Hydrology and
Earth System
Sciences
Discussions

Interactive comment on "Estimating water residence time distribution in river networks by boosted regression trees (BRT) model" *by* Meili Feng et al.

Anonymous Referee #1

Received and published: 4 December 2018

General Comments: This paper provides a new method, boosted regression trees (BRT) model, to estimate water residence time. 82 river reaches in Germany were used to assess the method by using an average of the 2008-2014 discharge data. Parallel studies for an extreme flood event and a dry month were assessed, although one main comment is that the specific analysis for these hydrological conditions needs to be clarified. Although this method has the potential to estimate water residence time more easily and with limited data, the advantages over current methods should be further emphasized within the text. Specific comments to clarify the methods and results and suggestions on how to redirect the discussion are included below in order to improve the overall message and strengthen the paper.

Specific Comments: 1. Clearly explain how this method is an advantage from current models to estimate water residence time and add this information throughout the text and in the abstract. a. At the end of introduction (L74) add the main goals of the paper and model analysis. L49-51 states that there is still a huge gap between detailed process-based models and over simplified empirical methods. State something similar at the end of the introduction and/or mention a main objective stated in the discussion (L190-191) that the author's aim to use predictive variables to facilitate the empirical estimation of WRT in a river on the basis of generally available information. b. Include in the abstract how the model was validated.

2. Improved explanations are needed of the main model equations, model inputs and outputs. Some specific examples include: a. Add a statement to justify the approach of model validation with the average discharge (2008-2014). b. L113-114 What information does this analysis provide? Although it may not be possible to fully describe the model, it is important to provide the reasoning for why and how partial dependence plots and fitted link functions for each variable operate and what information they provide. Basically, it is important to explain model inputs, a general description of what the model does and how it interprets and processes the data, and then a full description of the model output. The description on L115 is quite brief and it would be helpful to add here more information on how WRT is calculated. c. Equation 1 is WRT predict? This is stated on L152 but should be specified near to the equation to avoid confusion. d. L125 How is WRTobs originally calculated? This is important because it has some inherent bias in what the most important parameters/dependent parameters. e. L133, L138 (& previously in methods) further define Euclidean distance and "sum of all trees multiplied by the learning rate". Include a concise statement on each term and how each support the analysis. f. L182 states that parallel studies for extreme flood and dry months of 2 specific months. Was the same analysis conducted as for the average? Clarify that Figure 9 is WRTpred and if this value is compared to WRTobs as was done for the average.

3. Redirect the discussion to improve focus on the main results presented within the article: a. In order to link to the presented results, the following topics need to be further discussed 1) impacts of drought and flood events on WRT and implications of these hydrological conditions (i.e., extreme) and 2) how geomorphology attributes are more influential on small rivers and where the model capability was both best and worst at WRT estimates and why. The discussion should link to the main purpose of this analysis that is clearly stated at the end of the introduction (see comment #1). Elaborate on L198-199 WRT can be estimated even at low flows or based on this study only at high flows are accurate? b. Section 4.2 – it isn't obvious after reviewing the introduction through results why this is a main discussion point. Related to this point, the figures within the discussion should be removed or improved justification is required to include them but with a recommendation to move these into the results section.