Supplement of Earth Surf. Dynam., 7, 829–839, 2019
https://doi.org/10.5194/esurf-7-829-2019-supplement

Earth Surface Dynamics
Open Access

Supplement of

Seismic and geologic controls on spatial clustering of landslides in three large earthquakes

Claire Rault et al.

Correspondence to: Claire Rault (claire.rault@ens.fr)

SUPPLEMENTARIES

Methods and metrics

We aim at producing statistically robust information on the position of the landslides with regards to their respective distance to rivers and crests in a given portion of the landscape. This cannot be achieved by constructing a simple distribution of distances of the landslides cells (or centroids) to rivers and crests without normalizing for the relative frequency of crest and river in the area of interest. Indeed variations in the relative proportion of crests and rivers within a given area will strongly bias the resulting distribution of distance of any objects (landslides or others) randomly distributed in this area. Such bias may appear because of variable crest (or river) frequency caused by either landscape natural shape or methodological effects. For example, Fig. S1a-c illustrates the reduction of the crest network density with changing criterion of crest tagging. A given set of landslides will appear to be much closer to crests in the landscape represented in Fig. S1.a than in the one represented in Fig. S1.c, simply because crests are less frequent in the latter case. Note that a similar bias would appear with a variable frequency of river channels.

Thus, to be able to discuss physical control on the statistical location of landslides relative to rivers and ridges across large regions, it is necessary to find an adequate normalization for biases emerging from the definition of crests and rivers and for comparing areas with variable landscape shape.

First, we define, for each cell of the landscape, its normalized distance to the river network $|d_{st}|$ expressed as (Meunier et al., 2008):

$$|d_{st}| = \frac{d_{st}}{d_{st}+d_{tp}}, \qquad (1)$$

where d_{st} and d_{tp} are its flow distances to the nearest river and the nearest crest respectively. $|d_{st}|$ ranges from 0 for cells located in the floodplains (i.e., with drainage area above a threshold defined following Montgomery, 2001) to 1 for cells located on the crests (with a zero drainage area). The thresholds of drainage area we use to define channel heads vary from 0.02 to 0.5 km^2 in this study. Crests are mapped using a double criterion of null flow accumulation and a threshold of positive curvature (Fig. S1b). By definition, cells with $|d_{st}|>0.75$ stand in the upper quarter of the hillslopes, whereas cells with $|d_{st}|<0.25$ are in the lower quarter. A given portion of the landscape is characterized by its probability density function of occurrence of |dst| values, PDF_{topo} . In our analysis, we only consider this distribution over the interval (0,1), i.e. we exclude all cells located on crests and in the floodplains, making PDF_{topo} independent of variations of floodplain or crest width. Figure S2.e shows examples of PDF_{topo} for 3 synthetic catchments with straight, concave and convex hillslopes respectively. Whatever the hillslope curvature, PDF_{topo} is a monotonic function with no asymptotic behavior toward zero and can therefore be used for normalization. Landslide locations along hillslope are characterized by PDF_{ls}, the probability density function of $|d_{st}|$ derived only from cells affected by landslides. Then within portions of the epicentral areas (macrocells) we compute both PDF_{topo} and PDF_{ls} and define the ratio of probability R_p as:

$$R_p = \frac{PDF_{ls}}{PDF_{topo}} \qquad (2)$$

In this way, the distribution of locations of the landsliding cells along hillslopes (here expressed in $|d_{st}|$) is normalized by the distribution of occurrence of locations in the landscape of the macrocell, effectively removing physical or methodological biases (Fig. S2.f). If the landscape into the macrocell is uniformly sampled by landsliding, $R_p=1$ over (0,1). High values of R_p *(>>1)* for $|d_{st}|>0.75$ indicate a significant crest oversampling by landslides. Inversely, low values of R_p express undersampling. Similarly, large values of R_p for $|d_{st}|<0.25$ indicate hillslope toe oversampling. In our analysis, we have defined $Rp_{crest} = Rp_{|dst|>0.75}$ and $Rp_{toe} = Rp_{|dst|<0.25}$ as the mean value of R_p over the upper and the lower quarter of the hillslope respectively. Figure S1.f shows three plots of R_p computed from a given distribution of landslides (visible in Fig. S1.e) in a macrocell of varying density of crests, only due to methodological choice in the definition of ridges. In contrast to non-normalized landslide locations distribution plots (Fig. S1.e), the R_p curves appear almost identical for the three ridge definitions, demonstrating that R_p is independent of the mapping methods of crests and rivers.

Statistical robustness

The use of R_p to remove potential physical (due to landscape shape) or mapping biases but may still suffer from statistical bias. For example a macrocell with only one landslide would allow to define R_p, but would intuitively be suspected of not being representative. More generally, we want to quantify the probability for a given topography and landslides within it that the observed R_p could differ from one because of statistical fluctuations rather than for physical reasons. In other words we want to quantify the null hypothesis that $R_p>1$ (or $R_p<1$) is due to random fluctuations around PDF_{topo} and hense insure that we retain only statistically robust cases (macrocells) of landslide clustering.

If we select a random subset of N cells within a macrocell, that represent less than 10% of the macrocell area, this draw can be considered as a sampling with replacement (SWR) and their probability to be at a given $|d_{st}|_i$ follows a binomial law $\mathcal{B}$(n=N, p=$PDF_{topo}(|d_{st}|=|d_{st}|_i)$). Under these conditions, the distribution PDF_{rd} of a random sampling of cells within a given macrocell should converge toward PDF_{topo} for a large number of samples, i.e $lim_{n\to N_i} PDF_{rd}(|d_{st}|_i) = PDF_{topo}(|d_{st}|_i)$, with N_i the number of cells equals to $|d_{st}|_i$ in the macrocell.

Then, the Central Limit Theorem (CLT) gives the 90% prediction intervals of $PDF_{rd}(n,|d_{st}|_i)$ as

$$I_p = \left[p - 1.64\sqrt{\frac{p(1-p)}{n}}; p + 1.64\sqrt{\frac{p(1-p)}{n}}\right], \qquad (3)$$

for a given value $PDF_{topo}(|d_{st}|_i)=p$ and n independent random samples. The convergence towards a normal distribution centered on p also requires $n>30$, $np>5$ and $n(1-p)>5$ (CLT conditions). By construction, the 90% prediction intervals on $R_{p(rd)}$ is defined as :

$$I_{Rp} = \frac{I_p}{p} = \left[1 - 1.64\sqrt{\frac{(1-p)}{np}}; 1 + 1.64\sqrt{\frac{(1-p)}{np}}\right], \qquad (4)$$

entailing $lim_{n\to N_i} I_{Rp}(|d_{st}|_i) = 1$. Figure S2.f shows the 90% interval on R_p in the 3 synthetic catchments mentioned above for 500 cells randomly drawn in the *DEM*. Note than as PDF_{topo} monotonically grows with $|d_{st}|$, the prediction interval is generally smaller in the right region of the plot, *i.e* near the crests. As a result, if R_p, computed from landslide cells in a macrocell, is contained within I_{Rp}, there is more than 10% chances that the difference between R_p and 1 is due to chance rather than any physical effects, and we will refrain from interpreting this macrocell. In contrast, regions where R_p is found to be beyond the interval I_{rp} have less than 10% chance to be due to chance and can therefore be interpreted as resulting from physical processes. Figure S1.f shows an example of R_p computed in a macrocell affected by 500 landsliding cells (red) and exhibiting crest oversampling. By contrast, Fig. S2.g shows an example of R_p plot computed from a draw of 100 failing cells. The peak observed at $|d_{st}|$=0.8 cannot be interpreted as it remains confined within the interval of fluctuation of random draws of 100 cells.

