# Supplementary materials

## *A segmentation approach for the reproducible extraction and quantification of knickpoints from river long profiles*

Boris Gailleton, Simon M. Mudd, Fiona J. Clubb, Daniel Peifer and Martin D. Hurst

# 1. Overview

This document contains supplementary materials for the manuscript *A segmentation approach for the reproducible extraction and quantification of knickpoints from river long profiles*, with additional details about the analysis on landscapes and the sensitivity analysis for testing the effect of user-defined parameters on the method.

# 2. Example parameter files

We have provided example parameter files for running the analyses performed in the manuscript, which can be found in directory `Example_parameter_files`. It contains a detailed readme file with instructions on how to adapt it for another analysis.

# 3. DEM metadata and parameter files

We have provided the specific parameter files used for each analysis in order to make it fully reproducible. The structure of the directory is as follows:

```
|--Analysis_files
  |--Santa_Cruz_Island
  |--QF
```

The directory `Analysis_files` contains the parameter files used to run the analysis on the two test sites, where each site is contained its own subdirectory. Each file can be used to run the analysis and reproduce the results shown in the main manuscript. In addition to the parameter files for running the full knickpoint analysis, we also have included georeferencing information for the sites in the form of ENVI `hdr` files which contain the coordinate system and extent of each DEM analysed. Users can download lidar 1 m data for the Smugglers catchment within the dataset *2010 Channel Islands Lidar Collection* available on OpenTopography. After downloading the data, the user must project it into UTM coordinates and clip to the correct extent with the information in the `hdr` files. We have done this using GDAL. You can find instructions on the LSDTopoTools documentation website. We also provide the metadata for the Quadrilatero Feriferro site (Brazil), however we are not allowed to share the TanDEM-x 10m DEM following the terms of its licence. The SRTM 30m version of it can also be accessed on OpenTopography.

The source code required to run these parameter files are in the GitHub repository https://github.com/LSDtopotools/LSDTopoTools_ChiMudd2014. You will find links to instructions for installing and running the software, or you can see the documentation for "Channel steepness analysis with LSDTopoTools" at the main LSDTopoTools documentation website. Documentation specific to the knickpoint analysis can be found in the knickpoint section of the documentation.

# 4. Sensitivity to parameters in natural landscapes

In order to test the sensitivity of the knickpoint detection method to user-defined parameters, we ran a set of sensitivity analyses. The results are presented below, and the script used to generate the raw data and the figures are provided in the following directories:

```
|--Sensitivity_analyses
  |--Concavity_index_lambda
  |--Concavity_index
  |--Target_node_n_skip
  |--MC_iterations
  |--Combining_window
  |--Step_window_coeff
  |--A0
```

We provide the scripts that were used to run the sensitivity analyses on the parameters needed for different part of the algorithm toolchain: *Concavity_index_lambda*, *Target_node_n_skip*, *MC_iterations*, *Step_window*, *Combining_window*, *Grid_Resolution* and *A0*. These files can be found in the directory `Sensitivity_analyses` and are written in *Python* to provide basic control of the C++ code. Each file follows the same structure: they automatically write and execute the parameter files for ranges of parameters (combined or not). It then uses the Python *subprocess* module to call the analysis (compiled C++ LSDTopoTools) and plotting (LSDMappingTools) codes and automatically organises the results in subfolders using Python's *os* module. Running the sensitivity analyses can take a significant amount of computation time. We therefore used Python's *multiprocessing* module to take advantage of the now common multi-core CPUs and run several analysis at the same time. The scripts used to then generate violin plots and the z-score are available in the supplementary materials. They can be used to (i) reproduce our results or (ii) generate new range of tests or new combination of parameters in case of specific studies.

## 4.1. Sensitivity to the regulation parameter

The $\lambda$ parameter (*lamda_TVD*) is the regulation parameter that controls the denoising intensity in the Total Variation Denoising algorithm from Condat (2013). We tested a range of $\lambda$ values from 0.1 to 100000 for $k_{sn}$ calculated with a range of concavity indexes from 0.1 to 0.9, as $\theta$ controls the order of magnitude of $k_{sn}$. We then determined the best $\lambda$ for each value of $\theta$ by plotting the $k_{sn}$ extracted with Mudd et al. (2014) algorithm and the denoised signal against $\chi$, as shown on fig. S1.
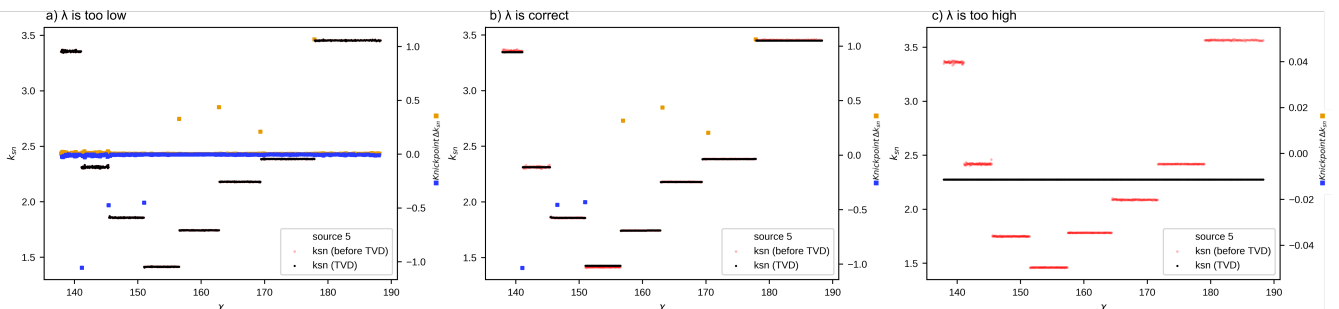
**Figure S1.** $\lambda$ *parameter selection for the regulation of* $k_{sn}$. *a) A* $\lambda$ *value which is too low: variations are not de-noised and plenty of low-magnitude knickpoints (blue squares for negatives and red squares for positive) are detected. The combining process cannot even clean these low-magnitude knickpoints as they are alterning between positive and negative magnitudes. b) A more appropriate value of* $\lambda$*: most of the small variations in* $k_{sn}$ *are flattened without altering the main signal and the few remaining are relevantly combined by the algorithm. c) A* $\lambda$ *which is too high: all the variations are flattened by the denoising algorithm. This data has been extracted from the Smugglers catchment described in the main manuscript.*

Results from this analysis showed that a default $\lambda$ value can be suggested for each $\theta$ (Table S1). Although we have implemented these default values into the algorithm, this parameter can still be adapted by the user if required.

| Concavity | 0.1 | 0.15 | 0.2 | 0.25 | 0.3 | 0.35 | 0.4 | 0.45 | 0.5 | 0.55 | 0.6 | 0.65 | 0.7 | 0.75 | 0.8 | 0.85 | 0.9 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Regulation | 0.1 | 0.3 | 0.5 | 2 | 3 | 5 | 10 | 20 | 40 | 100 | 200 | 300 | 500 | 1000 | 2000 | 5000 | 10000 |

