SUPPLEMENTARY METHODS

Experimental Setup

At the Saint Anthony Falls Laboratory (SAFL) in Minneapolis, Minnesota, we ran experiments of a river crossing a sandbox with dimensions of 3.9 m x 2.4 m x 0.4 m. The basin was lined with thick plastic to prevent leaking and filled with silica sand with a unimodal grain size distribution that has a mean grain size of 0.144 mm (Tofelde et al., 2019). A Gaussian curve fit to the sieved-sand data provides a mean and standard deviation of 0.140±0.04 mm; see Table S2 and Fig. S3. A pool controlled base level at the river outlet using a computer-controlled weir. Sediment and dyed water entered the basin at the river inlet through a gravel diffuser. This produced a braided river that flowed across the long axis of the basin (Fig. 1) towards the outlet, where a gap in the sandbox wall allowed water and sediment to exit into the basin (Savi et al., 2019; Tofelde et al., 2019).

We conducted six experiments (Table S1), each with a constant water discharge of 0.1 L/s and sediment discharge of 0.0022 L/s (including pore space). We tested base-level fall rates of 0 mm/hr (the control), 25 mm/hr, 50 mm/hr, 200 mm/hr, 300 mm/hr, and 400 mm/hr. Before each experiment, we cut a 10±2 cm wide channel to a depth that increased monotonically from 3±0.5 cm at the inlet to 10±1 cm at the outlet. We then activated the water and sediment feed and ran each experiment for four hours with a constant base level at 25 cm above the exterior basin floor. This “spin-up” phase allowed the river valley to widen and the slope to equilibrate (Mackin, 1948; Lane, 1955; Blom et al., 2016). For the control experiment, we retained a constant base level and continued to run the experiment for an additional 7.67 hours. For all other experiments, we began to lower base level at a constant rate from 25 cm to 10 cm following this initial four-hour “spin-up”. Depending on the rate of base-level fall, this required between 22.5 minutes and 6 hours. We continued each experiment until after the the river reached a new equilibrium slope.

During each experiment, we obtained laser topographic scans every 15–30 minutes and overhead photos every 20 seconds. The laser scans, used to construct 1-mm-horizontal-resolution DEMs, were taken every 30 minutes during the four-hour spin-up phase of each experiment and every 15 minutes thereafter. We orthorectified the overhead photos and georeferenced them to the DEMs using the basin boundaries and irregularities in the sand surface (e.g., boot prints) as control points with QGIS (QGIS Development Team, 2018) (to define the tie points) and command-line GDAL tools (Warmerdam, 2008) (to batch process the images) (Wickert, 2020). The final orthorectified and georeferenced images have 0.89-mm resolution. We then linked these photos and the topographic scans into a continuous timeline of seconds since the start of the experiment (Witte and Wickert, 2020), given in a seven-digit zero-padded number that is part of each filename (Beaulieu et al., 2020).

Mapping Landslides

We mapped landslides from each experiment by hand on sets of georeferenced overhead images. (Following Larsen et al. (2010), we use the term *Landslide* broadly to refer to mass failures.) Every time one or more landslides occurred, we produced a new time-stamped GIS vector file containing 2D polygons of each new landslide at the time corresponding to photo in which it was first observed (Figs. 1c, S1). We then determined landslide areas using Gauss' area formula. Landslide volumes were calculated by subtracting from each pre-landslide DEM cell's elevation the corresponding lowest point in the same "North–South" line, which crosses the river valley. The valley floor marks the presumed base of the landslides as they were triggered by fluvial undercutting, though by using the channel thalweg (Fig. 1e), we may slightly overestimate (by up to a ~1 cm flow depth) the actual thickness of the slides. This error is typically small compared to the height of the valley walls (~5-40 cm). We calculated landslide widths and lengths by fitting an ellipse to each landslide (using a least-squares approach) and then finding the distance across each ellipse along the *y* axis of Fig. 1, as this is perpendicular to the trend of the river valley. Through this process, we also calculated the major and minor axes of each ellipse and their orientations, as well as the midpoint of each landslide (Wickert, 2020).

