doi:10.26443/seismica.v4i2.1387

SEISMICA

ScarpLearn: an automatic scarp height measurement of normal fault scarps using convolutional neural networks

Léa Pousse-Beltran * [1], **Théo Lallemand** [1], **Laurence Audin** [1], **Pierre Lacan** [2,3], **Andres Nuñez-Meneses** [4], **Sophie Giffard-Roisin** [1]

[1]Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre, 38000 Grenoble, France, [2]Instituto de Geociencias, Universidad Nacional Autónoma de México, Blvd. Juriquilla, 3001, 76230 Juriquilla , Querétaro, México, [3]Géosciences Montpellier, University of Montpellier - CNRS, Montpellier, France., [4]Posgrado en Ciencias de la Tierra, Instituto de Geociencias, Universidad Nacional Autónoma de México, Blvd. Juriquilla, 3001, 76230 Juriquilla , Querétaro, México

Author contributions: *Conceptualization*: Pousse-Beltran, Giffard-Roisin. *Formal Analysis*: Pousse-Beltran, Lallemand, Giffard-Roisin, Nuñez-Meneses. *Writing - Original draft*: Pousse-Beltran, Giffard-Roisin, Audin, Lacan. *Funding acquisition*: Pousse-Beltran, Giffard-Roisin, Audin, Lacan.

Abstract Geomorphic markers such as displaced surfaces, offset rivers or scarps are witnesses to the neotectonic activity of the faults. The characterization (such as fault detailed surface trace, the scarp height, etc.) of these geomorphological markers is currently a time-consuming step with expert-dependent results, often qualitative and with uncertainties that are difficult to estimate. To overcome those issues, we present a proof of concept study for the use of deep learning in morphotectonics, specifically on fault markers. We developed a Bayesian supervised machine learning method using one-dimentional (1D) convolutional neural networks (CNN) trained on a database of simulated topographic profiles across normal fault scarps, called ScarpLearn. From a topographic profile, ScarpLearn is able to automatically give the cumulative scarp height with an uncertainty. We have developed two versions: one designed for more generalized cases involving profiles with multiple fault scarp (ScarpLearn), and another specifically trained to handle profiles featuring a single fault scarp (ScarpLearn_1F). We apply ScarpLearn for the characterization of active normal faults in extensional settings such as the Trans-Mexican Volcanic Belt and Malawi Rift system. From those specific case studies, we explore the progress (computation time, accuracy, uncertainties) that machine learning methods bring to the field of morphotectonics, as well as the current limits (such as bias). Our results show that we are able to develop a CNN model that is estimating scarp heights on topographic profiles from 5m resolution digital elevation model. We compared the results obtained with ScarpLearn and other non deep-learning methods. ScarpLearn achieves similar accuracy while being much faster and having smaller uncertainties. We invite readers to use and to extend our study: codes to build the synthetic scarp database and for the CNN model ScarpLearn are available at: https://gricad-gitlab.univ-grenoble-alpes.fr/poussel/scarplearn.

Production Editor:
Gareth Funning
Handling Editor:
Randy Williams
Copy & Layout Editor:
Théa Ragon

Received:
June 7, 2024
Accepted:
June 30, 2025
Published:
July 16, 2025

1 Introduction

Characterization of geomorphic markers recording fault deformation is crucial to understand past fault activity and future potential impact of earthquakes (i.e., Crone and Haller, 1991; Wells and Coppersmith, 1994; Schlagenhauf et al., 2008). Indeed, this activity is recorded in the landscape leaving a morphological trace documenting the historical physical processes that govern fault rupture (i.e., Zhang et al., 1991; McCalpin and Slemmons, 1998; Kurtz et al., 2018). Among the examples of characterization of faulting geomorphological evidence, the offset's quantification created by ruptures that have reached the surface is a parameter directly used to estimate fault rates, spatial patterns of past ruptures, and slip rates (i.e., Arrowsmith et al., 1998). This information is needed to model the past activity of the fault and estimate the potential hazard for society.

In this study, we focus on normal faults that are often responsible of shallow and destructive earthquakes in numerous inhabited regions of the world (i.e. Central Italy, Wasatch Mountains, Central Mexico). Those faults marks out the landscape through a vertical offsets leaving a typical trace: a scarp (Fig. 1). The scarp is the expression of earthquake in the landscape when the rupture reaches the surface. It is due to the slip along the fault plane that creates a free face which slope is greater than the angle of surrounding hillslopes. This scarp then undergoes erosive processes through times, altering its slope by degrading it (Wallace, 1977; Nash, 1980). Further rupture on the same fault splay may rejuvenate the scarp, which will be affected by erosion once again, altering its shape. Such normal fault scarps have been numerically modeled to characterize and decorrelate the forcing from seismic ruptures and erosional processes (e.g., Avouac and Peltzer, 1993; Hodge et al., 2020; Tucker et al., 2020; Gray et al., 2021; Holtmann et al., 2023). These models focus on the variation of elevation along scarp over time and both models and ob-

*Corresponding author: pousse.lea@ird.fr

servations show that one scarp can also be the sum of various slopes reflecting the complex history of the processes shaping the landscape, both constructive and destructive.

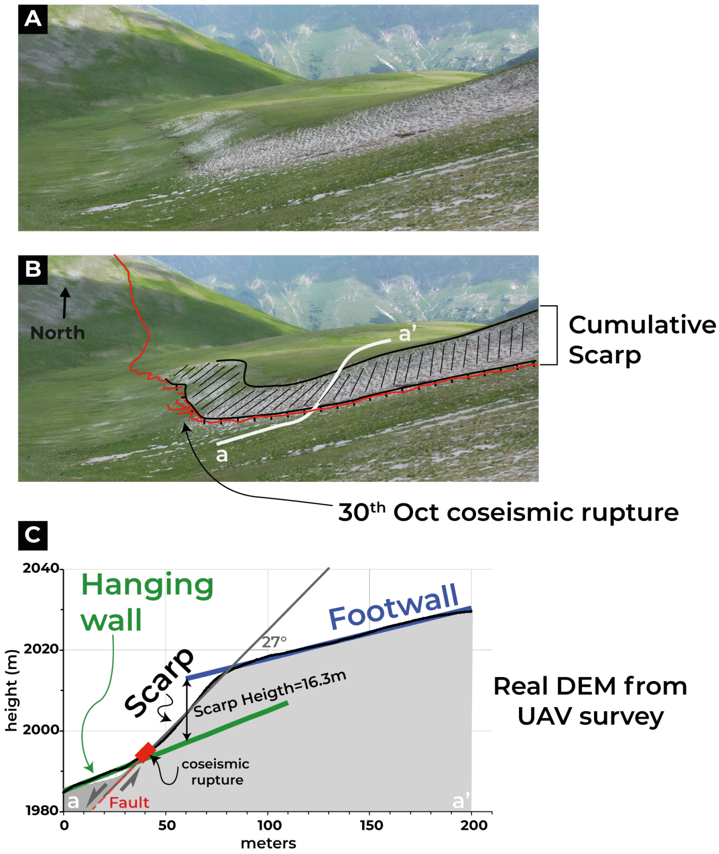

Figure 1 Example of a normal fault scarp in Italy in the Apennines which shows the co-seismic rupture of the 30th October 2016 Norcia earthquake at the base of the cumulative scarp created by previous ruptures (modified from Pousse-Beltran et al., 2022). A) Photo view without interpretation B) with interpretation C) AA' topographic profile across the DEM (Digital Elevation Model) showing footwall and hanging wall (real data).

For tectonic characterization purposes, the morphology of normal fault scarps is mainly analyzed through topographic profiling across the fault. Most studies focus on the vertical offset or the throw or the scarp height that are the most direct parameters signing the cumulative amount of seismic slip. In the very particular context of a flat and horizontal surfaces, the vertical offset, the throw or the scarp height would be the same thing. However, for an inclined surface or a more complex topography with slope breaks there can be quite important differences between these parameters and differences in definition according to authors in the bibliography (details in the following paragraph and in supplementary, Fig. S1). To avoid any confusion, the definitions that we consider for each of them in this study are:

- the vertical separation is the altitude difference between the projection of the planes that fit the upper and lower geomorphic surfaces at a point (x) along the topographic profile (Fig. 12-A), it is not directly related to a fault and allows avoiding implication of tectonics interpretation

- the throw is the vertical component of the slip displacement commonly measured at the base of the fault scarp (Fig. S1-B-C), here it is clearly implicit that the measurement is related to tectonic deformation (e.g., Pucci et al., 2021).

- the scarp height is the altitude difference between the projection of the planes that fit the upper and lower geomorphic surfaces at a point (x) along the topographic profile but need the existence of a fault. It thus corresponds to the the difference in altitude at a specific point (x) between a line that fits the footwall and another line that fits the hanging wall (Fig. S1 -D). The location of this point is detailed below. In some papers the term ”surface offset” or “vertical offset” is used instead (e.g., McCalpin, 2009; Campbell et al., 2015).

- the vertical offset can also be seen as a value comprised between 1) the difference in altitude between the hanging wall and the projection of the hanging wall on the footwall where the inflection due to the scarp begins and 2) the inverse, i.e. the difference in altitude between the footwall and the projection of the footwall on the hanging wall where the inflection due to the scarp begins (see Fig. S1-F).

It should also be noted, that when the scarp does not correspond to the fault free face: the slope of the scarp is not the slope of the fault (Fig. S1-C). Scarp height is more often used because it take into account a variability of the footwall and the hanging wall slopes and thus allows to take into account a more complex geomorphology (e.g., Johnson et al., 2018; Hodge et al., 2019a).

In the following sections of the manuscript we will then focus on the scarp height as it is a parameter signing the cumulative amount of seismic slip (Fig. 1). More in detail, the classic scarp height estimation can be divided into two stages:

- a mapping step which consists of delimiting three portions of the topographic profile that corresponds to the hanging wall, the footwall, and the scarp (Fig. 1). The complexity lies in the possible disturbance of the topography created by erosion, sedimentation, drainage, non-geologic related features (trees, anthropic disturbance, etc.). Also complexity comes from identification and reconstruction of the markers that are offset, that influences their projections to the fault.

- an estimation step where these portions are fitted to three lines, which are used to estimate the scarp height (Fig. 2). However, particular attention must be paid to where the scarp height measurement is performed. Some studies focus on the middle of the scarp (e.g., Johnson et al., 2018); others on the location where the scarp has a maximum slope (e.g., Hodge et al., 2019a; Scott et al., 2022); others

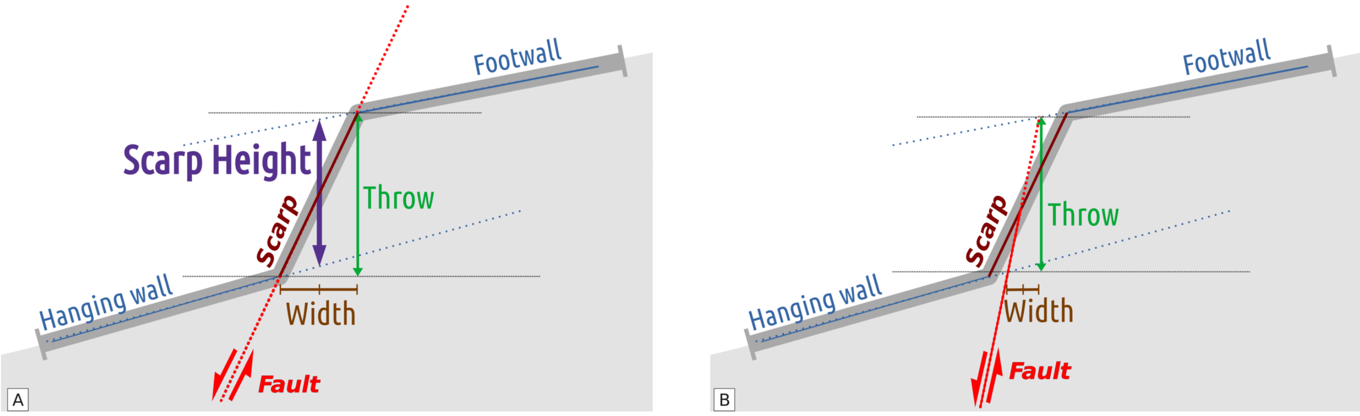

Figure 2 A) Sketch showing the scarp height's definition used in this manuscript. Here the scarp height is measured at the center of the width of the scarp. B) Case where the fault dip is different from the scarp dip.

project the hanging wall (or the footwall) on the inflexion between the footwall (or hanging wall, respectively) and the scarp to bracket the scarp height (see in supplementary the Fig. S1-C-D-E).

To make these measurements, a high-resolution topographic profile is therefore required. In the last 20 years, to get the accurate topographic measurements, researchers used to go into the field to measure Real-Time Kinematic positioning profiles and manually estimate the scarp height (e.g., Mitchell et al., 2001). For the last 10 years, thanks to the remote sensing democratization (drones, access to satellite data), researchers have computed digital elevation models (DEM) that cover several tens to hundreds of kilometres of fault zones at high resolution (<5 m). It therefore became necessary to create the tools to systematize the measurements. In the last 5 years, several research groups have developed methods to estimate the scarp height by empirical, semi-manual or semi automatic approaches (e.g., Stewart et al., 2018; Johnson et al., 2018; Hodge et al., 2019a; Wolfe et al., 2020; Scott et al., 2020; Salomon et al., 2021; Bello et al., 2021; Scott et al., 2022). We can group these approaches into six main categories:

- Manual methods: for each profile, once the portions of hanging wall, footwall and scarp are choose manually, a line is empirically fitted as best as possible through the three portions (identified visually). This manual fitting will exclude non-tectonic perturbations (tree, valleys). A measure of uncertainty can be estimated by identifying the maximum and minimum scarp heights.
- Semi-manual methods such as "Monte Carlo Slip Statistics Toolkit" (MCSST) by Wolfe et al. (2020): inspired by manual methods, here the fit is done by least square optimization. The manual part consists in choosing the limits of the three portions. The uncertainty can be estimated from Monte Carlo simulations which models all possibility by considering the least square fitting uncertainties of each three portions.
- Semi-automatic methods such as Scarp Parameter Algorithm (SPARTA, Hodge et al., 2019a), which needs a manual calibration by pointing manually the portion boundaries on some reference profiles to then automatically estimate the portions on other similar profiles. The topographic portions are finally fitted to lines using least-squares optimization. This method requires the user to choose the filter applied to the topographic profile and associated filter parameters. Moreover, there is no uncertainty estimation.
- Semi-automatic methods such as proposed by Scott et al. (2022), providing both a mapping and a scarp height estimation. This method is semi-automatic as it first requires a manual calibration on a restricted zone of the study area. Once this calibration is done, the algorithm can be run on the whole

Table 1 Overview of published approaches that focus on onshore normal fault scarp.

