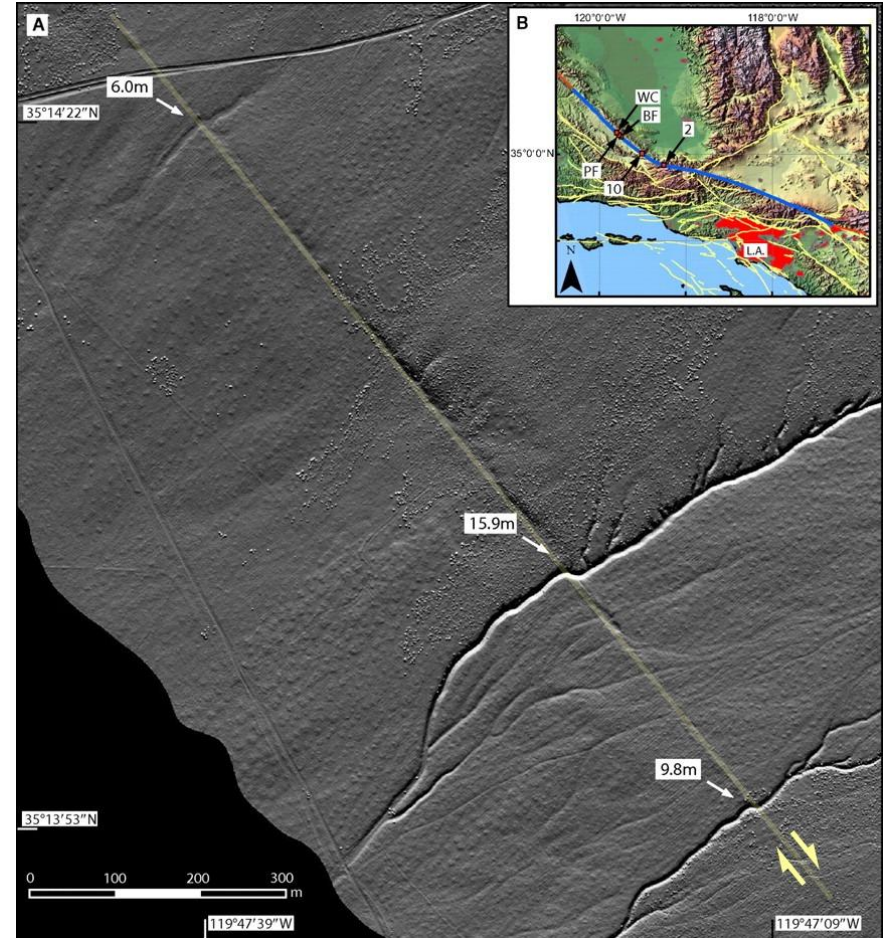


Characterization of strike-slip fault offsets using deep learning and simulated topographic data

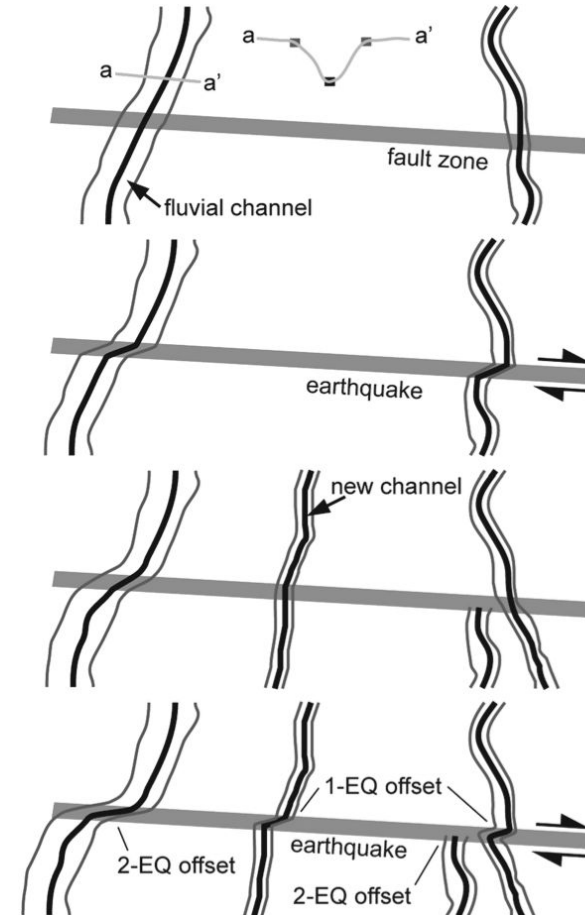
Inferring tectonic history from geomorphic offsets

- Displaced landscape features along faults
- Used to estimate size and recurrence of surface-rupturing past earthquakes



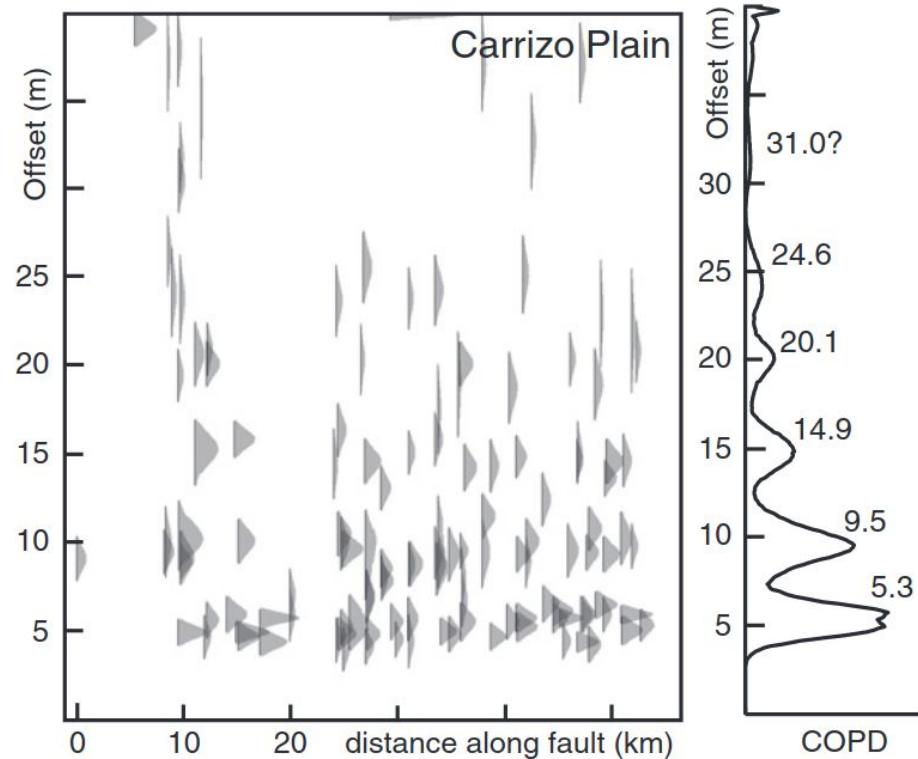
Inferring tectonic history from geomorphic offsets

- Displaced landscape features along faults
- Used to estimate size and recurrence of past surface-rupturing earthquakes



