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

Lea Pousse-Beltran * [1], **Théo Lallemand** [1], **Laurence Audin** [1], **Pierre Lacan** [2,3], **Andres David Nuñez-Meneses** [3], **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.

S1 Supplementary Texts

S1.1 SimScarp workflow

At the beginning, SimScarp chooses a random slip rate between realistic uniform distribution ($\mathcal{U}(0, Xyrs)$), and a realistic cumulative throw ($\mathcal{U}(0, Xyrs)$). Then the code estimates the number of events. SimScarp estimates the throw (u_i) for each event (i), according on the slip rate, the cumulative throw and the minimum and maximum throw per event defined. According to the slip rate, the number of event, the cumulative throw, SimScarp assigns periods between each events. To create the profile, SimScarp requires two slopes, one for the hanging wall (β_h) and one for the footwall (β_f), sampled from an uniform distribution ($\mathcal{U}(\beta_{min}, \beta_{max})$). The simulator SimScarp breaks a secondary fault branch, with a dip (δ) randomly set (uniform distribution,$\mathcal{U}(\delta_{min}, \delta_{min})$). Then, between each rupture a diffusive erosion is applied during the period between events. The diffusive erosion, is here a finite difference solution with a stability constraint is relative to dt fixed by the user. The simulator SimScarp breaks the main fault branch, with a dip (δsf) randomly set (uniform distribution, $\mathcal{U}(\delta sf_{min}, \delta sf_{min})$). The rupture location (X) is then randomly set to ± 5 % from the profile center (Gaussian distribution; $\mathcal{N}(mean\ profile, 5\% of\ the\ profile\ length)$). At each rupture a fault scarp is created at the bottom of the scarp, rejuvenating the scarp, with a throw per event. Then, between each rupture a diffusive erosion is applied during the period between events. The boundaries are set after the rupture and are stable through the diffusion time.

The total scarp height (S_h) is calculated by adding every scarp height (S_{hi}) created at each event (i). Here we measure the scarp height at the middle of the scarp, following this equation:

$$S_{hi} = u_i * (1 - \frac{\tan\beta_f + \tan\beta_h}{2 * \tan\delta_i}) \quad (1)$$

Once the ruptures are produced, SimScarp adds non-tectonic perturbations at random locations along the profile in order to create a realistic morphology using random parabolas or steps functions such as in Hodge et al. (2019). Those parabola attempt to represent narrow drainage, wide rivers, hills, steps functions attempts to represent trees. The number of parabolas or steps functions, theirs locations, heights and widths are chosen randomly in a uniform distribution (Table 2). Finally SimScarp adds a Gaussian noise accounting for an arbitrary perturbation affecting all the topographic profile.

S1.2 ScarpLearn_1F training

We train ScarpLearn_1F on our synthetic set (5000 samples) using a batch gradient descent of 32 samples per batch. We used 1000 epochs in which follow the evolution through the epochs of the ELBO loss. Concerning the confidence interval, 20% of the predicted target intervals are integrating the ground truth value. To convert this confidence interval into uncertainties, we thus multiply it by 5 to simulate a 1σ uncertainty.