Volumetric water content

Moisture content dictates both cohesive strength (due to matric water tension) and the weight of the failing block (as air within pores is replaced by water). Cohesion is maximized at intermediate values of water content and ranges from ~250 Pa to ~750 Pa for the Esperance sand (median grain diameter D_{50}=250 µm) (Lu et al., 2009), which is very close in texture to the sand in our experiments (D_{50}=144 µm: Fig. S2 and Table S2) though it is slightly coarser and less well sorted. This range is consistent with the results of experiments by Richefeu et al. (2006). Density of the combined sand–water mass is given by

$$\rho = \rho_s(1 - \lambda_p) + \rho_w(\lambda_p c_w), \tag{S1}$$

where ρ is the overall density, ρ_s is the density of the sand grains, ρ_w is the density of water, λ_p is the porosity, and c_w is the concentration (or fraction) of water in the pores, ranging from 0 (none) to 1 (full saturation). We neglect air density because the entire system is submerged in air at uniform pressure. Assuming a typical porosity of 35%, the density of the combined sand–water mass that fails can range from 1720 to 2070 kg m^{-3}. This 20% difference is much less than the significant variability in cohesion. Based on this analysis, we chose to pick a single representative value of c_w=0.5 for the density-affecting water content and compare the results against the full range of literature-derived cohesion values.

Block stability

We built and analyzed free-body diagrams (Fig. S3) of trapezoidal block slides, triangular block slides, and topples. We defined slides to be trapezoidal, driven by undercutting, unless they were too narrow for the given angle, α=76°. In these cases, we then redefined the failure as a triangular block slide and steepened the scarp accordingly.

We defined cohesive strength, σ_c, as follows for trapezoidal slides,

$$\sigma_c = \rho g \left(W - \frac{H}{2 \tan \alpha} \right) \sin \alpha \left(\sin \alpha - \mu \cos \alpha \right), \tag{S2}$$

and triangular slides,

$$\sigma_c = \frac{\rho g W^2 \left(\sin \alpha - \mu \cos \alpha \right)}{2 \sqrt{H^2 + W^2}}. \tag{S3}$$

Here, ρ is the bulk density of the landslide block, g is gravitational acceleration, α=76° is the mean measured landslide scarp angle (we use this for the dominant trapezoidal landslides; triangular landslides occur when the block is too tall and narrow to accommodate this angle), μ=0.6 is the coefficient of internal friction, W is landslide width (valley wall to landslide scarp), and H is landslide block height (thalweg to upper surface). The denominator provides the length of the failure plane. The bulk density reflects both sand grains and pore space, some of which is filled with water. Based on this criterion, 4% of failures occurred as triangular facets.

Similarly, we applied a torque balance to define the tensile strength, σ_t, involved in topples as

$$\sigma_t = \frac{\rho g W^2}{H}. \tag{S4}$$

A linear Mohr–Coulomb failure analysis, which incorporates the simplifying assumption that our system includes sufficient overburden to inhibit the grain-on-grain interlocking forces that give dry sand positive cohesion at low confining pressures (Kim, 2001; Lu et al., 2009), relates tensile strength to cohesion (which is the resistance to shear) as:

$$\sigma_t = \frac{\sigma_c}{\mu}. \tag{S5}$$

Based on this analysis, we plot the predicted tensile strength vs. cohesion of all failures along with a line defined by this Mohr–Coulomb criterion to separate the space into zones in which tensile (toppling) vs. shear (sliding) failures are more likely to occur (Fig. S4). We find that 11% of the failures are more likely to be topples. This is significantly greater than our measured 2% topple rate, and may result in either a failure to recognize small topples, or an overestimation of topples due to our use of a linear Mohr–Coulomb relationship.

We computed the predicted cohesive strength at failure for each of the identified mass-wasting events, and then sorted these into those that more likely occurred as topples or as slides (Fig. 4) in order to test the sensitivity of our cohesion estimates to ambiguity in failure mode. This ambiguity exists because we mapped the failure mode for only the subset of the total mass-wasting events that we could observe in the topographic scans. Based on the dominance of slides (98%), we assumed that all of the mass-wasting events occur as slides when calculating cohesion at failure. By treating topples as slides along the mean slide-type failure angle (76°), we find that they have mean and median predicted cohesive strengths that are 35% and 30% greater than the mean and median, respectively, for those expected to occur as slides. Therefore, these will tend to bias the inversion for cohesion towards higher values. Topples tended to be more dominant among the larger failures, with higher predicted cohesive strengths at failure for an equivalent slide event (Fig. S5), and therefore may have more limited impact on the position of the area–frequency and predicted cohesion–frequency peaks, while impacting to some extent the large-magnitude tail of the overall area–frequency distribution.