References	Approach	Fault Detection	Scarp height estimation method	Uncertainties
Classic manual estimation	Manual	No	Empirical	Minimum and Maximum
Wolfe et al. (2020): MCSST	Semi-Manual	No	Least-square	Monte Carlo
Hodge et al. (2019a): SPARTA	Semi-Automatic	No	Least-square	No
Sare et al. (2019)	Automatic	Yes	Template matching but it is not the focus	Overall quality of the fit
Scott et al. (2022)	Semi-Automatic	Yes	Least-square and grid search	Percentile
This study: ScarpLearn	Automatic	No	Convolution Neural Network	Bayesian Inference

study area. The height estimation is obtained by a parameter grid search and by fitting lines to the topographic flats bordering each fault using least-squares optimization. To obtain the uncertainty of the scarp height, the algorithm takes the 16^{th} - 84^{th} percentiles of the heights obtained from satisfactory setting conditions of confidence they have chosen.

- Automatic methods using an analytical solution such as Sare et al. (2019): it recovers the location and amplitude of the scarp through template matching. In this study the aim is mainly to test the detection capability, while the validation of the scarp amplitude estimation is only slightly discussed. The uncertainty in height is not directly estimated, but it is indicated by the overall quality of the fit of the template.

- Automatic method using Linear Discriminant Analysis (LDA) (such as Vega-Ramírez et al., 2021) for offshore settings. This approach consists in a segmentation operation to isolate the objects to classify (here faults). The object are profiles extracted from the bathymetry. Once identified as a fault, the method extracts the semi-scarp height by regressing the profiles against the equation of Nash (1980) as a scarp degradation model (that assumes the fault's vertical offset is accrued in a single rupture event). Although this approach seems to offer an automatic solution, it only focuses on offshore environment having a smoother topography. It seems difficult to transfer this approach for the onshore environment.

Most of these methods have systematized detection and/or height measurements (Tab. 1). However, among those who focus on estimating the scarp height, they are all time-consuming because still at least partly manual, and sometimes even needing a person-dependent calibration step. If this calibration is not frequently performed, the methods can either perform a wrong estimation or not provide any estimation. In other words, manual and semi-automatic methods are prone to operator error, though they retain human oversight, which can be beneficial for aspects of the estimation process. Fully automated methods, while more time-efficient, may introduce artifacts that require a level of critical evaluation to identify and interpret appropriately.

To overcome these issues, machine learning methods and in particular deep learning can represent an interesting solution. Today, artificial intelligence techniques have proven to be efficient in performing many automatic tasks in Geosciences (i.e., Ren et al., 2020), in particular using Convolutional Neural Networks (CNN), a deep learning architecture designed to process images or time series. Specifically in the field of morphotectonics, machine deep learning has only been scarcely used, such as for the automatic mapping of fractures and faults (Mattéo et al., 2021) or to quantify the rock trait distributions of rocky fault scarps (Chen et al., 2023).

Here we propose to automatize the fundamental task of scarp height estimation by evaluating the ability of a supervised CNN (ScarpLearn) trained on realistic synthetic topographic profile catalogs to characterize any normal fault scarp height within a second.

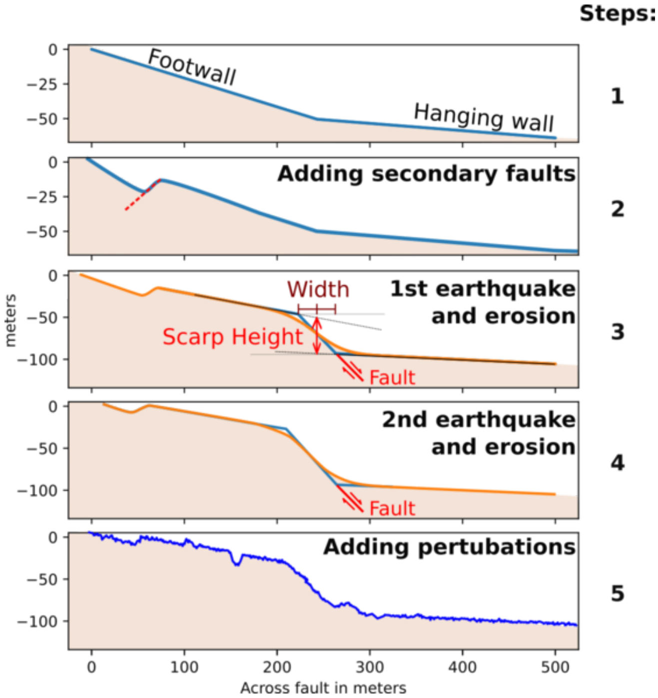

Figure 3 Synthetic normal fault scarp produced by our simulator SimScarp to train the CNN ScarpLearn. The secondary fault (created in Step 2) is subjected to diffusion. Step 4 is repeated as many times as required in order the follow the input parameters (here the total number of earthquakes). The total cumulative scarp height (in meters) is used as the ground truth label by ScarpLearn. At the Step 5, the perturbations are not subjected to diffusion, they are persistent.

2 Scope

The purpose of this investigation is to develop and evaluate an algorithm (ScarpLearn) that automatically estimates the cumulative scarp height for normal faults from a topographic profile with an uncertainty quantification. ScarpLearn targets natural cumulative normal fault scarps, i.e. scarps that may have been created by one or more earthquakes. The results are independent of the user, and thus reproducible with the trained ScarpLearn machine learning model. The profiles, perpendicular to the fault, are first extracted from terrain elevation models. Here ScarpLearn measures the scarp height with an uncertainty localized at the middle of the profile. ScarpLearn is able to ingest topographic profiles disturbed by erosion, drainage, vegetation, and other perturbations. We have developed two versions: ScarpLearn, a generic version for profiles with multiple fault scarps, and ScarpLearn_1F, specifically trained for profiles featuring a single fault scarp.

As there are not enough real data labelled in the literature (i.e. profiles with known ground truth scarp heights) to train the neural network, ScarpLearn is trained on synthetic topographic profiles created by our

Table 2 Parameters chosen from statistical distributions to create topographic profiles in SimScarp. See Supplementary Text S1.1 for more detail.

	Parameters	Distribution	Minimum	Maximum	Mean	Standard Deviation
Regional slopes	Hanging wall slope β_h	Uniform	-5°	10°	/	/
	Footwall slope β_f	Uniform	-5°	10°	/	/
Secondary faults	Number of secondary fault	Uniform	2	2	/	/
	Dip secondary fault δ	Uniform	25°	80°	/	/
	Secondary fault location	Uniform	Borders profile	5% away of the middle of the profile length	/	/
Main fault	Dip main fault δ	Uniform	25°	80°	/	/
	Main fault location	Gaussian	/	/	Middle of the profile	5% of the profile length
	Throw per event	Uniform	0.1 m	5 m	/	/
	Total cumulative throw	Uniform	1 m	50 m	/	/
	Diffusion	Uniform	0.5 m^2/Kyr	10 m^2/Kyr	/	/
	Slip rate	Uniform	0.05 mm/yr	20 mm/yr	/	/
	Minimum number of events	Uniform	1	1	/	/
Perturbations	Gaussian noise	Gaussian	/	/	0	(0.1-1)
	Parabolas A number	Uniform	0	1	/	/
	Parabolas A width	Uniform	0.1	150	/	/
	Parabolas A height	Uniform	-10	10	/	/
	Trees number	Uniform	0	10	/	/
	Trees width	Uniform	0.1	10	/	/
	Trees height	Uniform	1	15	/	/

simulator SimScarp. The chosen characteristics to create the catalog are crucial as it can restrict the scope of ScarpLearn. Synthetic topographic profiles are offset by a fault affecting the profile in its center (range of ±5 %, Fig. 3). This fault can rupture several times creating a cumulative fault scarp. At each inter-seismic period the scarp is subjected to some diffuse erosion, and random perturbations, such as trees, are also added to produce a realistic profile. For the generic version ScarpLearn, several secondary faults are also simulated in order to perturb the profile. Broadly, we are attempting to simulate first order geomorphologic imprints using theoretical knowledge. For example, we have excluded backtilting or rotation of the hanging wall, regolith mobilization, non-colluvial geomorphic processes, pedogenic processes.

In this manuscript we will then validate the algorithm with synthetic data not included in the training set. Then we will apply this algorithm on real cases from Mexico and Malawi in order to test ScarpLearn in real conditions. In addition, we have compared ScarpLearn's results with existing semi-manual and semi-automatic methods: MCSST (Wolfe et al., 2020) and SPARTA (Hodge et al., 2019a) both on synthetic and real data. These methods are selected because they use the same scarp height measurement convention as chosen in this paper (measured in the middle of the scarp width such as in Fig. 2) except for MCSST that measure it to the point of maximum slope on scarp and are representative of existing approaches for comparison (semi-manual and semi-automatic).

3 Methodology

3.1 Synthetics created with numerical model: SimScarp

Convolutional neural networks require a large and various (balanced) dataset for training, this is a challenge for morphotectonic studies because there are not enough real examples of normal fault scarp precisely characterized in the literature. This is due to the time-consuming and difficult task of building such a database (i.e., Nurminen et al., 2022). In fact studies summarizing the characterized normal fault scarps can be incomplete due to sparse measurements. Moreover, the height estimation can never be certain, as there is no cross-validation by multiple experts. In consequence, we have opted to create synthetic catalogs although it implies simplifications of natural processes. For this purpose we have developed a simulator **SimScarp**, which can create topographic profiles of synthetic normal fault scarp with random parameters resulting from robust statistical distributions (Fig. 3). These distributions are designed to reflect realistic morphologies (see Tab. 2, Fig. 3, and Supplementary Fig. S2) but also to represent a wide range of examples, therefore SimScarp is based on a set of parameter values picked from controlled uniform distributions.

For each training set, we can control the length and resolution of the profiles, as well as statistical distributions of the parameters used. For each profile, the simulator SimScarp randomly samples: the diffusion constant, the hanging wall slope, the footwall slope, the

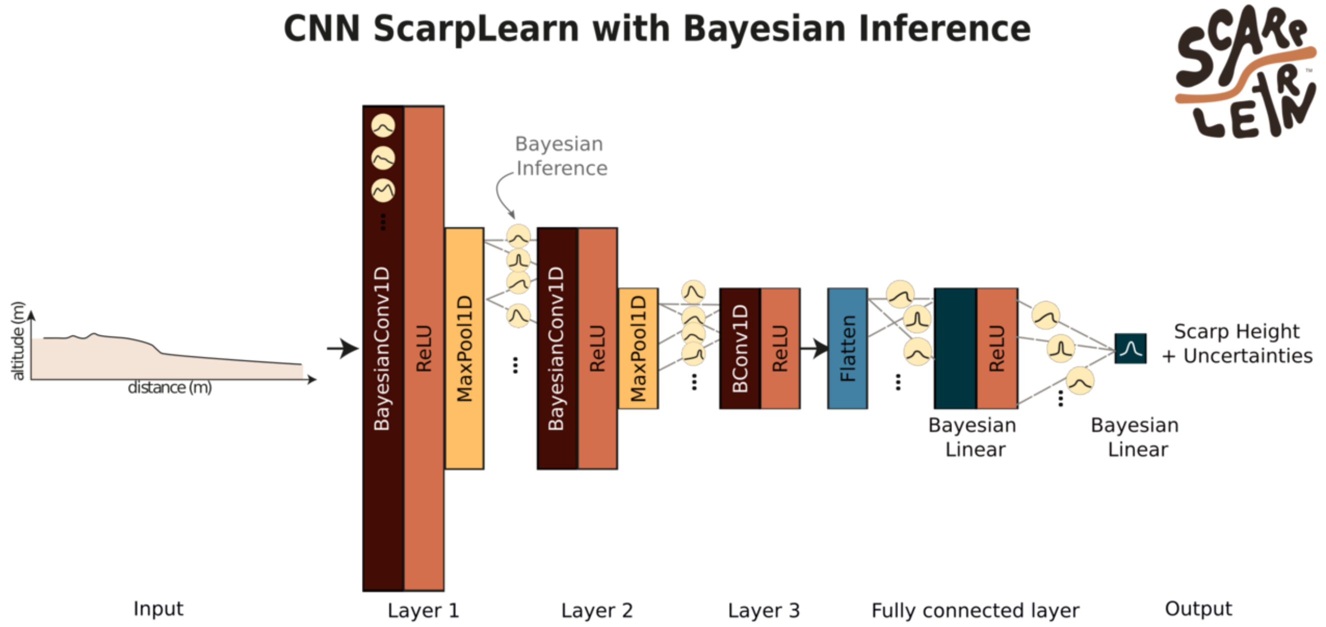

Figure 4 Schematic representation of the pipeline for scarp height characterization: ScarpLearn (1D convolutional neural networks). Between input layer and output layer, there are 3-convolutional layers fully connected layers including an Bayesian inference. The input is a topographic profile across the fault trace. The output of the ScarpLearn is the value of the scarp height with an uncertainties (at 1σ). The link to download ScarpLearn is https://gricad-gitlab.univ-grenoble-alpes.fr/poussel/scarplearn.

number of faults, the fault dip, the fault location, the total cumulative slip, the slip rate, the number of event and some perturbations parameters (see Tab. 2 and Supplementary Text S1.1). Using the slip rate, the total cumulative slip and the number of event, the model recalculates the throw per events and the period between each event. For each event the model creates a scarp at the center of the scarp. Then a diffusive erosion is applied during the inter-event period, following Smith and Bretherton (1972)'s equation simulated as proposed in Nash (1980):

$$\frac{dZ}{dt} = \kappa \frac{d^2 Z}{dx^2} \tag{1}$$

where Z is the elevation, t the time, x the horizontal distance and κ the diffusion constant (m^2/Kyr). We sample the random diffusion constant κ once, as a uniform distribution between 0.5 and 10 m^2/Kyr. This range includes arid conditions (0.5-5 m^2/Kyr) and semi arid to humid temperate condition (up to 10 m^2/Kyr) (Hanks et al., 1984; Andrews and Hanks, 1985; Arrowsmith et al., 1996; Hanks, 2000; Carretier et al., 2002; Kokkalas and Koukouvelas, 2005) . We also allow secondary fault scarps as perturbations both on the hanging wall and footwall (but not in the center), submitted to diffusion as well. The total scarp height S_H is finally calculated as the sum of scarp heights from each event (without taking into account the secondary scarps on the sides).

