



# Tracking Downstream Variability in Large Grain-Size Distributions in the South-Central Andes

Benjamin Purinton<sup>1</sup> and Bodo Bookhagen<sup>1</sup> <sup>1</sup>Institute of Geosciences, University of Potsdam, Potsdam-Golm, Germany**Key Points:**

- Complex grain-size distributions in dynamic mountain rivers can be computed via thousands of measurements from *PebbleCounts*
- Many measurements allow robust estimation of higher percentiles and we observe the most significant changes in the  $\geq 95$ th percentile
- Downstream grain-size variation is nonlinearly related to variations in topographic steepness and lithology

**Supporting Information:**

Supporting Information may be found in the online version of this article.

**Correspondence to:**B. Purinton,  
[purinton@uni-potsdam.de](mailto:purinton@uni-potsdam.de)**Citation:**Purinton, B., & Bookhagen, B. (2021). Tracking downstream variability in large grain-size distributions in the south-central Andes. *Journal of Geophysical Research: Earth Surface*, 126, e2021JF006260. <https://doi.org/10.1029/2021JF006260>

Received 14 MAY 2021

Accepted 28 JUL 2021

**Abstract** Mixed sand- and gravel-bed rivers record erosion, transport, and fining signals in their bedload size distributions. Thus, grain-size data are imperative for studying these processes. However, collecting hundreds to thousands of pebble measurements in steep and dynamic high-mountain river settings remains challenging. Using the recently published digital grain-sizing algorithm *PebbleCounts*, we were able to survey seven large ( $\geq 1,000 \text{ m}^2$ ) channel cross-sections and measure thousands to tens-of-thousands of grains per survey along a 100-km stretch of the trunk stream of the Toro Basin in Northwest Argentina. The study region traverses a steep topographic and environmental gradient on the eastern margin of the Central Andean Plateau. Careful counting and validation allows us to identify measurement errors and constrain percentile uncertainties using large sample sizes. In the coarse  $\geq 2.5 \text{ cm}$  fraction of bedload, only the uppermost size percentiles ( $\geq 95$ th) vary significantly downstream, whereas the 50th and 84th percentiles show less variability. We note a relation between increases in these upper percentiles and along-channel junctions with large, steep tributaries. This signal is strongly influenced by lithology and geologic structures, and mixed with local hillslope input. In steep catchments like the Toro Basin, we suggest nonlinear relationships between geomorphic metrics and grain size, whereby the steepest parts of the landscape exert primary control on the upper grain-size percentiles. Thus, average or median metrics that do not apply weights or thresholds to steeper topography may be less predictive of grain-size distributions in such settings.

**Plain Language Summary** Rock fragments on hillsides are transported to rivers, eventually becoming pebbles, sand, and mud as they are carried downstream by flowing water. The initial size of the pebbles, the way the size changes downstream, and the overprinting of the sizes with new pebbles from other hills and tributaries all form a complex process that can be difficult to disentangle. Yet studying the size of the pebbles at a given stream location or in a sedimentary deposit can provide insights into the conditions of their transport in terms of local upstream patterns of erosion, tectonics, and climate. We show that just looking at the size of the large pebbles on a riverbed can be used to infer the sources of material, but, since there are fewer large pebbles, they require more measurements to quantify. This necessitates new methods for pebble measurement using modern image-processing tools.

## 1. Introduction

It has long been recognized that the grain-size distribution in mixed sand- and gravel-bed channels—and the deposits they form—encode end transport conditions (e.g., Allen et al., 2017; Armitage et al., 2011; D'Arcy et al., 2017; Duller et al., 2010; Knighton, 1982). The initial sediment grain-size distribution is produced on hillslopes by myriad interacting processes including, but not limited to, landslides, chemical weathering, and biotic factors (e.g., Roda-Boluda et al., 2018; Sklar et al., 2017). This distribution is then modified during transport to and within the fluvial network, principally by some combination of abrasion and selective sorting, to produce patterns of downstream fining (e.g., Chatanantavet et al., 2010; Gomez et al., 2001; Harries et al., 2018; Miller et al., 2014; Rice & Church, 1998).

The relative importance of both size-reduction processes are debated, and this process-scale question of downstream fining is most often studied in models (e.g., Gasparini et al., 1999; Sklar et al., 2006), laboratory flumes (e.g., Paola et al., 1992), or small, confined channel reaches in relatively simple catchments with abundant data coverage (e.g., Ferguson et al., 1996; Neely & DiBiase, 2020; Sklar et al., 2017, 2020). Downstream grain-size patterns also have implications for studying landscape denudation at large scales

© 2021. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

related to basin history, lithology, tectonics, and climate (e.g., Attal & Lavé, 2006; Attal et al., 2015; Dingle et al., 2017; Harries et al., 2021; Parida et al., 2019).

Gathering grain-size distributions in steep high-mountain sand- and gravel-bed rivers is challenging, partially in terms of access to remote areas. Moreover, there is difficulty in collecting large sample sizes (typically via arduous manual labor) to robustly measure the full distribution in dynamic rivers where a wide range of material from fine gravels to large boulders can occur (e.g., Attal & Lavé, 2006; Bunte & Abt, 2001; Wolcott & Church, 1991). A number of algorithms have been developed for automated or semi-automated pebble counting via image segmentation on photos (Butler et al., 2001; Detert & Weitbrecht, 2012; Graham, Reid, & Rice, 2005; Graham, Rice, & Reid, 2005; Sime & Ferguson, 2003). These algorithms are accurate for interlocking, oblate grains with uniform inter-granular (interstice) shadows typical of many gravel-bed rivers (e.g., Graham et al., 2010). However, photos from dynamic rivers in high-mountain areas are typically more complex, showing overlapping grains, a large range of sizes including sand patches, irregularly shaped pebbles (non-ellipsoid), pebbles with intra-granular color variations or texture such as veins or fractures, and highly irregular inter- and intra-granular shadowing.

Building on others (e.g., Detert & Weitbrecht, 2012), the development of the open-source *PebbleCounts* digital grain-sizing algorithm (Purinton & Bookhagen, 2019a, 2019b) constitutes important progress in the accurate measurement of grain size in such challenging high-mountain settings. Using photos captured from handheld, mast-mounted, or UAV-mounted cameras, it is possible to quickly and efficiently survey large—spatially continuous, acre to hectare ( $\sim 4,000$ – $10,000$  m<sup>2</sup>)—channel areas at  $\sim 1$  mm/pixel resolution and later count tens-of-thousands of pebbles from the comfort of a computer screen.

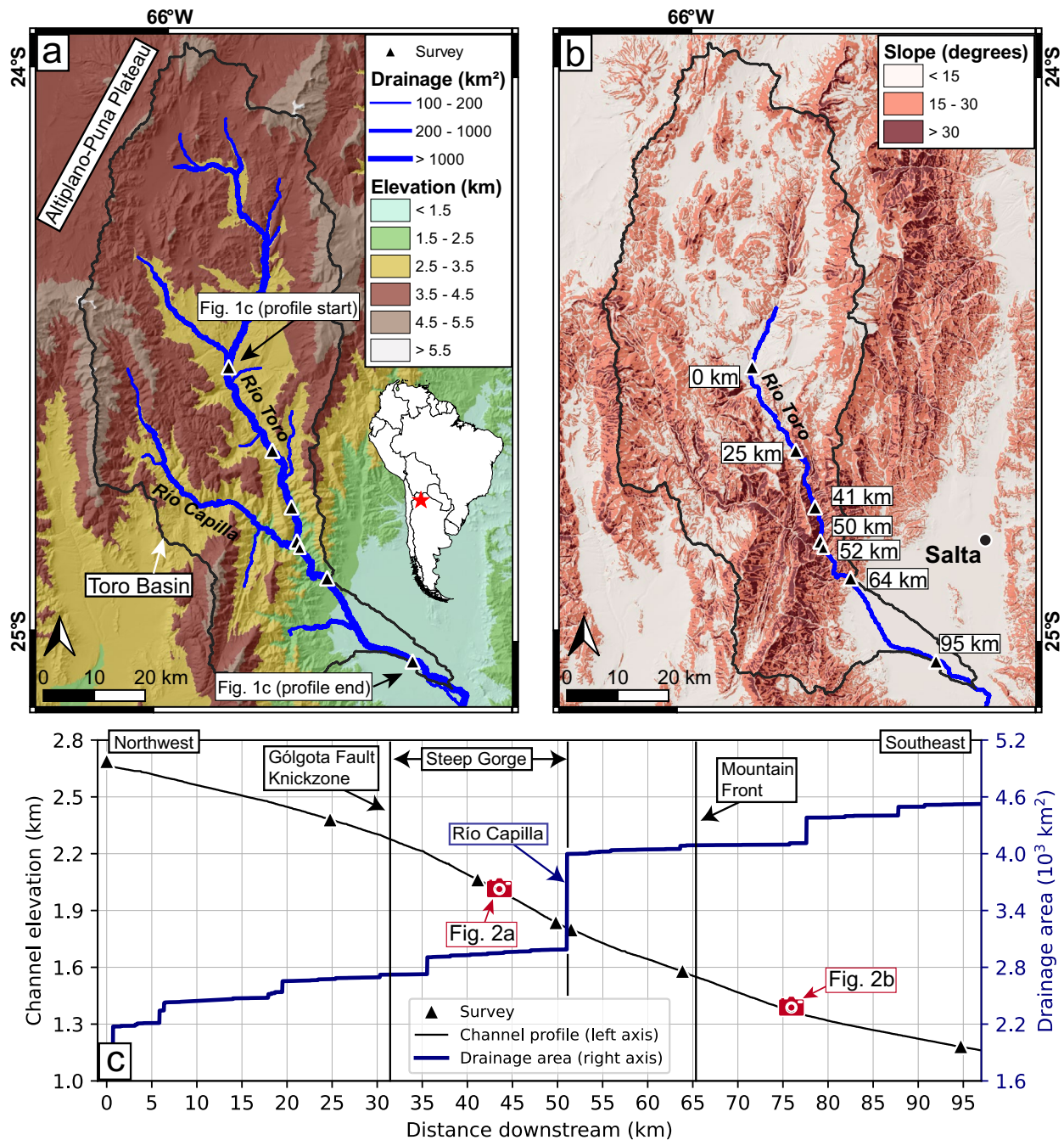
Here, we continue on previous work to develop *PebbleCounts* by applying it at large scale to the well-studied Toro Basin in the south-central Andes (Figure 1). First, we thoroughly outline the procedure for processing and validating channel cross-section photo surveys using *PebbleCounts*. Second, we examine the utility of large sample sizes to reduce percentile uncertainties at the  $\geq 95$ th percentiles using the methods of Eaton et al. (2019). Third, we use thousands of measurements to form robust percentile estimates to investigate the patterns of downstream changes in grain size. We emphasize a potentially strong nonlinear relation between geomorphic metrics and grain sizes in steep, high-mountain settings. This is reflected in a correlation between the upper percentiles of steepness indices (hill slope and channel) and downstream changes in the size of the coarsest bedload, and this signal is strongly influenced by lithology and geologic preconditioning.

## 2. Study Area

The Toro Basin (Quebrada del Toro) in the south-central Andes is a large ( $\sim 4,500$  km<sup>2</sup>) high-mountain catchment (Figure 1). This basin drains from headwaters to the north and west abutting the internally drained Altiplano-Puna Plateau (elevation  $> 3.5$  km) toward the eastern foreland (elevation  $< 1.5$  km). The seven *PebbleCounts* survey sites used in the present study are located along the trunk stream, and are referenced with respect to their downstream flow distance, such that the upstream-most site has a flow distance of 0 km (Figure 1b).

The trunk stream is the braided sand- and gravel-bed Río Toro, which has a drainage area of  $\sim 4,500$  km<sup>2</sup> at our lowest survey site. A major tributary junction 15-km upstream of the mountain front—the Río Capilla—accounts for  $\sim 1,000$  km<sup>2</sup> of this drainage area (Figure 1c). The Río Toro traverses a steep environmental and climatic gradient from arid, vegetation-free, low-slope upstream reaches, down through a steep bedrock gorge (cf. high relief along channel profile in Figure S1), and out to the humid ( $> 1$  m/y rainfall), densely vegetated foreland. We note that vegetation within the active channel is low (Figure 2).

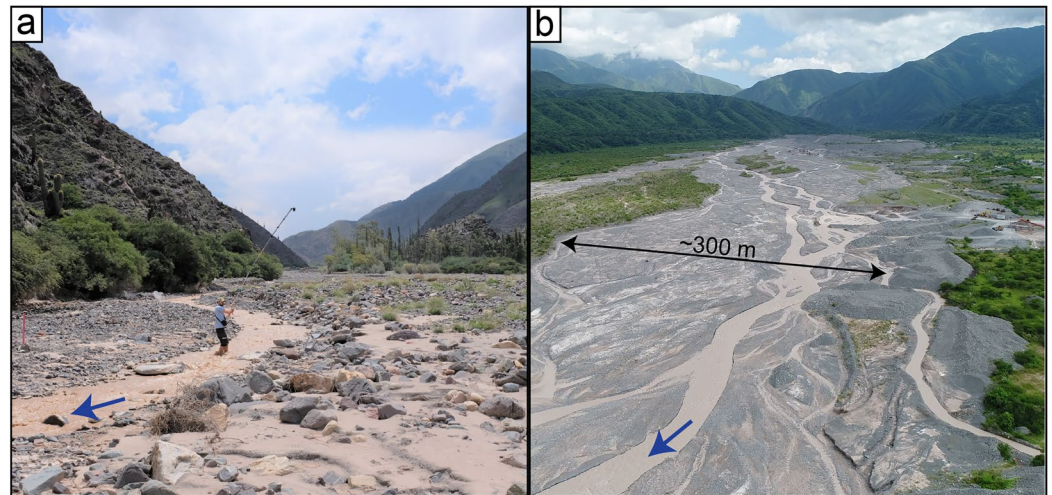
The Cenozoic and Quaternary basin history includes periods of connection and decoupling from the foreland sedimentary basins driven by tectonic and climatic factors (García et al., 2013; Hermanns et al., 2000; Hilley & Strecker, 2005; Marrett et al., 1994; Marrett & Strecker, 2000; Strecker et al., 2007; Tofelde et al., 2017). Cosmogenic radionuclide erosion rates from several tributaries show a general west to east (arid, high elevations to humid, low elevations) trend of increasing erosion (Bookhagen & Strecker, 2012; Tofelde et al., 2018). Contemporary hydro-meteorologic conditions are characterized by a wet and dry season (coinciding with the South American Summer Monsoon) with frequent, high-intensity rainfall events and recent trends of increasing rainfall and river discharge related to climate change (Castino et al., 2016a,



**Figure 1.** (a) Map of Toro Basin with the seven *PebbleCounts* survey sites, elevation, and major drainage network. (b) Slope map with flow distance location of surveys. Note, we set the upstream-most survey to 0 km and measure downstream flow distance from this site. (c) Longitudinal profile of Río Toro trunk stream with drainage area on the right axis. Locations of photos in Figure 2 are also shown. Vertical lines indicate major geomorphic transitions along the trunk stream, and are used in subsequent longitudinal profile plots. Elevation, slope, and drainage network are derived from the 12 m TanDEM-X DEM (Rizzoli et al., 2017).

2016b, 2017, 2020). Finally, high rates of bedload transport have been observed along the Río Toro (Purinton & Bookhagen, 2018), and associated spaceborne radar measurements of this bedload material indicate the wide range and variability of grain size (Purinton & Bookhagen, 2020), along with the seasonal motion of this material (Olen & Bookhagen, 2020).





**Figure 2.** (a) Field photo taken during *PebbleCounts* mast surveying (mast height  $\sim 5$  m) in the steep gorge. Note the low hillslope and channel vegetation and the large boulders in the active channel. (b) Field photo taken from a UAV downstream of the mountain front looking upstream. Note the higher vegetation on hillslopes and large quantity of gravel being transported out of the Toro Basin (cf. Figure 1c for photo locations).

The primary lithology for  $\sim 70\%$  of the basin is the Proterozoic low-grade metamorphic flysch Puncoviscana basement unit, with additional units including Precambrian/Cambrian plutonic granites, Cambrian quartzites, Cretaceous sandstones, Miocene to Pliocene conglomeratic units, and Quaternary fills (Figure 3; García et al., 2013; Marrett et al., 1994; Omarini et al., 1999; Reyes & Salfity, 1973; Schwab & Schäfer, 1976). The Río Toro and its tributaries traverse these units and cross a number of faults, with a broad knickzone at the top of the steep bedrock gorge associated with the Quaternary-active Gólgota fault (Figure 1c; Hilley & Strecker, 2005; Marrett et al., 1994).

