MODEL FUSION: A NEW APPROACH TO PROCESSING HETEROGENOUS DATA

OMAR OCHOA

Department of Computer Science

APPROVED:

Vladik Kreinovich, Chair, Ph.D.

Ann Q. Gates, Ph.D.

Aaron Velasco, Ph.D.

Benjamin C. Flores, Ph.D.
Dean of the Graduate School

MODEL FUSION: A NEW APPROACH TO PROCESSING HETEROGENOUS DATA

by

OMAR OCHOA

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Department of Computer Science

THE UNIVERSITY OF TEXAS AT EL PASO

July 2013

Abstract

In many practical situations, it is necessary to extract information from data of different types. For example, in geosciences, several sources of data can be used to determine the interior structure of the earth: the first arriving waves from earthquakes and from man-made sources, measurements of the earth's gravity field, measurements of the dispersion of surface waves generated by earthquakes, etc. At present, most existing data processing techniques deal only with data of one type.

A joint use of all the information derived from multiple types of data sources is an important theoretical and practical challenge. Such a joint use would provide the best description of the object of study; for example, in geosciences, such a joint "inversion" would represents the best model of the interior structure of the earth. While such combination methods are being developed, as a first step, this thesis proposes a practical solution: to fuse the models generated from different datasets by the existing data processing techniques.

In geosciences and in other applications, models generated from different datasets not only have different accuracy and coverage, but also different spatial resolution. This thesis describes how to optimally fuse such models under interval and probabilistic uncertainty. Additional ideas are described that take into into consideration the inherent discrete vs. continuous nature of different models, and how to gauge accuracy of each of the fused models. The resulting techniques can be used in various application domains to merge models of different accuracy, coverage, and spatial resolution.

Table of Contents

Chapter 1

Introduction

In many practical situations, it is necessary to extract information from data of different types. For example, in order to understand the interior structure of the earth, geophysicists apply a number of techniques to gather data about to the structure of the earth. At present, most existing data processing techniques deal only with data of one type. For example, different data processing algorithms are applied to different geophysical datasets to create *models*, multi-dimensional "images" of the earth. Each model is then analyzed to gain an understanding of the interior structure of the earth. In this way, scientists can use the data to find natural resources or to explain some natural phenomenon.

The resulting datasets are often complementary of each other, each of them contains information that is not available in other datasets. Therefore, the more datasets are used to generate a model, the more accurate the resulting model represents the interior structure of the earth. Ideally, it is desirable to use all different datasets. However, a joint use of all the information derived from multiple types of data sources remains an important theoretical and practical challenge. While such combination methods are being developed, as a first step, this thesis proposes a practical solution: to fuse the models generated from different datasets by the existing data processing techniques.

1.1 Motivation

In many areas of science and engineering, there are different sources of data. For example, in geophysics, there are many sources of data that are used to create models of the Earth:

- first-arrival passive seismic data (from actual earthquakes); see, e.g., [6];

- first-arrival active seismic data (from seismic experiments using man-made sources); see, e.g., [1, 4];
- gravity data; and
- surface waves; see, e.g., [7].

As stated before, datasets coming from different sources provide complimentary information. For example, different geophysical datasets contain different information on earth structure. In general:

- some of the datasets provide better accuracy and/or spatial resolution in some spatial areas and in some depths, while
- other datasets provide a better accuracy and/or spatial resolution in other areas or depths.

For example:

- each measured gravity anomaly at a point is the result of the density distribution over a relatively large region of the earth, so estimates based on gravity measurements have (relatively) low spatial resolution;
- in contrast, each seismic data point (arrival time) comes from a narrow trajectory (ray) a seismic wave travels within the earth, so the spatial resolution corresponding to this data is much higher.

Usually, there are several different geophysical datasets available. At present, each of these datasets is often processed separately, resulting in several different models reflecting different aspects of the studied phenomena. It is therefore desirable to combine data from different datasets.

1.2 Approach

The ideal approach would be to use all the datasets to produce a single model. At present, however, in many research areas – including geophysics – there are no efficient algorithms for simultaneously processing all the different datasets. The more datasets that this approach has to process, the more complex this approach becomes. Currently, designing such joint inversion techniques presents an important theoretical and practical challenge.

As an alternative, while research is being done on developing a practical and feasible join inversion techniques, this work proposes to combine all the datasets in an indirect way by fusing the resulting models from each technique. This practical approach called **model fusion** must take into consideration different attributes of the data, such as resolution, accuracy and coverage. This approach leverages all the current existing models into one unified model that merges the characteristics of each dataset.

1.3 Thesis Overview

This thesis is organized as follows: Chapter 1 introduces this research and the work described in this document. Chapter 2 presents the underlying idea of fusing data using both probabilistic and interval uncertainty. Chapter 3 presents the new proposed approach of model fusion. Chapter 4 describes the challenge of fusing continuous and discrete models. Chapter 5 describes the challenge of estimating the accuracy for models. Finally, Chapter 6 discusses the conclusions and future work of this research.

The work presented in this thesis has been published in [12, 13, 14].

