We are indebted with the referee for his constructive review. Our replies are embedded in bold italics in his original comments.

General comments:

I think the manuscript could benefit from at a more detailed comparison of the upper and lower reaches. At the moment the observations are presented with minimal discussion as to why sometimes reverse trends are observed (see specific comments below). As a starting point I would recommend (and very much look forward to seeing the results) treating the data cumulatively in conjunction with the flood to flood scale (such a treatment seems to parse through some of the inherent variability, see Bradley and Tucker, 2012 or Phillips and Jerolmack, 2014 for some recent examples). There is significant scatter in the data on the single flood scale (this is just the nature of variability in natural rivers and bed load transport, it can't really be avoided), much of this scatter may be substantially reduced by comparing the average cumulative displacement for each flood against the cumulative duration that the tracers have spent above the threshold of motion. It would be interesting to see the resulting trends and if the US and LS have different functional forms or similar functional forms. The resulting trends would be similar to figure 11, but would instead include time on the independent axis rather than the categorical variable of flood type, likewise displacement would replace virtual velocity.

It would be interesting to know if the two river segments are adjusted to pass the water and sediment supplied. In that are both river segments adjusted to near threshold conditions typical of gravel rivers (Parker et al., 2007) or do their particular positions within the formerly glaciated terrain cause contrasting or similar deviations from the generally expected form. In a sense is the glacially imprinted landscape actually present in the stream dynamics at large. The figures seem to suggest that increases in slope are being compensated for by increases in particle size, and that would suggest that the supply of particles may be more fundamentally important to the longterm evolution of the topography (similar patterns are observed in Attal et al., 2015 for a catchment in the Sierra Nevada mountains). If true it would suggest (to me) that the river may adjust rather rapidly to imposed sediment and water, but that a limiting factor could be well (or included in the supplementary), it is not clear which populations are used in the analysis. Questions to consider include: did the various populations behave in a similar manner or are newer populations more mobile because they have yet to be worked into the creek bed; do the first floods after tracer installation result in consistently larger displacements. Similarly the methods could benefit from a more detailed description of how the tracers were installed into the river (embedded by replacing a particle of like size, or simply placed on the bed).

When we conceived this project in the first place, we identified the main research objective as the evaluation of the contemporary coarse sediment transfer along a rugged formerly glaciated mountain stream. To pursue this objective, we decided to select US and LS as representative sites along Strimm Creek, and to monitor bedload transport within these sites via PIT-tagged tracers for three years. We believe that the analysis performed and the results presented in this manuscript have led to robust conclusions that address the original research question i.e, quantify the glacial conditioning on contemporary bedload transport.

We do appreciate the reviewer's suggestions, which for sure represent excellent ideas. However, we feel that the additional recommended analysis does not actually fit into the main objective of our paper; rather, it would entail writing a different paper, although very interesting and indeed planned for the next future. For this reason, we would like to stick to the original paper structure without re-analysing our data, hoping the reviewer and the editor will understand our perspective. Nonetheless, to answer the specific question on the behaviour of the different sub-sets of tracers seeded on the riverbed at different times, we did not observe significant differences between the "old" and the "new" cohorts of tracers. Specifically, after placing new tracers in the riverbed, we performed within the next few days an exploratory survey with the only aim to assess their actual starting positions after their possible rearrangement by the (low) flows. In this way, the bias induced by simply placing clasts in the riverbed (introducing an initial fictitious higher mobility) was eliminated, or at least reduced.

The methods section of the manuscript (p424, l12) now reads:

"PIT-tagged clasts were then released by placement on the streambed as transverse ribs along the study sites with a downstream spacing of 5 m in US and 3 m in LS.

The following comments and critiques are given in the order that they appear in the manuscript.

P.420 Ln.17: An appropriate reference for the various detection distances and how that is related to tracer size and orientation is Chapuis et al., 2014.

The suggested reference was added to the manuscript.

P.420 Ln.25: The work of Bradley and Tucker (2012) is relevant here as they record some of the highest recovery rates for yearly surveys.

The suggested reference was added to the manuscript.

P.425 Ln.20: If possible please provide the detection ranges for the RFID antennas used (horizontal and veritical).

The detection range (both horizontal and vertical) of the antennas is about 0.5 m. The manuscript was modified as follows:

"Bedload tracing was conducted by means of portable RFID antennas (one manufactured by AQUARTIS, one by Oregon RFID; vertical and horizontal detection range is approximately equal to 0.5 m)."

