Towards Combining Probabilistic, Interval, Fuzzy Uncertainty, and Constraints: An Example Using the Inverse Problem in Geophysics

V. Kreinovich, A. A. Velasco, M. G. Averill
R. Araiza, G. R. Keller, and G. Xiang
Departments of Computer Science and Geological Sciences
University of Texas at El Paso
El Paso, TX 79968, USA, vladik@utep.edu

1. Cyberinfrastructure: What It Does Now

- *Situation:*
  - large amounts of data are stored in different locations;
  - algorithms for processing this data are also implemented at different locations;
  - the ability to use the data and algorithms is not shared
- *Traditional solution:* centralization.
- *Drawback:* excessive workload.
- *New solution:* web services (cyberinfrastructure) allow users to submit requests without worrying about the geographic locations of different computational resources (databases and programs).
- Web services enable the user to receive:
  - the desired data $x_1, \ldots, x_n$ and
  - the results $y = f(x_1, \ldots, x_n)$ of processing this data.

2. Cyberinfrastructure: What It Should Do

- *Known fact:* data x_i usually come from measurements.
- *Uncertainty:* in general, the measured values $\widetilde{x}_i$ are different from the actual (unknown) values x_i:
$$\widetilde{x}_i \neq x_i$$
- *Result:* the result $\widetilde{y}$ of data processing is, in general, different from the actual value y of the desired quantity:
$$\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n) \neq y = f(x_1, \ldots, x_n).$$
- *Problem:* gauge this difference.

3. Combining Uncertainty Models Within Cyberinfrastructure – A Challenging Problem

- *The main objective of cyberinfrastructure* is to be able:
  - to seamlessly move data between different databases

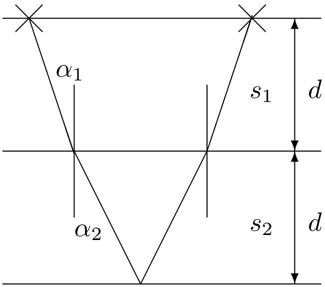
    (where this data is stored in different formats),
  - to feed the combined data into a remotely located program

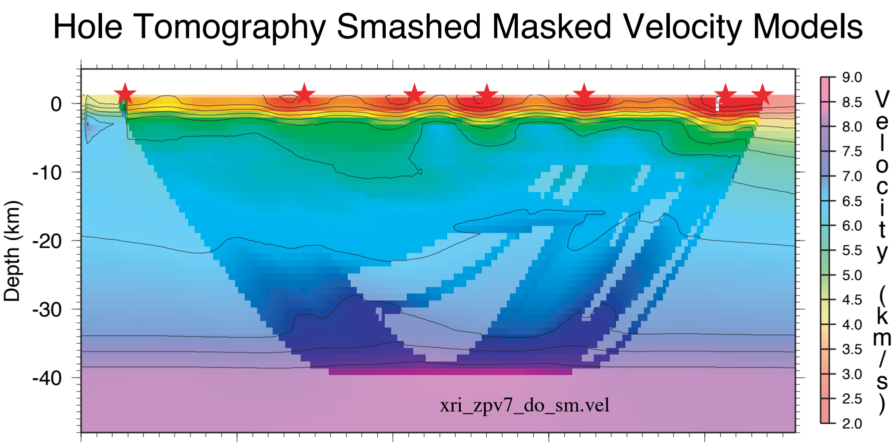
    (which may require yet another data format),

    and
  - to return the result to the user.
- *Additional objectives of cyberinfrastructure:* it is also important to gauge the quality and accuracy of the result.
- *Situation:* we may have different models for describing uncertainty of different databases and programs.
- *Conclusion:* it is important to be able to combine different uncertainty models.

4. Case Study: The Seismic Inverse Problem

- *We know:* traveltimes x_i between different locations (based on passive or active seismic measurements).
- *We would like to find:* slowness values.
- *Most widely used method:* Hole's tomographic inversion code
  - we start with the initial slowness model $s_j^{(0)}$;
  - based on the current model, we find the ray paths;
  - based on the ray paths, we predict the traveltimes $t_i = \sum\limits_j \ell_{ij} \cdot \Delta s_j$;
  - based on the differences $\Delta t_i \stackrel{\text{def}}{=} \widetilde{t}_i - t_i$, we adjust the slownesses: $s_j^{(k+1)} = s^{(k)} + \Delta s_j$, where Δs_j is the average of $\Delta s_{ij} \stackrel{\text{def}}{=} \dfrac{\Delta t_i}{L_i}$,
  - repeat until $\dfrac{1}{n} \cdot \sum (\Delta t_i)^2 \le \sigma^2$.





Hole Tomography Smashed Masked Velocity Models
Depth (km)
0
-10
-20
-30
-40
xri_zpv7_do_sm.vel
Velocity (km/s)
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0

5. Seismic Inverse Problem: Remaining Problems and Possible Solutions

- *Room for improvement:*
  - the resulting model is sometimes not very accurate;
  - it is often not clear how accurate the resulting model is;
  - the convergence is often slow – sometimes, requires time-consuming manual adjustment of the initial model;
  - the inversion is often unstable because known features are not included in the starting model.
- *Natural solution:*
  - The existing techniques only take into account the measurement results.
  - Expert geophysicists have prior knowledge about the structure of the region.
  - This additional knowledge can:
    * improve the accuracy of the geophysical structures;
    * speed up convergence.