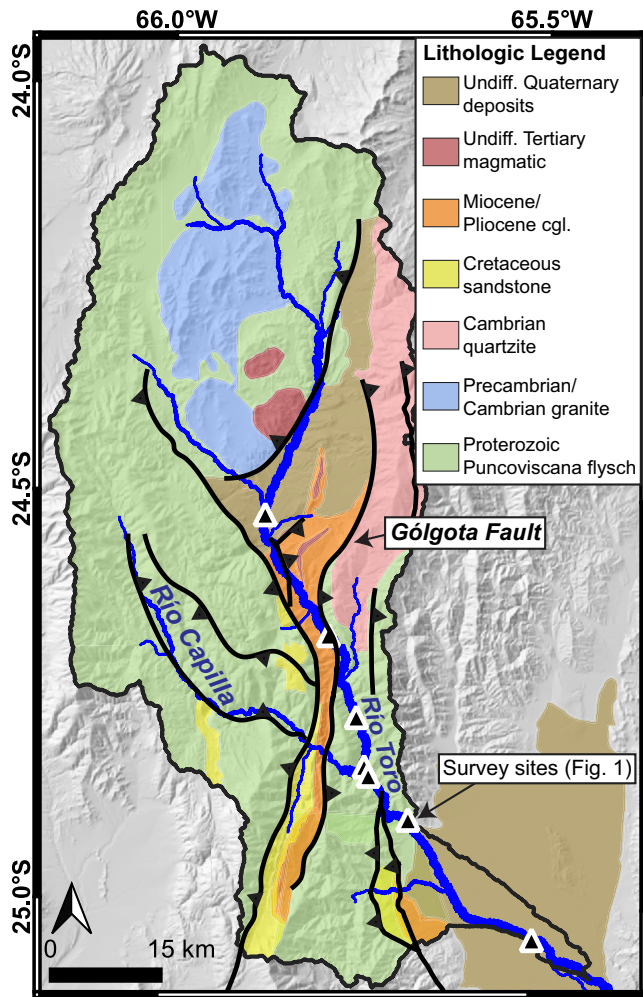
A significant portion of the bedload material in the Río Toro comes from gullying and associated debris flows, with additional delivery by scree production on steep hillslopes, and sparse and infrequent landslides (Tofelde et al., 2018). Upstream of the Gólgota fault is a low-slope region (Figure 1b) of large fluvial terraces, which experiences limited connectivity through the narrow entrance to the steep gorge. The upstream disconnectivity and transport-limited condition is highlighted by the lack of granite and quartzite clasts in the bedload material, despite large outcroppings of these resistant lithologies in the upstream basin (Figure 3). Marrett and Strecker (2000) note the dominance of the Puncoviscana unit in the bedload, particularly downstream of the Gólgota fault. Locally, we find other lithologies downstream of tributary junctions with these outcrops (e.g., Cretaceous outcrops in downstream Río Capilla; cf. Figure 3), but the Puncoviscana meta-sediments remain the primary channel-bed lithology.

### 3. Data and Methods

The data generated for this study are grain-size distributions containing thousands to tens-of-thousands of observations from the channel cross-section survey sites shown in Figure 1. The digital grain-sizing algorithms (*PebbleCounts*) are described in detail in Purinton and Bookhagen (2019a), and are available from the following repository: <https://github.com/UP-RS-ESP/PebbleCounts> (Purinton & Bookhagen, 2019b). Scripts and data associated with the present work are here: <https://github.com/UP-RS-ESP/PebbleCounts-Application> (Purinton, 2021).

*PebbleCounts* consists of two separate algorithms: A k-means manual selection (KMS) tool, which we herein refer to as the manual KMS tool, and an automatic image filtering (AIF) tool, which we herein refer to as the automatic AIF tool. We note the redundancy in the naming scheme, but choose this terminology for consistency with previous work (Purinton & Bookhagen, 2019a) and to remind the reader of the manual versus automatic approaches.





**Figure 3.** Simplified geologic map of the Toro Basin modified after others (García et al., 2013; Marrett et al., 1994; Tofelde et al., 2017, 2018). Of the mapped faults, only the knickzone-generating Gólgota fault is indicated with a label (cf. Figure 1c).

The manual KMS tool provides higher accuracy measurements via manual selection of correctly identified grains prior to ellipse fitting and measurement, whereas the automatic AIF tool provides faster results with less manual input, and higher measurement errors. As opposed to previous algorithms (e.g., Basegrain; Detert & Weitbrecht, 2012), *PebbleCounts* is especially tailored to high-energy mountain river settings.

Throughout the manuscript we refer to false measurements (i.e., misidentified or mismeasured grains) as measurement errors and the uncertainties in calculated percentiles as percentile uncertainties. Because our automatic sampling approach (higher measurement error) guides manual validation sampling (lower measurement error), we present the automatic measurement first.

### 3.1. Survey Pre-processing

*PebbleCounts* survey and processing statistics are summarized in Table 1. The surveys shown in Figure 1b were gathered via a mast-mounted camera (Figure 2a) and processed into seamless, georeferenced ortho-images using structure-from-motion processing in Agisoft (Agisoft, 2018; Purinton & Bookhagen, 2019a). Only the highest quality (lowest blur, sharpest image) parts of the survey were used by cropping the entire survey and buffering any holes (both by 1 m) caused by insufficient camera overlap. This resulted in measurable survey areas of ~1,000–3,500 m<sup>2</sup>, with ortho-image resolutions of 0.9–1.1 mm/pixel. We refer to Figure S2 for an example survey.

As processing many >1,000 m<sup>2</sup> survey sites with *PebbleCounts* requires significant effort (up to two weeks per site), we partially processed our surveys. We did so by first tiling the ortho-image into 2,000 × 2,000 pixel tiles (~2 × 2 m). The value of 2,000 pixels was selected as an efficient number to reduce processing time in *PebbleCounts* (which grows exponentially with the number of pixels in the image due to steps like non-local means de-noising and ellipse filtering), while limiting the number of tiles that needed to be processed.

### 3.2. Counting With *PebbleCounts* Automatic AIF Tool

Having prepared the survey for grain sizing (buffering and tiling), the tiles were fed into the automatic AIF tool in a checkerboard-like pattern, skipping every other row and every other column. This left us with four grids of tiles, and for each site we first processed one of these four grids (Figure S2). Prior to automatic processing, the tiles were pre-screened for quality assurance and skipped if there was excessive vegetation or woody debris, blurring, and/or strong shadowing obscuring grains.

Within *PebbleCounts*, the grains are segmented based on their color and interstices, and an ellipse is fit to this segmented area. The long and short axes of the ellipse are the apparent *a*- and *b*-axes of the pebble, respectively. Per the *PebbleCounts* lower-detection limit of *b*-axis length of 20 pixels, the minimum *b*-axis grain size that we truncated measurements to was 2.5 cm. This is greater than 20 pixels, which corresponds to 1.8–2.2 cm *b*-axis size given the imagery spatial resolutions. We chose 2.5 cm out of an abundance of caution. This conservative truncation allows us to make inter-site grain-size percentile comparisons, while reducing potential percentile uncertainties from under-counting the proportion of smaller grains (cf. Purinton & Bookhagen, 2019a). After counting using the default 20-pixel cutoff, any grains with a *b*-axis size < 2.5 cm were removed from further analysis.

**Table 1**  
*PebbleCounts* Survey Statistics for Grains  $\geq 2.5$  cm in b-Axis Length

| Survey flow distance (km) | Spatial resolution (mm/pixel) | Measurable area (m <sup>2</sup> ) (Tile count) | Automatic AIF tiles processed (Percent of total tiles) | Automatic AIF grains identified | Percentage grains/sand/unmeasured in automatic AIF <sup>a</sup> | Manual KMS validation tiles processed | Manual KMS grains measured (Percent of automatic AIF) |
|---------------------------|-------------------------------|--|--|---------------------------------|---|---------------------------------------|---|
| 0                         | 1.1                           | 3,470 (717)                                    | 178 (25%)  | 6,530                           | 3/82/15   | 14                                    | 1,001 (15%)   |
| 25                        | 1.0                           | 944 (236)                                      | 120 (50%)  | 7,000                           | 10/39/51  | 14                                    | 922 (13%)   |
| 41                        | 0.9                           | 1,331 (411)                                    | 204 (50%)  | 18,975                          | 12/36/52  | 26                                    | 1,985 (10%)   |
| 50                        | 1.0                           | 1,344 (336)                                    | 83 (25%)   | 11,021                          | 12/28/60  | 19                                    | 2,081 (19%)   |
| 52                        | 0.9                           | 1,225 (378)                                    | 186 (50%)  | 16,902                          | 13/37/50  | 20                                    | 1,663 (10%)   |
| 64                        | 1.0                           | 1,476 (369)                                    | 93 (25%)   | 18,173                          | 16/37/47  | 10                                    | 1,917 (11%)   |
| 95                        | 1.0                           | 1,488 (372)                                    | 93 (25%)   | 32,852                          | 27/2/71   | 15                                    | 2,863 (9%)  |

*Note.* Survey flow distance in km refers to Figure 1b. The tiling procedure is described in Section 3.1. The manual KMS identified grains in the final column are the reliable manually selected validation measurements used to assess the performance of the automatic AIF grain identifications.

Abbreviations: AIF, automatic image filtering; KMS, k-means manual selection.

<sup>a</sup>Percentage of tile area with: grain size ( $\geq 2.5$  cm) quantified, masked for smooth sand surfaces ( $< \sim 2$  mm), and unmeasured due to missed grains and/or vegetation and other debris.

We sought at least 50 grains identified in each percentile, including the 99th percentile. This set our counting requirement to  $\geq 5,000$  grains per site. After running the first grid, if  $< 5,000$  grains were counted we processed the next grid until reaching this quota. This was achieved after maximum two of the four grids of tiles for each site, and resulted in a number of sites with  $> 10,000$  measurements (Table 1). Between 1/4 and 1/2 of the cross-section was processed and coverage was evenly spread over the survey areas (Figure S2), which had heterogeneous arrangements of grain-size facies typical of mountain rivers. The number of tiles run through the automatic AIF tool for each site ranged from 83 to 204, representing an areal range of 332–661 m<sup>2</sup> and 6,530–32,852 individual grain measurements (Table 1).

In total, the time from cropping images to completion of automatic processing by a trained expert was between 3 and 6 h per site, depending on the number of tiles. These steps include rapid quality checks on each tile, selection of a shadow threshold, and clicking of a color mask representing smooth sandy areas between the grains. The automatic AIF tool provides not only many thousand grain-size measurements per site, but also an approximate percentage of sand surfaces ranging from 2% to 82% (Table 1).

The measured grains covered only 3%–27% of the total tile area measured, with much of the tile area remaining unmeasured for each site (15%–71%; Table 1). This unmeasured percentage represents total area where neither sand nor grains were measured, since the *PebbleCounts* algorithms do not attempt to measure all parts of the image, but only those grains that can be identified with high confidence. This missing measurement in each tile is made up for by counting many tiles over a large area.

### 3.3. Validation With *PebbleCounts* Manual KMS Tool

To gather accurate manual KMS measurements for validation of the automatic AIF tool, we required a careful tile-selection scheme to select a representative subset of the wide grain-size distribution. We began by considering the grain-size distribution provided by the automatically counted tiles. We first calculated the 95th/50th percentile ratio for each tile in a given site, which allows us to stratify the tiles into coarser (greater ratio value) and finer (lower ratio value, approaching 1) bins. We then created five bins of 95th/50th ratio, where the size (counts) of the histogram bins were equal to the number of automatically counted tiles of a given range. From this, we sub-selected from each 95th/50th ratio bin a relative proportion of 1/4 to 1/8 of the stratified tiles, resulting in anywhere from 0 to 8 tiles per bin (0 tiles were allowed to occur in some bins when the 95th/50th range across tiles was very low, typically  $< 0.5$ ). The bin-wise proportion was iteratively

selected to provide at least 10% of the individual automatic AIF grain measurements in the manual KMS validation tiles (except for the site at 95 km with 9%; Table 1).

We ran each of the 10–26 selected validation tiles through the manual KMS tool, resulting in 922–2,863 grains measured per site (Table 1). This tool requires manual selection (by a single click within the segmented grain boundary) of correctly identified grains on each tile, requiring up to 30 min per tile by a trained expert depending on the number of grains present.

This process ensured that the full range of grain sizes was proportionally sampled for validation and accuracy assessment. This validation scheme can be compared to grain-size stratified random sampling as opposed to random spatial sampling (Bunte & Abt, 2001), because the outlined steps are based on the expected range of grain sizes present, which may be heterogeneously arranged. The automatic AIF tool provides an initial assessment of the grain-size arrangement at a survey location, which is used to guide a manual KMS sampling strategy.

### 3.4. Percentile Measurements of the Coarse Fraction

We use our large sample sizes to produce robust percentiles, like the 50th or 84th percentiles (D50 and D84 in hydraulics notation). Throughout this work we refer explicitly to the percentiles (e.g., 50th) of the coarse ( $\geq 2.5$  cm) fraction of the grain-size distribution. We avoid the D-notation, since this implies a much finer cutoff in measured grains and usually refers to the full distribution of grain sizes identifiable by the naked eye. Calculations of dimensionless shear stress (Shields stress) and thresholds for erosion in fluvial incision laws are often based on specific D-values (e.g., D50 or D84), based on grain-size distributions with a lower truncation (e.g., particles in suspension vs. bedload) or no truncation. We stress that our percentiles will be different to D-values (e.g., D50 vs. our 50th percentile) due to our truncation at 2.5 cm.

Although we do acquire the areal percentage sand ( $< \sim 2$  mm) at each survey (cf. Table 1), it is difficult to justify inclusion of (e.g., through offsetting of the coarse distribution), since this is not an actual grain size, but rather a percentage of area based on color and smooth appearance (no grain edges) in the image. Studies that include the sand fraction do so through sieving in the same area as coarse clast counting (e.g., Attal et al., 2015). Including the sand fraction would still leave out the  $\sim 2$ –2.5 cm fraction, thus creating a large gap in the measurements, which we are not confident in interpolating.

In our case, the truncated distribution represents the coarser gravel, cobble, and boulder fraction of pebbles present in the channel bed. Downstream percentile changes in the sand and silt fraction may be far different from the pebbles, and this finer material is transported during low-flow conditions, whereas the coarser pebble fraction is only transported during sufficiently high discharges. Thus, here we are only interested in the fraction of the bedload, which is moved by stochastic events exceeding the threshold of motion of the pebbles. In making comparisons, we note this distinction to studies using D-notation, but we also note that previous studies are based on small (100–500) clast counts, and are thus unable to accurately determine the uppermost ( $> 95$ th) percentiles of the bedload.

### 3.5. Topographic and Lithologic Data

To draw connections between pebble size measured with *PebbleCounts* and drivers of downstream size changes, we analyzed upstream and local topographic steepness and lithology at each survey site.

#### 3.5.1. Channel Steepness

We first calculated the normalized channel steepness, given by:

$$k_{sn} = SA^{-\theta_{ref}} \quad (1)$$

where  $S$  is channel slope,  $A$  is drainage area, and  $\theta_{ref}$  is a fixed reference channel concavity (e.g., Wobus et al., 2006). We used the  $k_{sn}$  metric to relate with grain-size variability since bedrock fluvial channels in steep mountain environments are sensitive indicators of erosion (e.g., Hilley et al., 2019; Whipple, 2004), which may translate to channel bed grain size.



All drainage network extraction steps and  $k_{sn}$  calculations were done using the open-source LSDTopoTools package (Mudd et al., 2019). Specifically, we used the lsdtchi-mapping tool, which implements the procedures outlined in Mudd et al. (2014), based on the integral approach to channel-steepness mapping of Peron and Royden (2013). First, the drainage network was extracted using a minimum area of 0.1 km<sup>2</sup>. Then, the best fit  $\theta_{ref}$  value for the entire basin was calculated using a combination of chi-disorder and slope-area metrics (Mudd et al., 2018). For the Toro Basin, this was found to be  $\theta_{ref} = 0.4$  and this was used for further calculation. The high-quality, state-of-the-art 12 m TanDEM-X DEM with acquired coverage over the Toro Basin was used for this analysis (Purinton & Bookhagen, 2017; Rizzoli et al., 2017). The  $k_{sn}$  extraction applies smoothing steps to reduce topographic noise, resulting in a spatially averaged steepness along discrete (~100–1,000+ m) channel reaches.

In analyzing the upstream  $k_{sn}$  for each site, we reduced the spatial influence of distant  $k_{sn}$  values on the site's grain-size distribution by limiting the maximum upstream flow distance of considered pixels. We made assumptions based on Sternberg's law (Sternberg, 1875) of abrasion to set the limiting distance. This law states:

$$D = D_0 e^{-\alpha L} \quad (2)$$

where  $D$  is the final size,  $D_0$  is the initial size,  $\alpha$  is the abrasion constant, and  $L$  is the travel distance. This relationship was developed to describe the rate of fining by wear alone, ignoring the potential influence of selective transport (e.g., Paola et al., 1992), but has also shown a consistent relationship in settings with significant sediment sorting and supply changes (Hoey & Bluck, 1999). In the Río Toro, the dominant lithology of the channel bed is the Puncoviscana meta-sediments, and the largest  $b$ -axis measured was ~50 cm. Attal and Lavé (2006) report abrasion rates in similarly weak meta-sedimentary lithologies (schists, including phyllites, sericites, and micaschists) in the Himalaya of 16%/km. Using this abrasion rate in Sternberg's law, a 50 cm boulder of Puncoviscana would pass below our 2.5 cm measurement limit after a travel distance of only ~20 km.

Based on this assumption, we calculated the median and 95th to 100th percentile  $k_{sn}$  of every tributary > 0.1 km<sup>2</sup> that meets the trunk stream; only considering those pixels within 20-km flow distance of the trunk junction. We also tested doubling this limit, but found little difference in the downstream patterns.

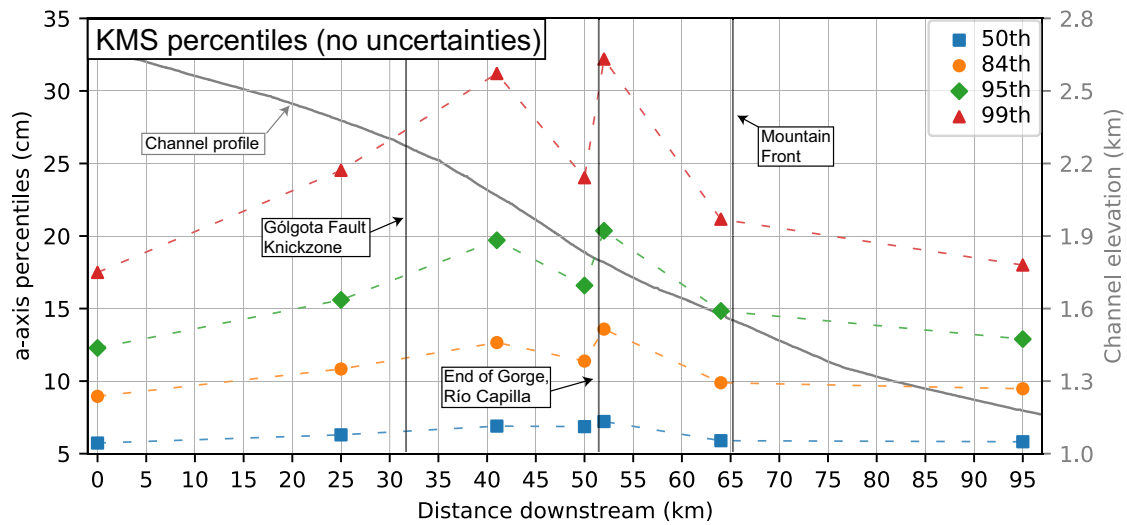
### 3.5.2. Hillslope Steepness

Since local hillslopes are known to significantly affect fluvial grain-size distributions (e.g., Neely et al., 2019; Neely & DiBiase, 2020), we also examined the relationship of grain-size variability to hillslope steepness adjacent to the surveyed trunk stream sites. We again used the 12 m TanDEM-X DEM to calculate slope in a 500-m buffered area on either side of the trunk stream. Prior to calculations, we applied a light Gaussian smoothing filter (standard deviation of 0.5) to reduce the influence of minor spaceborne DEM errors causing outliers in slope calculations (Purinton & Bookhagen, 2017). The 500-m buffer ensured that only the steepest slopes immediately adjacent to the channel were measured. To measure continuous downstream changes along the channel, we took 2-km flow-distance bins and calculated the median and upper-percentile (95th to 100th) hillslope angle in each bin.