Lastly, Simscarp adds non-tectonic perturbations at random locations along the profile in order to create a realistic morphologies using random parabolas or steps functions such as in Hodge et al. (2019a) to simulates hills, valleys or trees. More details are provided in the Supplementary Text S1.1 .

We simulate with SimScarp a database of 5000 different topographic profiles with their related scarp height S_H (the label), to be used as training set by the machine learning model ScarpLearn. Each profile is 1km long, with a resolution of 5m (it is a vector of size 200). The total scarp height S_H ranges between 0 and 50 m.

3.2 CNN: ScarpLearn

To learn the scarp height, we designed a 1-dimentional regression convolutional neural network (CNN) with 3 layers called ScarpLearn. This choice is based on the fact that each profile is an ordered vector, similar to a time series, which thus benefits from convolution operations able to extract meaningful features at different scales. Each of the 3 layers is a convolutional layer followed by a pooling layer and a ReLu activation function (Fig. 4). To have an uncertainty (or confidence interval) associated with each profile, crucial for morphotectonic analysis in particular for the scope of probabilistic seismic hazard models, we use variational Bayesian learning. We follow the method *Bayes by Backprop* of Blundell et al. (2015) incorporated in Pytorch by the package Blitz (Esposito, 2020) that allows to assign probability distributions on the weights of a neural network. During its training, the weights of the CNN will be iteratively optimized in order to reduce the error between predicted and real offsets while estimating consistent uncertainties (i.e. confidence interval). The balance between the two factor is adjusted by the complexity cost weight, here that we defined following Shridhar et al. (2019a,b) as a Blundell method.

3.3 Training using synthetic catalogs

We train ScarpLearn on our synthetic set (5000 samples) using a batch gradient descent of 32 samples per batch. For each batch, the model error is calculated using a loss function that is further back-propagated to update all the model parameters in order to minimise the Kullback-Leibler (KL) divergence with the true Bayesian posterior (Blundell et al., 2015). For each prediction, on the batch, we measure the accuracy by simulating the prediction distributions and extract a mean to compare with the correct label. This process is repeated for 300 iterations (i.e. epochs). After each epoch, we estimate the validation error of the validation set. We follow the evolution through the epochs of the ELBO loss which consists of the sum of the KL Divergence of the model with the mean squared error and the accuracy (here the mean absolute error) of the model optimization (Fig. 5). Loss and accuracy curves decline rapidly over epochs, indicating a good convergence of the model. Training ScarpLearn on the synthetic data yields a mean accuracy on the validation set of 3.8 m. The accuracy is here the Mean Absolute Error (MAE), i.e. the average of the absolute differences between the scarp height predicted by ScarpLearn and true the scarp height known for the synthetic dataset. Concerning the confidence interval, 10% (Fig. 5) of the predicted target intervals are integrating the ground truth value. To convert this confidence interval into uncertainties, we thus multiply it by 10 to simulate a 1σ uncertainty.

In addition, we also trained ScarpLearn_1F; another version that is based on a learning database consisting only of 1-fault profiles (see Supplementary Text S1.2). This version can be used when only a single fault trace affects the topography. To ensure this condition, an accurate geomorphological mapping is required. With this version, we reached a Mean Absolute Error on the validation set of 1.6 m.

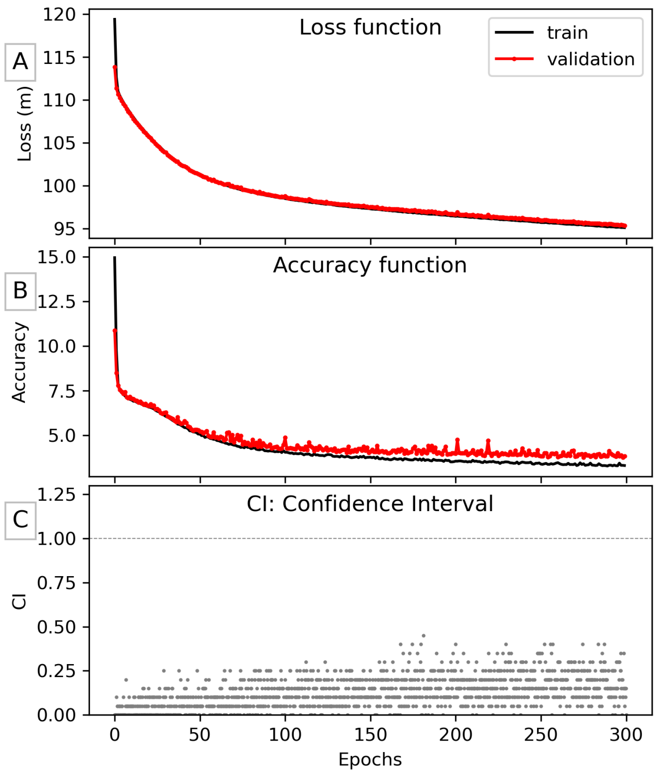

Figure 5 Loss (A) and accuracy (B) function through the epochs for the training and the validation. (C) Confidence Interval range prediction. Those plots show if labels (synthetic ground truths for the validation) fall in the predicted confidence interval for each epochs.

3.4 Application and comparison using synthetic and real study cases

First, using the same synthetic database (with known true scarp heights), we will estimate scarp heights using ScarpLearn, MCSST, and SPARTA. Additionally, we will compare the time required for each method to perform this estimation. Specifically, we ran MCSST on the same profiles, for each profile the user manually selects the boundaries of the hanging wall, footwall, and scarp, we slightly modified MCSST to also estimate scarp height. The MCSST then performs a least-squares fitting for each of the three sections. Then it performs a Monte Carlo simulations (10000 simulations) to generate a distribution of scarp heights. Similarly, SPARTA was applied to the same profiles. To do so, we conducted manual calibration by manually marking the boundaries of portions on reference profiles, allowing SPARTA to automatically estimate the boundaries on all the profiles. Finally, the topographic portions are fitted to lines using least-squares optimization, and only one scarp height estimate is obtained for each profile.

Testing ScarpLearn on real data is more challenging as there will always be unknowns due to the inherent nature of scarp measurement (no ground truth available). We would require a measurement just before and just after an earthquake (in terms of hours), which is an impossible task, especially for cumulative Holocene scarps. InSAR (Interferometric Synthetic Aperture Radar), optical or Lidar (Laser imaging detection and ranging) data before and after an earthquake are currently available with a revisit time of several days at most, and most frequently months. However, these measurements have either low spatial resolutions (>10m) for measurements with small temporal baselines (days, e.g. InSAR) or high spatial resolutions (cm) for measurements but with large temporal baselines (months, e.g. LiDAR), the latter being more likely to have undergone erosion processes.

Since it is not possible to validate with real data, we are limited to compare the results of real samples to existing methods. Performing a test by comparing with other methods is however challenging. Indeed studies on repeatability of geomorphology measurements or mapping show there is always a variability of the result, making it challenging to achieve precise unique validation (i.e., Arrowsmith et al., 2012; Salisbury et al., 2015; Kozacı et al., 2021; Scott et al., 2023). Therefore, scarp height measurements from manual, semi-manual, or semi-automatic methods include simplifications and errors, making them not a ground truth, yet the comparison is crucial for analyzing each method's benefits and limitations.

Table 3 Main results to compare ScarpLearn, MCSST, SPARTA using synthetic datasets (see Fig. 6). The relative error, corresponds to the median of absolute relative errors (see distributions in supplementary Fig. S3). RMSE is the Root Mean Squared Error. NLL is the Negative Log Likelihood, lower NLL is, the better the model fits the data in case of comparing predictions with uncertainties to a truth value. Relative uncertainties are expressed as mean $\pm$ std using 1σ. PICP is the Prediction Interval Coverage Probability, a 100% means that all truth values fall in the prediction interval. * SPARTA does not give results in all cases.

Sets	Metrics	ScarpLearn	ScarpLearn	MCSST	SPARTA*	SPARTA*
Simple Dataset	Number of profiles	100	(on 50)	50	52 over 100	(29 over 50)
	Time to process	<1 min	<1 min	3-4 hours	< 1 hour	
	Mean scarp height	23.3 m	24.6 m	23.8 m	32.3 m	32.5 m
	Mean Absolute error (MAE)	3.9 m	4.8 m	3.2 m	8.5 m	(9.0 m)
	Absolute Relative error (median)	11.3%	9.3%	6.0%	16.4%	19.1%
	RMSE	6.4 m	8.1 m	7.1 m	14.7 m	(15.2 m)
	PICP at 1σ, 2σ, 3σ	75%, 88%, 94%	68%, 84%, 88%	92%, 96%, 96%	-	
	Mean and std of uncertainties (at 1σ)	4.8 ± 1.9 m	4.5 ± 1.9 m	10.7 ± 9.1 m	-	
	NLL	3.9	4.8	3.3	-	
	Relative uncertainties (median)	22 ± 30%	20 ± 12%	31 ± 5559 %	-	
Complex Dataset	Number of profiles	100	(50)	on 50	21 over 100	(12 over 50)
	Time to process	<1 min	<1 min	3-4 hours	< 1 hour	
	Mean scarp height	23.5 m	22.7 m	19.0 m	30.5 m	34.0 m
	Mean Absolute error (MAE)	5.7 m	6.0 m	5.4 m	10.6 m	(13.6 m)
	Absolute Relative error (median)	25.0%	27.4%	15.6%	15.1%	14.8%
	RMSE	7.6 m	8.1 m	7.9 m	18.1 m	(22.5 m)
	PICP at 1σ, 2σ, 3σ	74%, 93%, 97%	68%, 92%, 96%	94%, 100%, 100%	-	
	Mean and std of uncertainties (at 1σ)	9.5 ± 5.0 m	9.5 ± 5.1 m	23.4 ± 19.3 m	-	
	NLL	3.8	3.8	3.7	-	
	Relative uncertainties (median)	36 ± 39 %	36 ± 32 %	100 ± 1101 %	-	

4 Results

4.1 Validation and comparison using synthetic cases

First, to compensate for the lack of ground truth data, we propose to compare scarp heights obtained with MCSST (semi-manual method), SPARTA (semi-automatic method) and ScarpLearn on synthetic tests. We test on two new test sets of synthetic samples of 100 profiles each:

- a simple set, with 1, 2 or 3 faults, with low regional slopes (between -5° and 10°), and few perturbations (see appendix Tab. S1)

- a complex set, also with 1, 2 or 3 faults, but with a wide range of regional slopes (between -10° and 25°), and more perturbations (see appendix Tab. S3)

4.1.1 Validation of ScarpLearn using synthetic cases

We apply ScarpLearn to the two test sets of synthetic data (as for the training set, each profile is 1km long at 5m resolution): the whole inference takes less than 1 minute. By comparing with the ground truth value, ScarpLearn yields a mean absolute error (MAE) of 3.9 m for the simple set and 5.7 m for the complex set. For absolute relative error (median), it yields 11.3 % and 25.0 % ,respectively, for the simple and complex set (Fig. 6-a and Tab. 3 for other metrics). Furthermore we observe that when the predictions are correct, the uncertainty bars are small, while the wrong predictions also show larger estimated uncertainties allowing to encompass the true values (Fig. 6). We obtain 4.8 ± 1.9 m (mean $\pm$ std) of uncertainty (at 1σ) for the simple test set and 9.5 ± 5.0 m (mean $\pm$ std) of uncertainty for the complex test set. The relative uncertainties obtained show a scattered distribution (22 ± 30 % and 36 ± 39 %)

We also analyzed the results by separating the samples containing with only one fault, only two faults, or only three faults (Supplementary Tab. S3, and Figs. S4, S5, 56). ScarpLearn yields, respectively, for the simple setting an MAE of 2.3 m, 3.6 m and 4.4 m. As the number of faults increases, the model becomes less accurate for simple setting. For the complex setting, the MAE not show the same trend as we obtain MAEs of 8.8 m, 5.7 m and 7.6 m.

4.1.2 Validation of ScarpLearn_1F using synthetic cases

Here we used 25 synthetics profiles containing only 1-fault trace (Supplementary Tab. S4 and Fig. S4). For the simple test set, ScarpLearn_1F achieves a mean absolute error (MAE) of 1.3 m and a median abs. relative error of 5.8%. The uncertainties are quantified as 2.9 $\pm$ 0.8 m (mean $\pm$ standard deviation) with a relative uncertainty averaging 16 $\pm$ 23%. For the complex test set, the MAE is 6.0 m, with a median absolute relative error of 31.4%. The uncertainties are measured at 7.4 $\pm$ 3.0 m, and the relative uncertainty averages 38 $\pm$ 206%. Comparing with the ScarpLearn generic model (Supplementary Tab. S4), the generic model does not perform as well on single-scarp (1-fault) profiles as a model that is trained exclusively on them (ScarpLearn_1F).