6. Probabilistic Prior Knowledge

- *Measurement uncertainty:* measurement results often come with probabilistic uncertainty.
- *Probabilistic uncertainty:* we know the probability distribution $\rho_i(\Delta x_i)$ of different values $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$ of measurement errors.
- *It is possible to use this probabilistic uncertainty:* as a result of processing such measurement results, we usually obtain
  - not only the estimate $\widetilde{y}$ of the desired quantity, but also
  - the standard deviation σ (and other statistical characteristics) of possible differences $\widetilde{y} - y$ between the estimated $\widetilde{y}$ and the actual y values of the desired quantity.
- *Some prior knowledge comes from measurements:* prior knowledge sometimes comes from previous measurement processing.
- *Probabilistic form of prior knowledge:* such prior knowledge has the form

$$y \approx y^{(0)}, \text{ with standard deviation } \sigma^{(0)}.$$

- *Probabilistic prior knowledge has been used:* this prior knowledge has been successfully used, e.g., in tomography of Central Asia.

7. The Need to Use Interval and Fuzzy Prior Knowledge

- *Prior knowledge as bounds:* in some situations, we have a different type of prior knowledge.
- *Example:* a geophysicist may know that the speed of sound at a certain depth must be between 6 and 8 km/s.
- *Natural idea:* assume that 7 is the most probable value and 1 is the standard deviation.
- *Problem:* the expert does not necessarily consider any value within the interval [6,8] to be more probable than others.
- *Alternative idea:* assume that the distribution is uniform.
- *Problem:* the probability to have a velocity close to 6 at all cells is very small, so statistical approach will dismiss it.
- However, from the geophysicist viewpoint, it is quite possible that $v \approx 6$ for all the cells.
- *New idea:* explicitly keep the interval as the prior data.
- *Fuzzy uncertainty:*
  - in addition to an interval that is guaranteed to contain the slownesses,
  - the expert can often produce several nested intervals corresponding to different degrees of certainty.

8. How We Can Use Interval Uncertainty

- *How algorithms work now:*
  - start with a reasonable velocity model;
  - predict traveltimes t_i between stations;
  - use the difference $\Delta t_i \stackrel{\text{def}}{=} \widetilde{t}_i - t_i$, where $\widetilde{t}_i$ are measured values, to adjust the velocity model:
    * divide Δt_i by the length L_i of the ray path;
    * add the $\Delta t_i / L_i$ to all slownesses along the ray path;
    * for each cell, take an average of the adjusted slownesses corresponding to different ray paths.
- *How to modify* when we know the interval $[\underline{s}_j, \overline{s}_j]$ of possible slownesses:
  - first, we compute the next approximation $s_j^{(k+1)}$ to the slownesses,
  - then, we replace $s_j^{(k+1)}$ with the nearest value within the interval $[\underline{s}_j, \overline{s}_j]$:
    * we take $s_j = \overline{s}_j$ if $s_j^{(k+1)} > \overline{s}_j$, and
    * we take $s_j = \underline{s}_j$ if $s_j^{(k+1)} < \underline{s}_j$.
- *Problem:* we do not get fully compensate Δt_i.
- *Better idea:* we distribute the whole difference Δt_i between cells in such a way that all resulting slownesses are within $[\underline{s}_j, \overline{s}_j]$.

9. How Accurate Is The Resulting Velocity Model

- *Sources of uncertainty:*
  - measurement errors in measuring traveltimes;
  - incomplete coverage.
- *First idea:* use standard statistical techniques (e.g., Monte-Carlo simulations).
- *For passive seismic data:* reasonable error bounds.
- *For active seismic data:* since error decreases as $\sim 1/\sqrt{n}$, where n is the number of measurements, we get accuracy 10 m/sec at 30 km depth.
- *Explanation:*
  - we implicitly assumed that all measurement errors are independent, but
  - all measurements may have the same systematic error component.
- *Natural idea:* worst-case analysis, without any independence assumptions.
- *Problem:* too pessimistic error estimates.
- *Combined idea:*
  - *bootstrap:* comparing results from the two halves of data;
  - *checkerboard:* determines the size of the cells about which we make accuracy estimates.

10. Work in Progress

- *Real-life situations:*
  - some prior knowledge comes from prior data processing,
  - some prior knowledge comes from prior interval constraints.
- *Conclusion:* we must combine different types of uncertainty.
- *The ultimate goal:* to provide visual presentations of the combined effects of different types of uncertainty.

11. Acknowledgments

This work was supported in part:

- by NSF grant EAR-0225670,
- by NASA under cooperative agreement NCC5-209,
- by NIH grant 3T34GM008048-20S1,
- by Army Research Lab grant DATM-05-02-C-0046,
- by Star Award from the University of Texas System,
- and by Texas Department of Transportation grant No. 0-5453.