Note that as the probability ratio R_p is built from the ratio of the normalized distributions of area of given $|d_{st}|$, *n* should be the number of cells affected by landsliding in the macrocell. But this method introduces a bias: as a landslide is composed of several cells, for any cell *i* affected by landsliding of given $|d_{st}|_i$, its neighboring cell *j* has a higher probability of being at $|d_{st}|_j \approx |d_{st}|_i$. In this approach, the draws are not independent anymore and the sampling with replacement hypothesis is not met. We can bypass this problem by defining *n* as the number of landslides included in the macrocell. Because the number of landslides per macrocell is much lower than the number of cells composing them, *n* is usually rather small, resulting in larger intervals, and more conservative interpretation (as the criterion for statistical significance is stricter).

In the epicentral area, crest-clustering is defined as macrocells where $Rp_{crest} > IRp_{crest_{max}}$ while toe-clustering corresponds to macrocells where $Rp_{toe} > IRp_{toe_{max}}$. In contrast, macrocells where $IRp_{crest_{min}} < Rp_{crest} < IRp_{crest_{max}}$ or where *n* is insufficient for the convergence criteria of the Central Limit Theorem (i.e., n<30, np<5 or n(1-p)<5) are not represented in Fig. 2 and should not be interpreted. Maps containing all Rp_{crest} values, statistically robust or not, are represented on Fig. S3 for comparison.

Figure S4 shows that crest clustering is generally equivalent to toe undersampling and *vice versa*. As a consequence, regions of toe-clustering are represented on Fig. 2 by macrocells where $Rp_{crest} < IRp_{crest_{min}}$ and where the CLT criteria are valid.

The size of the macrocell is chosen in order to image the clustering with the best resolution as possible within the epicentral area. There is too many areas not defined if the macrocells are too small and the special resolution is too low if the macrocells are too big (Fig. S5). The average value of Rp_{crest} over the whole epicentral area is converging with the macrocell size toward 1.22.

Validity of the metrics

The method we introduce aims at defining the landslide position independently of the distribution of area with $|d_{st}|$ in the landscape. This condition is satisfied since Rp_{crest} is uncorrelated to both $Ptopo_{crest} = PDF_{topo}(|dst| > 0.75)$ and $Ptopo_{toe} = PDF_{topo}(|dst| < 0.25)$ (Fig. S6).

The patterns presented in Fig. 2 can be biased by the landslide mapping technique. The inventories we use do not distinguish between landslide deposits and scars. As landslides move downslope, by definition they preferentially affect the lower parts of hillslopes. To test the robustness of our results, we have run the same analysis with the data from Taiwan, using the landslide centroids and estimated landslide scars. To do this, we determined the length and the width of individual landslides, and used the finding of Domej et al. (2017) that earthquake-triggered landslides scars have a stable width to length ratio of *Ar=0.6*. The length of a landslide is equal to the difference between its maximum and minimum distance to river. The width is calculated using the landslide length and area, assuming a rectangular shape. Then, the lower part of the landslide polygon is progressively removed until *Ar=0.6*. The Northridge and Wenchuan inventories contain too many instances of landslide amalgamation to perform a systematic, accurate scar extraction. For the Taiwanese case, the values of Rp_{crest} obtained from centroids and scars are plotted against the values obtained using the whole landslides in Fig. S7. The results from these three methods have a nearly 1:1 correlation. Therefore, the regional pattern of Rp_{crest} seems to be preserved, irrespective of whether we consider whole landslides or landslide scars.

Extraction of topographic features

In order to extract geometric features of the ridges, we simplify them considering they have triangular sections (Fig. S8).

The slope height h_{ri}, is defined as:

$$h_{ri} = H_{ri} - min(H_j)\,, \qquad (5)$$

where H_{ri} is the elevation of the crest cell r_i and H_j is the elevation of a river cell j distant of dst_{ri} from r_i. The half-base width of a hill for the section S_i through r_i, L_{ri}, is calculated as:

$$L_{ri}^{2} = \sqrt{dst_{ri}^{2} - h_{ri}^{2}}\,, \qquad (6)$$

The shape ratio at a given ridge point S_{ri} is defined as the ratio of ridge relief h_{ri} to the half-width L_{ri}.

Dependence on the dataset

Three landslide databases are available for the Wenchuan earthquake. The catalog produced by Xu et al, 2014 has a higher number of landslides and covers a larger area (Fig. S10 and Table 1). Furthermore, we notice that the number of landslides tends to converge to a maximum value for landslide densities above 10^{-2} in the Parker *et al.* 2015, and Gorum *et al.* 2011 inventories, while it still increases in the one from Xu *et al.* 2014 (Fig. S10). This difference is likely caused by landslide amalgations in the Parker *et al.* 2015, and Gorum *et al.* 2011 datasets (Marc and Hovius, 2015).

Figure S11 shows the Rp_{crest} maps obtained using these three inventories. The one resulting from the *Xu et al*, 2014 covers a larger area than the maps obtained from the two others catalogs. As a consequence of these two observations, we choose to use for our analysis the catalog produced by Xu *et al*. 2014.

Topographic amplification

We notice that in certain areas, topographic ground-motion amplification might explain the landslides crest-clustering. The topographic amplification can be approach by the smoothed curvature or the relative elevation (Maufroy *et al*, 2014, Rai *et al*, 2017, 2016).

Both of these parameters are computed from a 30m digital elevation model (DEM) resolution. The curvature is the second derivative of the topographic elevation, it is calculated using ArcGIS software (ESRI, 2011). The smoothed curvature matrix (*Cs*) is calculated as (Maufroy *et al*, 2015):

$$CS = \frac{1}{n^4}(C.\begin{pmatrix} 1_{11} & \cdots & 1_{1n} \\ \vdots & \ddots & \vdots \\ 1_{n1} & \cdots & 1_{nn} \end{pmatrix}\begin{pmatrix} 1_{11} & \cdots & 1_{1n} \\ \vdots & \ddots & \vdots \\ 1_{n1} & \cdots & 1_{nn} \end{pmatrix})$$

Where the curvature matrix C is convolved by a n x n unit matrix. The characteristic smoothing length is defined as $LS = 2.n.h$ where h is the resolution of the DEM in meter. The best correlation between smoothed curvature and amplification has been found for *LS* equal to the half of the seismic wavelength.

The relative elevation (topographic position index) at a DEM cell i (H_{di}) is calculated as:

$H_{di} = h_i - M_{hd}$

where h_i is the elevation of the cell i and M_{hd} is the mean of the topographic elevation of the cells distant from $d/2$ of the cell i.

Both methods show that finer features are visible at smaller scales *Ls* or *d*. As *Cs* and H_d are positively correlated (Fig. S17.b and Rai *et al*, 2015) we choose here to represent only the smoothed curvature as an example.

The distribution of the topographic half width (Fig. S17.c) in the area surrounded in white (Fig. S17.a) gives us an idea wavelength ranges at which ridges in that area may resonate and amplify ground-motion. Following Paolucci 2002, the resonance of those ridges should occur theoretically for a median wavelength slightly higher than 800m.

We compare the distribution of smoothed curvature of the upper slopes (*Cs*) of the topography to the *Cs* where the landslides are different scales (Ls=200m, Ls = 300m and Ls=500m) that would correspond to possible wavelengths at which topographic resonance could occurs (Fig. S17.e and S17.d). In the particular areas surrounded in white, the landslides seems to oversample areas with high smoothed curvature. Hence, topographic amplification may be an explanation for slope failures in that areas as the expected ground-motion should be stronger on high *Cs* (or H_d).