4.1.3 Evaluation of MCSST on synthetic cases

The semi-manual estimation by MCSST was performed on 50 profiles, and it required 3 to 5 min per profile, so for a fault segment it requires 3 to 4 manpower hours to process them. By comparing with the true values, MCSST yields an MAE of 3.2 m for the simple set and 5.4 m for the complex set. For abs. relative error, it yields 6.0 % and 15.6 % , respectively, for the simple and complex set (Fig. 6-b and Tab. 3 for other metrics). To be noted, fewer samples were processed compared to section 4.1.1, so the MAE cannot be directly compared. We obtain an uncertainty of 10.7 $\pm$ 9.1 m (mean $\pm$ std) (at 1σ) for the simple test set and 23.4 $\pm$ 19.3 m (mean $\pm$ std) for the complex test set. The high standard deviations show how the uncertainties have a scattered distribution.

We analyzed separately the MCSST results of the samples containing only a single fault (25 profiles for each simple and complex sets, see Supplementary Fig. S4 and Tab. S4 for others metrics). MCSST yields an MAE for the simple setting of 1.0 m and 7.2 m for the complex setting.

Regarding the impact of internal operator error in the application of the MCSST method, Wolfe et al. (2020) recognize that identifying fault scarp components (hanging wall, scarp, and footwall) involves human input, which introduces the potential for bias. They emphasize the importance of users verifying the accuracy of each selected component. To mitigate this in real cases, we typically select components while simultaneously verifying the hillshade DEM and the geomorphological mapping. But here, it is not possible because these are synthetic profiles, and we do not have associated DEM or mapping. Therefore, inter-operator comparisons could be interesting for the analysis of the impact of internal operator error.

4.1.4 Evaluation of SPARTA on synthetic cases

In less than an hour, we calibrated and applied the semi-automatic SPARTA method on both simple and complex synthetic sets. However, SPARTA does not provide uncertainties and out of 50 tested profiles, we obtained results only on 29 profiles (for the simple set) and on 12 profiles (for the complex set). By comparing with the true values for the few estimated profiles, SPARTA yielded an MAE of 8.5 m for the simple set and 10.6 m for the complex set. (Fig. 6-c and Tab. 3 for other metrics).

When analyzing the results of SPARTA on the 25 of 1-fault profiles only (see Supplementary Tab. S4 and Fig. S4 and Tab. S4 for others metrics), we obtain results for more profiles: 13 for the simple setting and 10 for the complex setting. We obtained also better MAE for the simple setting. Respectively for the simple and the complex settings, we obtained a MAE of: 6.4 m and 15.4 m. In all synthetic cases with our calibration, SPARTA yields less accurate results than ScarpLearn and MCSST.

4.1.5 Comparison of ScarpLearn, MCSST and SPARTA using synthetic cases

With our calibration, SPARTA was only able to provide results on 20% to 50% of the profiles. Moreover, in all tests, it gives higher mean absolute errors than MCSST and ScarpLearn (Tab. 3). In the synthetic cases with 1, 2 or 3 faults, by comparing MCSST with ScarpLearn on the same 50 profiles, we can observe that both codes give similar accuracy (Tab. 3). Compared to ScarpLearn, we can see that MCSST generally has a better absolute relative error 6.0% (vs 9.3% for ScarpLearn), because there are fewer outliers in the smaller scarps (Fig. 6). For the complex case, where MCSST is also better 15.6% (vs 27.4% for ScarpLearn), because despite the many outliers, there are more cases where MCSST is more accurate. Distribution for MCSST is here more heterogeneous. The main discrepancies come from the uncertainties, which are divided by two for ScarpLearn, but still allowing to reach the true value (the Prediction Interval Coverage Probability (PICP) between 68% and 68% at 1σ).

On the simple data set with only 1 fault (Supplementary Tab. S4), MCSST yields a lower MAE than ScarpLearn. However, ScarpLearn yields higher uncertainties at 1σ (5.4 m instead of 4.3 m). For the complex samples, MCSST and ScarpLearn are very similar (7.1 m for MCSST, 7.9 m for ScarpLearn in mean absolute error), yet the uncertainty of MCSST (15.3 $\pm$ 14.0 m at 1σ) is higher that the one obtained by ScarpLearn (11.4 $\pm$ 3.9 m).

In summary, ScarpLearn is much faster than MCSST with a speed gain factor of 2 orders of magnitude, achieves similar accuracy with a smaller uncertainty. To note, for the simple cases of 1-fault profiles, MCSST performs better. To obtain the better results for these cases with ScarpLearn, we have re-trained ScarpLearn with a learning database consisting only of 1-fault profiles. This new ScarpLearn_1F model gives globally better results in terms of uncertainties than MCSST for the set only capturing 1 fault branch (Supplementary

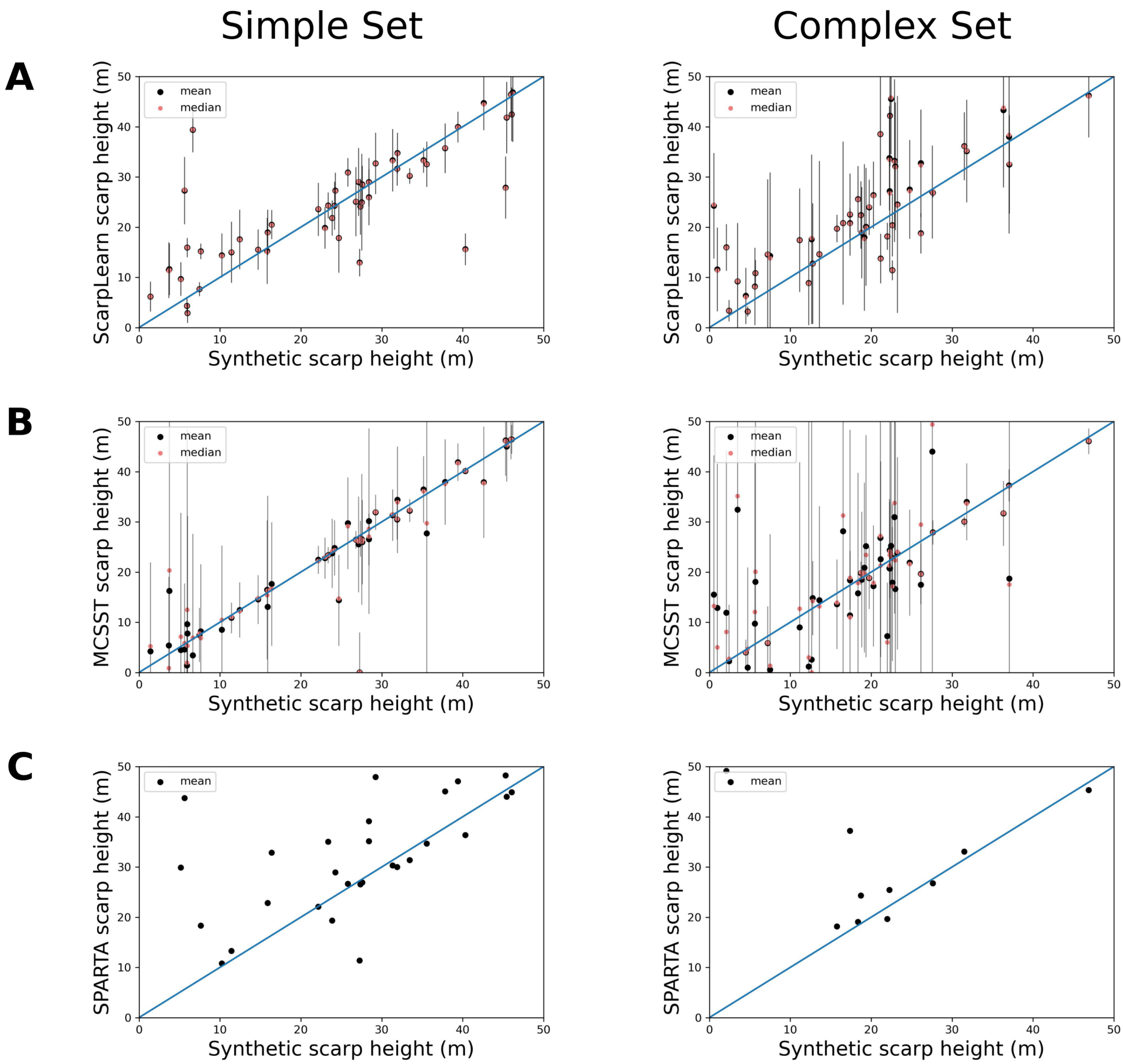

Figure 6 Labels (true values of scarp height) from synthetic dataset versus predictions (ScarpLearn in A, MCSST in B and SPARTA in C) for two set of synthetic datasets. We test on two sets of synthetic profiles : i) a simple set, with 1, 2 or 3 faults, with low regional slopes (between -5° and 10°), and few perturbations (see appendix Tab. S1) and ii) a complex set, also with 1, 2 or 3 faults, but with a wide range of regional slopes (between -10° and 25°), and more perturbations (see appendix Tab. S2). Left plots correspond to the simple setting and the right plots correspond to the complex setting. See Table 3 for comparison metrics. In both setting, we have the possibility of creating profiles with 1, 2 or 3 faults. In A) and B), uncertainty bars show 1σ. SPARTA does not provide uncertainties.

Tab. S4 and Fig. S4). For ScarpLearn_1F we have 5.8% of absolute relative error unlike MCSST where there is 2.1% of abs. relative error for easy samples. On the other hand, ScarpLearn_1F is better for complex samples 31.4% (vs 37.4%). We therefore recommend using ScarpLearn_1F in cases where the user is confident that the profile contains only one fault scarp.

4.2 Application and comparison using real study cases

We will compare the scarp heights obtained with ScarpLearn, MCSST (semi-manual method), SPARTA (semi-automatic method) on 2 real study sites.

We will thus extract topographic profiles perpendicular to the fault in different areas unaffected by significant disturbances not taken into account by ScarpLearn. This means areas that correspond to the conditions in which ScarpLearn has been trained, i.e. areas with :

- no or little anthropogenic infrastructure
- where the scarp is not totally degraded by gravitational erosion

The results of each method are compared and dis-

Table 4 Main results to compare ScarpLearn, MCSST, SPARTA using real fault datasets that corresponds to sampled profiles (see maps in Figs. 7 and 9 and results in Figs 8 and 10). See above Tab. 3 caption for metrics's definitions.

Sets	Metrics	ScarpLearn	ScarpLearn	MCSST	SPARTA*
Ameca Fault Dataset	Number of profiles	117 (all)	98 (where MCSST is)	98	17
	Time to process	<1 min	<1 min	6-8 hours	< 1 hour
	Mean scarp height	8.6 m	8.7 m	8.7 m	6.8 m
	Median scarp height	7.5 m	7.6 m	5.9 m	7.1 m
	Mean of uncertainties (at 1σ)	3.0 m	2.9 m	3.6 m	-
	Absolute difference with respect to ScarpLearn (mean and std)	-	-	2.9 ± 1.8 m	2.3 ± 18 m
	Absolute Relative difference with respect to ScarpLearn (median)	-	-	31 %	24 %
Bilila-Mtakataka Fault Dataset	Number of profiles	161 (all)	161 (where MCSST is)	161	89
	Time to process	<1 min	<1 min	6-8 hours	< 1 hour
	Mean scarp height	-	22.3 m	22.3 m	22.6 m
	Median scarp height	-	21.1 m	21.1 m	24.7 m
	Mean of uncertainties (at 1σ)	-	3.5 m	6.5 m	-
	Absolute difference with respect to ScarpLearn (mean and std)	-	-	6.1 ± 5.9 m	5.7 ± 5.7 m
	Absolute Relative difference with respect to ScarpLearn (median)	-	-	22 %	11 %

cussed.

4.2.1 Case study 1: Ameca Fault, Mexico

The Ameca Fault is located in the Trans-Mexican Volcanic Belt in Mexico (Fig. 7). This region is affected by more than 600 potentially active normal faults yet less than 5% have been correctly characterized by paleoseismological studies (Lacan et al., 2018; Núñez Meneses et al., 2021). In this context, a robust and automatic method to characterize the active normal fault scarp in a global, reproducible, robust (not expert-dependent) quantitative way is very valuable and a great step towards a better characterization of the regional seismic hazard. We focus on Ameca-Ahuisculco fault system (Fig. 7). This fault crosses three distinct geomorphic formations, distinguished by their age. First, there is an active alluvial fan, which is offset by the fault generating scarps of approximately 5 meters height. Further East, there is an older alluvial fan, also offset by the fault forming scarps of approximately 10 to 15 meters height. Finally, the fault crosses the base of the mountain front, marking the boundary between the intrusive basement of the Sierra Ameca and the sedimentary fill of the Ameca basin (Supplementary Fig. S9-A). Here, the cumulative displacement along the fault is estimated to exceed 20 meters. Due to the presence of multiscarps, we extract multiple profiles covering the same areas: in fact, for each parallel scarp the is one profile crossing it at their middle. ScarpLearn estimates the height of the scarp located near the center of the profile. To do so, we used a DEM from the Mexican National Institute of Statistics and Geography (INEGI), resulting from photogrammetric correlation processes to high resolution satellite images captured in stereoscopic mode. From this 5m resolution DEM, we sampled profiles every 100 meters, perpendicular to the Ameca-Ahuisculco fault system (Fig. 7) mapped in Núñez Meneses et al. (2021). These profiles are single pixel derived profiles and each of the 117 profiles is 1 km long.

