Response to comments of Referee #1 Heye Bogena

We thank Heye Bogena for reviewing our manuscript and for the helpful suggestions for improving the study. 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 response is in blue.

General Comments

Despite the novelty and interesting approach to studying preferential flow, I think this manuscript is not yet ready for publication and would benefit from a more clear description of methods (e.g. using flow charts) and from focusing its content to the most interesting parts.

The number of statistical analyses is rather excessive without providing much additional insights. For instance, the generalized linear regression model analysis is difficult to comprehend (especially for readers who are not familiar with this method) and does not provide clear results.

We will add a flow chart in the method section for a clear description of the analysis. Instead of analyzing small scale spatial patterns, we will focus on the temporal dynamics of soil moisture or rainfall characteristics of large scale spatial units (landscape units). We will remove information and statistical analysis in the text that is not supporting the main findings.

We agree that the methods section and explanation of the generalized linear model (GLM) is rather short. The current GLM results give little additional insight and was also criticized by the other reviewers (also in terms of sample size and pseudo-replicates). The motivation of the GLM was to derive the probability of an event to be a preferential flow (PF) event (temporal information), additional to a linear regression model (see e.g. Liu & Lin 2015, doi:10.2136/sssaj2014.08.0330), that can provide information about the spatial PF occurrence (e.g. % PF at one location). These probabilities can be very useful for soil hydrological models to estimate not only where, but also when to include PF. Therefore, we still think that the general intention of the GLM is valuable and we will modify and simplify the temporal PF occurrence models by using a generalized linear mixed effect model (GLMM, see RC2 and RC3), only focusing on the relevant temporal predictors (initial soil moisture and rainfall). Additionally, we will explain the method, data and benefits of the model in more detail. The applied GLMM shows a good agreement with observations using the spatial information (45 sites) as a random factor. If we know the site we can estimate the probability of the event to be a PF event just from soil moisture and rainfall data (conditional $R^2 \sim 0.7$; method of Nakagawa & Schielzeth 2013, doi: 10.1111/j.2041-210x.2012.00261.x). We think this is an interesting result since it shows that we can predict PF from temporal information if we know how the landscapes or soils behave. However, these random site factors that influence PF in the GLMM are probably small scale landscape properties at the individual sites that are unknown up to now.

Also the discussion section is too excessive should be focused on the most important results found in this study.

The result, discussion and conclusion section will be reorganized and shortened to focus on the main findings. We will also better connect the different sections to the hypothesis und thereby improve structure and readability.

A new aspect of this study is usage of calculated of water flow velocities on basis of 1D steady state flow assumptions to identify preferential flow events. However, although the velocity-based preferential flow assessment of events with a sequential order of sensor response times is appropriate, the derived matrix flow velocities may be prone to large errors. First, the steady state assumption is violated during infiltration events and second, the hydraulic parameters derived from field data which are influenced by preferential flow.

(1) Steady state assumption: During infiltration events a steady state is indeed not what we observe in nature. However, instead of a computationally intensive Richards based numerical 1D solution of all events, which also suffers from the uncertainty of parameters, the boundary conditions or discretization, we decided to use a steady state assumption of unsaturated flow. We tried to account for error that is based on the steady state assumption by using the maximum gradient during the event and are thus overestimating the driving forces. The hydraulic conductivity was previously calculated based on the median water content between the event begin and peak soil moisture. To be on the safe side, we will now change it to the maximum (peak) water content of the upper sensor. Hence, we overestimate matrix flow velocity rather than underestimating it, leading to a conservative estimate of preferential flow occurrence. Even though the absolute value is overestimated, it provides a maximum matrix flow velocity that can be used for the comparison to the magnitudes of measured flow velocities. We will clarify this in the manuscript.

(2) Parameters: The parameters will not be completely unaffected by PF. However, as we wrote in the manuscript the parameters are derived on a daily timestep over many years (including dry phases) and thus do not include soil moisture dynamics on event base (e.g. hours). This prevents optimization of the parameters to be able to account for fast flow in the range of 1000 cm/hour. Therefore, the parameters were seen as a valuable alternative to pedotransfer functions for estimating matrix flow. Additionally, the potential influence of fast flow in the estimated retention parameters of Sprenger et al. (2016) will lead again rather to an overestimate of matrix flow velocities and thus underestimate the frequency of occurrence of preferential flow.

