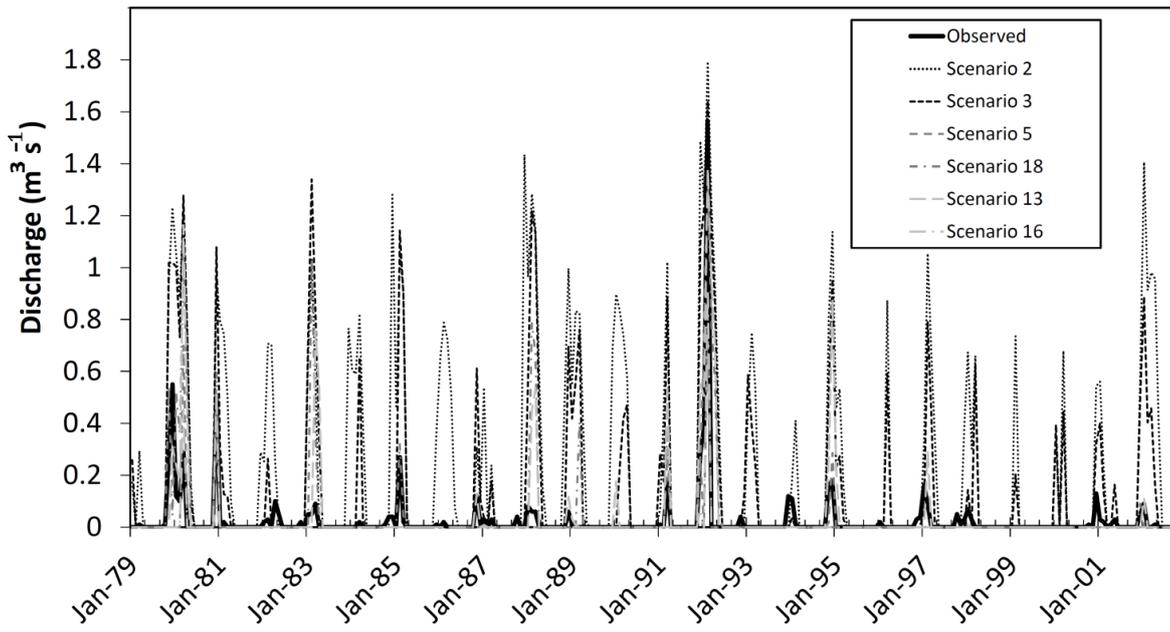
Figure 9. Simulated and observed average monthly discharge (m³s⁻¹) for scenarios 2, 3, 5, 18, 13 and 16.
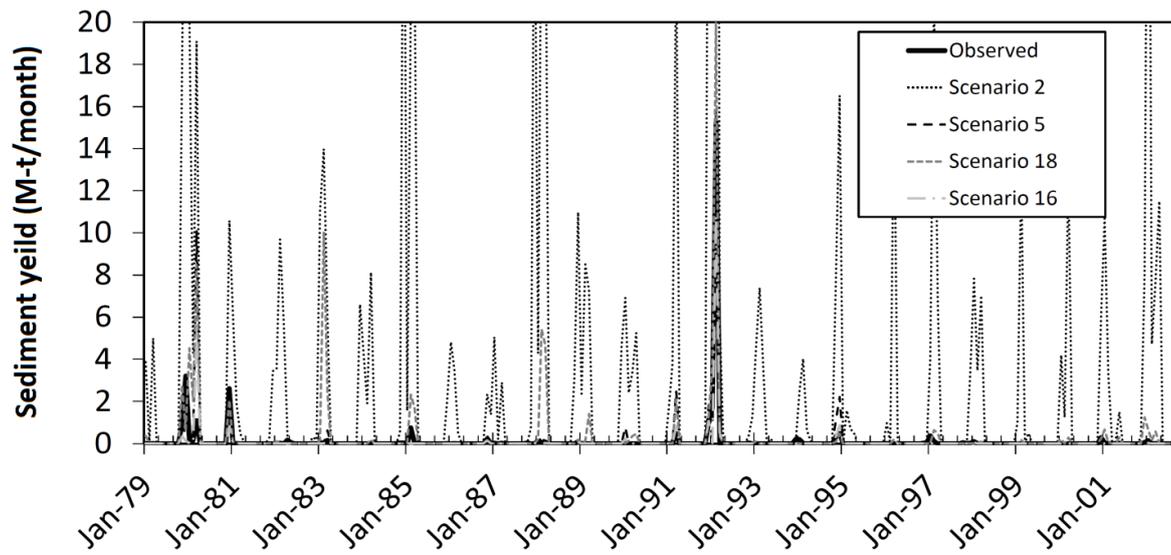
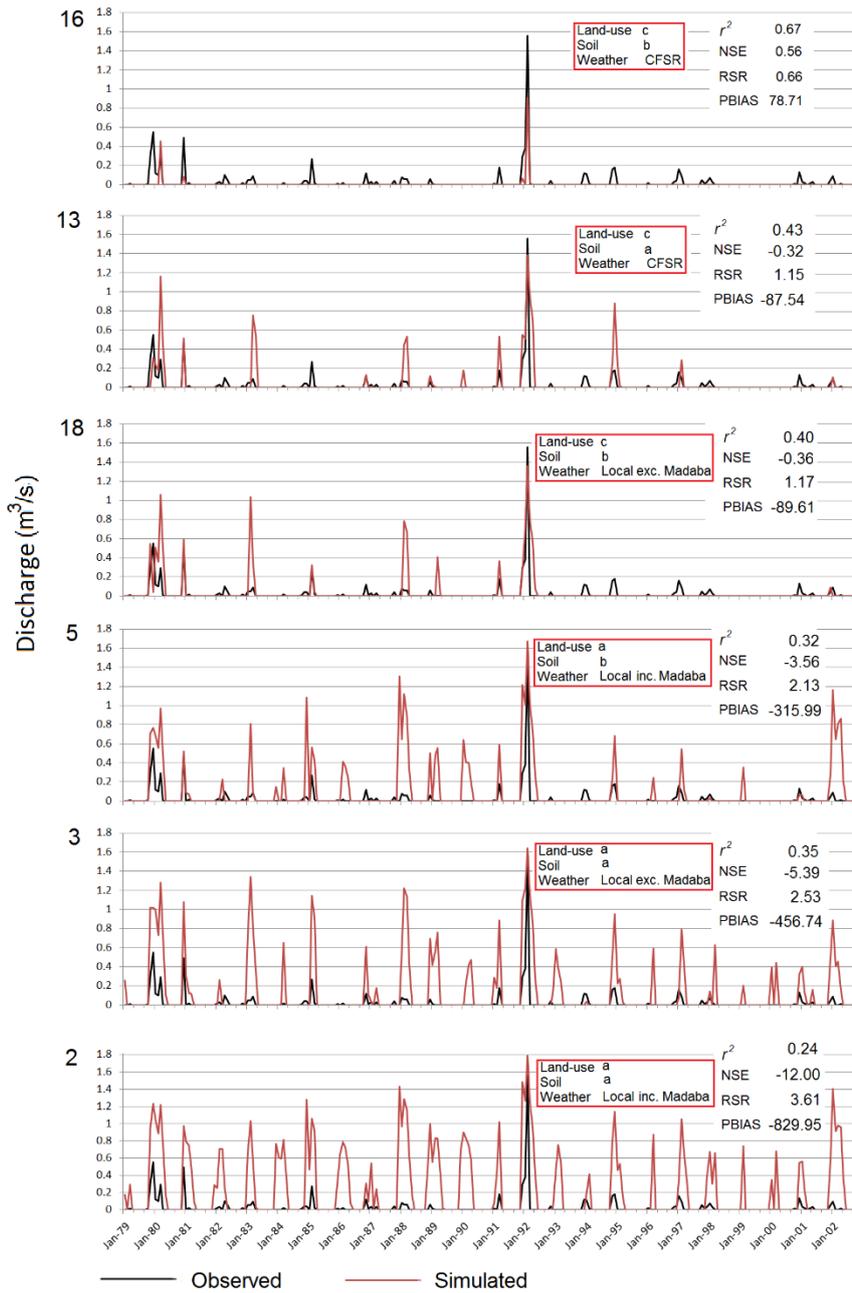**Figure 10. Simulated and observed monthly sediment yield (M-t/month) for scenarios 2, 5, 18 and 16.**

30

Figure 9. Simulated and observed average monthly discharge (m³s⁻¹) for scenarios 16, 13, 18, 5, 3 and 2 (arranged as model performance descends downward).
