Accuracy and Precision of Earthquake Location Methods: Insights from a Synthetic Controlled Experiment

Yifan Yu, William Ellsworth, Greg Beroza

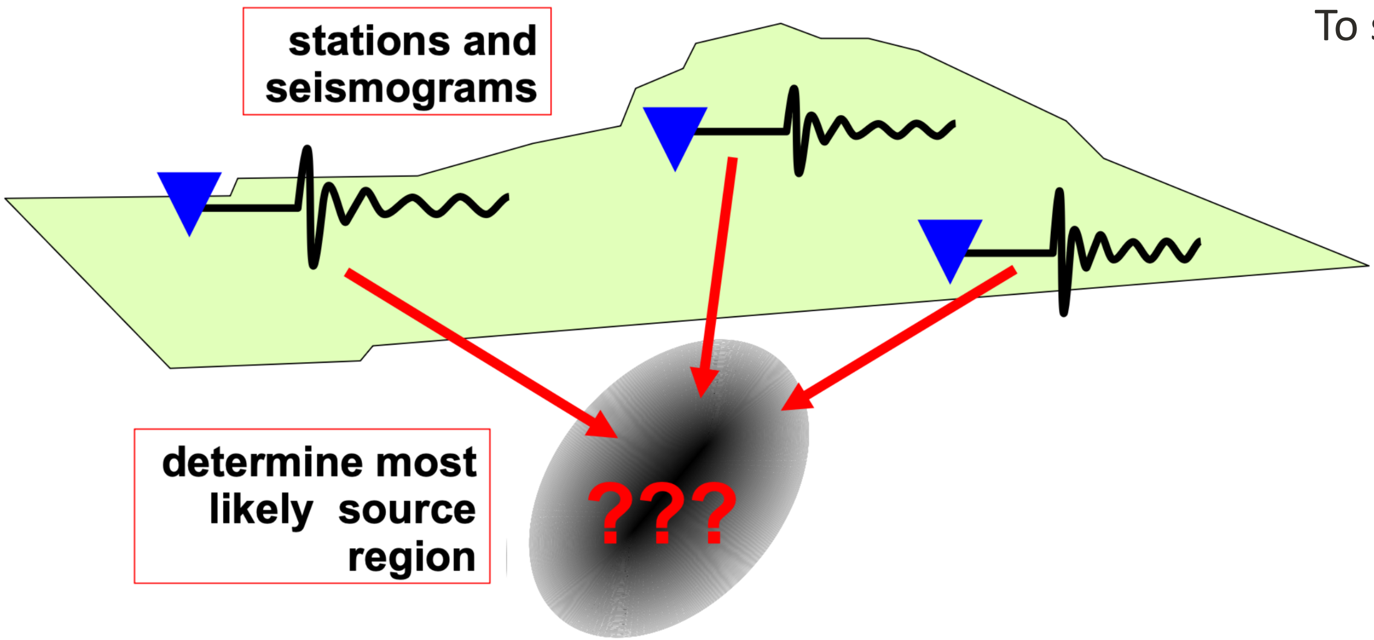
March 22th 2024

Stanford University

Introduction – Earthquake Location

A Well Defined Discrete, Nonlinear Inverse Problem -> hypocenter estimation

$$t_{arr} = t_0 + \int_r \frac{1}{v(s)} ds \qquad d_{obs} = (t_{arr}^1, \ldots, t_{arr}^n) = G(m) = G([x, y, z, t_0])$$



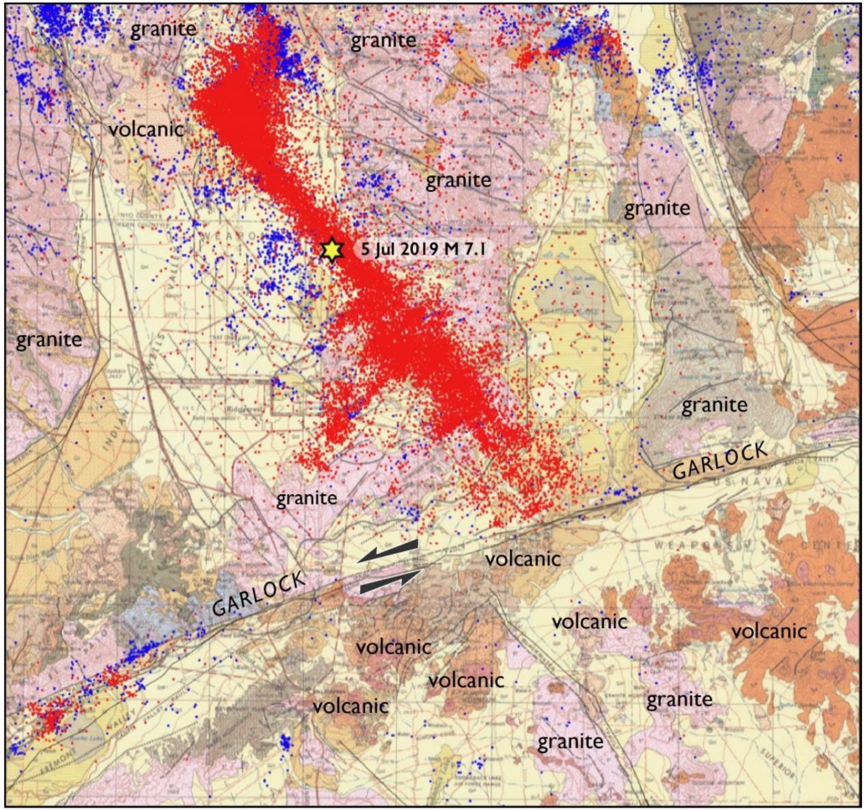
To solve we need:

- Traveltime
- Discretize
- Velocity

From: NonLinLoc

Introduction – Challenges

The unknown reality is complicated!



- Complex geology
- Varying Elevations
- Unmodeled 3D velocity structure
- Network Geometry
- Arrival time measurement errors
- ....

Geologic map of Ridgecrest area overlay with 2019 earthquake sequence.

From: Temblor.net

Various programs have been developed to correct these effects and recover locations

Q: How to Assess Location Accuracy and Error Estimates?