*Corresponding author: pousse.lea@ird.fr

S2 Supplementary Figures

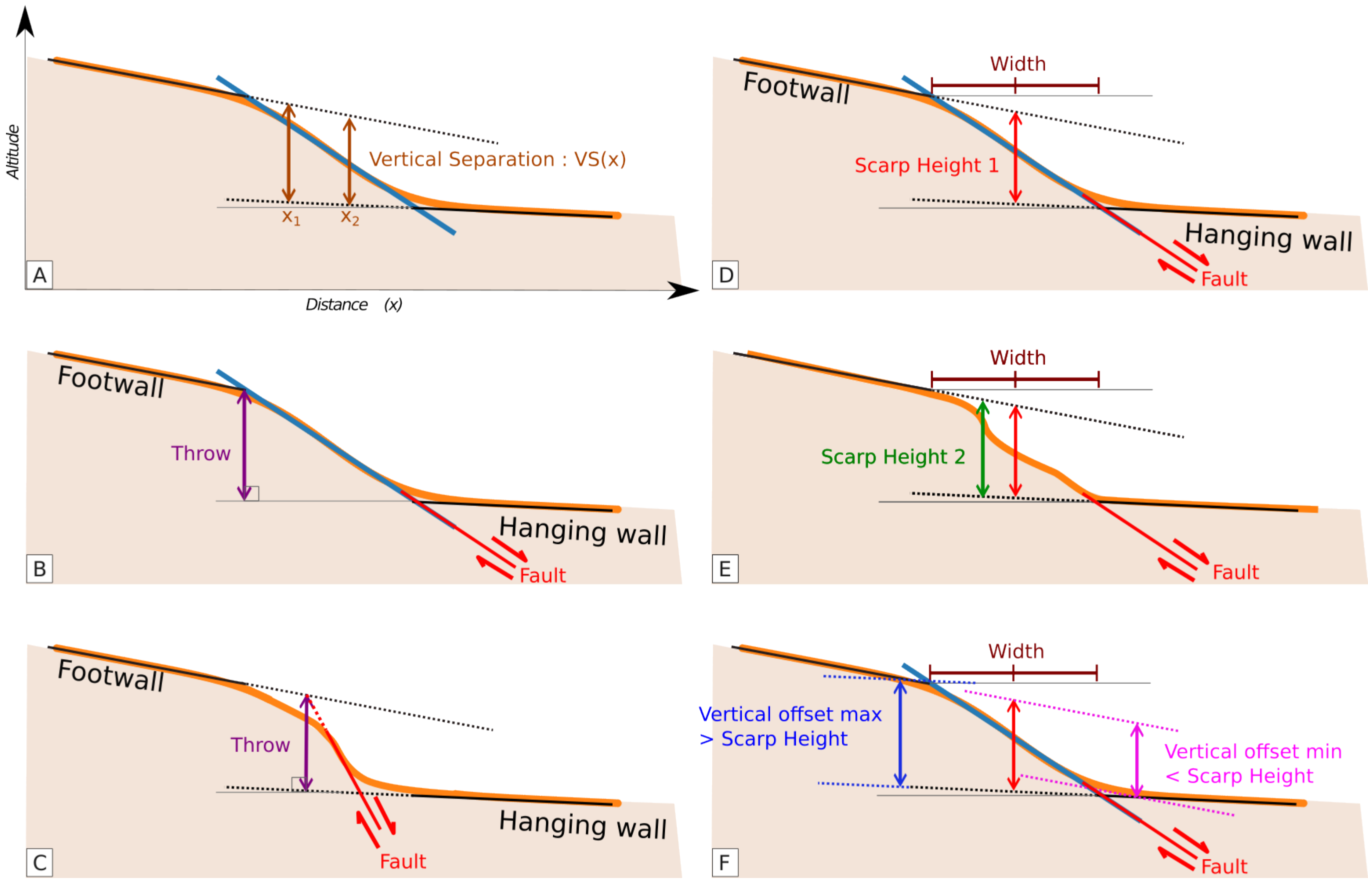

Figure S1 Difference between terminologies in morphotectonics to describe the measure of deformation across a topographic profile : A) Vertical separation, B-C) Throw depending on the fault trace location, D-E-F) Scarp Height and the different methods of measuring it. Also not that the scarp dip is not necessarily the fault dip. D) Some studies focus the measurement on the middle of the scarp. E) Some other focus the on the location where the scarp has a maximum slope. F) Some others projects the hanging wall (or the footwall) on the inflexion between the footwall (or hanging wall, respectively) and the scarp to bracket the vertical offset. In addition it can be noted that if the footwall and hanging wall are horizontal and the fault free face is the scarp, then the throw, the vertical separation and scarp height will be equal.

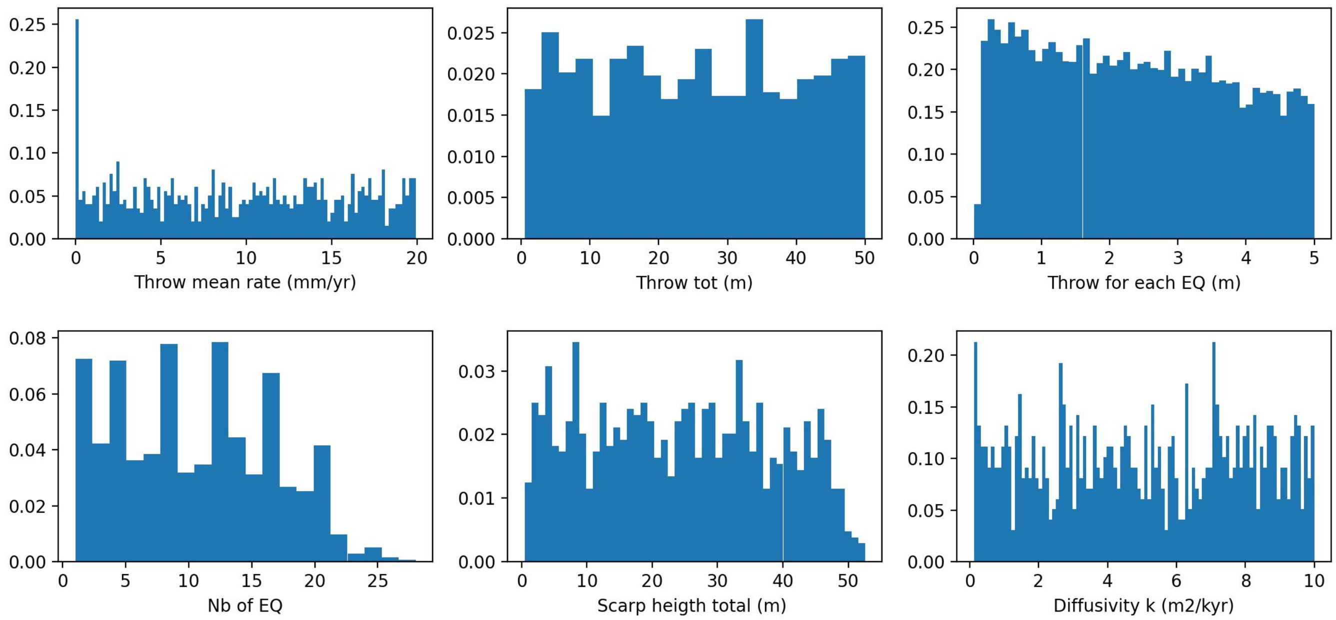

Figure S2 Distribution of SimScarp parameters obtained when generating synthetic datasets to train ScarpLearn. See Table 2 for the parameters used.