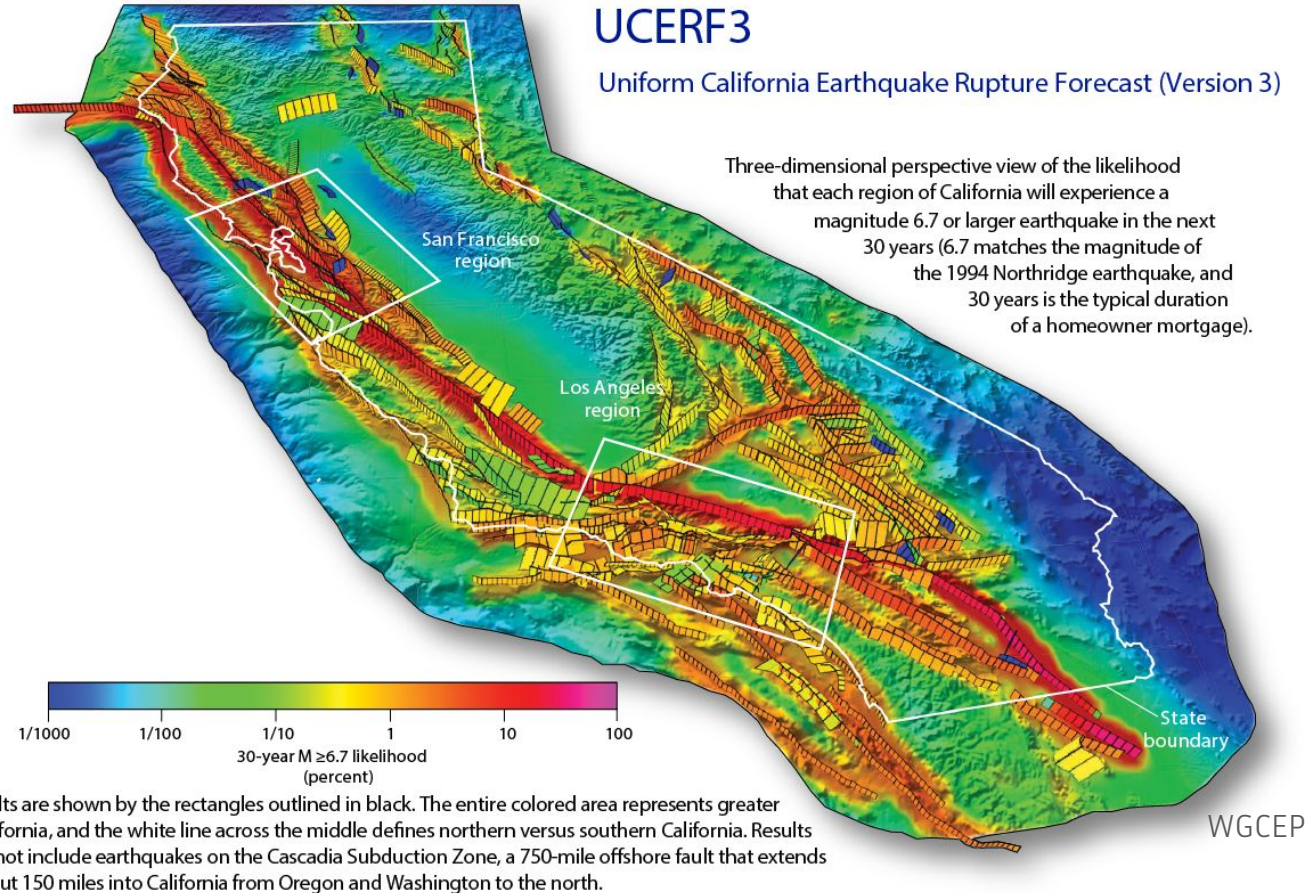
Distribution of offsets along strike

- Cumulative offset probability distributions (COPD)
- Constrain earthquake history and distribution of slip in space/time



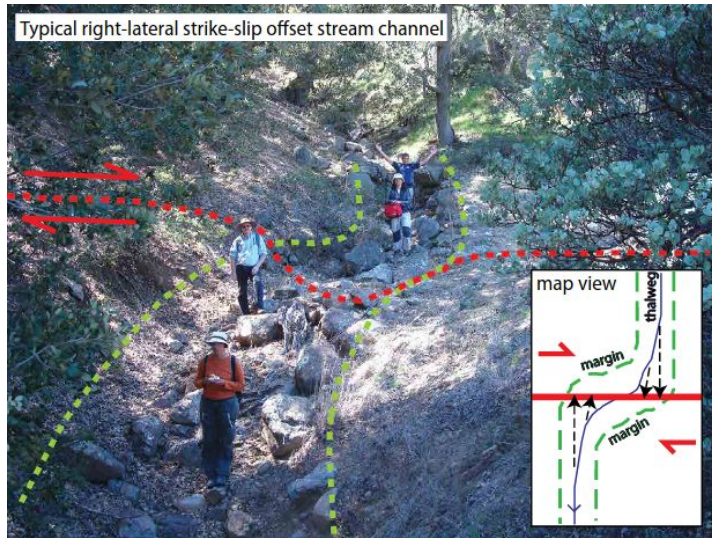
Along-strike offset distribution and cumulative offsets probability distribution (COPD), San Andreas fault, Carrizo Plain (Zielke et al. '15)

Seismic hazard assessment

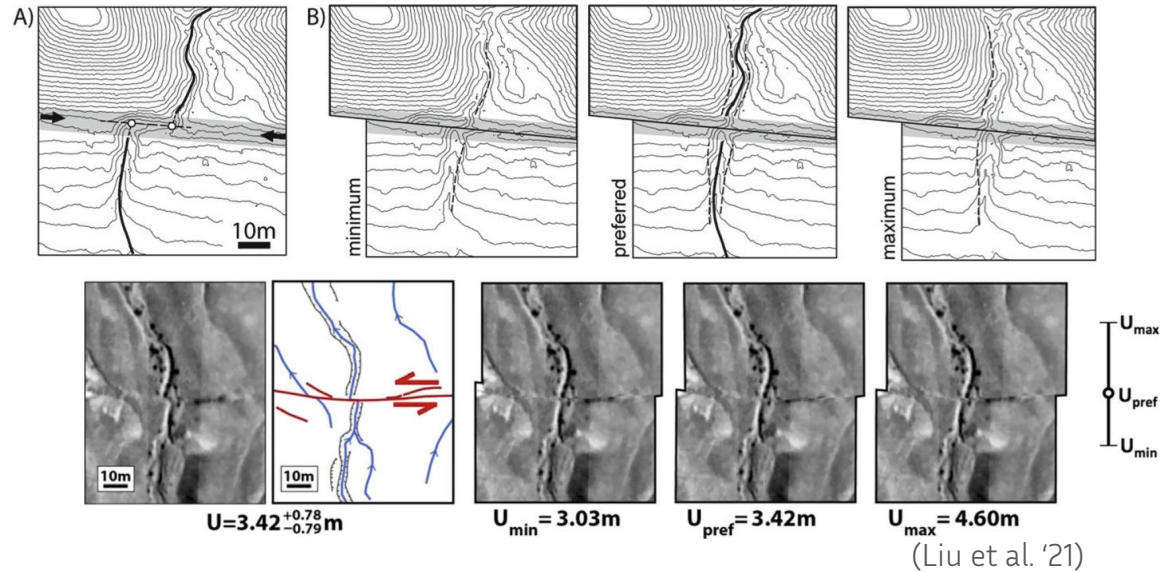


Methods for measuring displacements

- In-situ
- Remote-sensing based: using HR DEMs or optical images



(Salisbury et al. '15)

