Need for Expert Knowledge (and Soft Computing) in Geosciences

Vladik Kreinovich

Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA
vladik@utep.edu

http://www.cs.utep.edu/vladik

1. Determining Earth Structure Is Important

- *Importance:* civilization greatly depends on the things we extract from the Earth: oil, gas, water.
- *Need* is growing, so we must find new resources.
- *Problem:* most easy-to-access mineral resources have been discovered.
- *Example:* new oil fields are at large depths, under water, in remote areas – so drilling is very expensive.
- *Objective:* predict resources before we invest in drilling.
- *How:* we know what structures are promising.
- *Example:* oil and gas concentrate near the top of (natural) underground domal structures.
- *Conclusion:* to find mineral resources, we must determine the structure at diff. depths z and locations (x, y).

2. Data that We Can Use to Determine the Earth Structure

- *Available measurement results:* those obtained without drilling boreholes.
- *Examples:*
  - gravity and magnetic measurements;
  - travel-times t_i of seismic ways through the earth.
- *Need for active seismic data:*
  - passive data from earthquakes are rare;
  - to get more information, we make explosions, and measure how the resulting seismic waves propagate.
- *Resulting seismic inverse problem:*
  - we know the travel times t_i;
  - we want to reconstruct velocities at different depths.

3. Algorithm for the Forward Seismic Problem

- *We know:* velocities v_j in each grid cell j.
- *We want to compute:* traveltimes t_i.
- *First step:* find shortest (in time) paths.
- *Within cell:* path is a straight line.
- *On the border:* between cells with velocities v and v', we have Snell's law $\dfrac{\sin(\varphi)}{v} = \dfrac{\sin(\varphi')}{v'}$.
- *Comment:* if $\sin(\varphi') > 1$, the wave cannot get penetrate into the neighboring cell; it bounces back.
- *Resulting traveltimes:* $t_i = \sum\limits_j \dfrac{\ell_{ij}}{v_j}$, where ℓ_{ij} is the length of the part of i-th path within cell j.
- *Simplification:* use slownesses $s_j \stackrel{\text{def}}{=} \dfrac{1}{v_j}$; $t_i = \sum\limits_j \ell_{ij} \cdot s_j$.

4. Algorithm for the Inverse Problem: General Description

- *The most widely used:* John Hole's iterative algorithm.
- *Starting point:* reasonable initial slownesses.
- *On each iteration:* we use current (approximate) slownesses s_j to compute the travel-times $t_i = \sum\limits_j \ell_{ij} \cdot s_j$.
- *Fact:* measured travel-times $\widetilde{t}_i$ are somewhat different: $\Delta t_i \stackrel{\text{def}}{=} \widetilde{t}_i - t_i \neq 0$.
- *Objective:* find Δs_j so that $\sum \ell_{ij} \cdot (s_j + \Delta s_j) = \widetilde{t}_i$.
- *Problem:* we have many observations n, and computation time $\sim n^3$ – too long, so we need faster techniques.
- *Stopping criterion:* when average error $\dfrac{1}{n} \sum\limits_{i=1}^{n} (\Delta t_i)^2$ is below noise.

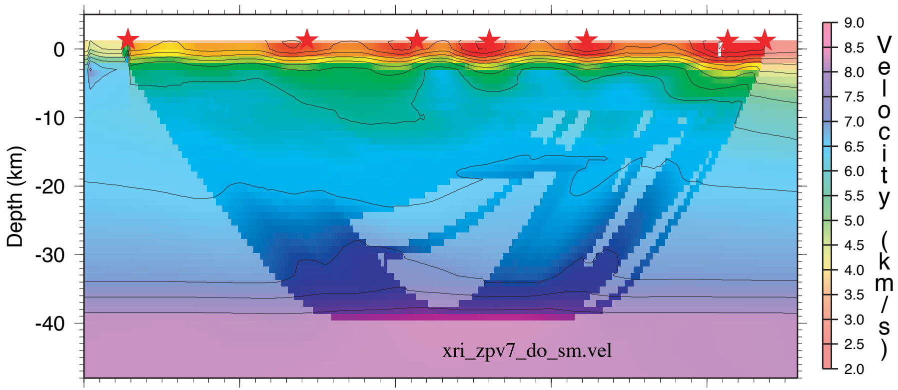
5. Algorithm for the Inverse Problem: Details

- *Objective (reminder):* find Δs_j s.t. $\sum \ell_{ij} \cdot \Delta s_j = \Delta t_i$.
- *Simplest case:* one path.
- *Specifics:* under-determined system: 1 equation, many unknowns Δs_j.
- *Idea:* no reason for Δs_j to be different: $\Delta s_j \approx \Delta s_{j'}$.
- *Formalization:* minimize $\sum\limits_{j,j'} (\Delta s_j - \Delta s_{j'})^2$ under the constraint $\sum \ell_{ij} \cdot \Delta s_j = \Delta t_i$.
- *Solution:* $\Delta s_j = \dfrac{\Delta t_i}{L_i}$ for all j, where $L_i = \sum\limits_j \ell_{ij}$.
- *Realistic case:* several paths; we have Δs_{ij} for different paths i.
- *Idea:* least squares $\sum\limits_i (\Delta s_j - \Delta s_{ij})^2 \to \min$.
- *Solution:* Δs_j is the average of Δs_{ij}.