Chapter 2

Data Fusion: Reminder

The underlying foundation of this approach is the fusing of data: how these values can be combined under both a probabilistic and interval approach.

In many real-life situations, there are several measurements and/or expert estimates $\widetilde{x}^{(1)}, \ldots, \widetilde{x}^{(n)}$ of the same quantity x.

- These values may come from the actual (direct) measurements of the quantity x.
- Alternatively, these values may come from *indirect* measurements of x, i.e., from different models, in which, based on the corresponding measurement results, the i-th model leads to an estimate $\widetilde{x}^{(i)}$ for x.

In such situations, it is desirable to fuse these estimates into a single more accurate estimate for x; see, e.g., [17].

2.1 Data Fusion: Probabilistic Uncertainty

Let's start with the case when each estimate $\widetilde{x}^{(i)}$ is known with the (traditionally described) probabilistic uncertainty, e.g., when each estimation error $\Delta x^{(i)} \stackrel{\text{def}}{=} \widetilde{x}^{(i)} - x$ is normally distributed with 0 mean and known standard deviation $\sigma^{(i)}$, and estimation errors $\Delta x^{(i)}$ corresponding to different models are independent.

In practice, the estimation errors are indeed often normally distributed. This empirical fact can be justified by the Central Limit Theorem, according to which, under certain reasonable conditions, the joint effect of many relatively small errors is (approximately) normally distributed; see, e.g., [19]. For each model based on measurements of a certain type

(e.g., gravity or seismic), not only the resulting error of each measurement comes from many different error sources, but also each estimate comes from several different measurements – thus further increasing the number of different error components contributing to the estimation error.

In this case, the probability density for each estimation error $\Delta x^{(i)}$ has the form

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

and the probability density $\rho(x)$ corresponding to all n estimates is (due to independence) the product of these densities:

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

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

As a single estimate x for the desired quantity, it is reasonable to select the value for which this probability (density) $\rho(x)$ is the largest (i.e., to use the *Maximum Likelihood* method). Since $\exp(z)$ is an increasing function, maximizing a function $A\cdot\exp(-B(x))$ is equivalent to minimizing $B(x)$, arriving at the following *Least Squares* approach: find x for which the sum $\sum\limits_{i=1}^{n}\frac{(\widetilde{x}^{(i)}-x)^2}{2\cdot(\sigma^{(i)})^2}$ is the smallest possible.

Differentiating this expression with respect to x and equating the derivative to 0, concludes that $x=\dfrac{\sum\limits_{i=1}^{n}\widetilde{x}^{(i)}\cdot(\sigma^{(i)})^{-2}}{\sum\limits_{i=1}^{n}(\sigma^{(i)})^{-2}}$. The accuracy of this fused estimate can be described by the standard deviation σ for which $\sigma^{-2}=\sum\limits_{i=1}^{n}(\sigma^{(i)})^{-2}$.

2.2 Data Fusion: Interval Uncertainty

In some practical situations, the value x is known with interval uncertainty, i.e., the only available information is the interval

$$\mathbf{x}^{(i)}=[\widetilde{x}^{(i)}-\Delta^{(i)},\widetilde{x}^{(i)}+\Delta^{(i)}]$$

containing the actual (unknown) value of x. This happens, e.g., when only the upper bound $\Delta^{(i)}$ is known on each estimation error $\Delta x^{(i)}$: $|\Delta x^{(i)}| \le \Delta^{(i)}$. In this case, from the fact that the estimate is $\widetilde{x}^{(i)}$, it can be concluded that $|x - \widetilde{x}^{(i)}| \le \Delta^{(i)}$, i.e., that $\widetilde{x}^{(i)} - \Delta^{(i)} \le x \le \widetilde{x}^{(i)} + \Delta^{(i)}$.

For interval uncertainty, it is easy to fuse several estimates. Based on each estimate $\widetilde{x}^{(i)}$, it is possible to conclude that the actual value x belongs to the interval $\mathbf{x}^{(i)}$. Thus, the (unknown) actual value x belongs to the intersection

$$\mathbf{x} \stackrel{\text{def}}{=} \bigcap_{i=1}^{n} \mathbf{x}^{(i)} = [\max(\widetilde{x}^{(i)} - \Delta^{(i)}), \min(\widetilde{x}^{(i)} + \Delta^{(i)})]$$

of these intervals.

Chapter 3

Model Fusion: A New Approach

3.1 Proposed Solution: Model Fusion

In many areas of science and engineering, there is a need to jointly process data coming from different data sources. Usually, the data from each source is processed separately, resulting in a separate model. It is therefore desirable to merge these models into a single model that incorporates all the available data.

Different models have different accuracy and different spatial resolution. For example, in the geosciences,

- seismic data leads to estimates of the density at different locations and depths which have higher spatial resolution (based on an empirical relationship between density and seismic velocity), while
- gravity data leads to estimates of the same densities which have lower spatial resolution.

Towards precise formulation of the problem. Estimates with higher spatial (spatio-temporal) resolution mean that estimates of the values correspond to small spatial (spatio-temporal) cells. An estimate with a lower spatial resolution means that its results are affected by several neighboring spatial cells, i.e., that the estimate is in effect, a weighted average of the values in several neighboring cells.