P.427 Ln.6: How was the integral treated when the threshold of motion doesn't fall on a sampling point? Was the hydrograph interpolated between sampling points and then integrated?

The portion of the hydrograph (recorded at 10 min intervals by the pressure transducers) relevant for the calculation of virtual velocity was determined from the values of the motion thresholds estimated by equation 1. Given the sufficient temporal resolution of the flow discharge data relevant to bedload transport duration in our study site, no interpolation was performed and we started the integral computation from the first value higher than the motion threshold.

P.427 Ln.9: Could you describe the digging tests in LS to determine the active layer.

Digging tests (n = 17) performed in US returned burial depths comprised between 1 and 14 cm (median depth = 9 cm), therefore within the D_{50} of the surficial grain size distribution (Houbrecht et al., 2012). In LS, where no digging was conducted, we did not observed any major morphological unit rearrangement, even after the most important bedload events, so that the values for the active layer used in the paper appear sensible and in line with those reported by other researchers working on steep mountain streams (excluding large flood events).

The new text in the manuscript reads as follows:
"This approximation, which is based on digging tests conducted on a subset of buried tracer stones in US (n = 17; burial depth range: 1-13 cm; median burial depth = 9 cm), is in agreement …"

P.429 Ln.1: Are all of the following Qmax values listed for intra or inter survey floods. For the intra-survey flood, was it flooding during the survey?

They were all inter-survey; intra-survey has been modified into inter-survey in the whole manuscript.

P.429 Ln.1-25 & P.430 Ln.16-28 + P.431 Ln.1-24 : Could these lines may be turned into a table with columns: survey date, Qmax, %mobile, mobile classes, maximum travel distance, rain/snow flood, percent of total number of tracers recovered (or other columns). At the moment the information is difficult to assimilate due to the large number of events and list like nature (I am not sure there is a good way other than a table to describe this information). Could you also report the percentage of the total number of tracers installed that were recovered for each flood. (These tables could also be put in the supplementary material along with the other supplementary material already present).

The reviewer was right, thanks for the suggestion. We added a tables (see amended Table 3) where data have been reorganized so that they could work as a summary for the in-text data.

"Table 3 – Recovery rates per survey in US and LS. Rates of the last two surveys are considerably lower than what previously recorded, due to several clasts moving beyond the outlet section of the study area."

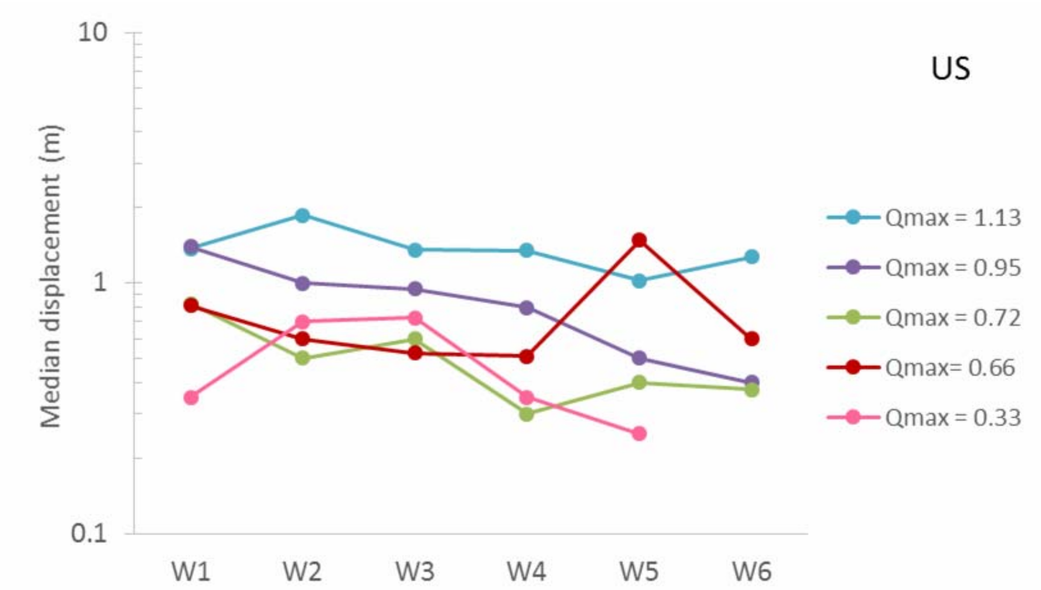
Survey (US)	Qmax [m^3s^{-1}]	Mobilized clasts [%]	Mobilized classes	Max travel distance [m]	Intra-survey dominant regime	Recovery rate
2011.09.28	0.36	3.8%	W2, W3, W5	1.2	Rainfall	100%
2012.06.14	0.66	27.7%	W1 to W6	4.5	Snowmelt	100%
2012.07.03	0.55	4.6%	W1, W2	3.4	Snowmelt	100%
2012.09.11	0.33	17.3%	W1 to W5	3.9	Rainfall	98.5%
2012.10.04	0.32	4.3%	W1 to W3	1.6	Rainfall	97.7%
2013.06.27	1.13	50.8%	W1 to W6	35.0	Snowmelt	93.4%
2013.10.01	0.72	38.1%	W1 to W6	12.0	Mixed	92.3%
2014.06.13	0.95	40.6%	W1 to W6	23.9	Snowmelt	90.3%