**Table S1.** List of default regulation parameter ($\lambda$) per each concavity ($\theta$).

## 4.2. Sensitivity to the concavity index theta



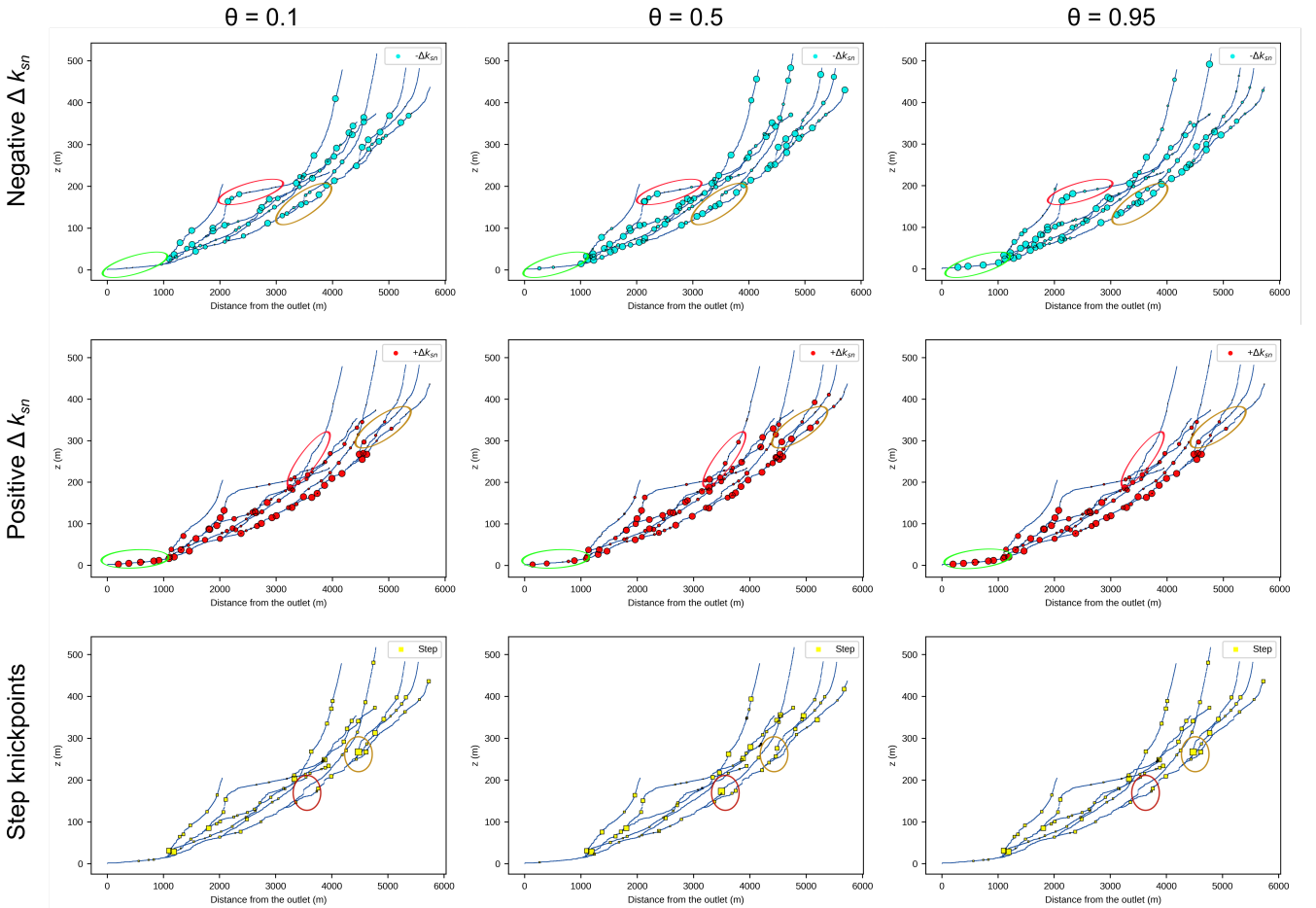**Figure S2.** *Effect of* $\theta$ *used to calculate* ($\chi$) *prior to the calculation of* $k_{sn}$. *These profiles, calculated for*

*various θ, show that the relative magniture of the knickpoints is affected. For each knickpoint morphology (i.e., a decrease of $k_{sn}$, an increase of $k_{sn}$ and stepped knickpoints), we circle example zones where the relative magnitude of knickpoints is significantly impacted by the change in θ. The largest knickpoints are the ones with an absolute $\Delta k_{sn} > 3^{rd}$ quartile of the population and the smallest knickpoints the ones with an absolute $\Delta k_{sn} < 1^{st}$ quartile of the population for each θ. This data has been extracted from the Smugglers catchment described in the main manuscript.*

We also ran a sensitivity analysis on the concavity index θ from Flint's Law (see main manuscript). This parameter describes the overall concavity of the river long profile and should be carefully chosen (e.g., Mudd et al., 2018). However, we ran a sensitivity analysis to assess the extent of its effect on the knickpoint extraction as discussed in the main manuscript. We ran the algorithm for a range of θ from 0.1 to 0.95 in 0.5 increments. We find that the relative magnitude of knickpoints is affected by the concavity value, but the location of knickpoints does not significantly change (see Figure S2).

## 4.3. Sensitivity to channel steepness extraction

As discussed in the main manuscript, *target_nodes* and *n_skip* control the size of the segments in the $k_{sn}$ extraction using the Mudd et al. (2014) algorithm. We ran a combined sensitivity analysis on these parameters to assess the effect of segment size in knickpoint extraction. Our results show that the segment size can impact the results of the knickpoint detection, and therefore we suggest that these parameters need to be carefully chosen by the user. Although reference values can be set (*e.g.* 80 target nodes and a skip value of 2), these parameters depends on data quality: smaller segments will fit the profiles with more fidelity but will therefore be noise sensitive, whereas larger segments can average noise at the risk of ignoring small variations. Therefore, these parameters cannot be set as default and should depend on the data source (*e.g.* SRTM 90m vs airborne lidar). Figure S3 shows results for a wide range of segment size jointly controlled by the 2 parameters. It shows that smaller segments fit the river profile with higher fidelity. It also impacts the knickpoint expression: smaller segments tend to catch steep reaches (*e.g.* waterfalls) by fitting a segment to it (and therefore 2 slope break knickpoints) when larger segments tend to express such reaches with a step (and therefore as a stepped knickpoint). It also slightly changes the knickpoint location on the river long profiles, as different river nodes are used depending on the segment size.
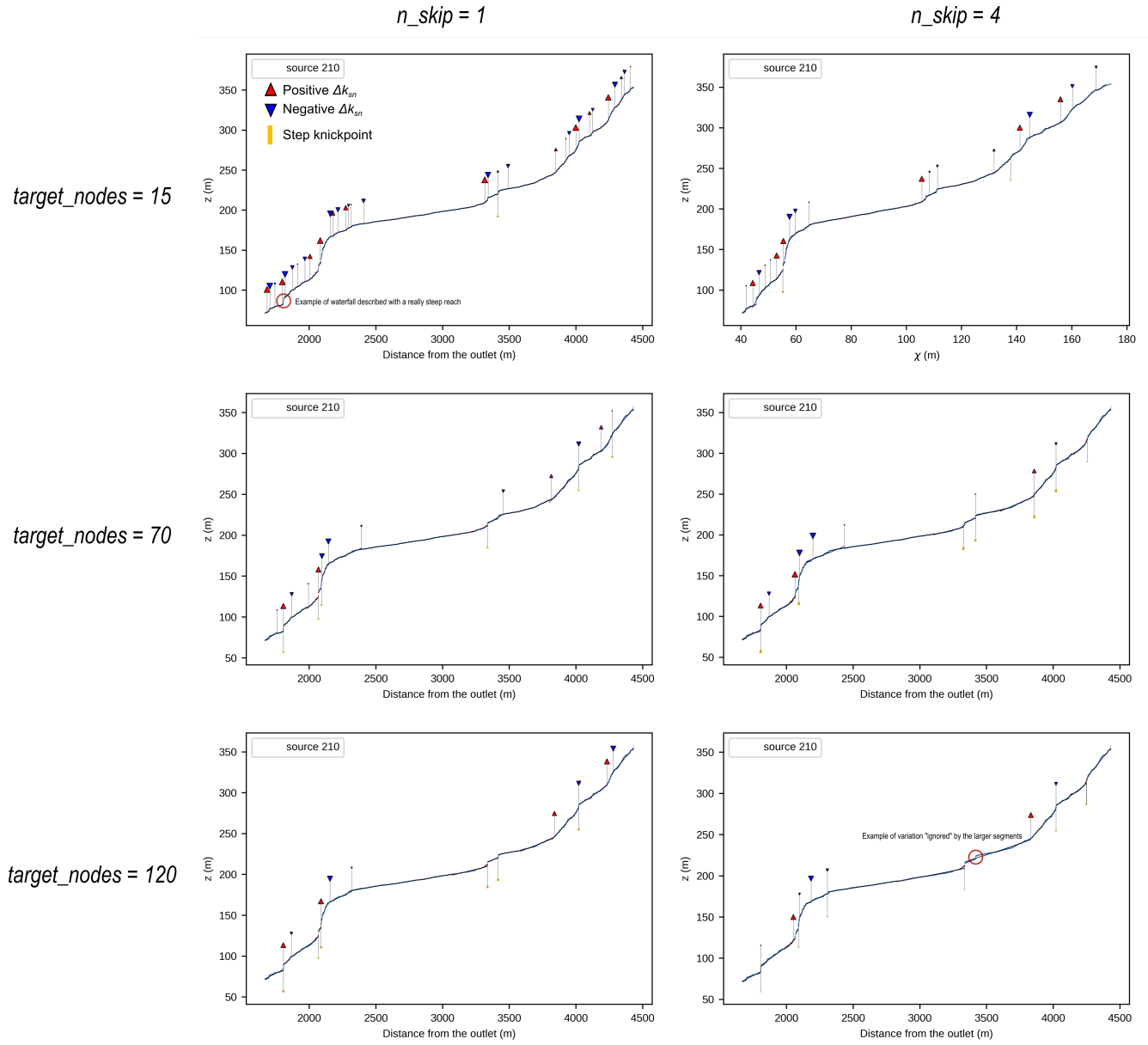
***Figure S3.** Effect of target_nodes and n_skip on knickpoint extraction. Smaller values (top-left) generate smaller segments than high values (bottom-right). River profiles are diplayed in blue, with the segments extracted with Mudd et al. (2014) algorithm on top in black. Knickpoints are represented with marker offset to the river profile in order to highlight the segment boundaries. This data has been extracted from the Smugglers catchment described in the main manuscript.*

