EAGE

INCLUDING SPE EUROPEC

78th EAGE Conference & Exhibition

VIENNA 2016

Efficient Use of Technology - Unlocking Potential

30 May – 2 June 2016

www.eage.org/events/vienna-2016

An Automatic Arrival Time Picking Method Based on RANSAC Curve Fitting

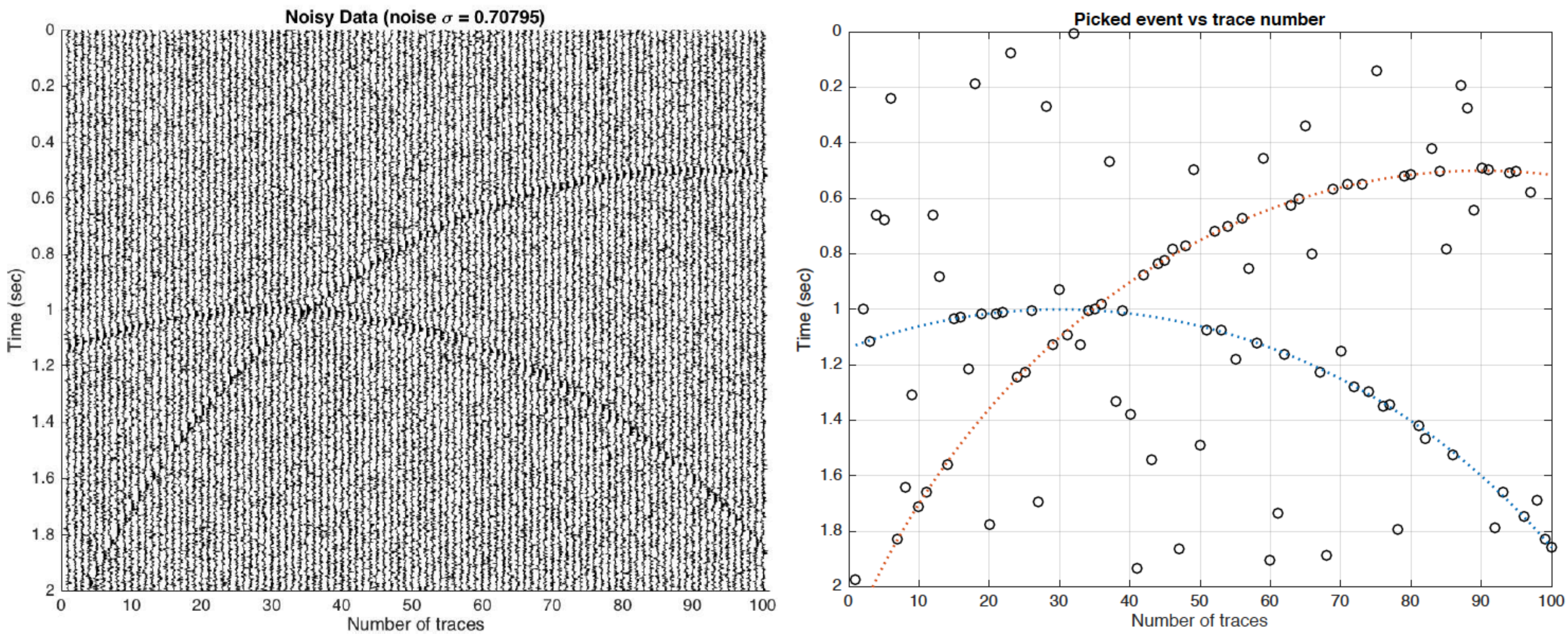
Lijun Zhu, Entao Liu, J. H. McClellan

Jun 2nd, 2016

Outline

- Motivation
- **RANSAC Method**
  - **Extensions for arrival time picking**
- Simulations for Linear Arrays
- 2-D Array Scenario
- Conclusion

Event Picking Problem



(a) (b)

Figure 1: Synthetic example: a) two seismic events with additive white Gaussian noise (σ = 0.708); b) picked cross-correlation maxima (dotted lines indicate the true moveout curve).

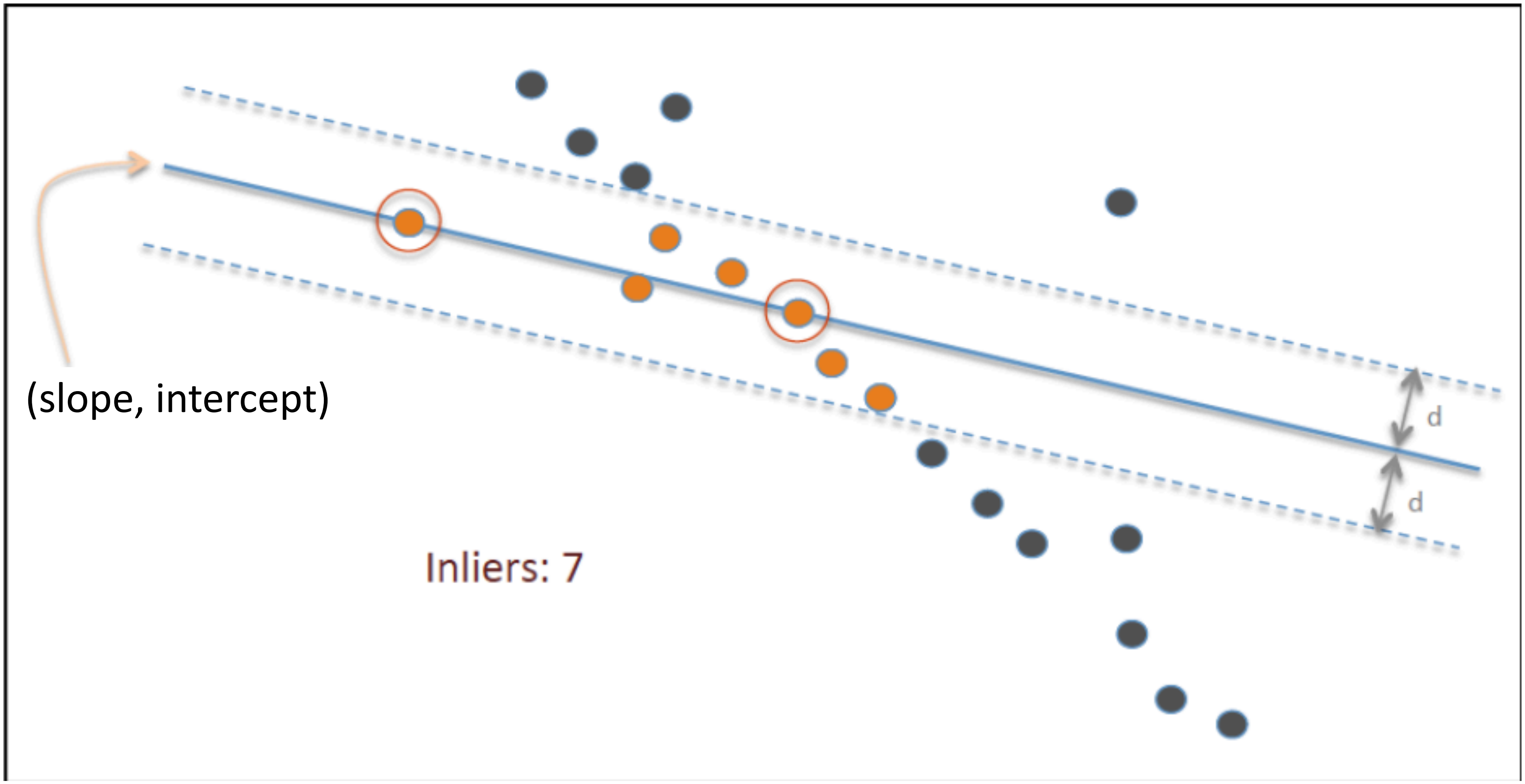
Challenge for Traditional Methods

- Missed (or extra) picks are common in low SNR
  - Need to deal with a large number of **<u>outliers</u>**
- Kalman filter based tracking algorithm has been investigated by [Deng* and Zhang, 2015]
  - Sensitive to the accuracy of the picks on the initializing traces (Good for active monitoring)
- Curve fitting after eliminating outliers
  - Fit low-order parametric model to the picks

<u>RAN</u>dom <u>SA</u>mple <u>C</u>onsensus (RANSAC)

- Goal: Find a small set of parameters of a simple model that fits the observed data, i.e., the picks
- Distinguish outliers from “inliers” in the data
- Five-step resample solution: [Fischler and Bolles, 1981]
    - Pick a small number of data points at random
    - Determine best fit model
    - Evaluate the current model over all traces
    - Count number of inliers within a specified inlier distance ***d***
    - Repeat many times and save inlier-ratio for each model
    - Choose the model with highest inlier-ratio

