Fusing Continuous and Discrete Data, on the Example of Merging Seismic and Gravity Models in Geophysics

Omar Ochoa[1,2], Aaron Velasco[1,3], Vladik Kreinovich[1,2]
[1]Cyber-ShARE Center of Excellence: A Center for Sharing Cyberresources to Advance Science and Education
Departments of [2]Computer Science and [3]Geological Sciences
University of Texas at El Paso, El Paso, TX 79968, USA
omar@miners.utep.edu, velasco@geo.utep.edu, vladik@utep.edu

***Abstract*—In many application areas, we need to fuse continuous and discrete models of the same phenomena. For example, in geophysics, we have two main models for describing how the sound velocity changes with location and depth: a discrete gravity-based model, in which we have several layers with abrupt transition between layers, and a seismic model, in which the velocity continuously changes with the change in location and depth – and a transition is represented by a steeper change. Due to inevitable uncertainty, in two fused models, the same actual transition is placed at slightly different depths.**

If we simply fuse these models, the fused model will inherit both nearby transitions and therefore, will, misleadingly, correspond to *two* nearby transitions instead of one. It is therefore necessary, before fusing, to first get a fused (more accurate) location of the transition surface.

In this paper, we show how to find such a location.

I. Formulation of the Problem

Need for model fusion. In many applications area, there are several different sources of data.

For example, in geophysics, one of the main problems is determining the density ρ at different locations and different depths. It is known that for most minerals, density is uniquely related to the speed of sound, so determining density ρ at different locations and depths is equivalent to determining the sound velocity v at different locations and depths.

In geophysics, the two main sources of data for determining the density ρ (or, equivalently, the sound velocity v):

- the *seismic data*, i.e., the arrival times of signal from earthquake (passive seismic data) and from the experimental explosions (active seismic data); and
- the *gravity* data, i.e., the values of the gravitational force at different locations.

Both data provide complementary information about the density:

- seismic data provides information about a narrow zone around a path from the source to the sensor, enabling us to determine density around this path with a high spatial resolution; on the other hand, seismic data only covers a direct vicinity of the paths, so it leaves many area barely covered
- on the other hand, the gravity value at a given location is affected by the densities in the wide area around this location; thus, the gravity data provides information about the larger area – but with much smaller spatial resolution.

It is therefore desirable to use both types of data when estimating the values of density at different locations and lengths – what geophysicists call a *density model*; see, e.g., [1], [2], [3], [5], [6].

At present, there are no efficient algorithms for processing both types of data. So, to use both types of models, we must fuse the results of processing these two types of data: a seismic model that is obtained by processing seismic data, and a gravity model that is obtained by processing gravity data; algorithms for model fusions are described in [8], [9].

Computational problem: need to fuse discrete and continuous models. Traditionally, seismic models are *continuous* in the sense that in these models, the velocity smoothly changes as we change the location and/or depth. In contrast, the gravity models are *discrete*: in these models, we have layers, in each of which the velocity is constant, with an abrupt transition between layers.

The abrupt transition corresponds to a steep change in the continuous model. The problem is that both models describe the location of the transition only approximately, the corresponding transitions are located at slightly different depths. So, if we simply combine the corresponding values value-by-value, e.g., by taking a weighted average of values corresponding to different locations and depths, then the resulting fused model will have *two* different abrupt transitions instead of one:

- one transition where the continuous model has it, and
- another transition nearby where the discrete model has it.

What we plan to do. To avoid the misleading double-transition models, it is desirable, before fusing the models, to first fuse the corresponding transition locations. In this paper, we provide an algorithm for such location fusion.

Specifically,

- we formulate the problem both in probabilistic terms (see, e.g., [11]) and in fuzzy terms (see, e.g., [4], [7]), and

- we show that both approaches result in the exact same transition location.

The fact that two different approaches lead to the same location reassures us that this location is reasonable.

II. Available Data: What is Known and What Needs to Be Determined

For each location, in the discrete model, we have the exact depth z_d at which we have a transition between the two layers. In contrast, for the continuous model, we do not have the abrupt transition; instead, we have velocity values $v(z)$ at different depths. We must therefore extract the corresponding transition value z_c from the velocity values.

To be more precise, we have values $v_1, v_2, \ldots, v_i, \ldots, v_n$ corresponding to different depths. We need to find i for which the transition occurs between the depths i and $i+1$.