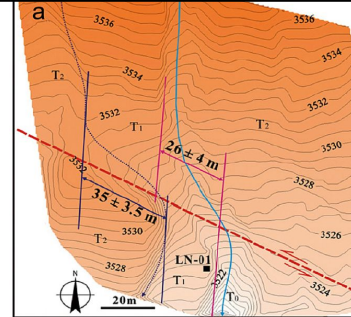
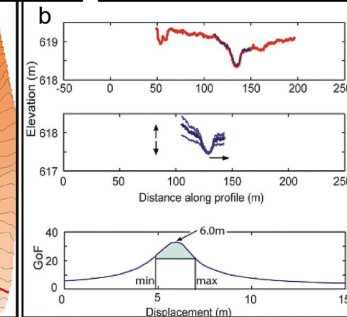


(Liu et al. '21)

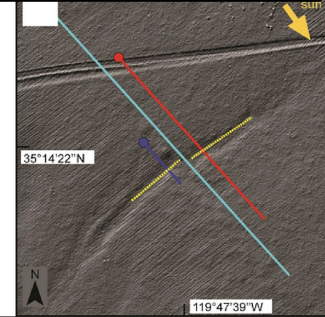
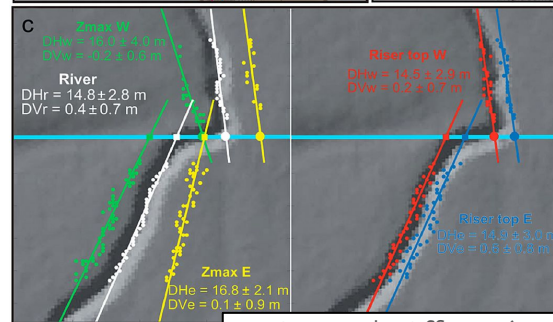
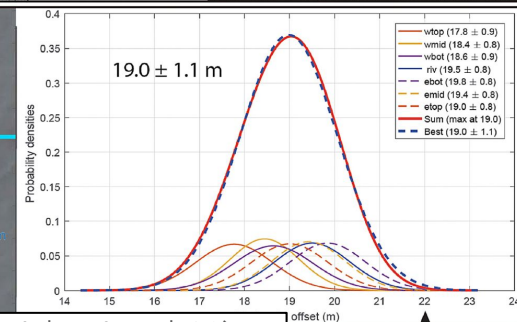
Topographic measurement methods and limitations

- Manual or semi-automatic methods: long process, expert-dependent, difficult to reproduce
- Need for a fully automatic approach: Machine learning model

Contour matching / backslipping



LaDiCaoz (Zielke et al' 12)



3D_Fault_Offsets (Manighetti et al' 20)

Liu et al '21

Machine Learning for Morphotectonics



Normal fault scarps
(Pousse et al., 2025)



Strike-slip fault
offsets
(this work)

- Automate systematized detection and/or measurement of tectono-geomorphic markers
 - fast, calibration-free, reproducible
- Main challenge: lack of labeled natural data for training models
- Need to generate synthetic topographic data, with controlled tectonic and geomorphic (e.g., erosion) processes.

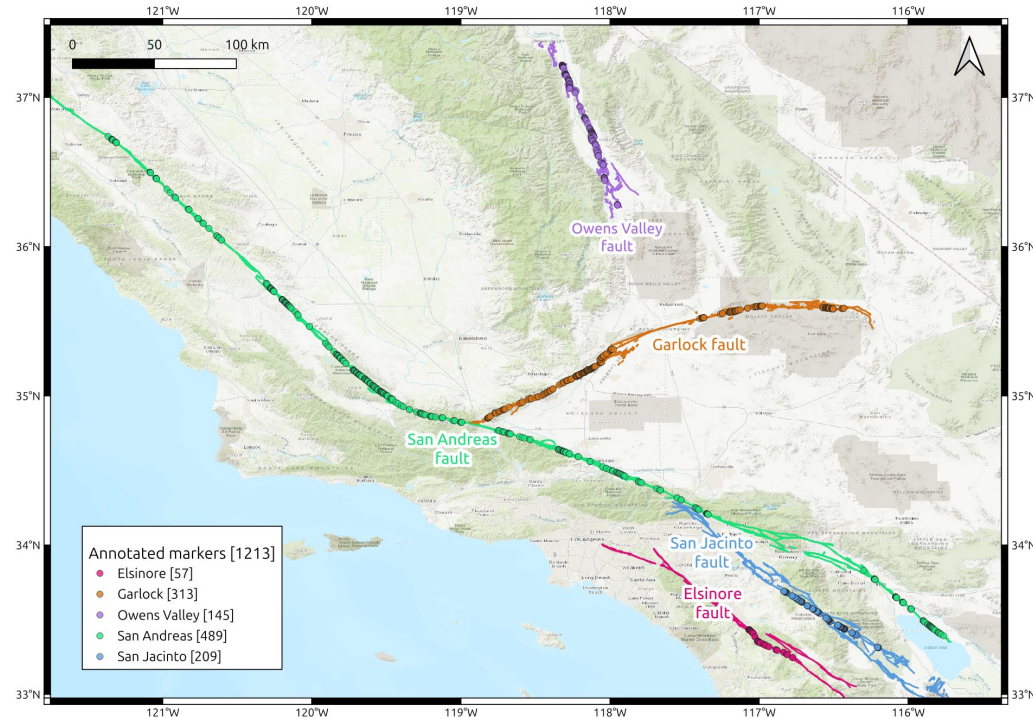
StrikeLearn approach



- Method to **automatically quantify** strike-slip fault offsets from high-resolution topography
- **2-D Tectono-geomorphic landscape simulation** for synthetic data generation
- **Deep neural network** with **simulated-to-real transfer learning**

Study data

- Offset markers from **five strike-slip faults in California** (UCERF3 database, Madden et al. '14), measured with semi-automatic (LaDiCaoz) and manual methods on topographic data
- **HR DEMs** (0.5–0.6 m/px) from airborne LiDAR campaigns (NCALM, EarthScope)