### 3.5.3. Lithology

In conjunction with channel and hillslope steepness, the nearby upstream lithology also influences channel-bedload characteristics (e.g., Dingle et al., 2017; Mueller & Pitlick, 2013; Mueller et al., 2016). To assess this, we calculated the percentage of each lithologic unit exposed upstream (cf. Figure 3). Under the same assumptions made with Sternberg's law for channel steepness, we limited the upstream pixel distance to 20-km upstream in each tributary. A clipping mask on the basin lithology was manually digitized using a drainage network of the ≤ 20-km (flow-distance to the trunk junction) channels, and hand-clicking the area including these channels and their hillslopes (up to the nearby ridge-crests).

In addition to this basin-wide lithology, we also calculated the approximate percentage of the dominant Puncoviscana meta-sediment lithology in each pebble survey. This lithology is relatively easy to identify from photos given its purple color, prominent quartz veins, and sub-angular to sub-rounded appearance. Lithology data were not collected in the field, since the goal was to measure very large areas and test new



**Figure 4.** Longitudinal evolution of  $a$ -axis percentiles downstream along the Río Toro. Measured  $a$ -axis percentiles are from the k-means manual selection (KMS) algorithm with 922–2,863 grains measured per site (Table 1). No percentile uncertainties are presented in this first assessment, but these are on the order of 5%–10% from previous calibration against hand-clicked control data (Purinton & Bookhagen, 2019a). Throughout the manuscript, the sites are connected with dashed lines to express possible between-site variability in the trend. Note the longitudinal river profile on the right axis and vertical lines of geomorphic transitions, included to maintain the spatial context of the sites (cf. Figure 1), and used in all subsequent longitudinal figures.

grain-sizing methods. Attempts were made at automatic lithology identification from the photos, but this requires further research and likely the use of multi- or hyperspectral information, since RGB color and segmented shape alone do not capture the variability in pebble lithology. We accomplished a simple lithology count by randomly selecting three tiles run in the automatic AIF tool, with two from either end of the survey and one from the middle. From these tiles, the non-Puncoviscana pebbles were identified (more rounded red, yellow, and white pebbles typically making up < 50% of the total), this number was subtracted from the total measurements in the tile, and a percentage of the Puncoviscana was calculated for each survey site. This percentage is taken from a total of at least 500 quickly viewed pebbles per site.

## 4. Results

The results consist of six analyses in the following order: First, we show initial results of the downstream evolution of grain-size percentiles using the manual KMS measurements. Second, measurement errors in the automatic AIF tool are shown, and we demonstrate validation of these percentiles using the manual KMS control. (To reiterate, we refer to false measurements as measurement errors and the uncertainties in calculated percentiles as percentile uncertainties.) Third, we show the importance of large sample sizes in reducing percentile uncertainties at upper percentiles via binomial theory presented in Eaton et al. (2019). Fourth, we present an alternative approach for reaching these large sample sizes based on log-normal fitting and distribution simulation. Fifth, the downstream evolution of grain-size percentiles is presented taking into account percentile uncertainties. Sixth, we highlight the drivers of grain-size changes through reach-scale analysis of topographic steepness and lithology.

### 4.1. Downstream Percentile Evolution From Manual KMS Tool

We begin by presenting the reliable manual KMS percentiles at our seven survey sites (Figure 4). This figure does not include uncertainties. Previous calibration efforts using >1,000 hand-clicked control grains show maximum percentile uncertainties of ~10% from the KMS tool, though often <5% (Purinton & Bookhagen, 2019a). We note that the increasing uncertainty of higher percentiles (e.g., 95th) shown in Purinton and Bookhagen (2019a) can partially be attributed to the small sample size at these upper values, which we discuss in detail below. The observed grain-size variability between sites is larger than these reported

uncertainties. From Figure 4, we note the high variability of the upper percentiles (95th and 99th), whereas the lower percentiles (50th and 84th) show less variability.

Error bars could be calculated for Figure 4 using a bootstrapping approach (e.g., D'Arcy et al., 2017), but this makes assumptions about the true grain size from the full population. We prefer to consider the uncertainty purely as a function of the sample number. Even 1,000 individual measurements constitute only 10 grains at the 99th percentile. This percentile is subject to considerable uncertainty from small sample sizes when a large range in grain size is present over a wide channel area.

## 4.2. Larger Sample Sizes From Automatic AIF Tool

One way to achieve larger pebble sample sizes is using the automatic AIF measurements. However, this method has possible significant percentile uncertainties due to few control steps during image segmentation. Therefore, these measurements need to be validated against the manual results prior to use.

### 4.2.1. Sources of Measurement Error

Poorer performance of *PebbleCounts* can be expected on lower quality imagery (less sharp, more blur). This can be mitigated by using high-quality camera lenses and quicker shutter speeds during photo capture. Additionally, the more steady the mount (be it handheld, tripod, mast, or UAV), the higher quality the photo result. Aside from image quality that can be controlled for, segmentation algorithms like *PebbleCounts* for grain sizing are error prone in images that contain overlapping grains, a large size range (including sand patches), angular pebbles (non-ellipsoid), pebbles with intra-granular color variations (veins or fractures), and in which shadowing is irregular (both inter- and intra-granular). These factors are all present in many high-mountain study areas like ours.

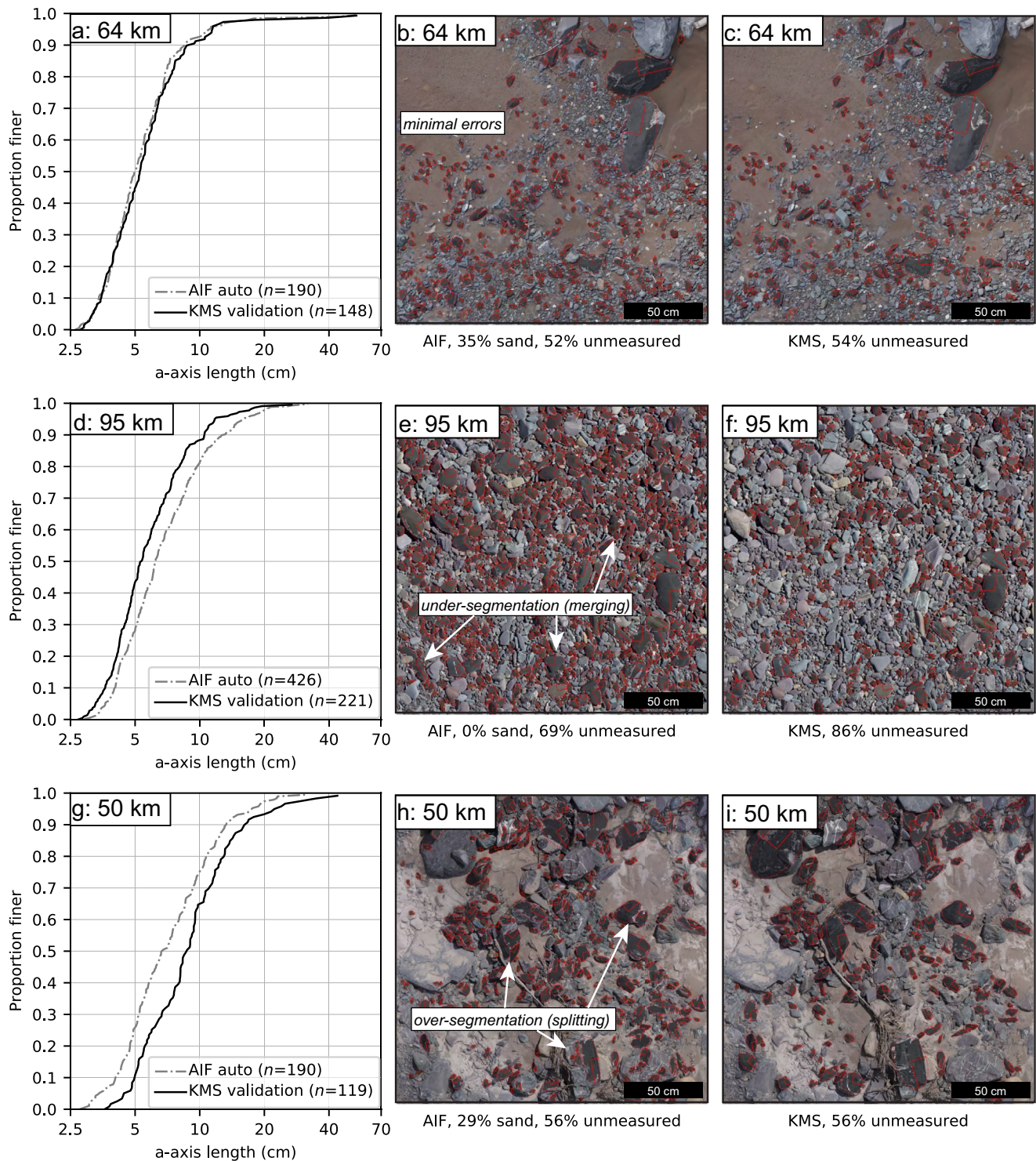
When comparing automatic AIF counted tiles with manual KMS counted tiles during validation steps, we found distinct error patterns (Figure 5). Many tiles performed very well, with the grain-size distribution matching almost perfectly between the two methods (Figures 5a–5c). However, in other cases, the tiles suffered from one of two distinct types of error.

One error type was found on image tiles with little sand between the grains to use for initial segmentation, a more uniform grain size and color, and significant grain overlap. Here, we found a positive bias in the automatic AIF grain-size distribution compared to the manual KMS validation result (Figures 5d–5f). We term this under-segmentation error, and it is caused when smaller grains are merged into larger grains, because the algorithm has trouble segmenting the similar colors with high overlap that do not always have distinct boundaries necessary for edge detection.

The other error type was found where the sand patches between grains were mottled in color—caused by a mixture of dry and wet sand patches with a gradation of tan to brown colors—making sand masking by color more difficult (i.e., some sand areas remained outside of the mask), and where the grains were more angular and contained quartz veins, both of which led to intra-granular edges. Here, we found a negative bias in the automatic AIF grain-size distribution compared to the manual KMS validation result (Figures 5g–5i). We term this over-segmentation error, and it is caused when larger grains are split into smaller grains (because of internal edges) and when many false finer grains are identified, which are actually remaining isolated sand patches.

Generating validation figures like those shown in Figure 5 for every tile run through the manual KMS and automatic AIF *PebbleCounts* tools demonstrated which tiles (and thus which sites) suffered from errors. The first type of error (under-segmentation, positive bias) was least prevalent along the Río Toro (e.g., Figure 5e), where there tended to be abundant sand patches and more angular grains. The second type of error (over-segmentation, negative bias), was somewhat more prevalent in some tiles (e.g., Figure 5h), caused by two factors: (a) the fracturing of the dominant Puncoviscana metamorphosed flysch, which tends to form angular fragments, as well as abundant quartz veins in these rocks, both leading to problematic internal edges; (b) partially dry sand patches mottled in color leading to difficulty forming a uniform-colored mask in the first stage of image segmentation.





**Figure 5.** Under- and over-segmentation errors in *PebbleCounts* shown with images and with grain-size distributions: (a–c) with minimal errors are from the 64-km site; (d–f) with under-segmentation errors are from the 95-km site; and (g–i) with over-segmentation errors are from the 50-km site. K-means manual selection (KMS) refers to manual control data and automatic image filtering (AIF) refers to automatic data.

#### 4.2.2. Validation of Automatic AIF Measurements

Although both types of error (under-/over-segmentation) can occur locally (between tiles) at a given site, we found that, these opposing errors either canceled out or were minimized compared to the correctly identified grains. This is demonstrated in the grain-size distributions and similarity of percentiles for the automatic AIF and manual KMS measurements from all validation tiles for a given site in Figure 6. We reiterate that *PebbleCounts* measures grain size via an ellipse-fitting approach (as shown in the examples in Figure 5), and thus provides both *a*- (long) and *b*- (intermediate) axes. The results for the *b*-axis percentiles are similar and shown in Figures S4 and S5.

Six of our seven survey sites (0, 25, 41, 50, 52, and 64 km; Figures 6a–6l) had excellent agreement between the automatic AIF and manual KMS validation percentiles, with  $r^2$  values (indicating deviations from a 1:1 relationship) of  $\geq 0.99$ . Thus, the automatic AIF *PebbleCounts* tool performed very well when considering percentiles, despite containing many misidentified grains. Furthermore, it is shown in the Figure S3 that similar results are obtained when including all 6,530–18,975 automatic AIF measurements for these six sites, indicating that the carefully chosen stratified sub-sample of validation tiles (comprising ~10%–20% of the total automatic AIF measurements; cf. Table 1) produces similar percentiles to the full automatic AIF distribution.

Only one survey site (95 km; Figures 6m and 6n) showed significant percentile deviations of the automatic AIF measurements. Here, the  $r^2$  value was 0.87 and the  $\geq 50$ th percentiles clearly deviate from the 1:1 relationship, showing greater values in the automatic AIF measurement. Upon visual inspection, we found the reason for the poorer AIF result at 95 km was due to similar colored grains with high overlap, which the automated segmentation and filtering struggled with, leading to under-segmentation.

We emphasize the importance of validation steps. At six sites we can confidently extract large sample sizes from the automatic AIF measurements, but at one site (95 km) we can only confidently use the manual KMS measurements ( $n = 2,863$ ).

#### 4.3. Percentile Uncertainties From Binomial Theory

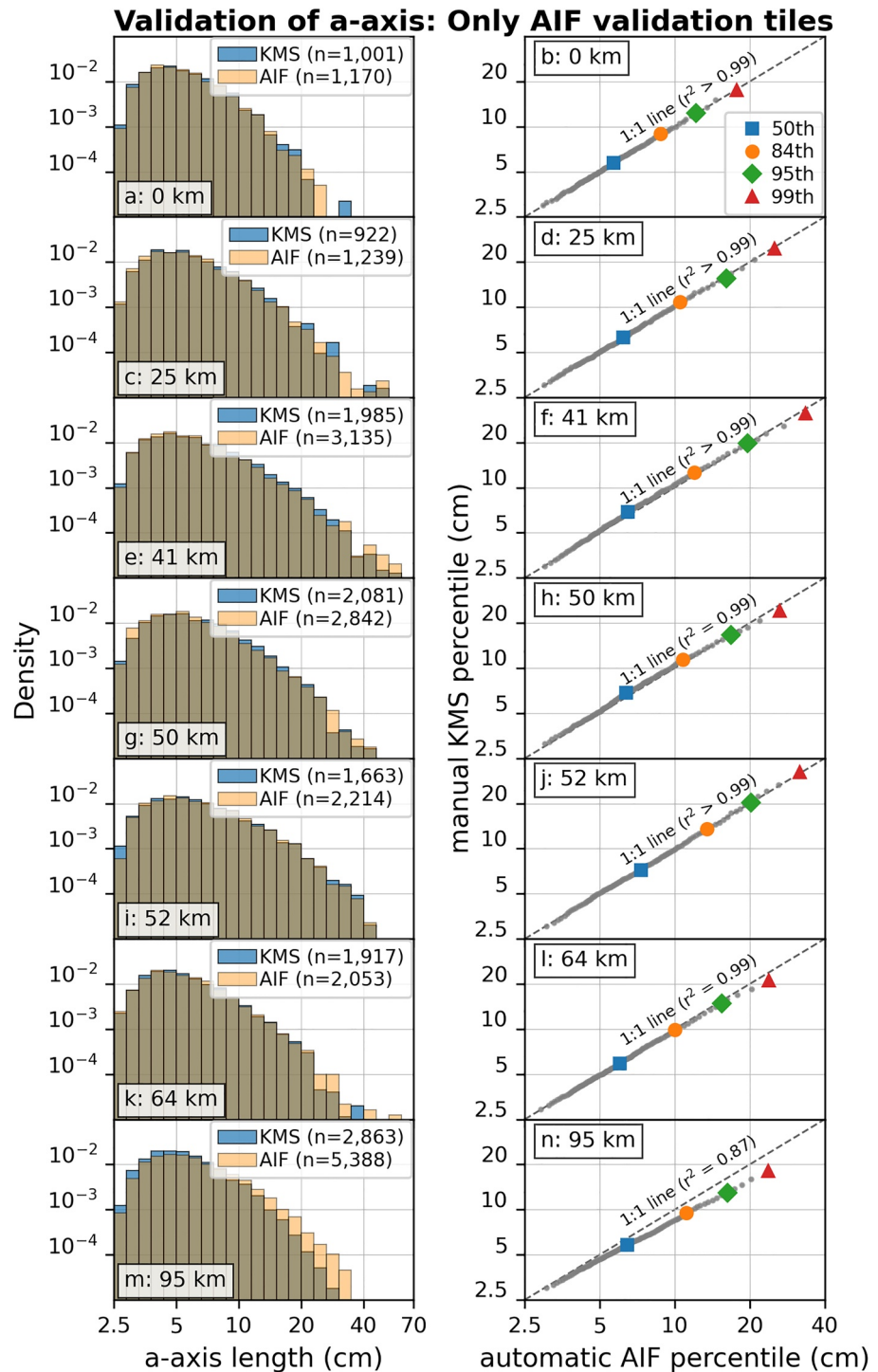
What is the utility of large sample sizes for studying grain-size variability in high-mountain settings? We answer this question by turning to recent work by Eaton et al. (2019) to robustly estimate uncertainties in grain-size percentiles. Their method treats every percentile measurement as a collection of coin flips with an equal probability of heads or tails (true or false), modeling the result with a binomial distribution. Modeling pebble counting as a binomial experiment (where every grain selection is a coin flip) relies on the fact that the probability of sampling a given grain size is proportional to the relative bed area covered by that size. Using only the number of measurements and the percentile of interest, percentile confidence intervals at a given level (e.g., 95%) can be estimated and these intervals can be transferred to a grain-size range using the cumulative frequency distribution. For details and derivation, we refer to the study of Eaton et al. (2019). For convenience, we translated the published *R* code (Eaton & Moore, 2019) to *Python*, and made it available here: <https://github.com/UP-RS-ESP/PebbleCounts-Application>.