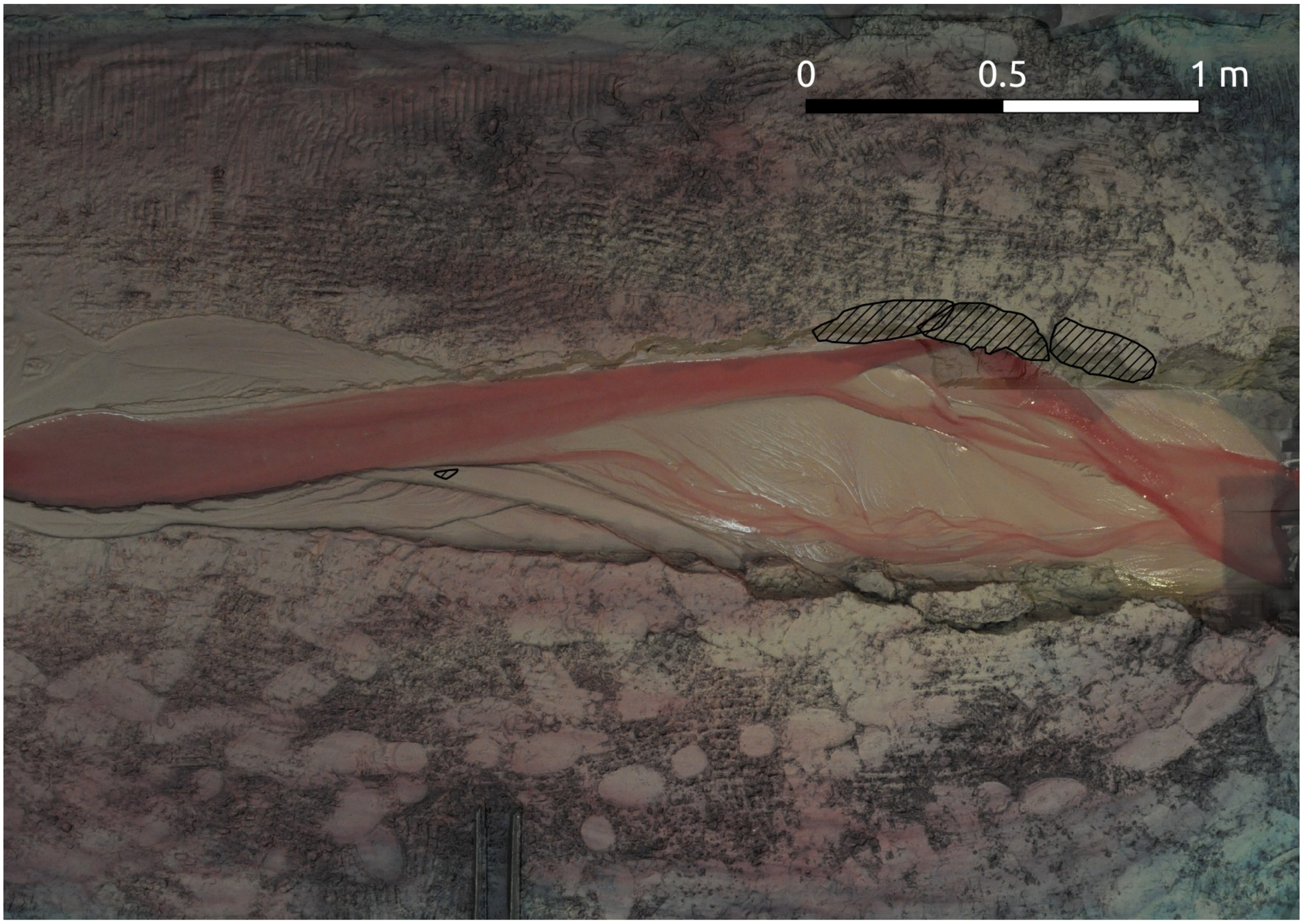

Figure S1. Experiment overview and observational approach. Landslides were identified on georeferenced overhead photos of the experiment. ImgSec_0043168 from the 25 mm/hr base-level fall experiment is shown here with semi-transparency to expose a shaded-relief map from the DEM generated immediately after this photo was taken (DEM_fullextent_0043188) beneath. Hatched regions indicate where landslides occurred after 43,168 seconds of experiment runtime; three lie on the bluff tops to the “northeast”, and a smaller one is on a terrace to the “southwest”. A linear shadow running left to right to the “south” of the cluster of three future landslides indicates via the shaded-relief map where the plywood basin walls along the sides of the outlet prevented the angled laser-topography scanner from casting light on the the valley bottom. This does