We use SPARTA, MCSST and ScarpLearn to process these profiles (Figs. 8, S7, S8, and Tab. 4). ScarpLearn and MCSST allow us to obtain results for all profiles, which is not the case with SPARTA (only 17 out of 117). SPARTA with our calibration is less accurate. When we compare MCSST and ScarpLearn, we get similar results (mean height around 9 m), and a t-student test shows that 74% of their results are in agreement (t-student value <1) and only 3% of results are in complete disagreement (t-student value >3) (Fig. 8-E). The results in disagreement are for cases where the scarps are either very small (<1m) or very large (>30m) (Fig. S7-A-B). The differences between the results give a distribution centered around 0 (mean -0.1 $\pm$ 4.4 m (std)), which means that neither MCSST nor ScarpLearn tend to under- or over-estimate the scarp heights relative to each other (Fig. 8-F). The mean absolute difference is 2.9 $\pm$ 1.8 m, but when we look at the cumulative distribution of this difference, it appears that 75% of absolute difference is less than 3.8 m (Fig. 8-G). So there are only strong outliers having large differences. When we look at the distribution of the absolute relative difference, it appears that half of the differences are less than 31% (Fig. S7-G). The uncertainties obtained by MCSST and ScarpLearn are similar (Tab. 4). Their distributions show, however, that MCSST has strong outliers (Fig. S7-C-D) and that ScarpLearn uncertainties tend to increase with the value of the scarp height (Fig. S7-C).

4.2.2 Case study 2: Bilila-Mtakataka Fault, Malawi

The second area studied is in Malawi (Fig. 9), along the Bilila-Mtakataka Fault that is part of the Malawi Rift sys-

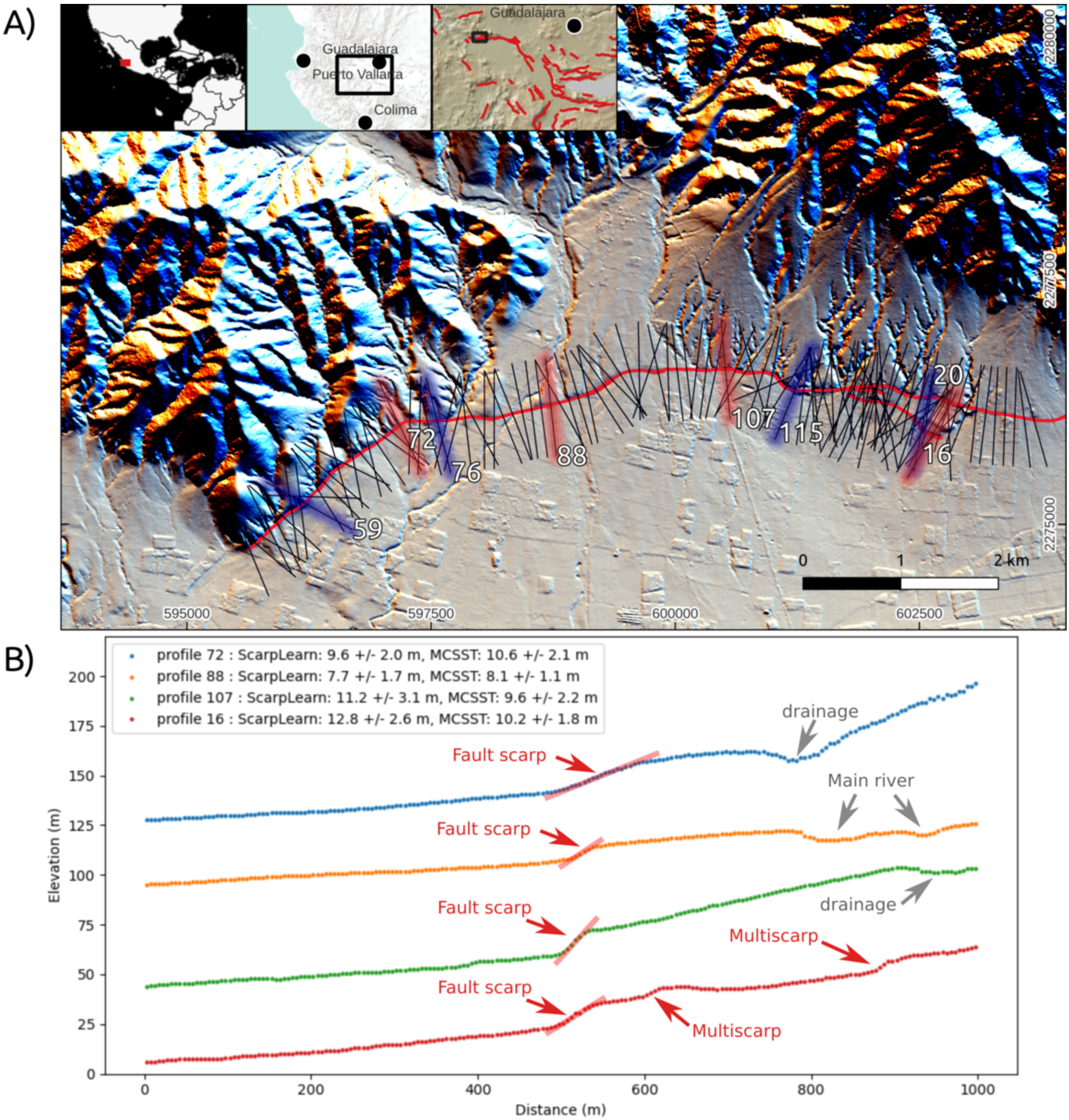

Figure 7 A) Map view of the Ameca fault system in Mexico. Insets show the localization of the studied site. In red, the fault mapped in Núñez Meneses et al. (2021). Black profiles are topographic profiles used for the comparison. Profiles are single pixel derived. Red profiles are plotted in plot B. Blue profiles are plotted in Fig. 11. The DEM is from the Mexican National Institute of Statistics and Geography (INEGI) (see Section "Data and code availability" for more detail). B) Four examples of profiles analyzed. Here the vertical axis values are shifted to provide a better visualization of the profiles.

tem belonging to the East African Rift System (e.g., Jackson and Blenkinsop, 1997). We extracted topographic profiles (single pixel derived) from the 5 meters resolution DEM. Hodge et al. (2019a,b) obtained the DEM by processing bi-stereo Pleiades imagery (50 cm pixel-1) using the Leica Photogrammetry Toolbox within ERDAS Imagine. We focus on the Ngodzi fault segment, here the orientation of the fault scarp follows a zigzag pattern due to the presence of transfer faults. This fault intersects the foliated gneissic bedrock and a Quaternary sedimentary fill (Hodge et al., 2018) (Supplementary Fig. S9-B). Profiles are perpendicular to the fault trace mapped in Hodge et al. (2019a). We extracted 161 profiles of 1 km long of 200 points each (Fig. 9).

We compared SPARTA, MCSST and ScarpLearn on these profiles (Figs. 10, S10, S11 and Tab. 4). ScarpLearn obtains on average a scarp height of 22 m. With SPARTA we obtain 89 results out of 161 on this study

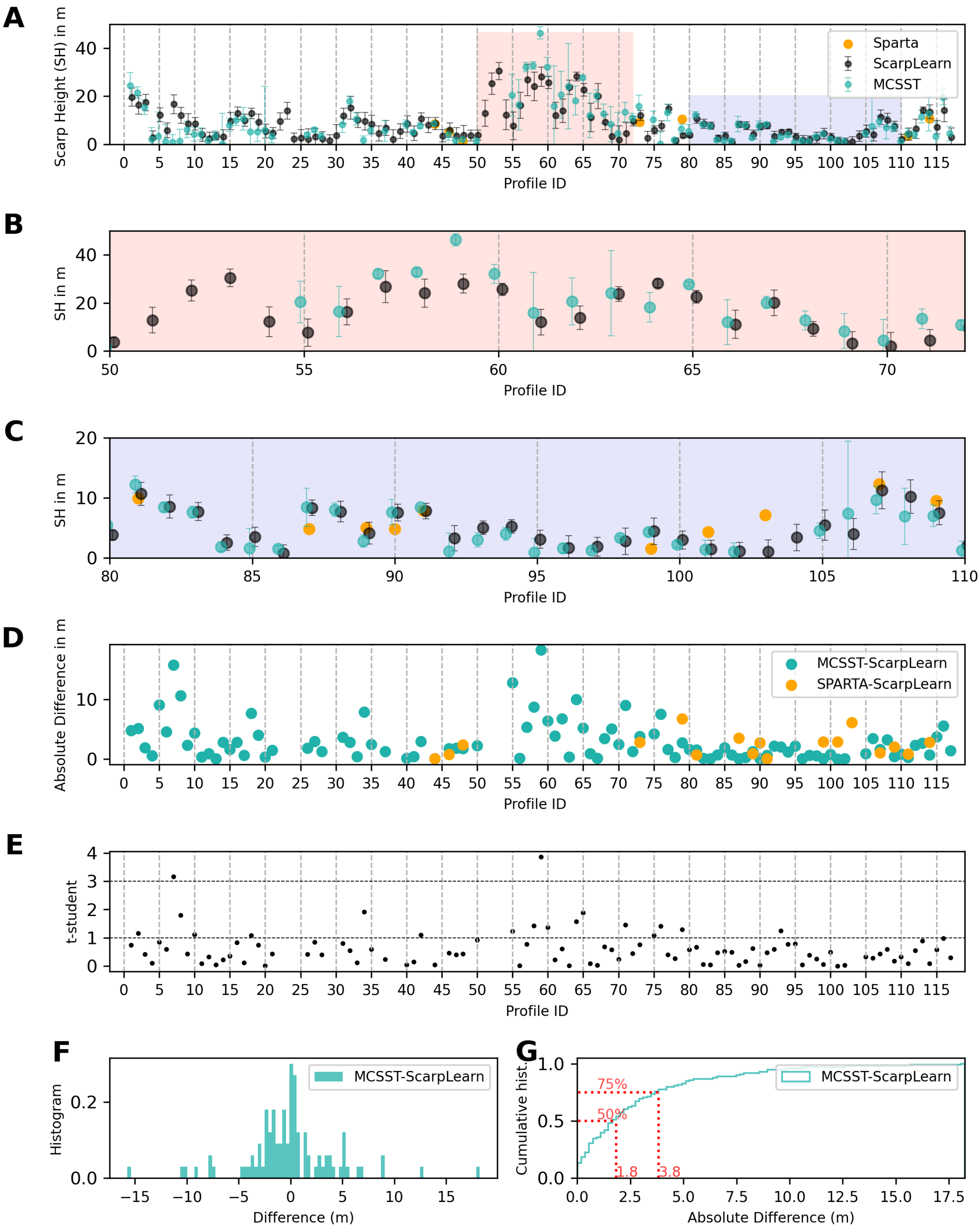

Figure 8 A) Scarp height results obtained for Ameca Fault, using Sparta (orange), ScarpLearn (black) ans MCSST (green) from the profiles sampled across the 5 m DEM (see Fig. 7). The DEM from the Mexican National Institute of Statistics and Geography (INEGI) (see Section "Data and code availability" for more detail). Uncertainty bars represent 1σ. B) Zoom in the pink area from the plot A. C) Zoom in the blue area from the plot A. D) Absolute difference between MCSST and ScarpLearn (in green) and between Sparta and ScarpLearn (in orange). E) T-student test between MCSST and ScarpLearn. Values below 1 mean that the distributions are in agreement. Values between 1 and 3 mean that the distributions are in tension, while values above 3 indicate that the distributions are in disagreement. F) Histogram of the difference between MCSST and ScarpLearn. G) Cumulative histogram of the absolute difference between MCSST and ScarpLearn results. See Tab 4 for a summary of the main results. See in appendix the Fig. S7 and Fig. S8 for more metrics.

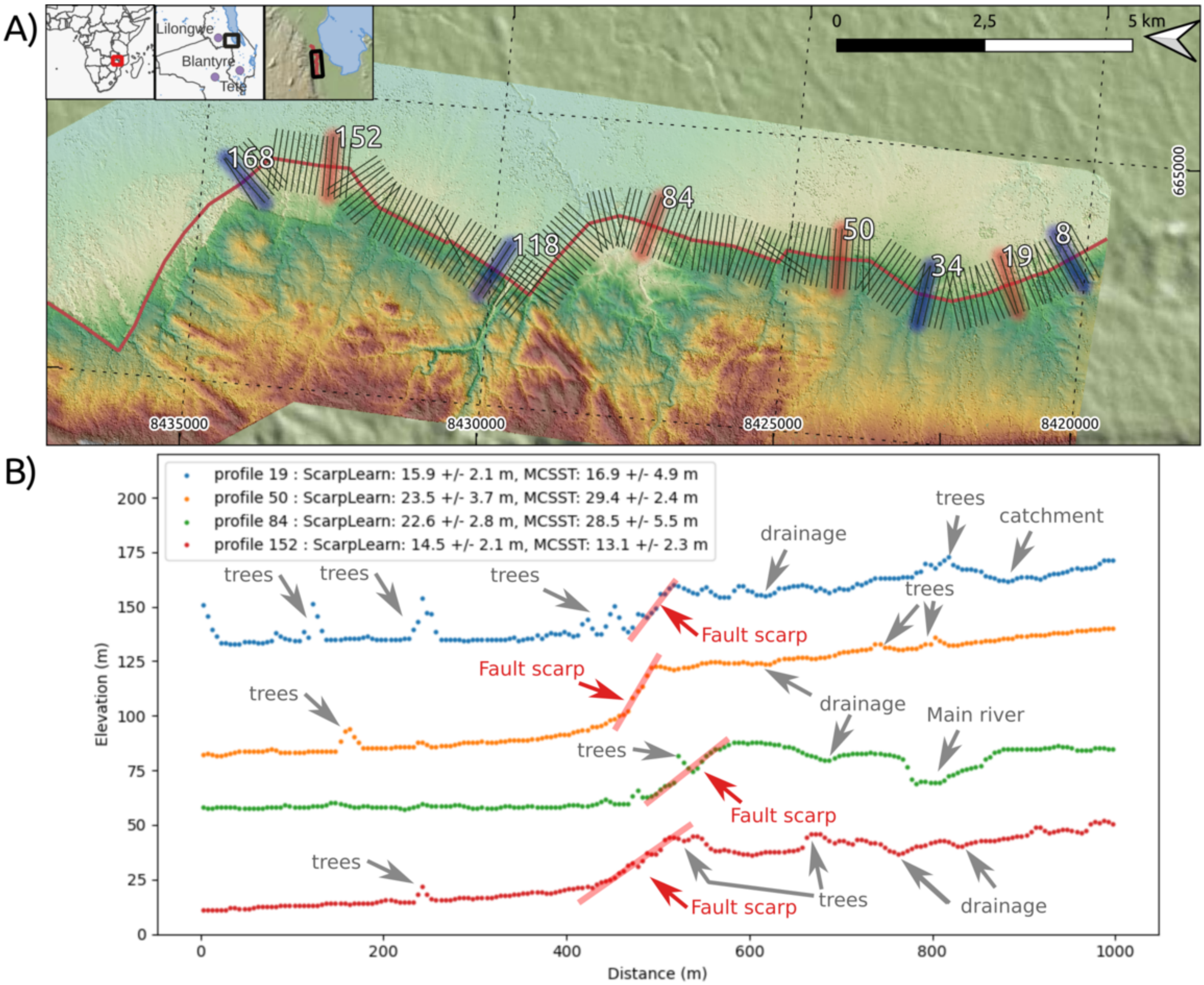

Figure 9 A) Map view of the Bilila-Mtakataka Fault. Insets show the localization of the studied site. In red fault mapped in (Hodge et al., 2019a). Black are topographic profiles used for the comparison. Profiles are single pixel derived. Red profiles are the ones plotted in the plot B (see below). Blue profiles are plotted in Fig. 11. The DEM is from Hodge et al. (2019a,b) (see Section "Data and code availability" for more detail). B) Four examples of profiles analyzed. Here the vertical axis values are shifted to provide a better visualization of the profiles.