Another parameter linked to $k_{sn}$ extraction is the number of iterations used in the segment-fitting algorithm (*e.g.* how many times a sub-set of nodes is picked following a Monte-Carlo sampling scheme). We ran a sensitivity analysis with a range from 1 to 1000. The results shows that more iterations tend to produce cleaner signals and we therefore recommend a minimum value of 50 for this parameter. However, increasing this parameter to above 50 will not significantly affect the algorithm output.

# 4.4. Sensitivity to the combining window

Despite using a denoising algorithm to clean our $k_{sn}$, some stepped knickpoints can still be expressed as a succession of slope break knickpoints. We therefore use a combining window to solve this problem (full explanation in the main manuscript). We ran a sensitivity analysis on the

size of this window, with a range of 0 to 100 nodes. Results showed in the main manuscript and in Figure S4 suggest that this parameter does not have a significant influence on the results, except when set with extreme values (<10 or >50). A combining window of 0 can result in underestimating knickpoint magnitudes in steepened reaches and therefore bias the statistical sorting of knickpoints. High values, on the other hand, can result in the merging of unrelated knickpoints.
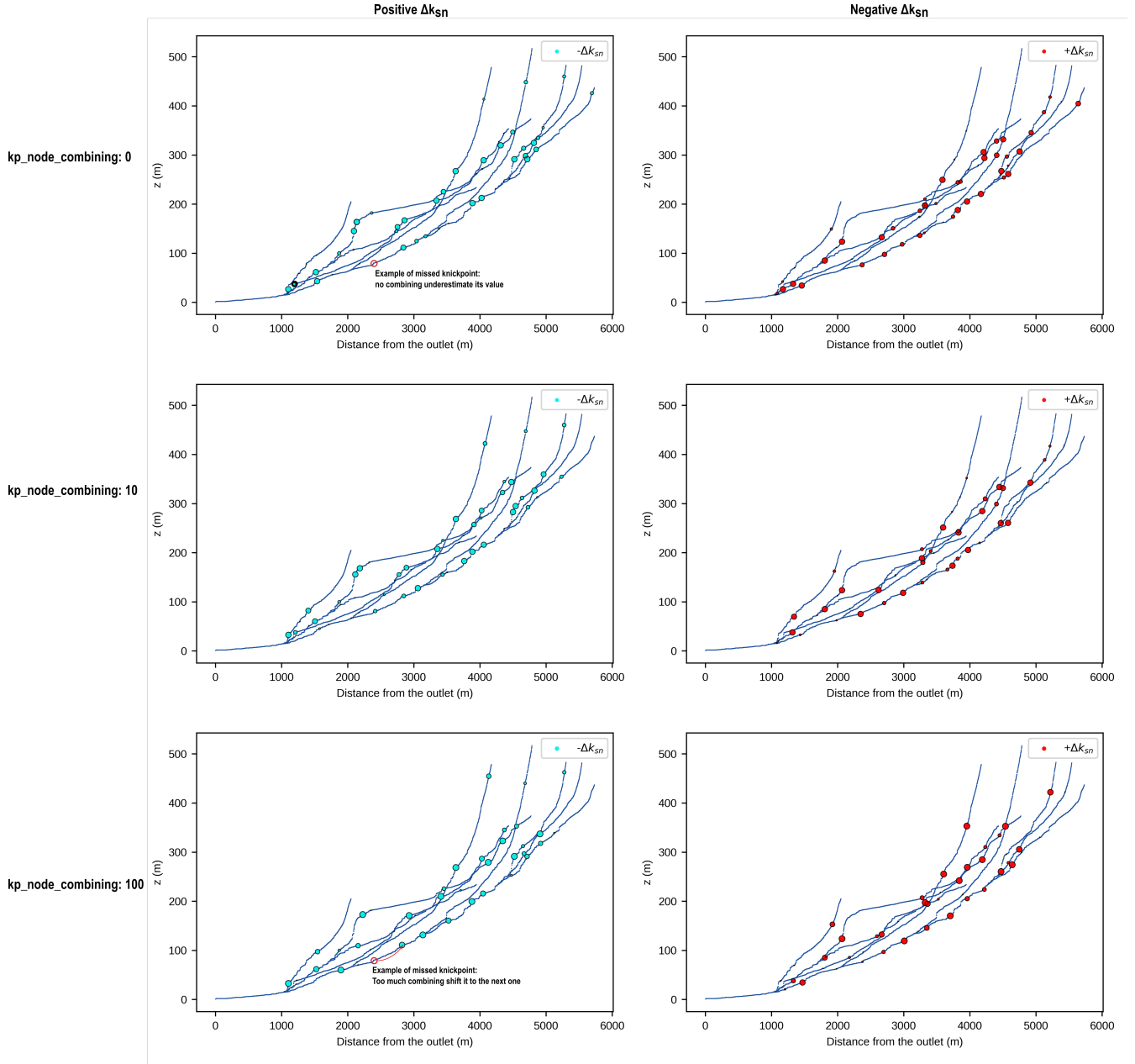


**Figure S4.** *Effect of kp_combining_node on knickpoint extraction. Values are calculated and sorted with the same threshold as in the main manuscript. It shows that the results are not significantly affected as long as the size of the combining window is within a reasonable range. This data has been extracted from the Smugglers catchment described in the main manuscript.*