Using only the sample size and the cumulative distribution of grain sizes, we can plot the change in percentile uncertainty as a percentage versus increasing sample size using Equation 3 in Eaton et al. (2019):

$$\epsilon_p = 0.5 \left( \frac{d_u - d_l}{d_p} \right) \quad (3)$$

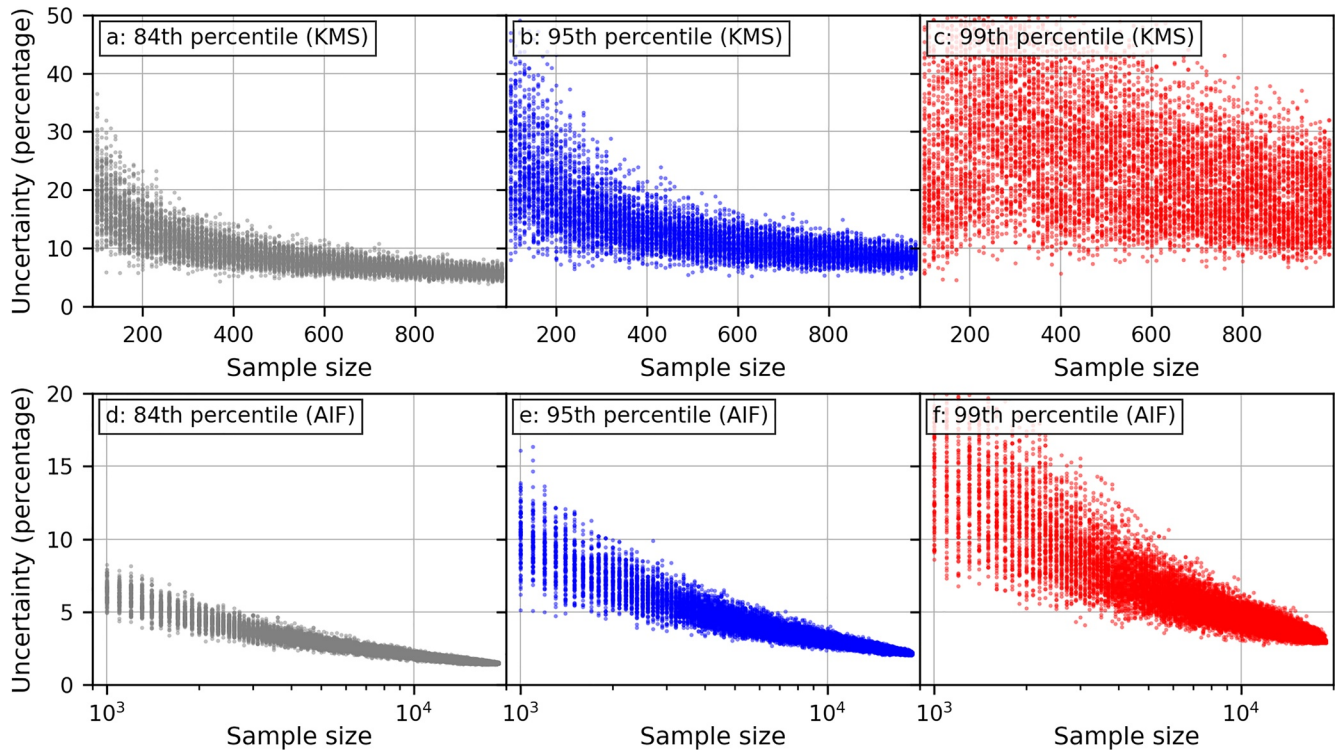
where  $\epsilon_p$  is the percentile uncertainty as a percentage about the percentile of interest  $d_p$ , and  $d_u$  and  $d_l$  are the upper and lower confidence intervals, respectively, calculated from the binomial distribution. Plots using binomial confidence intervals and this equation for the manual KMS and automatic AIF measurements are presented in Figure 7 for the 41-km site (results were comparable for all sites). Here, we randomly sampled the full distributions 100 times for each sample size and plotted each result.

In Figures 7a–7c, the decrease in percentile uncertainty for the 84th, 95th, and 99th percentiles are presented using 100–1,000 manual KMS measurements in 10 sample steps (i.e., 100, 110, ..., 990, 1,000). Beyond ~600 samples the 84th percentile uncertainties are consistently below  $\pm 10\%$ . However, even with 1,000 samples,



**Figure 6.** Grain-sizing validation results. Left column shows the binned distribution of *a*-axis grain size for the automatic image filtering (AIF) and k-means manual selection (KMS) tools over all the validation tiles for each site. The same log-spaced bins are used across all plots. Right column plots the 1st–99th percentiles against each other (gray points) for the two methods, highlighting the 50th, 84th, 95th, and 99th percentiles with different symbols and colors. A dashed line and the  $r^2$  indicates the deviation from a perfect 1:1 relationship.





**Figure 7.** Decreasing percentile uncertainty at the 84th, 95th, and 99th percentiles with increasing sample size calculated for 100–1,000 k-means manual selection (KMS) measurements (a–c) and 1,000–18,000 automatic image filtering (AIF) measurements (d–f) at the 41-km survey site. Note the change in y-limit from 0–50 to 0–20 and the change in x-scale from linear to logarithmic from the first to the second row.

the 95th and 99th percentile uncertainties remain high. The 99th percentile has percentile uncertainties of up to  $\pm 30\%$  over this entire range. In Figures 7d–7f, the decrease in percentile uncertainty for the 84th, 95th, and 99th percentiles are presented using 1,000–18,000 automatic AIF measurements in 100 sample steps. In this case, the upper percentiles (95th and 99th) approach or pass below  $\pm 5\%$  percentile uncertainty only after  $\sim 10,000$  samples.

This answers the posed question: The utility of large sample sizes for studying grain-size variability in high-mountain settings is to reduce percentile uncertainties and assess variability at the uppermost grain-size percentiles (e.g.,  $\geq 95$ th). In the next section, we present an alternative method for artificially simulating larger sample sizes from the manual KMS distribution without using the automatic AIF results, which can have significant measurement errors.

#### 4.4. Simulating Larger Samples Using Log-Normal Fitting

A common, verified assumption is that fluvial grain-size distributions follow a log-normal analytical shape (e.g., Allen et al., 2017; Bunte & Abt, 2001; Sklar et al., 2006). Therefore, we propose an additional method for achieving larger sample sizes and thus reducing percentile uncertainties at the upper percentiles: analytical distribution fitting to the empirical *PebbleCounts* measurements and simulation of larger samples from these smoothed fits.

The two parameter log-normal probability distribution function is defined as:

$$f(x, \sigma, \mu) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \quad (4)$$

with the shape of the distribution  $\sigma$  and the scale  $\mu$ .

**Table 2**  
Fit Statistics of the Log-Normal Distributions Shown in Figure S6

| Survey flow distance (km) | Log-normal fit to KMS shape parameter $\sigma \pm$ standard error | Log-normal fit to KMS KS-test $p$ -value | Log-normal fit to AIF shape parameter $\sigma \pm$ standard error | Log-normal fit to AIF KS-test $p$ -value |
|---------------------------|---|--|---|--|
| 0                         | 0.6368 $\pm$ 0.0029   | 0.96                                     | 0.6667 $\pm$ 0.0004   | 0.90                                     |
| 25                        | 0.7764 $\pm$ 0.0022   | >0.99                                    | 0.7550 $\pm$ 0.0004   | 0.85                                     |
| 41                        | 0.8146 $\pm$ 0.0010   | >0.99                                    | 0.8488 $\pm$ 0.0003   | 0.64                                     |
| 50                        | 0.6789 $\pm$ 0.0016   | 0.79                                     | 0.7712 $\pm$ 0.0002   | 0.90                                     |
| 52                        | 0.8095 $\pm$ 0.0011   | >0.99                                    | 0.8566 $\pm$ 0.0002   | 0.85                                     |
| 64                        | 0.7711 $\pm$ 0.0011   | >0.99                                    | 0.7262 $\pm$ 0.0001   | 0.99                                     |
| 95                        | 0.6840 $\pm$ 0.0009   | 0.97                                     | n/a   | n/a                                      |

Note. The analytical log-normal scale parameter  $\mu$  is included in Table S1. For the automatic AIF results, we exclude the 95-km survey site with greater measurement error. A  $p$ -value closer to one indicates high similarity between the empirical data and fitted distribution. Note that 95% confidence intervals for the parameters can be approximated as two times the standard error.

Abbreviations: AIF, automatic image filtering; KMS, k-means manual selection.

Assuming that measurement errors are negligible on the manual KMS grain-size distribution, we fit a log-normal distribution to each survey site with 922–2,863 measurements (cf. Table 1). This provides us with estimates for the shape and scale parameters. The goodness-of-fit is assessed by low (<5%) standard errors about the fit parameters, and the one sample Kolmogorov-Smirnov test (KS-test) is used to check the null hypothesis that the empirical measurements are drawn from the fitted analytical distribution. For the KS-test,  $p$ -values >0.05 indicate that we cannot reject the null hypothesis at the 95% confidence interval, with  $p$ -values approaching 1 indicating excellent agreement between the empirical and analytical distribution. After fitting, we randomly sample the analytical distribution to 50,000 samples (a number chosen to reduce all percentile uncertainties to well-below 5%; cf. Figure 7), from which we calculate percentiles of interest.

We could extract exact percentiles from the analytical distribution, but our usage is more nuanced. We seek to re-create the random sampling of a log-normal grain-size distribution akin to fieldwork on a real river, but with a very large sample size (50,000). This then allows us to apply the sample-size based percentile uncertainties, which exact analytical percentiles would not.

This fitting and simulation procedure is then repeated independently on the empirical AIF measurements, with the knowledge that misidentification errors may be present. Despite this, the validation results

(Figure 6) indicate that the empirical KMS and AIF distributions correspond very well, aside from the 95-km site. Therefore, we only carried out the log-normal fitting to the empirical automatic AIF measurements for the first six survey sites with 6,530–18,975 measurements (cf. Table 1).

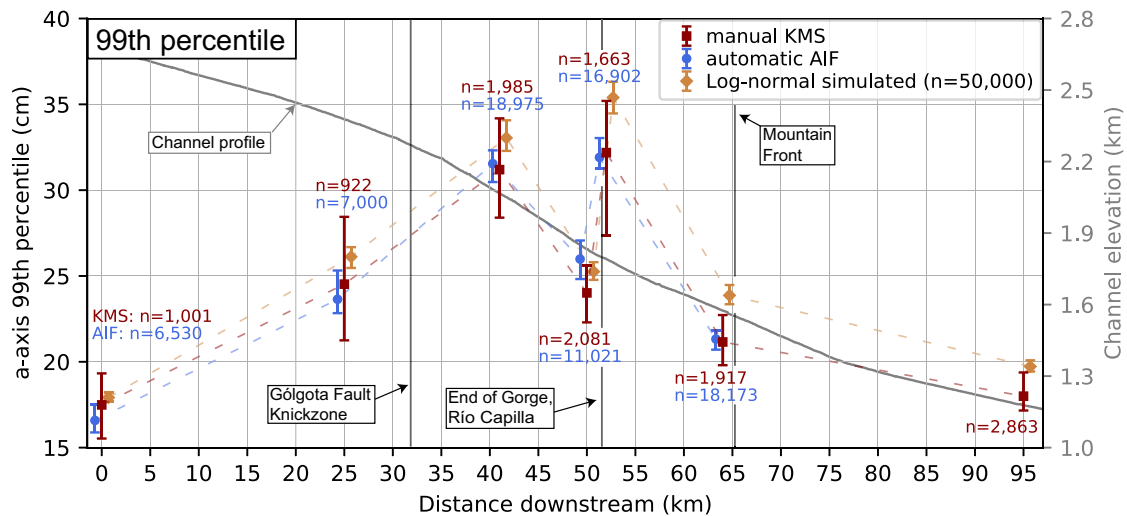
In Table 2 we list important fit statistics, and in Figure S6 we plot the results of this exercise for each site. Only the shape parameter  $\sigma$  is shown in Table 2, since this determines the stretching of the distribution toward higher magnitude upper percentiles (Table S1 includes the scale parameter  $\mu$ ). For all seven sites, the KS-test  $p$ -value was  $\geq 0.64$  for both the manual KMS and automatic AIF fits, and the standard errors about the fit parameters were below  $\pm 5\%$  (Table 2 and Table S1). In Figure S6, we show the correspondence between the empirical and analytical percentiles using a percentile-percentile plot of the 90th–99th percentiles, similar to the second column of Figure 6. For all fits (KMS and AIF) the  $r^2$  of a 1:1 fit between simulated and empirical percentiles was  $\geq 0.95$ . Furthermore, using the two different sets of fit parameters (AIF and KMS) to form random samples for each site, the results of a two-sample KS-test is consistently  $> 0.05$ . Therefore, the fit parameters for both sets of empirical results produce statistically similar distributions; but because they were created from different empirical data sources they are treated separately.

Notably, the  $b$ -axis results (Figure S7) demonstrated poorer log-normal fits compared with the  $a$ -axis. This is expected, since the grains were truncated by their  $b$ -axis measurements, which modifies the  $b$ -axis distribution and removes the log-normal shape via a sharp 2.5 cm cutoff, whereas the  $a$ -axes of these grains have a smoothly decreasing, log-normal tail. Therefore, the presented method for simulating larger sample sizes to reduce percentile uncertainties at higher percentiles is only valid on  $a$ -axis distributions, assuming that these pebbles were sharply truncated by their  $b$ -axis lengths.

#### 4.5. Downstream Percentile Evolution of Grain Size

Having both explored the measurement errors of grain sizing via *PebbleCounts* and highlighted the importance of large sample sizes in reducing percentile uncertainties in the uppermost percentiles, we now return to assessing the variability in downstream grain-size changes (or downstream evolution).

Figure 8 demonstrates the options for 99th percentile measurement: (a) the empirical manual KMS measurements with only high quality grains manually selected leading to sample sizes of 922–2,863; (b) the



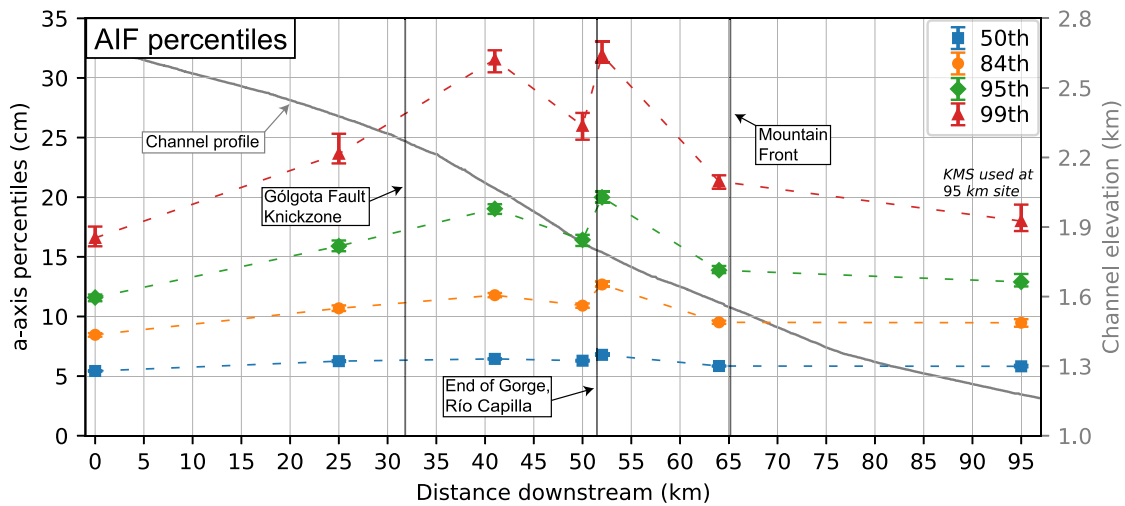
**Figure 8.** Longitudinal evolution of *a*-axis 99th percentile downstream along the Río Toro using the k-means manual selection (KMS) measurements, automatic image filtering (AIF) measurements (except at the 95-km site, where these measurements did not pass validation), and 50,000 log-normal simulated samples from the manual KMS fit. For visibility the 2–3 points at each site are offset by  $\pm 0.5$  km upstream and downstream. Percentile uncertainty error bars represent the 95% confidence intervals from binomial theory (cf. Section 4.3).

empirical automatic AIF measurements, including a number of misidentified grains, validated against the manual KMS measurements leading to sample sizes of 6,530–18,975; and (c) 50,000 simulated log-normal samples for each survey based on an analytical fit to the manual KMS measurements. In this figure, we do not include the fourth option: 50,000 simulated log-normal samples based on the automatic AIF fit for the sites excluding 95 km. As expected from validation (Figure 6) and log-normal fitting results (Figure S6 and Table 2), when simulating the larger sample sizes and plotting percentiles downstream, the automatic AIF and manual KMS log-normal fits provide similar, overlapping patterns (cf. Figure S9). We note that at 52, 64, and 95 km there appears to be a small over-estimation discrepancy at the 99th percentile with the log-normal approach. However, this is minor (often with overlapping error bars to the other methods) and the fit is expected to be slightly worse at the tails of the distribution, where the least number of measurements are found.