site, and when compared with ScarpLearn, the mean absolute difference is 5.7 m. Examining the distribution of relative difference between ScarpLearn, MCSST and SPARTA, ScarpLearn is a middle position, showing -8% (median) compared to MCSST and +7% compared to SPARTA (Supplementary Fig. S10-F). When considering the absolute relative difference with ScarpLearn, half of the profiles show less than 11% with SPARTA, while less than 22% with MCSST (Supplementary Fig. S10-G). When comparing MCSST and ScarpLearn scarp height estimations, the t-student test shows that 59% of results agree, while 6% of results disagree completely (Fig. 10-E). The difference between the results shows a distribution that appears to be symmetrical, although the mean difference of 0.5 ± 8.5 m (std) shows that MCSST gives slightly higher scarp heights than ScarpLearn (Fig. 10-F). The mean absolute difference between MCSST and ScarpLearn is 6.1 ± 5.9 m, and the cumulative absolute difference distribution shows that 50% of results have an absolute difference < 5.0 m (Fig. 10-G). MCSST gives higher uncertainties than ScarpLearn, and is not correlated with scarp height (Supplementary Figure S10 -C-D).

5 Discussion

5.1 Comparison of ScarpLearn, MCSST and SPARTA

In our tests with synthetic data, ScarpLearn yields results comparable to MCSST. Yet, ScarpLearn demonstrates significantly faster processing times ($\sim$ 2 orders of magnitude faster) and provides smaller uncertainties compared to MCSST. Specifically, ScarpLearn appears to be slightly more accurate for in scenarios involving 2 or 3 faults than MCSST. This is because multiscarp cases assign shorter hanging wall and footwall

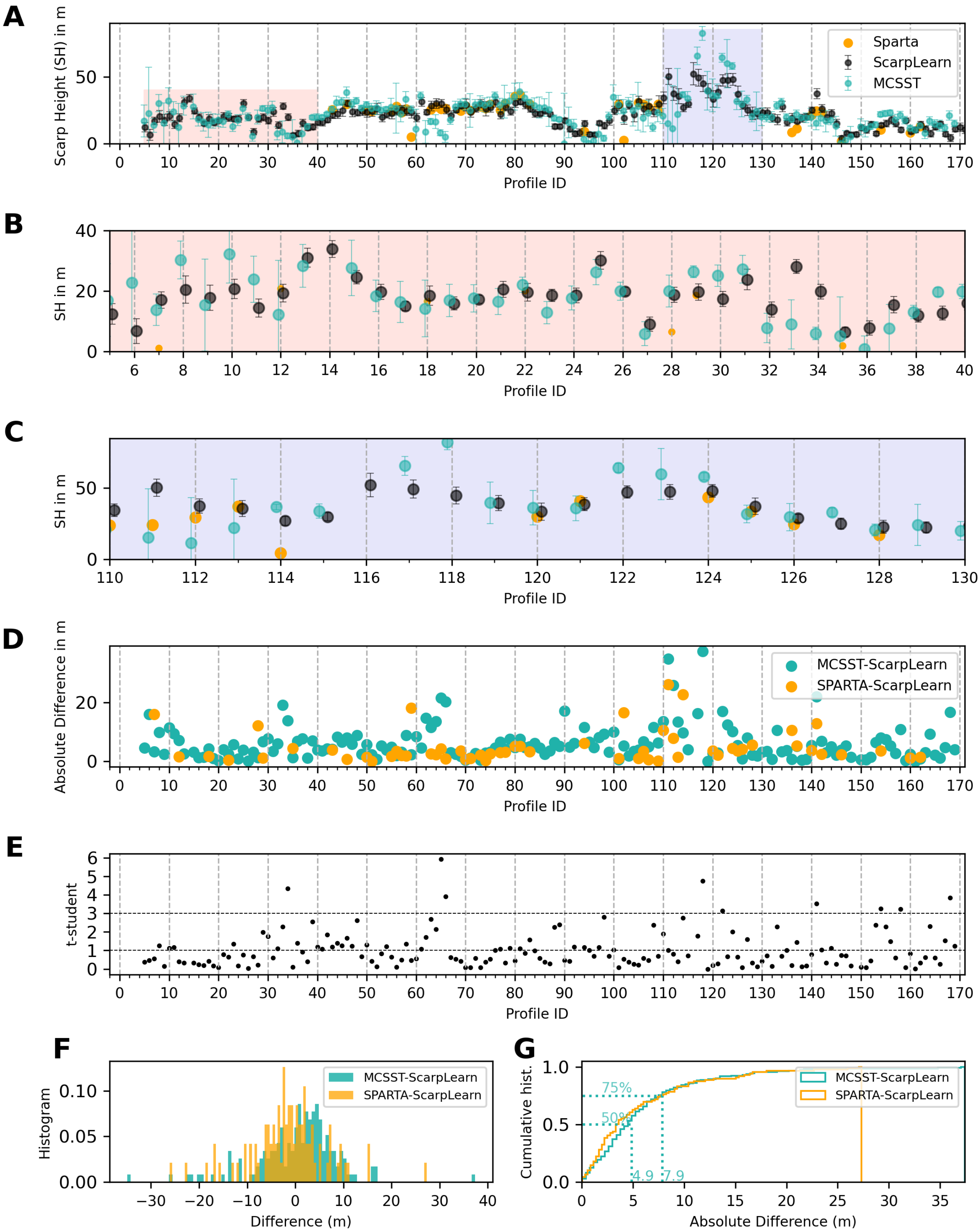

Figure 10 Bilila-Mtakataka Fault results. See legend in Fig. 8 and Table 4 for a summary of the main results. See in Supplementary Material the Fig. S10 and Fig. S11 for more metrics.

surface, which pose challenges for precise fitting in MCSST. Conversely, MCSST is more precise for the 1-fault case, likely due to its effective fit on larger hanging wall and footwall slopes. For this reason, we have trained a specialized version of ScarpLearn just for the 1-fault case, ScarpLearn_1F, giving then better results to MCSST for these cases. ScarpLearn is a thus good alternative that achieves a compromise between rapidity and accuracy while providing uncertainties.

SPARTA was not able to provide an estimation for a majority of profiles, especially from the synthetic test set and from the Ameca fault. This can be explained by the fact that SPARTA is not designed for multiscarp profiles. It can also be explained by the calibration. Indeed, on the Ameca F. site, manual calibration would have to be performed separately for each fault segment, as the profiles cross several geomorphologies (long term, alluvial fans of different ages, etc.). In addition, a generic calibration is impossible on our synthetics, as we randomly parameterize the profiles (slopes, diffusion, dip, etc.). However, on the Bilila-Mtakataka Fault zone, its performance is higher, probably because the code has been designed, tested and published on these data. In this case study, ScarpLearn and SPARTA show less discrepancy than ScarpLearn and MCSST. In the few profiles where SPARTA worked, namely the synthetic profiles with one fault (Supplementary Fig. S4-C) and the real case of the Bilila-Mtakataka Fault Zone (Fig. 10-F), we observe that SPARTA tends to underestimate the values of scarp height. One explanation is that SPARTA uses the point of maximum inflection as the measurement reference rather than the middle of the scarp.

5.2 Perturbation sensitivity

The synthetic database allows us to train ScarpLearn effectively, since in the real cases we obtain similar results than MCSST. However, we can study the sensitivity of MCSST and ScarpLearn to certain types of perturbations by analyzing real-case profiles where MCSST and ScarpLearn give different scarp heights. Among the profiles where the results differ (Fig. 11), we can identify different reasons:

- For cases with many trees, MCSST seems to be perturbed to find the scarp height. This is probably because trees perturb the fit of the hanging wall and footwall, MCSST thus yields large uncertainties (e.g. profile 8 in Fig. 11-B)
- For cases with cumulative long-term scarps (scarp height > 50 m) (e.g. profile 59 in Fig. 11-A or profile 118 in Fig. 11-B), there is often a slope's change in the scarp (see profile 118). The upper part of the scarp is less steep than the lower part. This difference could result from a change in surface processes. A scarp with two slopes seems to pose a problem for ScarpLearn, since it has only learned cases with one slope, and thus seems to only take into account the steeper slope. Moreover, for semi-manual methods (MCSST), it is difficult to know which slopes to take into account (only one slope or both). Here, we have taken the whole scarp (with the two slopes), which explains why MCSST gives higher scarp heights.
- For particular cases, such as flat-bottomed rivers close to the foot of the scarp, they were not included in ScarpLearn (we have only used hyperbola-shaped valleys). This prevents ScarpLearn from differentiating between a flat river-bottom surface and the slope of the hanging wall (see profile 76 in Fig. 11-A and profile 34 in Fig. 11-B).
- Cases where fault mapping is poorly done, such as profile 168 in Fig. 11-B, where the scarp is far from the center of the profile. In this case, ScarpLearn estimates the scarp height at the wrong location.
- Multiscarp cases, as with synthetic data, this configuration makes the fit in MCSST of hanging wall and footwall slopes more complicated (shorter zones) (e.g. profile 20 in Fig. 11-A)
- Cases where erosion is not only due to diffusion, for example the profile 115 in Fig. 11-A which is affected by a landslide exhibits high uncertainty in MCSST. In such instances, landslides may also compromise the performance of ScarpLearn. Furthermore, the Bilila-Mtakataka scarps are bedrock scarps which are subject to gravitational processes involving rock-fracturing. Then in such instances ScarpLearn can also be compromise since the scarp won't diffuse in the way ScarpLearn was trained.

MCSST and ScarpLearn methods are more consistent for Ameca F. study than Bilila-Mtakataka F. study. We explain this because:

- the fault is better mapped in the case of Ameca, in fact in Malawi we used a simplified mapping from a study of a regional scale, whereas in Ameca the mapping was obtained from a local paleoseismological study.
- the presence of trees in Malawi disturbs MCSST, which has difficulties in fitting slopes, while ScarpLearn can probably better filter out high-frequency noise.

5.3 Operator errors

MCSST and SPARTA, being manual and semi-automatic methods, are prone to operator error. Human input is required in SPARTA during calibration, and in MCSST to identify fault scarp components (hanging wall, scarp, and footwall).

For synthetic cases, profiles cannot be cross-checked using the geomorphological mapping, a process that can help reduce operator error in real cases. Consequently, results for synthetic cases with SPARTA and MCSST could greatly benefit from inter-operator comparisons to quantify the operator-induced variability. This would also enhance the precision of comparisons between MCSST, SPARTA, and ScarpLearn.

For real cases, operator error in SPARTA calibration remains significant (see section 5.1), while in MCSST it is mitigated by thorough verification of hillshade DEMs and geomorphological mapping for each profile. This process makes MCSST more time-consuming for real cases than for synthetic ones. Even so, inter-operator comparisons would also provide a more precise estimation of uncertainties for real cases.

Overall, this measurement redundancy would enable us to account for ”operator” uncertainties (epistemic uncertainties) in addition to those directly calculated by MCSST (aleatoric uncertainties) (e.g. Salisbury et al., 2015). However, Bond et al. (2007) and Salisbury et al. (2015) explain that the epistemic uncertainties of experts often are due to confirmation bias. A potential approach to achieve more accurate epistemic uncertainty estimates would be to use blind measurements, as demonstrated by Zielke et al. (2015).