These circumstances may also explain why the measured NSR is lower for Marl grassland sites than predicted by the 1D steady state flow model.

For the hypothesis testing of NSR ($P_{int} > K_{mat}$) in the Marl grassland the hood infiltrometer K_{mat} values were used (not the predicted values by the 1D steady state flow). We apologize for the unclear description and will clarify the methodology of the different analysis in the revised manuscript.

Previous sensor based preferential studies (e.g. Wiekenkamp et al., 2016) used infiltrometer measurements to assign a meaningful threshold for the matrix flow velocity. Given that fact that hood infiltrometer measurements are available for the study area I suggest to use this data to define maximum matrix flow velocities for the different landscape units.

We will use the hood infiltrometer data on maximum matrix flow velocities as an additional method to estimate fast preferential flow and will add this to the analysis. However, using matrix saturated hydraulic conductivities for dividing into a fast/slow response only yields a binary information that results from one static threshold for all conditions. Therefore, we will also keep our analysis with the individual thresholds for every event (based on the 1D steady state flow). This allows for demonstrating the magnitude of deviation from matrix flow only and hence allows us to estimate the extent of impact of PF on the flow velocities.

Specific comments:

P3L5: Change to "...PF was more frequent during higher rainfall intensities.

We changed the sentence as suggested.

”P4L7: Change to “. . . and that therefore infiltration . . .

We changed the sentence as suggested.

”P4L16: Add number of sites.

We added the number of sites to the section.

P5L5: Change to “. . . mostly exhibit loamy texture.”

We changed the sentence as suggested.

P5L5-6: Sentence reads awkward. Please reformulate and add landuse percentages.

We reformulated the sentence and land use percentage was added.

P5L6: How was the macroporosity defined/determined?

We specified this in the text. We counted visually observable pores > 2 mm in diameter found by digging of horizontal soil profiles.

P5L15: Change to “METER Group Inc., USA”

We changed the sentence as suggested.

P6L8-9: You are actually calculating a “difference” and not a “macropore portion” of saturated hydraulic conductivity.

We agree and changed it to “macropore fraction of saturated hydraulic conductivity”.

P7L6: Why 6.7 mm?

The threshold was based on an intensity measurement of 80 mm/h ≈ 6.7 mm/5 min.

P7L10-11: This should be reformulated in a less unconditional way, e.g. “The manufacturer gives an accuracy of...”

We changed the sentence as suggested.

P7L20: You should change “diel” into “diurnal” as this term is more common.

We changed the sentence as suggested.

P7L24-25: But later you decide to use 12 hour rainfall breaks, which is somewhat confusing. In order to make this more comprehensive I suggest to present the rainfall event delineation methodology completely in the method section (thus moving P11L5-15 to the method chapter).

Rainfall events were divided by 12-hour breaks, but soil moisture was tracked for additional 48 hours after the end of a rainfall event. However, we do agree that the different time steps for event definitions are confusing and will move the mentioned section (P11 L5-15 and Table 2) to the methods.

P8L6-8: This is not clear to me (e.g. what is the meaning of "0.4 %"?). Please be more specific.

We clarified the unit (0.4 Vol.%).

P9L08-15: There exists a vast literature on showing that preferential flow cannot be described with classical capillary theory. Why do you still pursue this analysis although this approach is obviously prone to fail?

Good question, however, models based on classical capillary theory are still very common. Anyhow, many people still have the misconception that preferential flow occurs only at saturation or at rainfall intensities exceeding infiltration capacity. We wanted to reiterate that this is not the case while at the same time the motivation was to see the how often classical theory fails, how strong the deviation is and if there are differences between different landscape units (probably some regions could be described by matrix flow). We will provide a clarified explanation in the revised manuscript and will further simplify the comparison of capillary theory and measurements.