6. Successes, Limitations, Need for Prior Knowledge

- *Successes:* the algorithm usually leads to reasonable geophysical models.
- *Limitations:* often, the resulting velocity model is not geophysically meaningful.
- *Example:* resulting velocities outside of the range of reasonable velocities at this depth.
- *It is desirable:* incorporate the expert knowledge into the algorithm for solving the inverse problem.

Hole Tomography Smashed Masked Velocity Models



7. Need for Expert Knowledge in Data Processing is Ubiquitous

- We need to reconstruct the values of the quantities of interest from the measurement results.
- Geosciences example: reconstructing density at different depths and different locations.
- Often, several drastically different density distributions are consistent with the same observations.
- Such problems are called "ill-posed".
- Out of all these distributions, we need to select the physically meaningful one(s).
- This is where expert knowledge is needed, to describe what "physically meaningful" means.
- On the example of the above geophysical problem, we show how to use this expert knowledge.

8. Case of Interval Prior Knowledge

- *Idea:* for each cell j, a geophysicist provides an interval $[\underline{s}_j, \overline{s}_j]$ of possible values of s_j.
- *Hole's code:* along each path i, we find corrections Δs_{ij} that minimize

$$\sum_{j,j'} (\Delta s_{ij} - \Delta s_{ij'})^2$$

under the constraint

$$\sum_{j=1}^{c} \ell_{ij} \cdot \Delta s_{ij} = \Delta t_i.$$

- *Modification:* we must minimize under the additional constraints

$$\underline{s}_j \le s_j^{(k)} + \Delta s_{ij} \le \overline{s}_j.$$

- *What we designed:* an $O(c \cdot \log(c))$ algorithm for solving this new problem.

9. Main Idea of an Algorithm

- *Idea* – method of alternating projections:
  - first, add a correction that satisfy the first constraint,
  - then, the additional correction that satisfies the second constraint,
  - etc.
- *Specifics:*
  - first, add equal values Δs_{ij} to minimize Δt_i;
  - restrict the values to the nearest points from $[\underline{s}_j, \overline{s}_j]$,
  - repeat until converges.
- *Comment:* this way, we can also use other prior knowledge (e.g., probabilistic).

10. New Algorithm: For Each Path on Each Iteration

- *Case:* $\Delta t_i > 0$; for $\Delta t_i < 0$, we have similar formulas.
- Compute, for each cell j,

$$\underline{\Delta}_j = \underline{s}_j - s_j^{(k-1)} \text{ and } \overline{\Delta}_j = \overline{s}_j - s_j^{(k-1)}.$$

- Sort values $\overline{\Delta}_j$ into

$$\overline{\Delta}_{(1)} \le \overline{\Delta}_{(2)} \le \ldots \le \overline{\Delta}_{(c)}.$$

- For every p from 0 to c, compute:

$$A_0 = 0,\ \mathcal{L}_0 = L_i,\ A_p = A_{p-1} + \ell_{i(p)} \cdot \overline{\Delta}_{(p)},\ \mathcal{L}_p = \mathcal{L}_{p-1} - \ell_{i(p)}.$$

- Compute $S_p = A_p + \mathcal{L}_p \cdot \Delta_{(p+1)}$, and find p s.t.

$$S_{p-1} \le \Delta t_i < S_p.$$

- Take $\Delta s_{i(j)} = \overline{\Delta}_j$ for $j \le p$, and $\Delta s_{(j)} = \dfrac{\Delta t_i - A_p}{\mathcal{L}_p}$ else.
- Then, average Δs_{ij} over paths i.

11. Explicit Expert Knowledge: Fuzzy Uncertainty

- Experts can usually produce a wide interval of which they are practically 100% certain.
- In addition, experts can also produce narrower intervals about which their degree of certainty is smaller.
- As a result, instead of a *single* interval, we have a *nested* family of intervals corr. to diff. levels of uncertainty.
- In effect, we get a *fuzzy* interval (of which different intervals are α-cuts).
- *Previously:* a solution is satisfying or not.
- *New idea:* a satisfaction *degree* d.
- *Specifics:* d is the largest α for which all s_i are within the corresponding α-cut intervals.

12. How We Can Use Fuzzy Uncertainty

- *Objective:* find the largest possible value α for which the slownesses belong to the α-cut intervals.
- *Possible approach:*
  - try $\alpha = 0$, $\alpha = 0.1$, $\alpha = 0.2$, etc., until the process stops converging;
  - the solution corresponding to the previous value α is the answer.
- *Comment:*
  - this is the basic straightforward way to take fuzzy-valued expert knowledge into consideration;
  - several researchers successfully used fuzzy expert knowledge in geophysics (Nikravesh, Klir, et al.).