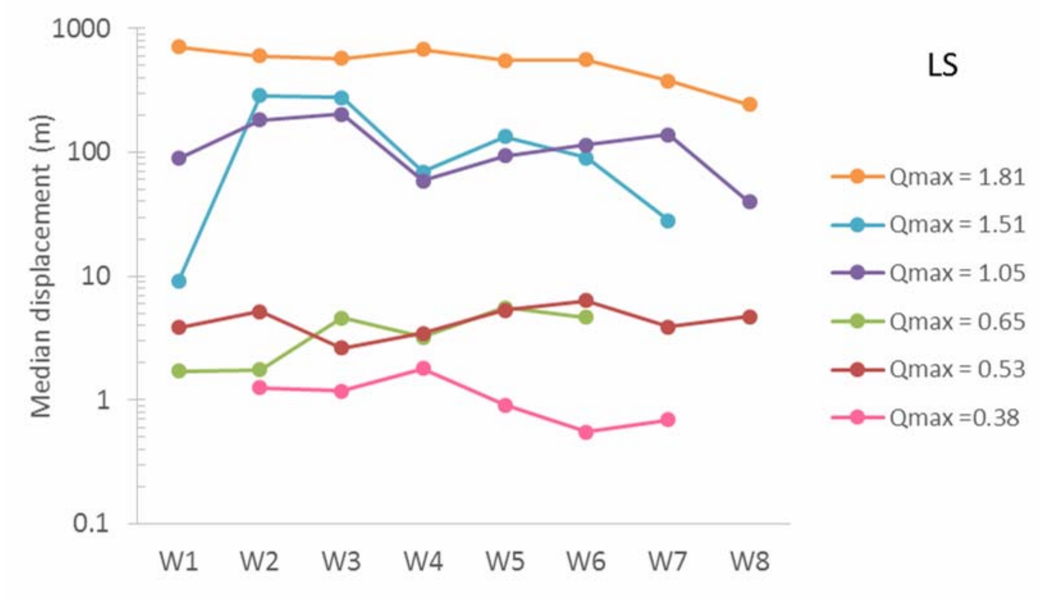
Survey (LS)	Qmax [m^3s^{-1}]	% mobile	Mobilized classes	Max travel distance [m]	Intra-survey dominant regime	Recovery rate
2011.09.27	0.58	14.1%	W2, W3, and W5 to W8	33.9	Rainfall	100%
2012.05.18	0.53	49.4%	W1 to W8	17.2	Mixed	100%
2012.06.28	1.05	86.5%	W1 to W8	936.8	Snowmelt	100%
2012.08.21	0.65	37.6%	W1 to W8	21.7	Mixed	97.1%
2012.10.25	0.51	36.6%	W1 to W8	21.1	Rainfall	96.5%
2013.05.21	0.38	26.8%	W2 to W7	10.8	Snowmelt	96.5%
2013.07.02	1.81	96.3%	W1 to W8	926.9	Snowmelt	96.5%
2013.10.02	1.01	60.0%	W1 to W8	959.3	Mixed	56.0%
2014.07.16	1.51	64.6%	W1 to W8	958.3	Snowmelt	54.7%