Figures

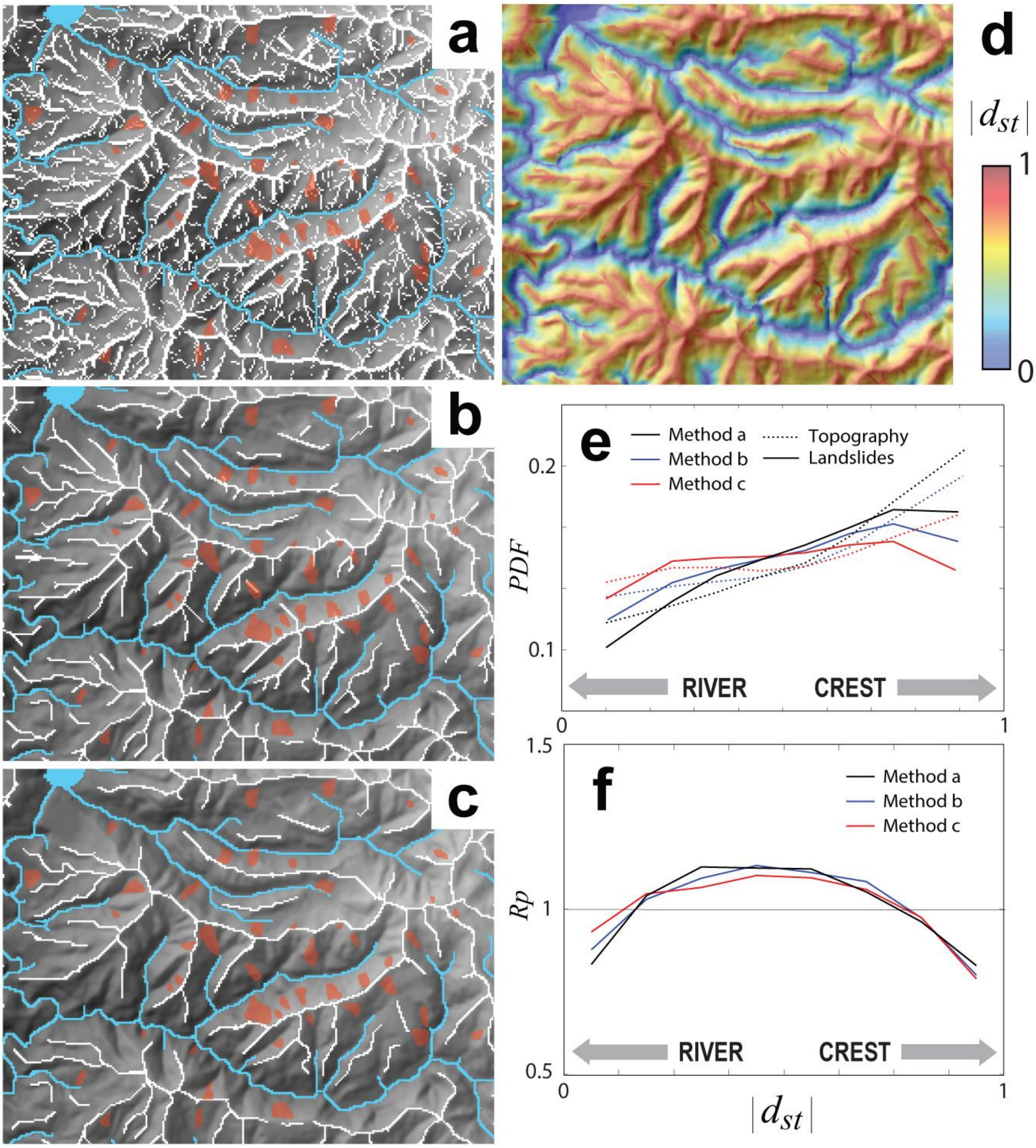

Figure S1: a-c Network of crests (white) and rivers (blue) in a 25km^2 portion of landscape in Japan. Channel heads are evaluated after Montgomery, 2001, using a threshold of 0.07 km^2 of drainage area. Crests are generated with three different methods: Crest cells are mapped as a) cells of null flow accumulation (NFA) b) cells of NFA and above a positive curvature threshold (PCT) (used in this study) and c) cells of NFA, PCT and separating identified watershed (including a first order river). d- Map of the normalized distance to stream $|d_{st}|$ in the area for the landscape b. *e*- Plots of PDF_{topo} and PDF_{ls} computed in the three landscapes. Although PDF_{ls} is derived from the same set of landslides (mapped in red polygons), its distribution with $|d_{st}|$ co-varies with the distribution of the topography

and is biased by the decreasing density of crests (from a to c). f) After normalization, the plots of R_p in the three landscapes are almost identical.