## 4.5. Sensitivity to step knickpoint detection

Stepped knickpoint detection is controlled by two parameters: *window_stepped_kp_detection* noted $r_w$ and *std_dev_coeff_stepped_kp* noted $T_\sigma$ (we refer to the main manuscript for a full explanation). We ran a combined sensitivity analysis to determine their values as discussed on the main manuscript and shown in Figure S5. $r_w$ tends not to significantly alter the results as long as it is set

with a minimum size of 80 nodes (we even recommend 100 to 150 based on our tests). $T_\sigma$ needs to be picked more carefully: a low value would generates high number of artifacts (*e.g.* each node of a steep reach would be picked) and a high value would inhibit the detection. However results show that effective values are within a relatively narrow range (form 6 to 8) making this parameter easy to constrain.
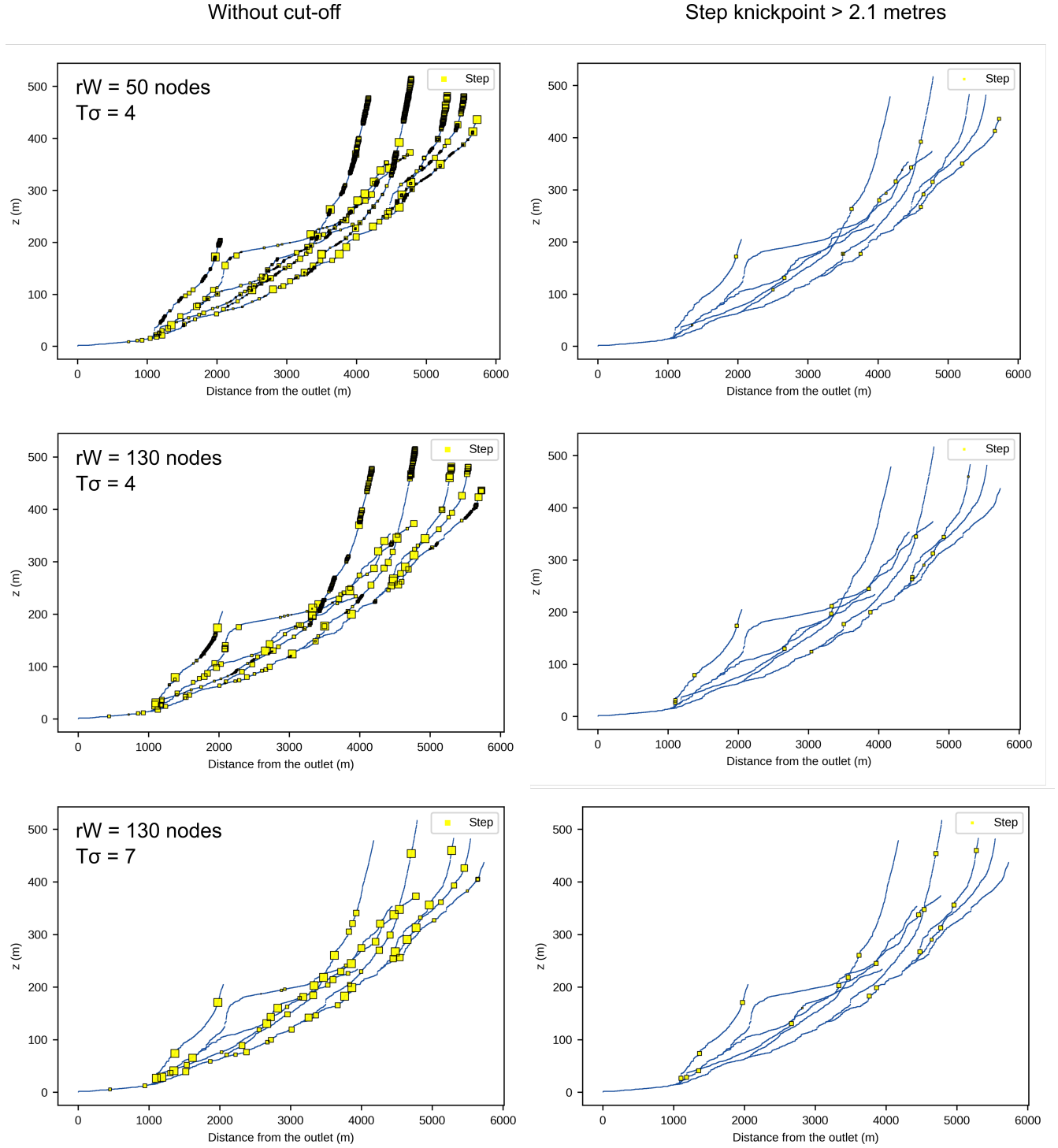


***Figure S5.*** *Sensitivity result to $r_w$ and $T_\sigma$ controlling stepped knickpoint detection. On the left the basin-wide river profiles show the raw data and on the right after applying the same threshold as in the main manuscript. Note that a correct selection of $r_w$ and $T_\sigma$ improve the step-knickpoint detection by reducing the number of artifacts detected. However, applying a threshold to select the main features also removes of the artifacts. This data has been extracted from the Smugglers catchment described in the main manuscript.*
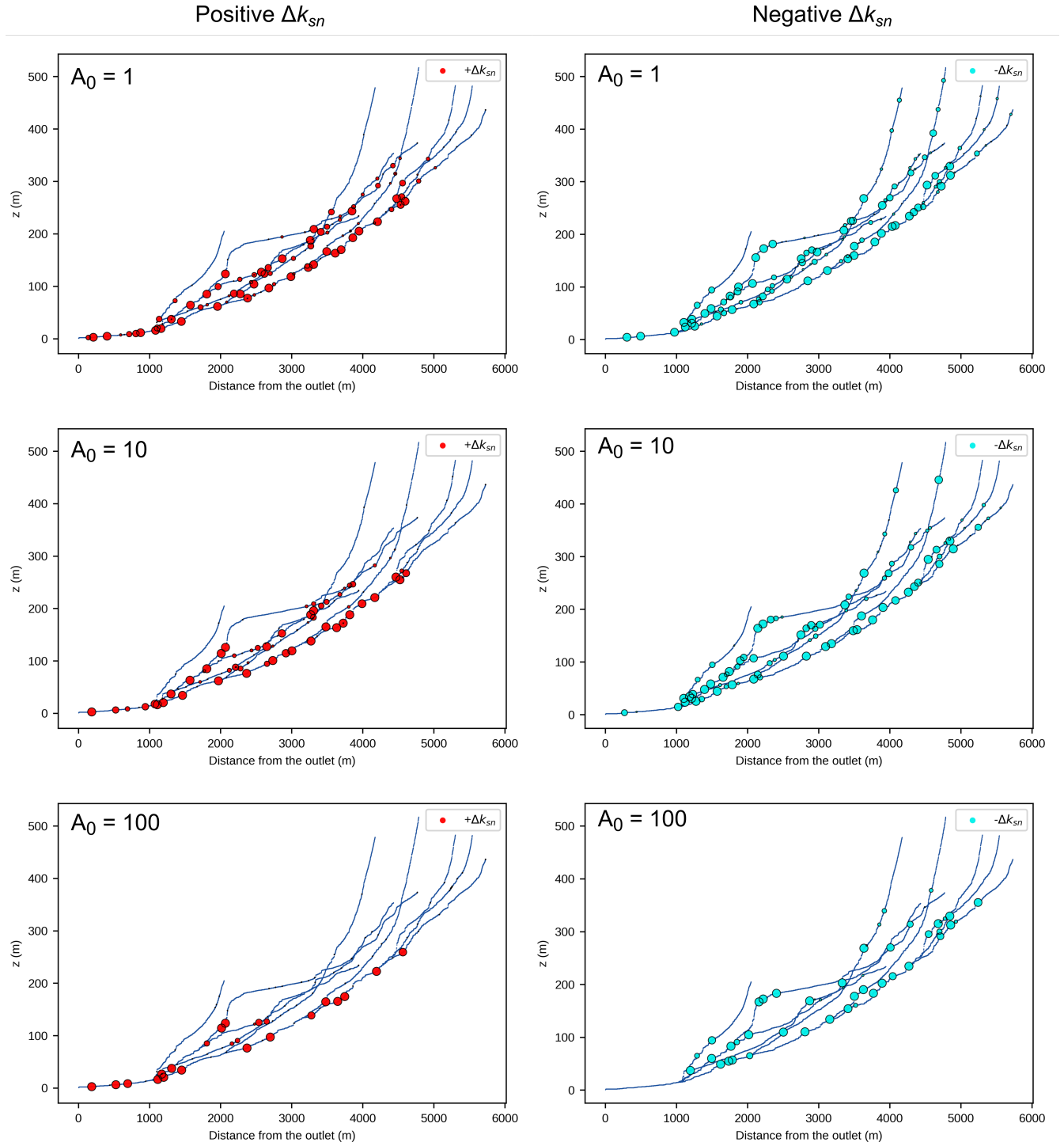
# 4.6. Sensitivity to A0