13. Need for Data Fusion

- In many practical situations, we have several results $\widetilde{x}^{(1)}, \ldots, \widetilde{x}^{(n)}$ of measuring the same quantity x.
- These results are different since measurements are never 100% accurate.
- It is know that by combining different measurement results, we increase accuracy.
- Simplest case: we use the same measuring instrument for all measurements.
- In this case, an arithmetic average reduces the st. dev. by a factor of $\sqrt{n}$:

$$\widetilde{x} = \frac{\widetilde{x}^{(1)} + \ldots + \widetilde{x}^{(n)}}{n}.$$

- When we fuse measurements of *different* accuracy, we need to use different weights for *different* values $\widetilde{x}^{(i)}$.

14. Case of Probabilistic Prior Knowledge

- *Description:* from prior observations, we know $\widetilde{s}_j \approx s_j$, and we know the st. dev. σ_j of this value.
- *Minimize:* $\sum\limits_{j,j'} (\Delta s_{ij} - \Delta s_{ij'})^2$ s.t. $\sum\limits_{j=1}^{c} \ell_{ij} \cdot \Delta s_{ij} = \Delta t_i$ and

$$\frac{1}{n} \cdot \sum_{j=1}^{c} \frac{((s_j^{(k)} + \Delta s_{ij}) - \widetilde{s}_j)^2}{\sigma_j^2} = 1.$$

- *Solution* (Lagrange multipliers): $\overline{\Delta s} \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum\limits_{j=1}^{c} \Delta s_{ij}$,

$$\frac{2}{n} \cdot \Delta s_{ij} - \frac{2}{n} \cdot \overline{\Delta s} + \lambda \cdot \ell_{ij} + \frac{2\mu}{n \cdot \sigma_j^2} \cdot (s_j^{(k)} + \Delta s_{ij} - \widetilde{s}_j) = 0.$$

- *Fact:* Δs_{ij} is an explicit function of λ, μ, $\overline{\Delta s}$.
- *Algorithm:* solve 3 non-linear equations (above one + 2 constraints) with unknowns λ, μ, $\overline{\Delta s}$.

15. Combination of Different Types of Prior Knowledge

- *Need:* we often have both:
  - prior measurement results – i.e., *probabilistic* knowledge, and
  - expert estimates – i.e., *interval* and *fuzzy* knowledge.
- *Minimize:* $\sum\limits_{j,j'}(\Delta s_{ij}-\Delta s_{ij'})^2$ s.t. $\sum\limits_{j=1}^{c}\ell_{ij}\cdot\Delta s_{ij}=\Delta t_i$,

$$\frac{1}{n}\cdot\sum_{j=1}^{c}\frac{((s_j^{(k)}+\Delta s_{ij})-\widetilde{s}_j)^2}{\sigma_j^2}\le 1,$$

and $\underline{s}_j\le s_j^{(k)}+\Delta s_{ij}\le\overline{s}_j$.

- *Idea:* we minimize a convex function under convex constraints; efficient algorithms are known.

16. Combination of Different Types of Prior Knowledge: Algorithm

- *Idea* – method of alternating projections:
  - first, add a correction that satisfy the first constraint,
  - then, the additional correction that satisfies the second constraint,
  - etc.
- *Specifics:*
  - first, add equal values Δs_{ij} to minimize Δt_i;
  - restrict the values to the nearest points from $[\underline{s}_j, \overline{s}_j]$,
  - find the extra corrections that satisfy the probabilistic constraint,
  - repeat until converges.

17. Data Fusion: Challenge

- When we fuse measurements of *different* accuracy, we need to use different weights for *different* values $\widetilde{x}^{(i)}$.
- Sometimes, we can find the actual values and thus, estimate the accuracy of different measurements.
- In other cases – e.g., in geosciences – it is difficult to find the actual density at depth 40 km.
- Hence, in geosciences, it is difficult to gauge the accuracy of seismic, gravity, and other techniques.
- In this case, we need to estimate the accuracies from the observations.
- We will show that in this case, seemingly reasonable statistical methods do not work well.
- Thus, statistical methods need to be supplemented with expert knowledge.

18. Traditional Statistical Methods: Reminder

- In many cases, the measurement error is caused by many different causes.
- It is known that the distribution of the sum of many small random variables is $\approx$ normally distributed.
- So, we can conclude that the measurement errors are normally distributed, with probability density

$$\rho(\widetilde{x}) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{(\widetilde{x} - x)^2}{2\sigma^2}\right).$$

- If we have n results $\widetilde{x}^{(i)}$ of independent measurements, then prob. is prop. to $\rho = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{(\widetilde{x}^{(i)} - x)^2}{2\sigma^2}\right).$
- Maximum Likelihood Method: select most probable x and σ, for which prob. (hence ρ) is the largest.