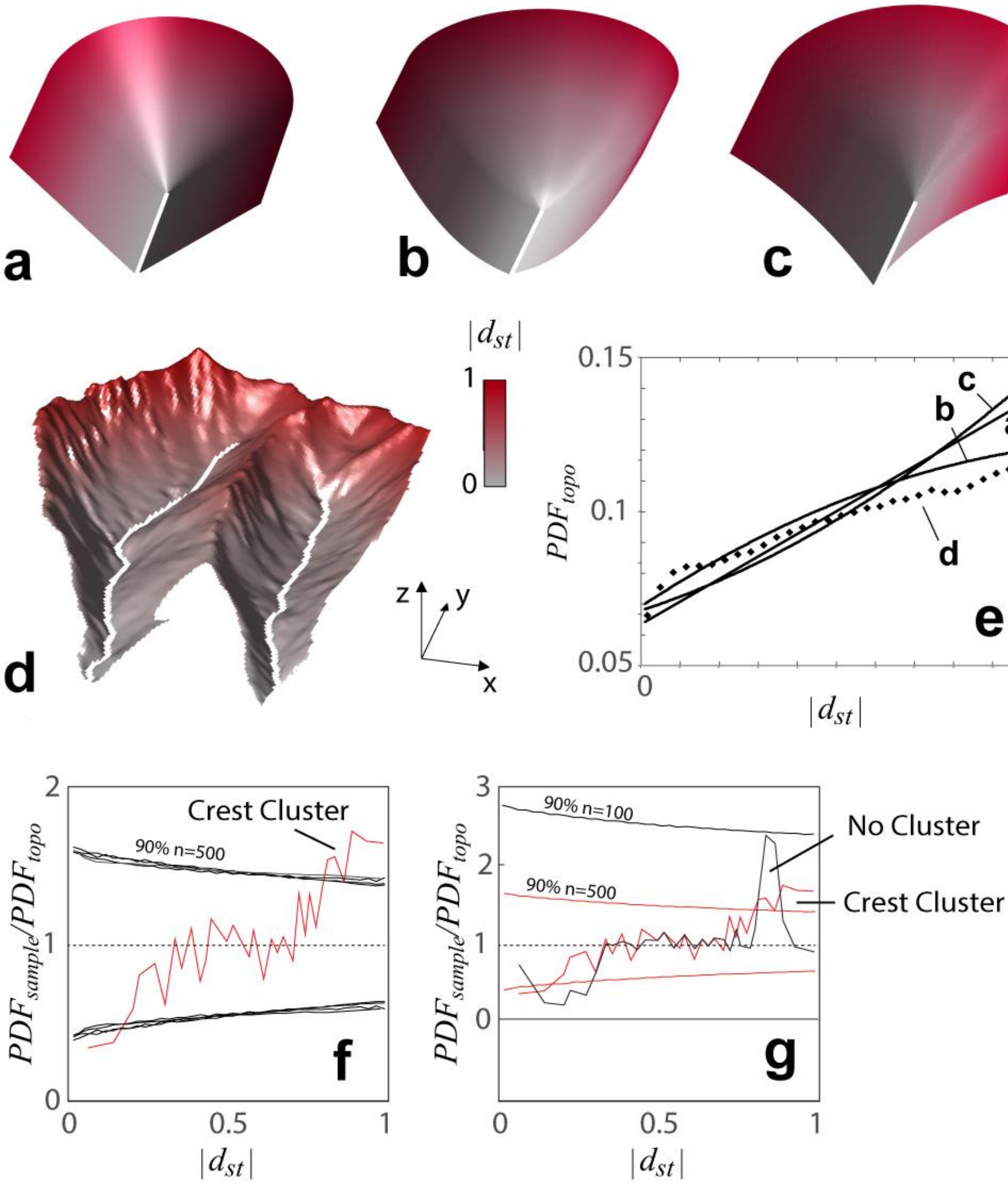

Figure S2: e) Plots of the Probability Density Function PDF_{topo} computed in 3 synthetic catchments of varying hillslope convexity (a-c) and in a pair of real catchments in Taiwan (d). f) 90% Prediction interval Irp in the 4 catchments associated with a random draw of 500 cells (black). Rp in a macrocell of the taiwanese foothills affected by 500 landslide cells (red). Two cases Rp plots in a macrocell of the taiwanese foothills affected by 500 (red) and 100 (black) landslide cells respectively. The peak observed in the second case is not interpreted as a cluster since it remains confined within the interval of fluctuation of random draws of 100 cells.

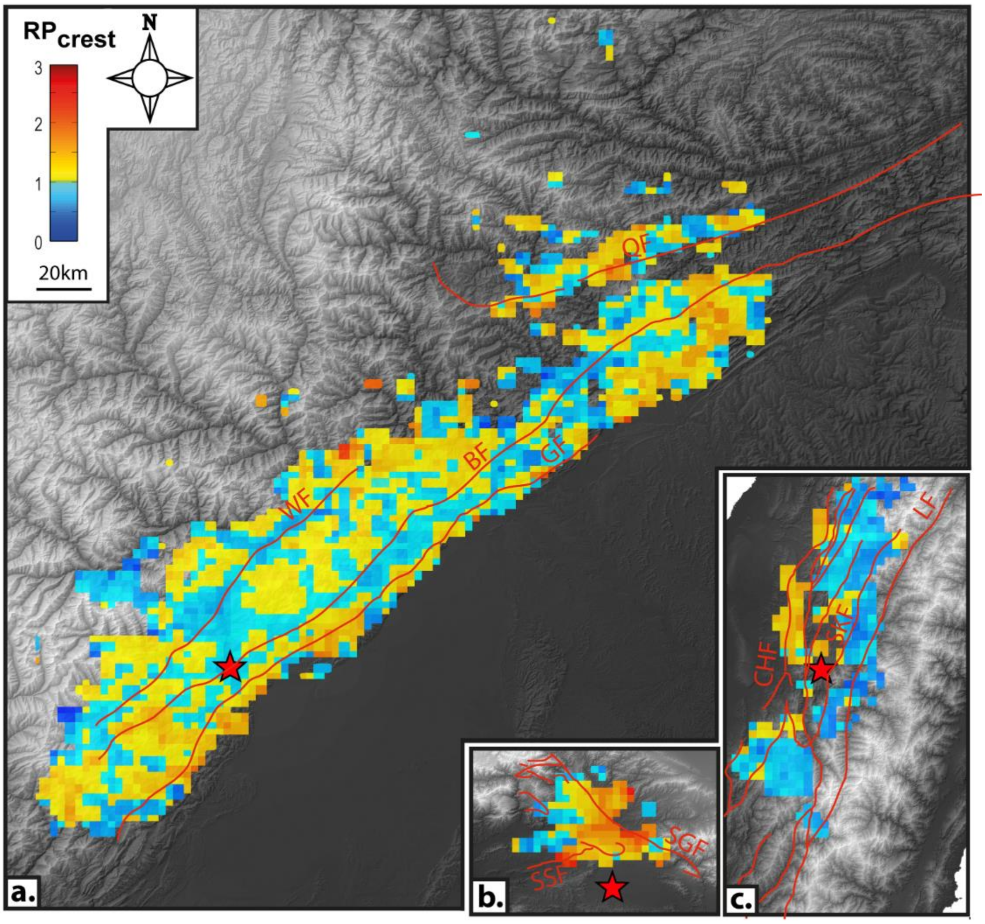

Figure S3: ***Rp_{crest}*** **values in the a. Wenchuan, b. Northridge and c. Chi-Chi epicentral areas. The 3 maps are at the same scale. All macrocells, including statistically meaningless ones, are represented. Main faults are represented by red lines and epicenters by red stars. WF: wenchuan fault; BF: Beichuan fault; GF Guanxian fault. b. SSF: Santa Susanna fault; SGF: San Gabriel fault .c. CHF Chelungpu thrust fault; SKF: Shuilikeng fault; LF: Lishan fault.**

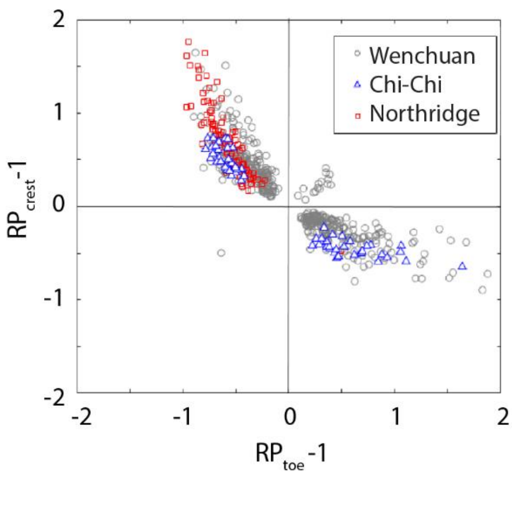

Figure S4: Values of ***RP_{crest}*****-1 plotted along Values of** ***RP_{toe}*****-1 in every macrocells of the 3 inventories. The inverse correlation shows the absence of double clustering.**

