Probabilistic and Interval Uncertainty of the Results of Data Fusion, with Application to Geosciences

Christian Servin[1], Omar Ochoa[1], and Aaron A. Velasco[2]
[1]Department of Computer Science
[2]Department of Geological Sciences
University of Texas at El Paso
El Paso, TX 79968, USA
contact christians@miners.utep.edu

1. Data Fusion under Interval Uncertainty: Reminder

- *Frequent practical situation:*
  - we are interested in a quantity u;
  - we have several measurements and/or expert estimates $u_1, \ldots, u_n$ of u.
- *Objective:* fuse these estimates into a single more accurate estimate.
- *Interval case:* each u_i is known with interval uncertainty.
- *Formal description:* for each i, we know the interval $\mathbf{u}_i = [u_i - \Delta_i, u_i + \Delta_i]$ containing u.
- *Solution:* u belongs to the intersection $\mathbf{u} \stackrel{\text{def}}{=} \bigcap\limits_{i=1}^{n} \mathbf{u}_i$ of these intervals.

2. Data Fusion under Probabilistic Uncertainty: Reminder

- *Probabilistic uncertainty:* each measurement error $\Delta u_i \stackrel{\text{def}}{=} u_i - u$ is normally distributed w/0 mean and known σ_i.
- *Technique:* the Least Squares Method (LSM)

$$\sum_{i=1}^{n} \frac{(u-u_i)^2}{2\sigma_i^2} \to \min .$$

- *Resulting estimate:* is

$$u = \frac{\sum\limits_{i=1}^{n} u_i \cdot \sigma_i^{-2}}{\sum\limits_{i=1}^{n} \sigma_i^{-2}}.$$

- *Standard deviation:*

$$\sigma^2 = \frac{1}{\sum\limits_{i=1}^{n} \sigma_i^{-2}}, \qquad \text{with } \sigma^2 \ll \sigma_i^2.$$

3. New Problem: Different Resolution

- *Traditional data fusion:* fusing measurement results with different accuracy.
- *Additional problem:* different measurements also have different resolution.
- *Case study – geosciences:* estimating density $u_1, \ldots, u_n$ at different locations and depths.
- *Examples* of different geophysical estimates:
  - *Seismic data* leads to higher-resolution estimates $\widetilde{u}_1, \ldots, \widetilde{u}_n$ of the density values.
  - *Gravity data* leads to lower-resolution estimates, i.e., estimates $\widetilde{u}$ for the weighted average

$$u = \sum_{i=1}^{n} w_i \cdot u_i.$$

4. Why This Is Important

- *Reminder:* there are many sources of data for Earth models:
  - first-arrival passive seismic data (from the actual earthquakes),
  - first-arrival active seismic data (from the seismic experiments),
  - gravity data,
  - surface waves, etc.
- *At present:* each of these datasets is processed separately, resulting in several different Earth models.
- *Fact:* these models often provide complimentary geophysical information.
- *Idea:* all these models describe the properties of the same Earth, so it is desirable to combine them.

5. New Idea: Model Fusion

- *Objective:* to combine the information contained in multiple complementary datasets.
- *Ideal approach:* it is desirable to come up with techniques for joint inversion of these datasets.
- *Problem:* designing such joint inversion techniques is an important theoretical and practical challenge.
- *Status:* such joint inversion methods are being developed.
- *Practical question:* what to do while these methods are being developed?
- *Our practical solution:* fuse the Earth models coming from different datasets.

6. Model Fusion: Statistical Case

- *Objective:* find the values $u_1, \ldots, u_n$ of the desired quantity in different spatial cells.
- *Geophysical example:* u_i is the density at different

$$1\text{ km } \times 1\text{ km } \times 1\text{ km cells.}$$

- *Input:* we have
  - high-resolution measurements, i.e., values $\widetilde{u}_i \approx u_i$ with st. dev. σ_i;
  - lower-resolution measurements, i.e., values $\widetilde{u}^{(k)}$ corresponding to blocks of neighboring cells:

$$\widetilde{u}^{(k)} \approx \sum_i w_i^{(k)} \cdot u_i, \text{ with st. dev. } \sigma^{(k)}.$$

- *Additional information:* a lower-resolution measurement result is representative of values within the block.