not significantly affect the landslidevolume estimates because the laser scanner was able to record topography in the center of the channel and use this to calculate landslide volumes. On the “south” side of the valley from the three larger future landslides lie detached blocks from recent landsliding failures. This image is also stored at the Data Repository for the University of Minnesota (DRUM)[31] and at the Sediment Experimentalists Network (SEN) Knowledge Base.

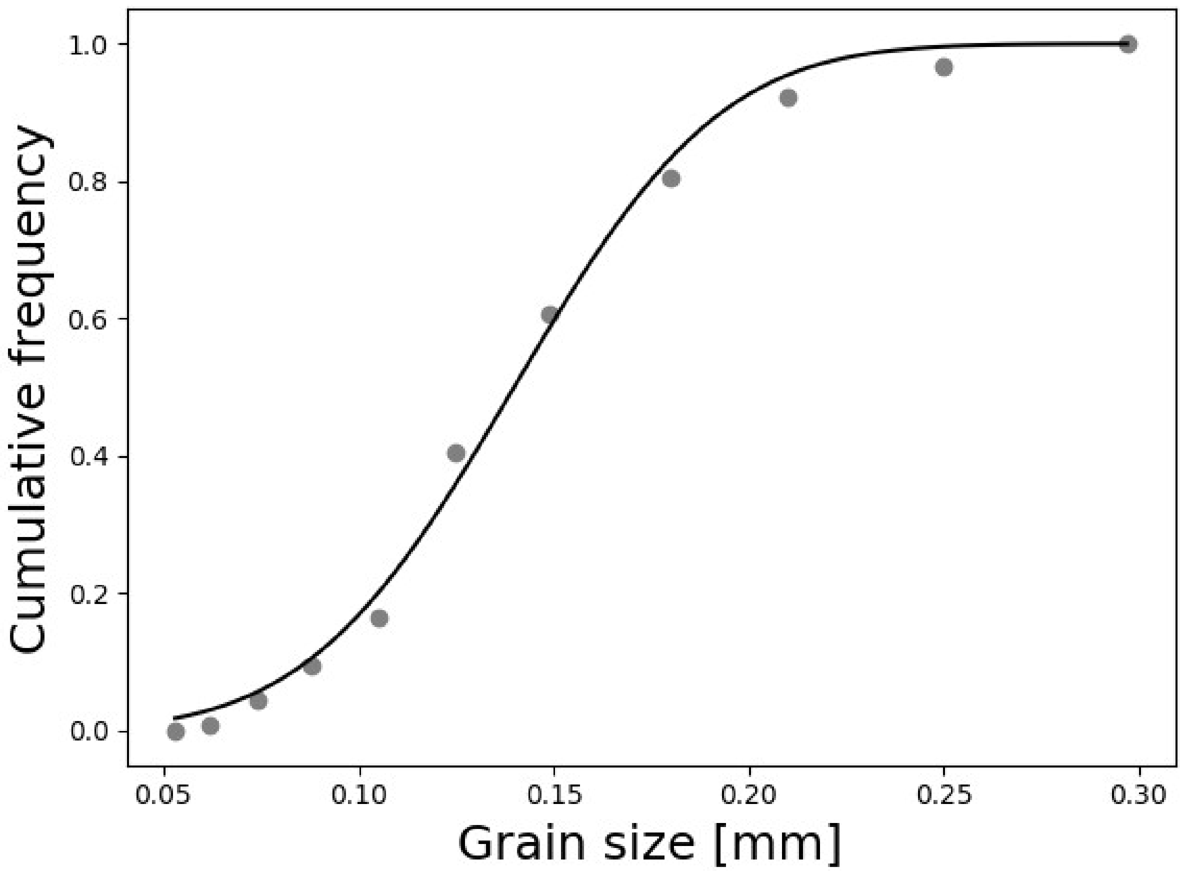

Figure S2. Grain-size distribution of sand used in the experiment[49]; Table 2 contains the raw sieve data, corresponding to the points on the plot. The black line is a Gaussian cumulative probability distribution, indicating that the sand-grain sizes are approximately normally distributed. Median grain size, D_{50}, is 144 μm. The standard deviation of grain sizes is 40 μm.