**Figure S6.** *Sensitivity result to $A_0$. Extraction has been performed with a high $\theta = 0.85$ in order to reproduce the non linearity effect. Note that the high-magnitude knickpoints on the steepened reach tend to decrease in magnitude and even disappear when incresing $A_0$. This data has been extracted from the Smugglers catchment described in the main manuscript, but with a significatly different $\theta$ for the specific aim of this sensitivity analysis.*

As discussed in the main manuscript, deriving $k_{sn}$ from $\chi$-elevation profiles can be affected by the difference in order of magnitude bewtween those two metrics. For instance, especially with high $\theta$ (*e.g.* 0.85), the majority of $\chi$ values vary between 0 and 0.5. $k_{sn}$ is the gradient of such a relationship and therefore will be attributed extremely high values in case of steep segments in such cases (*e.g.*

$d\chi$ can be 0.001 and $d$elevation 10 within a single segment). We compare the slope break knickpoints by their drop or increase of $k_{sn}$, and therefore these non-linearities can bias data sorting by overestimating the importance of such knickpoints. We therefore tested the sensitivity of $A_0$ from the $\chi$ equation. Increasing the value of $A_0$ correspondingly increases the absolute value of the $\chi$ coordinates, and therefore removes this bias. However, if $A_0$ is not equal to 1, the magnitude of the knickpoint can no longer be described in terms of $k_{sn}$. In these cases it is equivalent to $M_\chi = (\frac{E}{K*A_0^m})^{1/n}$ as described in Mudd et al. (2014). Therefore, we suggest that $A_0$ should only be modified from 1 if this problem occurs.

## 4.7. Sensitivity to grid resolution:

As suggested in hte main manuscript, grid resolution affect the knickpoint extraction. The segmentation process used to calculate $k_{sn}$ depends on the number of nodes describing each river. For an equal number of nodes targetted during the segmentation process, lower grid resolution leads to lower number of nodes describing each rivers and therefore a lower number of segments. Knickpoints tending to be located around segment boundaries, less segments will describe less knickpoints. This behaviour is illustrated on the following figure S7, where we extracted the knickpoint with $n_{tg} = 10$ and $n_{sk} = 1$ for a grid resolution downgraded using GDAL from the original Smugglers 1m Lidar DEM from 5 meters, 10 meters (e.g. TanDEM-x) and 30 meters (e.g. SRTM, Aster).
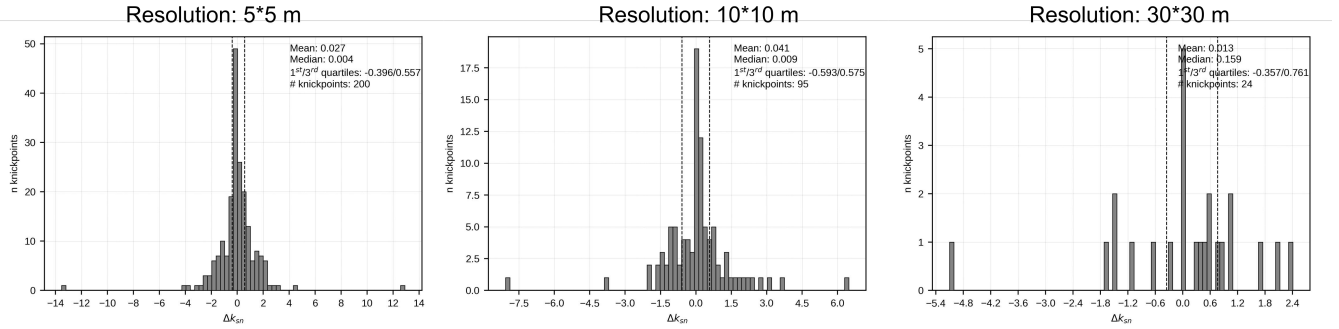


*Figure S7.* *Sensitivity results to grid resolution. Knickpoints have been extracted for a range of downgraded DEM grids. The statistical distribution and the number of slope-break knickpoints extracted shows that coarder grid resolution etract significantly less knickpoints than finer resolutions.*

# 5. Additional methodological details

The main manuscript tests the algorithm on two sites: The Smugglers catchment on Santa Cruz Island (California, USA), chosen because of its well constrained topography, freely available high resolution topographic data (lidar) and independently picked knickpoints; and the Caraca catchment (Quadrilatero Ferrifero, Mina Gerais, Brazil), chosen because of field knowledge from the area. We present and discuss the results in the main manuscript, but we provide here more details of the different outputs produced with the algorithms.

## 5.1. The Smugglers catchment (California, USA)

## 5.1.1. Sources Extraction

We took advantage of the high-precision 1-metre airborne lidar DEM to extract source with the Pelletier method (from Pelletier et al., 2013). The extracted sources and the resulting channel network are presented in Figure S8 below.
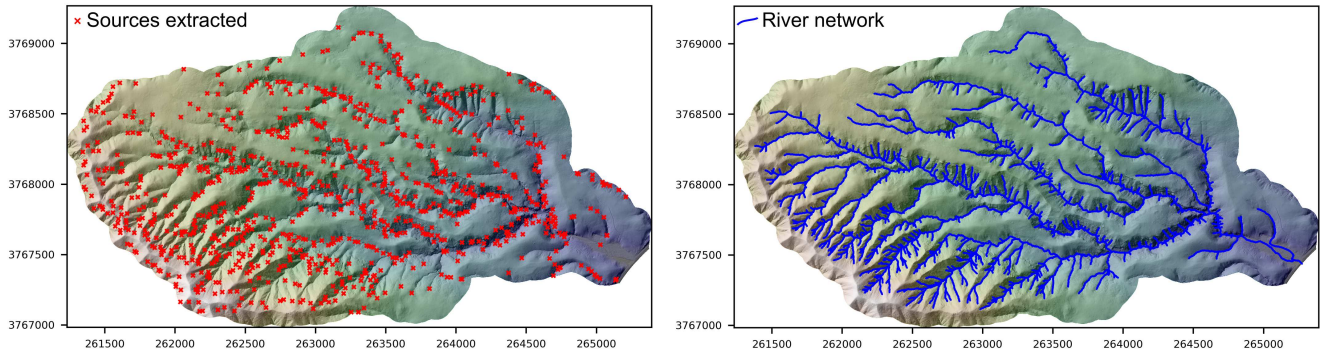


*Figure S8. Extraction of the sources (left) and the subsequent channel network (right). All the sources have an ID which allowed us to select the ones matching with the calibration dataset.*

## 5.1.2. Concavity

After extracting the river network, we determine a best-fit concavity using a method implemented in Mudd et al. (2018). The best-fit range of concavities is shown in Figure S9. For this paper we used a value of 0.25 to match with previous studies (e.g., Neely et al., 2017) which was within the range of best-fit suggested by the results of this algortihm.
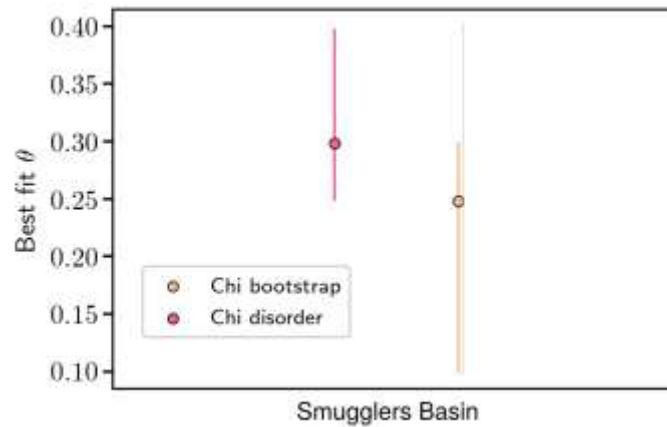