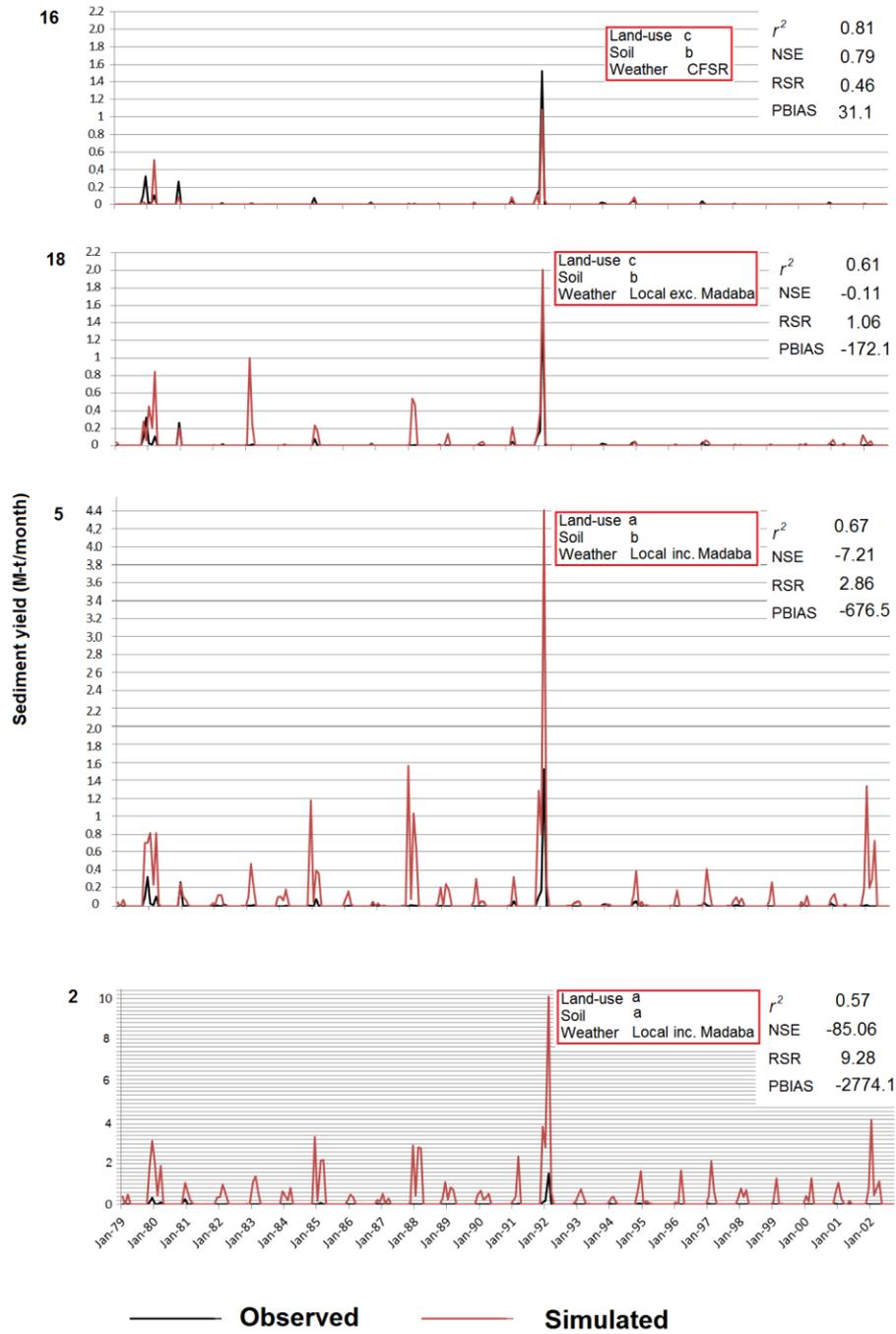
31

Scenario No.