19. Traditional Statistical Methods (cont-d)

- Maximizing $\rho = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\cdot\sigma} \cdot \exp\left(-\frac{(\widetilde{x}^{(i)} - x)^2}{2\sigma^2}\right)$ is equivalent to minimizing

$$\psi = -\ln(\rho) = \text{const} + n \cdot \ln(\sigma) + \sum_{i=1}^{n} \frac{(\widetilde{x}^{(i)} - x)^2}{2\sigma^2}.$$

- W.r.t. x, we get the Least Squares method which leads to the arithmetic average $x = \frac{1}{n} \cdot \sum_{i=1}^{n} \widetilde{x}^{(i)}$.
- Differentiating ψ w.r.t. σ and equating to 0, we get

$$\frac{n}{\sigma} - \sum_{i=1}^{n} \frac{(\widetilde{x}^{(i)} - x)^2}{\sigma^3} = 0.$$

- So, we get the usual estimate $\sigma^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (\widetilde{x}^{(i)} - x)^2$.

20. Case of Different Measuring Instruments (MI): Surprising Problem

- *Situation:* for different quantities x_j, $j = 1, \ldots, m$, we have measurement results $\widetilde{x}_j^{(i)}$ corr. to diff. MI, w/diff. σ_i.
- The resulting probability is proportional to

$$\rho = \prod_{i=1}^{n} \prod_{j=1}^{m} \frac{1}{\sqrt{2\pi} \cdot \sigma_i} \cdot \exp\left(-\frac{(\widetilde{x}_j^{(i)} - x_j)^2}{2\sigma_i^2}\right).$$

- *Seemingly natural idea:* use Maximum Likelihood method, i.e., find x_j and σ_i for which $\rho \to \max$.
- *We tried*, and found that at maximum, one of σ_i is 0.
- *We then theoretically confirmed:* that maximum $\rho_{\max} = \infty$ is attained:
  - when $\sigma_{i_0} = 0$ for some i_0, and
  - when $x_j = \widetilde{x}_j^{(i_0)}$ for all j.

21. Analysis of the Problem

- *We know:* that all the measuring instruments are imperfect, i.e., $\sigma_i > 0$.
- *From the mathematical viewpoint:* we get $\sigma_{i_0} = 0$ for some i_0.
- This mathematical solution is not physically meaningful.
- To avoid this non-physical solution, we need to explicitly add the requirement that $\sigma_i > 0$ for all i.
- This *crisp* requirement does not help: by taking smaller and smaller σ_{i_0}, we can get ρ as large as possible.
- Intuitively, what we need is a *fuzzy* requirement – that all σ_i are not too small.
- This fuzzy requirement enables us to avoid non-physical values of σ_i.

22. Towards an Algorithm

- $\rho = \prod_{i=1}^{n} \prod_{j=1}^{m} \frac{1}{\sqrt{2\pi} \cdot \sigma_i} \cdot \exp\left(-\frac{(\widetilde{x}_j^{(i)} - x_j)^2}{2\sigma_i^2}\right) \to \max.$
- $\psi = -\ln(\rho) = n \cdot \sum_{i=1}^{n} \ln(\sigma_i) + \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{(\widetilde{x}_j^{(i)} - x_j)^2}{2\sigma_i^2} \to \min.$
- Differentiating ψ w.r.t. x_j and σ_i, we get:

$$x_j = \frac{\sum_{i=1}^{n} \sigma_i^{-2} \cdot \widetilde{x}_j^{(i)}}{\sum_{i=1}^{n} \sigma_i^{-2}}; \quad \sigma_i^2 = \frac{1}{m} \cdot \sum_{j=1}^{m} (\widetilde{x}_j^{(i)} - x_j)^2.$$

- We first take $\sigma_i = \text{const}$, then iteratively compute:

  (1) x_j from σ_j, (2) σ_j from x_i, (3) x_j from σ_j, ...
- We stop when one of σ_i becomes too small (2-3 cycles).
- We return results of the previous cycle (cf. astrometry).

23. Need for Cyberinfrastructure

- A large amount of data has been collected and stored at different locations.
- Researchers and practitioners need easy and fast access to all the relevant data.
- For example, a geoscientist needs access to:
  - a state geological map (which is usually stored at the state's capital),
  - NASA photos (stored at NASA Headquarters and/or at one of corresponding NASA centers),
  - seismic data stored at different seismic stations,
  - satellite radar data, etc.

24. What Is Cyberinfrastructure

- Cyberinfrastructure is a general name for hardware/software tools that facilitate such data transfer/processing.
- Ideally, this data transfer and processing should be as easy and convenient as a google search.
- At present, the main challenges in cyberinfrastructure design are related to the actual development of:
  - the corresponding hardware tools and
  - the corresponding software tools.
- Most existing cyberinfrastructure tools use existing well defined algorithms.
- The results of using cyberinfrastructure are exciting.
- However, there is still room for improvement.