P.430 Ln.6-7 & figure 7a: My interpretation of Figure 7 is that for larger floods the degree to which a lighter particle travels the farthest increases. This doesn't really say anything about selective entrainment. How representative is the maximum travel distance given that there are very few heavy tracers, in that could much of this trend be due to sample size? Tracer displacements tend to be well described by exponential (Phillips and Jerolmack, 2014) or gamma functions (Bradley and Tucker, 2012), for a small number of tracers we are not likely to observe large displacements because they are rare. For larger populations of tracers we are more likely to be able to sample the tail of the distributions and observe large displacements. I am not sure if what I am suggesting is the case, but caution should be used when drawing conclusions based on a small number (compared to the number of lighter tracers) of heavy tracers. A less biased indicator would be the 90th or 95th percentile displacement, but the bulk of the data seem to show relatively equal displacement distances. It would be useful to provide the number of tracers within each boxplot in figure 7b (a number above or below the top whisker would suffice I think).

In order to address the reviewer's concern we have replaced old Fig 7a and 8a with the two figures below. New Fig 7a and 8a report median travel distance across weight classes for selected inter-survey periods. In consideration of this new representation, comments to the figures have been changed accordingly.

The new text reads as follows (p430, l6ff):
"For each five intra-survey values of Qmax, weight classes were plotted against the median travel distance during that period (Fig. 7a). These plots suggest that median transport lengths are only weakly influenced by peak discharge (all values within one order of magnitude). Boxplots of tracer travel distance (d) across weight classes (W) (Fig. 7b) show that median values (and maximum values even more) decrease progressively (with the exception of class W5) for heavier particles."





The text at p.432, line 3-11 has been changed as:

"Median travel distances at LS are more sensitive to Qmax (Fig. 8a) compared to what observed at US (Fig. 7a), as they range over three orders of magnitude. In contrast, less pronounced than at LS is the decrease in travel distance (max, median, quartiles) with increasing particle weight (Fig. 7b). Indeed, displacement data at LS indicate that all the clasts within the analyzed size range (representing the "ordinary" bedload flux) move along similar distances for a given flow condition."

The number of tracers for each box in Fig 7b & 8b was added as suggested.

P.432 Ln.10: Could you add in the discussion a possible explanation for why the W2 weight class has such a high median transport distance as the rest of the weight classes are remarkably consistent with each other.

We agree with the reviewer. The median value in W2 is much higher than in the other weight classes. Currently, we have no explanation for a similar result.

P.432 Ln.11: Without additional information figure 8 does not necessarily demonstrate equal mobility. It does demonstrate that transport distance does not depend on particle weight. Are similar percentages of the tracer weight classes mobile. Both figure 7 and 8 are very intriguing, but it is figure 6 that tends to show if there are mobility differences.

The reviewer's comment is indeed pertinent and we have modified the text as follows: (p432, l11-12):

"Overall, within the explored range of sediment, size, Figure 8 (along with Fig. 6) clearly shows how the dependency of bedload travel distance on particle weight is rather low in both segments of the Strimm Creek, corroborating the general tendency to near-equimobile conditions (see Lenzi, 2004) already observed analysing the critical discharges for tracers motion. However, such a behavior is stronger in LS than at US, as in the latter transport data feature more variability across the different particle weight classes, as well as a consistent decreasing trend of the maximum travel distance."

P.432 Ln.15-19: It is not clear to me why the threshold discharges should be the same. Are the cross sections the same or has the slope and particle size adjusted just so that the same amount of water is now capable of moving larger particles? Please if possible estimate what these discharges would be in terms of shields stress for a representative cross section or lacking a stage discharge curve what are the representative shields stresses for the bankfull discharges?

As we specifiedin the text, we prefer not to include the critical shear stress approach in the ms (see e.g. Comiti and Mao, 2012). However, to answer the reviewers' question, bankfull geometries at US and LS are different (see methods section), as well as are their bed sediment size. Overall this leads to similar critical flow conditions in the two reaches. Dimensionless (relative to D_{50} of the bed sediment) shear stress are in the range of 0.1 – 0.3.

P.432 Ln.21-23: Are immobile tracers included in the mean transport distance? It is not clear from the text so far where they are included (other than figure 6) and excluded in the analysis.

The boxplots contain data from mobile tracer particles only (hence, observed displacement greater than 20 cm in US and 50 cm in LS). The non inclusion of zero movement clasts in the analysis was due to the fact that we wanted to evaluate the sediment fluxes towards the outlet of the basin. The presence of immobile clasts does not change the results; hence they were not included in the boxplots.