RANSAC for Best Line Fit



Required Number of Iterations

- The number of iterations, ***N***, needs to be high enough to ensure that at least one set of random samples does not include any outliers, with probability ***p*** (usually ***p*** = 0.99)
- Let ***u*** be the probability that any chosen sample is an inlier and ***v = 1 - u*** be the probability it is an outlier
- Given by [Fischler and Bolles, 1981], for ***m*** parameters,

$$1 - p = (1 - u^m)^N$$

Assume $u \cong 0.6$ and $m = 5$

- Thus,

$$N = \frac{\log(1-p)}{\log(1-u^m)}$$

$N = 57$

- Also, the standard deviation of ***N*** is given as

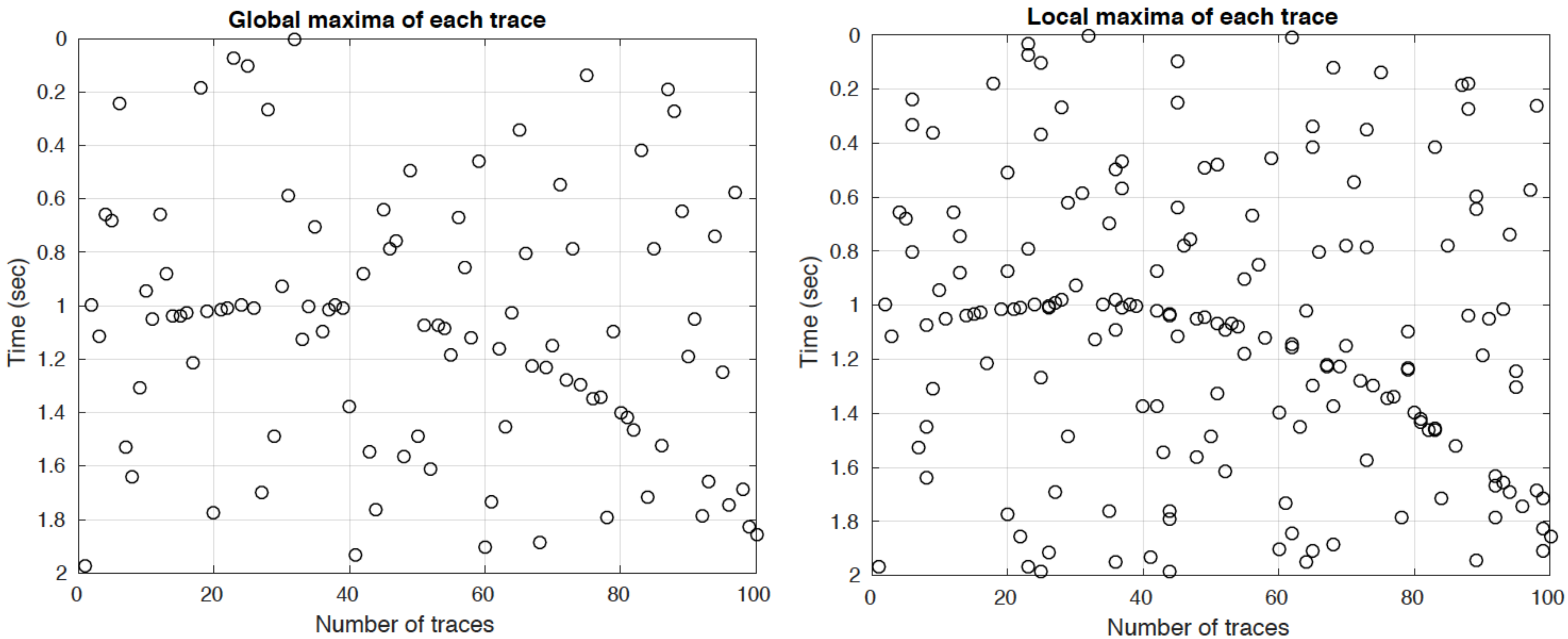
$$\text{std}(N) = \frac{\sqrt{1-u^m}}{u^m}$$

std(N) = 12

RANSAC for the Time Picking Problem

- Feature extraction: Multiple picks per trace will encourage feature extraction to find more inliers (while generating more outliers as well)
- Model selection: use enough parameters to describe the set of possible expected curve shapes
- Fitting method: Pseudo-inverse if ill-conditioned
- Accept model: if "inlier ratio" is greater than 10%
- Extend RANSAC to include refitting with all inliers by over-determined least squares
- Extend RANSAC to deal with multiple models for simultaneous events

Feature Extraction



(a) (b)

Figure 2: Feature extraction (PSNR = 3dB) using (a) global maximum of each trace, (b) local maxima above a threshold (θ = 0.95) on each trace.

Model Selection

- Hyperbolic curve/surface for perfect layered model
- Quadratic equation models rotation and stretch

$$Ax^2 + Bxy + Cy^2 + Dx + Ey = 1$$

- Works for all linear arrays (uniform/non-uniform)
- Vectorize observed data (x1; y1) ... (xn; yn):

$$A = \begin{bmatrix} x_1^2 & x_1 y_1 & y_1^2 & x_1 & y_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n^2 & x_n y_n & y_n^2 & x_n & y_n \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \text{ and } \quad \boldsymbol{x} = \begin{bmatrix} A \\ \vdots \\ E \end{bmatrix}$$

- Thus, $A\boldsymbol{x} = \boldsymbol{b}$

Fitting, Refitting and Multiple Models

- Use pseudo-inverse for stable solution:

$$\hat{\boldsymbol{x}} = (A^T A)^{-1} A^T \boldsymbol{b}$$

- Refit all inliers (with respect to the selected parameters) to reduce misfit due to random noise.

- After finding the first model (curve), remove all inliers in that model and re-run the algorithm for more possible models.

Algorithm Summary

Algorithm 1: RANSAC finds the best curve that fits a hyperbola out of (x, y) samples

Input: Data points from feature extraction

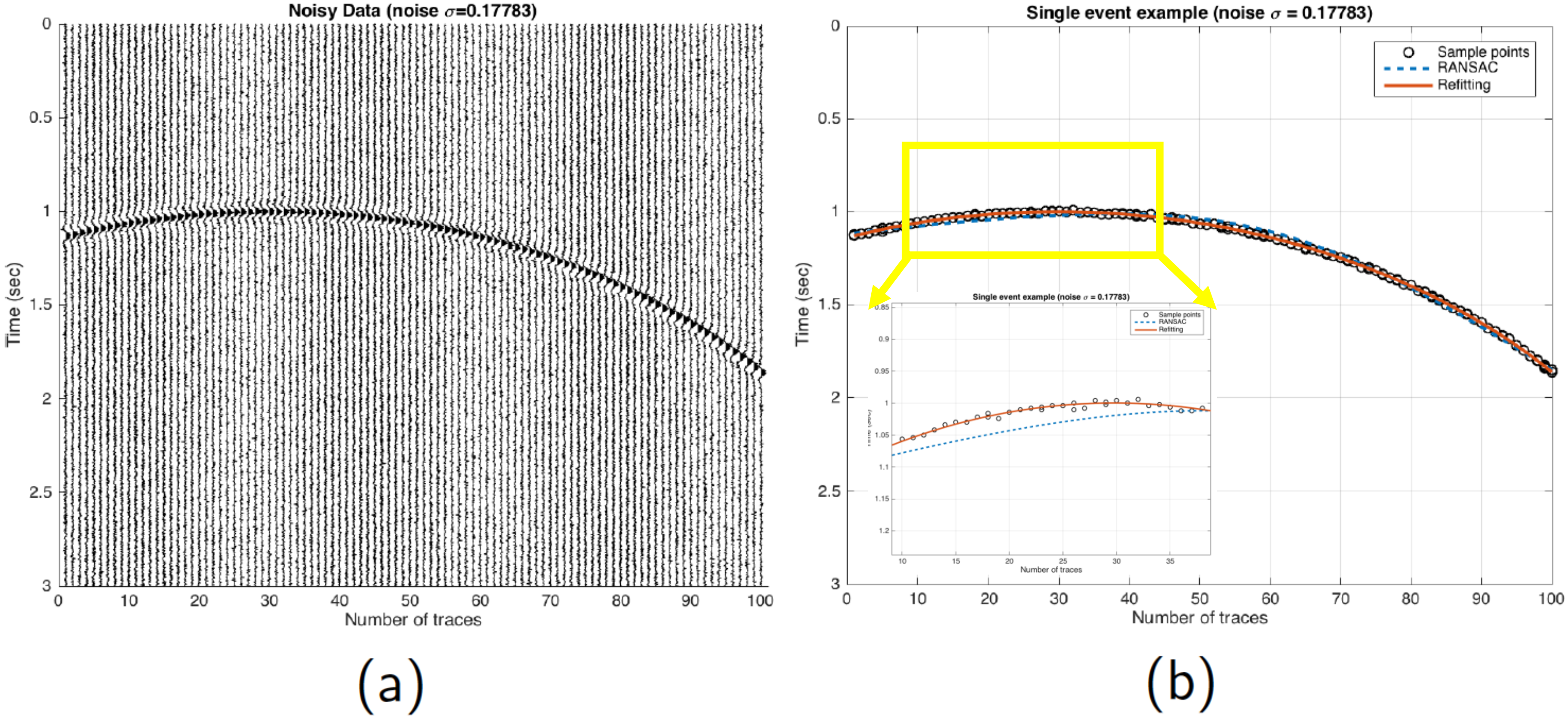
Output: Parameters vector x^* for quadratic equation in (4)

```
Best fit curve x* ← [0, 0, 0, 0, 0]^T
Constant vector b ← [1, 1, 1, 1, 1]^T
Number of inliers of best fit curve n* ← 0
for i ← 1 to # of iteration do
    Randomly select five sample points and form data matrix A as in (5)
    Least-squares fitting: x_i ← (A^T A)^{-1} A^T b
    Count number of inliers n_i based on distance between fitted curve and all sample points
    if n_i > n* then
        Save all inliers
Use saved inliers to form data matrix A* as in (5)
Use least-squares approximation to fit all saved inliers:
    x* ← (A*^T A*)^{-1} A*^T b
return x*
```