25. Cyberinfrastructure: Expert Knowledge Is Needed

- Current cyberinfrastructure results are based only on data processing.
- Some of these results do not make geological sense.
- It is necessary to take into account expert knowledge.
- Specifically, we must incorporate expert knowledge directly into the cyberinfrastructure.
- Some expert knowledge is formulated in precise terms; these types of knowledge are easier to incorporate.
- A large part of expert knowledge is formulated by using *imprecise* (fuzzy) words (like "small").
- To deal with such knowledge, fuzzy techniques have been invented.
- So, to incorporate this knowledge, it is natural to use fuzzy techniques.

26. Need for Uncertainty Propagation, and for Provenance of Uncertainty

- *Need for uncertainty propagation.*
  - main reasons for data processing and data fusion: accuracy is not high enough;
  - we must make sure that after the data processing (data fusion), we get the desired accuracy.
- *In cyberinfrastructure* this is especially important:
  - accuracy varies greatly, and
  - we do not have much control over these accuracies.
- *Need for the provenance of uncertainty:*
  - sometimes, the resulting accuracy is still too low;
  - it is desirable to find out which data points contributed most to the inaccuracy.

27. Uncertainty of the Results of Direct Measurements: Probabilistic and Interval Approaches

- Manufacturer of the measuring instrument (MI) supplies Δ_i s.t. $|\Delta x_i| \le \Delta_i$, where $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$.
- The actual (unknown) value x_i of the measured quantity is in the interval $\mathbf{x}_i = [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$.
- *Probabilistic uncertainty:* often, we know the probabilities of different values $\Delta x_i \in [-\Delta_i, \Delta_i]$.
- *How probabilities are determined:* by comparing our MI with a much more accurate (standard) MI.
- *Interval uncertainty:* in two cases, we do not determine the probabilities:
  - cutting-edge measurements;
  - measurements on the shop floor.
- In both cases, we only know that $x_i \in [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$.

28. Typical Situation: Measurement Errors are Reasonably Small

- *Typical situation:*
  - direct measurements are accurate enough;
  - the resulting approximation errors Δx_i are small;
  - terms which are quadratic (or of higher order) in Δx_i can be safely neglected.
- *Example:* for an error of 1%, its square is a negligible 0.01%.
- *Linearization:*
  - expand f in Taylor series around the point $(\widetilde{x}_1, \ldots, \widetilde{x}_n)$;
  - restrict ourselves only to linear terms:

$$\Delta y = c_1 \cdot \Delta x_1 + \ldots + c_n \cdot \Delta x_n,$$

where $c_i \stackrel{\text{def}}{=} \dfrac{\partial f}{\partial x_i}$.

29. Uncertainty of the Result of Data Processing

- *Propagation (probabilistic case):* if Δx_i are independent with st. dev. σ_i (and 0 mean), then Δy has st. dev.

$$\sigma^2 = c_1^2 \cdot \sigma_1^2 + \ldots + c_n^2 \cdot \sigma_n^2.$$

- *Provenance:*
  - we know which component σ^2 comes from the i-th measurement;
  - we can predict how replacing the i-th measurement with a more accurate one ($\sigma_i^{\text{new}} \ll \sigma_i$) will affect σ^2.
- *Propagation of interval uncertainty:*

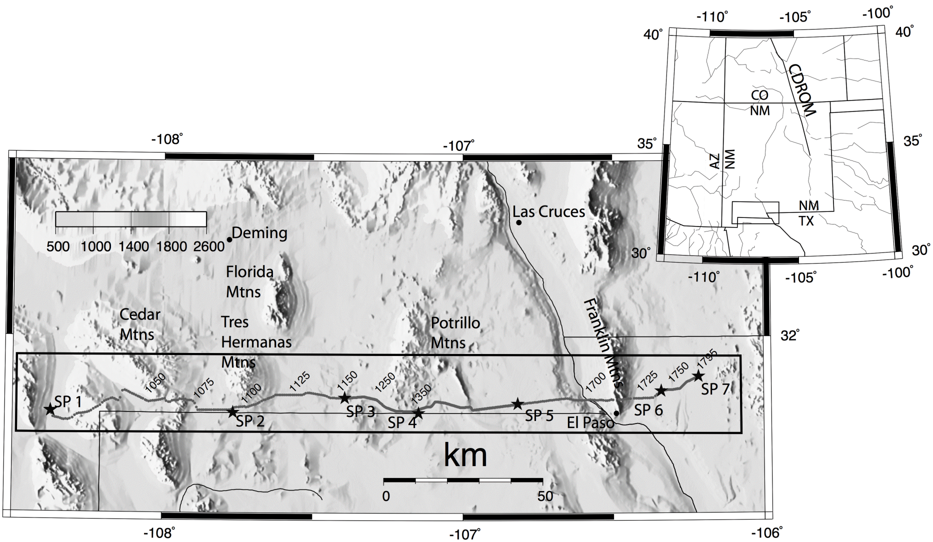
$$\Delta = |c_1| \cdot \Delta_1 + \ldots + |c_n| \cdot \Delta_n.$$

- We can predict how replacing the i-th measurement with a more accurate one ($\Delta_i^{\text{new}} \ll \Delta_i$) will affect Δ.

