Open Access

Earth System
Science
Data

Discussions

Interactive comment on "Vegetation, ground cover, soil, rainfall simulation, and overland flow experiments before and after tree removal in woodland-encroached sagebrush steppe: the hydrology component of the Sagebrush Steppe Treatment Evaluation Project (SageSTEP)" *by* C. Jason Williams et al.

Anonymous Referee #2

Received and published: 23 February 2020

The manuscript presents extensive data on numerous parameters characterizing surface and shallow subsurface hydrology at three locations within the western U.S. These data are concise and relevant for future hydrological and sedimentary analysis, and potential inclusion to various land surface models. The manuscript is available for download via the URL provided by the authors.

1. The description of plot scales should be consistent throughout the manuscript. In the Abstract, only 'overland flow' plots are mentioned explicitly; this changes to rainfall simulations at various plot sizes and overland flow plots in Lines 111-113, and finally to four plot scales in Lines 148-150, hillslope plots added. Besides, a small figure showing locations for each plot could be useful for non-U.S. readership.

2. This inconsistency is brought further to the text, Section 3, where field methods description starts with hillslope-scale plots, the largest, and continues with small- and large-scale plots etc. Though there might be a certain logic in such description order, I would suggest to follow either top-down or bottom-up approach.

3. Lines 287-288, the sediment concentration is said to be calculated from runoff samples by weighing; what is a 'runoff sample'? Is it a liquid volume - and if yes, was it just dried to full sample evaporation? If not, was any filtration system used, and if yes, then what were its parameters - pore size etc?

3. The dataset is well-organized, but several technical corrections are needed: 3.1. Data Dictionary - data types should be presented as standard notation, i.e. integer, real, character etc; same, variable sizes should be given, i.e. as INT/LONG INT/DOUBLE/CHAR(X) etc. 3.2. Categorical variables are multiple in the Data Dictionary, and are particularly poorly described; possible categories are listed as 'Acceptable values', which is not the best way to present them. No explanation on whet does, e.g. 'Tracked_LowMulch' mean, is given in the dataset itself. A separate table explaining your categorical variables is needed, or you might suggest a better way of presentation. 3.3. Same, 'Yes/No' is not a character variable, but has LOGICAL type, therefore acceptable values are 0/1, Y/N, or T/F, each is valid. 3.4. Dataset contains some info on treatment area and date, but I've found no clear descriptors for treatment type for each dataset in the plot characteristics table. This raises the question on whether the variables are correctly distributed between various dataset tables. 3.5. Table 3 contains no info on either plot type (small vs large vs overland etc) or plot area. 3.6. I find it difficult to browse through data with visual inspection, since: PLOT_ID is a

last column, e.g. in Table 4, and is hard to find in other tables as well; in several tables, PLOT_ID is not unique since two rows contain data for differerent years; treatment date repeats in Tables 3 and 4. In general, column sequence is not entirely logical, and can be enhanced.

The dataset structure, I believe, should be subject to technical inspection. I suggest the authors to read your dataset to R/RStudio environment and check dataset usability / statistical analysis performance.

_____

C3