III. Probabilistic Approach

Description of the model. It is reasonable to assume that, with the exception of the transition point, for all other values j, the difference $\Delta v_j \stackrel{\text{def}}{=} v_j - v_{j+1}$ is small. This difference is caused by many different factors, so it is reasonable to invoke the Central Limit Theorem and assume that this difference is normally distributed with 0 mean and some standard deviation σ; see, e.g., [11]. The corresponding probability density is equal to

$$p_j \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\cdot\pi}\cdot\sigma}\cdot\exp\left(-\frac{1}{2\cdot\sigma^2}\cdot(\Delta v_j)^2\right).$$

We assume that differences corresponding to different depths j are independent.

The value Δv_i at the transition depth i is *not* described by the normal distribution, it has to be given separately.

The resulting model is described by three parameters:

- the standard deviation σ,
- the transition depth i, and
- the transition values Δv_i.

Due to independence of different depth, the overall likelihood L_i of the model with given values of these parameters is determined by the formula

$$L_i = \prod_{j\ne i} p_j = \prod_{j\ne i}\frac{1}{\sqrt{2\cdot\pi}\cdot\sigma}\cdot\exp\left(-\frac{1}{2\cdot\sigma^2}\cdot(\Delta v_j)^2\right).$$

How to find the location: the general idea of the Maximum Likelihood Approach. In the probabilistic approach, we usually select the parameters for which the likelihood of the observed data is the largest; see, e.g., [11]. In other words, in this Maximum Likelihood Approach, we select the values of the parameters for which the likelihood L attains the largest possible value.

How to find the optimal location i_0: analysis of the corresponding optimization problem. Due to the fact that $\exp(a)\cdot\exp(b) = \exp(a+b)$, the expression L_i can be represented as

$$L_i = \frac{1}{(\sqrt{2\cdot\pi}\cdot\sigma)^{n-2}}\cdot\exp\left(-\frac{1}{2\cdot\sigma^2}\cdot\sum_{j\ne i}(\Delta v_j)^2\right).$$

The factor in front of the exponent does not depend on the location i at all, so L_i is the largest if and only if the exponential term is the largest:

$$\exp\left(-\frac{1}{2\cdot\sigma^2}\cdot\sum_{j\ne i}(\Delta v_j)^2\right)\to\max_i.$$

The function $\exp(-z)$ is strictly decreasing, so it attains its largest possible values when z is the smallest. Thus, to find the optimal location i, we must find the value i for which the following expression is the smallest:

$$\frac{1}{2\cdot\sigma^2}\cdot\sum_{j\ne i}(\Delta v_j)^2\to\min_i.$$

Again, the factor in front of the sum does not depend on i, so this expression is the smallest if and only the sum attains its smallest value:

$$\sum_{j\ne i}(\Delta v_j)^2\to\min_i.$$

This sum can be represented as

$$\sum_{j\ne i}(\Delta v_j)^2 = \sum_{j=1}^{n-1}(\Delta v_j)^2 - (\Delta v_i)^2.$$

The first term in this expression does not depend on i at all. Thus, the above difference is the smallest if and only if the value $(\Delta v_i)^2$ is the largest. This, in turn, is equivalent to $|\Delta v_i|$ being the largest.

Thus, we arrive at the following conclusion.

Resulting location. As the most probable location of the transition point, we select the depth i_0 for which the absolute value $|\Delta v_i|$ of the difference $\Delta v_i = v_{i+1} - v_i$ is the largest possible.

Comment. This conclusion seems to be very reasonable: the most probable location of the actual abrupt transition between the layers is the depth at which the measured difference is the largest.

IV. Fuzzy Approach

Formulation of the model. Intuitively, for each depth i, our confidence that this is a transition point depends on the actual value of the corresponding difference $|\Delta v_i|$:

- the smaller the difference, the less confident we are that this is the actual transition depth, and
- the larger the difference, the more confident we are that this is the actual transition depth.

In general, we can therefore assume that the degree of confidence d_i that the transition occurs at the depth i is equal to $f(|\Delta v_i|)$, for some monotonically increasing function $f(z)$.

How to find the location: the general idea. In fuzzy techniques, if we need to select a single location i_0, it is reasonable to select a value for which our degree of confidence is the largest

$$d_i = f(|\Delta v_i|) \to \max .$$

How to find the optimal location i: analysis of the corresponding optimization problem. Since the function $f(z)$ is strictly increasing, it attains its largest possible values when z is the largest.

Thus, to find the optimal location i_0, we must find the value i for which the expression $|\Delta v_i|$ is the largest possible.

Resulting location is the same as in the probabilistic approach. It is worth mentioning that both probabilistic and fuzzy approaches lead to the same location.