A: Create a benchmark data set of known locations that includes:

- Complex 3D velocity structure
- Measurement errors including outliers
- Missing data/stations
- Elevation variations and correlations with velocity structure

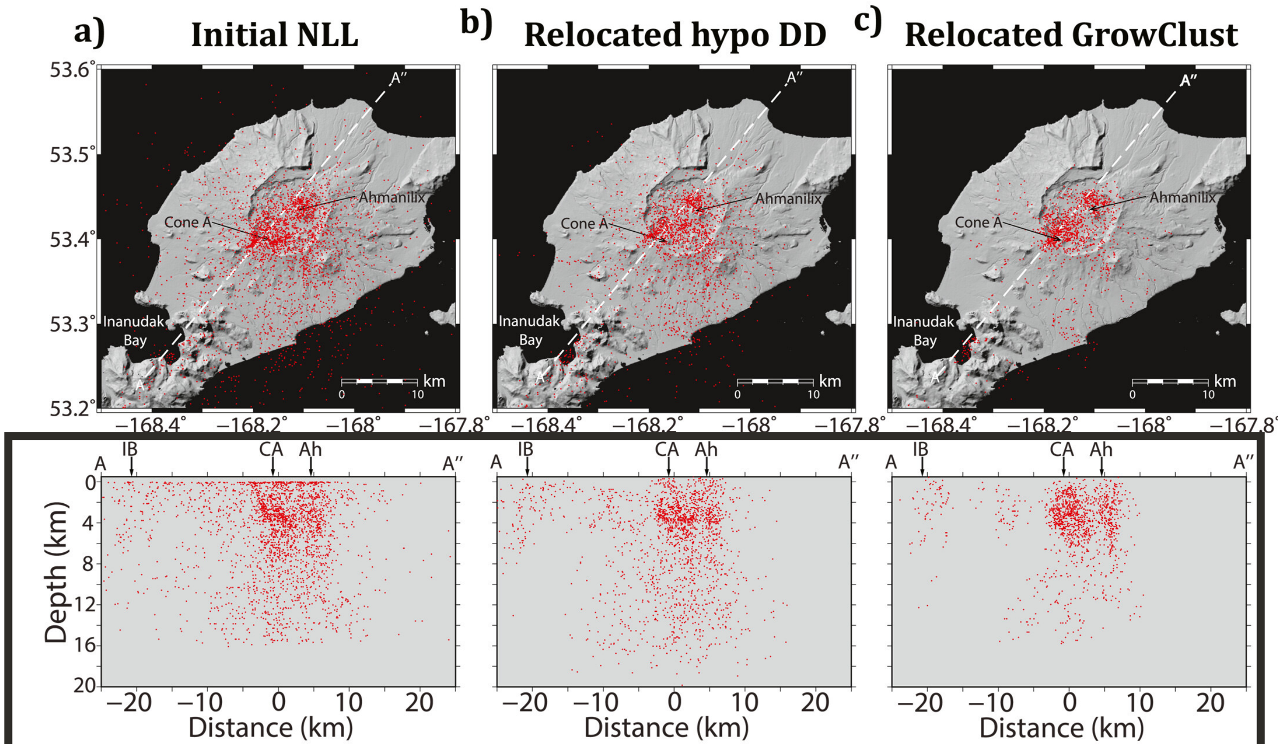
Apply different location algorithms and compare the results:

- locations
- error estimates

Apply location algorithms and compare against synthetic ground truth

Introduction – Understand the earthquake location results

The discrepancy comes from the different approaches!



Events' depths and cluster features vary from methods

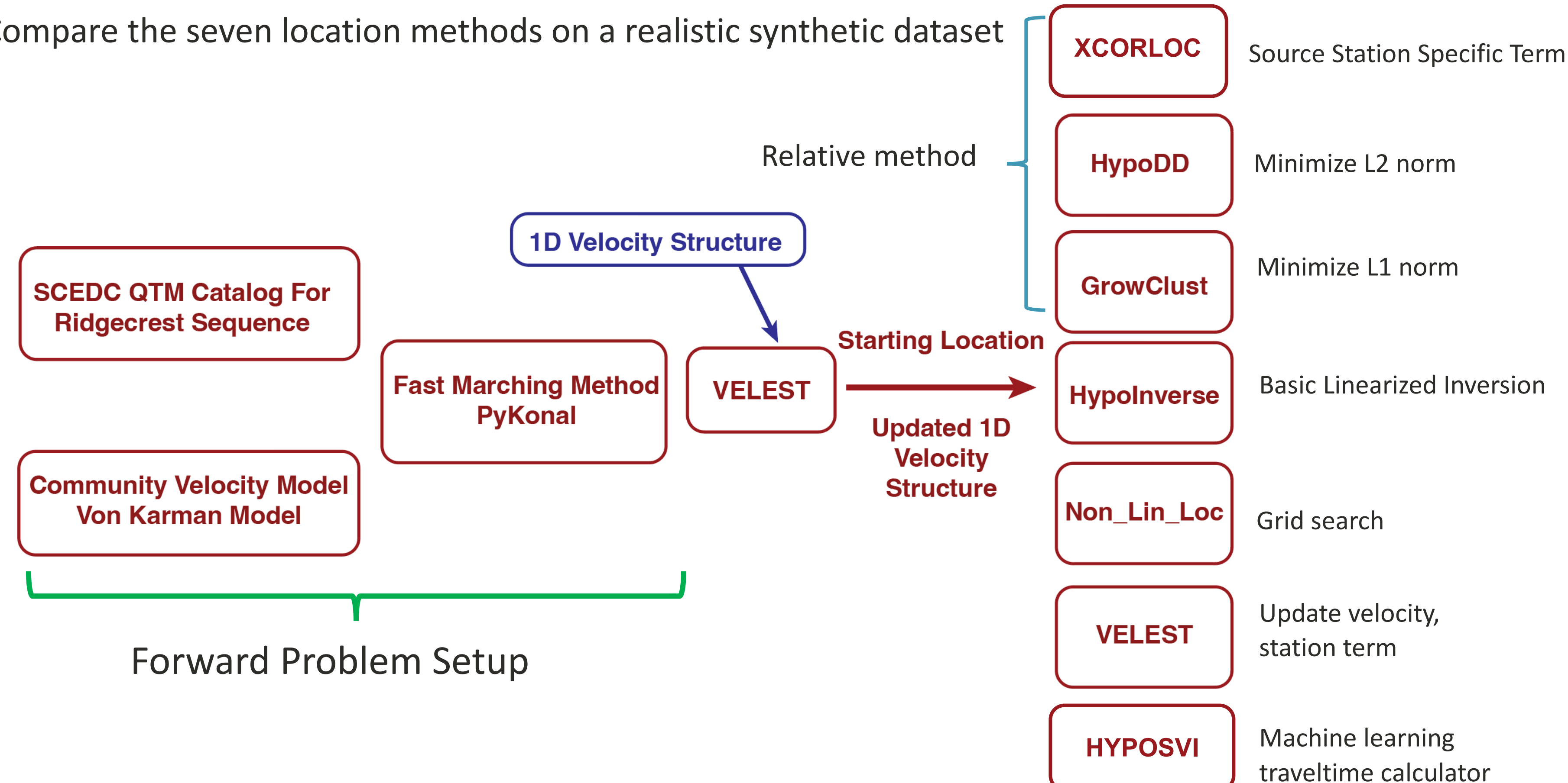
Earthquake locations in Okmok Volcano sequence output from three location programs.