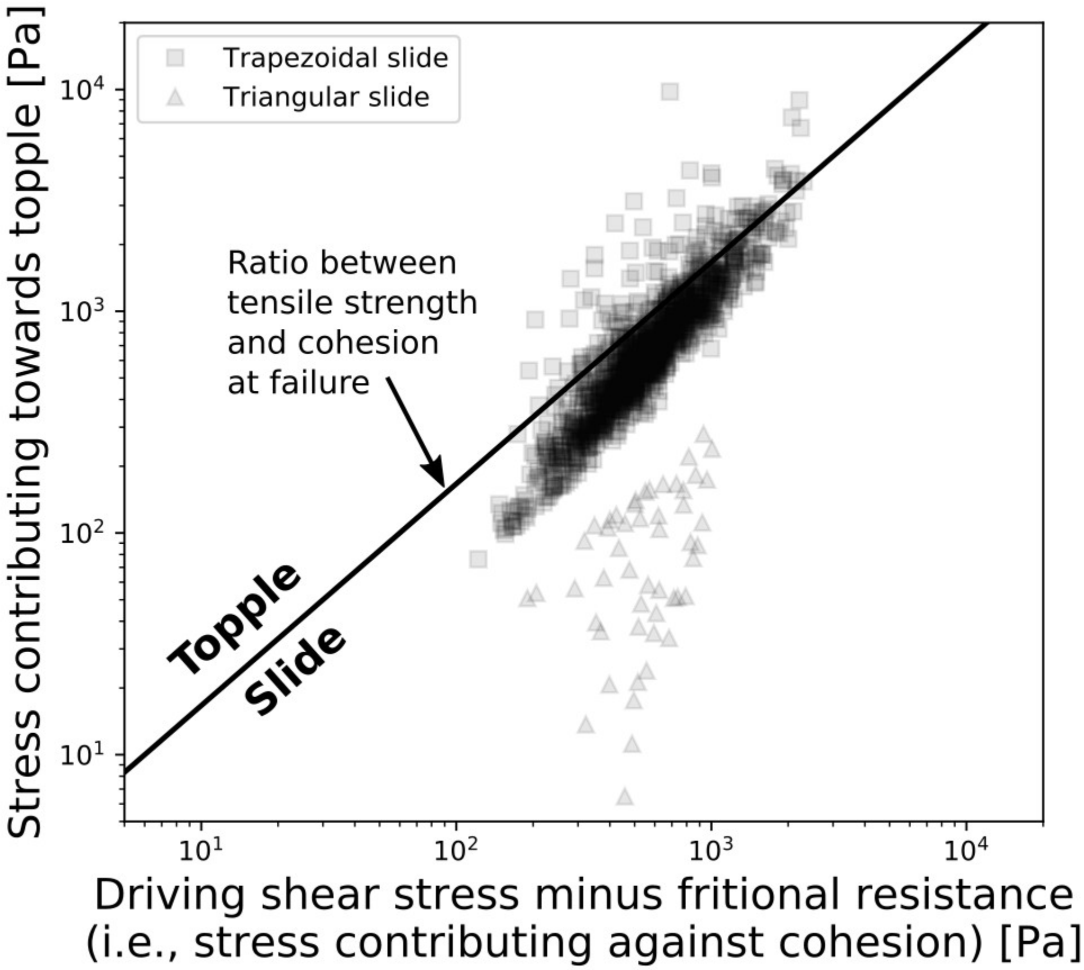

Figure S3. Calculated stress at failure assuming rotational (topple, *y* axis) and translational (cohesion:
145 frictional effects subtracted out, *x* axis) mass-wasting events (Fig. S3). Increasing stress at failure indicates an increasing chance that the failure would have overcome the internal tensile strength and/or cohesion, thereby making it more likely to occur as a topple (tensile failure) or slide (shear failure). The line separating the tensile and slide domains is given assuming a linear Mohr–Coulomb relationship with an internal friction coefficient of μ=0.6.