Examples

- Synthetic examples
  - Ricker wavelet + additive noise
  - Single event
  - Multiple events
- Seismic event with P- and S-waves
  - Synthetic delays
  - Additive noise (AWGN)
- 2-D array synthetic example

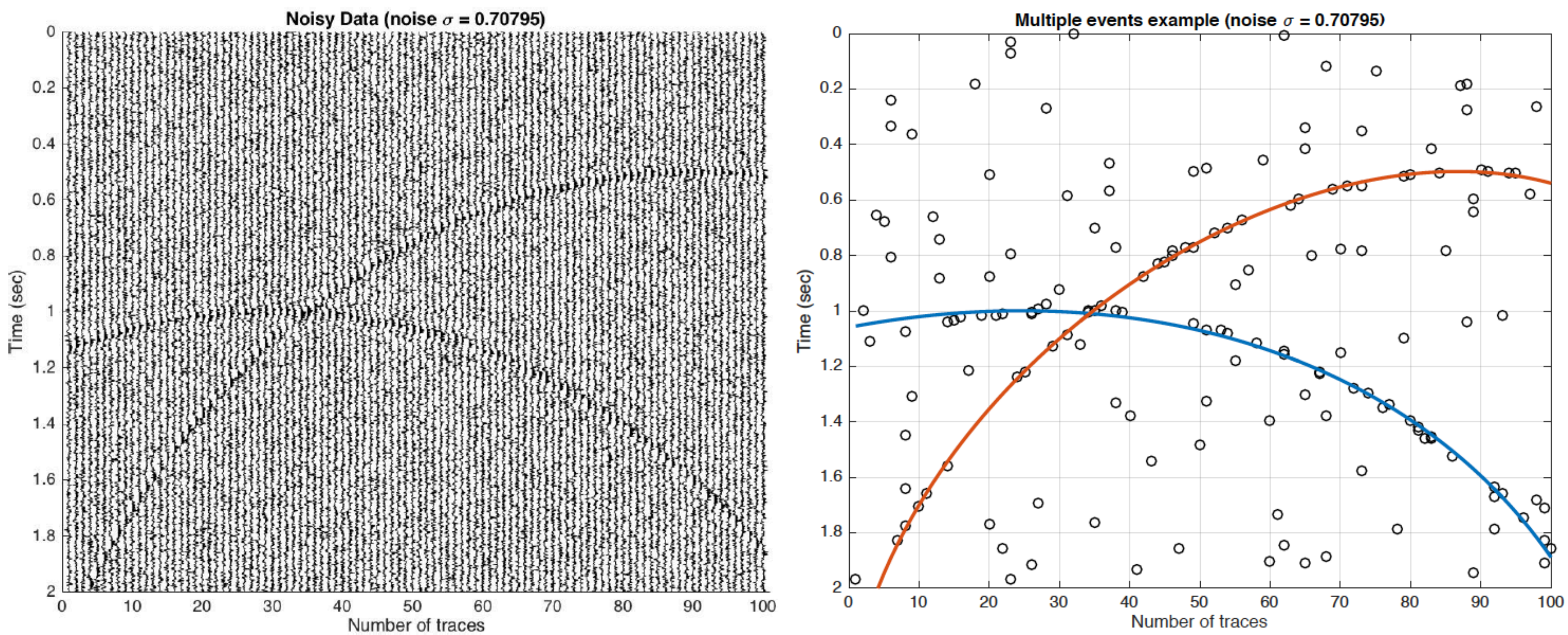
Single event with low noise



(a) (b)

Figure 3: Curve fitting results: a) input data with noise level at σ = 0.178, and picks with threshold θ = 0.95, b) single event fitting results: RANSAC fitted curve (blue) and refitting curve (red).

Multiple Events Simulation



(a) (b)

Figure 4: Curve fitting results: (a) input with noise level of σ = 0.708, and picks above θ = 0.95, (b) multiple events: red (1st event) and blue (2nd event) curves indicate two distinct events found by RANSAC.

