Response to comments of anonymous Referee #3

We thank the anonymous referee #3 for reviewing our manuscript and his suggestion for improving the temporal occurrence model of preferential flow (PF). We answer below to each comment in a point-by-point reply. For clarity, the comments of the referee were copied in black and our comments are in blue.

General Comments

Primarily the structure and selection of results should be reconsidered, but also a more defined storyline could assist the reader to extract the main novelties of this study. In general, the manuscript could benefit from reconsidering what information is necessary to broadcast the main message. I recommend to consider a few key figures that conveniently show the reader the approach and main interesting findings, instead of a long sequence of tables and graphs. Lastly, the readability would greatly increase if the authors consider a key phrase in each paragraph that, perhaps almost trivially, highlights what should be learned from the given information.

For the structure of the paper, I would recommend to consider separation of the hypothesis and throughout the paper clearly indicate which sections address information for which hypothesis. I miss this in the paper. The hypothesis could possibly be broken up in two sections. For example: 1) PF is the dominant process during infiltration, and 2) capillary theory does not suffice to explain infiltration. These can be tested for the given explanatory factors, such as land cover, geology etc. Which also gives more structure in the result and discussion section.

We will restructure the results and focus on the main findings, the temporal dynamics of PF in different large scale spatial units (landscape units). By removing most of the small-scale spatial analysis we will highlight the main storyline. Additionally, we will add introduction phrases and summaries for the single sections. We will clearly address which section of the results and discussion contributes to which hypothesis.

The generalized linear model (GLM) provided insight in the explanatory power of a large set of variables. However, as anonymous referee #2 addressed, there are some limitations to this approach. I will not re-evaluate these points, but I instead would recommend the authors to consider the use of mixed effect models. This approach allows to include random factors that potentially explain variability but are not directly incorporated in the study design. Seen the authors use R, the packages ‘lme4’, ‘lmerTest’, and ‘nlme’ could relatively easily allow to explore the use of mixed effect models.

We thank you for the suggestion and have changed the GLM to a generalized linear mixed effect model (GLMM) that incorporates the spatial site information as a random effect (see response to RC1 and RC2).

On a final note, I wonder if there is any indication that the contributing area of each site is independent of the occurrence of NSR? A correlation could guide towards rising groundwater tables and associated capillary rise, or horizontal flow. Especially with high antecedent soil moisture groundwater response could be relatively fast when contributing area is large.

We have calculated the upslope contributing area for each site and compared it against NSR occurrence. The Spearman R is 0.1 and hence, influence of groundwater in the upper 0.5 m of soil seems to be small for our sites. This is supported by the fact that % NSR and distance to stream does not show a correlation (P15L1).

Specific Comments

P2L17 Seems out of context to mention hotspots or hot moments, especially as a final statement of the section. The statement needs further elaboration and references.

We agree that the sentence is out of context at the end of this section and removed it.

P2L26 ’…scale (~ km2) and’ Is this referring to 1 km2 to be considered large scale, or is a number missing?

It means “on a kilometer scale”, and is considered to be large scale for PF, since spatial and temporal information on PF occurrence is usually only known on a plot scale (centimeters to meters).

P3L11 This section seems out of place, considering reorganizing with earlier paragraphs covering methods.

We will move this section to a different paragraph.

P6L1 Appendix A: consider presenting standard errors of the K measurements

We will add the standard error.

P8L15 How can observations at a single depth be considered sequential?

A sensor response in 10 cm only, is neither "no response" nor a "non-sequential response". Since we know that the upper sensor reacted first during a 10 cm only reaction (similar than during a sequential response), we included it to the category "sequential responses".

P11L5 I would start with the most interesting finding of this study, although it could be strictly seen as a result, I could see this information to be more suited in the methods section.

We will move the analysis of the rainfall event separation to the methods.

P24L10 The range of reported flow velocities both in this study and other reported studies generally seems extremely large. If the range is large to begin with, how is it remarkable that they fall in the same range? Perhaps I miss a part of the reasoning.

We will clarify the sentence. It is not remarkable that all these different soils fall in the same large range, but it is remarkable that they all show the same large variation. In other words there is no systematic difference of flow velocity with a certain type of soil and independent of the soil one can observe flow velocities from a few cm/day up to ~100000 cm/day.

P25L15 Awkward sentence structure.

We will rephrase the sentence.

P25L29 Although this seems like an insightful comment, are there any examples how this could be implemented, or is it readily tested on small scale? A reference would be useful here.

The hillslope parametrization of Loritz et al. 2017 (doi:10.5194/hess-21-1225-2017), which includes parametrization of fast flow path based on macropore field observations, could be seen as a recent example on a larger scale. The reference will be added.