Although the overall downstream grain-size patterns are similar between the methods, there are clear differences in the percentile uncertainties, which depend on sample size (Figure 7). In order of decreasing percentile uncertainty (worst to best) are the manual KMS measurements, then the automatic AIF measurements, and finally the 50,000 log-normal simulated samples. The confidence intervals for the manual KMS measurements cover a range of  $\sim 3$ – $8$  cm ( $\sim 8\%$ – $20\%$ ), whereas the larger sample methods have a maximum range of  $\sim 3$  cm ( $\sim 4\%$ – $8\%$ ). As opposed to the uncertainty-free manual results presented in Figure 4, it is clear the inclusion of percentile uncertainties warrants caution in interpreting the downstream signals from the manual KMS data. On the other hand, the larger sample size methods allow confident deciphering of upper percentile variability well outside the range of percentile uncertainty.

We continue using the automatic AIF measurements, except at the 95-km site, where we use the smaller manual KMS samples. This choice of the automatic AIF measurements versus the log-normal simulated is based on our confidence in the validation shown in Figure 6 and our goal to work with empirical data. Using the 50,000 simulated log-normal samples would not change the patterns or discussion of the results (cf. Figures S8 and S9).

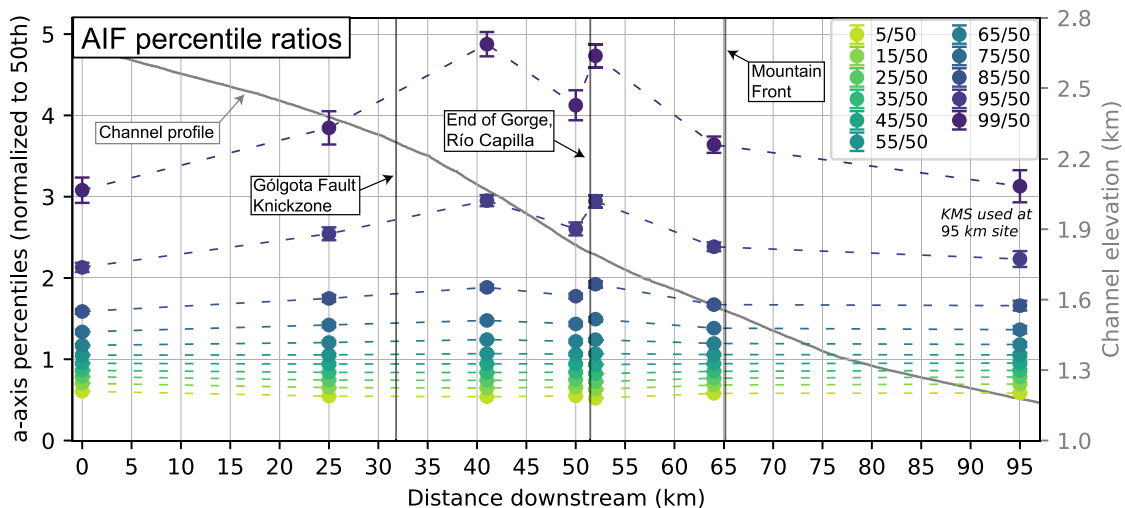
Figure 9 shows representative percentiles (50th, 84th, 95th, and 99th) for the automatic AIF *a*-axes. This is similar to our initial results in Figure 4, but in Figure 9 we use percentile uncertainties and the larger automatic AIF measurements (except at 95 km). The increasing range of values at increasing percentiles is clear. The downstream range of 50th percentile values vary by  $< 1$  cm ( $\sim 20\%$ ), the 84th percentile by  $< 5$  cm ( $\sim 50\%$ ), the 95th percentile by  $\sim 9$  cm ( $\sim 70\%$ ), and the 99th percentile by  $\sim 16$  cm ( $\sim 100\%$ ). The *b*-axis results differ only in magnitude and not in downstream pattern (Figure S10).



**Figure 9.** Longitudinal evolution of *a*-axis percentiles downstream along the Río Toro. Measured *a*-axis percentiles are from the automatic image filtering (AIF) algorithm with 95% confidence intervals from binomial theory (cf. Section 4.3).

The increased variability toward higher percentiles is highlighted by the ratio plots in Figure 10. Here, every tenth percentile (from 5 to 95), plus the 99th percentile, is normalized to the median (50th percentile). The *b*-axis results are similar and presented in Figure S11. Notably, the normalized lines show an identical range between the *a*- and *b*-axes (e.g., 3–5 range of upper line in Figures 10 and S11), indicating that the pebble shape does not change downstream (the *a*- and *b*-axis ratios remain approximately constant with respect to one another).

The relative difference in percentiles as their distance from the median is referred to herein as stretching or compression of the grain-size distribution. Values closer to one indicate compression and farther from one indicate stretching. Consider the 95th/50th or 99th/50th ratio lines (Figure 10). These show stretching factors of 2–3 and 3–5, respectively, meaning the upper-percentile is 2–3 or 3–5 times larger than the median. These descriptive terms are preferred, since the term sorting in fluvial geomorphology more often references statistics around the central distribution (e.g.,  $D_{75}/D_{25}$ ) in which finer material (<1 cm) has been measured and the lower and upper quartiles of the distribution are considered.



**Figure 10.** Longitudinal evolution of *a*-axis percentile ratios downstream along the Río Toro. Ratios are median normalized with the 95% confidence intervals from binomial theory (cf. Section 4.3 and Figure 9). Values closer to one indicate compression and farther from one indicate stretching.



Some downstream patterns are apparent when considering Figure 10. Note that the sites are connected with dashed lines to highlight the fact that between-site variability may differ from the observed trend at the endpoints of each reach. Between the 0- and 41-km sites (toward the steep gorge), there is an increase in percentile stretching. Within the steep bedrock gorge between the Gólgota fault and Río Capilla junction (cf. Figure 1), the stretching is high, but fluctuates to a somewhat lower value at the 50-km site. Toward the mountain front, the percentiles show rapid compression at the 64-km site. This pattern of compression is continued toward the 95-km site, but is lower in magnitude.

#### 4.6. Reach-Scale Characteristics Driving Grain-Size Variability

To explore the landscape drivers of stretching and compression, we continue plotting only the 99th/50th ratio, since this has the largest variability. Figure 11 shows the geomorphic setting that generates the pattern of the 99th percentile grain-size variability. Although there is local fluctuation, the trunk stream has high  $k_{sn}$  values above  $100 \text{ m}^{0.8}$ , and is thus able to transport coarse material. Local hillslope angles also increase into the gorge downstream of the Gólgota fault, which generates a broad channel knickzone, then remain high to the mountain front, with a consistent median of  $40^\circ$ ; the angles show some variability at the 95–100th percentile, but this largely mirrors the median trend.

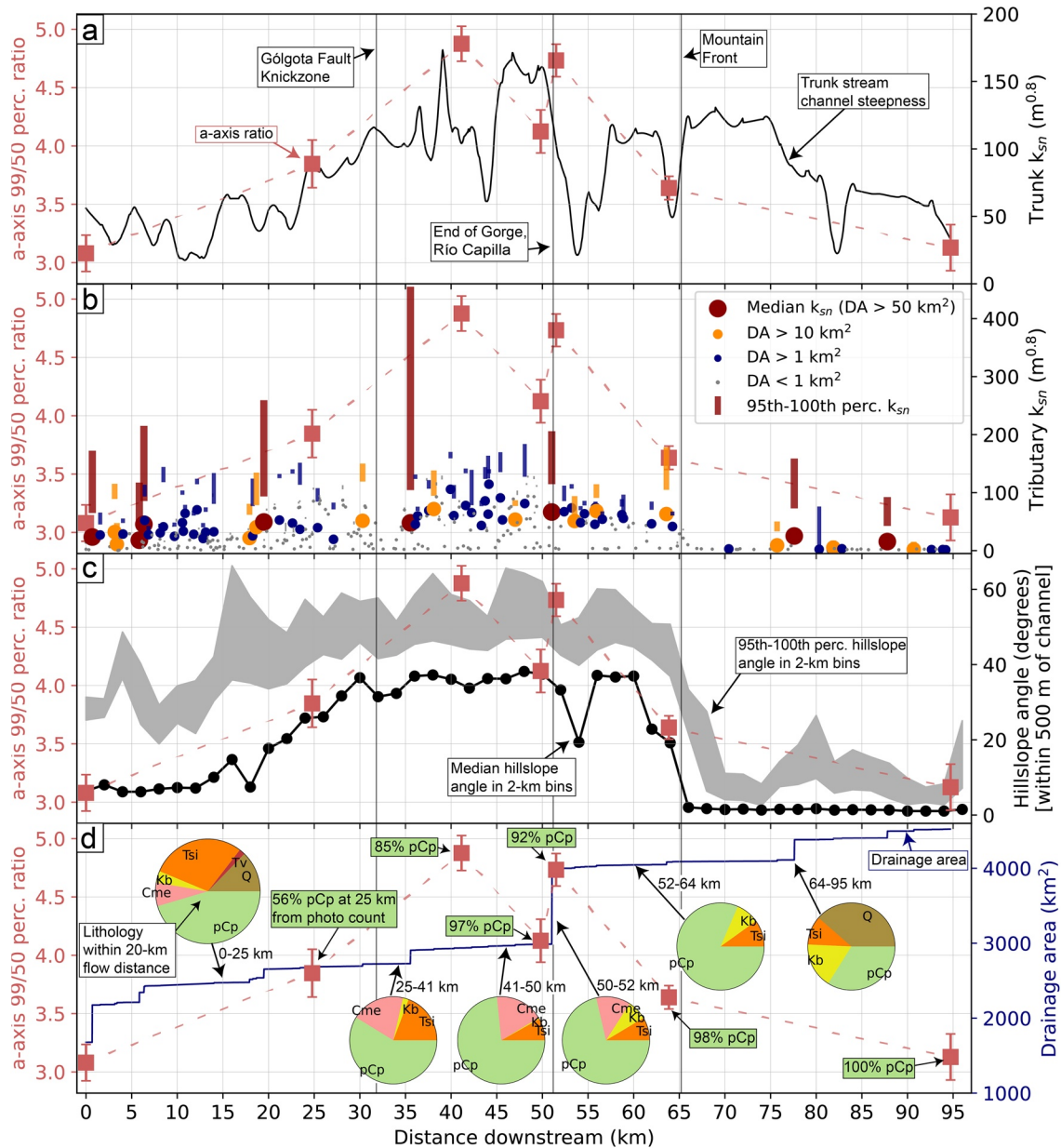
For survey sites with increased stretching of the grain-size distribution to higher percentiles, we observe upstream tributaries with drainage areas  $>50 \text{ km}^2$  and high 95–100th percentile  $k_{sn}$  values within 20-km flow distance of their junction with the trunk stream. In the upstream Toro Basin, diverse lithologies outcrop and hillslope angles and  $k_{sn}$  (trunk and tributary; Figures 11a and 11b) increase toward the gorge. We observe percentile stretching from the 0- to the 25-km site, and this trend continues downstream to the 41-km site, where a large upstream tributary (junction at  $\sim 35 \text{ km}$ ) with  $k_{sn}$  values over  $400 \text{ m}^{0.8}$  brings Cambrian quartzites and large Puncoviscana boulders into the Río Toro (Figures 11b and 11d).

At the 50-km site the  $a$ -axis ratio sharply compresses. Upstream of 50 km no large tributaries with high  $k_{sn}$  values join the trunk and the percentage of Puncoviscana in the lithologic pie chart increases (Figure 11d). The Cambrian unit and steep upstream tributary at 35 km may still somewhat contribute to the grain-size distribution at 50 km, but this junction is 15-km upstream and the Cambrian outcrops are even farther upstream of the junction. At 50 km, we posit that the delivery of finer material in landslide and scree deposits compresses the 99th percentile in the absence of greater sediment supply from large and steep tributaries. This compression can also be related to the supply of initially finer grained Puncoviscana meta-sediments from steep hillslopes with many  $50$ – $60^\circ$  slopes upstream. In the gorge, cataclastic zones of heavily fractured Puncoviscana are generated by tectonic processes (cf. cross-cutting and parallel faults to gorge in Figure 3), and this material thus has a finer grain size to start out with.

The sudden increase in  $a$ -axis ratio (stretching of the distribution) at the 52-km site after crossing the Río Capilla junction ( $<1 \text{ km}$  upstream of this site) is expected given its similar characteristics to the 41-km site. At 52 km, the Río Capilla has high upstream  $k_{sn}$  (95th–100th percentile values of  $\sim 150$ – $200 \text{ m}^{0.8}$ ), and introduces Cretaceous sandstones to the Río Toro (yellow Kb sliver in the lithology pie chart in Figure 11d). 12-km downstream of this site at the mountain front (64-km site), percentile compression occurs, akin to the 41- to 50-km site transition. At 64 km, this coincides with no large and steep upstream tributary junctions (Figure 11b), increases in the upstream Puncoviscana (Figure 11d) on steep hillslopes (Figure 11c), and fining of the coarse Cretaceous sandstone brought in by the Río Capilla.

Despite large tributaries and Cretaceous sandstone and Miocene/Pliocene conglomeratic supply from 64 to 95 km, we observe continued percentile compression at the 95-km site. This can be explained by the low hillslope angles in the Lerma Valley downstream of the mountain front, as well as the fact that the large tributary supplying these units meets the trunk stream over 15-km upstream of the survey, thus abrading and/or selectively sorting these units significantly (cf.  $\sim 500 \text{ km}^2$  drainage area increase in Figure 11d at  $\sim 77$ -km distance).

The pie charts in Figure 11d only contain the lithology from trunk-stream adjacent hillslopes and tributary hillslopes within a flow distance of 20-km from their trunk junction. Thus, this is the lithology likely making up the coarse fraction of the bedload ( $\geq 2.5 \text{ cm}$ ) that we are able to measure. Additional material from



**Figure 11.** (a) 99th/50th percentile ratio for  $a$ -axis surveys with Río Toro trunk stream  $k_{sn}$  ( $\theta_{ref} = 0.4$ ) plotted on right axis. The  $a$ -axis ratio is plotted on each sub-plot on the left axis, with dashed lines connecting the surveys. (b) Median  $k_{sn}$  of every tributary with drainage area (DA)  $> 0.1$  km<sup>2</sup> meeting the trunk stream with symbols sized by DA and vertical bars above each symbol showing the 95th to 100th (maximum) percentile of  $k_{sn}$ . These are only the  $k_{sn}$  values within a  $\leq 20$ -km flow distance of the trunk junction. (c) Hillslope angle in a 500-m buffer on either side of the trunk stream, with 2-km flow distance bins at the median and the 95th to 100th percentile values shown by gray envelope. (d) Drainage area along the trunk stream. Lithologic pie charts show the exposed bedrock within 20-km flow distance. Green boxes pointing to each survey site indicate percentage Puncoviscana from photo counting on three image tiles per site. Colors correspond to Figure 3 geologic map with symbols: Q—undifferentiated Quaternary; Tv—undifferentiated Tertiary magmatics; Tsi—Miocene/Pliocene conglomerates; Kb—Cretaceous sandstone; Cme—Cambrian quartzite; pCq—Precambrian/Cambrian granite; pCp—Proterozoic Puncoviscana flysch.

bedrock at greater flow distances is still a large fraction of the bedload at these survey sites, but this material is likely mostly below the measurement limit and is otherwise masked by the abundant coarse material introduced nearer to the survey sites. The approximate proportion of Puncoviscana in the  $\geq 2.5$  cm bedload at each survey indicated by the green boxes in Figure 11d is taken from photo counting on three tiles randomly selected from each location, as described in the methods. The relative trends in these values track with the relative trends of Puncoviscana in the pie charts for upstream lithology, aside from the 95-km site, where

the Quaternary alluvium in the intermontane basin of the Lerma Valley (Hain et al., 2011) includes a large proportion of reworked Puncoviscana material.

## 5. Discussion

Accurate grain-size distributions from dynamic high-mountain rivers are challenging to acquire. This is largely related to the issue of gaining representative sample sizes to capture the tails of the distribution. Our method allows many measurements on complex gravel arrangements, and our steep and dynamic sand- and gravel-bed river in the south-central Andes provides an ideal environment to test the method, with implications for downstream fining in similar environments.

Our results demonstrate the measurement errors in *PebbleCounts* and how the automatic image filtering (automatic AIF) data can be validated with k-means manual selection (manual KMS) control data. Large sample sizes from empirical or log-normal simulated data reduce uncertainties at the uppermost percentiles and allow observations of between-site variability well outside of uncertainty bounds. When plotted downstream, the grain-size distributions exhibit the most significant changes at the highest percentiles. Topographic steepness, local lithology, and sediment supply can explain the observed patterns.

This discussion focuses on the limits and caveats of the proposed methods, including consideration of additional uncertainties. Following this, the landscape drivers of grain-size variability are further justified, and the results are placed into context of other studies on downstream grain-size evolution in high-mountain settings.

### 5.1. Benefits, Caveats, and Limits of *PebbleCounts*

Challenges in measuring grain size in dynamic mountain rivers are due to difficulties attaining sufficient sample sizes that represent the site-specific grain-size range. The suggested number varies in the range of ~100–500 individual clasts (e.g., Bunte & Abt, 2001; Eaton et al., 2019; Rice & Church, 1996, 1998; Wolcott & Church, 1991). Added to this manual effort is the selection of a sampling strategy covering a wide area of the channel bed and sampling all represented facies (Wolcott & Church, 1991). This is often done with Wolman-style gridded random sampling (Wolman, 1954), along a line orthogonal to flow direction (e.g., Wohl et al., 1996), or using all grains in many small ~1 m<sup>2</sup> areas (e.g., Bunte & Abt, 2001). Even with careful planning and collection of hundreds of measurements, the uncertainties at upper percentiles can remain significant (Figure 7).