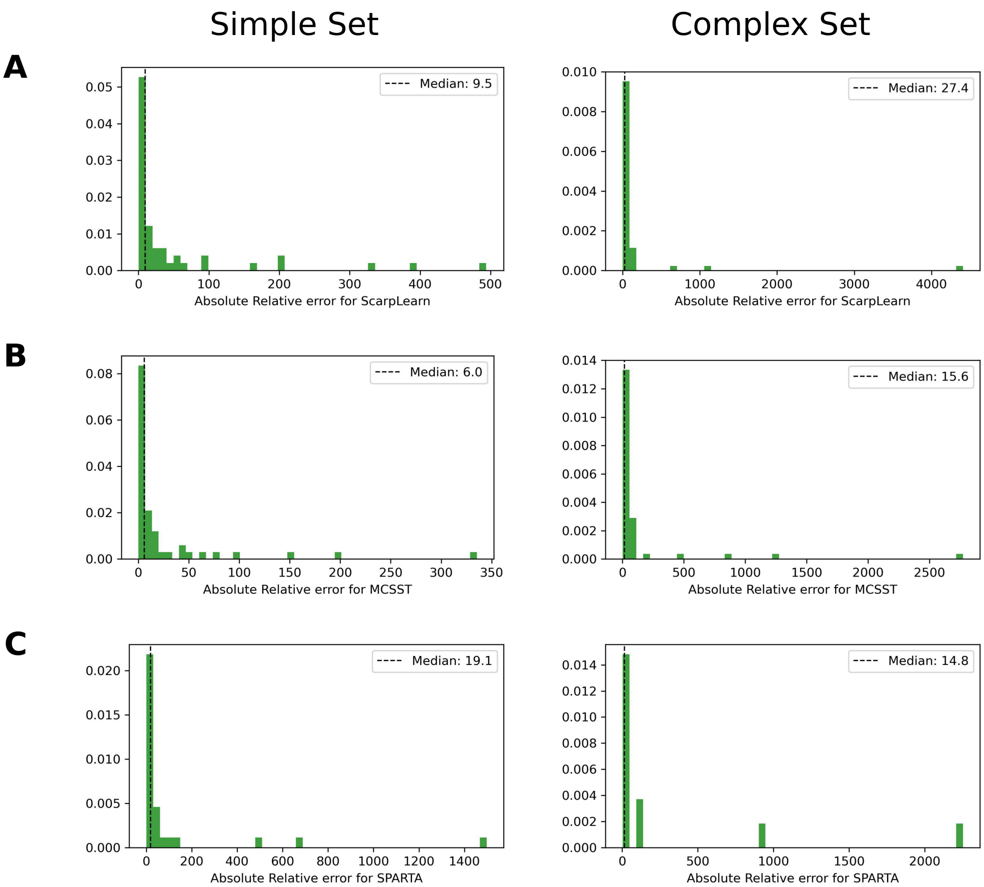

Figure S3 Absolute Relative error distributions for the simple and the complex data sets. Distribution obtained with ScarpLearn in A, with MCSST in B and with SPARTA in C.

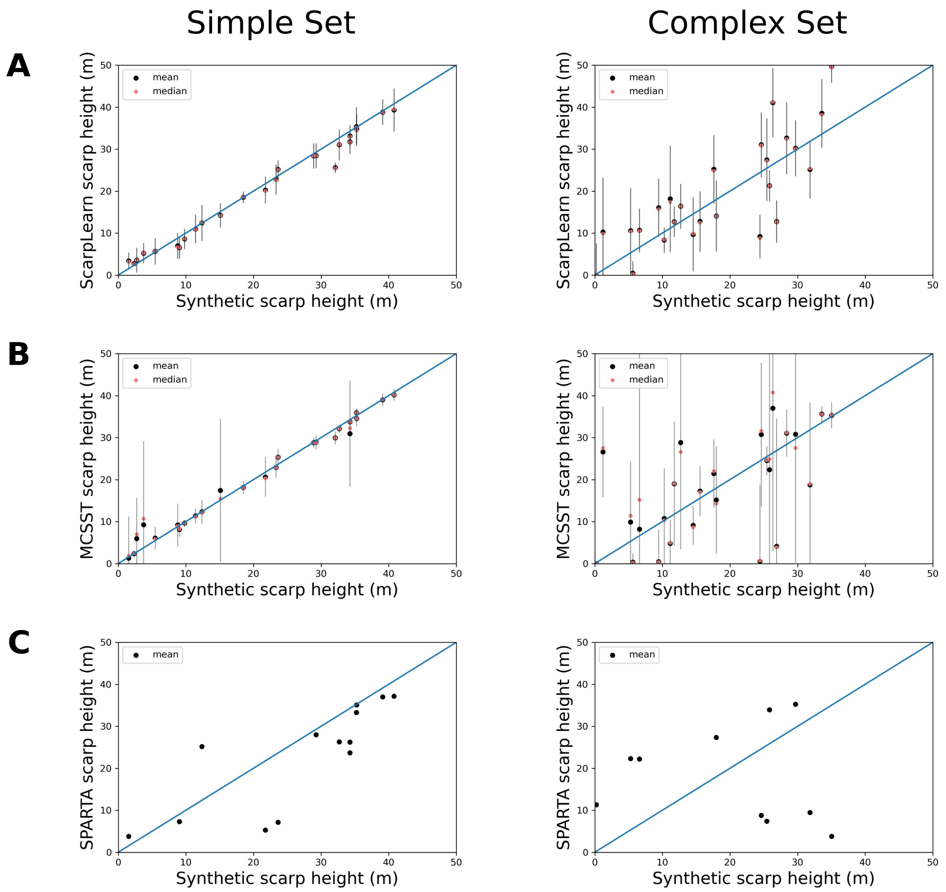

Figure S4 Synthetic setting with only one fault. Labels (true values of scarp height) versus predictions (**ScarpLearn_1F** in A, MCSST in B and SPARTA in C) for two set of synthetic datasets. The left plot corresponds to the simple setting and the right plot corresponds to the complex setting. In A and B, uncertainty bars show 1σ. See Table S4 for metrics.

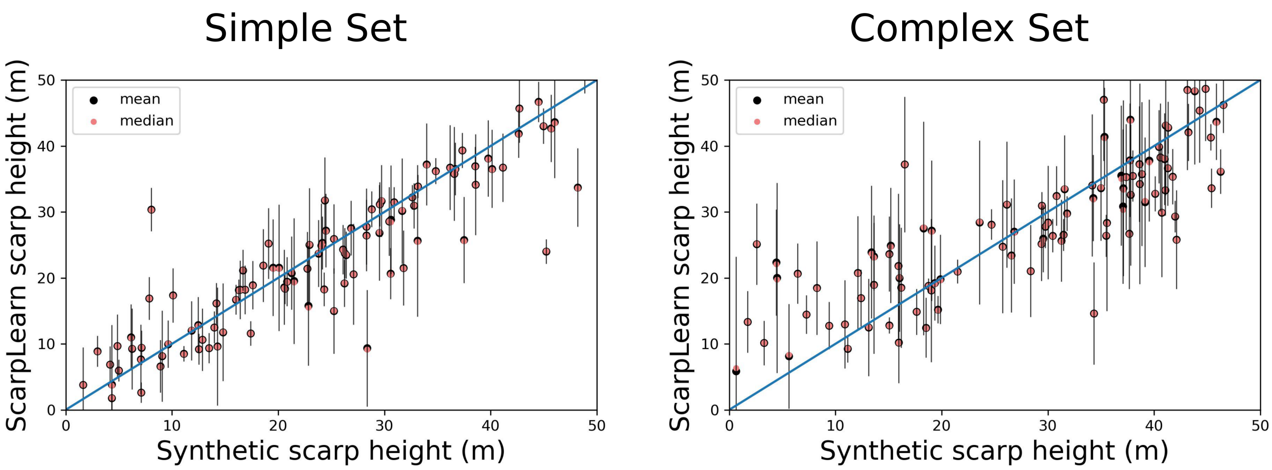

Figure S5 Synthetic setting test where the topographic profiles include only two faults. The figure shows the labels (true values of the scarp height from the synthetics) versus ScarpLearn scarp height predictions. The left plot corresponds to the simple setting and the right plot corresponds to the complex setting. See Table S3 for metrics. See legend in Fig. S4 for more details.