To avoid any possible misunderstanding, the manuscript will be changed as follows:

"Boxplots of tracer travel distance (d) across weight classes (W) (Fig. 7b), in which "no motion" counts are not included, show that median values decrease progressively for heavier particles […]"

Following the same logic, on p432 l6ff, the manuscript will be changed accordingly:

"***Boxplots of clast displacement (d) across weight classes (W) during the study period (Fig. 8b; "no motion" counts are not included) show comparable ranges of variability, both in terms of inner (25–75 %) and outer (< 25 and > 75 %) quartiles."***

Captions of figures 7 and 8 will be changed as follows:

Figure 7 – (a) Tracer mean travel distance as a function of weight and stratified by inter-survey peak discharge (Q_{max}) in Upper Strimm (US). (b) Boxplot showing the distribution of tracer travel distances across weight classes in Upper Strimm (US). No motion counts are not included.

Figure 8 – (a) Tracer mean travel distance as a function of weight and stratified by inter-survey peak discharge (Q_{max}) in Lower Strimm (LS). (b) Boxplot showing the distribution of tracer travel distances across weight classes in Lower Strimm (LS). No motion counts are not included.

P.434 Ln.18-26: Could you specify or hypothesize as to the origin of the valley step, it seems to be an essential piece in decoupling the upstream and downstream sections of the river. Is it glacial in origin?

We realize there was a misleading statement. Further discussion with Dr. Corrado Morelli, the geologist that oversees the geological mapping in this part of the Italian Alps has ruled out the structural origin of the valley step. The origin of the valley step is therefore glacial; the new text now reads (l21-23):

"In the case of Strimm Creek, the main valley step is of glacial origin, as it is not associated to any fault line or lithological contact/discontinuity (Corrado Morelli, Geological Survey of Bolzano, personal communications). Specifically, it is the result of differential glacial erosion associated with the main relict ice flow of the Adige Valley, and the tributary cirque glacier located at the headwaters of Strimm Creek (i.e., the hanging valley in US)".

p. 435 , line 17: "highly" will be changed into "weakly"

P.437 Ln.12-19: This is a rather long and confusing sentence. These points would be better emphasized as several separate sentences.

Comment accepted. The original sentence has been simplified and split in two. The new text now reads:

"Despite the limited length of record and the uncertainty associated with the calculation of bedload transport volumes, the combination of debris flow/flood activity together with estimates of 15 bedload transport events document contrasting rates of landscape down-wasting in the upper and lower portions of the investigated basin. These findings point to a postglacial evolution in which glacial topographic signatures (e.g., Brardinoni and Hassan, 2006) will likely amplify rather than reduce."

P.437 Ln.20: These ideas are not integrated into the body of the work. It could be true, and I think the authors may be able to make a case for the claim that reduced transport in the US reach will provide fewer tools to erode the knick zone (valley step) separating the LS and US reaches and thus the decoupling between the two reaches will persist, but this argument is not clearly made.

Comment addressed. Implications for long-term landscape evolution have been added to the discussion on page 436, line 26:

"The implications for long-term landscape evolution in Strimm Creek are intriguing as we are looking at two portions of the same basin that are functioning at markedly different speeds, and therefore characterized by a different geomorphic sensitivity to change (Brunsden and Thornes, 1979). Upper Strimm Creek is a typically "slow response" sub-system, in which, due to generalized hillslope-channel disconnection and to inefficient bedload transport, any perturbation through the hillslopes and the drainage network is transmitted relatively slow. By contrast, Lower Strimm Creek represents a highly sensitive, fast responding system thanks to the high degree of hillslope-to-channel sediment delivery, together with much higher bedload transport. In consideration of the limited bedload transport recorded at US, which implies a limited supply of clasts to the rocky valley step separating US from LS, we postulate that the valley step will undergo extremely slow bedrock incision, due to the limited tool effect acting on the sub-vertical rocky channel (e.g., Finnegan et al., 2007; Turowski et al, 2007). This hypothesis is supported by the little bedrock incision achieved to date by Strimm Creek in the last 12,000 years since deglaciation."

Figure 2. A scale bar or mention of scale in the caption for each photograph would be useful.