The primary benefit of *PebbleCounts* surveying and counting, is to gain large sample sizes of thousands to tens-of-thousands of grain-size measurements over an entire channel cross-section (Figure S2). Increasing sample sizes via analytical log-normal fitting and simulation provides an alternative, smoothed method, but this assumes a model for the data rather than relying on the empirical results. Both methods lead to narrow confidence intervals using binomial theory (Eaton et al., 2019), and, when viewing the downstream evolution of percentiles, significant differences are clear (Figure 9). The percentiles of these large sample sizes better represent the true percentiles at a given survey location. The 99th percentile only contains 1–5 measurements for manual counting of 100–500 pebbles, whereas here they contain 29–190 measurements (from 2,863 total manual KMS measurements used at the 95-km site to 18,975 total automatic AIF measurements used at the 41-km site; cf. Figure 8).

Digital grain sizing is minimally invasive, requiring only photos, and pebble counting can be done later from indoors. Even if validation reveals low quality of the automatic AIF measurements (e.g., for the 95-km site in Figures 6m and 6n), the automatic AIF grain-sizing approach provides a first pass at the grain sizes in the area. This can be used to guide a sampling strategy for higher quality manual KMS grain-sizing (cf. Section 3.3 and Figure S2) without the need for field decisions about representative facies (e.g., Bunte & Abt, 2001; Wolcott & Church, 1991). Instead, we sample all facies through a spatially continuous survey. Additionally, as opposed to traditional field recording, digital grain sizing is repeatable and reproducible using the photographs collected.

Other recent efforts use lidar or structure-from-motion point clouds and/or imagery to measure grains (e.g., Carbonneau et al., 2018; Chardon et al., 2020; Lang et al., 2020; Woodget et al., 2018). These are useful,

but they still rely on texture-based rather than segmentation-based approaches to grain sizing (Purinton & Bookhagen, 2019a). As opposed to the *PebbleCounts* method of segmenting grains in an image of known resolution (scale) and providing the full grain-size distribution, texture-based methods require site-specific manual calibration, only provide a few percentiles of interest rather than the full grain-size distribution, and are subject to greater percentile uncertainties compared to individual grain measurements. Ultimately, texture-based methods provide a spatially integrated measurement of grain size with limited use, whereas segmenting images provides much more detail of grain-size variability. We note that the results from *PebbleCounts* could be used to better calibrate texture-based methods to explore a given percentile over a larger area (e.g., Detert et al., 2018).

Measurement errors (misidentifications) do occur with the automatic AIF tool, particularly in more complex channel areas (Figures 5d–5i). However, these errors can be tracked back to obvious conditions where segmentation algorithms have trouble, leading to either over- or under-segmentation. In the majority of sites surveyed, validation showed close agreement between percentiles of the manual KMS control measurements and the larger sample size automatic AIF measurements (Figures 6 and S3–S5).

*PebbleCounts* requires large sample areas since it misses many grains in favor of only measuring high-confidence segmented grains. Our study took place over a 100-km flow distance on a channel bed that is, frequently >100 m wide, with maximum *b*-axis grain sizes of ~50 cm. This allowed us to survey large areas (>1,000 m<sup>2</sup>) without concern for overlap with other grain-size populations at different survey sites several km away. In shorter and/or narrower channel reaches, survey areas must necessarily be smaller so as not to overlap and blur the signal between sites. Also, in short and narrow headwaters, the pebble size and upper-percentiles of the grain-size distribution can be in excess of several m. Thus, the user is restricted to smaller areas (shorter and narrower channel reaches), while also requiring more grains covered, since not all will be counted by *PebbleCounts* and many counts are needed to form a robust percentile measurement at the coarser grain sizes. As one approaches the headwaters where streams narrow and grain sizes increase, the use of the *PebbleCounts* method may be precluded, and other methods (e.g., Wolman-style counts) may be required.

Another limit of *PebbleCounts* is the lower truncation on grain sizing of 20 pixels in *b*-axis size (Purinton & Bookhagen, 2019a). Yet, a lower truncation must always be made. Manual counting on photos or in the field may allow very small grains down to ~0.8 cm to be counted (Graham et al., 2010; Rice & Church, 1998). On the other hand, digital grain sizing is limited by the imagery resolution (~1 mm/pixel in the present study, cf. Table 1). If all sites are truncated to the same lower limit (here  $\geq 2.5$  cm), then we create a new distribution of only the coarse bedload fraction. From this new distribution, inter-site comparisons can be made to assess the relative downstream changes in a subset of the coarse bedload. To reiterate: we avoid the *D*-notation (e.g., D50), since this implies that much finer material has been measured, which is not the case here.

Measuring the coarsest grains in a river reach has garnered great interest in recent years given the suggested importance of these outliers in controlling channel roughness, critical stresses, and long-term reach-scale morphology (e.g., MacKenzie et al., 2018; Williams et al., 2019). In this work, we lay out a fundamental new method to explore the coarsest fraction of the bedload distribution in large, high-mountain rivers. Higher resolution imagery could be acquired to measure finer grain sizes with *PebbleCounts*, and a possible synergy could be achieved between manual sieving of the finer material (which requires smaller sampling areas given its more homogeneous mixing) and photo surveying of the coarse fraction over a larger survey area. However, this is beyond the scope of the current study.

## 5.2. Assessment of Additional Errors and Uncertainties

In addition to the proposed limits and caveats of the outlined methods, two other considerations should be made: The inherent measurement error in photo-sieving using *PebbleCounts*, and the possible uncertainties from sampling location.

### 5.2.1. Measurement Errors in *PebbleCounts*

Digital grain sizing has inherent differences to traditional manual counting. Since grains are not physically handled and only viewed top-down, tilting, different axes exposed, difficult lighting, partial burial,



and/or imbrication can all cause discrepancies in manual versus digital results, potentially propagating into percentile uncertainties. These effects have been recognized and examined extensively since the first photo-sieving attempts (e.g., Ibbeken & Schleyer, 1986; Kellerhals & Bray, 1971).

Graham et al. (2010) provide detailed analysis of this and find that manual digital grain sizing (i.e., hand-clicking) tends to slightly over-estimate the percentiles of interest. This is caused by a combination of: (a) manual sampling over-counting finer grains compared to digital methods under-counting these finer grains; and (b) selective structural effects of imbrication, burial, tilting, etc. on the finer grain sizes, thus skewing the digital results to coarser percentiles. Despite this, Graham et al. (2010) found that semiautomatic digital grain sizing via segmentation (similar to *PebbleCounts*) had very low biases and high precision, similar to manual field measurement. Given the high quality of results, many researchers continue to use photo-sieving to gather grain-size distributions (e.g., Attal et al., 2015; D'Arcy et al., 2017; Harries et al., 2018; Neely & DiBiase, 2020).

In Purinton and Bookhagen (2019a), we used over 1,000 hand-clicked and manual KMS results to compare the automatic AIF tool in test patches. We found that at most 70% of the grains identified with the automatic AIF tool had matching grains in the validation data set, and typically nearer to 50%. This does not, however, imply that 50% of the AIF measurements are false, since the matching method required that the centroids of the automatic AIF and validation grain be within five pixels (6 mm) of one another. Manual KMS image segmentation or hand-clicked vector differences versus the automatic AIF tool can lead to larger offsets than five pixels, but still capture the same grain and same approximate axis dimensions. Both *PebbleCounts* tools have some error when measuring each grain via an ellipse fit to the segmented area; however, we found that these errors are generally low on length measurements (~1–5 pixels or ~1–5 mm given ~1 mm/pixel resolution; cf. Figure 9 in Purinton & Bookhagen, 2019a).

As with similar digital grain sizing methods (e.g., Graham, Reid, & Rice, 2005; Graham, Rice, & Reid, 2005; Graham et al., 2010), despite missing measurements and some small errors on individual grains, percentiles can be accurately retrieved from *PebbleCounts* (Purinton & Bookhagen, 2019a). Importantly, binomial percentile uncertainty estimation assumes that each individual grain is measured without error; thus, including measurement errors would require an extension of the method or alternative approaches.

To summarize: inaccuracies (measurement errors) exist in the form of slight (~1–5 pixel) length errors and in misidentified grains. The first measurement error is minimal and likely comparable or better than manual grain-sizing errors in the field. The second measurement error (misidentification) is only an issue for the automatically counted grain-size distribution, which is why the automatic AIF results must be treated with caution and validated using manual KMS control data. In any case, inclusion of mm-scale grain-by-grain measurement errors present in *PebbleCounts* would not affect the observed downstream patterns, which vary over several cm in the 99th percentile.

### 5.2.2. Uncertainties Related to Survey Location

Sampling gravel-bed mountain rivers to study downstream trends is complicated by the local heterogeneity of pebbles. Some studies at similar scales ( $\geq 100$  km) take denser sampling to smooth the downstream variability that occurs from this local heterogeneity (e.g., Attal & Lavé, 2006; Gomez et al., 2001), as well as from the smaller sample sizes taken via traditional grain-size measurement. These surveys rely on operator expertise in selecting areas representative of the long-term, baseline bedload transport, and not stochastic flood or landslide deposits. We followed these strategies in the field and carefully selected the surveys to be far from hillslopes and near the active channel, excluding any direct sampling of landslide or scree deposits, or excessively large boulder bars associated with flood deposits. An example survey and *PebbleCounts* sampling strategy can be found in the Figure S2. Rather than sampling a discrete patch or a single line across the channel, our approach covers a much wider area, and is thus more representative of the channel characteristics at a given site. The grain-size trends that we show (e.g., Figure 10) likely contain additional variability between the survey sites (locally discrete fining and coarsening trends), and this is why we connect the surveys with dashed lines. Despite this, our carefully chosen sites and large survey areas, as well as our local topographic and lithologic analyses, connect the observed grain-size changes with clear drivers upstream of each site, and we restrict our interpretations to the local site characteristics.

### 5.3. Upper Percentile Variability and Landscape Drivers

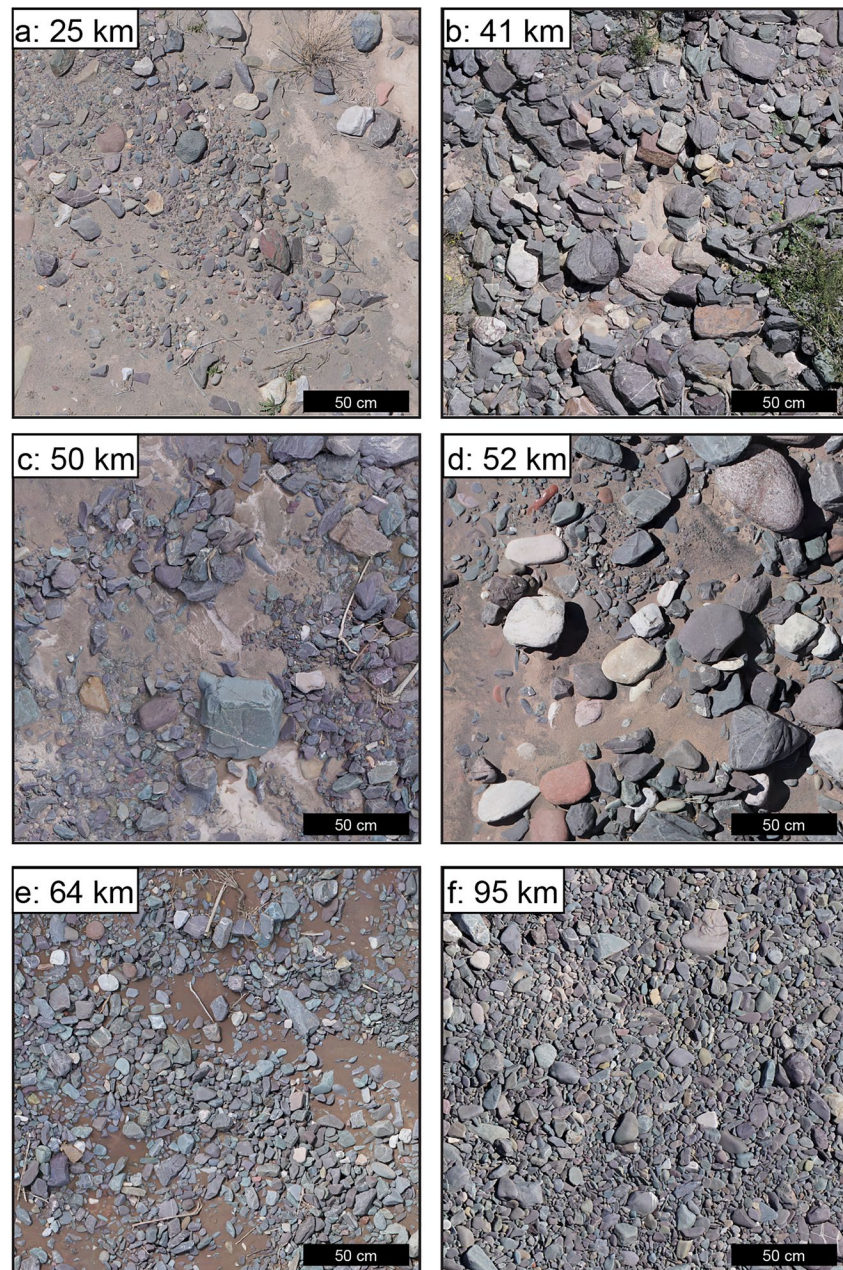
The 50th and 84th percentiles are commonly used to characterize grain size (e.g., Attal & Lavé, 2006; Attal et al., 2015), assess hydraulic conditions (e.g., Lague, 2014; Pfeiffer et al., 2017), and determine tectono-climatic conditions during transport and deposition (e.g., Armitage et al., 2011; Duller et al., 2010). Higher percentiles like the 90th and 95th are less commonly used in studying grain-size patterns in mountain rivers (e.g., Litty et al., 2017; Parida et al., 2019; Rice & Church, 1998), likely because the number of grains describing these percentiles is very low. We again note that the referenced studies all measure a finer grain size (based on a physical cutoff in suspension vs. bedload) to acquire the true D50 and D84. Here, we are only measuring the coarsest pebbles, and thus refer explicitly to percentiles (not *D*-notation) in this distribution of the subset of bedload. Our data show that in at least some mountain settings in highly erosive catchments like the Toro Basin, the uppermost percentiles (e.g., 95th and 99th) of the coarse bedload (here  $\geq 2.5$  cm) are more sensitive to downstream changes in material delivered to and moved by the trunk stream.

Despite our limited number of sample sites, the general trend of coarsening from a plateau into a steep bedrock gorge and then fining further downstream is in agreement with other observations in similar settings (e.g., Attal & Lavé, 2006; Attal et al., 2015; Whittaker et al., 2010). Here, we note the local landscape controls that modify these general trends, especially with regard to the connectivity of steep tributaries and steep, adjacent hillslopes that deliver diverse lithologies to the channel bed. In detailed studies, Mueller and Pitlick (2013) and Mueller et al. (2016) also found close ties between bedload sediment and catchment lithology and drainage area. Our results agree with these controls, and we also point out the importance of oversteepened tributary segments and heavily fractured hillslope lithology (Neely et al., 2019; Neely & DiBiase, 2020).

From Figure 11, we contend that large and steep tributaries delivering clasts of variable lithology to the trunk stream drive large-scale patterns in 99th percentile variability, and the influence of steeper hillslopes in the fractured Puncoviscanca meta-sediments adjacent to the trunk stream provides additional, local modification. The dependence of the bedload grain size on the supply of Cambrian quartzite and Cretaceous sandstone lithologies (making up  $< \sim 30\%$  of bedrock geology; Figure 3) from these large tributaries is clear in Figure 11d. Where steep and large tributaries do not continue this supply, local Puncoviscanca supply dominates the bedload. The general patterns in site-wise Puncoviscanca proportion (percentage of Puncoviscanca at each site in Figure 11d) follow the patterns in nearby upstream basin lithology, supporting this argument. We note that the photo counting for the percentage Puncoviscanca at each survey site (cf. Section 3.5.3) is an approximation, and more detailed data require new methods, possibly utilizing multispectral UAV-borne instruments. The results in Figure 11 can be reinforced by inspection of representative tiles from the surveys.

In Figure 12, we observe that the Puncoviscanca meta-sediments dominate the bedload (purple color). At 25, 41, and 52 km, there are also a number of lighter red, white, and yellow colored pebbles sourced from the Miocene/Oligocene conglomerates, Cambrian quartzites, and Cretaceous sandstone units, which outcrop in steep tributaries just upstream of these sites (Figure 11d). On the other hand, at 50, 64, and 95 km, we see more uniform Puncoviscanca lithology (the photo-counted lithology is  $\geq 97\%$  Puncoviscanca at these sites; cf. Figure 11d). In particular, we note that at 50 km the clasts are sub-angular to sub-rounded, indicating short transport distances in the fluvial network and the impact of steep nearby hillslopes delivering this material, whereas at for example, 52 km the clasts are rounded to well-rounded with few angular corners, indicating transport from the steep Río Capilla tributary. Angularity, like lithology, is not captured by *PebbleCounts*, and this would require further research using dense point clouds (Purinton & Bookhagen, 2019a).

We observe additional anecdotal evidence of fluvial transport of larger boulders from an upstream source as opposed to a nearby hillslope in Figure 2a. This photo shows the location of the 41-km survey site in the Puncoviscanca-dominated bedrock gorge. The boulders are significantly rounded, and a few are of the Cambrian quartzite lithology, which only outcrops in the upstream tributary at  $\sim 35$  km (Figure 3). Thus, it is likely that these boulders originate from this  $\sim 6$ -km upstream tributary (relative to the 41-km survey site) with  $k_{gr}$  values in excess of  $400 \text{ m}^{0.8}$  (Figure 11b).