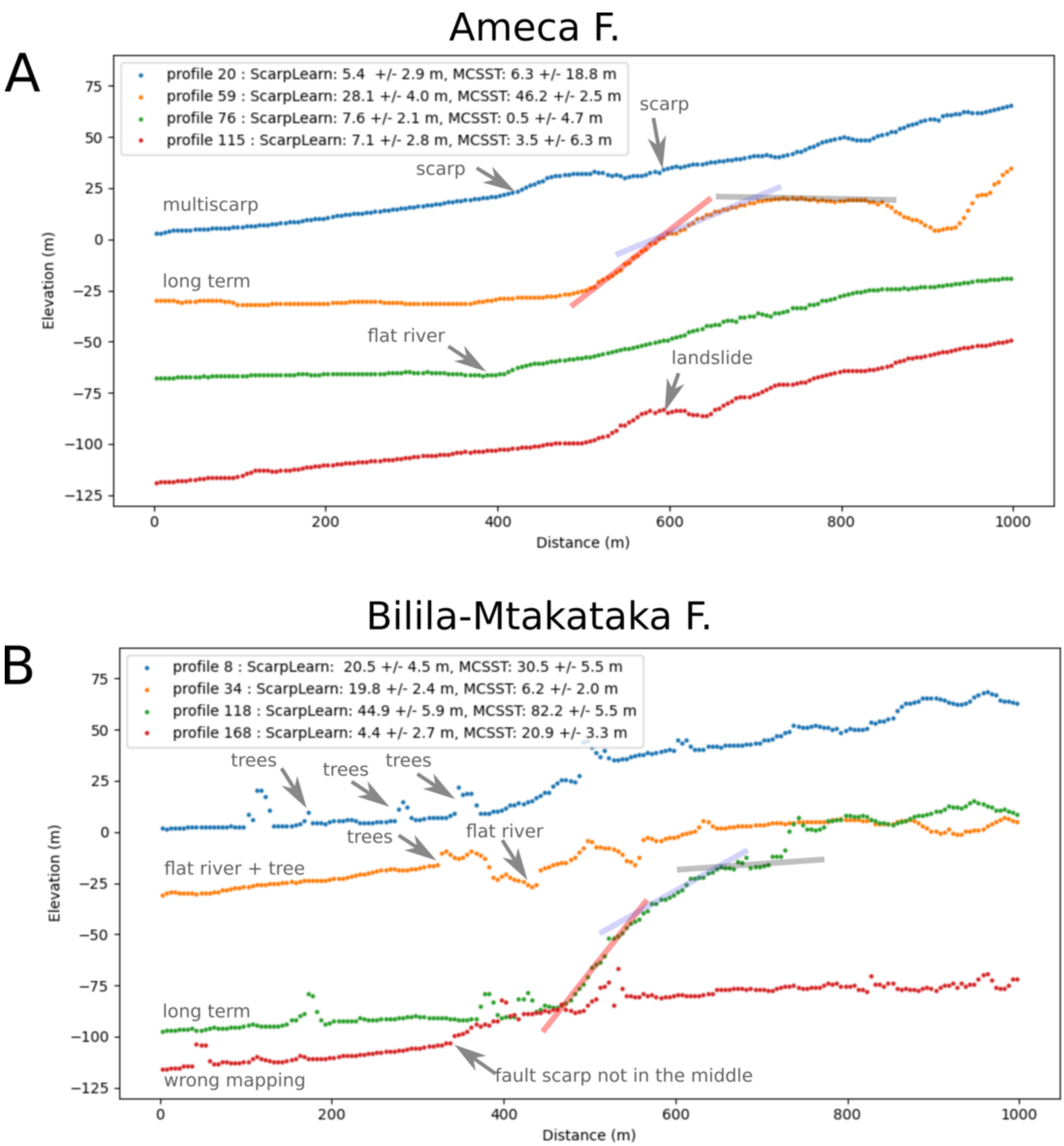

Figure 11 Profiles whose scarp heights are not in agreement between MCSST and ScarpLearn. See profiles localization in Figs. 7 and 9. Profiles 59 and 118 are those that pass through long-term scarps (> 50m). In those profiles, lines are showing the marker ties. Here several interpretations can be made: red for the scarp, gray for the footwall surface and blue for the slope that can be either consider as a scarp that undergone more erosion or either as a footwall. Here the vertical axis values are shifted to provide a better visualization of the profiles.

ScarpLearn is largely exempt from these epistemic biases, as the measurements do not depend on the operator. However, what remains operator-dependent, is the selection of the input profile for ScarpLearn. For example, ScarpLearn measurements can be influenced by the quality of fault mapping (e.g., if the fault is not centered in the profile), which again relies on significant interpretive work by the operator. That said, if outliers, strong uncertainties, or unexpected results are identified, it is straightforward and quick to rerun ScarpLearn with adjusted inputs to obtain new results.

5.4 Scope of ScarpLearn

Using ScarpLearn, for the first time we can calculate the scarp height continuously over the whole fault in just a few seconds, giving us much more information about the fault. ScarpLearn presents thus as a robust alternative; however, it is important to ensure its use under appropriate conditions. In fact, there's a tradeoff between larger number of scarp height measurements and the average noise in the measurements (or just in appropriate measurements). In other words, when conditions deviate from ScarpLearn's training parameters (e.g., the fault is not centered in the profile or there is the presence of a landslide), the measurements are likely to be less accurate or show a high degree of uncertainty. ScarpLearn will always provide a result value even when the profile has too much noise, whereas a geomorphologist can evaluate when to discard a result if it seems unreliable. Nonetheless, ScarpLearn can be useful even in less suitable conditions. It then requires a following step of critical evaluation and numerous measurements. In such cases, the ScarpLearn results could be used for a preliminary, rapid assessment of an area to identify potential targets for more detailed work. If the user want to use ScarpLearn robustly, it is important to ensure its usage good operating conditions. To ascertain these conditions, meticulous expert mapping is required. This mapping should encompass fault traces, flat river areas, landslide contours, and other potential scenarios to verify under which conditions ScarpLearn can be used. Operators cannot work blindly but need to understand the context. In fact this was also true for any previous methodology (MCSST, SPARTA, etc...). In the future, it will be interesting to complete the learning database, either with real cases, or with more complex processes that will enable ScarpLearn to be effective on more various scenarios.

6 Conclusion

We have developed a machine learning model called ScarpLearn capable of estimating the cumulative scarp height of normal faults as well as estimating its uncertainty based on 1-dimensional topographic profiles (extracted from Digital Elevation Models). Training with synthetic data has enabled us to obtain a efficient CNN model that can be applied to a variety of real datasets (here on case study DEMs of 5m resolution in Mexico and Malawi). In our tests with synthetic data, ScarpLearn gives similar results as existing semi-manual methodology (MCSST). On the other hand, ScarpLearn is two order of magnitudes faster and achieves smaller uncertainties. The same applies to real data: ScarpLearn is comparable to semi-manual method and only disagrees on less than 6% of the cases, completely agrees on 74% of the cases in Mexico and 59% in Malawi, leaving the rest of the cases (33% and 35% respectively) in ambiguity. Although the distribution of residuals is centered around zero, there are complicated cases where the ScarpLearn differs from the MCSST. It's reflecting the fact that ScarpLearn has been trained by synthetic data that does not take into account some complex field configurations: long term cumulative scarp (with diffusion rates variations, flat rivers, etc). Although ScarpLearn is automatic, it is still necessary to have an expert overview on the fault mapping, the geomorphological mapping and on the local climatic and topographic context in order to verify if ScarpLearn can be applied or not, depending on the fault scarp training model. Nonetheless, once these conditions are fulfilled, ScarpLearn allows to: 1) gain a considerable expert time (few minutes instead of multiple hours), 2) quickly scan an area and identify targets for more detailed work, 3) obtain reproducible results not user-dependant, and 4) obtain high resolution estimations with realistic uncertainties. This provides therefore a reliable method to perform fault scarp analysis, to be developed for strike-slip or reverse faults as well.

Acknowledgements

We acknowledge the reviewers Ramon Arrowsmith and an anonymous reviewer for their thoughtful and constructive comments. We acknowledge also the editor Randy Williams. This work benefited from government aid managed by the Agence Nationale de la Recherche under the France 2030 programme, reference ANR-23-IACL-0006. This research was partially supported by MIAI@Grenoble Alpes (ANR-19-P3IA-0003). This research was also partially supported by ISTerre (BQR intern call). Thanks to GRICAD infrastructure (gricad.univ-grenoble-alpes.fr), which is supported by the Grenoble research communities, for the computations. ISTerre is part of Labex OSUG@2020 (Investissements d'avenir – ANR10 LABX56). We thank the CNES R&T Call 2022 "Hybridation des donnees" N°34500075632 and the call PNTS program of INSU CNRS to award Léa Pousse. We also acknowledge the PAPIIT grant IN108220 and IG101823 awarded to Pierre Lacan. Partial support was received from the France-Mexico collaborative project SEP-CONACYT-ANUIES-ECOS N°321193 and the IGCP-669 Ollin Project of UNESCO-IUGS. This work was supported by the Universidad Nacional Autónoma de México under PASPA – DGAPA grant (P. Lacan's academic stay). We thank the CNES for providing high-resolution optical images. Access to topographic data was granted through the DINAMIS program (https://dinamis.teledetection.fr/). This work is based on data services provided by the OpenTopography Facility with support from the National Science Foundation under NSF Award Numbers 1948997, 1948994 & 1948857. The data corresponds to the point cloud for the Bilila-Mtakataka Fault and Mua Segment from Hodge et al. (2019a,b, see https://portal.opentopography.org/dataspace/dataset?opentopoID=OTDS.062019.32736.2).
We acknowledge INEGI (Mexican National Institute of Statistics and Geography) for providing DEM.

Data and code availability

The codes developed and data sets used in this manuscript are available at https://gricad-gitlab.univ-grenoble-alpes.fr/poussel/scarplearn . The data corresponding to the point cloud for the Bilila-Mtakataka Fault and Mua Segment comes from Hodge et al. (2019a,b, see https://portal.opentopography.org/dataspace/dataset?opentopoID=OTDS.062019.32736.2), obtained by the processing of bi-stereo Pleiades imagery (50 cm pixel-1). The DEM covering the Ameca fault system comes from the Mexican National Institute of Statistics and Geography (INEGI), and results from photogrammetric correlation processes to high resolution satellite images captured in stereoscopic mode, see the INEGI Elevation Maps of Mexico webpage https://www.inegi.org.mx/temas/relieve/continental/.

Competing interests

The author declares that there is no conflict of interest.

References

Andrews, D. J. and Hanks, T. C. Scarp Degraded by Linear Diffusion: Inverse Solution for Age. *Journal of Geophysical Research: Solid Earth*, 90(B12):10193–10208, Oct. 1985. doi: 10.1029/JB090iB12p10193.

Arrowsmith, J. R., Pollard, D. D., and Rhodes, D. D. Hillslope Development in Areas of Active Tectonics. *Journal of Geophysical Research: Solid Earth*, 101(B3):6255–6275, Mar. 1996. doi: 10.1029/95JB02583.

Arrowsmith, J. R., Rhodes, D. D., and Pollard, D. D. Morphologic Dating of Scarps Formed by Repeated Slip Events along the San Andreas Fault, Carrizo Plain, California. *Journal of Geophysical Research: Solid Earth*, 103(B5):10141–10160, 1998. doi: 10.1029/98JB00505.

Arrowsmith, J. R., Rockwell, T. K., Madugo, K. S., and Zielke, O. Repeatability, Accuracy, and Precision of Surface Slip Measurements from High-Resolution Topographic Data: Collaborative Research with Arizona State University and San Diego State University. *Final Technical Report US Geological Survey National Earthquake Hazards Reduction External Research Program*, 2012.

Avouac, J.-P. and Peltzer, G. Active Tectonics in Southern Xinjiang, China: Analysis of Terrace Riser and Normal Fault Scarp Degradation along the Hotan-Qira Fault System. *Journal of Geophysical Research: Solid Earth*, 98(B12):21773–21807, Dec. 1993. doi: 10.1029/93JB02172.

Bello, S., Scott, C. P., Ferrarini, F., Brozzetti, F., Scott, T., Cirillo, D., de Nardis, R., Arrowsmith, J. R., and Lavecchia, G. High-Resolution Surface Faulting from the 1983 Idaho Lost River Fault Mw 6.9 Earthquake and Previous Events. *Scientific Data*, 8(1):68, Feb. 2021. doi: 10.1038/s41597-021-00838-6.

Blundell, C., Cornebise, J., Kavukcuoglu, K., and Wierstra, D. Weight Uncertainty in Neural Network. In *International Conference on Machine Learning*, pages 1613–1622. PMLR, 2015.

Bond, C., Gibbs, A., Shipton, Z., and Jones, S. What Do You Think This Is? “Conceptual Uncertainty” in Geoscience Interpretation. *GSA Today*, 17(11):4, 2007. doi: 10.1130/GSAT01711A.1.

Campbell, G. E., Walker, R. T., Abdrakhmatov, K., Jackson, J., Elliott, J. R., Mackenzie, D., Middleton, T., and Schwenninger, J.-L. Great Earthquakes in Low Strain Rate Continental Interiors: An Example from SE Kazakhstan. *Journal of Geophysical Research: Solid Earth*, 120(8):5507–5534, Aug. 2015. doi: 10.1002/2015JB011925.

Carretier, S., Ritz, J.-F., Jackson, J., and Bayasgalan, A. Morphological Dating of Cumulative Reverse Fault Scarps: Examples from the Gurvan Bogd Fault System, Mongolia: Morphological Dating of Cumulative Reverse Fault Scarps. *Geophysical Journal International*, 148(2):256–277, Feb. 2002. doi: 10.1046/j.1365-246X.2002.01007.x.

Chen, Z., Scott, C., Keating, D., Clarke, A., Das, J., and Arrowsmith, R. Quantifying and Analysing Rock Trait Distributions of Rocky Fault Scarps Using Deep Learning. *Earth Surface Processes and Landforms*, 48(6):1234–1250, 2023. doi: 10.1002/esp.5545.

Crone, A. J. and Haller, K. M. Segmentation and the Coseismic Behavior of Basin and Range Normal Faults: Examples from East-Central Idaho and Southwestern Montana, U.S.A. *Journal of Structural Geology*, 13(2):151–164, Jan. 1991. doi: 10.1016/0191-8141(91)90063-O.

Esposito, P. BLiTZ - Bayesian Layers in Torch Zoo (a Bayesian Deep Learing Library for Torch), 2020.

