We want to thank the anonymous reviewer #1 for the constructive review. The comments are a great help to improve the manuscript. Below you can find our detailed responses (in red) to the comments (black):

General comments

This paper reports the results of comprehensive sediment sampling campaigns conducted in a small river located within the Swiss Plateau. The instrumentation and methods used for sampling suspended sediment, infiltration, and bed load are soundly presented, compared and evaluated. The bibliographic review, the scientific objectives and the statistical analyses of the results are clearly, thoroughly and precisely stated.

In my opinion, the interpretation of the results is the weakest part of the study. Indeed, despite comprehensive statistical analysis of the data is performed, spatial and temporal trends in the measurements cannot be explained clearly, since instrumental and sampling errors cannot be distinguished from the spatial and temporal heterogeneity of the sediment processes. In their conclusions, the Authors are often not able to decide and honestly express this. Clearly, this impressive experimental study calls for i) laboratory tests to assess the instrumental biases in controlled hydraulic and sedimentary conditions; ii) more advanced studies of the hydrological processes at the catchment scale and of the hydraulic processes at the local reach scale.

Author reply: We agree with the reviewer. We are often not able to distinguish instrumental and sampling errors from spatial and temporal heterogeneity. We will emphasize this better in the discussion and conclusion.

However, I think that this experimental field study is scientifically sounded and of sufficient interest to the community for being published with some moderate revisions.

Specific comments

Overwhelming statistical analyses could be reduced by eliminating some developments that do not bring real advances in the understanding of sediment processes and instrumental biases.

Author reply: We agree. We will eliminate the equations 2 to 7 and only indicate the R^2 and p of the relationships. As such, we will reduce the emphasis on the statistics and rather discuss potential biases.

p.11325: why don't you study the influence of grain size changes on your turbidity measurements? Was the grain size of SS samples used for the turbidity-meter calibration measured?

Author reply: No, we did not use the grain size of the SS caught by the SS samplers to calibrate the turbidity measurements. Since these samples are time integrated samples over a week it is not possible to use them to calibrate turbidity measurements, which have a much higher temporal resolution (10 min). Ideally, we could have used the water samples taken for the turbidity calibration to study the influence of grain size composition on NTU values. However, the amount of SS in those samples was too small to accurately determine their grain size distribution.

section 3.4: here typically, some study of the local hydraulic conditions (bed shear stress) would help interpret the spatial patterns of the bedload results. Some illustration of the site layout, flow patterns and disposal of the separate bedload samplers at each site would be helpful.

Author reply: The flow patterns changed regularly after flood events, thus a general illustration is not possible. We could add an illustration of the distribution of the bedload traps. However, we believe that the information gained would not be appropriate compared to the space used (i.e. three graphics of the three sites).

section 3.5.2 and section 3.5.3: brave statistical analysis and empirical fits do not bring real answers to the questions raised about the possible bias of sediment infiltration baskets, bedload traps, and accumulation baskets... That the infiltration rate is almost twice of the trapped bedload remains without explanation to me. Could you draw some perspectives of further studies in order to solve the point experimentally?

Author reply: As noted, we will eliminate the equations 2 to 8 and only indicate the R^2 and p of the relationships. As such, we will reduce the emphasis on the statistics and rather discuss potential biases. The differences in the infiltration rate and the bedload rate are due to differences in the method. We discuss this in more detail. The differences between the two methods could be assessed under defined conditions in a stream channel.

The number of tables could be considered too high for a research paper. In an additional figure, some pictures or technical designs of the instruments, especially baskets, integrative SS samplers and bedload traps would be helpful and informative to the reader.

Author reply: Yes, we have included many tables in the manuscript. Since we are evaluating and comparing methods under field condition, we feel this is necessary to enable the readers to judge for themselves in addition to our interpretation. Additionally, we will add figures of the different traps to the manuscript as you requested.

Fig.2. a) It is not clear to me how the 95% confidence intervals were established. Could you assess the uncertainty in the SSC samples? as well as the uncertainty in the SSC derived from the calibrated NTU measurements? Usually, the calibration curve is forced to (0,0) since turbidity should be proportional to SSC.

Author reply: The confidence intervals of the linear model were predicted from the regression model fit to the data. We only took one weekly SSC sample, thus the uncertainty in the samples can not be assessed. Sources of errors in the SSC samples could be due to vertical and horizontal heterogeneity of SSC in the water. Additionally, the imprecision of weighing the filters or measuring the exact volume of the water samples could trigger some uncertainty. We tried to keep these errors as small as possible. We believe that these errors are very small compared to the uncertainty of the general turbidity measurements. The standard error of the slope of the linear regression line (= 0.76) is 0.029.

We do not believe that forcing the calibration curve to (0,0) is necessarily correct. Our calibration curve shows the uncertainty in the method and is thus more honest than forcing it through (0,0).

Technical corrections

page 11318, line 17: dynamics

Author reply: Yes, thanks

11320, 7: mean instead of median?

Author reply: No, the OBS sensor was programmed to calculate the median value. One can not expect a normal distribution of the measured values and the mean would be a

mis-leading summary statistic for quantification. Hence, the median value provides the best approximation of the true NTU value since big particle can cause errors that would skew the average (Campbell, 2008: Operator manual OBS-3+).

11323, 8: clay <63µm (2µm?)

Author reply: Yes of course, you're right. Clay < 2 µm. Thanks.

11330, 11: exponentially

Author reply: Yes, thanks.

11332, 2: quite

Author reply: Yes, thanks.

11332, 7: Table 9 not available

Author reply: That must be a typesetting error. We wanted to refer to Table 8. Thanks.

11335, 2: dynamics

Author reply: Yes, thanks.