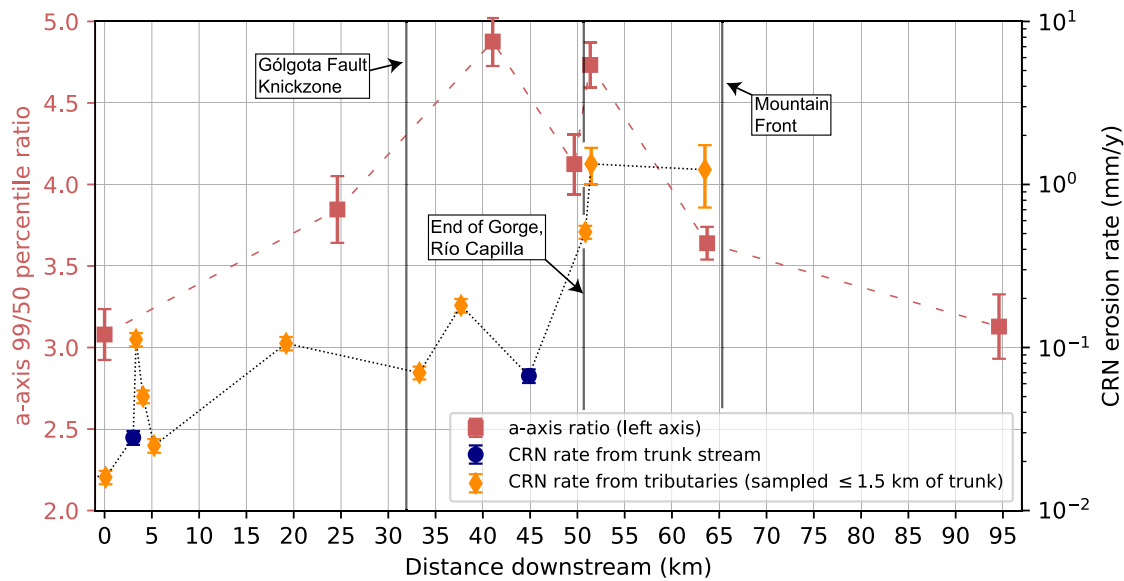


**Figure 12.** Representative 2,000×2,000 pixel ortho-image tiles (about 2×2 m) from each survey site, excluding the upstream-most survey at 0 km. Purple pebbles with quartz veins that make up most of the material are the Puncoviscana meta-sediments. Lighter colored pebbles (white, yellow, red) include the Miocene/Oligocene conglomerates, Cambrian quartzites, and Cretaceous sandstone units.

Differences in connectivity, lithology, and channel steepness as drivers of downstream grain-size changes are supported by others (Attal & Lavé, 2006; Attal et al., 2015; Harries et al., 2021; Mueller & Pitlick, 2005), and these relationships are complex, leading to local coarsening and fining trends. For instance, as opposed to this study, Attal et al. (2015) found that the 50th percentile (D50 in their study) in the Feather River basin increased faster with increasing stream power compared with the maximum grain size (D100). This result highlights the variability we may expect in channel-bed grain-size responses in different settings.

Notably, Hilley et al. (2019) proposed a  $k_{sn}$  value of  $200 \text{ m}^{0.8}$  (also using  $\theta_{ref} = 0.4$ ) as setting the global relief limit of Earth's topography. In Figure 11b, we note that the observed zones of percentile stretching appear





**Figure 13.** Downstream grain-size percentile stretching, shown with the 99th/50th percentile  $a$ -axis line on the left axis, compared with trunk (blue) and tributary (orange; sampled within 1.5 km of the trunk) cosmogenic radionuclide (CRN) erosion rates (right axis). Note the logarithmic CRN scaling and two orders of magnitude range in values. Erosion rates are from Tofelde et al. (2018), with some recalculated values from Bookhagen and Strecker (2012). The CRN values are connected with a dotted line only to visually aid the downstream order of the values where samples are closely located around 5 and 50 km.

to roughly coincide with large tributaries with  $k_{sn}$  values approaching or above this suggested relief limit. This lends credence to the notion that the uppermost percentiles of the coarse grain-size fraction are more sensitive and heavier influenced by the upper percentiles of channel steepness. That said, the influence of steep hillslopes and bedrock structure (fracturing) on channel-bed grain size—as is shown by recent work (Neely et al., 2019; Neely & DiBiase, 2020; Sklar et al., 2017, 2020)—certainly sets some local patterns in the Toro Basin. However, we note that these studies in bedrock channels assumed that there was no selective transport, and abrasion alone created the patterns of fining. In the Río Toro, many tributaries show bedrock-exposed channels, however the large quantity of material delivered to the trunk stream leads to an alluvial channel where a mix of abrasion and selective sorting likely operates. Furthermore, these studies have all been carried out in much smaller catchments ( $\leq 10 \text{ km}^2$ ), whereas here we are examining patterns over a 100-km flow distance in a  $4,500 \text{ km}^2$  catchment.

Previous important studies and concepts have pointed out the balances and feedbacks between grain size, channel steepness and other controlling factors (e.g., Lane, 1955). In this study, we support this, but also demonstrate the changing spatial scales of the involved parameters. We suggest that changes in grain sizes, especially at the upper percentiles, may respond to impacts several kilometers upstream and may thus dilute the conceptual relations in sediment transport. Furthermore, while large grains can impact local channel gradients, the large-scale gradient of the Río Toro is controlled by tectonic uplift and Quaternary faulting (e.g., García et al., 2013; Hilley & Strecker, 2005), and, at shorter time scales, by Pleistocene pluvial periods and associated landslide dams (e.g., Trauth et al., 2003). We do not anticipate that grain-size variations significantly impact channel steepness at these large scales in our study area.

In Figure 13, we compare the grain-size stretching with cosmogenic radionuclide erosion rates from the study area (Bookhagen & Strecker, 2012; Tofelde et al., 2018), which are commonly associated with topographic steepness (e.g., Granger et al., 1996; Ouimet et al., 2009). We limited comparisons to only 10 tributary samples within 1.5 km of the trunk stream and two trunk-stream samples. The general trend we observe is that the two orders of magnitude increase in erosion rate downstream (from  $\sim 0.02 \text{ mm/y}$  to  $\sim 1 \text{ mm/y}$ ) is met with an  $\sim 3$  to  $\sim 5$  times 99th/50th percentile grain-size stretching difference. Significant differences do exist in the trends, and important caveats include the fact that cosmogenic radionuclide samples are taken from sands, whereas the grain-size measurements are in the  $\geq 2.5 \text{ cm}$  pebbles, and the cosmogenic and *PebbleCounts* sites are neither spatially nor temporally coincident. Nevertheless, overall



trend similarities indicate the possible importance of more erosive parts of the catchment in influencing upper percentile grain-size variability.

In conjunction with these conditional erosion-rate data, the spatial coincidence of the increases in grain-size percentile stretching with the steepest ( $\geq 95$ th percentile) parts of the catchment (Figure 11) suggest a possible nonlinear relationship between geomorphic metrics and grain size. The steepest areas likely exert disproportionate control on increases in the coarsest bedload fraction, leading to distribution stretching. This signal is heavily influenced by lithology, which cannot be ignored. In any case, average values that do not weight steeper topography may not explain grain-size patterns in similar high-mountain settings. As we move downstream of junctions with large, steep tributaries in the Río Toro, the largest grains are both abraded and selectively sorted leading to compression of the grain-size distribution. This pattern is then overprinted and further modified by the delivery of finer, fractured material of tectonically ground bedrock on steep adjacent hillslopes.

## 6. Conclusions

Efficient methods of gathering large sample sizes—both in terms of counts and areal coverage—and making accurate measurements to form robust grain-size distributions are vital road blocks to understanding sediment transport and erosion processes in steep, complex, and remote high-mountain settings. Using field-based photo surveying and the application of the recently published pebble counting algorithm *PebbleCounts*, we have demonstrated that seven channel cross-sections with  $\sim 3,000$ – $20,000$  pebble measurements in the size fraction  $\geq 2.5$  cm can be used to draw catchment-scale conclusions about spatial patterns of sediment transport.

In a first result, we have shown that  $> \sim 1,000$  manual grain-size measurements follow an analytical log-normal distribution, which allows better statistical analysis of grain-size percentiles via larger simulated sample sizes. Larger sample sizes can also be achieved by careful validation of automatic empirical measurements, without the need for a log-normal model. Nevertheless, the manual, automatic, and log-normal results all demonstrate similar downstream patterns in grain-size variability.

Second, we showed the most abundant grain-size changes are recorded at the higher percentiles ( $\geq 95$ th). The 50th and 84th  $a$ -axis grain-size percentiles vary by  $\sim 20\%$  and  $\sim 50\%$ , respectively, over a 100-km flow distance, whereas the 95th and 99th percentiles vary by  $\sim 70\%$  and  $\sim 100\%$ , respectively. A large sample size, like those achieved with *PebbleCounts*, is required to reduce percentile uncertainties and meaningfully assess higher percentiles in dynamic mountain rivers.

Finally, we found that the observed grain-size changes can be related with upper percentiles of topographic steepness, in conjunction with consideration of upstream lithology. The correlation between upper percentiles suggests a highly nonlinear relationship between steep parts of the landscape and channel grain-size changes. While sediment-transport processes are complex and more work is necessary, we argue that uppermost grain-size percentiles are most sensitive to the oversteepened upstream catchment, convolved with a strong lithologic signal.

## Data Availability Statement

The *PebbleCounts* tools are available at <https://github.com/UP-RS-ESP/PebbleCounts> (Purinton & Bookhagen, 2019b). Some scripts associated with the analysis are available at: <https://github.com/UP-RS-ESP/PebbleCounts-Application>, and are also included in the online data (grain size,  $k_{sn}$ , and hillslope angle) repository: <https://zenodo.org/record/5089789> (Purinton, 2021).

## References

- Agisoft (2018). *Agisoft photoscan professional*. Retrieved from <http://www.agisoft.com/downloads/installer/>
- Allen, P. A., Michael, N. A., D'Arcy, M., Roda-Boluda, D. C., Whittaker, A. C., Duller, R. A., & Armitage, J. J. (2017). Fractionation of grain size in terrestrial sediment routing systems. *Basin Research*, 29(2), 180–202. <https://doi.org/10.1111/bre.12172>
- Armitage, J. J., Duller, R. A., Whittaker, A. C., & Allen, P. A. (2011). Transformation of tectonic and climatic signals from source to sedimentary archive. *Nature Geoscience*, 4(4), 231–235. <https://doi.org/10.1038/ngeo1087>

### Acknowledgments

TanDEM-X DEMs were received through grant DEM\_CALVAL1028 for B. Purinton through the DLR. Additional funding was sourced from DFG BO 2933/3-1 (B. Bookhagen), DFG funded IRTG-StRATEGy (IGK2018) to M. R. Strecker and B. Bookhagen, and BMBF LIDAR and NEXUS funded through the MWFK Brandenburg, Germany, both for B. Bookhagen. Taylor Smith, Stephanie Olen, Stefanie Tofelde, and Aljoscha Rheinwalt are thanked for conversations and suggestions throughout the analysis and manuscript preparation. Reviews and comments from Amy East, Mikael Attal, Rebekah Harries, and three anonymous reviewers improved the manuscript. Open access funding enabled and organized by Projekt DEAL.

- Attal, M., & Lavé, J. (2006). Changes of bedload characteristics along the Marsyandi River (central Nepal): Implications for understanding hillslope sediment supply, sediment load evolution along fluvial networks, and denudation in active orogenic belts. *Geological Society of America Special Paper*, 398, 143–171. [https://doi.org/10.1130/2006.2398\(09\)](https://doi.org/10.1130/2006.2398(09))
- Attal, M., Mudd, S., Hurst, M., Weinman, B., Yoo, K., & Naylor, M. (2015). 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 Surface Dynamics*, 3(1), 201–222. <https://doi.org/10.5194/esurf-3-201-2015>
- Bookhagen, B., & Strecker, M. R. (2012). Spatiotemporal trends in erosion rates across a pronounced rainfall gradient: Examples from the southern central Andes. *Earth and Planetary Science Letters*, 327–328, 97–110. <https://doi.org/10.1016/j.epsl.2012.02.005>
- Bunte, K., & Abt, S. T. (2001). Sampling surface and subsurface particle-size distributions in wadable gravel- and cobble-bed streams for analyses in sediment transport, hydraulics, and streambed monitoring. In *Sampling surface and subsurface particle-size distributions in wadable gravel- and cobble-bed streams for analyses in sediment transport, hydraulics and streambed monitoring* (Vol. RMRS-GTR-74; Tech. Rep.). Rocky Mountain Research Station US Forest Service. <https://doi.org/10.2737/RMRS-GTR-74>
- Butler, J. B., Lane, S. N., & Chandler, J. H. (2001). Automated extraction of grain-size data from gravel surfaces using digital image processing. *Journal of Hydraulic Research*, 39(5), 519–529. <https://doi.org/10.1080/00221686.2001.9628276>
- Carbonneau, P., Bizzi, S., & Marchetti, G. (2018). Robotic photosieving from low-cost multirotor suavs: A proof-of-concept. *Earth Surface Processes and Landforms*, 43(5), 1160–1166. <https://doi.org/10.1002/esp.4298>
- Castino, F., Bookhagen, B., & de la Torre, A. (2020). Atmospheric dynamics of extreme discharge events from 1979 to 2016 in the southern central Andes. *Climate Dynamics*, 55(11), 3485–3505. <https://doi.org/10.1007/s00382-020-05458-1>
- Castino, F., Bookhagen, B., & Strecker, M. (2016). River-discharge dynamics in the southern central Andes and the 1976–77 global climate shift. *Geophysical Research Letters*, 43(22), 11679–11687. <https://doi.org/10.1002/2016GL070868>
- Castino, F., Bookhagen, B., & Strecker, M. R. (2017). Oscillations and trends of river discharge in the southern central Andes and linkages with climate variability. *Journal of Hydrology*, 555, 108–124. <https://doi.org/10.1016/j.jhydrol.2017.10.001>
- Chardon, V., Schmitt, L., Piégay, H., & Lague, D. (2020). Use of terrestrial photosieving and airborne topographic lidar to assess bed grain size in large rivers: A study on the Rhine River. *Earth Surface Processes and Landforms*, 45(10), 2314–2330. <https://doi.org/10.1002/esp.4882>
- Chatanantavet, P., Lajeunesse, E., Parker, G., Malverti, L., & Meunier, P. (2010). Physically based model of downstream fining in bedrock streams with lateral input. *Water Resources Research*, 46(2), W02518. <https://doi.org/10.1029/2008WR007208>
- D'Arcy, M., Whittaker, A. C., & Roda-Boluda, D. C. (2017). Measuring alluvial fan sensitivity to past climate changes using a self-similarity approach to grain-size fining, death valley, California. *Sedimentology*, 64(2), 388–424. <https://doi.org/10.1111/sed.12308>
- Detert, M., Kadinski, L., & Weitbrecht, V. (2018). On the way to airborne gravelometry based on 3D spatial data derived from images. *International Journal of Sediment Research*, 33(1), 84–92. <https://doi.org/10.1016/j.ijrsr.2018.02.001>
- Detert, M., & Weitbrecht, V. (2012). Automatic object detection to analyze the geometry of gravel grains—a free stand-alone tool. In *River Flow 2012: Proceedings of the International Conference on Fluvial Hydraulics, San José, Costa Rica September 5–7, 2012*, (pp. 595–600). Taylor & Francis Group.
- Dingle, E. H., Attal, M., & Sinclair, H. D. (2017). Abrasion-set limits on himalayan gravel flux. *Nature*, 544(7651), 471–474. <https://doi.org/10.1038/nature22039>
- Duller, R. A., Whittaker, A. C., Fedele, J. J., Whitchurch, A. L., Springett, J., Smithells, R., & Allen, P. A. (2010). From grain size to tectonics. *Journal of Geophysical Research*, 115(F3), F03022. <https://doi.org/10.1029/2009jf001495>
- Eaton, B., & Moore, R. (2019). *bceaton/GSDtools: GSDtools: Analyse and compare River bed surface grain size distributions*. Zenodo. Retrieved from <https://doi.org/10.5281/zenodo.3229387>
- Eaton, B., Moore, R. D., & MacKenzie, L. G. (2019). Percentile-based grain size distribution analysis tools (gsdtools) – Estimating confidence limits and hypothesis tests for comparing two samples. *Earth Surface Dynamics*, 7(3), 789–806. <https://doi.org/10.5194/esurf-7-789-2019>. Retrieved from <https://esurf.copernicus.org/articles/7/789/2019/>
- Ferguson, R., Hoey, T., Wathen, S., & Werritty, A. (1996). Field evidence for rapid downstream fining of river gravels through selective transport. *Geology*, 24(2), 179–182. [https://doi.org/10.1130/0091-7613\(1996\)024<0179:FEFRDF>2.3.CO;2](https://doi.org/10.1130/0091-7613(1996)024<0179:FEFRDF>2.3.CO;2)
- García, V. H., Hongn, F., & Cristallini, E. O. (2013). Late Miocene to recent morphotectonic evolution and potential seismic hazard of the northern Lerma valley: Clues from lomas de medeiros, Cordillera Oriental, NW Argentina. *Tectonophysics*, 608, 1238–1253. <https://doi.org/10.1016/j.tecto.2013.06.021>
- Gasparini, N. M., Tucker, G. E., & Bras, R. L. (1999). Downstream fining through selective particle sorting in an equilibrium drainage network. *Geology*, 27(12), 1079–1082. [https://doi.org/10.1130/0091-7613\(1999\)027<1079:dftsps>2.3.co;2](https://doi.org/10.1130/0091-7613(1999)027<1079:dftsps>2.3.co;2)
- Gomez, B., Rosser, B. J., Peacock, D. H., Hicks, D. M., & Palmer, J. A. (2001). Downstream fining in a rapidly aggrading gravel bed river. *Water Resources Research*, 37(6), 1813–1823. <https://doi.org/10.1029/2001WR900007>
- Graham, D. J., Reid, I., & Rice, S. P. (2005). Automated sizing of coarse-grained sediments: Image-processing procedures. *Mathematical Geology*, 37(1), 1–28. <https://doi.org/10.1007/s11004-005-8745-x>
- Graham, D. J., Rice, S. P., & Reid, I. (2005). A transferable method for the automated grain sizing of river gravels. *Water Resources Research*, 41(7), W07020. <https://doi.org/10.1029/2004WR003868>
- Graham, D. J., Rollet, A.-J., Piégay, H., & Rice, S. P. (2010). Maximizing the accuracy of image-based surface sediment sampling techniques. *Water Resources Research*, 46(2), W02508. <https://doi.org/10.1029/2008WR006940>
- Granger, D. E., Kirchner, J. W., & Finkel, R. (1996). Spatially averaged long-term erosion rates measured from in situ-produced cosmogenic nuclides in alluvial sediment. *The Journal of Geology*, 104(3), 249–257. <https://doi.org/10.1086/629823>
- Hain, M. P., Strecker, M. R., Bookhagen, B., Alonso, R. N., Pingel, H., & Schmitt, A. K. (2011). Neogene to quaternary broken foreland formation and sedimentation dynamics in the Andes of NW Argentina (25s). *Tectonics*, 30(2), TC2006. <https://doi.org/10.1029/2010tc002703>
- Harries, R. M., Gailleton, B., Kirstein, L. A., Attal, M., Whittaker, A. C., & Mudd, S. M. (2021). Impact of climate on landscape form, sediment transfer and the sedimentary record. *Earth Surface Processes and Landforms*, 46(5), 990–1006. <https://doi.org/10.1002/esp.5075>
- Harries, R. M., Kirstein, L. A., Whittaker, A. C., Attal, M., Peralta, S., & Brooke, S. (2018). Evidence for self-similar bedload transport on Andean alluvial fans, iglesia basin, south central Argentina. *Journal of Geophysical Research: Earth Surface*, 123(9), 2292–2315. <https://doi.org/10.1029/2017jf004501>
- Hermanns, R. L., Trauth, M. H., Niedermann, S., McWilliams, M., & Strecker, M. R. (2000). Tephrochronologic constraints on temporal distribution of large landslides in northwest Argentina. *The Journal of Geology*, 108(1), 35–52. <https://doi.org/10.1086/31438310.1086/314383>
- Hilley, G. E., Porder, S., Aron, F., Baden, C. W., Johnstone, S. A., Liu, F., et al. (2019). Earth's topographic relief potentially limited by an upper bound on channel steepness. *Nature Geoscience*, 12(10), 828–832. <https://doi.org/10.1038/s41561-019-0442-3>