What is given. In precise terms:

- the high-resolution estimates $\widetilde{x}_1, \ldots, \widetilde{x}_n$ of the values $x_1, \ldots, x_n$ are given for several small spatial cells; these estimates correspond to models with a higher spatial resolution
- additionally, the estimates $\widetilde{X}_j$ are given for the weighted averages $X_j = \sum\limits_{i=1}^{n} w_{j,i} \cdot x_i$; these estimates correspond to models with a lower spatial resolution.

It is assumed that the values of the weights $w_{j,i}$ are known. This assumption makes sense for geophysical problems, because in these problems, these weights are indeed known. For example:

- It is known how exactly the gravity at a given point depends on the densities at different spatial locations.
- Additionally, it is know how the travel time depends on the density distribution: specifically, how exactly the travel time of a seismic signal depends on the velocity distribution, and the empirical velocity-density relationship is also known.

In some applications, however, the corresponding weights are only approximately known. In such situations, when fusing the models, it must also be taken into account the uncertainty with which these weights are known. For these applications, it is desirable to extend the techniques – to accommodate such more complex situations.

What the objective is. The values x_i are of interest. Therefore, based on the estimates $\widetilde{x}_i$ and $\widetilde{X}_j$, more accurate estimates must be provided for x_i.

Example. In the geophysical example, the values of interest are the densities x_i.

The proposed approach. In this chapter, it is described how to fuse estimates with different accuracy and spatial resolution:

- In the case of probabilistic uncertainty, the Least Squares Method is used to derive explicit formulas for combining the estimates $\widetilde{x}_i$ and $\widetilde{X}_j$.

- In the case of interval uncertainty, an efficient algorithm is provided for estimating the ranges of x_i.

3.2 Model Fusion: Case of Probabilistic Uncertainty

3.2.1 General Case

Main idea. The solution to the model fusion problem is to take into account several different types of approximate equalities:

- Each estimate $\widetilde{x}_i$ from a model with a high spatial resolution is approximately equal to the actual value x_i in the corresponding (smaller size) cell i, with the known accuracy $\sigma_{h,i}$:

$$\widetilde{x}_i \approx x_i.$$

- Each estimate $\widetilde{X}_j$ from (one of the) models with a lower spatial resolution is approximately equal to the weighted average of values of all the smaller cells $x_{i(1,j)}, \ldots, x_{i(k_j,j)}$ within the corresponding larger size cell, with a known accuracy $\sigma_{l,j}$:

$$\widetilde{X}_j \approx \sum_i w_{j,i} \cdot x_i,$$

for known weights $w_{j,i} \geq 0$ for which $\sum\limits_{i=1}^{n} w_{j,i} = 1$. In the simple case, these weights are equal:

$$\widetilde{X}_j \approx \frac{x_{i(1,j)} + \ldots + x_{i(k_j,j)}}{k_j}.$$

- Usually, there is also a prior knowledge of the values x_i. It is reasonable to assume that this prior knowledge can also be described by a normal distribution, with the mean $x_{pr,i}$ and the standard deviation $\sigma_{pr,i}$:

$$x_i \approx x_{pr,i}.$$

(The case when for some i, there is no prior information at all is equivalent to setting $\sigma_{pr,i} = \infty$.)

- Finally, each estimate $\widetilde{X}_j$ from a model with a lower spatial resolution is approximately equal to the value within each of the constituent smaller size cells $x_{i(l,j)}$, with the accuracy corresponding to the (empirical) standard deviation $\sigma_{e,j}$ of the smaller-cell values within the larger cell:

$$\widetilde{X}_j \approx x_{i(l,j)},$$

where $\sigma_{e,j}^2 \stackrel{\text{def}}{=} \dfrac{1}{k_j} \cdot \sum_{l=1}^{k_j} \left(\widetilde{x}_{i(l,j)} - E_j\right)^2$, and $E_j \stackrel{\text{def}}{=} \dfrac{1}{k_j} \cdot \sum_{l=1}^{k_j} \widetilde{x}_{i(l,j)}$.

The Least Squares technique can be used to combine these approximate equalities, and to find the desired combined values x_i by minimizing the resulting sum of weighted squared differences.

Relation between different standard deviations. As mentioned earlier, there is usually a trade-off between accuracy and spatial resolution:

- estimates with a higher spatial resolution usually have lower accuracy, i.e., higher values of the standard deviation $\sigma_{h,i}$;
- on the other hand, estimates with a lower spatial resolution, i.e., estimates corresponding to a larger spatial area, have higher accuracy can be attained, i.e., lower values of the standard deviation $\sigma_{l,j} \ll \sigma_{h,i}$.

From the mathematical viewpoint, this trade-off makes sense. An estimate for a model with a low spatial resolution is an average of the values corresponding to high spatial resolution, and averaging usually decreases the approximation error: $\sigma_{l,j} \ll \sigma_{h,i} \ll \sigma_{e,j}$.