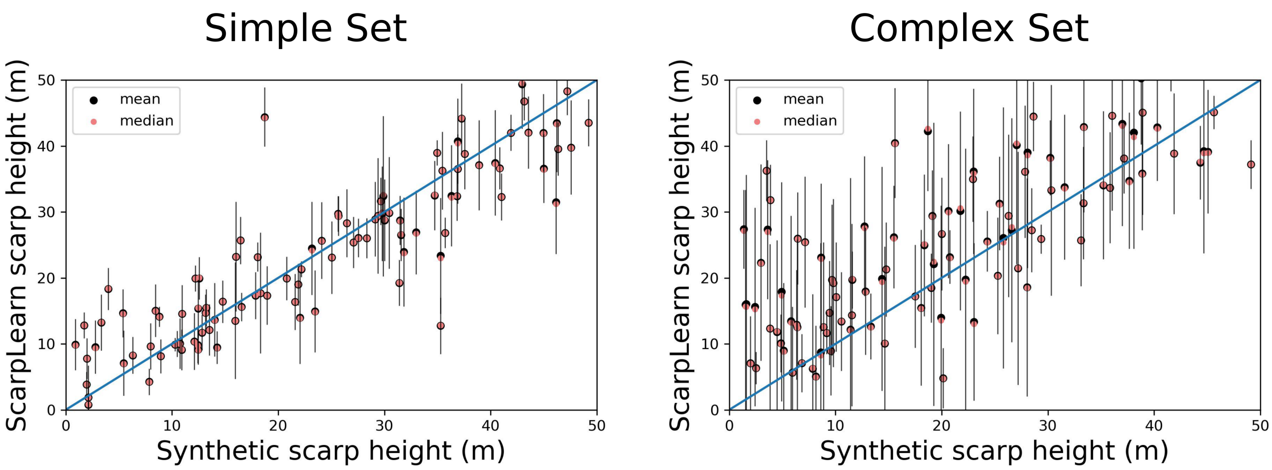

Figure S6 Synthetic setting test where the topographic profiles include only three faults. The figure shows the labels (true values of the scarp height from the synthetics) versus ScarpLearn scarp height predictions. The left plot corresponds to the simple setting and the right plot corresponds to the complex setting. See Table S3 for metrics. See legend in Fig. S4 for more details.