- Hilley, G. E., & Strecker, M. R. (2005). Processes of oscillatory basin filling and excavation in a tectonically active orogen: Quebrada del toro basin, NW Argentina. *The Geological Society of America Bulletin*, 117(7–8), 887–901. <https://doi.org/10.1130/B25602.1>
- Hoey, T. B., & Bluck, B. J. (1999). Identifying the controls over downstream fining of river gravels. *Journal of Sedimentary Research*, 69(1), 40–50. <https://doi.org/10.2110/jsr.69.4010.2110/jsr.69.40>
- Ibbeken, H., & Schleyer, R. (1986). Photo-sieving: A method for grain-size analysis of coarse-grained, unconsolidated bedding surfaces. *Earth Surface Processes and Landforms*, 11(1), 59–77. <https://doi.org/10.1002/esp.3290110108>
- Kellerhals, R., & Bray, D. I. (1971). Sampling procedures for coarse fluvial sediments. *Journal of the Hydraulics Division*, 97(8), 1165–1180. <https://doi.org/10.1061/jycej.0003044>
- Knighton, A. (1982). Longitudinal changes in the size and shape of stream bed material: Evidence of variable transport conditions. *Catena*, 9(1–2), 25–34. [https://doi.org/10.1016/s0341-8162\(82\)80003-9](https://doi.org/10.1016/s0341-8162(82)80003-9)
- Lague, D. (2014). The stream power river incision model: Evidence, theory and beyond. *Earth Surface Processes and Landforms*, 39(1), 38–61. <https://doi.org/10.1002/esp.3462>
- Lane, E. W. (1955). Importance of fluvial morphology in hydraulic engineering. *Proceedings (American Society of Civil Engineers)*. v. 81. paper no. 745.
- Lang, N., Irniger, A., Rozniak, A., Hunziker, R., Wegner, J. D., & Schindler, K. (2020). Grainet: Mapping grain size distributions in River beds from uav images with convolutional neural networks. *Hydrology and Earth System Sciences Discussions*, 2020, 1–38. <https://doi.org/10.5194/hess-2020-196>
- Litty, C., Schlunegger, F., & Viveen, W. (2017). Possible threshold controls on sediment grain properties of Peruvian coastal river basins. *Earth Surface Dynamics*, 5(3), 571–583. <https://doi.org/10.5194/esurf-5-571-2017>
- MacKenzie, L. G., Eaton, B. C., & Church, M. (2018). Breaking from the average: Why large grains matter in gravel-bed streams. *Earth Surface Processes and Landforms*, 43(15), 3190–3196. <https://doi.org/10.1002/esp.4465>
- Marrett, R., Allmendinger, R., Alonso, R., & Drake, R. (1994). Late Cenozoic tectonic evolution of the puna plateau and adjacent foreland, northwestern Argentine Andes. *Journal of South American Earth Sciences*, 7(2), 179–207. [https://doi.org/10.1016/0895-9811\(94\)90007-8](https://doi.org/10.1016/0895-9811(94)90007-8)
- Marrett, R., & Strecker, M. R. (2000). Response of intracontinental deformation in the central Andes to late cenozoic reorganization of south american plate motions. *Tectonics*, 19(3), 452–467. <https://doi.org/10.1029/1999tc001102>
- Miller, K. L., Szabó, T., Jerolmack, D. J., & Domokos, G. (2014). Quantifying the significance of abrasion and selective transport for downstream fluvial grain size evolution. *Journal of Geophysical Research: Earth Surface*, 119(11), 2412–2429. <https://doi.org/10.1002/2014jf003156>
- Mudd, S. M., Attal, M., Milodowski, D. T., Grieve, S. W. D., & Valters, D. A. (2014). A statistical framework to quantify spatial variation in channel gradients using the integral method of channel profile analysis. *Journal of Geophysical Research: Earth Surface*, 119(2), 138–152. <https://doi.org/10.1002/2013jf002981>
- Mudd, S. M., Clubb, F. J., Gailleton, B., & Hurst, M. D. (2018). How concave are river channels? *Earth Surface Dynamics*, 6(2), 505–523. <https://doi.org/10.5194/esurf-6-505-2018>
- Mudd, S. M., Clubb, F. J., Grieve, S. W. D., Milodowski, D. T., Hurst, M. D., Gailleton, B., & Valters, D. A. (2019). *Lsdtopotools2*. Zenodo. <https://doi.org/10.5281/zenodo.3245041>
- Mueller, E. R., & Pitlick, J. (2005). Morphologically based model of bed load transport capacity in a headwater stream. *Journal of Geophysical Research*, 110(F2), F02016. <https://doi.org/10.1029/2003jf000117>
- Mueller, E. R., & Pitlick, J. (2013). Sediment supply and channel morphology in mountain river systems: 1. relative importance of lithology, topography, and climate. *Journal of Geophysical Research: Earth Surface*, 118(4), 2325–2342. <https://doi.org/10.1002/2013jf002843>
- Mueller, E. R., Smith, M. E., & Pitlick, J. (2016). Lithology-controlled evolution of stream bed sediment and basin-scale sediment yields in adjacent mountain watersheds, Idaho, USA. *Earth Surface Processes and Landforms*, 41(13), 1869–1883. <https://doi.org/10.1002/esp.3955>
- Neely, A. B., & DiBiase, R. A. (2020). Drainage area, bedrock fracture spacing, and weathering controls on landscape-scale patterns in surface sediment grain size. *Journal of Geophysical Research: Earth Surface*, 125(10), e2020JF005560. <https://doi.org/10.1029/2020JF005560>
- Neely, A. B., DiBiase, R. A., Corbett, L. B., Bierman, P. R., & Caffee, M. W. (2019). Bedrock fracture density controls on hillslope erodibility in steep, rocky landscapes with patchy soil cover, southern California, USA. *Earth and Planetary Science Letters*, 522, 186–197. <https://doi.org/10.1016/j.epsl.2019.06.011>
- Olen, S., & Bookhagen, B. (2020). Applications of SAR interferometric coherence time series: Spatiotemporal dynamics of geomorphic transitions in the south-central Andes. *Journal of Geophysical Research: Earth Surface*, 125(3), e2019JF005141. <https://doi.org/10.1029/2019JF005141>
- Omarini, R., Sureda, R., Götz, H., Seilacher, A., & Pflüger, F. (1999). Puncoviscana folded belt in northwestern Argentina: Testimony of late Proterozoic rodinia fragmentation and pre-gondwana collisional episodes. *International Journal of Earth Sciences*, 88, 76–97. <https://doi.org/10.1007/s005310050247>
- Ouimet, W. B., Whipple, K. X., & Granger, D. E. (2009). Beyond threshold hillslopes: Channel adjustment to base-level fall in tectonically active mountain ranges. *Geology*, 37(7), 579–582. <https://doi.org/10.1130/g30013a.1>
- Paola, C., Parker, G., Seal, R., Sinha, S. K., Southard, J. B., & Wilcock, P. R. (1992). Downstream fining by selective deposition in a laboratory flume. *Science*, 258(5089), 1757–1760. <https://doi.org/10.1126/science.258.5089.1757>
- Parida, S., Singh, V., & Tandon, S. (2019). Sediment connectivity and evolution of gravel size composition in Dehra dun? An intermontane valley in the frontal zone of NW Himalaya. *Zeitschrift für Geomorphologie*, 62(2), 83–105. <https://doi.org/10.1127/zfg/2019/0568>
- Perron, J. T., & Royden, L. (2013). An integral approach to bedrock river profile analysis. *Earth Surface Processes and Landforms*, 38(6), 570–576. <https://doi.org/10.1002/esp.3302>
- Pfeiffer, A. M., Finnegan, N. J., & Willenbring, J. K. (2017). Sediment supply controls equilibrium channel geometry in Gravel Rivers. *Proceedings of the National Academy of Sciences*, 114(13), 3346–3351. <https://doi.org/10.1073/pnas.1612907114>
- Purinton, B. (2021). *Data and scripts for journal of geophysical research: Earth surface publication: Tracking downstream variability in large grain-size distributions in the South-central Andes*. Zenodo. <https://doi.org/10.5281/zenodo.5089789>
- Purinton, B., & Bookhagen, B. (2017). Validation of digital elevation models (dems) and comparison of geomorphic metrics on the southern central Andean plateau. *Earth Surface Dynamics*, 5(2), 211–237. <https://doi.org/10.5194/esurf-5-211-2017>
- Purinton, B., & Bookhagen, B. (2018). Measuring decadal vertical land-level changes from srtn-c (2000) and tandem-x in the south-central Andes. *Earth Surface Dynamics*, 6(4), 971–987. <https://doi.org/10.5194/esurf-6-971-2018>
- Purinton, B., & Bookhagen, B. (2019). Introducing PebbleCounts: A grain-sizing tool for photo surveys of dynamic gravel-bed rivers. *Earth Surface Dynamics*, 7(3), 859–877. <https://doi.org/10.5194/esurf-7-859-2019>

- Purinton, B., & Bookhagen, B. (2019). *Pebblecounts: A python grain-sizing algorithm for gravel-bed river imagery*. GFZ Data Services. <https://doi.org/10.5880/figeo.2019.007>
- Purinton, B., & Bookhagen, B. (2020). Multiband (x, c, l) radar amplitude analysis for a mixed sand- and gravel-bed river in the eastern central andes. *Remote Sensing of Environment*, 246, 111799. <https://doi.org/10.1016/j.rse.2020.111799>
- Reyes, F., & Salfity, J. (1973). Stratigraphy considerations about the cretaceous (pirgua subgroup) in northwestern Argentina. *Quinto Congreso Geológico Argentino*, 3
- Rice, S., & Church, M. (1996). Sampling surficial fluvial gravels; the precision of size distribution percentile sediments. *Journal of Sedimentary Research*, 66(3), 654–665. <https://doi.org/10.2110/jsr.66.654>
- Rice, S., & Church, M. (1998). Grain size along two gravel-bed rivers: Statistical variation, spatial pattern and sedimentary links. *Earth Surface Processes and Landforms*, 23(4), 345–363. [https://doi.org/10.1002/\(SICI\)1096-9837\(199804\)23:4<345::AID-ESP850>3.0.CO;2-B](https://doi.org/10.1002/(SICI)1096-9837(199804)23:4<345::AID-ESP850>3.0.CO;2-B)
- Rizzoli, P., Martone, M., Gonzalez, C., Wecklich, C., Tridon, D. B., Brättingam, B., et al. (2017). Generation and performance assessment of the global tandem-x digital elevation model. *ISPRS Journal of Photogrammetry and Remote Sensing*, 132, 119–139. <https://doi.org/10.1016/j.isprsjprs.2017.08.008>
- Roda-Boluda, D. C., D'Arcy, M., McDonald, J., & Whittaker, A. C. (2018). Lithological controls on hillslope sediment supply: Insights from landslide activity and grain size distributions. *Earth Surface Processes and Landforms*, 43(5), 956–977. <https://doi.org/10.1002/esp.4281>
- Schwab, K., & Schäfer, A. (1976). Sedimentation und tektonik im mittleren abschnitt des rio toro in der Ostkordillere NW-Argentinien. *Geologische Rundschau*, 65, 175–194. <https://doi.org/10.1007/BF01808462>
- Sime, L., & Ferguson, R. (2003). Information on grain sizes in gravel-bed rivers by automated image analysis. *Journal of Sedimentary Research*, 73(4), 630–636. <https://doi.org/10.1306/112102730630>
- Sklar, L. S., Dietrich, W. E., Fofoula-Georgiou, E., Lashermes, B., & Bellugi, D. (2006). Do gravel bed river size distributions record channel network structure? *Water Resources Research*, 42(6), W06D18. <https://doi.org/10.1029/2006WR005035>
- Sklar, L. S., Riebe, C. S., Genetti, J., Leclere, S., & Lukens, C. E. (2020). Downvalley fining of hillslope sediment in an alpine catchment: Implications for downstream fining of sediment flux in Mountain rivers. *Earth Surface Processes and Landforms*, 45(8), 1828–1845. <https://doi.org/10.1002/esp.4849>
- Sklar, L. S., Riebe, C. S., Marshall, J. A., Genetti, J., Leclere, S., Lukens, C. L., & Mercus, V. (2017). The problem of predicting the size distribution of sediment supplied by hillslopes to rivers. *Geomorphology*, 277, 31–49. <https://doi.org/10.1016/j.geomorph.2016.05.005>
- Sternberg, H. (1875). *Untersuchungen über längen-und querprofil geschiefeführende flüsse*. Zeitschrift für Bauwesen, 25, 483–506.
- Strecker, M., Alonso, R., Bookhagen, B., Carrapa, B., Hilley, G., Sobel, E., & Trauth, M. (2007). Tectonics and climate of the southern central Andes. *Annual Review of Earth and Planetary Sciences*, 35(1), 747–787. <https://doi.org/10.1146/annurev.earth.35.031306.140158>
- Tofelde, S., Duesing, W., Schildgen, T. F., Wickert, A. D., Wittmann, H., Alonso, R. N., & Strecker, M. (2018). Effects of deep-seated versus shallow hillslope processes on cosmogenic 10be concentrations in fluvial sand and gravel. *Earth Surface Processes and Landforms*, 43(15), 3086–3098. <https://doi.org/10.1002/esp.4471>
- Tofelde, S., Schildgen, T. F., Savi, S., Pingel, H., Wickert, A. D., Bookhagen, B., et al. (2017). 100 kyr fluvial cut-and-fill terrace cycles since the middle pleistocene in the southern central Andes, NW Argentina. *Earth and Planetary Science Letters*, 473, 141–153. <https://doi.org/10.1016/j.epsl.2017.06.001>
- Trauth, M. H., Bookhagen, B., Marwan, N., & Strecker, M. R. (2003). Multiple landslide clusters record quaternary climate changes in the northwestern Argentine Andes. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 194(1), 109–121. [https://doi.org/10.1016/S0031-0182\(03\)00273-6](https://doi.org/10.1016/S0031-0182(03)00273-6)
- Whipple, K. X. (2004). Bedrock Rivers and the geomorphology of active orogens. *Annual Review of Earth and Planetary Sciences*, 32, 151–185. <https://doi.org/10.1146/annurev.earth.32.101802.120356>
- Whittaker, A. C., Attal, M., & Allen, P. A. (2010). Characterising the origin, nature and fate of sediment exported from catchments perturbed by active tectonics. *Basin Research*, 22(6), 809–828. <https://doi.org/10.1111/j.1365-2117.2009.00447.x>
- Williams, R. D., Reid, H. E., & Brierley, G. J. (2019). Stuck at the bar: Larger-than-average grain lag deposits and the spectrum of particle mobility. *Journal of Geophysical Research: Earth Surface*, 124(12), 2751–2756. <https://doi.org/10.1029/2019jfe005137>
- Wobus, C., Whipple, K. X., Kirby, E., Snyder, N., Johnson, J., Spyropolou, K., & Sheehan, D. (2006). Tectonics from topography: Procedures, promise, and pitfalls. *Special Papers - Geological Society of America*, 398, 55. [https://doi.org/10.1130/2006.2398\(04\)](https://doi.org/10.1130/2006.2398(04))
- Wohl, E. E., Anthony, D. J., Madsen, S. W., & Thompson, D. M. (1996). A comparison of surface sampling methods for coarse fluvial sediments. *Water Resources Research*, 32(10), 3219–3226. <https://doi.org/10.1029/96WR01527>
- Wolcott, J., & Church, M. (1991). Strategies for sampling spatially heterogeneous phenomena; the example of river gravels. *Journal of Sedimentary Research*, 61(4), 534–543. <https://doi.org/10.1306/D4267753-2B26-11D7-8648000102C1865D>
- Wolman, M. G. (1954). A method of sampling coarse river-bed material. *Eos, Transactions American Geophysical Union*, 35(6), 951–956. <https://doi.org/10.1029/TR035i006p00951>
- Woodget, A. S., Fyffe, C., & Carbonneau, P. E. (2018). From manned to unmanned aircraft: Adapting airborne particle size mapping methodologies to the characteristics of sUAS and SfM. *Earth Surface Processes and Landforms*, 43(4), 857–870. <https://doi.org/10.1002/esp.4285>