It should be mentioned that while usually, higher spatial resolution estimates have lower accuracy, sometimes, a higher-resolution model has more accuracy in some places. For example, in the geosciences,

- the measurements from a borehole provide the most accurate estimates of the corresponding quantities,

- and for these measurements, the spatial location is also known with a very high accuracy.

Resulting formulas for the general case. According to the Least Squares approach, in the general case, minimizing the following expression:

$$\sum_{i=1}^{n} \frac{(x_i - \widetilde{x}_i)^2}{\sigma_{h,i}^2} + \sum_{j=1}^{m} \frac{1}{\sigma_{l,j}^2} \cdot \left(\widetilde{X}_j - \sum_{i=1}^{n} w_{j,i} \cdot x_i \right)^2 +$$

$$\sum_{i=1}^{n} \frac{(x_i - x_{pr,i})^2}{\sigma_{pr,i}^2} + \sum_{j=1}^{m} \sum_{l=1}^{k_j} \frac{(\widetilde{X}_j - x_{i(l,j)})^2}{\sigma_{e,j}^2}.$$

In this general case, differentiation with respect to x_i leads to the following system of linear equations:

$$\frac{x_i - \widetilde{x}_i}{\sigma_{h,i}^2} + \sum_{j:j \ni i} \frac{w_{j,i}}{\sigma_{l,j}^2} \cdot \left(\sum_{i'=1}^{n} w_{j,i'} \cdot x_{i'} - \widetilde{X}_j \right) + \frac{x_i - x_{pr,i}}{\sigma_{pr,i}^2} + \sum_{j:j \ni i} \frac{x_i - \widetilde{X}_j}{\sigma_{e,j}^2} = 0,$$

where $j \ni i$ means that the j-th estimate corresponding to a model with a low spatial resolution covers the i-th cell.

Towards simplification of fusing prior estimates with estimates from a model with a high spatial resolution. For each cell i for which there are both a prior estimate $x_{pr,i}$ and an estimate $\widetilde{x}_i$ from a model with a higher spatial resolution, these two estimates can be fused by using the above-described standard data fusion technique. As a result, instead of the two terms $\sigma_{h,i}^{-2} \cdot (x_i - \widetilde{x}_i) + \sigma_{pr,i}^{-2} \cdot (x_i - x_{pr,i})$, there is a single term $\sigma_{f,i}^{-2} \cdot (x_i - x_{f,i})$, where $x_{f,i} \stackrel{\text{def}}{=} \dfrac{\widetilde{x}_i \cdot \sigma_{h,i}^{-2} + x_{pr,i} \cdot \sigma_{pr,i}^{-2}}{\sigma_{h,i}^{-2} + \sigma_{pr,i}^{-2}}$ and $\sigma_{f,i}^{-2} \stackrel{\text{def}}{=} \sigma_{h,i}^{-2} + \sigma_{pr,i}^{-2}$. The same formula can be used if there is only a high spatial resolution estimate or if there is only a prior estimate:

- If there is only a high spatial resolution estimate but no prior estimate, then it is possible to use $\sigma_{pr,i}^{-2} = 0$ (i.e., $\sigma_{pr,i} = \infty$).
- If there is only a prior estimate but no high spatial resolution estimate, then it is possible to use $\sigma_{h,i}^{-2} = 0$ (i.e., $\sigma_{h,i} = \infty$).

As a result of this fusion, the following simplified formulas are generated.

Resulting formulas as simplified equations.

$$\frac{x_i - x_{f,i}}{\sigma_{f,i}^2} + \sum_{j:j\ni i} \frac{w_{j,i}}{\sigma_{l,j}^2} \cdot \left(\sum_{i'=1}^{n} w_{j,i'} \cdot x_{i'} - \widetilde{X}_j \right) + \sum_{j:j\ni i} \frac{x_i - \widetilde{X}_j}{\sigma_{e,j}^2} = 0.$$

Solving this system of linear equations. Known algorithms for solving this system of linear equations can be used.

It is worth mentioning that usually, these algorithms require that the system be represented in the standard form $Ax = b$. To represent these system of equations in this form, it is necessary move, to the right-hand side, all the terms which do not contain unknowns.

3.2.2 Case of a Single Estimate with Low Spatial Resolution

Description. Consider the simplest case, when when there is exactly one estimate $\widetilde{X}_1$ from a model with a low spatial resolution. There are also prior estimates and the estimates with high spatial resolution for *some* of the cells.

This situation is typical in geosciences: e.g.,

- there is an estimate originating from the gravity measurements (with a lower spatial resolution) which covers a huge area in depth, and
- there are estimates originating from seismic measurements (corresponding to higher spatial resolution) which only cover depths above the Moho surface (the base of the earth's crust).

For convenience, the cells will be numbered in such a way that the cells for which there is either a prior estimates or estimates from a high spatial resolution model are listed first. Let h denote the total number of such cells.

For these h cells, as the result of combining prior estimates and estimates corresponding to high spatial resolution model(s), h values $x_{f,1}$, $x_{f,2}$, ..., $x_{f,h}$ are obtained.