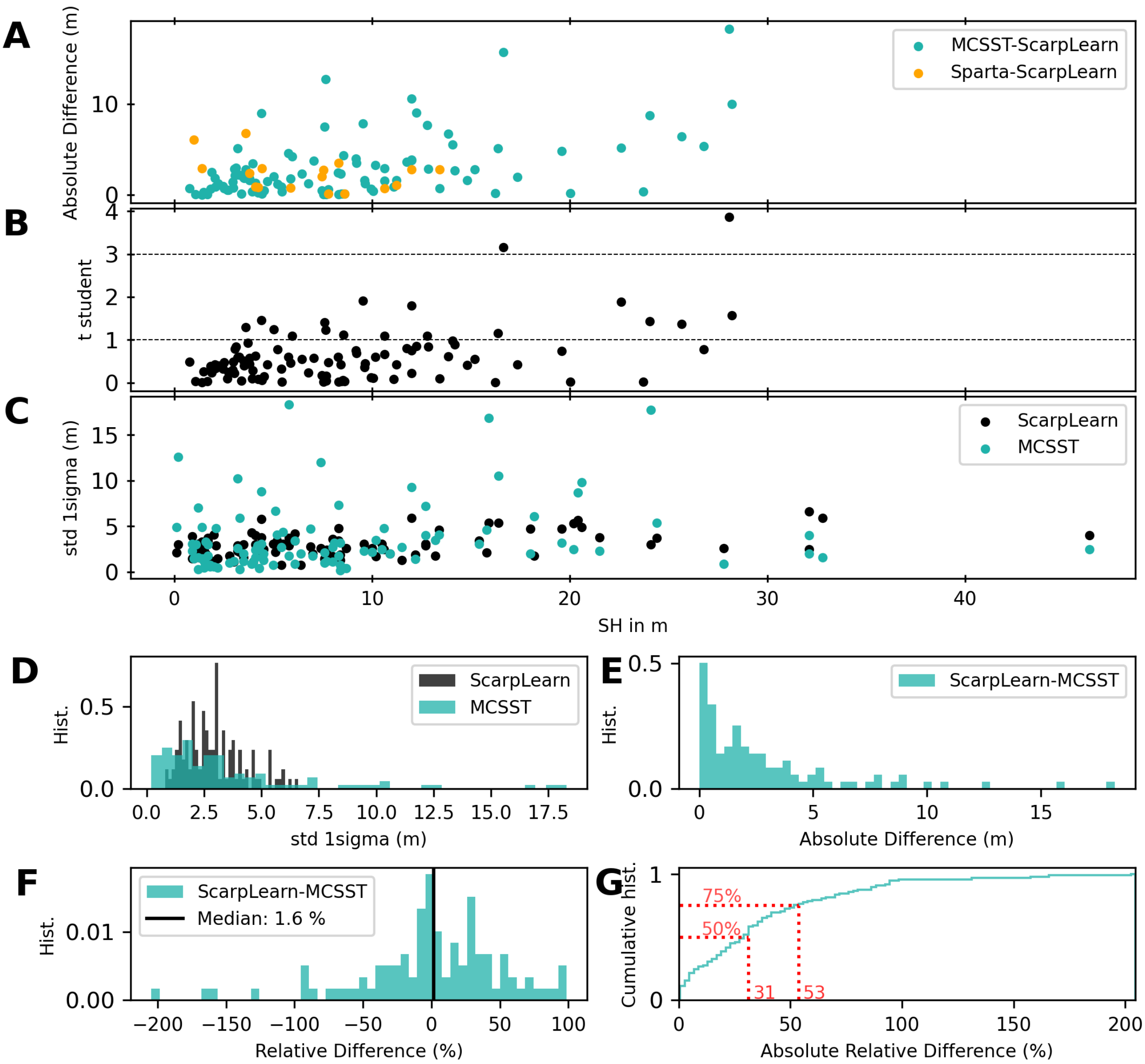

Figure S7 Ameca Fault results. A: Absolute difference between MCSST and ScarpLearn (in green) and between Sparta ans ScarpLearn (in orange) versus the the scarp height. B: T-student tests between MCSST and ScarpLearn agreement of the scarp height distributions versus scarp height. C: Standard deviation of ScarpLearn (in black) and of MCSST (in green) versus scarp height. D: Distribution of ScarpLearn standard deviation (in black) and of MCSST standard deviation (in green). E: Distribution of the absolute difference in m between MCSST and ScarpLearn (in green). F: Distribution of the relative difference in % between MCSST and ScarpLearn, which is the difference between MCSST and ScarpLearn divided by the scarp height estimated with ScarpLearn. G: Cumulative distribution of relative difference plotted in F.

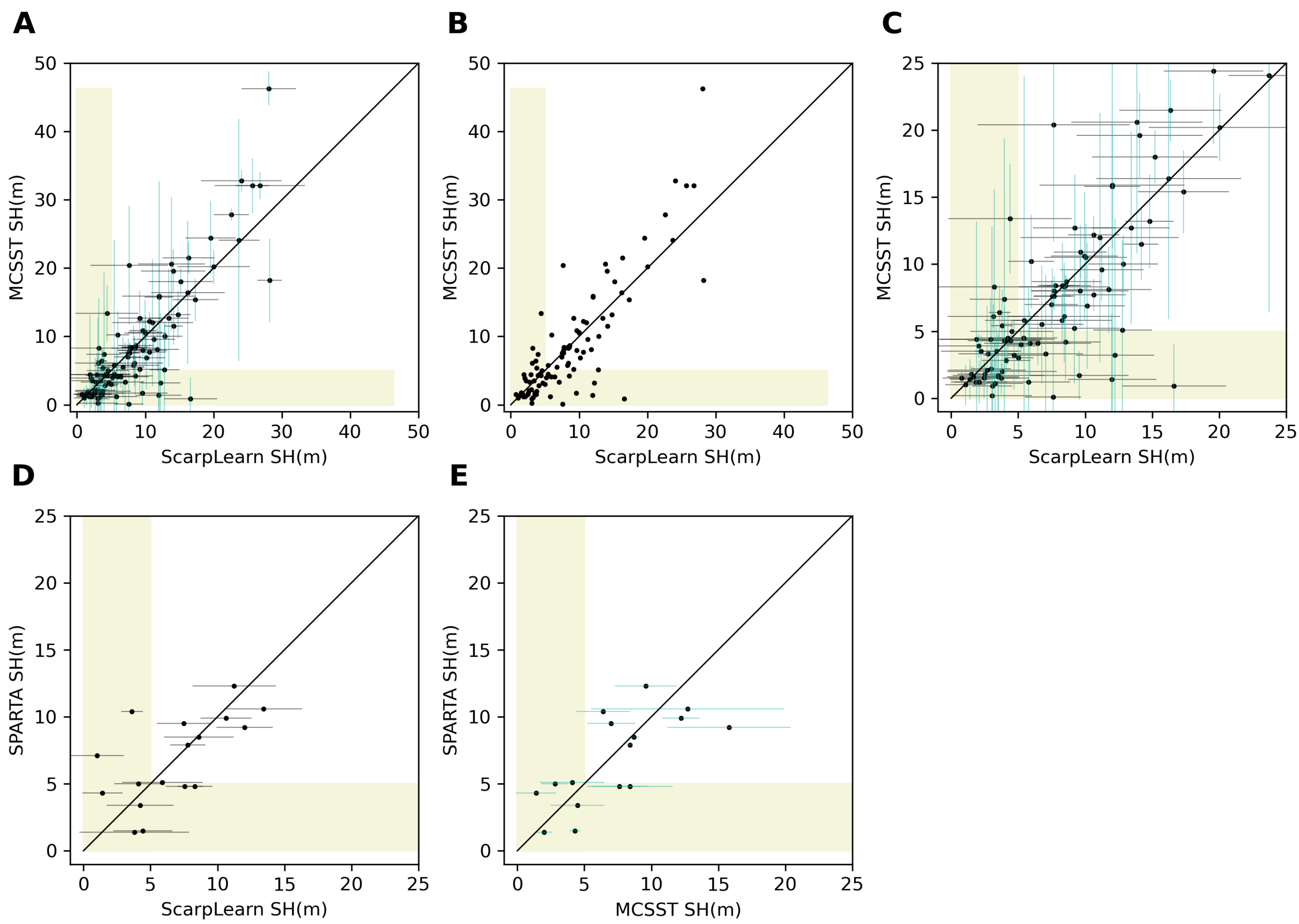

Figure S8 Ameca Fault results. A,B,C: Scarp Height estimations with MCSST versus estimations with ScarpLearn. On the X-axis (light black), we display the uncertainties for the ScarpLearn measurements, while on the Y-axis (light blue), we present the uncertainties for the MCSST measurements. (A: with uncertainties at 1σ, B: without uncertainties, C: zoom below 25m). D-E: Scarp Height estimations with Sparta versus estimations with ScarpLearn (plot D) or with MCSST (plot E). Beige banding represent 5-m, which is the DEM pixel size.