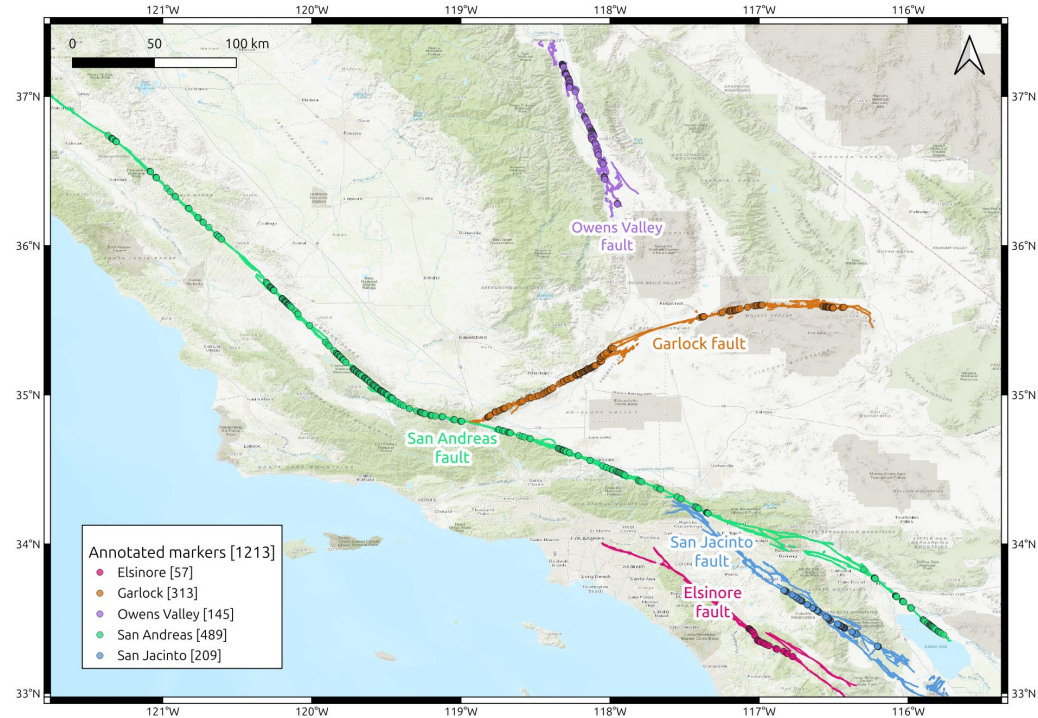


Natural dataset creation

- <700 samples of natural measurements

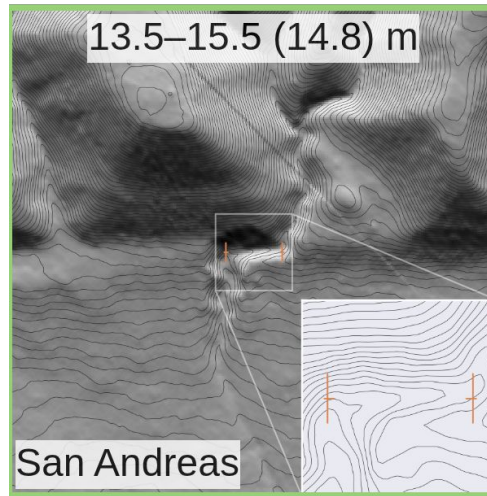
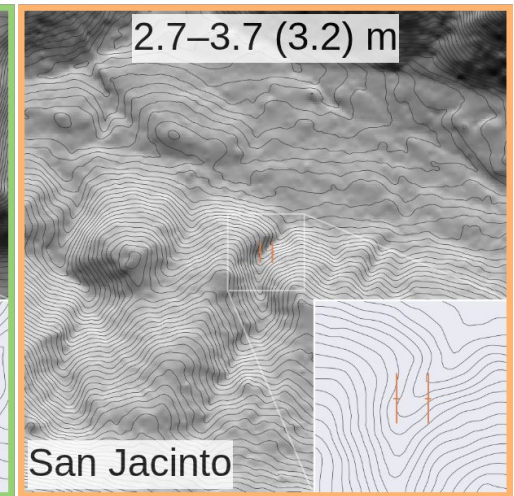
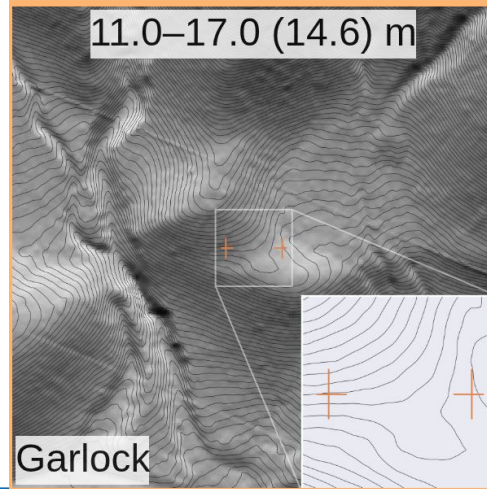
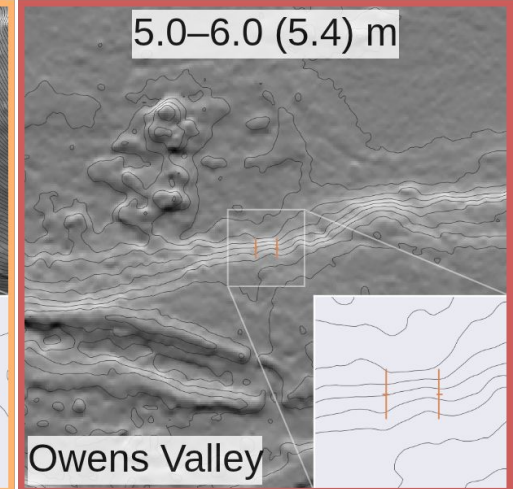
Total dataset size	Training set size	Test set size
666	422	244

- Lack of sufficient ground-truth data for supervised learning



Natural examples

- Variable quality: epistemic uncertainty
- Range of plausible offset values: aleatoric uncertainty



Synthetic data generation

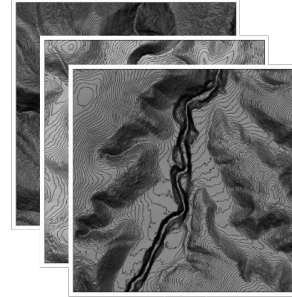
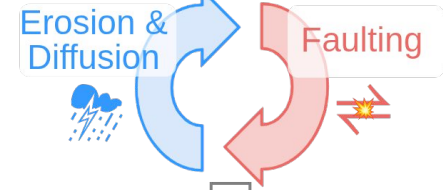
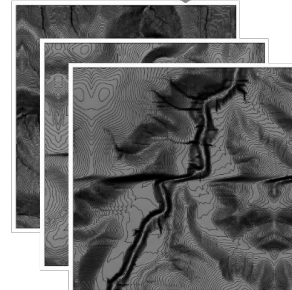
Simulating strike-slip faulting and
landscape evolution
(Reitman et al. '19)

Simulation of landscape evolution

Simulator parameters:

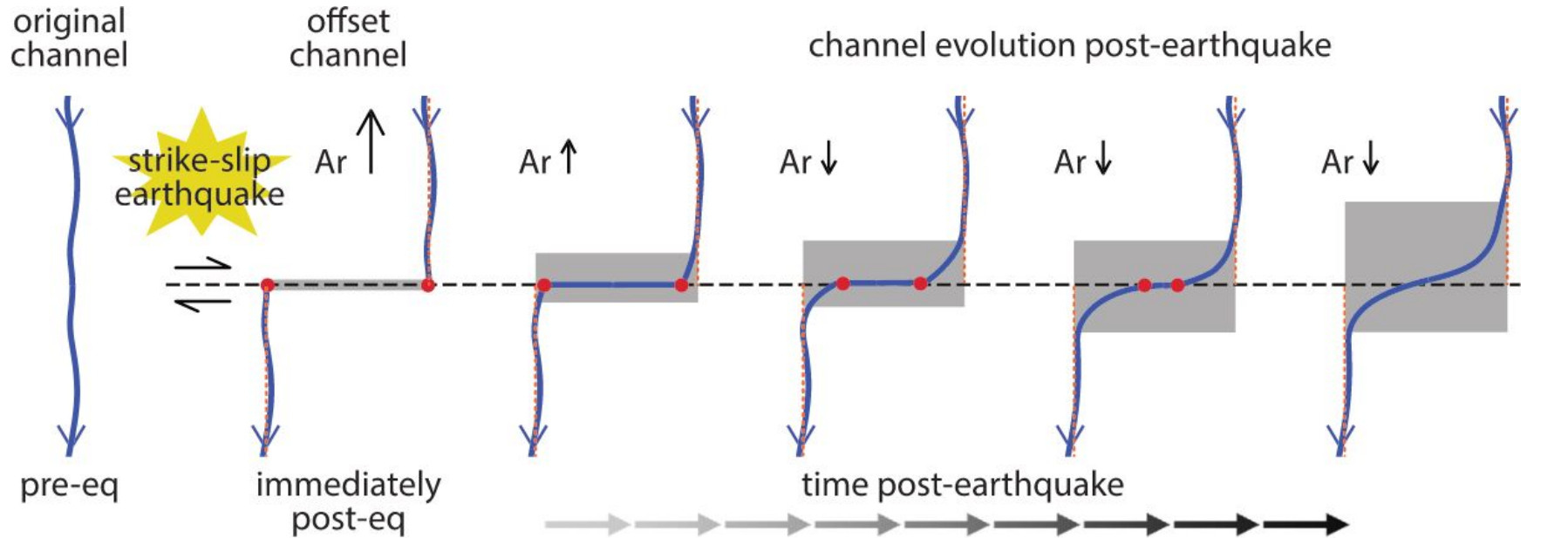
total time [yr]
slip rate [m/yr]
recurrence interval [yr]
fault zone width [m]
stream power coeffs.
diffusivity coeffs.
uplift rate [m/yr]
fault - river angle [deg]

Natural DEMs with
non-faulted stream channels



Synthetic DEMs with
displaced stream channels

Synthetic data generation

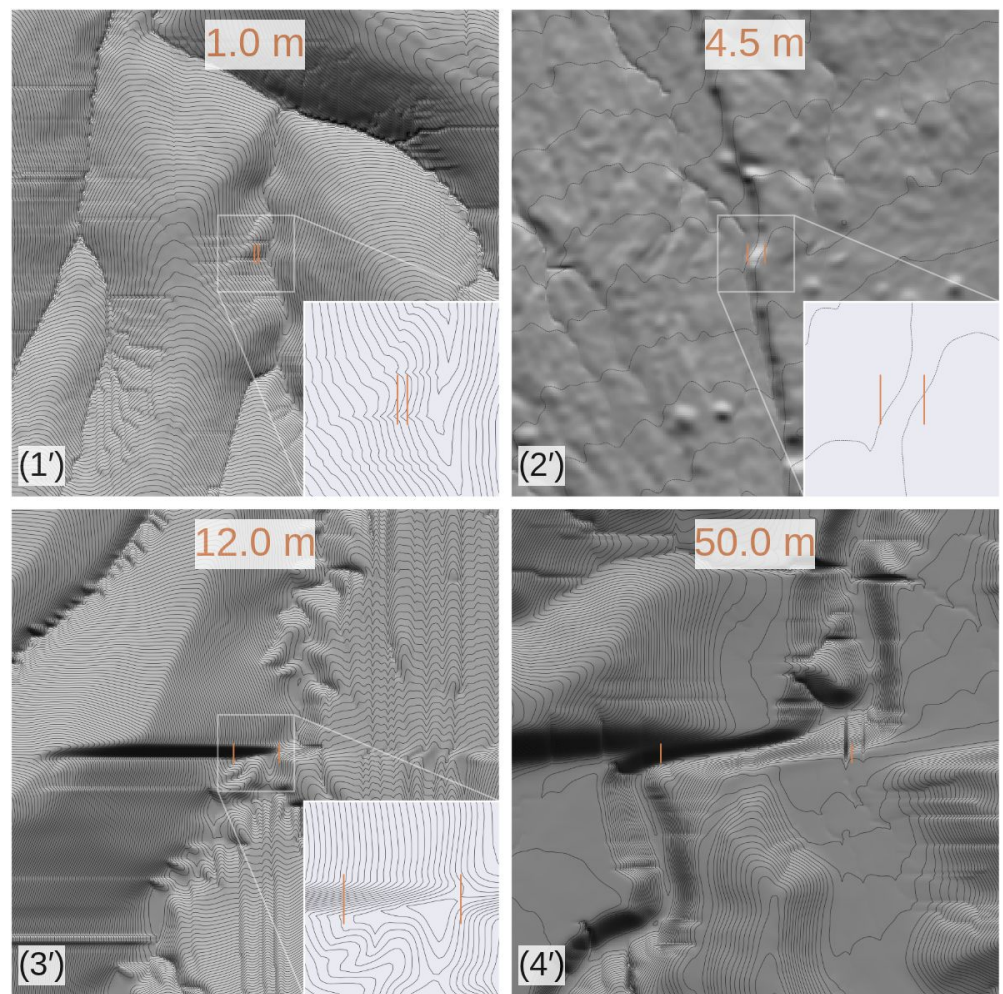


(Reitman et al. '19)