Garza-Girón et al., 2023

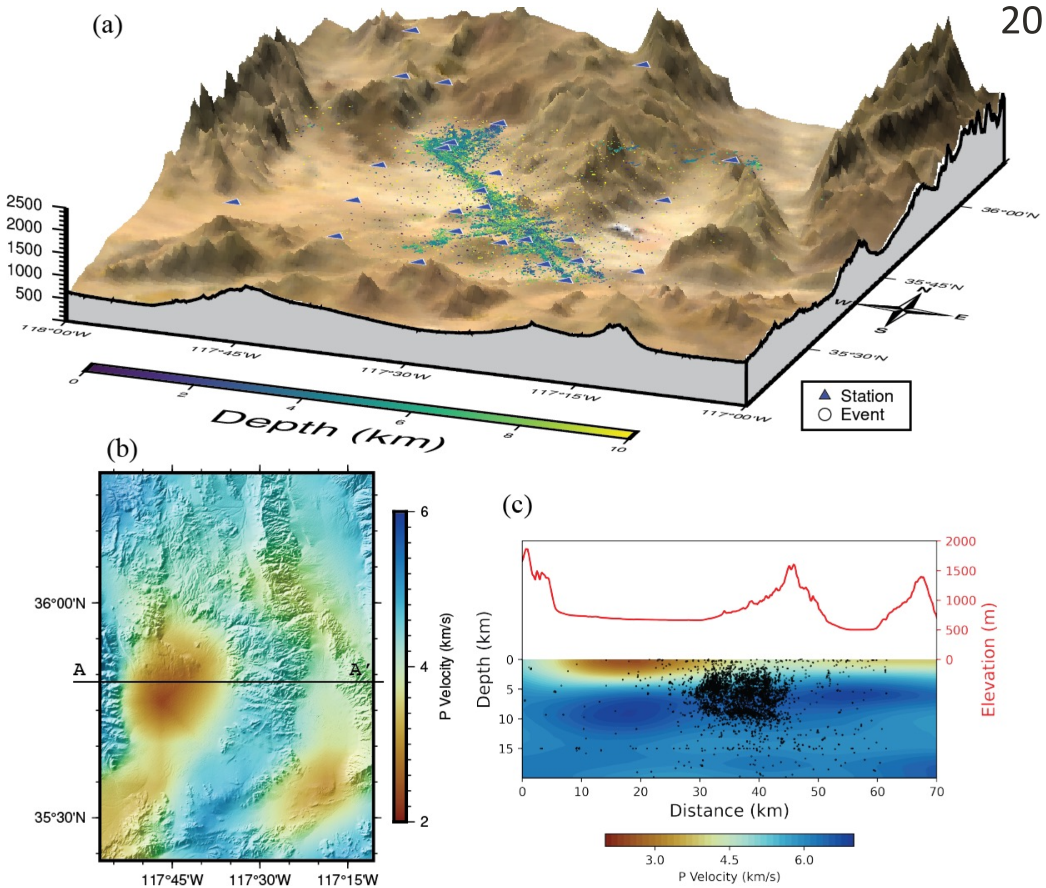
A realistic controlled experiment with ground truth can be used to assess the effectiveness of different algorithms for <u>estimating location</u> and <u>quantifying uncertainty</u>.

How to compare their performance?

Compare the seven location methods on a realistic synthetic dataset



Setup – Earthquake Location



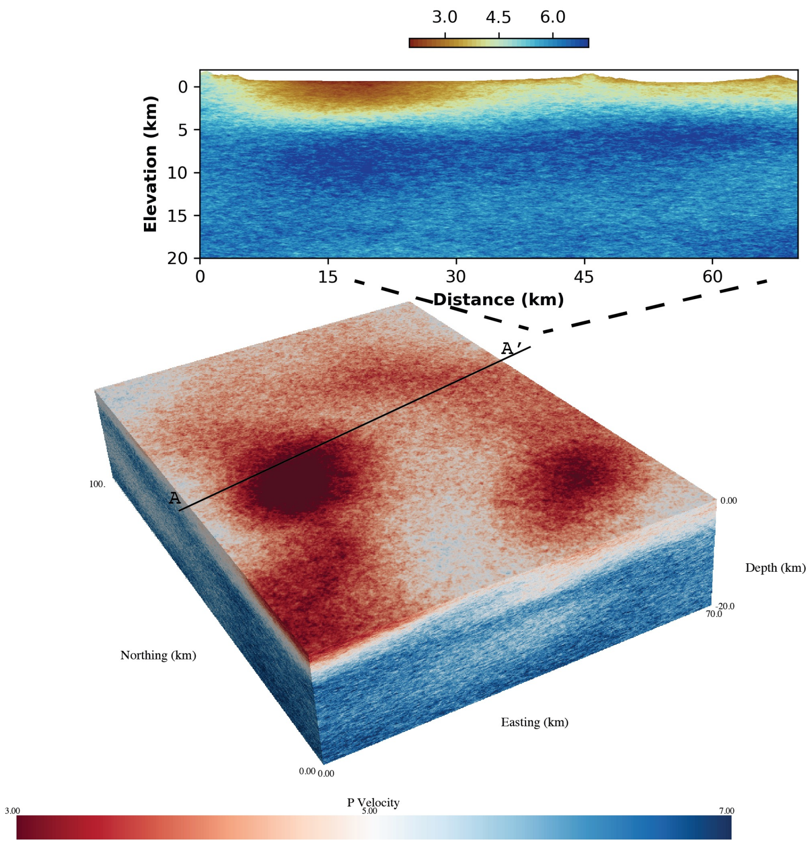
2019 Ridgecrest Earthquake Sequence – Data Mine