Derivation. In this case, the above system of linear equations takes the following form: for $i = 1, \ldots, h$,

$$\sigma_{f,i}^{-2} \cdot (x_i - x_{f,i}) + \frac{1}{\sigma_{l,1}^2} \cdot w_{1,i} \cdot \left(\sum_{i'} w_{1,i'} \cdot x_{i'} - \widetilde{X}_1 \right) + \frac{1}{\sigma_{e,1}^2} (x_i - \widetilde{X}_1) = 0;$$

and for $i > h$,

$$\frac{1}{\sigma_{l,1}^2} \cdot w_{1,i} \cdot \left(\sum_{i'} w_{1,i'} \cdot x_{i'} - \widetilde{X}_1 \right) + \frac{1}{\sigma_{e,1}^2} (x_i - \widetilde{X}_1) = 0.$$

For $i \le h$, multiplying both sides by $\sigma_{f,i}^2$ results in

$$x_i - x_{f,i} + \frac{\sigma_{f,i}^2}{\sigma_{l,1}^2} \cdot w_{1,i} \cdot \left(\sum_{i'} w_{1,i'} \cdot x_{i'} - \widetilde{X}_1 \right) + \frac{\sigma_{f,i}^2}{\sigma_{e,1}^2} \cdot (x_i - \widetilde{X}_1) = 0.$$

Introducing an auxiliary variable

$$\mu \stackrel{\text{def}}{=} \frac{1}{\sigma_{l,1}^2} \cdot \left(\sum_{i'} w_{1,i'} \cdot x_{i'} - \widetilde{X}_1 \right)$$

results in the equation

$$x_i - x_{f,i} + w_{1,i} \cdot \sigma_{f,i}^2 \cdot \mu + \frac{\sigma_{f,i}^2}{\sigma_{e,1}^2} \cdot (x_i - \widetilde{X}_1) = 0.$$

Keeping terms proportional to x_i in the left-hand side and moving all the other terms to the right-hand side results in $\left(1 + \dfrac{\sigma_{f,i}^2}{\sigma_{e,1}^2}\right) \cdot x_i = x_{f,i} - w_{1,i} \cdot \sigma_{f,i}^2 \cdot \mu + \dfrac{\sigma_{f,i}^2}{\sigma_{e,1}^2} \cdot \widetilde{X}_1$, hence

$$x_i = \frac{x_{f,i}}{1 + \dfrac{\sigma_{f,i}^2}{\sigma_{e,1}^2}} - \frac{w_{1,i} \cdot \sigma_{f,i}^2}{1 + \dfrac{\sigma_{f,i}^2}{\sigma_{e,1}^2}} \cdot \mu + \widetilde{X}_1 \cdot \frac{\dfrac{\sigma_{f,i}^2}{\sigma_{e,1}^2}}{1 + \dfrac{\sigma_{f,i}^2}{\sigma_{e,1}^2}}.$$

For $i > h$, similarly $x_i - \widetilde{X}_1 + w_{1,i} \cdot \sigma_{e,1}^2 \cdot \mu = 0$, hence $x_i = \widetilde{X}_1 - w_{1,i} \cdot \sigma_{e,1}^2 \cdot \mu$.

To make this expression practically useful, μ must be described in terms of the given values $\widetilde{x}_i$ and $\widetilde{X}_1$. The auxiliary quantity μ is defined in terms of the weighted average of the values x_i, namely, in terms of the combination

$$\sum_{i=1}^{n} w_{1,i} \cdot x_i = \sum_{i=1}^{h} w_{1,i} \cdot x_i + \sum_{i=h+1}^{n} w_{1,i} \cdot x_i,$$

where

$$\sum_{i=1}^{h} w_{1,i} \cdot x_i = \sum_{i=1}^{h} \frac{w_{1,i} \cdot x_{f,i}}{1 + \frac{\sigma_{f,i}^2}{\sigma_{e,1}^2}} - \mu \cdot \sum_{i=1}^{h} \frac{w_{1,i}^2 \cdot \sigma_{f,i}^2}{1 + \frac{\sigma_{f,i}^2}{\sigma_{e,1}^2}} + \widetilde{X}_1 \cdot \sum_{i=1}^{h} \frac{w_{1,i} \cdot \frac{\sigma_{f,i}^2}{\sigma_{e,1}^2}}{1 + \frac{\sigma_{f,i}^2}{\sigma_{e,1}^2}}.$$

Similarly,

$$\sum_{i=h+1}^{n} w_{1,i} \cdot x_i = \left(\sum_{i=h+1}^{n} w_{1,i} \right) \cdot \widetilde{X}_1 - \left(\sum_{i=h+1}^{n} w_{1,i}^2 \right) \cdot \frac{\sigma_{e,1}^2}{\sigma_{l,1}^2} \cdot \mu.$$