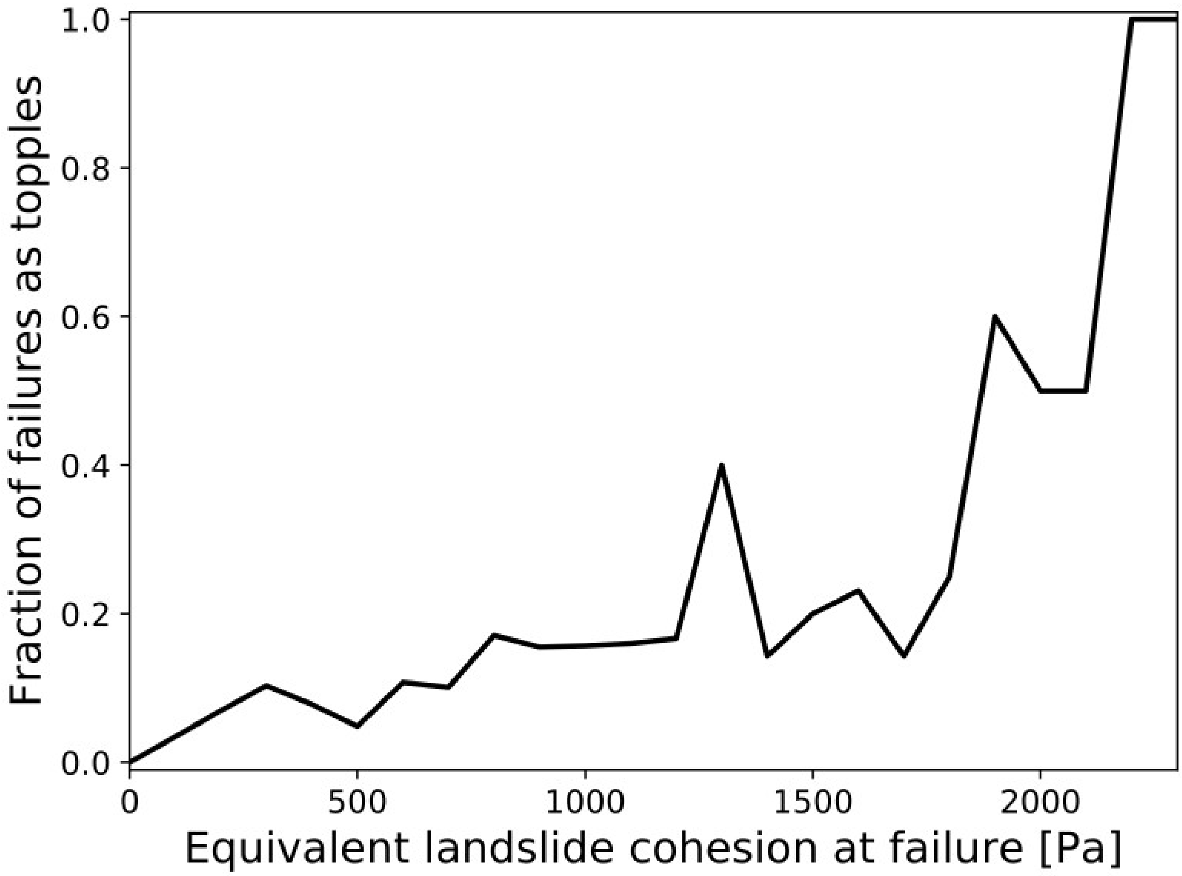

Figure S4. Frequency of predicted topples based on the analysis in Fig. S4. These values (11% predicted topples) are higher than what we observed from mass-wasting events within 3 minutes prior to each topographic scan (2%). The trend towards larger failures being more dominantly via a toppling mechanism is consistent with our observations of the experiment, indicating that some of the tail of the area–frequency distribution may be influenced by this toppling mechanism.