SCEDC QTM Catalog For Ridgecrest

We randomly draw 1000 sources out of the catalog (cluster > 400) as true locations

Real stations distribution from SCEDC

Setup – Velocity Structure



3D, Heterogeneous

- SCEC Community Velocity
- Von Karman Perturbation
  (a_z=100m, a_h=500m, $\epsilon = 0.05$)
- $70km \times 100km \times 20km\ depth +$
  $(max.2km\ elevation)$
  (dx=dy=100m, dz=50m)

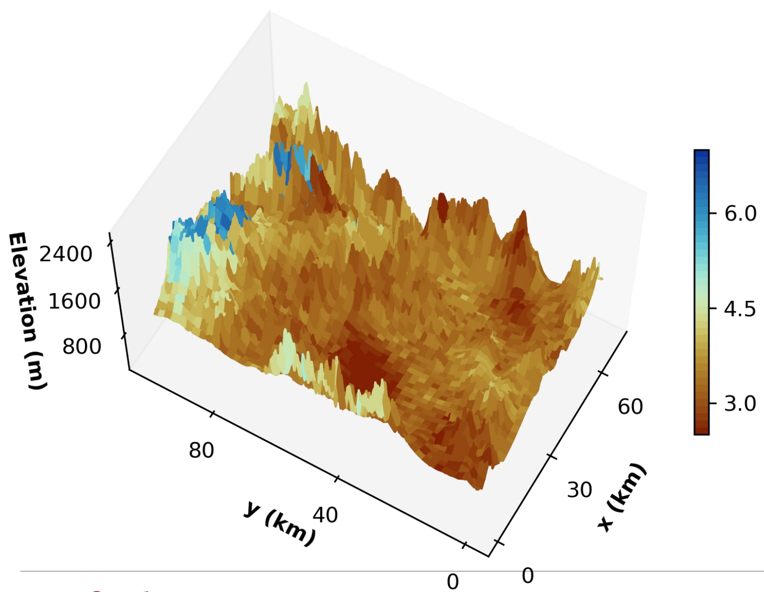
Setup – Elevation Effect and correlation with velocity

1. Classify as Granite or Sediments

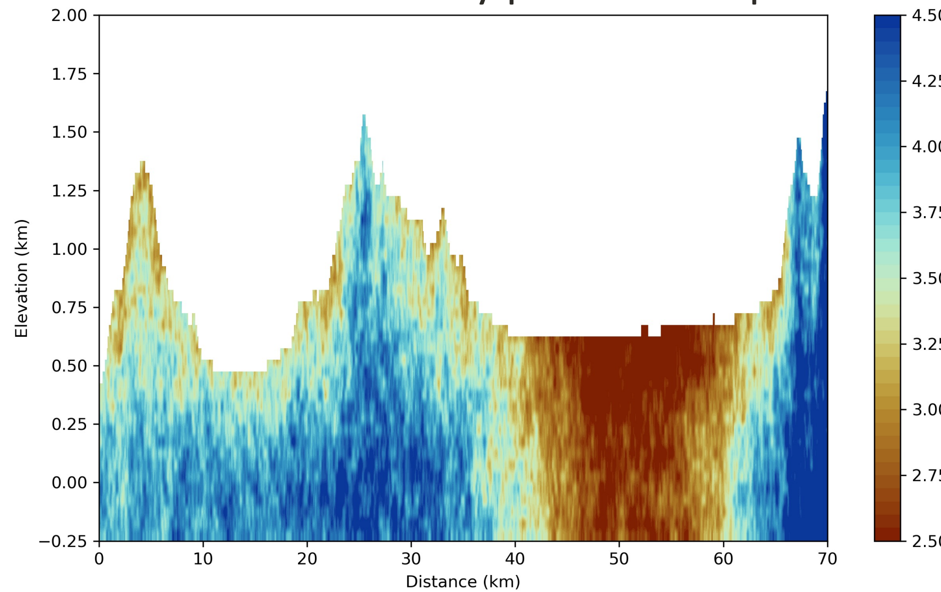
2. Velocity decrease with elevation in different slope based on empirical relations from experiment (Brocher, 2008)

Geologically realistic velocity model

Elevation Velocity example



Velocity profile example



Setup – Traveltime Calculator

Eikonal Function Based
Fast Marching Method: **PyKonal**

RESEARCH ARTICLE | JUNE 03, 2020

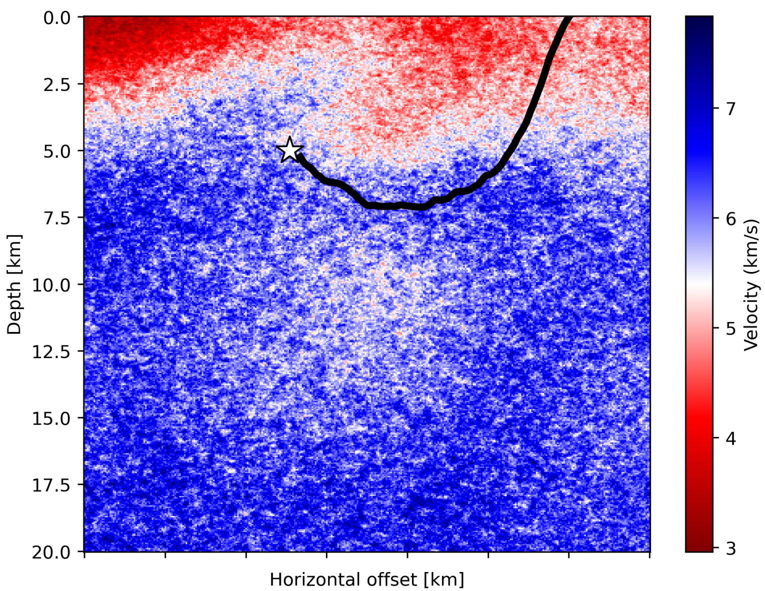
PyKonal: A Python Package for Solving the Eikonal Equation in Spherical and Cartesian Coordinates Using the Fast Marching Method