By adding these two sums and subtracting $\widetilde{X}_1$, it follows that

$$\sigma_{l,1}^2 \cdot \mu = \sum_{i=1}^{n} w_{1,i} \cdot x_i - \widetilde{X}_1 = \sum_{i=1}^{h} w_{1,i} \cdot x_i + \sum_{i=h+1}^{n} w_{1,i} \cdot x_i - \widetilde{X}_1 =$$

$$\sum_{i=1}^{h} \frac{w_{1,i} \cdot x_{f,i}}{1 + \frac{\sigma_{f,i}^2}{\sigma_{e,1}^2}} - \mu \cdot \sum_{i=1}^{h} \frac{w_{1,i}^2 \cdot \sigma_{f,i}^2}{1 + \frac{\sigma_{f,i}^2}{\sigma_{e,1}^2}} + \widetilde{X}_1 \cdot \sum_{i=1}^{h} \frac{w_{1,i} \cdot \frac{\sigma_{f,i}^2}{\sigma_{e,1}^2}}{1 + \frac{\sigma_{f,i}^2}{\sigma_{e,1}^2}} +$$

$$\left(\sum_{i=h+1}^{n} w_{1,i} \right) \cdot \widetilde{X}_1 - \left(\sum_{i=h+1}^{n} w_{1,i}^2 \right) \cdot \sigma_{e,1}^2 \cdot \mu - \widetilde{X}_1.$$

From $\sum\limits_{i=1}^{n} w_{1,i} = \sum\limits_{i=1}^{h} w_{1,i} + \sum\limits_{i=h+1}^{n} w_{1,i} = 1$, it follows that

$$\left(\sum_{i=h+1}^{n} w_{1,i} \right) \cdot \widetilde{X}_1 - \widetilde{X}_1 = - \left(\sum_{i=1}^{h} w_{1,i} \right) \cdot \widetilde{X}_1$$

thus,

$$\widetilde{X}_1 \cdot \sum_{i=1}^{h} \frac{w_{1,i} \cdot \frac{\sigma_{f,i}^2}{\sigma_{e,1}^2}}{1 + \frac{\sigma_{f,i}^2}{\sigma_{e,1}^2}} + \left(\sum_{i=h+1}^{n} w_{1,i} \right) \cdot \widetilde{X}_1 - \widetilde{X}_1 =$$

$$\widetilde{X}_1 \cdot \sum_{i=1}^{h} \frac{w_{1,i} \cdot \frac{\sigma_{f,i}^2}{\sigma_{e,1}^2}}{1 + \frac{\sigma_{f,i}^2}{\sigma_{e,1}^2}} - \left(\sum_{i=1}^{h} w_{1,i} \right) \cdot \widetilde{X}_1 = -\widetilde{X}_1 \cdot \sum_{i=1}^{h} \frac{w_{1,i}}{1 + \frac{\sigma_{f,i}^2}{\sigma_{e,1}^2}}.$$

So, the equation for μ takes the following simplified form:

$$\sigma_{l,1}^2 \cdot \mu =$$

$$\sum_{i=1}^{h} \frac{w_{1,i} \cdot x_{f,i}}{1 + \dfrac{\sigma_{f,i}^2}{\sigma_{e,1}^2}} - \mu \cdot \sum_{i=1}^{h} \frac{w_{1,i}^2 \cdot \sigma_{f,i}^2}{1 + \dfrac{\sigma_{f,i}^2}{\sigma_{e,1}^2}} - \widetilde{X}_1 \cdot \sum_{i=1}^{h} \frac{w_{1,i}}{1 + \dfrac{\sigma_{f,i}^2}{\sigma_{e,1}^2}} - \left(\sum_{i=h+1}^{n} w_{1,i}^2 \right) \cdot \sigma_{e,1}^2 \cdot \mu.$$

By moving all terms containing μ to the left-hand side and all other terms to the right-hand side, the following explicit equation for μ is obtained.

Resulting formulas. First, compute the auxiliary value μ as $\mu = \dfrac{N}{D}$, where

$$N = \sum_{i=1}^{h} \frac{w_{1,i} \cdot (x_{f,i} - \widetilde{X}_1)}{1 + \dfrac{\sigma_{f,i}^2}{\sigma_{e,1}^2}} \text{ and } D = \sigma_{l,1}^2 + \sum_{i=1}^{h} \frac{w_{1,i}^2 \cdot \sigma_{f,i}^2}{1 + \dfrac{\sigma_{f,i}^2}{\sigma_{e,1}^2}} + \left(\sum_{i=h+1}^{n} w_{1,i}^2 \right) \cdot \sigma_{e,1}^2.$$

Then, compute the desired estimates for x_i, $i = 1, \ldots, h$, as

$$x_i = \frac{x_{f,i}}{1 + \dfrac{\sigma_{f,i}^2}{\sigma_{e,1}^2}} - \frac{w_{1,i} \cdot \sigma_{f,i}^2}{1 + \dfrac{\sigma_{f,i}^2}{\sigma_{e,1}^2}} \cdot \mu + \widetilde{X}_1 \cdot \frac{\dfrac{\sigma_{f,i}^2}{\sigma_{e,1}^2}}{1 + \dfrac{\sigma_{f,i}^2}{\sigma_{e,1}^2}},$$

and the estimates x_i for $i = h+1, \ldots, n$ as $x_i = \widetilde{X}_1 - w_{1,i} \cdot \sigma_{e,1}^2 \cdot \mu$.