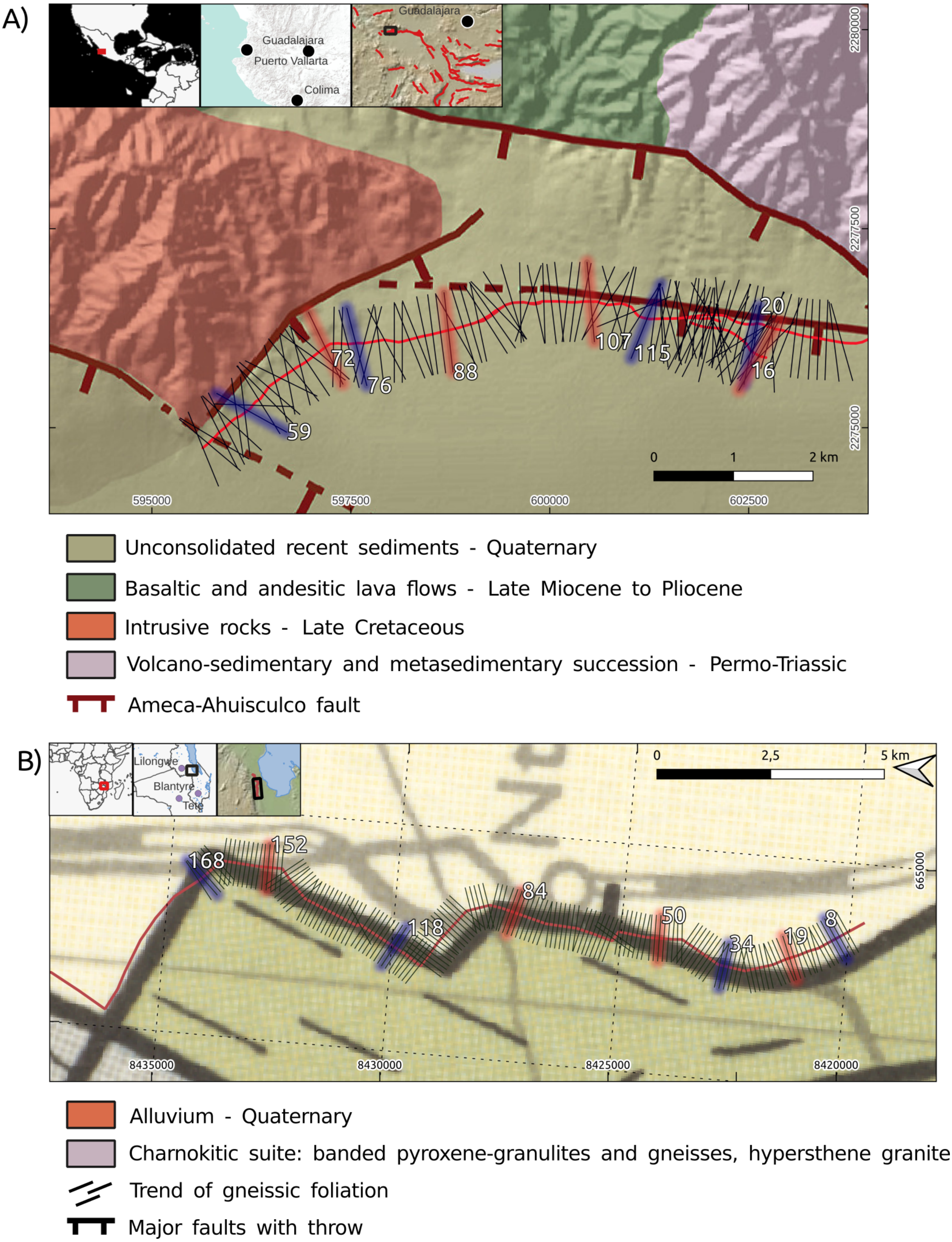

Figure S9 Geological map of (A) the Ameca fault study area (Núñez Meneses et al., 2021; Ferrari et al., 2018) and (B) Bilila-Mtakataka fault study area (Bloomfield and Mason, 1966).