Malcolm C. A. White; Hongjian Fang; Nori Nakata; Yehuda Ben-Zion

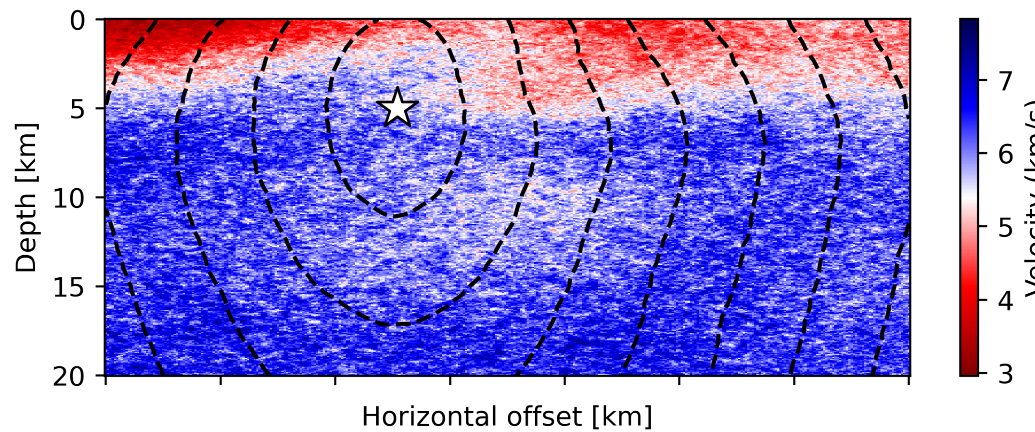
+ Author and Article Information

Seismological Research Letters (2020) 91 (4): 2378–2389.
https://doi.org/10.1785/0220190318 | Article history

Ray tracing verification on heterogeneous velocity model

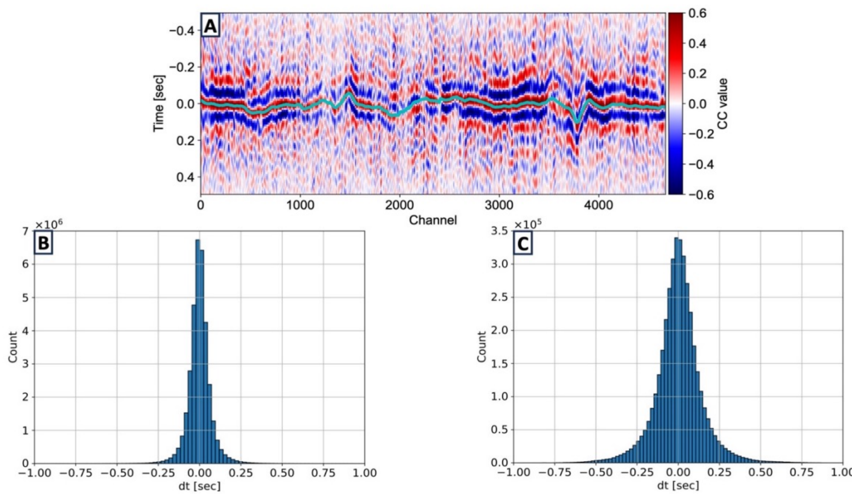


Traveltime contour verification



Setup – Realistic Error

Measurements are not perfect.