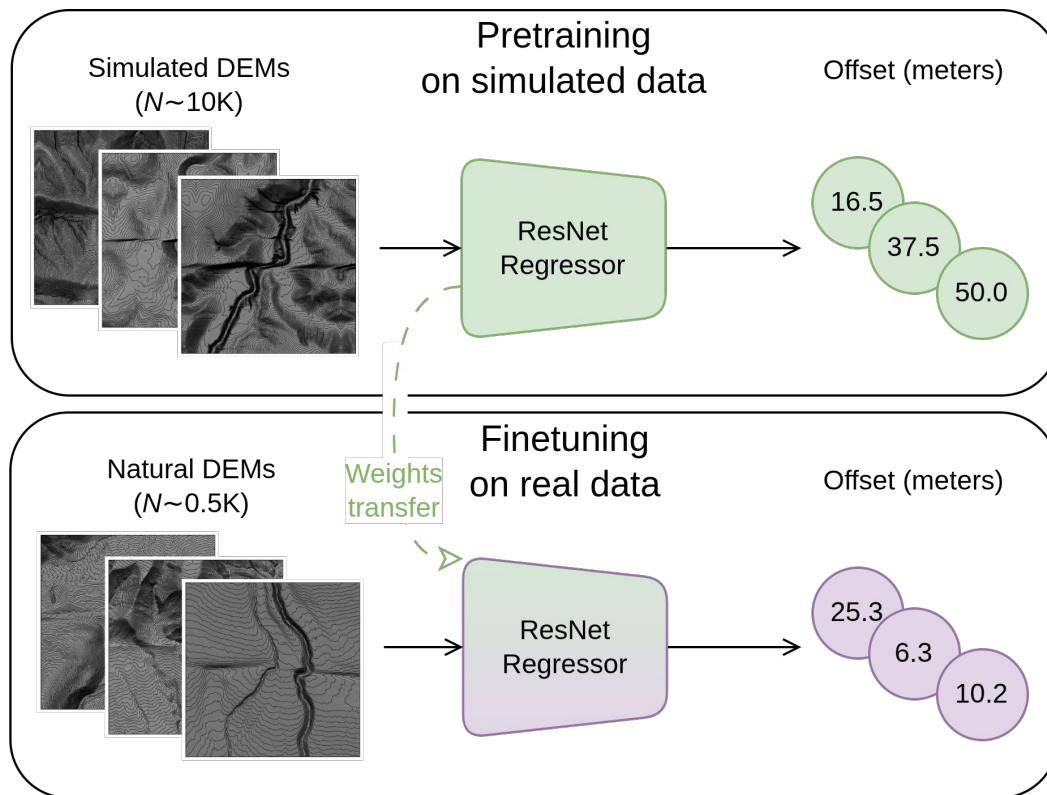
Synthetic data examples

- Generation of 10K samples
- Domain randomization approach: sample simulation parameter combinations

total time [yr]
slip rate [m/yr]
recurrence interval [yr]
fault zone width [m]
stream power coeffs.
diffusivity coeffs.
uplift rate [m/yr]
fault - river angle [deg]

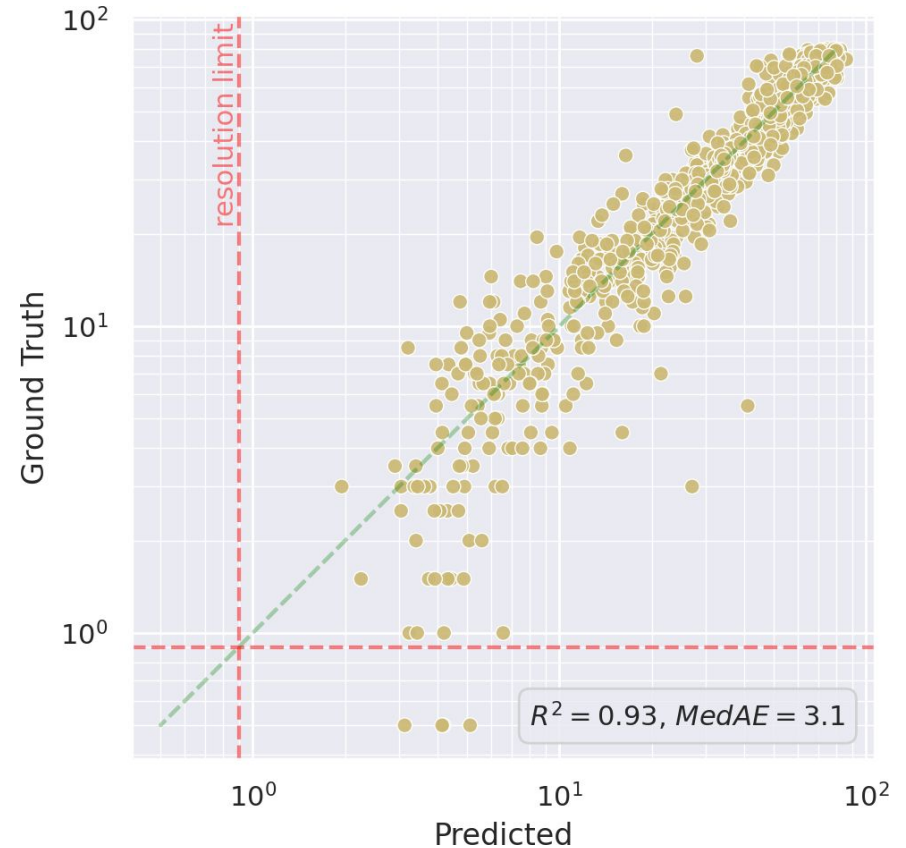


Simulated-to-real transfer learning



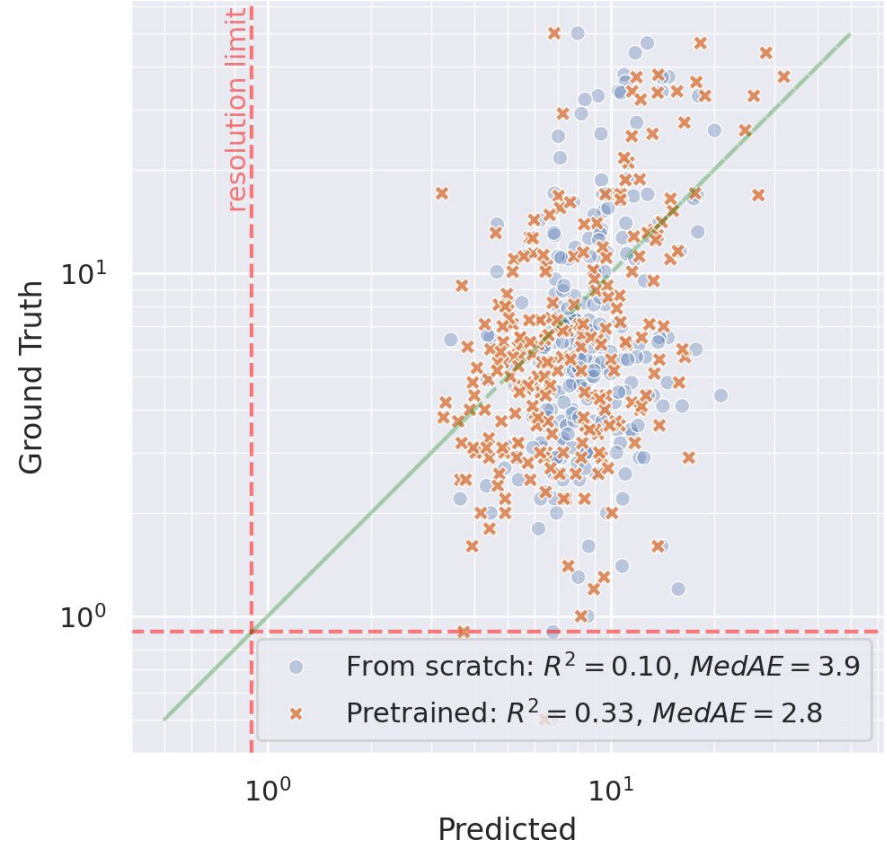
Results on synthetic test data

- Very good performance
- Model is able to generalize to new samples in the synthetic domain
- Resolution limit