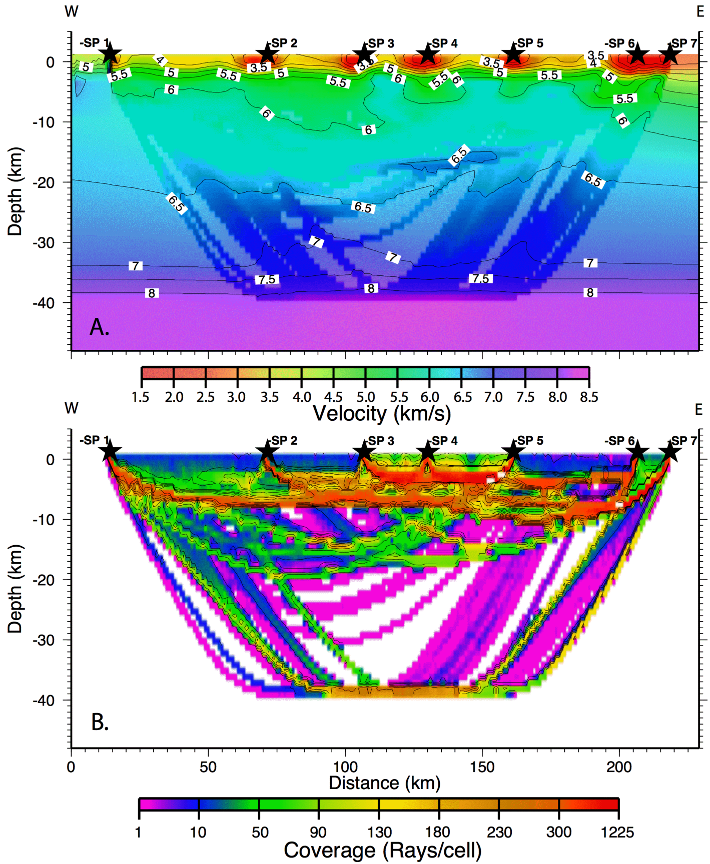
30. Beyond Probabilistic and Interval Uncertainty

- *Up to now:* we considered two extreme situations:
  - *probabilistic* uncertainty, when we know all the probabilities;
  - *interval* uncertainty, when we have no information about the probabilities.
- *Fact:* probabilistic situation is a particular case of the interval situation.
- *Conclusion:* interval bounds are wider.
- *In practice:* often, we have partial information about probabilities.
- *As a result:*
  - probabilistic bounds are too narrow,
  - interval bounds are too wide.
- *We need:* intermediate bounds.

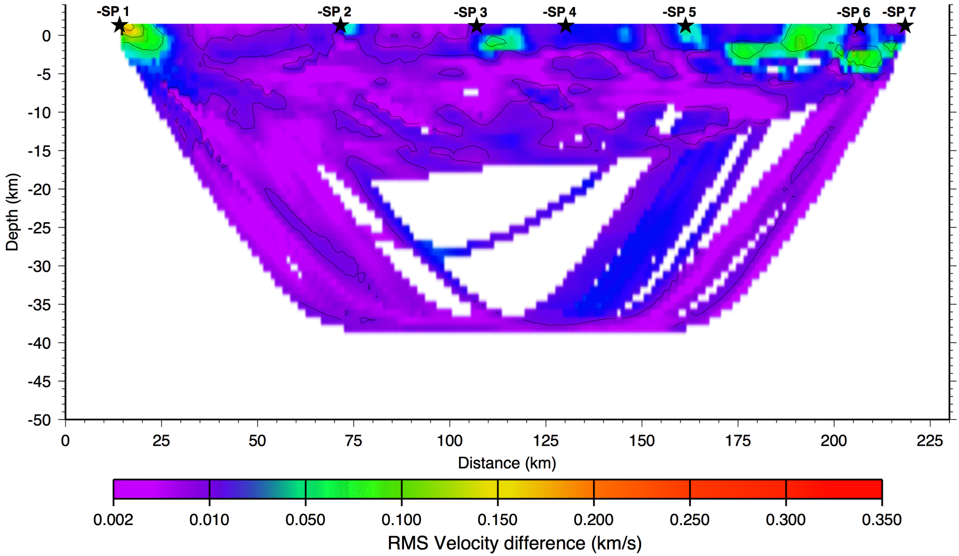
31. Case Study: Seismic Inverse Problem in the Geosciences