Seismic Trace with P- and S-waves

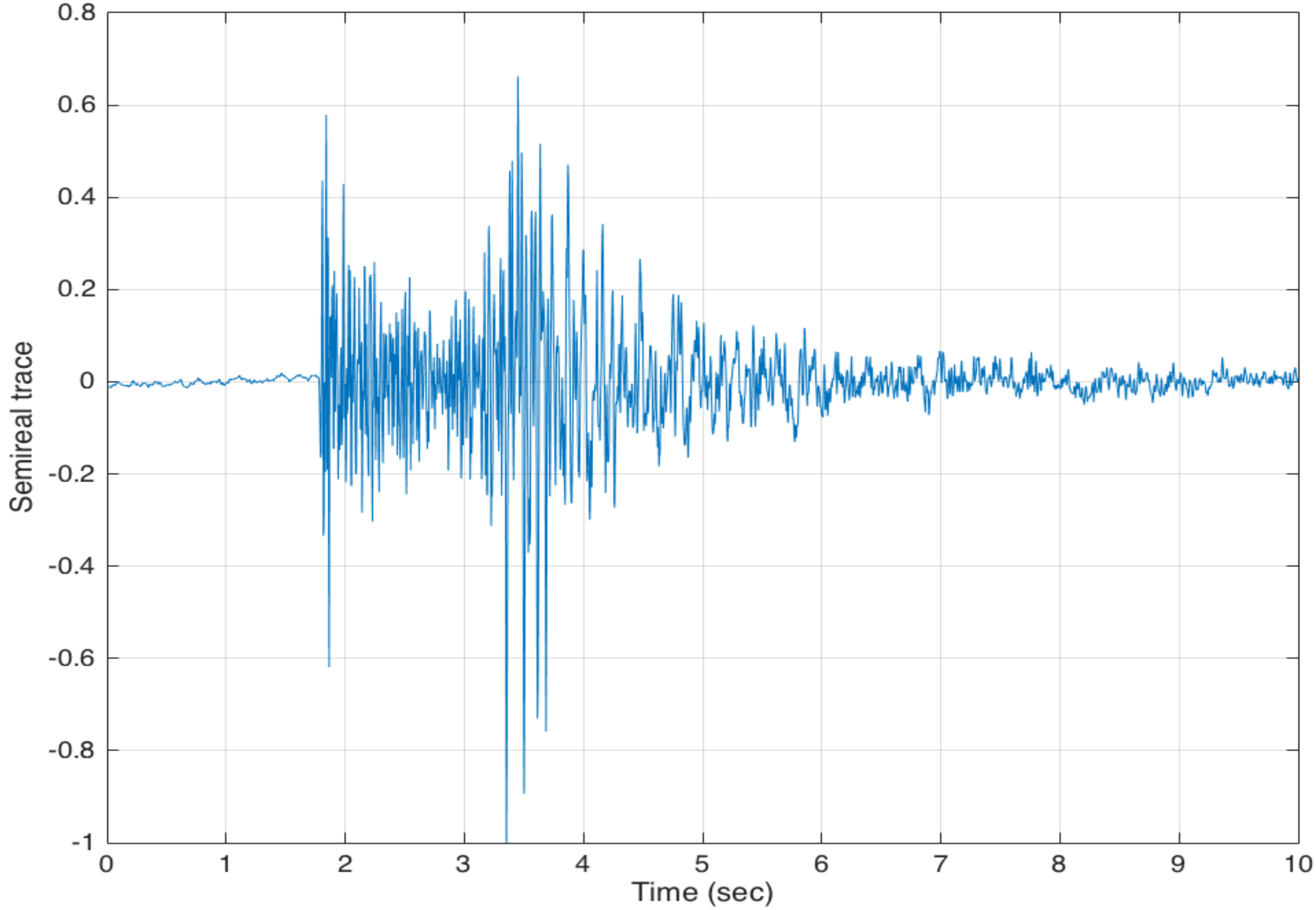
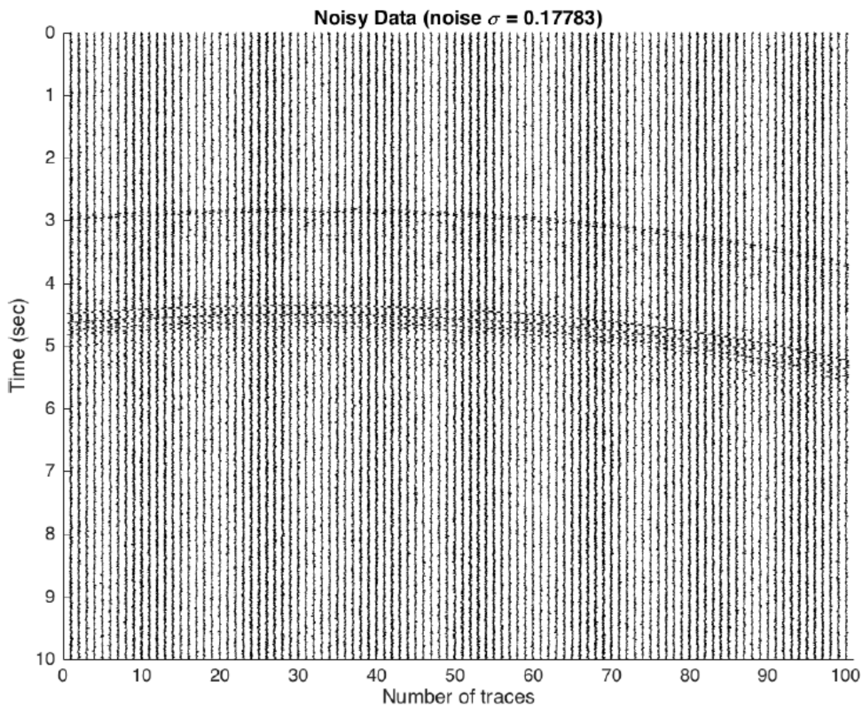
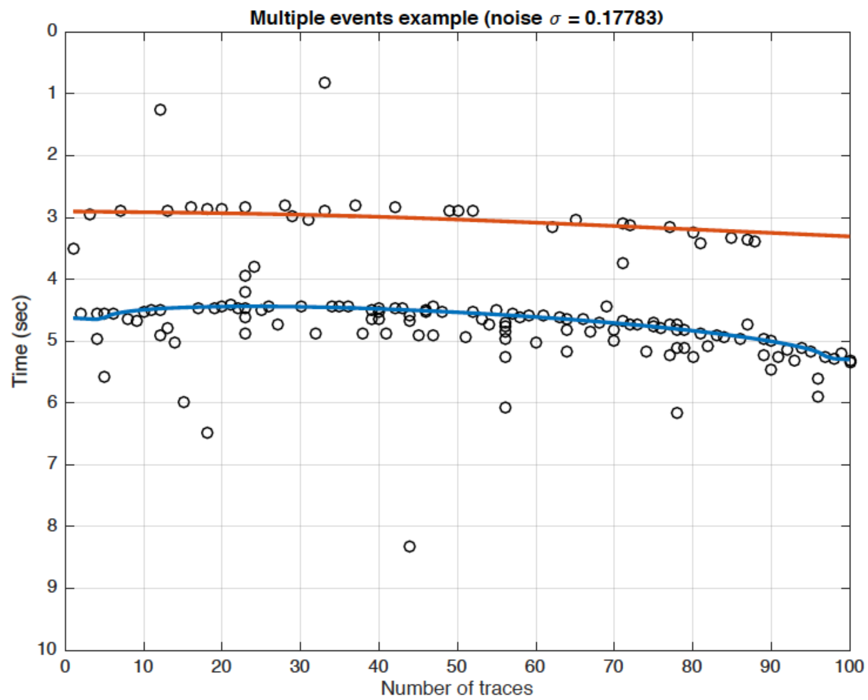


Figure 5: Seismic trace with two events.

Seismic Traces Example



(a)



(b)

Figure 6: Curve fitting results: (a) semi-real seismic traces as input data with noise level of σ = 0.178 and pick threshold θ = 0.9, (b) P-wave (red) and S-wave (blue, first event) phases of arrival event.

Extension to 2-D Array

- Arrival times lie on a **hyperbolic surface** assuming a layered velocity model
- Quadratic model requires nine parameters

- Synthetic example using Ricker wavelet and AWGN
- Surface array (15 by 15) with 400m spacing
- Target source (2 km deep) below the array center