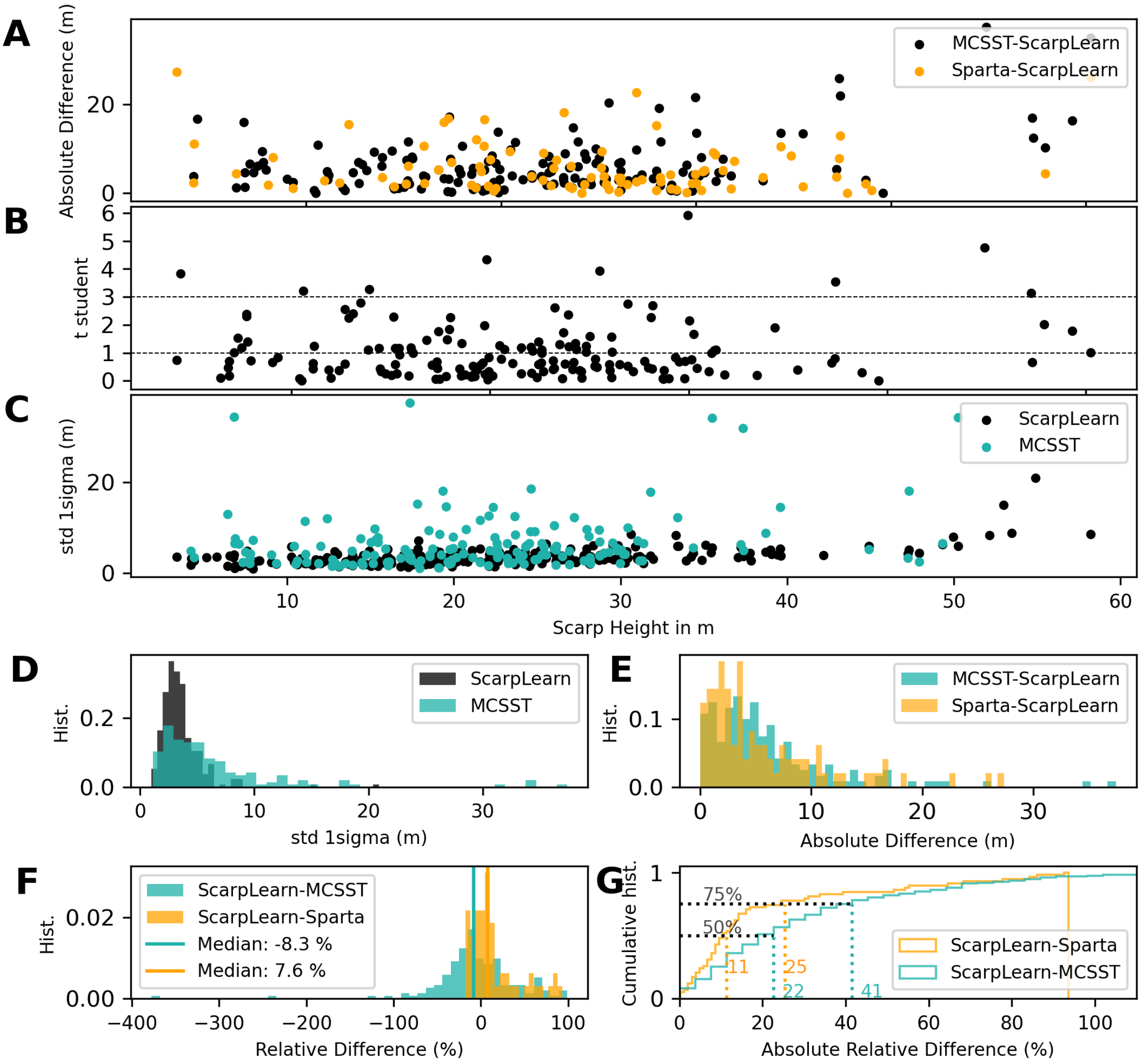

Figure S10 Bilila-Mtakataka Fault results. See legend in Fig. S7.

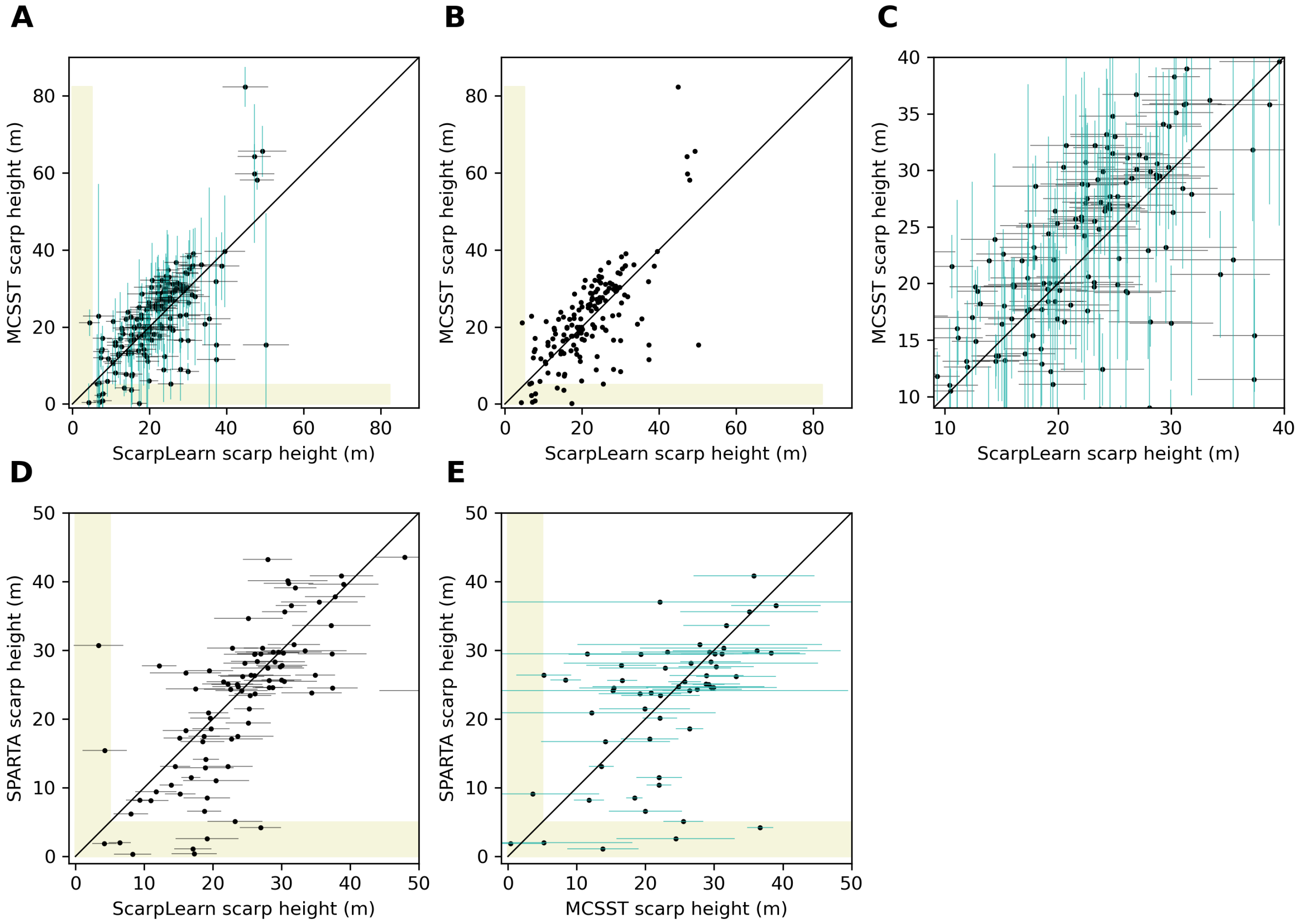

Figure S11 Bilila-Mtakataka Fault results. See legend in Fig. S8.

S3 Supplementary Tables

Table S1 Parameters chosen from statistical distributions to create topographic profiles in SimScarp for **simple** dataset. This simple dataset is used to create 100 synthetic topographic profiles, on which ScarpLearn, MCSST and SPARTA will be tested (see Table 3 and Fig. 6).