In pictures a, c, d a person is portrayed to give an idea of the actual dimensions of the river; as for picture b, boulders have an approximate diameter of 0.8 m. But we agree with the reviewer, a mention of the channel width was added in the caption, that now reads:

Figure 2 – (a) Upper Strimm flowing across the hanging valley floor in decoupled conditions; and (b) detail of rapids channel-reach morphology (i.e., transverse ribs on a shallow bed) at low flow conditions (channel width = 4 m). (c) Lower Strimm in strongly coupled conditions characterized by step-pool and boulder-cascade morphology at low flow conditions with banks eroded by the 2010 debris flow (channel width = 4.5 m); and (d) the water gauging station (WS1) to the left of the picture.

Figure 3a. It would be useful to have the CDF for the tracers on this plot as well. At the moment it is not as insightful to compare the tracer grain size histograms to the stream cdf (it is not the point of 3b to compare to the stream, but it is nice to be able to graphically understand what part of the bed distribution the tracers sample).

Comment accepted. CDFs for the tracers have been added to Figure 3a. The modified version of Figure 3a is reported below.

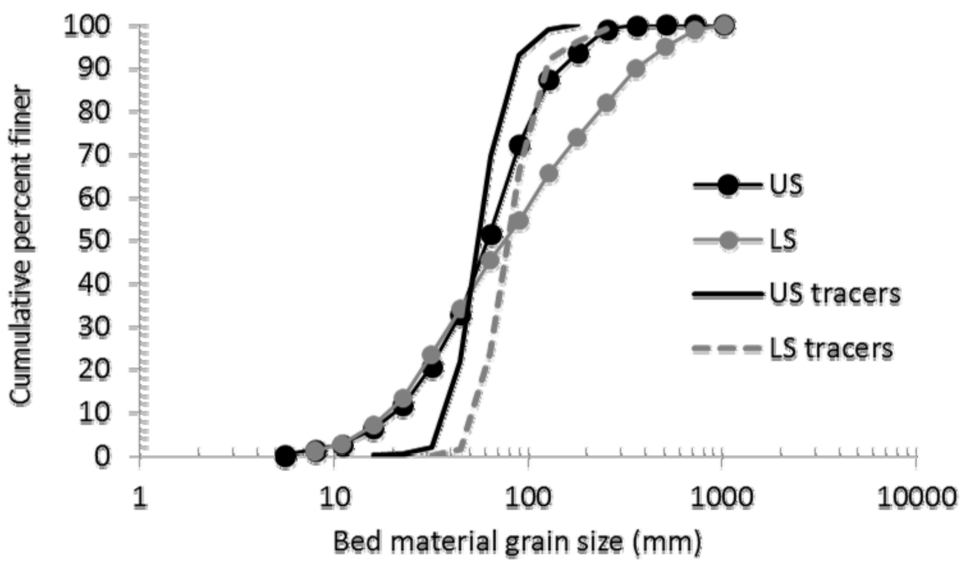


Figure 3c. Is there a reason why the fitted functional relationship is not shown?

The figure was modified and now is showing the fitted functional relationship.

Figure 4. Please add a line delineating where (approximately) the threshold is located.

A line was added to show the US activation threshold.

Figure 6. Could you put the discharge or another metric associated with the flood (shear stress, stream power, or duration) in each plot. A number in the upper right would be useful i.e. Q=nn in the upper right of each plot (or better yet the shields stress or width normalized stream power). Where are plots J and K or are they mislabeled?

Plots in Fig 6 were mislabelled. They have been relabelled so that letters a-h are used for the US diagrams in the left-hand side panels and letters i-q for the LS diagrams in the right hand side; Qmax has been added to each plot.

Figure 10. Consider adding to the discussion of this figure why the mixed events result in the lowest virtual velocities. It would be logical (maybe) to think that they should be between the snow melt and the rainfall floods. Is the trend for the rainfall induced flood in part A significant? Overall these are very interesting results in that it seems that particle virtual velocity is nearly independent of particle weight, this result is, to my knowledge, novel and undersold in the manuscript.

We thank the reviewer for offering us the opportunity to give a deeper thought to the issue of virtual velocities in general, and to the values associated with "mixed" periods in particular. To better illustrate what causes comparatively lower virtual velocities during "mixed" inter-survey periods we have decided to add to the manuscript two extra tables (now labelled Table 4 and Table 5) in which we report for each inter-survey period Qmax, inter-survey length, median virtual velocity and virtual transport duration.