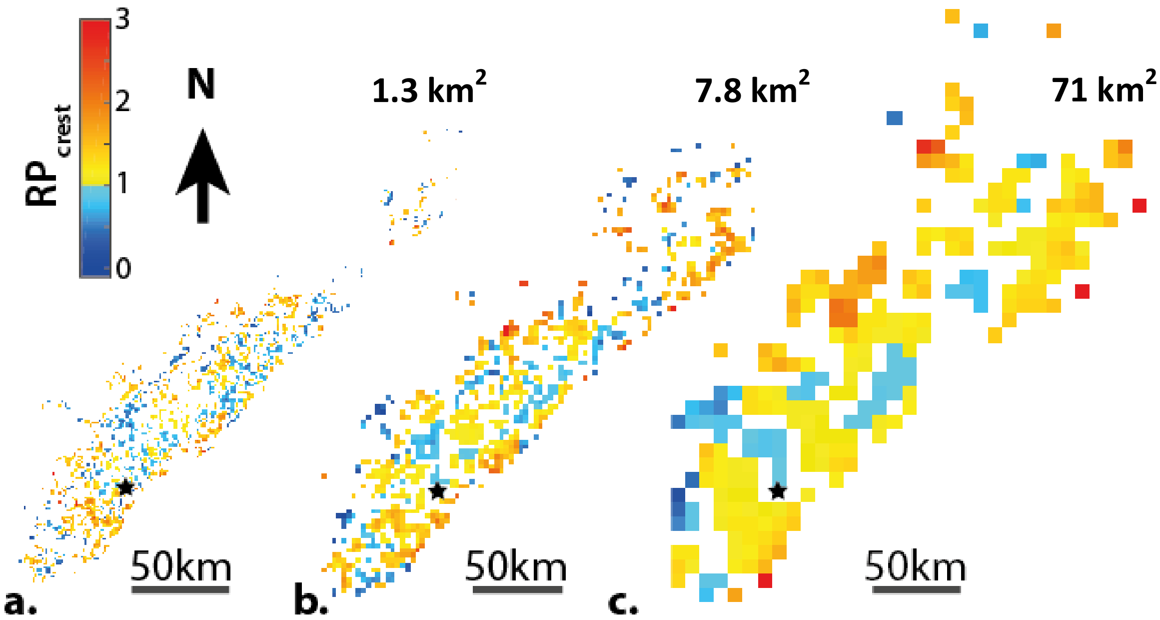

Figure S5: Rp_{crest} map in the Wenchuan epicentral area with varying macrocell size. The black stars indicate the epicenter of the 2008 Wenchuan earthquake. Note that the clustering patterns remain similar. The mean value of Rp_{crest} over the whole epicentral area is 1.15, 1.17, and 1.23 (see table 2 for the other cases).

Table 2: Mean value of Rp_{crest} with varying macrocell size.

Macrocell size	**1.3 km^2**	**7.8 km^2**	**71 km^2**	**282 km^2**
Chi-Chi Earthquake	0.82	0.87	0.86	0.84
Typhoon Morakot	0.37	0.5	0.54	0.51
Wenchuan Earthquake	1.15	1.17	1.23	1.22
Northridge Earthquake	1.68	1.61	1.47	1.6

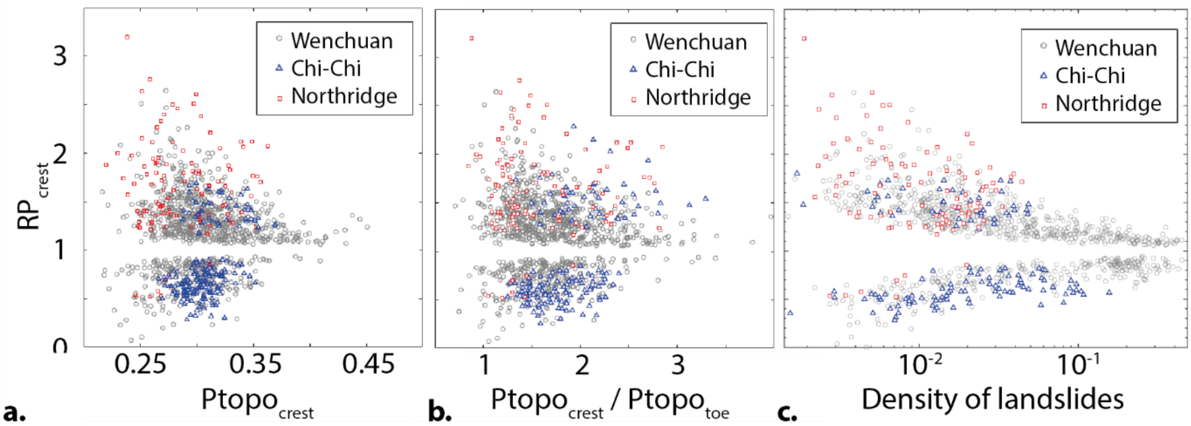

Figure S6: *RP_{crest}* plotted along a. *$Ptopo_{crest}$* and b. *$Ptopo_{crest}/Ptopo_{toe}$* and c. landslide density for the three study areas. The plots a. and b. show no correlation, insuring that crest clustering is independent of the amount of landscape standing along crests/rivers in the landscape. The *Rp_{crest}* is converging toward one for high density of landslides, hense there is not macrocells with false clustering due to the landslide concentration.

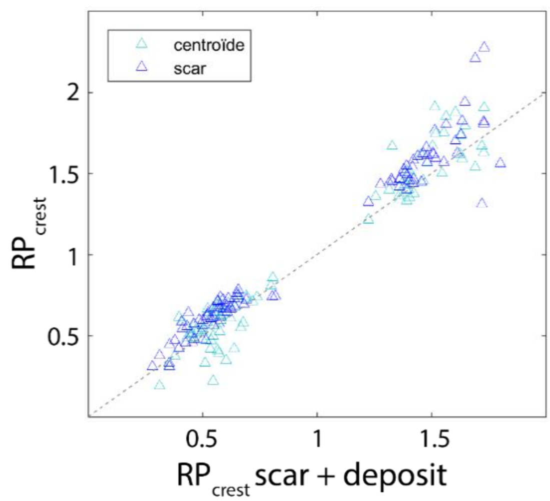

Figure S7: Comparison of the crest clustering values (*RP_{crest}*) in Chi-Chi obtained using the total landslide surface with the one obtained using the landslide centroid (light blue triangles), and the landslide scar (dark blue triangles) considering Domej et al, 2017 observations. A 1:1 linear relation is represented by the black dotted line. A quasi 1:1 relation is observed between the methods.

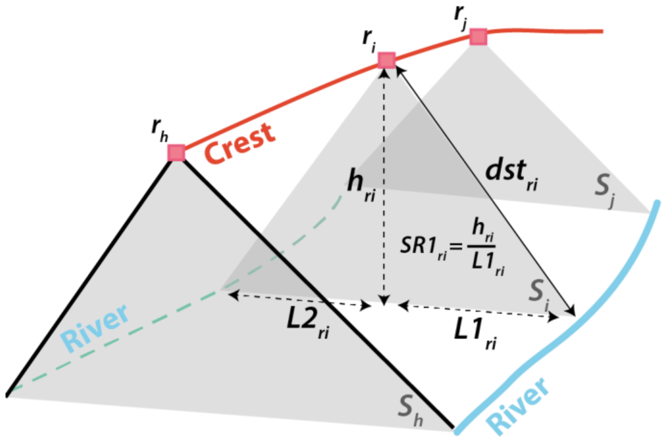

Figure S8: Extraction of geometric topographic features on each ridge point. The section of the ridge through the ridge point *ri* (*Si*) is simplified by a triangular shape (gray). h_{ri} is the slope height, dst_{ri} is the distance to the stream and $L2_{ri}$ and $L1_{ri}$ are the half base widths measured at a given ridge point *ri*. The shape ratio $S1_{ri}$ is defined as the ratio of h_{ri} to $L1_{ri}$.