Base-level fall [mm/hr]	Runtime [s]	Number of landslides
0	42024	197
25	46788	233
50	36013	235
200	43091	276
300	36004	176
400	40484	242
All	**244404**	**1359**

Table S1: Experiments. We ran each experiment for four hours at steady base level, and then changed base level (or held it constant) as indicated until the experiment reached a new steady state. All data from these experiments are provided in the data repository (Beaulieu et al., 2020).

`

Sieve mesh size [mm]	Percentage of grains smaller than sieve mesh
0.297	100
0.250	96.8
0.210	92.2
0.180	80.6
0.149	60.8
0.125	40.5
0.105	16.6
0.088	9.6
0.074	4.3
0.062	0.8
0.053	0.0

Table S2: Grain Size. Distribution of sand and coarse silt grains that formed the substrate for the experiments. These data are plotted in Figure S2 and are available via the GitHub and Zenodo (Wickert, 2020).

SUPPLEMENT: REFERENCES CITED

Beaulieu, O.P., Wickert, A.D., Witte, L.D., and Tofelde, S., 2020, Experimental alluvial-river and landsliding response to base-level fall: Data Repository for the University of Minnesota, University of Minnesota Libraries Digital Conservatory, https://doi.org/10.13020/zw6r-am46.

Blom, A., Viparelli, E., and Chavarrías, V., 2016, The graded alluvial river: Profile concavity and downstream fining: Geophysical Research Letters, v. 43, p. 6285–6293, doi:10.1002/2016GL068898.

Kim, T.-H., 2001, Moisture-induced tensile strength and cohesion in sand [Ph.D. Thesis]: University of Colorado Boulder.

Lane, E.W., 1955, importance of fluvial morphology in hydraulic engineering: Proceedings (American Society of Civil Engineers) ; v. 81, paper no. 745, http://agris.fao.org/agris-search/search.do?recordID=US201400000288 (accessed July 2018).

Larsen, I.J., Montgomery, D.R., and Korup, O., 2010, Landslide erosion controlled by hillslope material: Nature Geoscience, v. 3, p. 247–251, doi:10.1038/ngeo776.

Lu, N., Kim, T.-H., Sture, S., and Likos, W.J., 2009, Tensile Strength of Unsaturated Sand: Journal of Engineering Mechanics, v. 135, p. 1410–1419, doi:10.1061/(ASCE)EM.1943-7889.0000054.

Mackin, J.H., 1948, Concept of the Graded River: Geological Society of America Bulletin, v. 59, p. 463, doi:10.1130/0016-7606(1948)59[463:COTGR]2.0.CO;2.

QGIS Development Team, 2018, QGIS Geographic Information System: Open Source Geospatial Foundation Project, http://qgis.osgeo.org.

Richefeu, V., El Youssoufi, M.S., and Radjaï, F., 2006, Shear strength properties of wet granular materials: Physical Review E, v. 73, doi:10.1103/PhysRevE.73.051304.

Savi, S., Tofelde, S., Wickert, A.D., Bufe, A., Schildgen, T.F., and Strecker, M.R., 2019, Interactions between channels and tributary alluvial fans: channel adjustments and sediment-signal propagation: Earth Surface Dynamics Discussions, doi:10.5194/esurf-2019-73.

Tofelde, S., Savi, S., Wickert, A.D., Bufe, A., and Schildgen, T.F., 2019, Alluvial channel response to environmental perturbations: fill-terrace formation and sediment-signal disruption: Earth Surface Dynamics, v. 7, p. 609–631, doi:10.5194/esurf-7-609-2019.

Warmerdam, F., 2008, The Geospatial Data Abstraction Library, *in* Hall, G.B. and Leahy, M.G. eds., Open Source Approaches in Spatial Data Handling, Berlin, Heidelberg, Springer Berlin Heidelberg, Advances in Geographic Information Science, v. 2, p. 87–104, doi:10.1007/978-3-540-74831-1_5.

Wickert, A.D., 2020, umn-earth-surface/landslide-experiment-analysis: v0.3.0: Revised submission:, http://doi.org/10.5281/zenodo.4026739.

Witte, L.D., and Wickert, A.D., 2020, umn-earth-surface/fluvial-experiment-time-series-helper: v0.1.1: Base-level fall experiment:, http://doi.org/10.5281/zenodo.3758678.