P9L11: The term "pore water pressure" only applies to subsurface water. Ponded water on the soil surface shows even positive pressures.

We agree and changed the sentence.

P9L13: better: "during a rainfall event infiltration capacity decreases as the soil approaches saturation"

We changed the sentence as suggested.

P9L27: Why should the aspect has an effect on the frequency of NSR occurrence?

Some studies found an effect of the slope aspect on the soil water properties (e.g Geroy et al. 2011, doi: 10.1002/hyp.8281). However, to focus on the main findings of our study this section was removed.

P10L4-5: better: ...as the velocity determined from the first responses of two sensors"

We changed the sentence as suggested.

P10L9-11: During infiltration events soil water flow should be governed by non-stationary conditions. Why do you believe that your stationarity assumption can be applied?

Please see our explanation under "General Comments".

P10L19-20: What does "in combination with parameter sets of Sprenger et al. (2016)" exactly means?

The formulation is confusing and was changed. van Genuchten-Mualem equation was used and parametrized with the parameters from Sprenger et al. (2016).

P10L23-26: Nevertheless, as the inversely derived Ks-parameter of Sprenger et al.(2016) are derived from field data, they will still be affected by preferential flow and thus will be higher compared to a Ks derived from pure matrix flow.

Please see our explanation under "General Comments".

P11L22: Either use proportion or percentage

We changed the sentence.

P12L13: Add explanations of the abbreviations

We added an explanation of the abbreviations.

P13L18-26: This section comes somewhat out of the blue as it is not well related to the previous analysis. I suggest separating both sections and adding a short introduction to new one concerning soil water content changes.

We separated both sections and added a short introduction as suggested.

P13L20-21: This is difficult to understand. Please try to rephrase in a more comprehensive way.

We rephrased the sentence.

P15L2-3: This is an interesting finding. Does this correspond with seasonally varying precipitation properties?

The NSR pattern corresponds to the seasonal pattern in maximum precipitation intensity (highest from June – September) and soil moisture (lowest from July – October). We will add a graph to Figure 4 showing this.

P15L5-6: What are the possible reasons?

This result is discussed on P22L21-29. Higher macroporosity, stemflow or hydrophobicity are possible reasons.

P16L6: The formulation "of up to~25 % of events" can be misinterpreted.

We changed the phrasing.

P16L10-11: Does this finding indicate that the inversely derived Ks-parameter of Sprenger et al. (2016) overestimate pure matrix flow?

Please see the explanation under "General Comments".

P17L1-17: I find this section not very meaningful as it cannot be well reproduced and the results are not very much enlightening. Therefore, for the sake of comprehensibility I suggest removing it.

Please see the explanation under "General Comments".

P18L11: It should "increasing" instead of "decreasing"

We apologized for this unclear phrasing. We meant "velocity is decreasing with decreasing water content." This was corrected.

P18L15: I guess it should be again "increasing" instead of "decreasing"

Again we apologize for this phrasing and changed it accordingly.

P19L10: It would be interesting to see how choosing other rain gaps would influence the results (e.g. the proportion of NSR to SR events).

This was included in an earlier version of the manuscript but was removed since it made the results rather complicated and it did not add any additional information. The changes in the proportions (NSR/SR) of the reactions are relatively small (8% for Marl forest, ± 3-5% for all other landscape units), increasing with longer rain gaps for most landscape units (comparing for example 6h, 12h, and 24h rain gaps). However, there is not a clear trend of increasing proportions with longer rain gaps for all landscape units. In general the number of rain events is decreasing with longer rain gaps and events last longer (see Table 2). This leads to a decrease of soil moisture events without a reaction (NR), while SR and NSR are increasing. However, the patterns between the landscape units stay similar.

P21L26: Did you compare pre-response analysis with entire event analysis? This would be interesting with respect to the comparison with the other studies.

Similar patterns are observed using total event rainfall amount or maximum rainfall intensity of the entire event. We added a sentence to give this information for comparison. However, since the response classification is not affected by the rainfall amount or intensity after the first soil moisture sensor response we keep the pre-response in our analysis and figures.