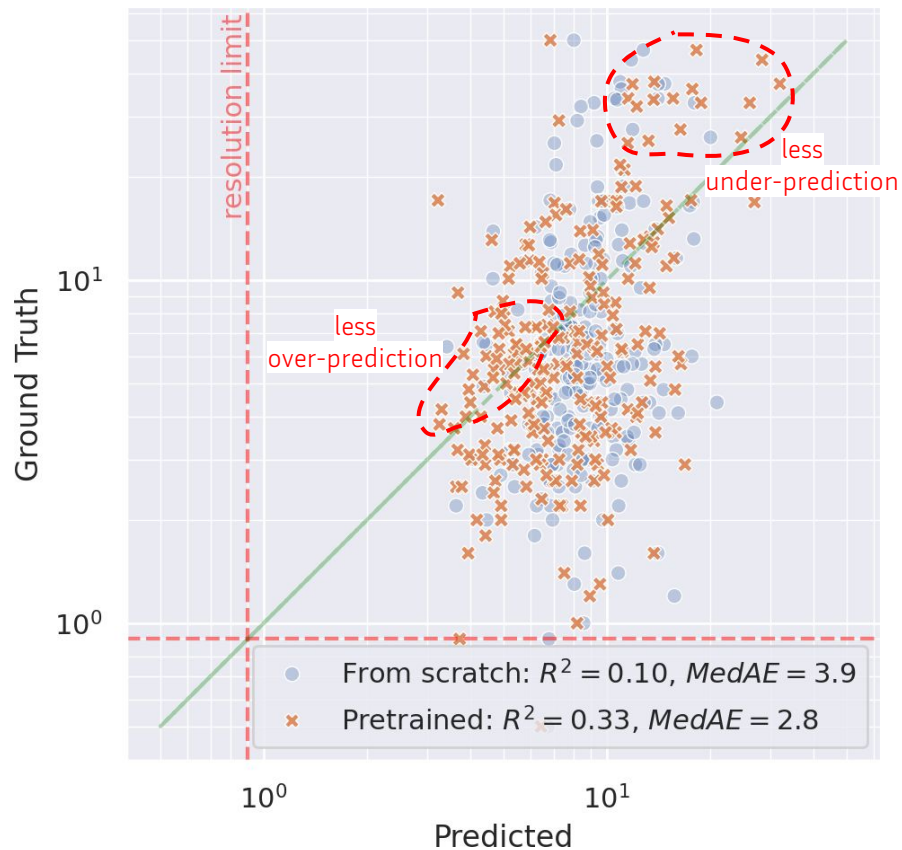
Results on natural test data

- From scratch model performs poorly
 - Lack of labeled data for training
- Pretraining improves R^2 by over 300% and MedAE by 28%
 - Representations learned on synthetic data transfer to some extent
- Still... A large performance gap with synthetic data !
 - Consequence of synthetic/natural data domain gap



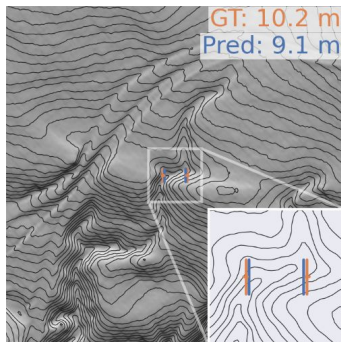
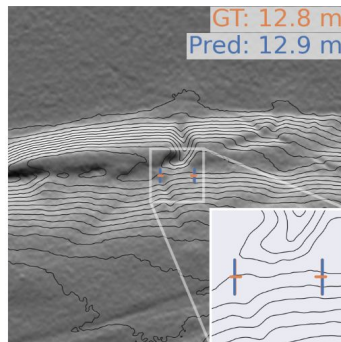
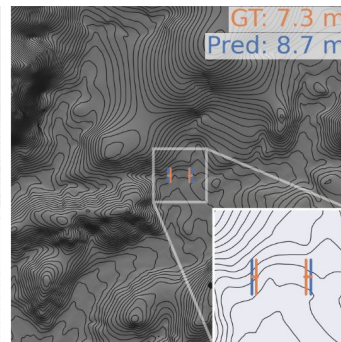
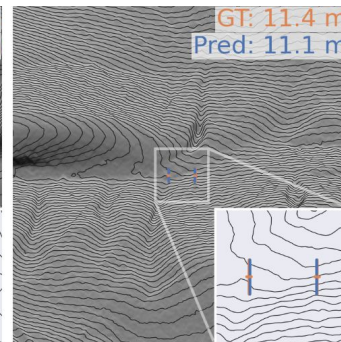
Results on natural test data

- From scratch model performs poorly
 - Lack of labeled data for training
- Pretraining improves R^2 by over 300% and MedAE by 28%
 - Representations learned on synthetic data transfer to some extent
- Still... A large performance gap with synthetic data !
 - Consequence of synthetic/natural data domain gap

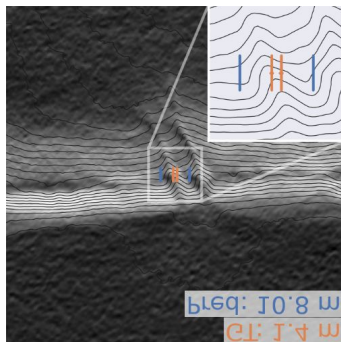
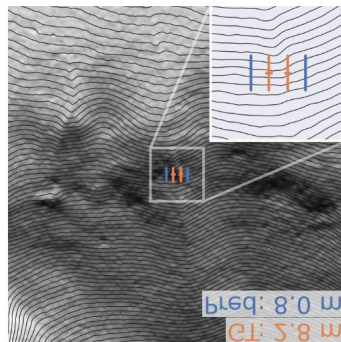
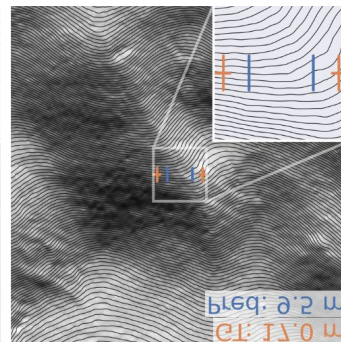
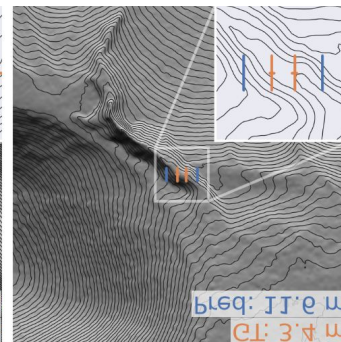