7. Model Fusion: Statistical Case (cont-d)

- *Formal description:* when $w_i^{(k)} \neq 0$, we have $\widetilde{u}^{(k)} \approx u_i$, with st. dev. $\delta^{(k)}$.
- *How to estimate $\delta^{(k)}$:* as the empirical st. dev. within the block.
- *High-resolution values (reminder):* $\widetilde{u}_i \approx u_i$ w/st. dev. σ_i.
- *Lower-resolution values (reminder):*

$$\widetilde{u}^{(k)} \approx \sum_i w_i^{(k)} \cdot u_i, \text{ with st. dev. } \sigma^{(k)}.$$

- *LSM Solution:* minimize the sum

$$\sum_i \frac{(u_i - \widetilde{u}_i)^2}{\sigma_i^2} + \sum_i \sum_k \frac{(u_i - \widetilde{u}^{(k)})^2}{(\delta^{(k)})^2} + \sum_k \frac{\left(\widetilde{u}^{(k)} - \sum_i w_i^{(k)} \cdot u_i\right)^2}{(\sigma^{(k)})^2}.$$

- *How:* differentiating w.r.t. u_i, we get a system of linear equations with unknowns $u_1, \ldots, u_n$.

8. Model Fusion: Interval Case

- *Quantities of interest:* values $u_1, \ldots, u_n$ of the desired quantity in different spatial cells.
- *Objective:* find the ranges $\mathbf{u}_1, \ldots, \mathbf{u}_n$ of possible values of $u_1, \ldots, u_n$.
- *High-resolution measurements:* values $\widetilde{u}_i \approx u_i$ with bound Δ_i:
$$\widetilde{u}_i - \Delta_i \le u_i \le \widetilde{u}_i + \Delta_i.$$
- *Lower-resolution measurements:* values $\widetilde{u}^{(k)}$ corresponding to blocks of neighboring cells:
$$\widetilde{u}^{(k)} \approx \sum_i w_i^{(k)} \cdot u_i, \text{ with bound } \Delta^{(k)}.$$
- *Resulting constraint:*
$$\widetilde{u}^{(k)} - \Delta^{(k)} \le \sum_i w_i^{(k)} \cdot u_i \le \widetilde{u}^{(k)} + \Delta^{(k)}.$$

9. Model Fusion: Interval Case (cont-d)

- *Additional information:* a priori bounds on u_i:

$$\underline{u}_i \le u_i \le \overline{u}_i.$$

- *Additional information:* a priori bounds on the changes between neighboring cells:

$$-\delta_{ij} \le u_i - u_j \le \delta_{ij}.$$

- *High-resolution measurements (reminder):*

$$\widetilde{u}_i - \Delta_i \le u_i \le \widetilde{u}_i + \Delta_i.$$

- *Lower-resolution measurements (reminder):*

$$\widetilde{u}^{(k)} - \Delta^{(k)} \le \sum_i w_i^{(k)} \cdot u_i \le \widetilde{u}^{(k)} + \Delta^{(k)}.$$

- *Objective:* minimize and maximize each u_i under these constraints.

10. Model Fusion: Interval Solution

- *Problem.* Minimize (Maximize) u_i under the following constraints:
  - $\underline{u}_i \le u_i \le \overline{u}_i$.
  - $-\delta_{ij} \le u_i - u_j \le \delta_{ij}$.
  - $\widetilde{u}_i - \Delta_i \le u_i \le \widetilde{u}_i + \Delta_i$.
  - $\widetilde{u}^{(k)} - \Delta^{(k)} \le \sum\limits_i w_i^{(k)} \cdot u_i \le \widetilde{u}^{(k)} + \Delta^{(k)}$.
- *Current solution method:* linear programming.
- *Objective:* provide more efficient algorithms for specific geophysical cases.
- *Preliminary results:* some such algorithms have been developed.

11. Preliminary Experiments

- *What we have done:* preliminary proof-of-concept experiments.
- *Simplifications:*
  - equal weights w_i;
  - simplified datasets.
- *Conclusion:* the fused model improves accuracy and resolution of different Earth models.

12. Acknowledgments

This work was supported in part:

- by NSF Cyber-ShARE grant HRD-0734825;
- by the Alliance for Graduate Education and Professoriate (AGEP);
- by the Louis Strokes Alliance for Minority Participation (LSAMP).