P21L28-29: I cannot follow this argument. Please explain in more detail.

The argument will be clarified. In the sandstone grassland PF seems to be more often initialized at higher initial saturation, simply because infiltration capacity is lower and saturation is achieved faster compared to dry conditions. This is in contrast to the other landscape units with higher clay content, where more NSR is found under dry conditions with soil structure formation or hydrophobicity being the driving mechanism. We will clarify the argument.

P23L7-11: In my view, the results of this study rather suggest that the occurrence of preferential flow is governed by unresolved small-scale structures and processes. The study of Wiekenkamp et al. (2016) used an even denser soil moisture sensor network and still could not find landscape properties to explain their results.

We totally agree. However, these small-scale structures and processes can probably be attributed to landscape properties. Different combinations of the landscape properties could lead to similar flow reactions making it hard to distinguish. We hypothesize that due to the high heterogeneity of soils it would need much more sensors (even more than in Wiekenkamp et al. (2016)) to identify them. We rephrased the section.

P23L17-18: Why should the lower k_mat values of the Marl site lead to more NSR events and how do you know that the matrix infiltration capacity was underestimated?

See P9L13-15. The saturated matrix hydraulic conductivity (K_{mat}) was estimated using the hood infiltrometer that corresponds to the infiltration capacity at full saturation. We did not measure infiltration capacity at various moisture contents. Since the infiltration capacity is increasing with lower initial soil water content we rather underestimate the infiltration capacity under field conditions using K_{mat} (because soils are rarely saturated in our catchment). P9L13-15 was moved to a different section to clarify this.

Our capillary-based estimation of NSR is again a conservative approach using this minimum infiltration capacity (K_{mat}). NSR is overestimated, because the infiltration capacity is underestimated (the threshold is more often exceeded then with a higher infiltration capacity). However, to have a clearer structure of the study (especially results) and focus on the main analysis we will remove the comparison of the observed NSR responses with the estimated preferential flow reaction based on matrix hydraulic conductivity.

P23L18-19: Why should the overestimation of NSR by capillary theory in the Marl grassland be an indication of more vertical macropore flow?

We estimated more events with infiltration capacities (in our case K_{mat}) lower than maximum rainfall intensities in the Marl grassland and hence we should observe PF. Since we measured NSR less frequently than estimated by this approach, we hypothesized that these events have probably not resulted in PF with a NSR (non-homogenous flow), but rather in fast SR, which is supported by the high wetting front velocities. As correctly stated by the referee it has not to be vertical, but we do not observe a break in the sensor reaction sequence. We clarified the section.

P24L23-28: These arguments are rather dubious.

We agree that these arguments are rather vague and speculative, hence the section was removed.

P24L23-P25L11: This section is a rather excessive discussion that does not provide much additional insights.

We partly removed this section and some sentences were moved to a different section. In general, the discussion will be restructured and will focus on the main findings.

P24L29: I guess it should be “increase” instead of “decrease”.

We apologize again for causing confusion by our repeated erroneous phrasing.

P25L2-4: Remove repetition.

We removed the repetition.

P25L16: “. . . showed . . . ”

We changed the sentence as suggested.

P25L17: “. . . in deeper . . . ”

We changed the sentence as suggested.

P25L20: “showed” instead of “had”

We changed the sentence as suggested.

P25L21: You did not prove the occurrence of “non-homogeneous wetter fronts”. There are other mechanisms that can lead to preferential flow (e.g. by-pass flow).

We used NSR only as a proxy for preferential flow. However, we think that NSR in the first place only proved that there was a non-homogeneous wetting front that could be generated by preferential flow (see P8L22-23). This non-homogeneous wetting front can be generated by various preferential flow process (including by-pass flow). We have clarified that.

P25L23-P26L7: Please focus on presenting the main results in the conclusion section and avoid vague speculations.

We will rewrite the conclusion section and will focus on the main findings.

Figures

Figure 9: Dots are difficult to discern.

The figure will be changed.