	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	1	3	/	/
	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	0.1 m^2/Kyr	10 m^2/Kyr	/	/
	Slip rate	Uniform	0.05 mm/y	20 mm/y		/
	Maximum number of events	Uniform	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	/	/

Table S2 Parameters chosen from statistical distributions to create topographic profiles in SimScarp for **complex** dataset. This complex dataset is used to create 100 synthetic topographic profiles, on which ScarpLearn, MCSST and SPARTA will be tested (see Table 3 and Fig. 6).

	Parameters	Distribution	Minimum	Maximum	Mean	Standard Deviation
Regional slopes	Hanging wall slope β_h	Uniform	**-10°**	**25°**	/	/
	Footwall slope β_f	Uniform	**-10°**	**25°**	/	/
Secondary faults	Number of secondary fault	Uniform	1	3	/	/
	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	0.1 m^2/Kyr	10 m^2/Kyr	/	/
	Slip rate	Uniform	0.05 mm/y	20 mm/y		/
	Maximum number of events	Uniform	1	-	/	/
Perturbations	Gaussian noise	Gaussian	/	/	0	(0.1-1)
	Parabolas A number	Uniform	0	**3**	/	/
	Parabolas A width	Uniform	0.1	150	/	/
	Parabolas A height	Uniform	-10	10	/	/
	Trees number	Uniform	0	**50**	/	/
	Trees width	Uniform	0.1	10	/	/
	Trees height	Uniform	1	15	/	/

Table S3 Main metrics to compare ScarpLearn using synthetic datasets. RMS is the Root Mean Square, MSE is the Mean Square 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 PICP of 100% means that all truth values fall in the prediction interval. The parameters for SimScarp to create the simple and the complex datasets are in Tables S1 and S2. See plots in Fig. S4, S5 and S6 in appendix.

Sets	Metrics	ScarpLearn 1 fault dataset	ScarpLearn 2 faults dataset	ScarpLearn 3 faults dataset
Simple Dataset	Number of profiles	100	100	100
	Time to process	<1 min	<1 min	<1 min
	Mean scarp height	19.3 m	23.0 m	23.7 m
	Mean Absolute error	2.3 m	3.6 m	4.4 m
	RMSE	3.6 m	5.4 m	6.4 m
	PICP at 1σ, 2σ, 3σ	89%, 98%, 99%	75%, 87%, 96%	61%, 83%, 92%
	Mean and std of uncertainties (at 1σ)	5.2 $\pm$ 1.8 m	4.8 $\pm$ 2.3 m	4.5 $\pm$ 2.1 m
	NLL	2.8	2.9	3.7
	Relative uncertainties	28 $\pm$ 32 %	20 $\pm$ 31 %	19 $\pm$ 43 %
Complex Dataset	Number of profiles	100	100	100
	Time to process	<1 min	<1 min	<1 min
	Mean scarp height	22.8 m	28.0 m	25.1 m
	Mean Absolute error	8.8 m	5.7 m	7.6 m
	RMSE	11.5 m	7.8 m	10.2 m
	PICP at 1σ, 2σ, 3σ	67%, 87%, 99%	64%, 81%, 95%	72%, 89%, 93%
	Mean and std of uncertainties (at 1σ)	11.2 $\pm$ 4.4 m	7.1 $\pm$ 3.5 m	10.5 $\pm$ 4.7 m
	NLL	4.0	3.8	4.2
	Relative uncertainties	52 $\pm$ 60%	24 $\pm$ 32 %	40 $\pm$ 42 %

Table S4 Main metrics to compare ScarpLearn, MCSST and SPARTA using synthetic datasets only having 1 fault (see Fig. S4 in appendix). See legend of the Table S3 for metrics definitions.

Sets	Metrics	ScarpLearn	MCSST	SPARTA*	ScarpLearn_1F train for 1 fault
		1 fault Dataset	1 fault Dataset	1 fault Dataset	1 fault Dataset
Simple Dataset	Number of profiles	only on 25	25	13 (over 25)	only on 25
	Time to process	<1 min	1-2 hours	<1 hour	<1 min
	Mean scarp height	18.3 m	20.5 m	22.7 m	19.7 m
	Mean Absolute error	3.3 m	1.0 m	6.4 m	1.3 m
	Abs. Relative error	11.6%	2.1%	19.6%	5.8%
	RMS	4.8 m	1.7 m	8.6m	1.8 m
	PICP at 1σ, 2σ, 3σ	84%, 88%, 92%	96%, 100%, 100%	-	92%, 96%, 96%
	Mean and std of uncertainties (at 1σ)	5.4 ± 4.0 m	4.3 ± 5.2 m	-	2.9 ± 0.8 m
	NLL	3.5	1.9	-	2.6
	Relative uncertainties (median)	32 ± 28 %	10 ± 154 %	-	16 ± 23 %
Complex Dataset	Number of profiles	only 25	25	(10 on 25)	25
	Time to process	<1 min	1-2 hours	<1 hour	<1 min
	Mean scarp height	21.7 m	18.1 m	18.2 m	19.3 m
	Mean Absolute error	7.9 m	7.1 m	15.4 m	6.0 m
	Abs. Relative error	30.6%	37.4%	70.5%	31.4%
	RMS	11.2 m	10.2 m	21.1 m	7.4 m
	PICP at 1σ, 2σ, 3σ	76%, 88%, 100%	76%, 92%, 100%	-	76%, 88%, 96%
	Mean and std of uncertainties (at 1σ)	11.4 ± 3.9 m	15.3 ± 14 m	-	7.4 ± 3.0 m
	NLL	3.8	3.6	-	3.7
	Relative uncertainties (median)	58 ± 60 %	92 ± 730 %	-	38 ± 206 %

References

Bloomfield, K. and Mason, T. P. R. Geological Map of Malawi, 1966, 1966.

Ferrari, L., Orozco-Esquivel, T., Navarro, M., López-Quiroz, P., and Luna, L. Digital Geologic Cartography and Geochronologic Database of the Trans-Mexican Volcanic Belt and Adjoining Areas. *Terra Digitalis*, 2(1):1–8, Apr. 2018. doi: 10.22201/igg.terradigitalis.2018.1.34.

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. 2019. doi: 10.5194/se-10-27-2019.

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.