2-D Array Layout

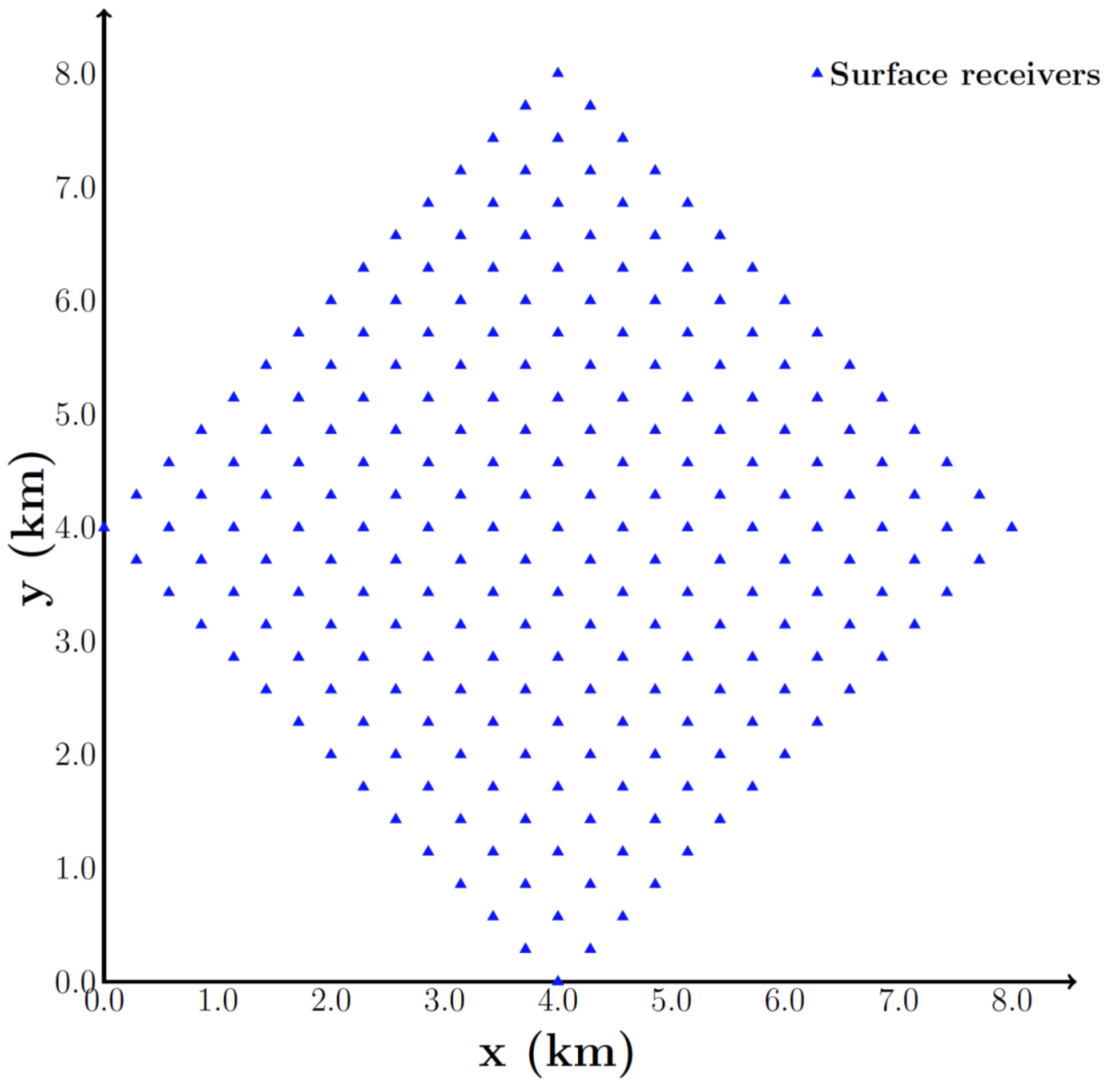
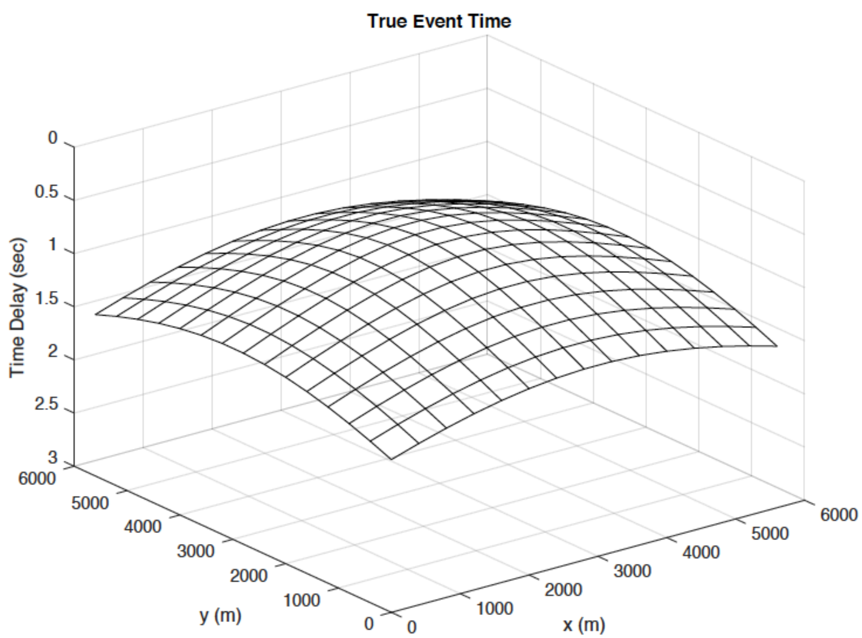
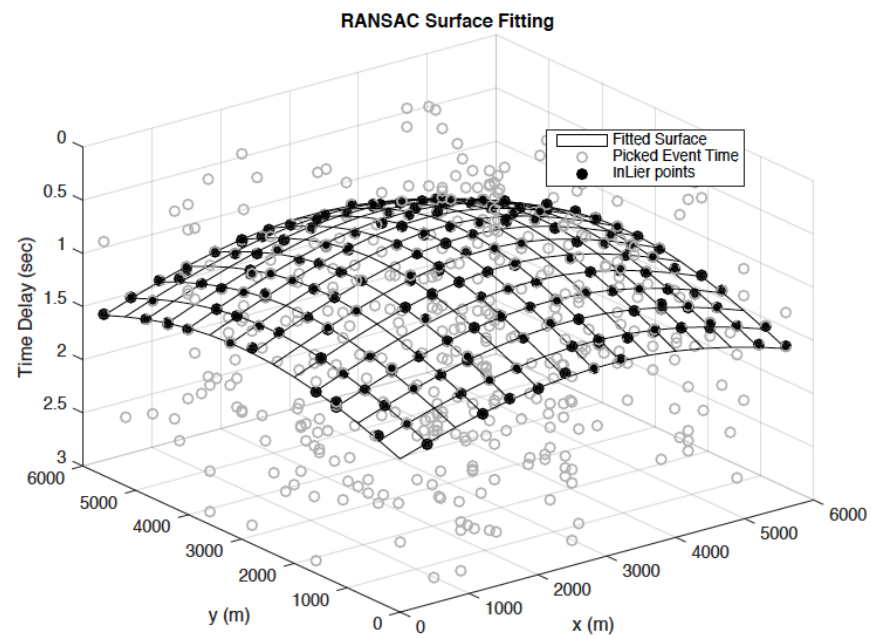


Figure 7:

2-D Array Surface Fitting (6 dB)



(a)



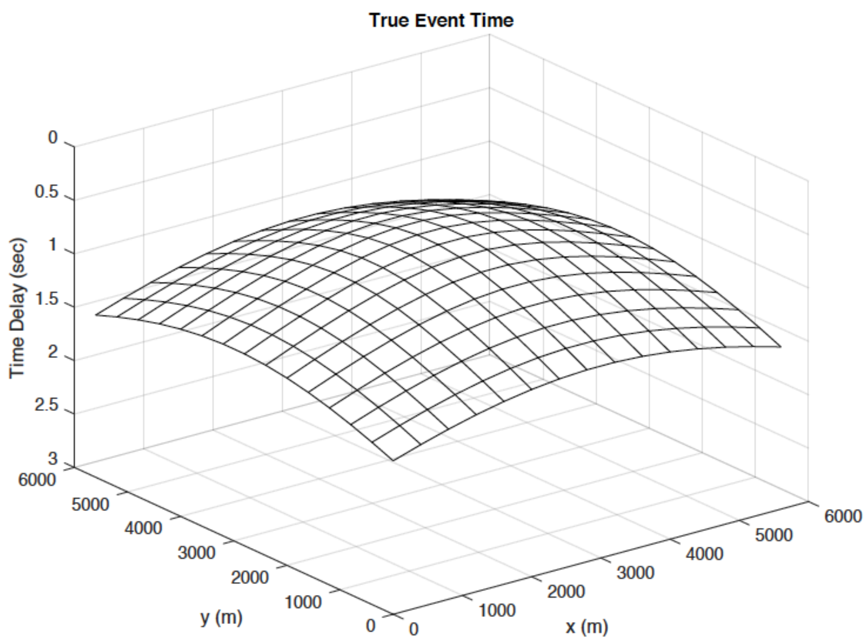
(b)

Figure 8: Surface fitting results: (a) True event times of 2D array monitoring a single event, (b) picked event times (o) with noise level of σ = 0.501 and pick threshold θ = 0.9. RANSAC fitting result shown as a surface.

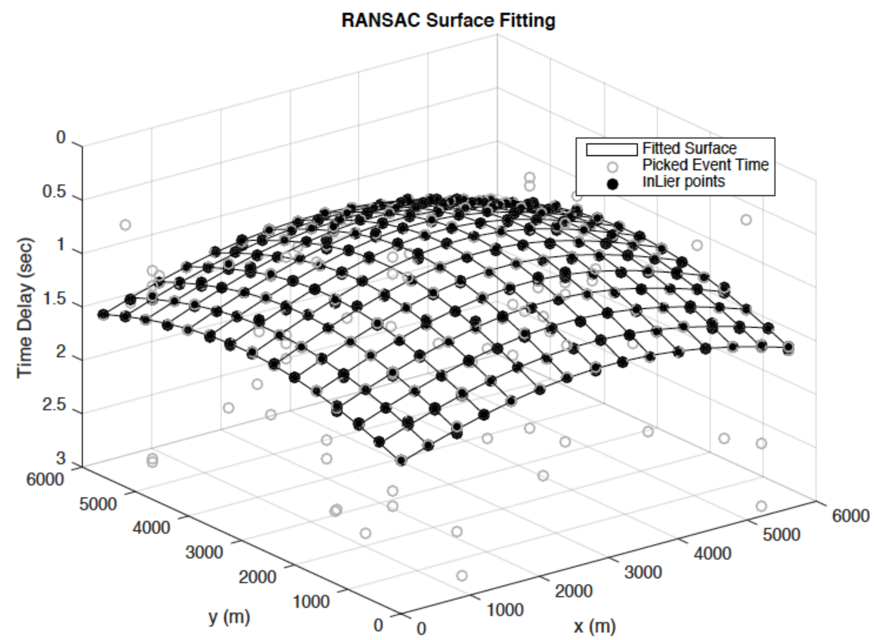
Conclusion

- Under low SNR, RANSAC is able to learn most of the inliers and effectively increases the overall time picking accuracy
- Fast processing: try many simple models
- Multiple events are handled sequentially
- RANSAC has been extended to 2-D arrays and 3-D data by using more parameters in the model

2-D Array Surface Fitting (10 dB)



