This paper tried to evaluate the impact of rainfall features and antecedent soil moisture on occurrence of preferential flow on slope in north China by interpreting response of soil moisture to rainfall. The result showed that occurring frequency of preferential flow was averagely 40.7. Nevertheless, it is my feeling that the authors did not stress enough the limitation of previous researches and their relations with the major objectives. And I also concern about the method that the authors used to analyze the correlation between rainfall features, antecedent soil moisture and frequency of preferential flow. It is not clear to me that how each rainfall feature was divided into 15 sub-ranges with non-uniform increments (P8L5). In my opinion, the method would determine the fitting curves and R2, and so the results depend on the inclination of the authors. Therefore, the authors should provide more explanation. As a general comment, I think that the paper requires major revision before being published.

Reply: Thanks a lot for your comments.

We will illustrate more to fill the gap between previous studies and our objectives. Basically, this study was initiated from two considerations. (1) It would be helpful in understanding the processes of subsurface hydrology, if we get the key factors that control the occurrence of preferential flow. Lots of studies have been carried out on this topic. However, contradictory results were obtained in different cases, e.g., the cases of Wu et al. (2014) and Hardie et al. (2013). And to our knowledge, no study on this topic has been carried out in northern China with sub-humid climate and poorly developed underlying soil. Hence, we think this study could be a complementary to the understanding of controlling factors of preferential flow; meanwhile, it would be helpful in understanding hydrological processes of the study area. (2) By far, there are many methods for the detection of preferential flow, but in-suit method is rather limited. The method using wetting front as an indicator, which was proposed by Lin and Zhou (2008) and later improved by Hardie et al. (2013), could be an alternative option. Since this method has been on applied in only two or three cases to our knowledge, it would be of interest to apply it in our study area, where climate and surface condition are different from previous cases.

We agree that our way to set sub-ranges is not absolutely robust. Nevertheless, Given that the values of features are not uniformly distributed, we think it's inappropriate to divide the sub-ranges by uniform increments. If the increment were small, only one or even none of the data points would be contained by the sub-ranges with high values, and statistic of frequencies in these sub-ranges would be meaningless. If the increment were large, only one or two sub-ranges could be obtained at low values, the increase trend of frequency would be less visible, and some key information may be lost. In this circumstance, we think the non-uniform sub-ranges would be a reasonable choice. In order to compare among the features, values of the 4 features are all divided into the same number of sub-ranges. Moreover, ranges of the sub-ranges all increase with values by generally the same rates in all the 4 scenarios of features (there are several exceptions in the scenario of "maximum intensity", and it is caused by the limit of value distribution). Therefore, though there are other ways to set sub-ranges, we think our interpretation of data is reasonable, and we will illustrate more to explain the sub-ranges.

I have listed in the following a number of issues that should be addressed in this paper before publication:

1. "Eighty four groups of soil samples were collected from the profiles at all sites 5 except FH4. "(P4L5). The reason should be explained why FH4 was excluded.

Reply: Information about the soil properties of FH4 was not deliberately excluded, but because of sample missing during particle analysis. We are going to re-sample the soils at this site and analysis their properties.

2. What is the accuracy of the rain gauges?

Reply: According to the instruction manual, accuracy of the rain gauge is $\pm$1.0% at rainfall rate lower than 50mm/hr. This information will be added in the revised manuscript.

3. The slope gradient and aspect, canopy coverage and elevation of each site are suggested to add to Table1, which will help readers to understand the differences of the sites. And more explanations should be given why the sites S1H1-S1H5 and S2H1-S2H3 were set, which seem very close to each other according to Fig1.

Reply: Thanks for your suggestion. We will add the suggested information into the manuscript. Generally, sites FH1-FH4 are located at flat areas without canopy cover, while sites S1H1- S1H5 and sites S2H1-S2H3 are located at two typical hillslopes, respectively, and both slopes are covered by canopy, especially when leaves are exuberant.

Sites S1H1- S1H5 are on the same hillslope and are close to each other, thus they should be in similar conditions and could be grouped together. So are sites S2H1-S2H3. In this case, comparison could be carried out in statistic, and some discussions on the influences of slope and surface cover could be carried out, e.g., the discussions in Section 4.2.

4. What the measurement radius of the probes of TDR? The information is important because only the preference flow occurred in this range could be interpreted by the variation of observed soil moisture.

Reply: The CS616 TDR is a 2-rod probe. Rod diameter is 3.2 mm, rod space is 32 mm, and rod length is 30cm. According the manual of the product, measurement radius of the probe is 5cm, and measurement accuracy is ±2.5% VWC.

5. What is theoretical basis that a 0.002 cm3/cm3 threshold was set to quantify the responses of water content to infiltration according to a bunch of studies, given accuracies of the applied TDR probes were 0.025cm3/cm3. Whether did previous studies (Blume et al., 2009; Lin and Zhou, 2008) use the TDR probes of same accuracy?

Reply: In Lin and Zhou's (2008) study, accuracy of the probe is ±4% and relevant resolution is 0.2%. Accuracy of the probes is about±3% in Blume's et al. (2009) study and ±2.6% in Hardie's et al. (2013) study. They all suggested 0.2% as the threshold of the responses of water content. While in this study, accuracy of probe is 2.5%, and resolution is 0.1%, thus the 0.2% threshold should be reasonable.

6. The null hypothesis of the Kolmogorov-Smirnov test is usually defined as that the sample is drawn from the reference distribution (in the one-sample case) or that the samples are drawn from the same distribution (in the two-sample case) (such as in XLSTAT). However, "significant difference between every two distribution was set as the null hypothesis in this paper"(Line102-103). Which software was used to carry out the tests?

Reply: Thanks a lot for your remanding. But theoretically, it would be also reasonable to state that distributions are the same by rejecting the hypothesis of significant difference. We will conduct the test according to your suggestion, to see if the results will be different.

All calculations of this study are carried out on a MATLAB R2014b compiler.

7. It is difficult to read Fig4 and I suggest change it to a table.

Reply: Thanks for your suggestion. We will change the figure into a table.

8. "Contents of organic matter, clay and silt generally decreased with depth, leading to higher saturated hydraulic conductivity. Detailed information at each depth was not listed in Table 2, but was covered by the ranges."(P4L11-12). What is the sampling depth of the data in Table 2?

Reply: Sampling depth was limited by the depth of underlying bedrock at each site. It was 60cm at FH1-FH3 and 80cm at other sites, except that was 90cm at S2H1 and S2H3. Samples were obtained by 10cm intervals of depth.

9. The rainfall amount difference between site FH3 and S2H1 is larger than 180mm from 2014/8/22 to 2014/10/31 (Table3) but the distance seems only about 500m. Is it because the logger at S2H1 failed from 2014/10/14 to 2014/10/31(P5L27)? This should be added as notes of Table3.

Reply: We would not attribute this difference to the failure of data logger during 2014/10/14 to 2014/10/31, since only 0.5mm of rain was precipitated in this period. Given that the difference at the same slope could be up to about 60cm (S1h1 and S1h2), space heterogeneity of rainfall itself and canopy interception should have played a more prominent role.

10. Fig1 is not contour map but a DEM map.

Reply: Thanks for your suggestion. We will change Figure 1 into a contour map.