*Figure S9. Best-fit concavity using the boostrap and disorder methods described in Mudd et al. (2018). Both of these include 0.25 in their best-fit range. These methods are two different ways to test the collinearity of a main trunk and its tributaries in $\chi$-elevation space.*

## 5.1.3. Knickpoint extraction

The knickpoint extraction was performed using the following parameters (all the non-specified parameters have been set to default values, we refer to the full documentation) presented in Table S2:

| Parameter | Value | Parameter | Value |
|---|---|---|---|
| channel heads fname | smugglers_Psources | min_slope_for_fill | 0.0001 |
| m_over_n | 0.25 | ksn_knickpoint_analysis | true |
| force_skip_knickpoint_analysis | 1 | force_n_iteration_knickpoint_analysis | 500 |
| target_nodes | 30 | n_nodes_to_visit | 30 |
| TVD_lambda | 1.7 | kp_node_combining | 10 |
| window_stepped_kp_detection | 120 | std_dev_coeff_stepped_kp | 7 |

**Table S2.** List of non-default parameters used for knickpoint extraction.

The extraction first produces a dense network of knickpoints, shown in Figure S10. Knickpoints are sized by magnitude as described in the main manuscript.



*Figure S10. Raw knickpoint extraction displayed by type of knickpoints: left column shows the knickpoints expressing a drop of $k_{sn}$, middle column the knickpoints expressing an increase of $k_{sn}$ and the right column the step knickpoints.*

We sorted this dataset using the statistical distribution of these knickpoints, in order to select relevant cut-off values. These cut-offs have been selected in the specific aim to (i) match with the calibration dataset for comparison purposes, (ii) remove non-significant knickpoints (those with very low magnitude). The knickpoint distributions and cut-off values are shown in Figure S11.
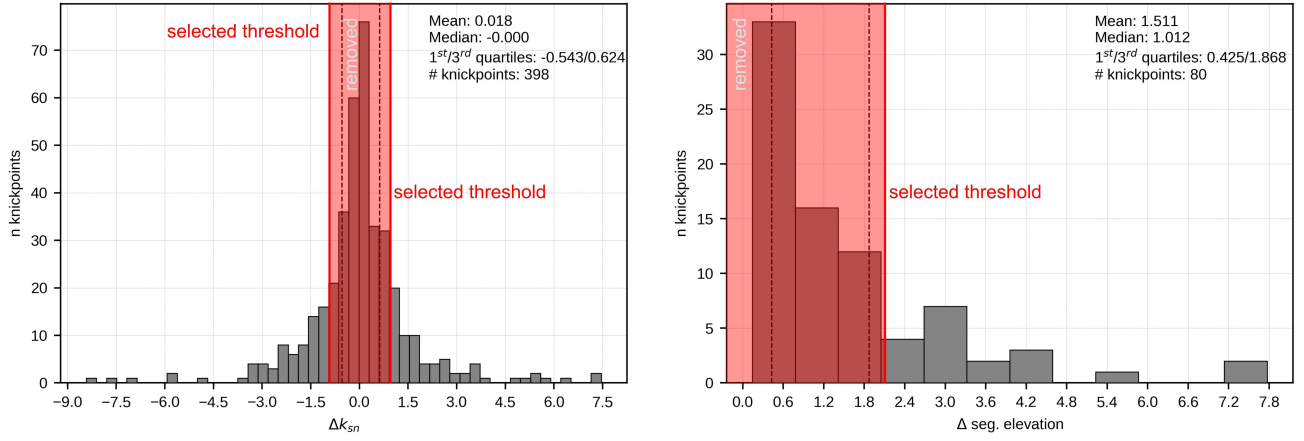
**Figure S11.** *Distribution of knickpoint magnitude per knickpoint type: on the left the $\Delta k_{sn}$ and on the right the increase of $z_{seg}$. The chosen cut-off values are illustrated by vertical bar. The red areas are represent the removed knickpoints.*

The main results after the sorting procedure are shown in the main manuscript, but here we present an extended version (Figure S12) where knickpoints are separated by morphology.
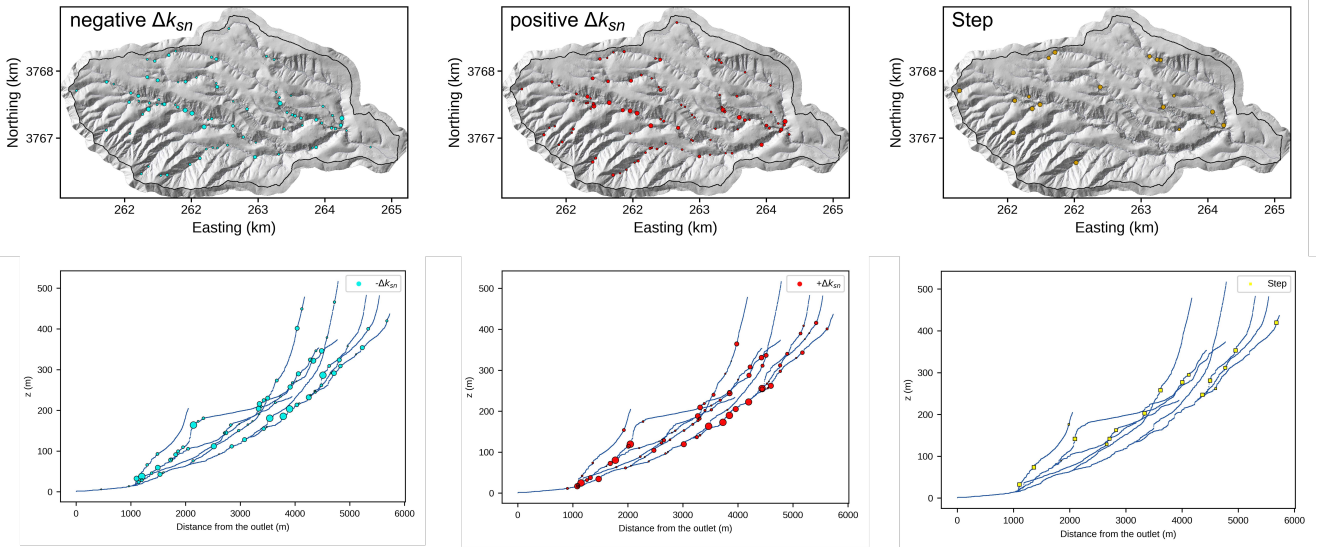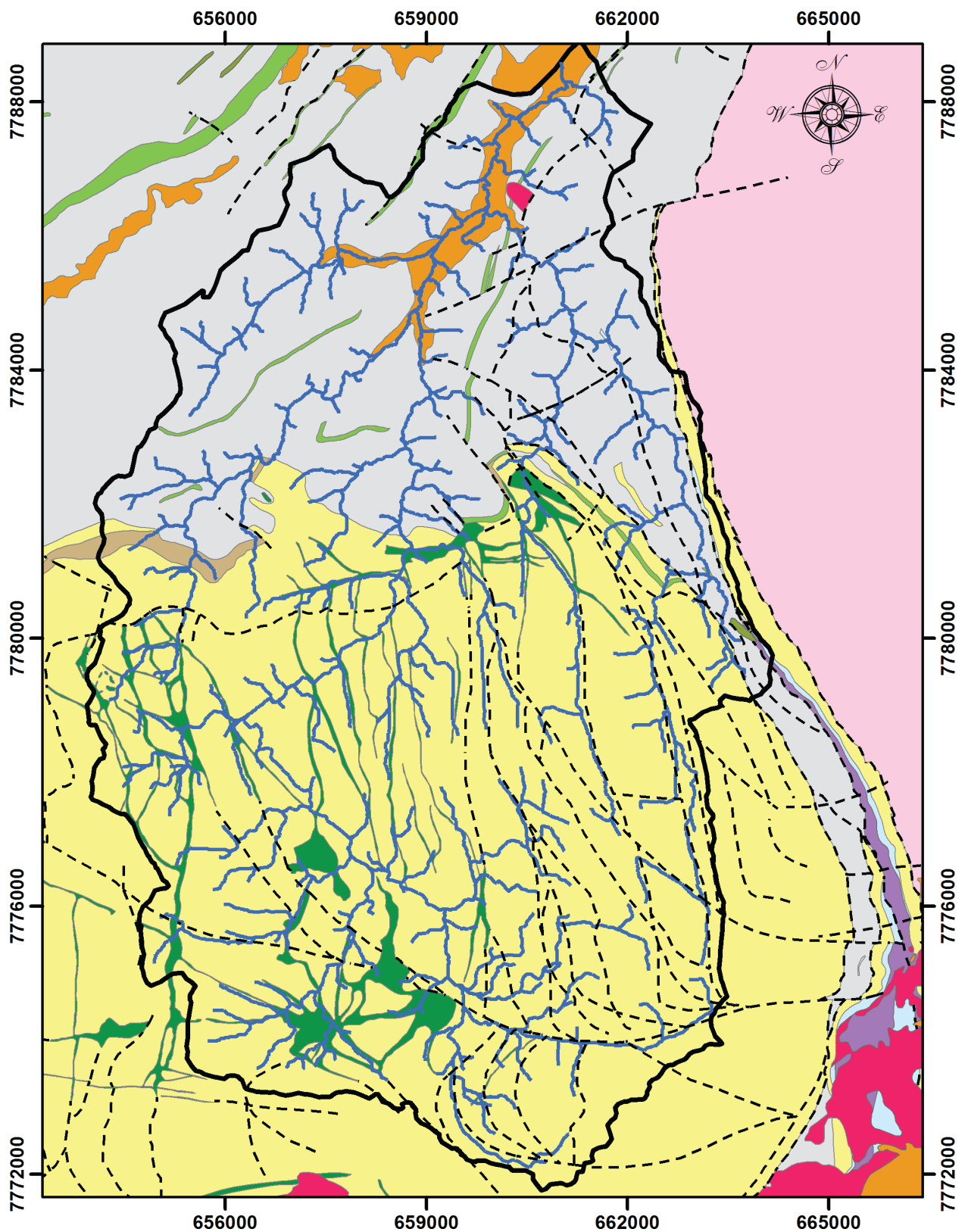