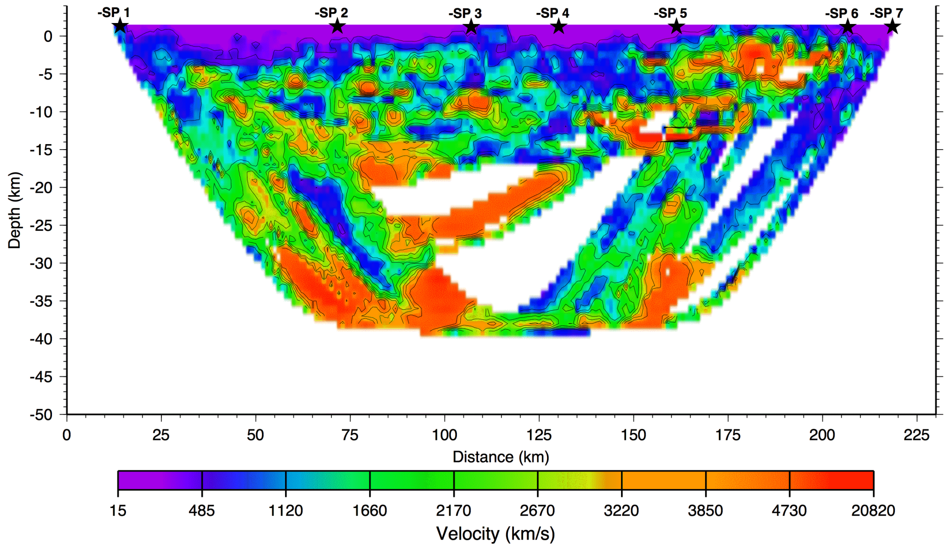
Determining Earth . . .
Need for Expert . . .
Main Idea of an Algorithm
Need for Data Fusion
Towards an Algorithm
Need for . . .
Uncertainty of the . . .
Beyond Probabilistic . . .
Case Study: Seismic . . .
Home Page
Title Page
◀◀ ▶▶
◀ ▶
Page 33 of 43
Go Back
Full Screen
Close
Quit


W
E
SP 1
SP 2
SP 3
SP 4
SP 5
SP 6
SP 7
Depth (km)
A.
1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5 7.0 7.5 8.0 8.5
Velocity (km/s)
B.
0 50 100 150 200
Distance (km)
1 10 50 90 130 180 230 300 1225
Coverage (Rays/cell)

32. Estimating Uncertainty, First Try: Probabilistic Approach



33. Estimating Uncertainty, Second Try: Interval Approach



34. Towards a Better Estimate: Revisiting Estimation Algorithms Under Probabilistic and Interval Uncertainty

- *Linearization:* $\Delta y = \sum\limits_{i=1}^n c_i \cdot \Delta x_i$, where $c_i \stackrel{\text{def}}{=} \dfrac{\partial f}{\partial x_i}$.
- *Formulas:* $\sigma^2 = \sum\limits_{i=1}^n c_i^2 \cdot \sigma_i^2$, $\Delta = \sum\limits_{i=1}^n |c_i| \cdot \Delta_i$.
- *Numerical differentiation:* n iterations, too long.
- *Monte-Carlo approach:* if Δx_i are Gaussian w/σ_i, then $\Delta y = \sum\limits_{i=1}^n c_i \cdot \Delta x_i$ is also Gaussian, w/desired σ.
- *Advantage:* # of iterations does not grow with n.
- *Interval estimates:* if Δx_i are Cauchy, w/$\rho_i(x) = \dfrac{\Delta_i}{\Delta_i^2 + x^2}$, then $\Delta y = \sum\limits_{i=1}^n c_i \cdot \Delta x_i$ is also Cauchy, w/desired Δ.

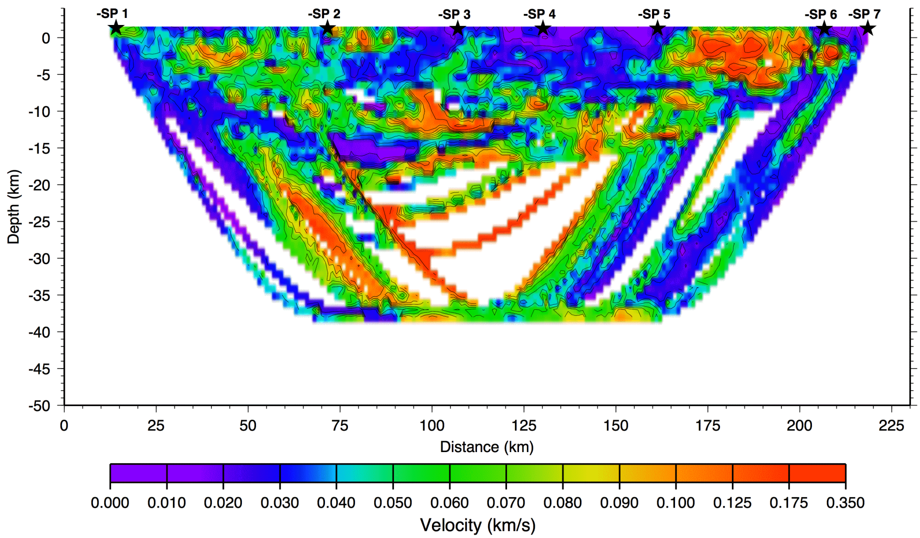
35. Resulting Fast (Linearized) Algorithm for Estimating Interval Uncertainty