Since fuzzy logic describes commonsense reasoning, the fact that fuzzy approach leads to the same location as the probabilistic approach means that the original probabilistic model is in good accordance with common sense – and this increase our confidence in this result.

V. How Accurate Is This Location Estimate?

Formulation of the problem. The location i that we obtained is approximate. How accurate is this location estimate?

Which approach should we use to solve this problem? We know that – at least for location – both models leads to the same result. It is therefore reasonable to select one of the models.

Since probability theory have been developed for centuries, so more methods and techniques have been developed – we will use the probabilistic approach.

Auxiliary result: estimating σ. In the probabilistic model, in addition to the location i, we also need to select the standard deviation σ.

We have already shown that the value i can be determined by the Maximum Likelihood method. A similar Maximum Likelihood approach can be used to determine σ. Specifically, we can find σ for which the expression

$$L_i = \frac{1}{(\sqrt{2\cdot\pi}\cdot\sigma)^{n-2}}\cdot\exp\left(-\frac{1}{2\cdot\sigma^2}\cdot\sum_{j\neq i}(\Delta v_j)^2\right)$$

takes the largest possible value over all possible i and σ. We already know that with respect to i, the largest value is attained when i is equal to the above estimate i_0, so we can simply plus in i_0 into the above expression:

$$L_{i_0} = \frac{1}{(\sqrt{2\cdot\pi}\cdot\sigma)^{n-2}}\cdot\exp\left(-\frac{1}{2\cdot\sigma^2}\cdot\sum_{j\neq i_0}(\Delta v_j)^2\right)$$

and maximize the result over σ.

Since the function $-\ln(z)$ is strictly increasing, this is equivalent to finding σ for which the value $\psi \stackrel{\text{def}}{=} -\ln(L_{i_0})$ is the smallest possible. This value has the form

$$\psi = (n-2)\cdot\ln(2\cdot\pi) + (n-2)\cdot\ln(\sigma) + \frac{1}{2\cdot\sigma^2}\cdot\sum_{j\neq i_0}(\Delta v_j)^2.$$

Differentiating this expression by σ and equating the derivative to 0, we conclude that

$$(n-2)\cdot\frac{1}{\sigma} - \frac{1}{\sigma^3}\cdot\sum_{j\neq i_0}(\Delta v_j)^2 = 0.$$

Multiplying both sides by σ^3, dividing both sides by $n-2$, and moving the term σ^2 to the other side, we conclude that

$$\sigma^2 = \frac{1}{n-2}\cdot\sum_{j\neq i_0}(\Delta v_j)^2.$$

Resulting probability distribution. Now, that we know the values of all the parameters, the probability distribution is uniquely determined: the probability P_i that the actual transition is at location i is proportional to

$$L_i \sim \exp\left(-\frac{1}{2\cdot\sigma^2}\cdot\sum_{j\neq i}(\Delta v_j)^2\right).$$

By using the above formula

$$\sum_{j\neq i}(\Delta v_j)^2 = \sum_{j=1}^{n-1}(\Delta v_j)^2 - (\Delta v_i)^2$$

and the fact that $\exp(a-b) = \exp(a)\cdot\exp(b)$, we conclude that

$$\exp\left(-\frac{1}{2\cdot\sigma^2}\cdot\sum_{j\neq i}(\Delta v_j)^2\right) = A\cdot S_i,$$

where we denoted

$$A \stackrel{\text{def}}{=} \exp\left(-\frac{1}{2\cdot\sigma^2}\cdot\sum_{j=1}^{n-1}(\Delta v_j)^2\right),$$

and

$$S_i \stackrel{\text{def}}{=} \exp\left(\frac{(\Delta v_i)^2}{2\cdot\sigma^2}\right).$$

Thus, $P_i \sim S_i$, i.e.,

$$P_i = c\cdot S_i$$

for some constant c. This constant c can be determined from the fact that the transition has to be somewhere, so $\sum\limits_{j=1}^{n-1} P_j = 1$. Thus,

$$1 = \sum_{j=1}^{n-1} P_j = c\cdot\sum_{j=1}^{n-1} S_j = 1,$$

hence

$$c = \frac{1}{\sum\limits_{j=1}^{n-1} S_j},$$

and, finally, the probability $P_i = c \cdot S_i$ takes the form

$$P_i = \frac{S_i}{\sum\limits_{j=1}^{n-1} S_j} = \frac{\exp\left(\dfrac{(\Delta v_i)^2}{2\cdot\sigma^2}\right)}{\sum\limits_{j=1}^{n-1} \exp\left(\dfrac{(\Delta v_j)^2}{2\cdot\sigma^2}\right)}.$$