**Figure S12.** *Knickpoint extraction separated by knickpoint morphology: left column shows the knickpoints expressing a drop of $k_{sn}$, middle column the knickpoints expressing an increase of $k_{sn}$ and the right column the knickpoints expressing step knickpoints.*

# 5.2. Caraca, Quadrilatero Ferrifero (Brazil)

Our second test site in the Caraca Basin (QF, Brazil), we provide below more details about the analysis.

## 5.2.1. Geological context:

One of the main interest of the Caraca region is its heterogeneous lithology that exerts a control on the drainage network as discussed in the main paper. We therefore provide here the lithological map of the Caraca basin to illustrate this complexity, on the following Figure S13.

## Legend

- - - - - pre-Paleozoic faults

──── Hydrography

☐ Ribeirão Caraça basin

**Metasedimentary and metamorphic rocks**

☐ Quartizites
☐ Schists
☐ Phyllites
☐ Banded iron formation

☐ Conglomerates
☐ Fluvio-lacustrine deposits
☐ Iron duricrust

**Igneous rocks**

☐ Gneisses, and granitic rocks
☐ Metabasalts

UTM projection
Datum: WGS 84 / Zone: 23 South
Numerical Scale 1 : 100,000

## 5.2.2. Knickpoint extraction

Using the same procedure as the previous analysis on the Smugglers catchment, we first produced a statistical distribution of extracted knickpoints. Figure S14 shows histogram distributions of the detected slope-break and stepped knickpoints.
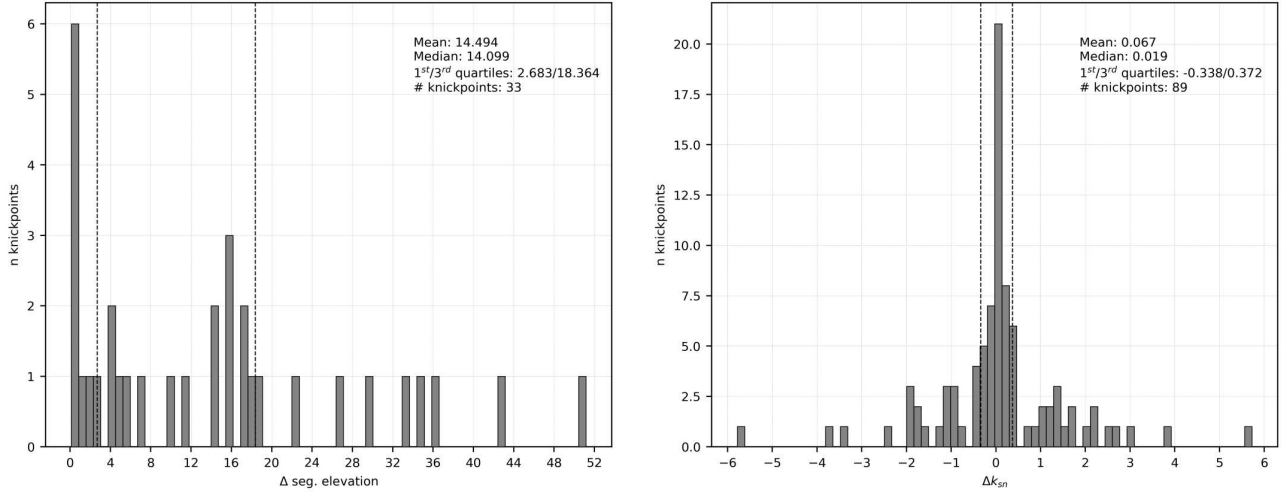