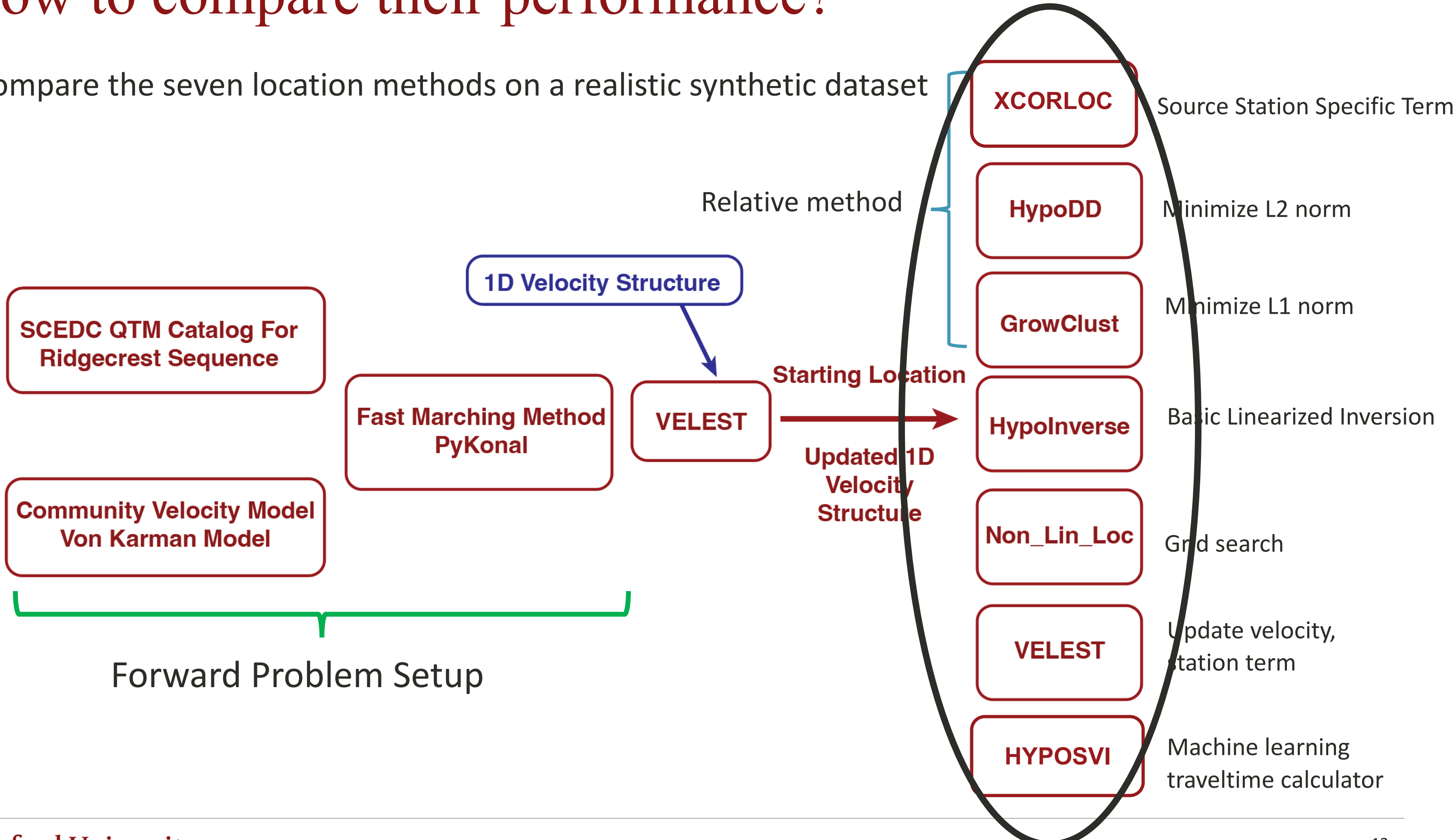
Biondi et al., 2023

- Realistic traveltime reading error following Laplacian distribution
- Variable phase availability for each stations: 0.67 for P, 0.5 for S
- 1% P phase and 4 % S phase outliers

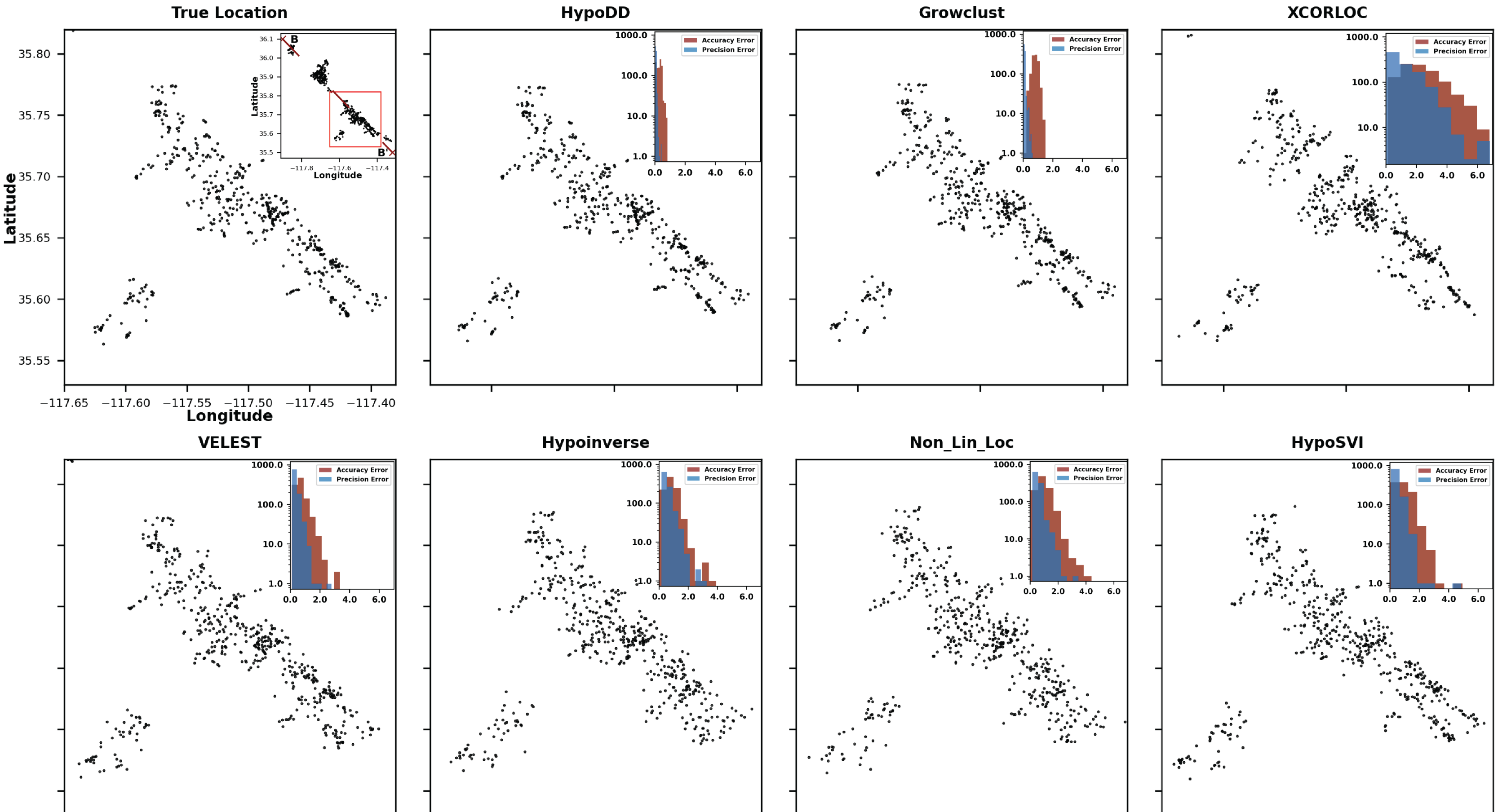
While both cross correlation and direct arrival phases error can be considered in Laplacian distributions, CC is more accurate than direct. And P phase is more accurate than S.