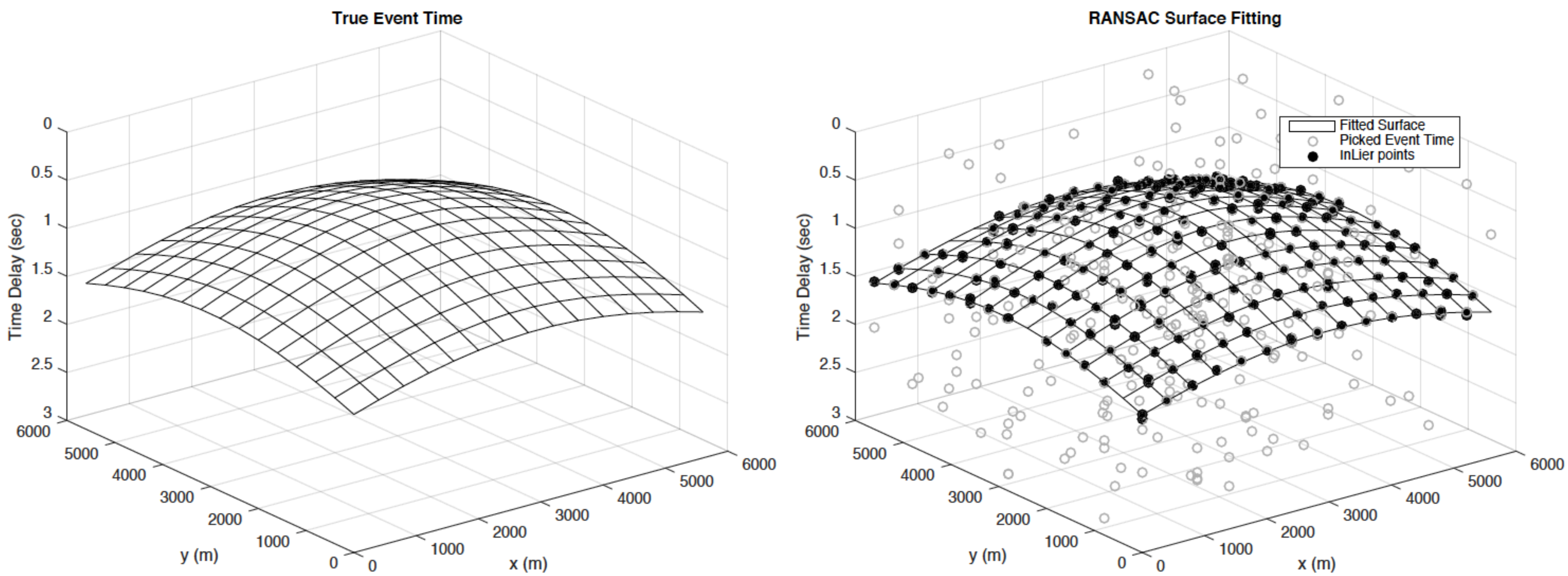
(a)



(b)

Figure 9: Surface fitting results: (a)True even time of 2D array monitoring single event data, (b) Picked event time with noise level of σ = 0.316 and pick threshold θ = 0.9 followed by RANSAC fitting results after sorting.

2-D Array Surface Fitting (8 dB)



(a) (b)

Figure 9: Surface fitting results: (a)True even time of 2D array monitoring single event data, (b) Picked event time with noise level of σ = 0.398 and pick threshold θ = 0.9 followed by RANSAC fitting results after sorting.

2-D Array Surface Fitting (4 dB)

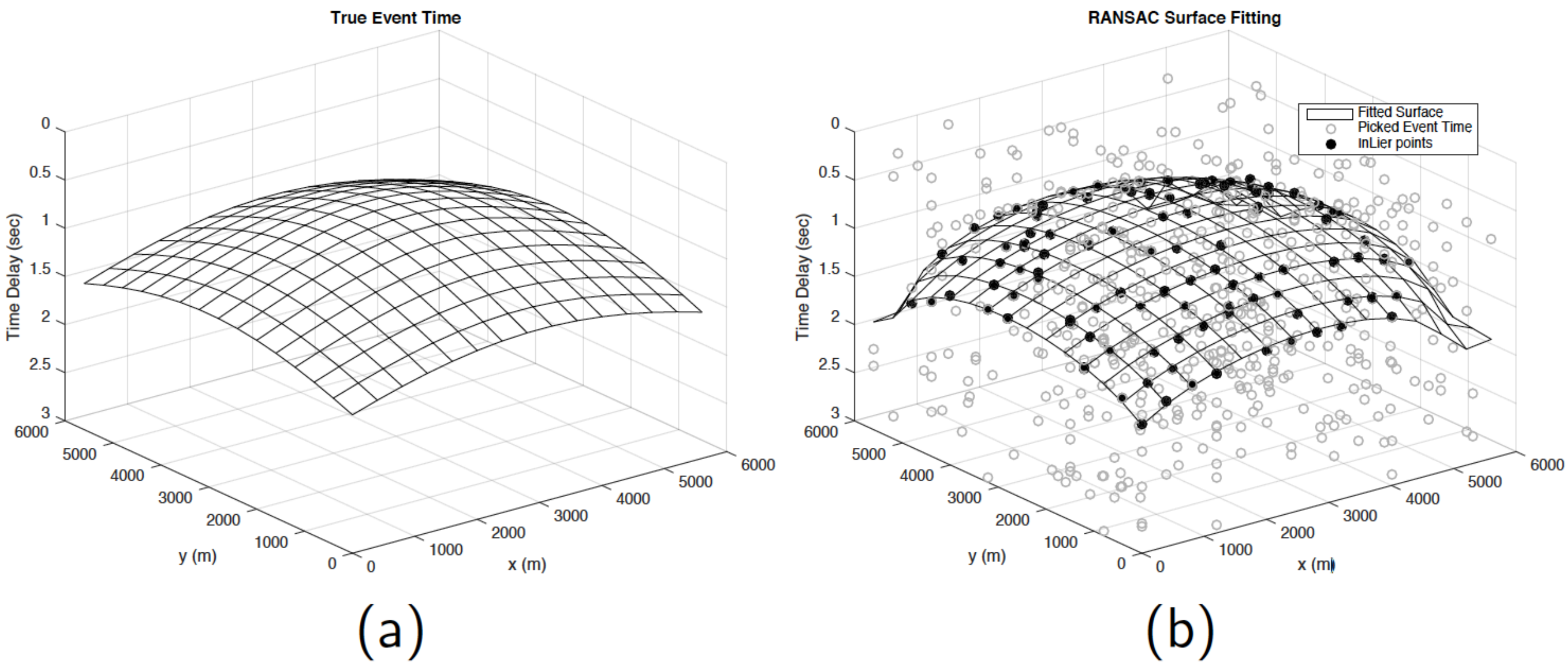


Figure 9: Surface fitting results: (a)True even time of 2D array monitoring single event data, (b) Picked event time with noise level of σ = 0.631 and pick threshold θ = 0.9 followed by RANSAC fitting results after sorting.

Sort 2-D Array

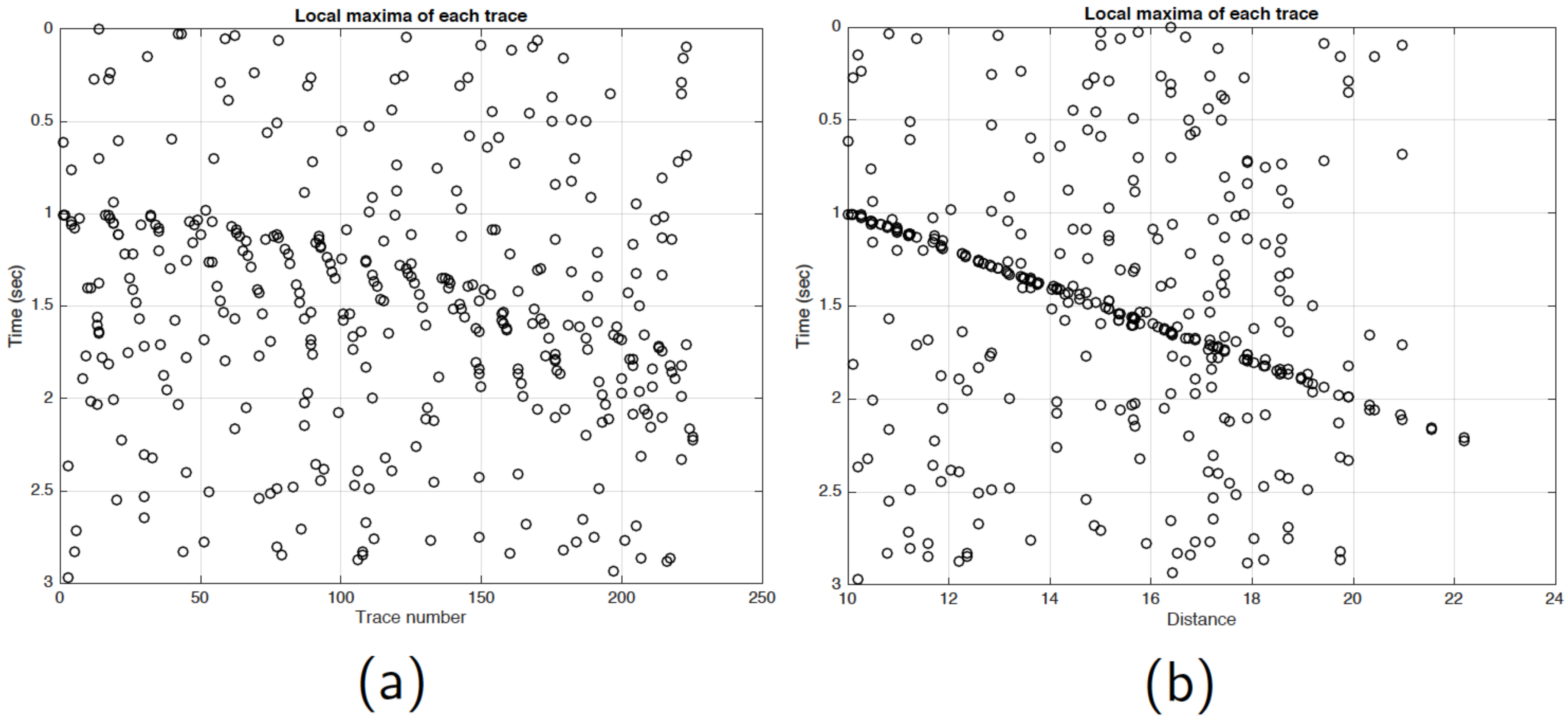


Figure 7: Sorting results: (a) local maxima of each traces; (b) local maxima with traces sorted by distance.