How to estimate accuracy: the idea. The mean square deviation σ_0^2 of the actual (unknown) transition depth from our estimate i_0 is, by definition, equal to

$$\sigma_0^2 = \sum_{i=1}^{n-1} (i - i_0)^2 \cdot P_i.$$

Substituting the above expression for P_i into this formula, we conclude that

$$\sigma_0^2 = \frac{\sum\limits_{i=1}^{n-1} (i-i_0)^2 \cdot \exp\left(\dfrac{(\Delta v_i)^2}{2\cdot\sigma^2}\right)}{\sum\limits_{j=1}^{n-1} \exp\left(\dfrac{(\Delta v_j)^2}{2\cdot\sigma^2}\right)}.$$

Resulting algorithm. First, we compute

$$\sigma^2 = \frac{1}{n_2} \cdot \sum_{j \neq i_0} (\Delta v_j)^2,$$

and then estimate σ_0 by using the above formula.

This algorithm leads to a reasonable result. We applied this algorithm to the seismic model of El Paso area derived in [1] (see also [10]); for this map, we got $\sigma_0 \approx 1.5$ km, which fits well with the model.

VI. How to Fuse the Estimates of the Transition Depth and How to Fuse the Corresponding Models

Available estimates for the transition depth. Now, we have two estimates for the transition depth:

- the estimate i_d from the discrete (gravity) model, and
- the estimate i_0 from the continuous (seismic) model.

Accuracy of the available estimates for the transition depth. The estimate i_d corresponding to the discrete model comes from a standard statistical analysis – as one of the parameters of the model. So, we can use the usual statistical techniques to estimate the standard deviation σ_d of this estimate.

For the continuous estimate i_0, we already know how to compute its standard deviation σ_0.

How to fuse estimates of the transition depth: analysis of the problem. We would like to use the Maximum Likelihood Method to find the best fused estimate i_f for the actual (unknown) transition depth i.

It is reasonable to assume that both differences $i_d - i$ and $i_0 - i$ are normally distributed and independent. The probability densities corresponding to $i_d - i$ and to $i_0 - i$ are therefore proportional to

$$\exp\left(-\frac{(i_d - i)^2}{2\cdot\sigma_d^2}\right) \text{ and } \exp\left(-\frac{(i_0 - i)^2}{2\cdot\sigma_0^2}\right).$$

Since these uncertainties are independent, the likelihood of i being the actual transition depth is proportional to the product

$$\exp\left(-\frac{(i_d - i_f)^2}{2\cdot\sigma_d^2}\right) \cdot \exp\left(-\frac{(i_0 - i_f)^2}{2\cdot\sigma_0^2}\right) =$$

$$\exp\left(-\left(\frac{(i_d - i)^2}{2\cdot\sigma_d^2} + \frac{(i_0 - i)^2}{2\cdot\sigma_0^2}\right)\right).$$

Maximizing this likelihood expression is equivalent to minimizing the argument of the decreasing function $\exp(-z)$, i.e., minimizing the expression

$$\frac{(i_d - i)^2}{2\cdot\sigma_d^2} + \frac{(i_0 - i)^2}{2\cdot\sigma_0^2}.$$

Differentiating this expression by i and equating the derivative to 0, we get the following result.

How to fuse estimates of the transition depth: resulting formula. Based on the estimates i_d and i_0 for the transition depth, as the optimal estimate i_f for the actual transition depth, we take the following value:

$$i_f = \frac{i_d \cdot \sigma_d^{-2} + i_0 \cdot \sigma_0^{-2}}{\sigma_d^{-2} + \sigma_0^{-2}}.$$

Towards fusing actual maps. The fused value i_f is our best estimate for the transition depth, i.e., for the border between the lower and upper zones.

In the discrete model:

- values corresponding to $i < i_d$ correspond to the upper zone, while
- values corresponding to the depths $i > i_d$ correspond to the lower zone.

Similarly, in the continuous model:

- values corresponding to $i < i_0$ correspond to the upper zone, while
- values corresponding to the depths $i > i_0$ correspond to the lower zone.

So, for depths $i \le \min(i_0, i_d)$ and $i \ge \max(i_0, i_d)$, both models correctly describe the zone, and we can simply fuse the values from both models – e.g., similarly to how we fused the estimates for the transition depth.

For intermediate depths, we need to adjust the models, by replacing the values corresponding to the wrong zone by the nearest value from the correct zone. As a result, we get the following procedure.

How to fuse the actual maps: general idea.

- First, we adjust both models so that they both have a transition at depth i_f.