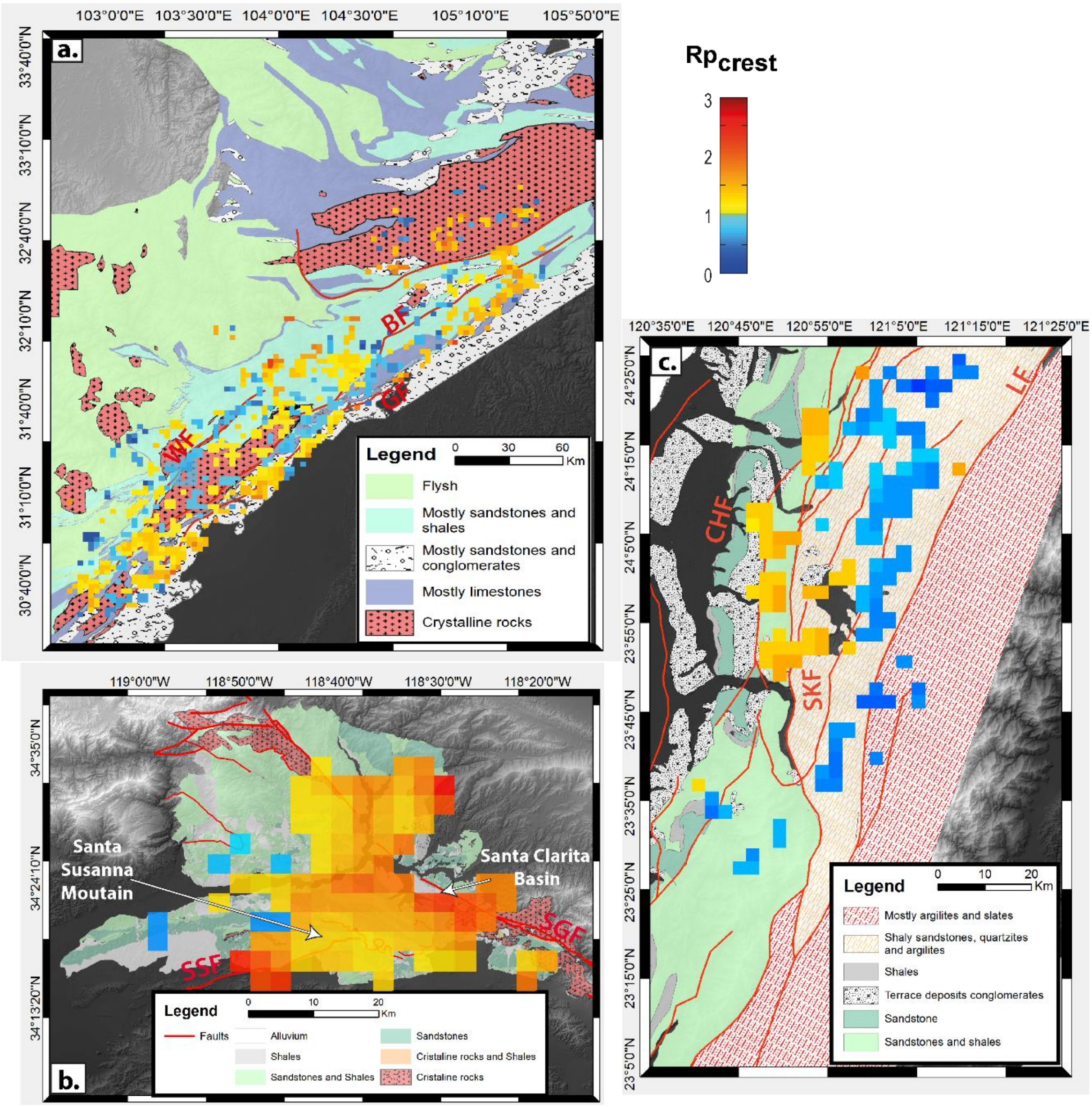

Figure S9: Lithological unit maps superimposed by Rp_{crest} maps in the a. Wenchuan, b. Northridge, c. Chi-Chi epicentral areas. The main faults are represented by red lines (a. WF: Wenchuan fault; BF: Beichuan fault; GF Guanxian fault. b. SSF: Santa Susanna fault; SGF: San Gabriel fault .c. CHF Chelungpu thrust fault; SKF: Shuilikeng fault; LF: Lishan fault). The upper slope clustering 90% maps are represented in transparency (Rp_{crest}>1: crest-clustering, Rp_{crest}<1: toe-clustering).

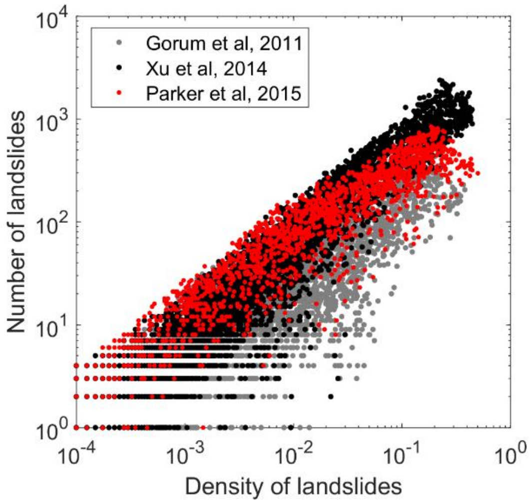

Figure S10: Number of individual landslides plotted with landslide density (ratio of surface covered) computed in each macrocell using the 3 landslide databases of the Wenchuan case: Gorum et al., 2011; Parker et al., 2015; Xu et al., 2014. The more precise is the catalog, the more small landslides there are. Amalgation and over mapping are observed in Parker et al., 2015 and Gorum et al., 2011 inventories.

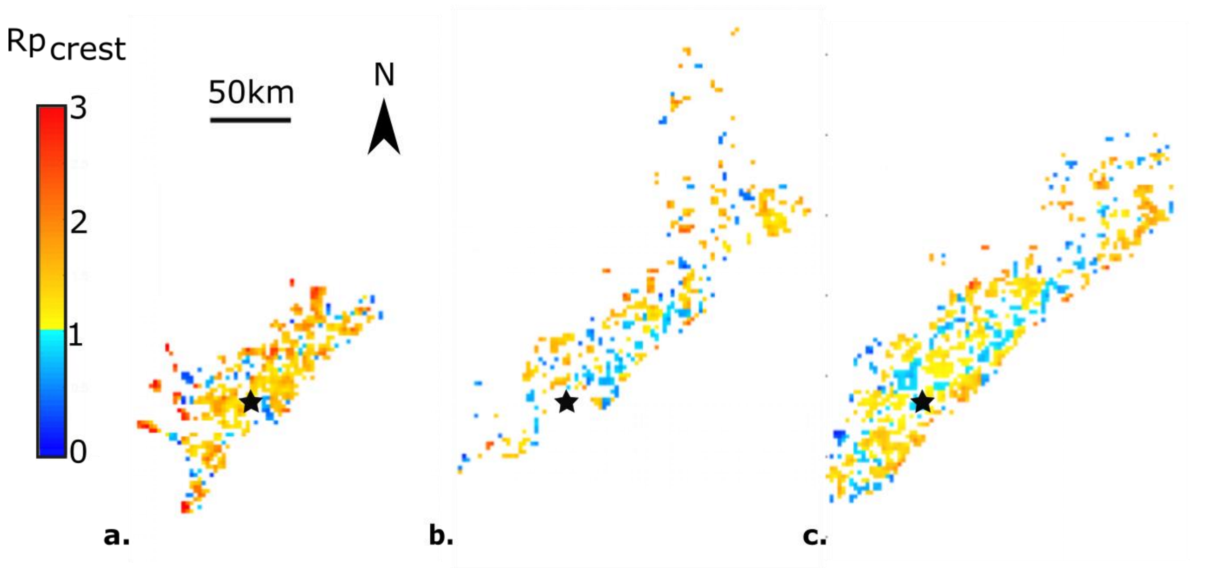

Figure S11*:* Rp_{crest} maps obtained using the a. Parker et al. 2015, b. Gorum et al. 2011 and c. Xu et al. 2014 landslide databases. The Wenchuan earthquake epicenter is represented by the black star. The Xu et al, 2014 one covers a larger area and has more statistically valid macrocells than obtained with the two others.