2-D Array Synthetic Example

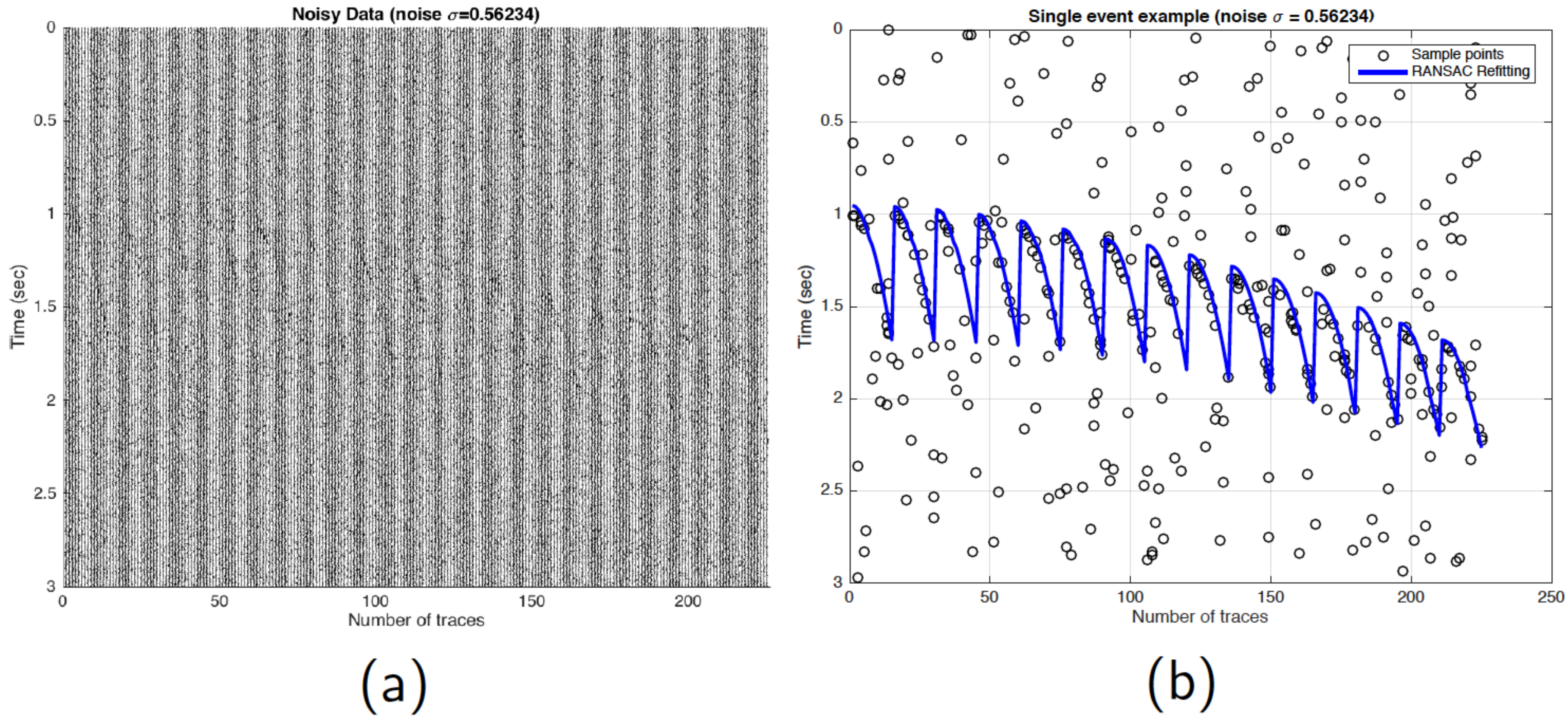


Figure 8: Curve fitting results: a)2D array monitoring single event data with noise level of σ = 0.562 and pick threshold θ = 0.95; b) Picked peaks and RANSAC fitting results after sorting.

RANSAC for 2-D Surface Arrays

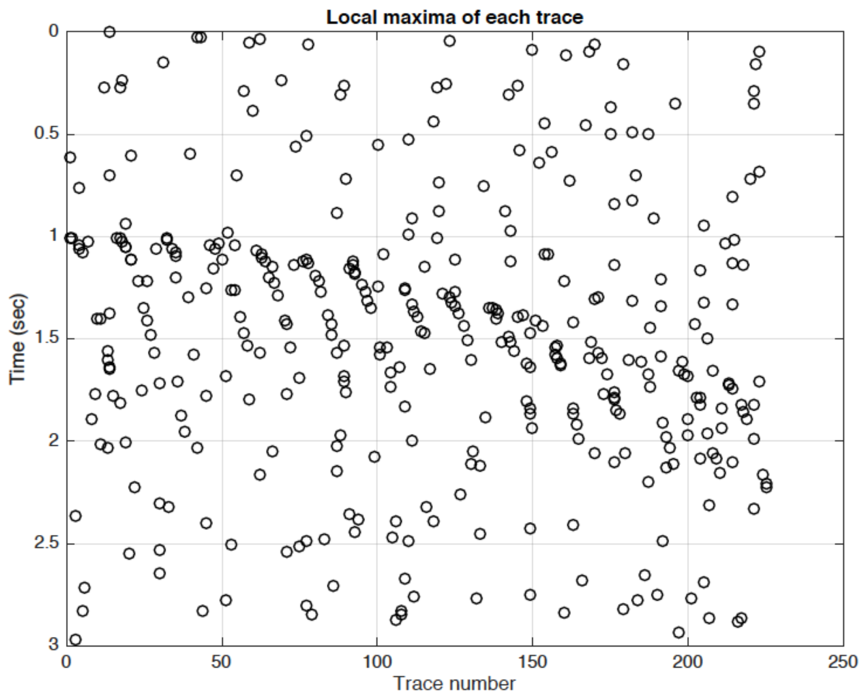
- In general, a quadratic surface would be needed
- In one case, the problem is fitting a line

- Sort received traces according to the distance between receiver array and target area
- Fit a RANSAC curve for the sorted hyperbola
  - Exploit RANSAC fitting for non-uniform sampling
- Map hyperbola curve back to the 2D receiver array setting

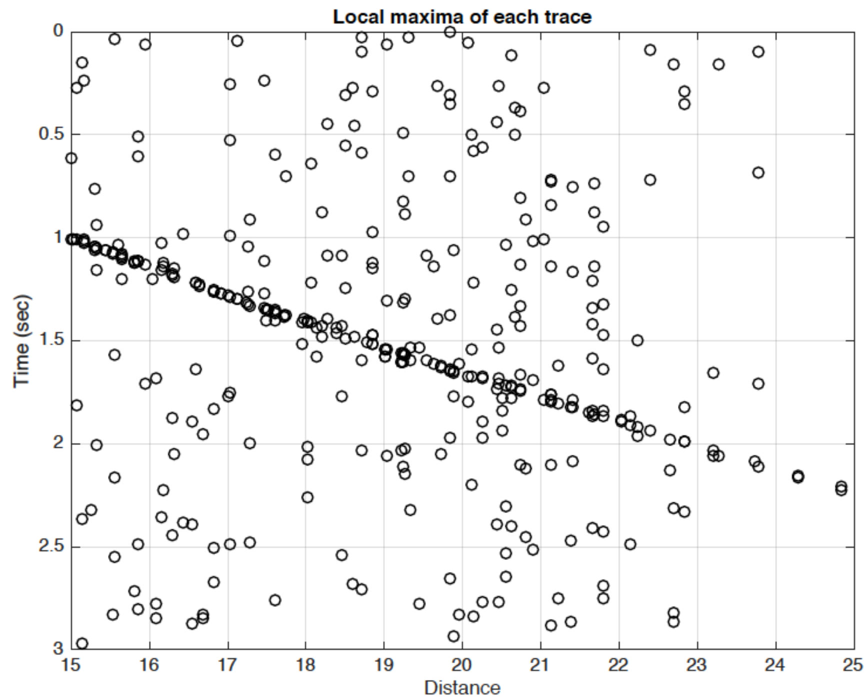
Strategy for sorting 2-D arrays

- Use linear sub-arrays to estimate source location as half-circles on orthogonal vertical planes
- Use the intersection of half-circles as predicted source location if possible; otherwise, use the mid point
- Use the predicted source location to sort 2-D array

Sort 2-D array (50% depth error)



(a)



(b)

Figure 9: Sorting results: a) local maxima of each traces; b) local maxima with traces sorted by distance.