How to compare their performance?

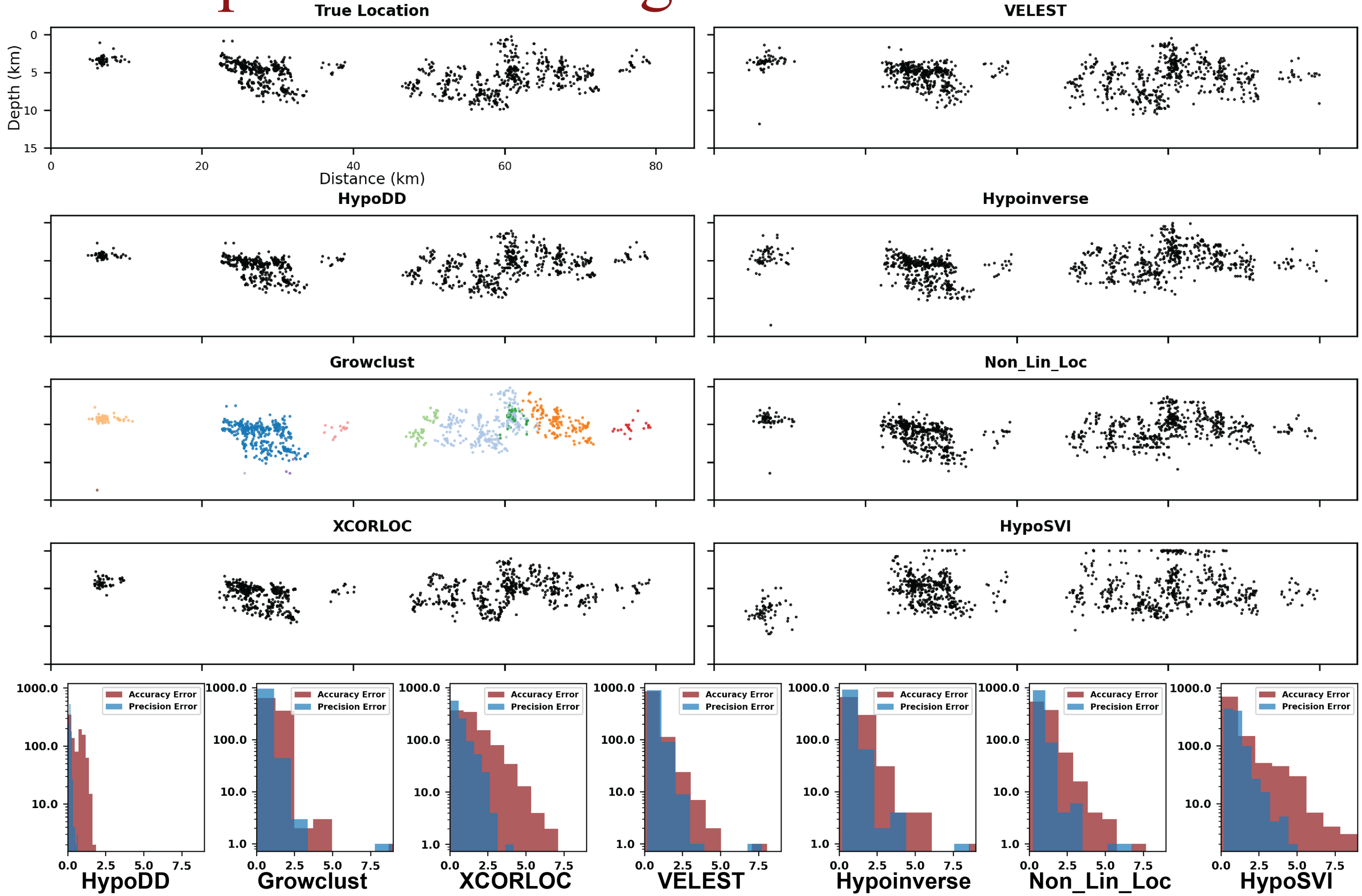
Compare the seven location methods on a realistic synthetic dataset