3.2.3 Numerical Example

Description of the simplified case. To illustrate the above formulas, consider the simplest possible case, when there is exactly one estimate $\widetilde{X}_1$ from a lower spatial resolution model, and when:

- this estimate covers all n cells;
- all the weights are equal: $w_{1,i} = 1/n$;
- for each of n cells, there is an estimate corresponding to this cell that comes from a high spatial resolution model (i.e., $h = n$);
- all estimates coming from a high spatial resolution model have the same accuracy $\sigma_{h,i} = \sigma_h$;

- the estimate corresponding to a low spatial resolution model is much more accurate than the estimates corresponding to higher spatial resolution models $\sigma_{l,1} \ll \sigma_h$, so it can be safely assumed that $\sigma_l = 0$; and
- there is no prior information, so $\sigma_{pr,i} = \infty$ and thus, $x_{f,i} = \widetilde{x}_i$ and $\sigma_{f,i} = \sigma_h$.

To cover the cells for which there are no estimates from a high spatial resolution model, heuristic rule is defined in which the estimate from a lower spatial resolution model is approximately equal to the value within each of the constituent smaller size cells, with the accuracy corresponding to the (empirical) standard deviation $\sigma_{e,j}$. In the current simplified example, there are individual estimates for each cell, so there is no need for this heuristic rule. The corresponding heuristic terms in the general least squares approach are proportional to $\frac{1}{\sigma_{e,1}^2}$, so ignoring these terms is equivalent to taking $\sigma_{e,1}^2 = \infty$. Thus, $\frac{\sigma_{f,i}^2}{\sigma_{e,1}^2} = 0$ and $1 + \frac{\sigma_{f,i}^2}{\sigma_{e,1}^2} = 1$.

Because of this, and because of the fact that $w_{1,i} = \frac{1}{n}$ and $x_{f,i} = \widetilde{x}_i$, the formula for N takes the form

$$N = \sum_{i=1}^{n} \frac{1}{n} \cdot (\widetilde{x}_i - \widetilde{X}_1).$$

Opening parentheses and taking into account that the sum of n terms equal to $\frac{1}{n} \cdot \widetilde{X}_1$ is simply $\widetilde{X}_1$, results in

$$N = \frac{1}{n} \cdot \sum_{i=1}^{n} \widetilde{x}_i - \widetilde{X}_1.$$

Similarly, due to the simplifying assumptions $\sigma_{l,1} = 0$, $w_{1,i} = \frac{1}{n}$, $\sigma_{f,i} = \sigma_h$, $\sigma_{e,1} = 0$, and $h = n$, the formula for the denominator D takes the form

$$D = \sum_{i=1}^{n} \left(\frac{1}{n}\right)^2 \cdot \sigma_h^2 = \frac{1}{n} \cdot \sigma_h^2.$$

Thus,

$$\mu = \frac{N}{D} = \frac{\frac{1}{n} \cdot \sum_{i=1}^{n} \widetilde{x}_i - \widetilde{X}_1}{\frac{1}{n} \cdot \sigma_h^2}.$$

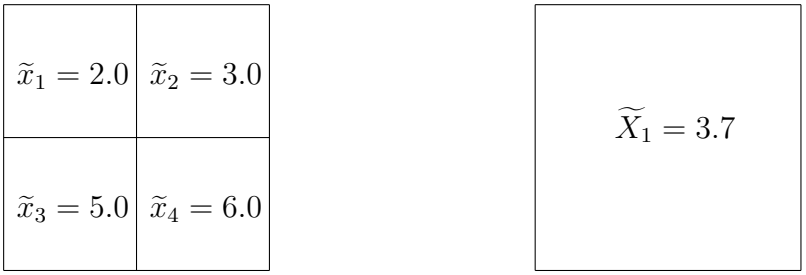

Figure 3.1: Higher and lower spatial resolution estimates

The formula for x_i now turns into

$$x_i = \widetilde{x}_i - \frac{1}{n} \cdot \sigma_h^2 \cdot \mu.$$

Substituting the above expression for μ into this formula results in

$$x_i = \widetilde{x}_i - \lambda,$$

where

$$\lambda \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} \widetilde{x}_i - \widetilde{X}_1.$$

Numerical example of the simplified case. Assume that there are $n = 4$ cells, and that the high spatial resolution estimates for these cells are $\widetilde{x}_1 = 2.0$, $\widetilde{x}_2 = 3.0$, $\widetilde{x}_3 = 5.0$ and $\widetilde{x}_4 = 6.0$. Additionally, assume that each of these estimates has the same accuracy $\sigma_h = 0.5$. And also assume that there is an estimate $\widetilde{X}_1 = 3.7$ for the average X_1 of these four values. Assume that this estimate has a much higher accuracy $\sigma_l \ll \sigma_h$ so that $\sigma_l \approx 0$.