Gray, H., DuRoss, C., Nicovich, S., and Gold, R. A Geomorphic-Process-Based Cellular Automata Model of Colluvial Wedge Morphology and Stratigraphy. *Earth Surface Dynamics Discussions*, pages 1–34, Oct. 2021. doi: 10.5194/esurf-2021-70.

Hanks, T. C. The Age of Scarplike Landforms from Diffusion-Equation Analysis. *Quaternary Geochronology: Methods and Applications*, 4:313–338, 2000. doi: 10.1029/rf004p0313.

Hanks, T. C., Bucknam, R. C., Lajoie, K. R., and Wallace, R. E. Modification of Wave-cut and Faulting-controlled Landforms. *Journal of Geophysical Research: Solid Earth*, 89(B7):5771–5790, July 1984. doi: 10.1029/JB089iB07p05771.

Hodge, M., Fagereng, Å., Biggs, J., and Mdala, H. Controls on Early-Rift Geometry: New Perspectives From the Bilila-Mtakataka Fault, Malawi. *Geophysical Research Letters*, 45(9):3896–3905, 2018. doi: 10.1029/2018GL077343.

Hodge, M., Biggs, J., Fagereng, Å., Elliott, A., Mdala, H., and Mphepo, F. A Semi-Automated Algorithm to Quantify Scarp Morphology (SPARTA): Application to Normal Faults in Southern Malawi. *Solid Earth*, 10(1):27–57, Jan. 2019a. doi: 10.5194/se-10-27-2019.

Hodge, M., Biggs, J., Fagereng, Å., and Wedmore, L. Bilila-Mtakataka Fault - Mua Segment. *Distributed by OpenTopography*, 2019b. doi: 10.5069/G92R3PSV.

Hodge, M., Biggs, J., Fagereng, Å., Mdala, H., Wedmore, L. N. J., and Williams, J. N. Evidence From High-Resolution Topography for Multiple Earthquakes on High Slip-to-Length Fault Scarps: The Bilila-Mtakataka Fault, Malawi. *Tectonics*, 39(2): e2019TC005933, 2020. doi: 10.1029/2019TC005933.

Holtmann, R., Cattin, R., Simoes, M., and Steer, P. Revealing the Hidden Signature of Fault Slip History in the Morphology of Degrading Scarps. *Scientific Reports*, 13(1):3856, Mar. 2023. doi: 10.1038/s41598-023-30772-z.

Jackson, J. and Blenkinsop, T. The Bilila-Mtakataka Fault in Malaŵi: An Active, 100-Km Long, Normal Fault Segment in Thick Seismogenic Crust. *Tectonics*, 16(1):137–150, 1997. doi: 10.1029/96TC02494.

Johnson, K. L., Nissen, E., and Lajoie, L. Surface Rupture Morphology and Vertical Slip Distribution of the 1959 Mw 7.2 Hebgen Lake (Montana) Earthquake From Airborne Lidar Topography. *Journal of Geophysical Research: Solid Earth*, 123(9):8229–8248, 2018. doi: 10.1029/2017JB015039.

Kokkalas, S. and Koukouvelas, I. Fault-Scarp Degradation Modeling in Central Greece: The Kaparelli and Eliki Faults (Gulf of Corinth) as a Case Study. *Journal of Geodynamics*, 40(2-3): 200–215, Sept. 2005. doi: 10.1016/j.jog.2005.07.006.

Kozacı, Ö., Madugo, C. M., Bachhuber, J. L., Hitchcock, C. S., Kottke, A. R., Higgins, K., Wade, A., and Rittenour, T. Rapid Postearthquake Field Reconnaissance, Paleoseismic Trenching, and GIS-Based Fault Slip Variability Measurements along the Mw 6.4 and Mw 7.1 Ridgecrest Earthquake Sequence, Southern California. *Bulletin of the Seismological Society of America*, 111 (5):2334–2357, Oct. 2021. doi: 10.1785/0120200262.

Kurtz, R., Klinger, Y., Ferry, M., and Ritz, J. F. Horizontal Surface-Slip Distribution through Several Seismic Cycles: The Eastern Bogd Fault, Gobi-Altai, Mongolia. *Tectonophysics*, 734–735:167–182, June 2018. doi: 10.1016/j.tecto.2018.03.011.

Lacan, P., Ortuño, M., Audin, L., Perea, H., Baize, S., Aguirre-Díaz, G., and Zúñiga, F. R. Sedimentary Evidence of Historical and Prehistorical Earthquakes along the Venta de Bravo Fault System, Acambay Graben (Central Mexico). *Sedimentary Geology*, 365:62–77, Mar. 2018. doi: 10.1016/j.sedgeo.2017.12.008.

Mattéo, L., Manighetti, I., Tarabalka, Y., Gaucel, J.-M., van den Ende, M., Mercier, A., Tasar, O., Girard, N., Leclerc, F., Giampetro, T., Dominguez, S., and Malavieille, J. Automatic Fault Mapping in Remote Optical Images and Topographic Data With Deep Learning. *Journal of Geophysical Research: Solid Earth*, 126(4), Apr. 2021. doi: 10.1029/2020JB021269.

McCalpin, J. P. *Paleoseismology*, volume 95. Academic press, 2009.

McCalpin, J. P. and Slemmons, D. B. *Statistics of Paleoseismic Data*. The Company, 1998.

Mitchell, S. G., Matmon, A., Bierman, P. R., Enzel, Y., Caffee, M., and Rizzo, D. Displacement History of a Limestone Normal Fault Scarp, Northern Israel, from Cosmogenic 36Cl. *Journal of Geophysical Research: Solid Earth*, 106(B3):4247–4264, 2001. doi: 10.1029/2000JB900373.

Nash, D. B. Morphologic Dating of Degraded Normal Fault Scarps. *The Journal of Geology*, 88(3):353–360, May 1980. doi: 10.1086/628513.

Núñez Meneses, A., Lacan, P., Zúñiga, F. R., Audin, L., Ortuño, M., Rosas Elguera, J., León-Loya, R., and Márquez, V. First Paleoseismological Results in the Epicentral Area of the Sixteenth Century Ameca Earthquake, Jalisco – México. *Journal of South American Earth Sciences*, 107:103121, Apr. 2021. doi: 10.1016/j.jsames.2020.103121.

Nurminen, F., Baize, S., Boncio, P., Blumetti, A. M., Cinti, F. R., Civico, R., and Guerrieri, L. SURE 2.0 – New Release of the Worldwide Database of Surface Ruptures for Fault Displacement Hazard Analyses. *Scientific Data*, 9(1):729, Nov. 2022. doi: 10.1038/s41597-022-01835-z.

Pousse-Beltran, L., Benedetti, L., Fleury, J., Boncio, P., Guillou, V., Pace, B., Rizza, M., Puliti, I., and Socquet, A. 36Cl Exposure Dating of Glacial Features to Constrain the Slip Rate along the Mt. Vettore Fault (Central Apennines, Italy). *Geomorphology*, page 108302, May 2022. doi: 10.1016/j.geomorph.2022.108302.

Pucci, S., Pizzimenti, L., Civico, R., Villani, F., Brunori, C. A., and Pantosti, D. High Resolution Morphometric Analysis of the Cordone Del Vettore Normal Fault Scarp (2016 Central Italy Seismic Sequence): Insights into Age, Earthquake Recurrence and Throw Rates. *Geomorphology*, 388:107784, Sept. 2021. doi: 10.1016/j.geomorph.2021.107784.

Ren, C. X., Hulbert, C., Johnson, P. A., and Rouet-Leduc, B. Chapter Two - Machine Learning and Fault Rupture: A Review. In Moseley, B. and Krischer, L., editors, *Advances in Geophysics*, volume 61 of *Machine Learning in Geosciences*, pages 57–107. Elsevier, Jan. 2020. doi: 10.1016/bs.agph.2020.08.003.

Salisbury, J. B., Haddad, D., Rockwell, T., Arrowsmith, J. R., Madugo, C., Zielke, O., and Scharer, K. Validation of Meter-Scale Surface Faulting Offset Measurements from High-Resolution Topographic Data. *Geosphere*, 11(6):1884–1901, Dec. 2015. doi: 10.1130/GES01197.1.

Salomon, G. W., New, T., Muir, R. A., Whitehead, B., Scheiber-Enslin, S., Smit, J., Stevens, V., Kahle, B., Kahle, R., Eckardt, F. D., and Alastair Sloan, R. Geomorphological and Geophysical Analyses of the Hebron Fault, SW Namibia: Implications for Stable Continental Region Seismic Hazard. *Geophysical Journal International*, 229(1):235–254, Dec. 2021. doi: 10.1093/gji/ggab466.

Sare, R., Hilley, G. E., and DeLong, S. B. Regional-Scale Detection of Fault Scarps and Other Tectonic Landforms: Examples From Northern California. *Journal of Geophysical Research: Solid Earth*, 124(1):1016–1035, 2019. doi: 10.1029/2018JB016886.

Schlagenhauf, A., Manighetti, I., Malavieille, J., and Dominguez, S. Incremental Growth of Normal Faults: Insights from a Laser-Equipped Analog Experiment. *Earth and Planetary Science Letters*, 273(3-4):299–311, Sept. 2008. doi: 10.1016/j.epsl.2008.06.042.

Scott, C., Bello, S., and Ferrarini, F. Matlab Algorithm for Systematic Vertical Separation Measurements of Tectonic Fault Scarps, Nov. 2020. doi: 10.5281/zenodo.4247586.

Scott, C., Adam, R., Arrowsmith, R., Madugo, C., Powell, J., Ford, J., Gray, B., Koehler, R., Thompson, S., Sarmiento, A., Dawson, T., Kottke, A., Young, E., Williams, A., Kozaci, O., Oskin, M., Burgette, R., Streig, A., Seitz, G., Page, W., Badin, C., Carnes, L., Giblin, J., McNeil, J., Graham, J., Chupik, D., and Ingersoll, S. Evaluating How Well Active Fault Mapping Predicts Earthquake Surface-Rupture Locations. *Geosphere*, 19(4):1128–1156, July 2023. doi: 10.1130/GES02611.1.

Scott, C. P., Giampietro, T., Brigham, C., Leclerc, F., Manighetti, I., Arrowsmith, J. R., Laó-Dávila, D. A., and Mattéo, L. Semiautomatic Algorithm to Map Tectonic Faults and Measure Scarp Height from Topography Applied to the Volcanic Tablelands and the Hurricane Fault, Western US. *Lithosphere*, 2021(Special 2): 9031662, Feb. 2022. doi: 10.2113/2021/9031662.

Shridhar, K., Laumann, F., and Liwicki, M. A Comprehensive Guide to Bayesian Convolutional Neural Network with Variational Inference. *arXiv:1901.02731 [cs, stat]*, Jan. 2019a.

Shridhar, K., Laumann, F., and Liwicki, M. Uncertainty Estimations by Softplus Normalization in Bayesian Convolutional Neural Networks with Variational Inference. *arXiv:1806.05978 [cs, stat]*, May 2019b.

Smith, T. R. and Bretherton, F. P. Stability and the Conservation of Mass in Drainage Basin Evolution. *Water Resources Research*, 8 (6):1506–1529, 1972. doi: 10.1029/WR008i006p01506.

Stewart, N., Gaudemer, Y., Manighetti, I., Serreau, L., Vincendeau, A., Dominguez, S., Mattéo, L., and Malavieille, J. "3D_Fault_Offsets," a Matlab Code to Automatically Measure Lateral and Vertical Fault Offsets in Topographic Data: Application to San Andreas, Owens Valley, and Hope Faults. *Journal of Geophysical Research: Solid Earth*, 123(1):815–835, 2018. doi: 10.1002/2017JB014863.

Tucker, G. E., Hobley, D. E. J., McCoy, S. W., and Struble, W. T. Modeling the Shape and Evolution of Normal-Fault Facets. *Journal of Geophysical Research: Earth Surface*, 125(3):e2019JF005305, 2020. doi: 10.1029/2019JF005305.

Vega-Ramírez, L. A., Spelz, R. M., Negrete-Aranda, R., Neumann, F., Caress, D. W., Clague, D. A., Paduan, J. B., Contreras, J., and Peña-Dominguez, J. G. A New Method for Fault-Scarp Detection Using Linear Discriminant Analysis in High-Resolution Bathymetry Data From the Alarcón Rise and Pescadero Basin. *Tectonics*, 40(12):e2021TC006925, 2021. doi: 10.1029/2021TC006925.

Wallace, R. E. Profiles and Ages of Young Fault Scarps, North-Central Nevada. *GSA Bulletin*, 88(9):1267–1281, Sept. 1977. doi: 10.1130/0016-7606(1977)88<1267:PAAOYF>2.0.CO;2.

Wells, D. L. and Coppersmith, K. J. New Empirical Relationships among Magnitude, Rupture Length, Rupture Width, Rupture Area, and Surface Displacement. *Bulletin of the Seismological Society of America*, 84(4):974–1002, Jan. 1994.

Wolfe, F. D., Stahl, T. A., Villamor, P., and Lukovic, B. Short Communication: A Semiautomated Method for Bulk Fault Slip Analysis from Topographic Scarp Profiles. *Earth Surface Dynamics*, 8(1): 211–219, Mar. 2020. doi: 10.5194/esurf-8-211-2020.

Zhang, P., Slemmons, D., and Mao, F. Geometric Pattern, Rupture Termination and Fault Segmentation of the Dixie Valley—Pleasant Valley Active Normal Fault System, Nevada, U.S.A. *Journal of Structural Geology*, 13(2):165–176, Jan. 1991. doi: 10.1016/0191-8141(91)90064-P.

Zielke, O., Klinger, Y., and Arrowsmith, J. R. Fault Slip and Earthquake Recurrence along Strike-Slip Faults — Contributions of High-Resolution Geomorphic Data. *Tectonophysics*, 638:43–62, Jan. 2015. doi: 10.1016/j.tecto.2014.11.004.