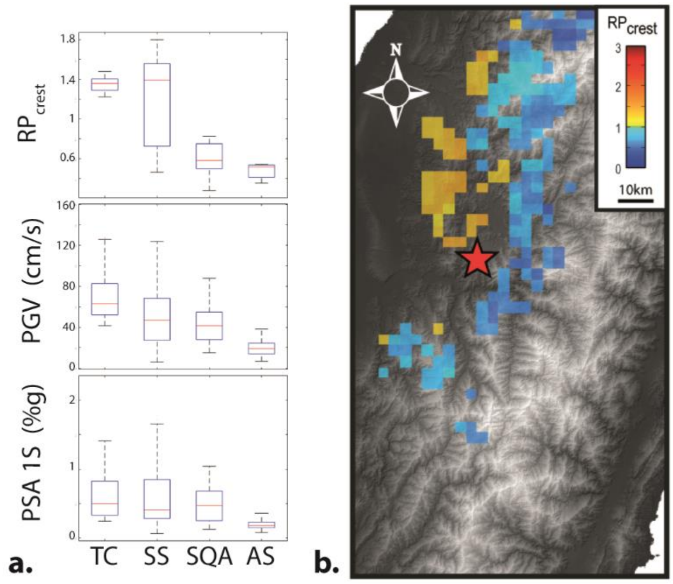

Figure S12: *Rp_{crest}, PGV* and *PSA* distributions plotted in a. lithological groups and b. *Rp_{crest}* map in the Chi-Chi epicentral area. TC: terrace and conglomerates SS: sandstones and shales SQA: shaly sandstones quartzite and argillites AS: argillites and slates. Both *PGV* and *PSA 1s* decrease with rock strength. Consequently, it is uneasy to dissociate ground motion control from lithological control on the reduction of *Rp_{crest}* toward the east.

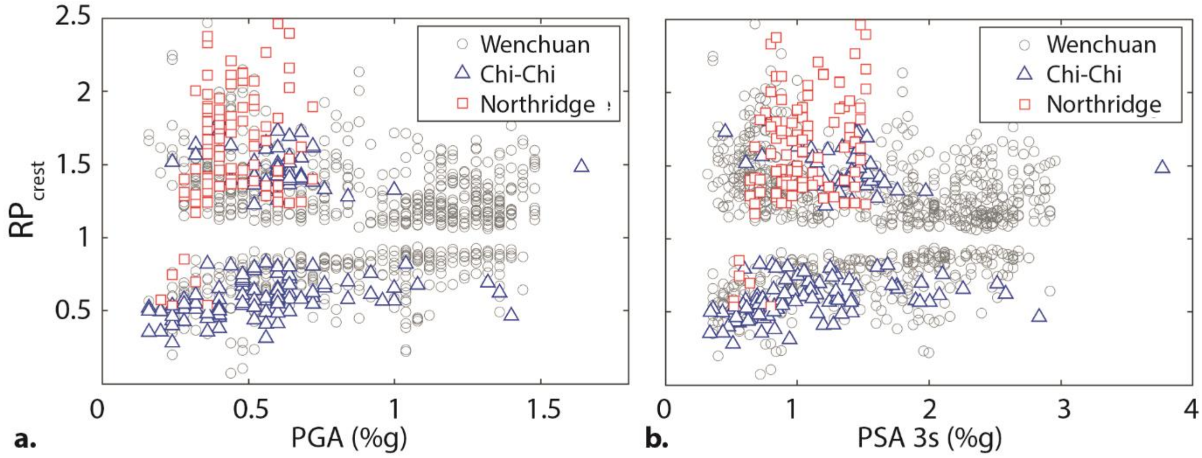

Figure S13: *Rp_{crest}* as a function of seismic features: a. Peak Ground Acceleration (PGA) (%g), b. Pseudo Spectral Acceleration at 3s (PSA 3s) (%g) for the Wenchuan, Northridge and Chi-Chi epicentral areas. Regional seismic parameters do not seem to explain landslide position along hillslope.

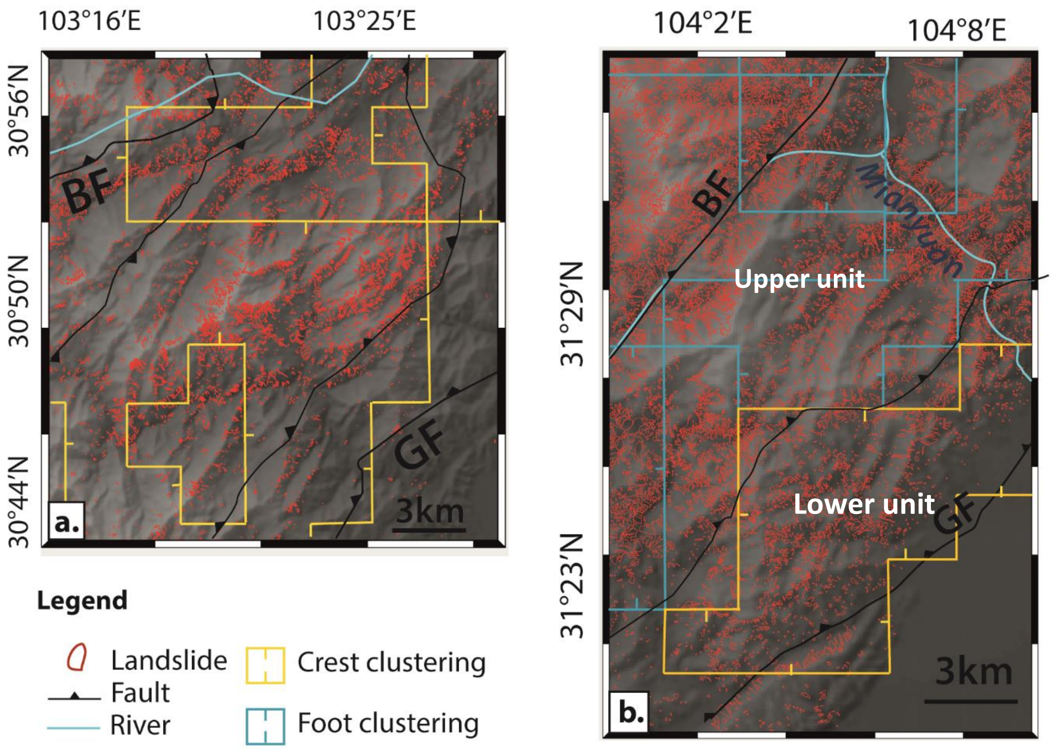

Figure S14: Snapshots of the landslide maps in a. the Sanjiang klippe and the b. foothills. The locations of a. and b. are reported in Fig. 7a. In the lower unit of the central zone and the Sanjiang klippe the landslides cluster around the crests. In the upper unit of the central zone the landslides cluster downslope along the Beichuan Fault. BF: Beichuan fault; GF Guanxian Fault.

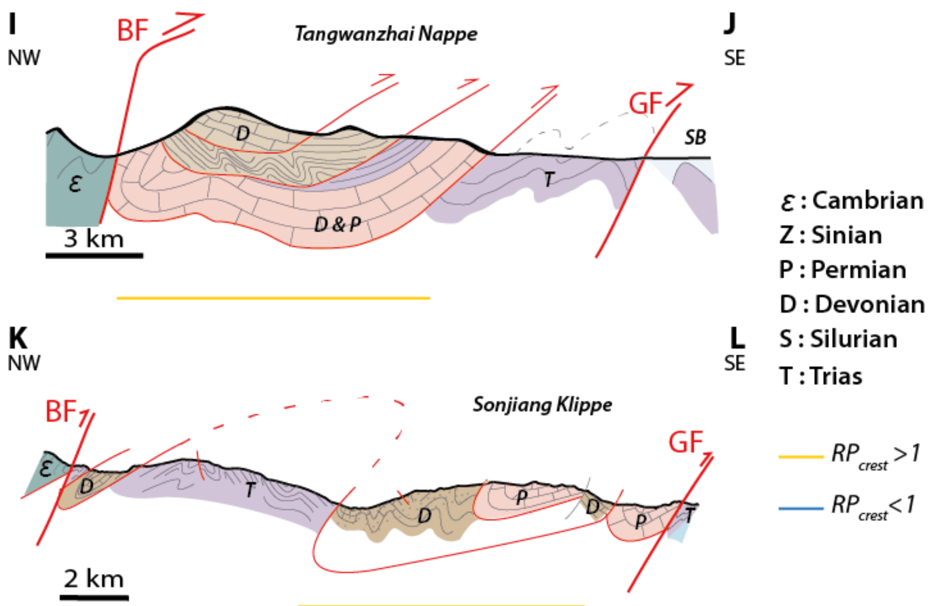

Figure S15: Cross sections of the South of the Tangwanzhai nappe (I-J) and of the Songjiang Klippe (K-L). Cross section locations are reported in Fig. 8a. BF: Beichuan fault; GF Guanxian Fault (after Robert 2011).