The expanded new text reads as follows (p 433, line 18):

"Mixed snowmelt-rainfall events (Qmax = 0.53, 0.65 and 1.01 m^3 s^{-1}) lie consistently lower than snowmelt- and rainfall-dominated periods. Inspection of the hydrograph shape (Figure 4) in and the virtual transport duration values (Tables 4 and 5) (i.e., time for which Q > threshold Qmax for bedload entrainment) is instructive. From the premise that "mixed" and purely snowmelt periods display comparable virtual transport durations (Tables 4 and 5), we argue that mixed periods are characterized by low virtual velocities because they integrate the second half of the snowmelt season (i.e., past the main snowmelt peak). Accordingly, at this time of the year, even though water discharge is well above the threshold Qmax associated with flashy summer storms (i.e., 0.32 and 0.38 m^3 s^{-1}), the nearly flat hydrograph shape of the snowmelt falling limb promotes channel bed stabilization (e.g., Hassan et al., 2006), with reduced bedload transport (e.g., Green et al., 2013) and lowest virtual velocities."

Table 4. Median virtual velocities (Vv) and virtual transport durations at US.

Type	Qmax (m^3 s^{-1})	Inter-survey length (days)	Median Vv (cm min^{-1})	Virtual transport duration (minutes)
Rain	0.36	28	0.6600	60
Snowmelt	0.66	260	0.0020	29570
Snowmelt	0.55	19	0.0018	27060
Rain	0.33	70	0.3461	130
Rain	0.32	23	4.4000	10
Snowmelt	1.13	266	0.0074	20890
Mixed	0.72	96	0.0011	52370
Snowmelt	0.95	255	0.0045	22220

Table 5. Median virtual velocities (Vv) and virtual transport durations at LS.

Type	Qmax (m^3 s^{-1})	Inter-survey length (days)	Median Vv (cm min^{-1})	Virtual transport duration (minutes)
Rain	0.58	32	0.0768	6575
Mixed	0.53	234	0.0092	41670
Snowmelt	1.05	41	0.1938	58870
Mixed	0.65	54	0.0168	22590
Rain	0.51	65	0.3991	470
Snowmelt	0.38	208	1.8432	60
Snowmelt	1.81	42	1.7501	33330
Mixed	1.01	92	0.0029	108970
Snowmelt	1.51	287	0.1654	103080

Figure 11. Typo in caption ("Snomelt").

The typo was fixed.

References cited in review:

Attal, M., Mudd, S. M., Hurst, M. D., Weinman, B., Yoo, K., and Naylor, M.: Impact of change in erosion rate and landscape steepness on hillslope and fluvial sediments grain size in the Feather River Basin (Sierra Nevada, California), Earth Surf. Dynam. Discuss., 2, 1047-1092, doi:10.5194/esurfd-2-1047-2014, 2014.

Bradley, D. N. and Tucker, G. E.: Measuring gravel transport and dispersion in a mountain river using passive radio tracers, Earth Surf. Process. Landf., 37(10), 1034–1045, doi:10.1002/esp.3223, 2012.

Chapuis, M., Bright, C. J., Hufnagel, J. and MacVicar, B.: Detection ranges and uncertainty of passive Radio Frequency Identification (RFID) transponders for sediment tracking in gravel rivers and coastal environments, Earth Surf. Process. Landforms, 39(15), 2109–2120, doi:10.1002/esp.3620, 2014.

Li, C., Czapiga, M. J., Eke, E. C., Viparelli, E. and Parker, G.: Variable Shields number model for river bankfull geometry: bankfull shear velocity is viscosity dependent but grain size-independent, Journal of Hydraulic Research, 0(0), 1–13, doi:10.1080/00221686.2014.939113, 2014.

Parker, G., Wilcock, P. R., Paola, C., Dietrich, W. E. and Pitlick, J.: Physical basis for quasi-universal relations describing bankfull hydraulic geometry of single-thread gravel bed rivers, J. Geophys. Res., 112, 21 PP., doi:200710.1029/2006JF000549, 2007.

Phillips, C. B. and Jerolmack, D. J.: Dynamics and mechanics of bed-load tracer particles, Earth Surface Dynamics, 2(2), 513–530, doi:10.5194/esurf-2-513-2014, 2014.

Trampush, S. M., Huzurbazar, S. and McElroy, B.: Empirical assessment of theory for bankfull characteristics of alluvial channels, Water Resour. Res., 50(12), 9211–9220, doi:10.1002/2014WR015597, 2014.