- Second, for each depth i, we merge the values v'_i and v''_i corresponding to the adjusted models.

Let us describe this fusion in more detail.

Adjusting the discrete model. Adjusting the discrete model is (relatively) easy: we just replace the original depth i_d with the new (more accurate) fused value i_f.

Adjusting the continuous model. When the more accurate transition depth i_f is smaller than the transition depth i_0 corresponding to the continuous model ($i_f < i_0$), this means that the values at depths i between i_f and i_0 are erroneously assigned to the the upper zone. In this case, the values v_i for this i must be replaced by the the value of the nearest point at the lower zone, i.e., by the value v_{i_0+1}.

When the more accurate transition depth i_f is larger than the transition depth i_0 corresponding to the continuous model ($i_f > i_0$), this means that the values at depths i between i_0 and i_f are erroneously assigned to the the lower zone. In this case, the values v_i for this i must be replaced by the the value of the nearest point at the upper zone, i.e., by the value v_{i_0}.

How to merge the adjusted models. For each depth i, we now have two adjusted values v'_i and v''_i corresponding to two adjusted models. Let σ' and σ'' be the corresponding standard deviations. Then, similarly to what we have described earlier, we can compute the fused value $\widetilde{v}_i$ as follows:

$$\widetilde{v}_i = \frac{v'_i \cdot (\sigma')^{-2} + v''_i \cdot (\sigma'')^{-2}}{(\sigma')^{-2} + (\sigma'')^{-2}}.$$

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation grants HRD-0734825 (Cyber-ShARE Center) and DUE-0926721, by Grant 1 T36 GM078000-01 from the National Institutes of Health, by Grant MSM 6198898701 from MŠMT of Czech Republic, and by Grant 5015 “Application of fuzzy logic with operators in the knowledge based systems” from the Science and Technology Centre in Ukraine (STCU), funded by European Union.

REFERENCES

[1] M. G. Averill, *A Lithospheric Investigation of the Southern Rio Grande Rift*, University of Texas at El Paso, Department of Geological Sciences, PhD Dissertation, 2007.

[2] M. G. Averill, K. C. Miller, G. R. Keller, V. Kreinovich, R. Araiza, and S. A. Starks, “Using Expert Knowledge in Solving the Seismic Inverse Problem” *International Journal of Approximate Reasoning*, 2007, Vol. 45, No. 3, pp. 564–587.

[3] J. A. Hole, “Nonlinear High-Resolution Three-Dimensional Seismic Travel Time Tomography”, *Journal of Geophysical Research*, 1992, Vol. 97, pp. 6553–6562.

[4] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Upper Saddle River, New Jersey: Prentice Hall, 1995.

[5] J. M. Lees and R. S. Crosson, “Tomographic inversion for three-dimensional velocity structure at Mount St. Helens using earthquake data”, *Journal of Geophysical Research*, 1989, Vol. 94, pp. 5716–5728.

[6] M. Maceira, S. R. Taylor, C. J. Ammon, X. Yang, and A. A. Velasco. “High-resolution Rayleigh wave slowness tomography of Central Asia”, *Journal of Geophysical Research*, 2005, Vol. 110, Paper B06304.

[7] H. T. Nguyen and E. A. Walker, *First Course on Fuzzy Logic*, CRC Press, Boca Raton, Florida, 2006.

[8] O. Ochoa, A. A. Velasco, V. Kreinovich, and C. Servin, “Model fusion: a fast, practical alternative towards joint inversion of multiple datasets”, *Abstracts of the Annual Fall Meeting of the American Geophysical Union AGU'08*, San Francisco, California, December 15–19, 2008.

[9] O. Ochoa, A. A. Velasco, C. Servin, and V. Kreinovich, “Model Fusion under Probabilistic and Interval Uncertainty, with Application to Earth Sciences”, In: M. Beer, R. L. Muhanna, and R. L. Mullen (Eds.), *Proceedings of the 4th International Workshop on Reliable Engineering Computing REC'2010*, Singapore, March 3–5, 2010, pp. 81–100.

[10] C. Servin, M. G. Averill, N. Del Rio, L. Longpré, and V. Kreinovich, “Propagation and Provenance of Probabilistic and Interval Uncertainty in Cyberinfrastructure-Related Data Processing and Data Fusion”, In: R. L. Muhanna and R. L. Mullen (eds.), *Proceedings of the International Workshop on Reliable Engineering Computing REC'08*, Savannah, Georgia, February 20–22, 2008, pp. 199–234.

[11] D. G. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall/CRC Press, Boca Raton, Florida, 2007.