Sample predictions

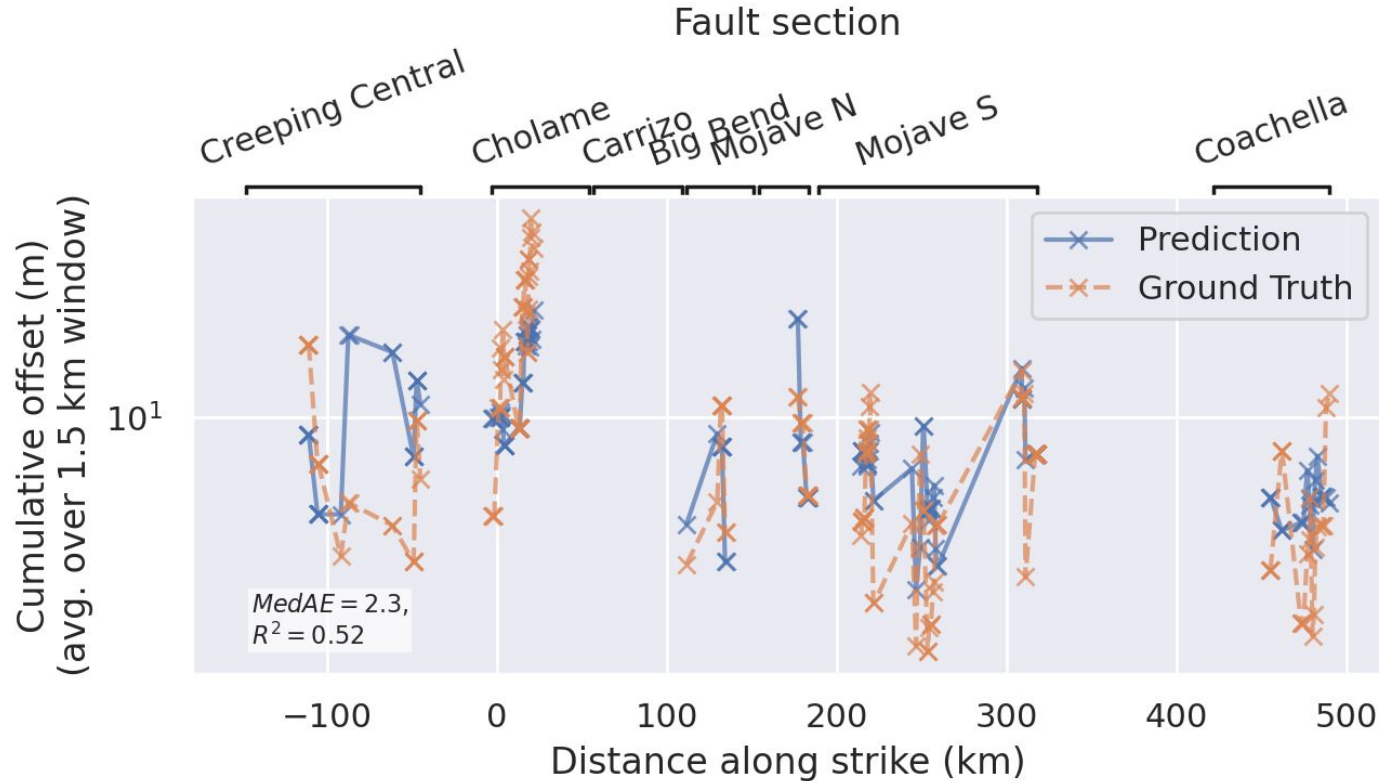
Correct predictions



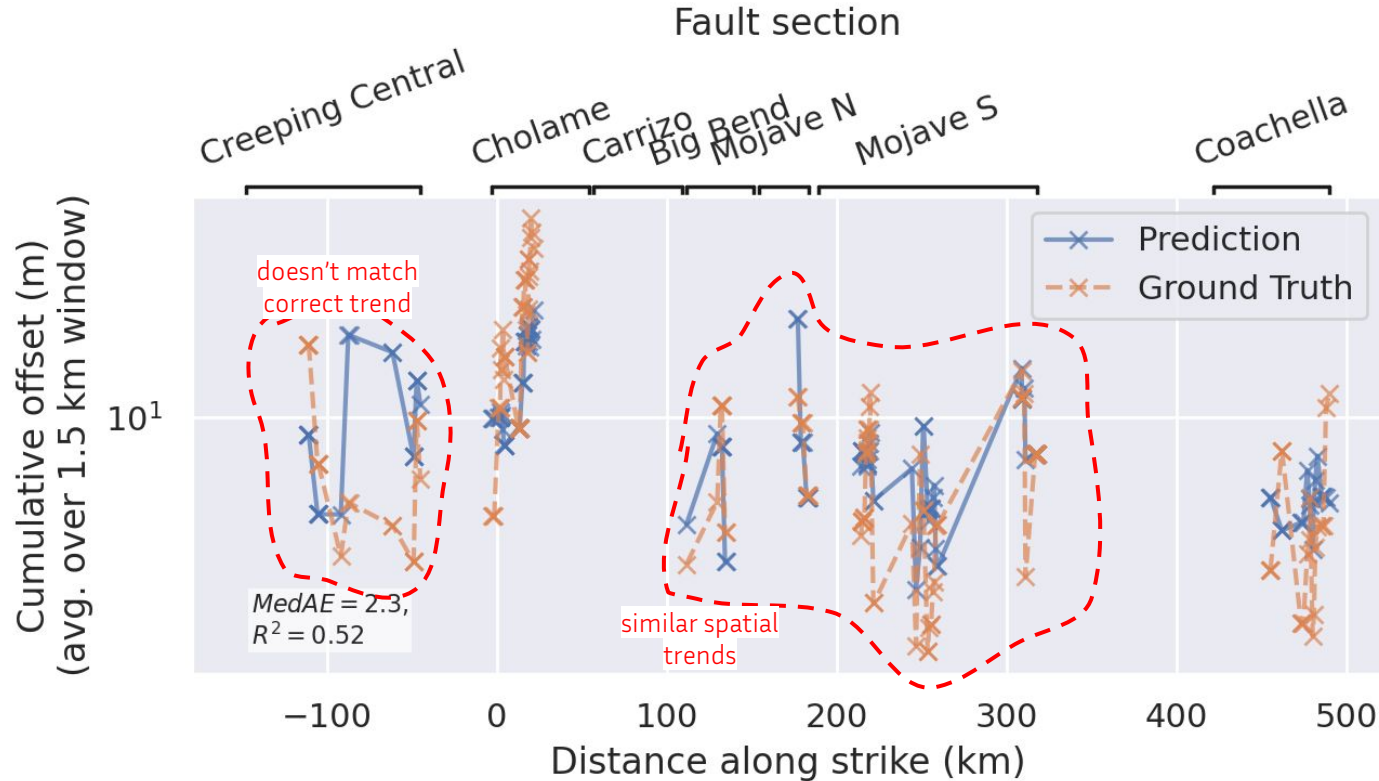
Incorrect predictions



Prediction of model along strike (San Andreas fault)



Prediction of model along strike (San Andreas fault)



Conclusion

- **Synthetic data** generation with **tectono-geomorphic landscape simulation**
- Deep neural network with **simulated-to-real transfer learning** to **overcome label scarcity**
- Potential of the method for **automatic strike-slip fault offsets characterization**
- Limitations: synthetic/natural **data domain gap, poorly-constrained problem**

Perspectives

- **Improving realism** of the synthetic data generation
- **Uncertainty quantification** (aleatoric and epistemic)
- Evaluation across **diversified faults** (e.g., New Zealand faults)
- Use **optical imagery** as a fallback modality
- **Detection + regression** of offset markers

Co-authors

Sarah Visage^{1,2} Sofia Goncalves¹ Sarah Perrinel¹ Margaux Mouchéné¹
Sophie Giffard-Roisin¹ Léa Pousse-Beltran¹

Thank you !