Figure 10. Simulated and observed monthly sediment yield (M t/month) for scenarios 16, 18, 5 and 2 (arranged as model performance descends downward).
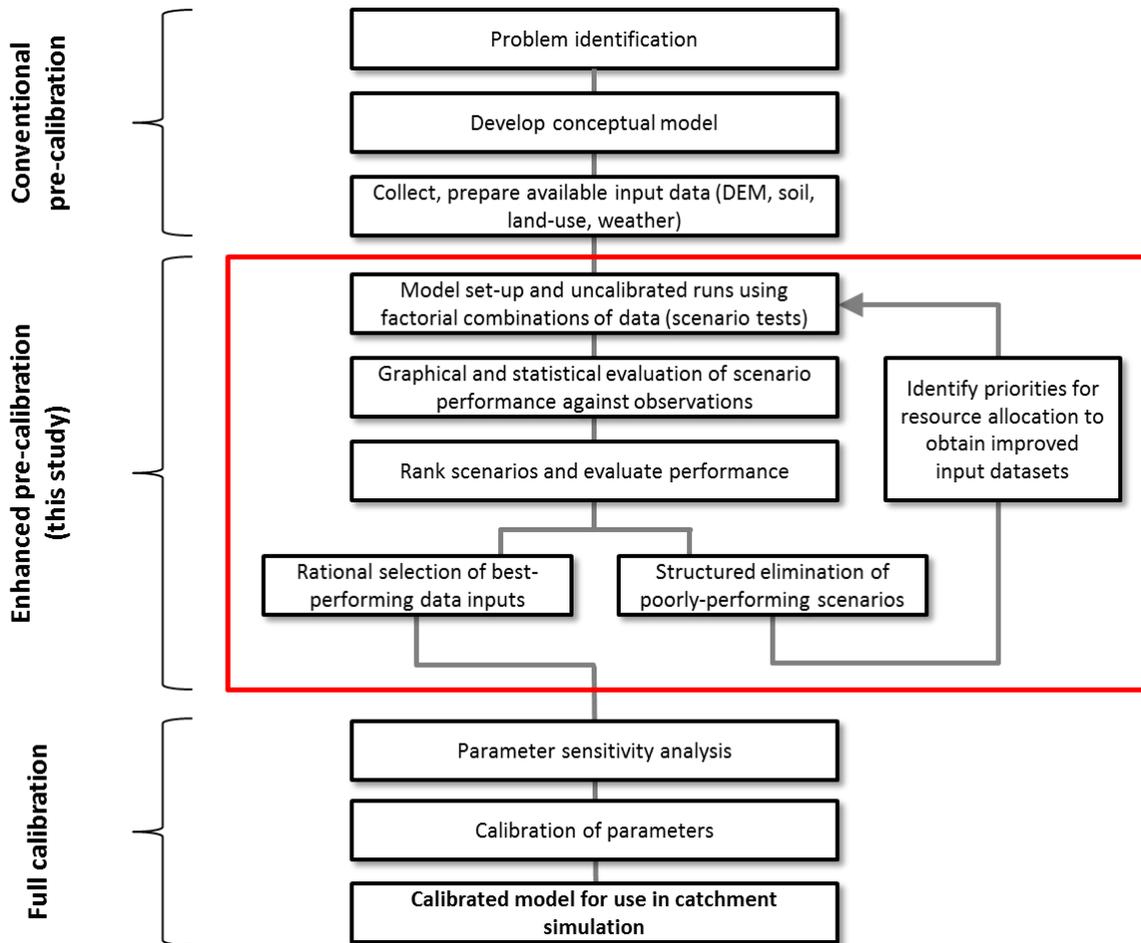
Figure 11. Workflow illustrating a generic pre-calibration approach based on the methodology outlined in this study.

33