2-D Array (50% depth error)

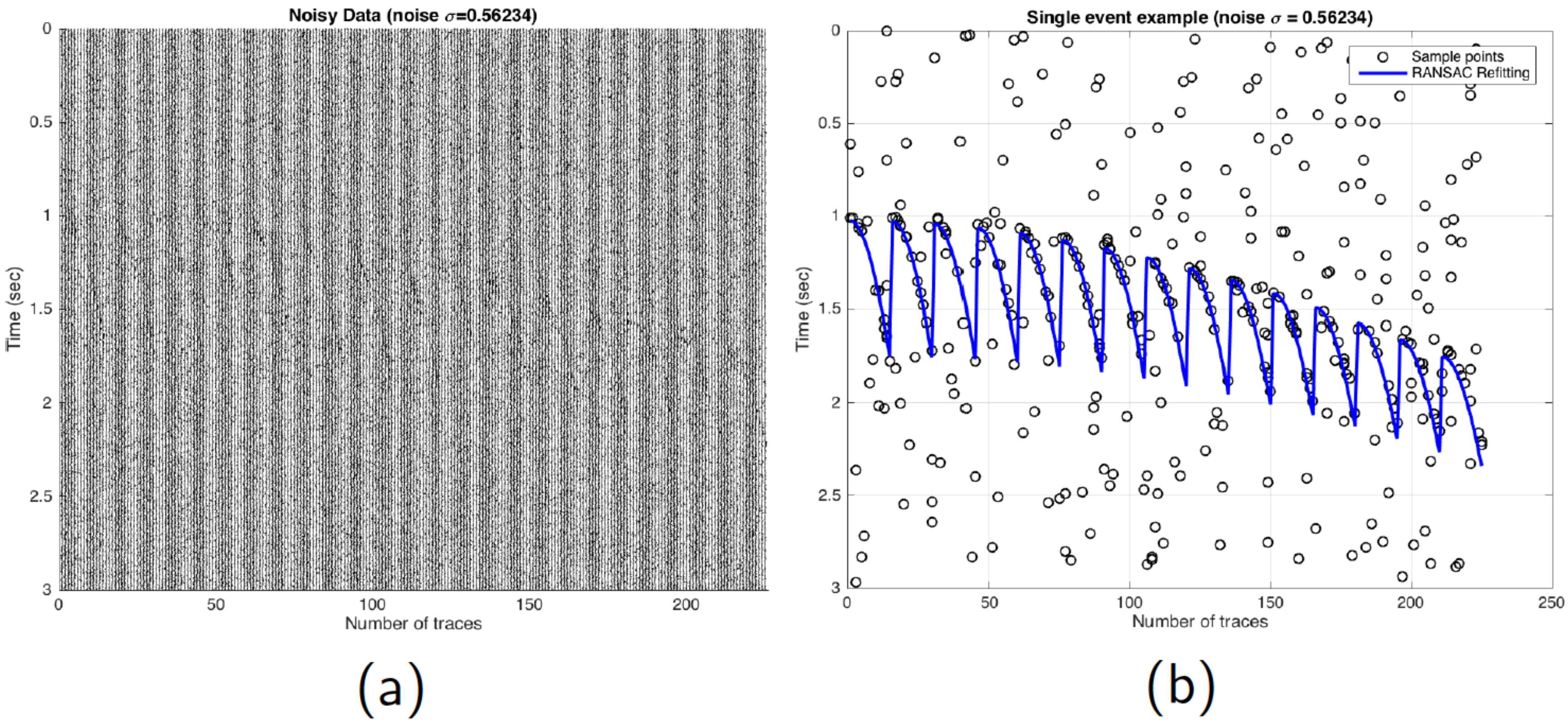


Figure 10: Curve fitting results: (a) 2D array monitoring single event data with noise level of σ = 0.562; (b) Picked peaks and RANSAC fitting results after sorting.

Single Event Simulation

Is it true that
Least squares
Uses strongest
Pick ??

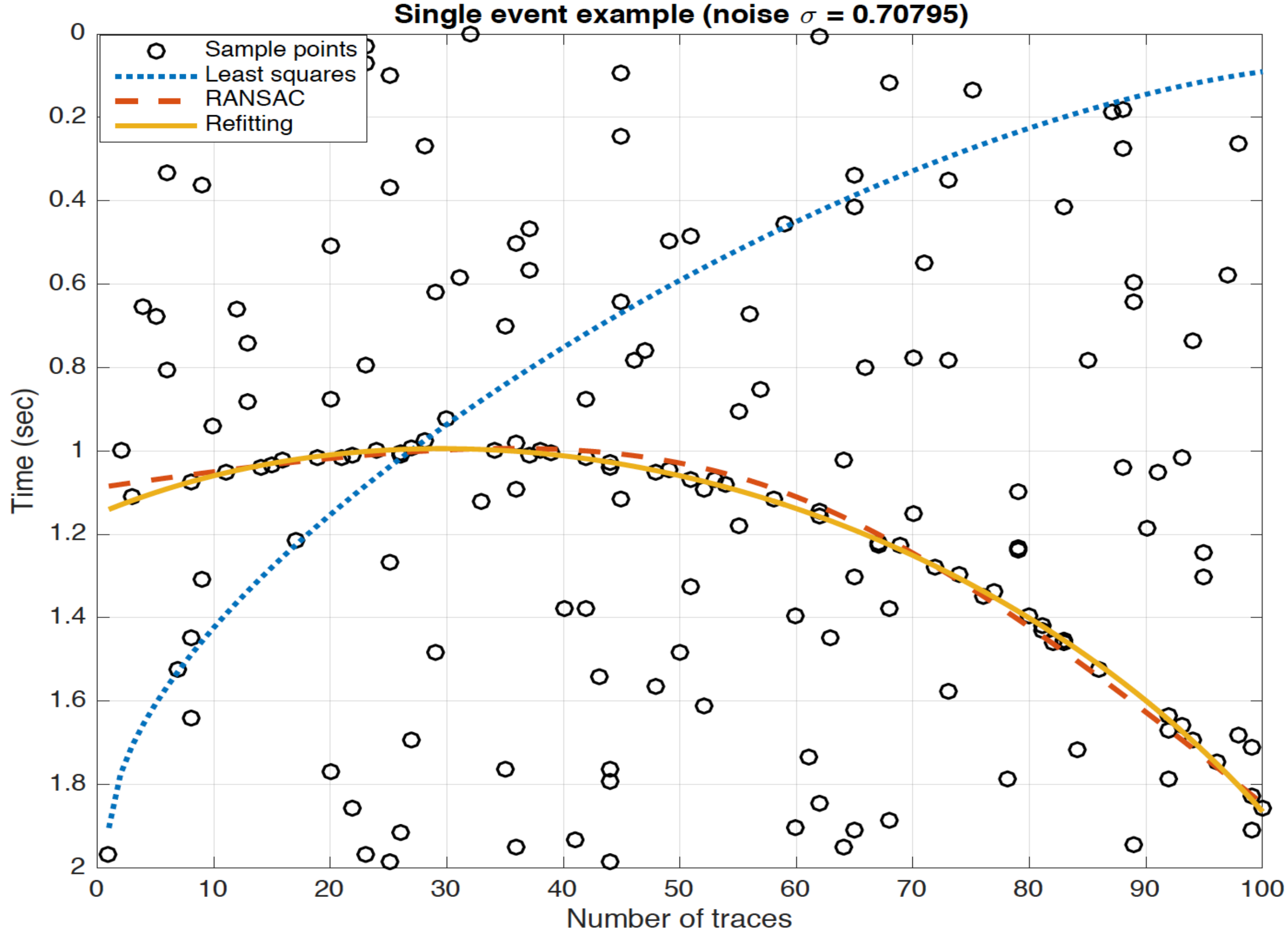


Figure 3: Single event curve fitting results comparing with direct least square.