- Apply f to $\widetilde{x}_i$: $\widetilde{y} := f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$;
- For $k = 1, 2, \ldots, N$, repeat the following:
  - use RNG to get $r_i^{(k)}$, $i = 1, \ldots, n$ from $U[0,1]$;
  - get st. Cauchy values $c_i^{(k)} := \tan(\pi \cdot (r_i^{(k)} - 0.5))$;
  - compute $K := \max_i |c_i^{(k)}|$ (to stay in linearized area);
  - simulate "actual values" $x_i^{(k)} := \widetilde{x}_i - \delta_i^{(k)}$, where $\delta_i^{(k)} := \Delta_i \cdot c_i^{(k)}/K$;
  - simulate error of the indirect measurement:
  $$\delta^{(k)} := K \cdot \left(\widetilde{y} - f\left(x_1^{(k)}, \ldots, x_n^{(k)}\right)\right);$$
- Solve the ML equation $\sum\limits_{k=1}^{N} \frac{1}{1 + \left(\frac{\delta^{(k)}}{\Delta}\right)^2} = \frac{N}{2}$ by bisection, and get the desired Δ.

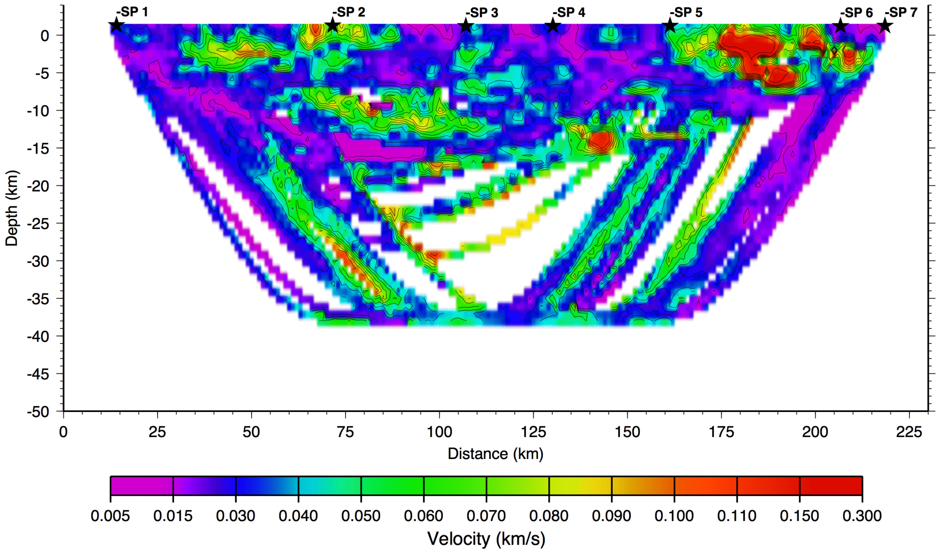
36. A New (Heuristic) Approach

- *Problem:* guaranteed (interval) bounds are too high.
- *Gaussian case:* we only have bounds guaranteed with confidence, say, 90%.
- *How:* cut top 5% and low 5% off a normal distribution.
- *New idea:* to get similarly estimates for intervals, we "cut off" top 5% and low 5% of Cauchy distribution.
- *How:*
  - find the threshold value x_0 for which the probability of exceeding this value is, say, 5%;
  - replace values x for which $x > x_0$ with x_0;
  - replace values x for which $x < -x_0$ with $-x_0$;
  - use this "cut-off" Cauchy in error estimation.
- *Example:* for 95% confidence level, we need $x_0 = 12.706$.

37. Heuristic Approach: Results with 95% Confidence Level



38. Heuristic Approach: Results with 90% Confidence Level



39. Conclusions

- *In the past:* communications were much slower.
- *Conclusion:* use centralization.
- *At present:* communications are much faster.
- *Conclusion:* use cyberinfrastructure.
- *Related problems:*
  - gauge the the uncertainty of the results obtained by using cyberinfrastructure;
  - which data points contributed most to uncertainty;
  - how an improved accuracy of these data points will improve the accuracy of the result.
- *We described:* algorithms for solving these problems.
- *Additional problem:* what if interval estimates are too wide and probabilistic estimates are too narrow.

40. Acknowledgment

This work was supported in part:

- by the National Science Foundation grants:
  - HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and
  - DUE-0926721,
- and by Grant 1 T36 GM078000-01 from the National Institutes of Health.

My special thanks to the Ministry of Communications and Information Technologies of the Republic of Azerbaijan.