Results – Mapview



Results – Depth Profile along the Fault



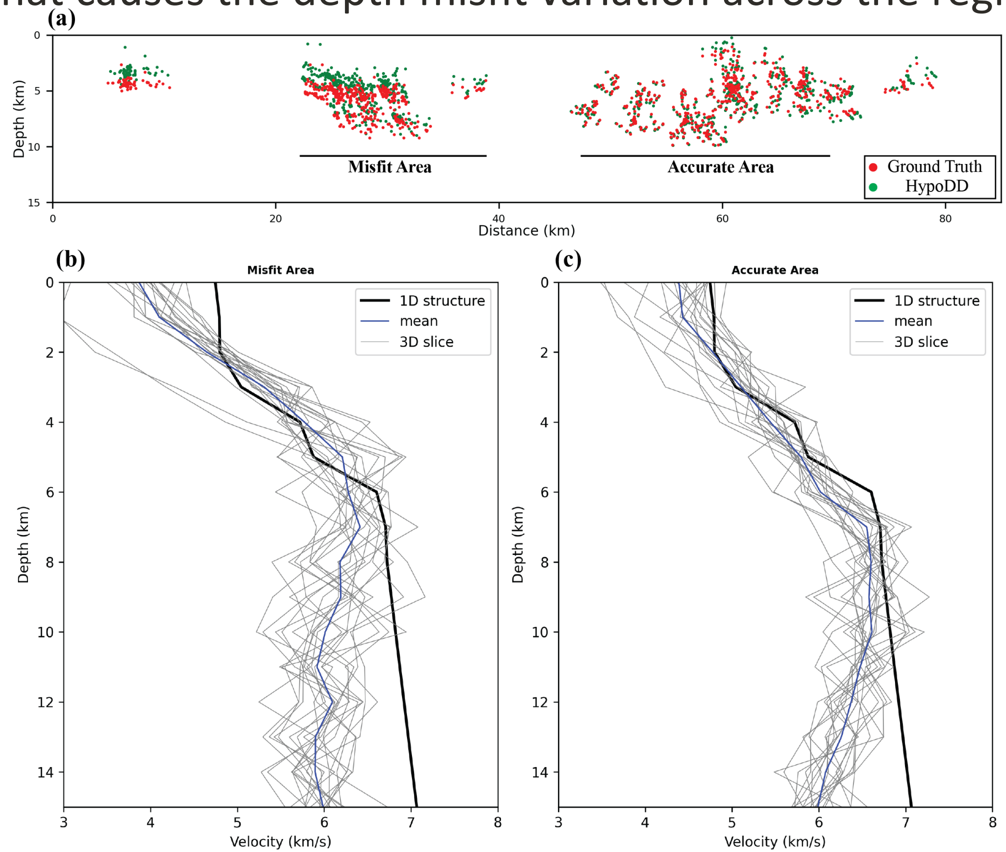
Results – Overall Performance

Table 1: Performance Summary

Method	Mean Accuracy Error (km)		Chamfer	Mean Precision Error (km)	
	Horizontal	Depth	Distance	Horizontal	Depth
Hypoinverse	0.824	1.118	1.617	0.571	0.684
VELEST	0.696	0.559	1.170	0.380	0.656
NonLinLoc	0.953	0.969	1.750	0.580	0.694
HypoSVI	0.911	1.152	1.854	0.463	0.970
HypoDD	0.289	1.124	0.946	0.081	0.151
XCORLOC	2.370	1.432	1.629	1.252	0.681
Growclust	0.802	1.071	1.702	0.113	0.303

Discussion – Limits of 1D Velocity Structure

What causes the depth misfit variation across the region?



- The velocity structure
- The station coverage

In accurate area -> 1D velocity model better approximates the 3D model.

Discussion – Uncertainty Analysis

Different Approaches:

1. Bootstrapping: Growclust, XCORLOC
2. Estimating from likelihood function: NonLinLoc, HypoSVI
3. Covariance Matrix: Hypoinverse

Table 2: Ratio of events/pairs within 95% confidence error ellipse of each program

Program	Hypocenters	Event Pairs	Chamfer Distance
Hypoinverse	14.6 %	62.8 %	1.617
VELEST	64.4 %	98.3 %	1.170
NonLinLoc	80.5 %	87.4 %	1.750
HypoSVI	28.1 %	79.5 %	1.854
HypoDD	n/a	80.6 %	0.946
XCORLOC	n/a	0.3 %	1.629
Growclust	n/a	91.2 %	1.702

Conclusion

- This controlled experiment proves the effectiveness of the relative location methods over absolute one, while VELEST and HypoDD stands out.

- 1D velocity structure is limited to represent the realistic situation.

- Our findings should motivate revisiting earthquake location uncertainty assessment for both accuracy and precision.

- Such synthetic tests can be easily extended to other scenarios at local/regional/teleseismic distances.

All the code is on my github page:
https://github.com/YuYifan2000 ,
Try do the locations yourself!

Thank You

Discussion – Near Source S Phase

Gomberg et al. (1990) argues that the S phase within distance of 1.4 * source depth is important for constraining the depth

without one random near S phase V.S. without one random far S phase

VELEST	Mean Accuracy Error (km)		Mean Precision Error (km)	
	Horizontal	Depth	Horizontal	Depth
Without near S	0.64	0.61	0.39	0.70
Without far S	0.58	0.51	0.35	0.62

Introduction – Hypocenter Locators

$$R = G * dM$$

$$dM = (G^T G)^{-1} G^T R$$

Hypoinverse An iterative linearized inversion scheme

VELEST

$$t_{res} = \underbrace{H * h}_{\text{Hypocenter}} + \underbrace{M * m}_{\text{Velocity Model}} + \overbrace{e}^{\text{Constant Error}} = Gm + e$$

Non_Lin_Loc Nonlinear inversion: Grid Search

EDT (Equal Differential Time)

$$pdf(x) \propto \boxed{\sum_{obs_a,\, obs_b}} e^{-\frac{\left[\boxed{\left[Tobs_a(x) - Tobs_b(x)\right]} - \boxed{\left[TTcalc_a(x) - TTcalc_b(x)\right]}\right]^N}{\sigma^2}}$$

"satisfy ***the most pairs*** of observations"

- no origin time estimate required

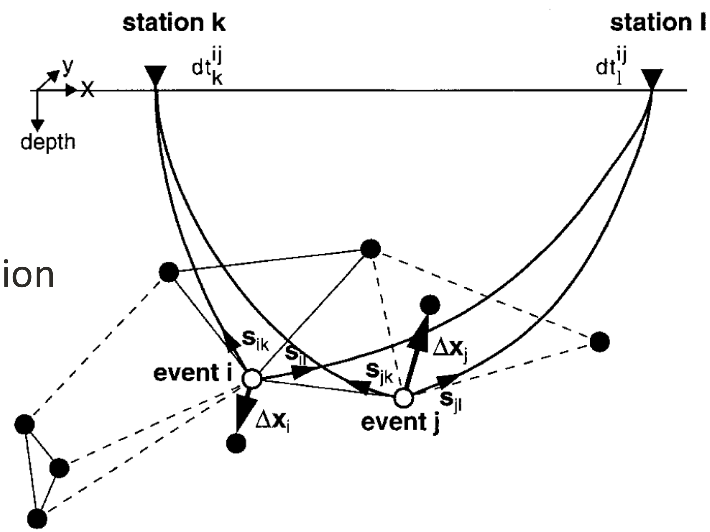
HypoSVI Machine Learning model!
Grid search with ML traveltime calculator

Introduction – Hypocenter Locators

HypoDD

$$dr_k^{ij} = (t_k^i - t_k^j)^{obs} - (t_k^i - t_k^j)^{cal}$$

$$\frac{\partial t_k^i}{\partial \boldsymbol{m}} \Delta \boldsymbol{m}^i - \frac{\partial t_k^i}{\partial \boldsymbol{m}} \Delta \boldsymbol{m}^j = dr_k^{ij}$$



Growclust

L1 Norm, more resilient to outliers
Traveltime table search instead of matrix inversion

XCORLOC

Similar to Growclust
But with Source Station Specific Term