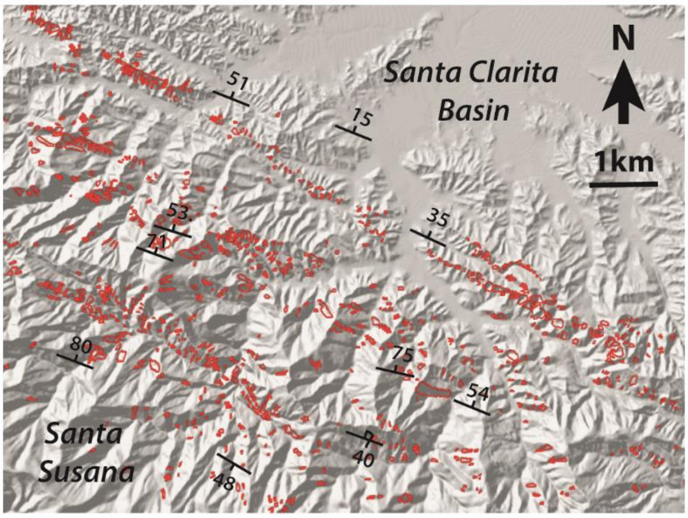

Figure S16: Northridge earthquake-induced landslides in the Northern flank of the Santa Susana Mountain. Landslides are represented by red polygons. Most of the landslides are concentrated on the top of the scarp slopes in the north the Santa Susana Mountain

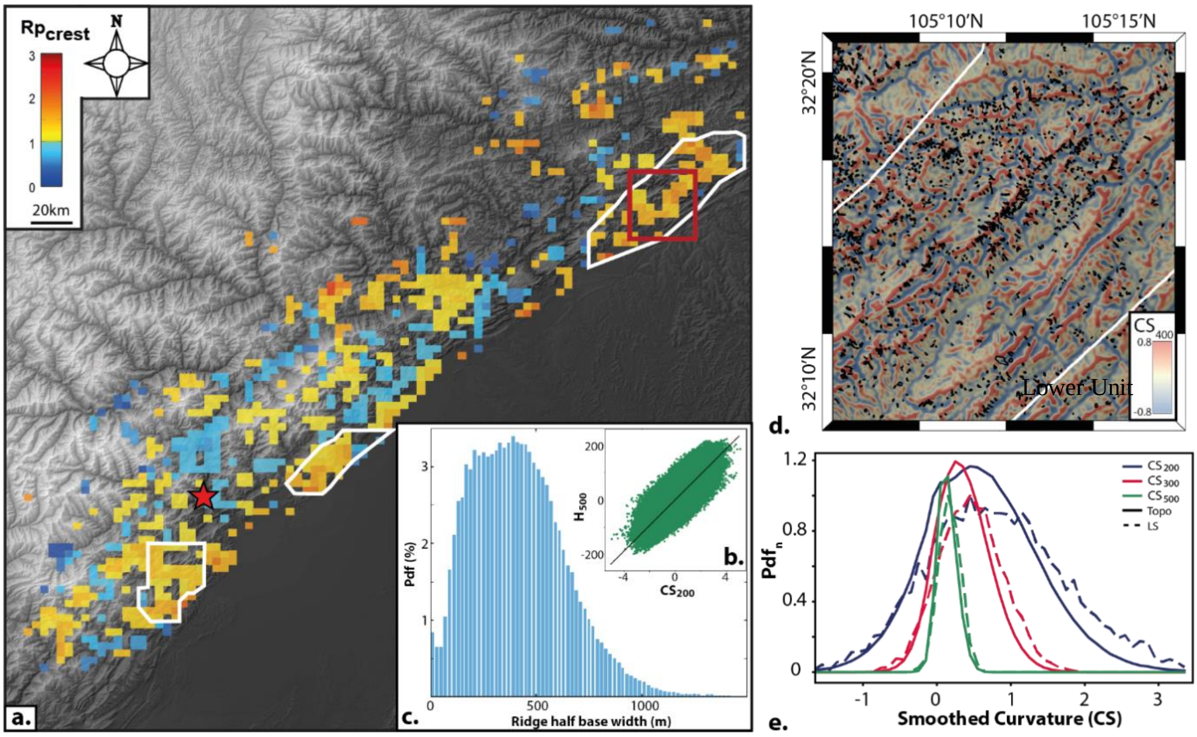

Figure S17. Variation of ridge geometric features in areas of crest clustering that could be related to topographic ground motion amplification. A. Location of areas where crest clustering could be explained by topographic

amplification. B. Smoothed curvature and relative elevation are positively correlated. C. Normalized distribution of ridge half-base width in the area surrounded by white line in a. d. Snapshot of curvature smoothed at a scale L=400 superimposed by the landslides polygons (black lines). Most of the landslides seem to occur in area where CS_{400} is the highest. E. Probability density function (Pdf) of smoothed curvature at different scale (200m, 300m and 500m) of the upper slope ($|d_{st}|$>0.75) of all the topography and only the smoothed curvature pdf of the upper slope cells covered by landslides. For each scale, the pdfs have been normalized by the maximum of the pdf of the landslides. The landslides tend to be located on area with higher CS.

Complementary references

Domej, G., Bourdeau, C., Lenti, L., Martino, S. and Pluta, K.: Mean landslide geometries inferred from globla database, 94 Italian Journal of Engineering Geology and Environment, 2, 87–107, doi:10.4408/IJEGE.2017-02.O-05, 2017.

Durante, F., Di Giulio, G., Tallini, M., Milana, G. and Macerola, L.: A multidisciplinary approach to the seismic characterization of a mountain top (Monteluco, central Italy), Physics and Chemistry of the Earth, 98, 119–135, doi:10.1016/j.pce.2016.10.015, 2017.

ESRI. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute, 2011.

Maufroy, E., Cruz-Atienza, V. M., Cotton, F. and Gaffet, S.: Frequency-Scaled Curvature as a Proxy for Topographic Site-Effect Amplification and Ground-Motion Variability, Bulletin of the seismological society of America, 105(1), 354–367, 2015.

Montgomery, D. R. Slope distributions, threshold hillslopes, and steady-state topography. American Journal of science, *301*(4-5), 432-454, 2001.

Rai, M., Rodriguez-Marek, A., & Yong, A.. An empirical model to predict topographic effects in strong ground motion using California small-to medium-magnitude earthquake database. Earthquake Spectra, *32*(2), 1033-1054, 2016.

Rai, M., Rodriguez-Marek, A., & Chiou, B. S. Empirical terrain-based topographic modification factors for use in ground motion prediction. Earthquake Spectra, 33(1), 157-177,2017.