*Figure S14. Histogram distribution of the different knickpoint types.*

We also provide a detailed basin-wide profile view per knickpoint type (Fig. S15) and in map view (Fig. S16).
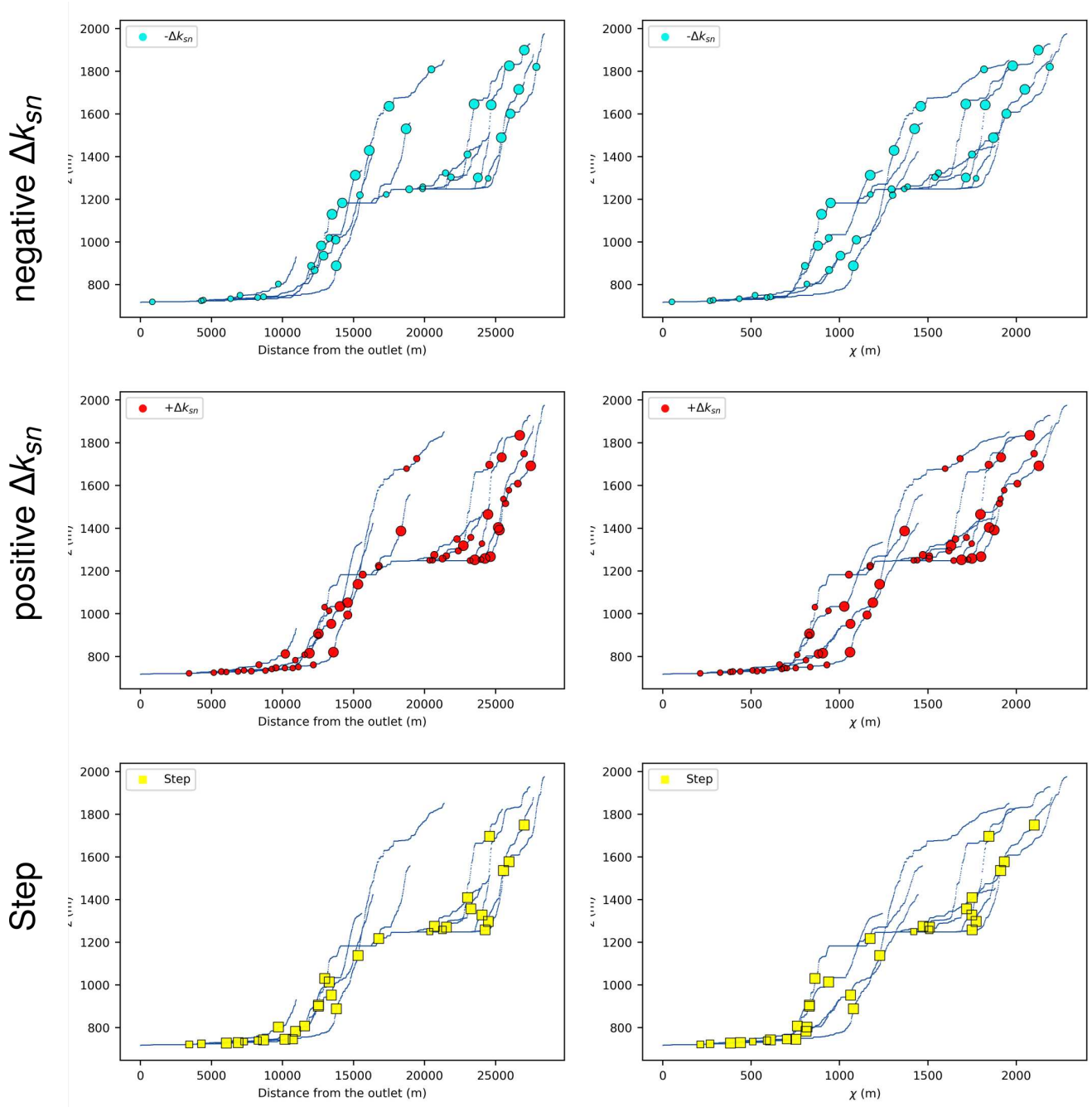
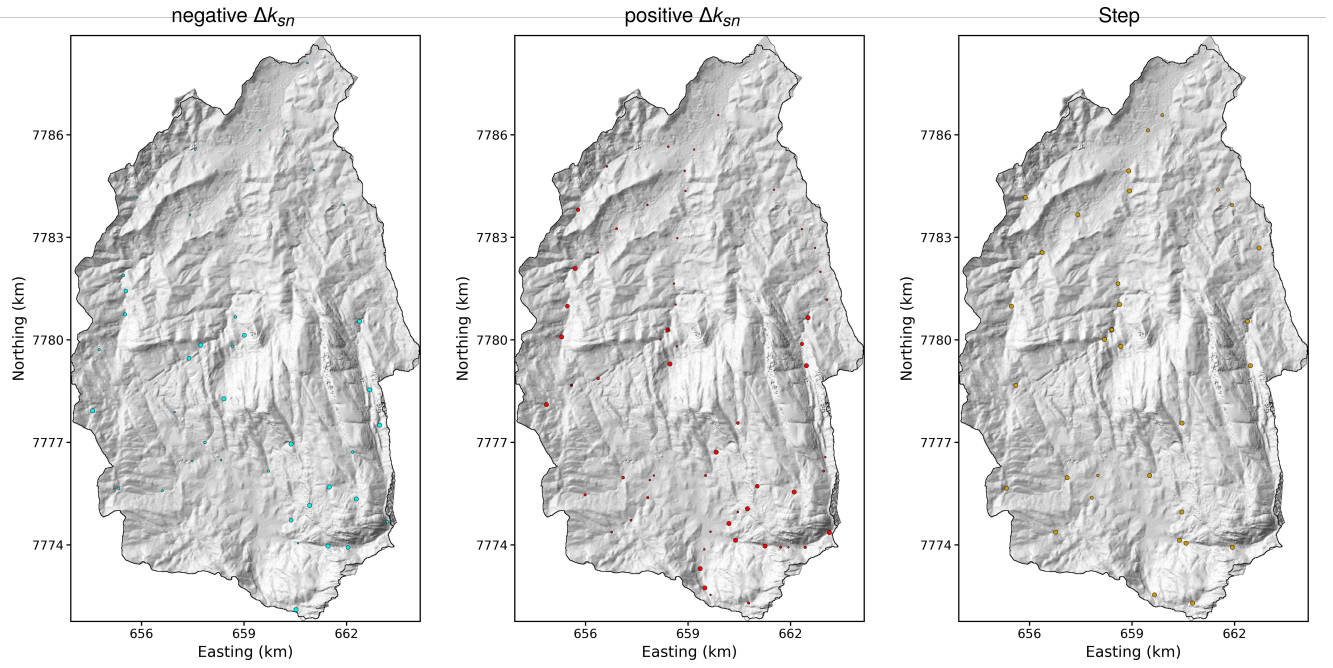***Figure S15.*** *Long and χ profile for the Caraca catchment for each type of knickpoint.*

**Figure S16.** *Map view of each type of knickpoint, extracted for the Caraca basin before any thinning.*

This first extraction has been used to select cut-off values to thin the dataset and extract significant knickpoints, as discussed in the main manuscript. We also provide here an additional figure comparing the detected knickpoints to known field-mapped locations.
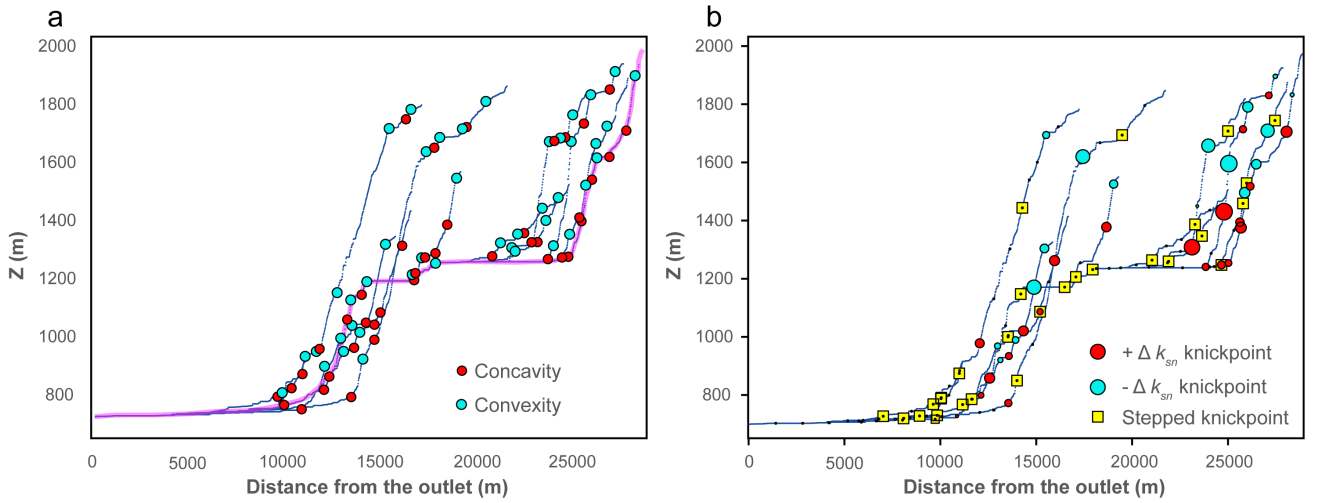


**Figure S17.** *Additional comparison between knickpoints picked by the algorithm described in the main manuscript and those picked by field mapping.*