Since it is assumed that the low spatial resolution estimates are accurate ($\sigma_l \approx 0$), it is therefore assumed that the estimated quantity, i.e., the arithmetic average of the four cell values, is practically exactly equal to this estimate $\widetilde{X}_1 = 3.7$:

$$\frac{x_1 + x_2 + x_3 + x_4}{4} \approx 3.7.$$

For the high spatial resolution estimates $\widetilde{x}_i$, the average is slightly different:

$$\frac{\widetilde{x}_1 + \widetilde{x}_2 + \widetilde{x}_3 + \widetilde{x}_4}{4} = \frac{2.0 + 3.0 + 5.0 + 6.0}{4} = 4.0 \neq 3.7.$$

$\widetilde{x}_1 = 1.7$	$\widetilde{x}_2 = 2.7$
$\widetilde{x}_3 = 4.7$	$\widetilde{x}_4 = 5.7$

Figure 3.2: The result of model fusion: simplified setting

This difference is caused by the fact that, in contrast to accurate low spatial resolution estimates, higher spatial resolution measurements are much less accurate: the corresponding estimation error has a standard deviation $\sigma_h = 0.5$. Therefore, as described above, the information from the low spatial resolution estimates to can be used to "correct" the high spatial resolution estimates.

In this particular example, since $\sigma_l \approx 0$, the correcting term takes the form

$$\lambda = \frac{\widetilde{x}_1 + \ldots + \widetilde{x}_n}{n} - \widetilde{X}_1 =$$

$$\frac{2.0 + 3.0 + 5.0 + 6.0}{4} - 3.7 = 4.0 - 3.7 = 0.3,$$

so the corrected ("fused") values x_i take the form:

$$x_1 = \widetilde{x}_1 - \lambda = 2.0 - 0.3 = 1.7; \quad x_2 = \widetilde{x}_2 - \lambda = 3.0 - 0.3 = 2.7;$$

$$x_3 = \widetilde{x}_3 - \lambda = 5.0 - 0.3 = 4.7; \quad x_4 = \widetilde{x}_4 - \lambda = 6.0 - 0.3 = 5.7;$$

For these corrected values, the arithmetic average is equal to

$$\frac{x_1 + x_2 + x_3 + x_4}{4} = \frac{1.7 + 2.7 + 4.7 + 5.7}{4} = 3.7,$$

i.e., exactly to the low spatial resolution estimate.

Taking $\sigma_{e,j}$ into account. What if, in the above numerical example, there is a requirement that the actual values in each cell are approximately equal to $\widetilde{X}_1$, with the accuracy $\sigma_{e,1}$ equal to the empirical standard deviation?

In this case, the above formulas take the form

$$N = \frac{1}{1 + \dfrac{\sigma_h^2}{\sigma_{e,1}^2}} \cdot \left(\frac{\widetilde{x}_1 + \ldots + \widetilde{x}_n}{n} - \widetilde{X}_1 \right)$$

and

$$D = \frac{1}{1 + \dfrac{\sigma_h^2}{\sigma_{e,1}^2}} \cdot \frac{1}{n} \cdot \sigma_h^2,$$

resulting in the exact same expression for μ:

$$\mu = \frac{N}{D} = \frac{\dfrac{1}{n} \cdot \sum_{i=1}^{n} \widetilde{x}_i - \widetilde{X}_1}{\dfrac{1}{n} \cdot \sigma_h^2}.$$

The formulas for the fused values x_i are now somewhat more complex:

$$x_i = \frac{\widetilde{x}_i - \lambda}{1 + \dfrac{\sigma_h^2}{\sigma_{e,1}^2}} + \widetilde{X}_1 \cdot \frac{\dfrac{\sigma_h^2}{\sigma_{e,1}^2}}{1 + \dfrac{\sigma_h^2}{\sigma_{e,1}^2}}.$$

Taking $\sigma_{e,j}$ into account for the numerical example. There is a requirement that the actual values in each cell are approximately equal to $\widetilde{X}_1$, with the accuracy $\sigma_{e,j}$ equal to the empirical standard deviation. In the example, the lower spatial resolution estimate $\widetilde{X}_1$ covers all four cells. In this example, the above condition takes the form $x_i \approx \widetilde{X}_1$, with the accuracy

$$\sigma_{e,1}^2 = \frac{1}{4} \cdot \sum_{i=1}^{4} (\widetilde{x}_i - E_1)^2,$$

where

$$E_1 = \frac{1}{4} \cdot \sum_{i=1}^{4} \widetilde{x}_i.$$

For the numerical example, as shown,

$$E_1 = \frac{1}{4} \cdot \sum_{i=1}^{4} \widetilde{x}_i = \frac{\widetilde{x}_1 + \widetilde{x}_2 + \widetilde{x}_3 + \widetilde{x}_4}{4} = 4.0$$

and thus,

$$\sigma_{e,1}^2 = \frac{(2.0-4.0)^2+(3.0-4.0)^2+(5.0-4.0)^2+(6.0-4.0)^2}{4} = \frac{4+1+1+4}{4} = \frac{10}{4} = 2.5,$$

hence $\sigma_{e,1} \approx 1.58$.

Now, the following formula can be used

$$x_i = \frac{1}{1+\dfrac{\sigma_h^2}{\sigma_{e,1}^2}} \cdot (\widetilde{x}_i - \lambda) + \frac{\dfrac{\sigma_h^2}{\sigma_{e,1}^2}}{1+\dfrac{\sigma_h^2}{\sigma_{e,1}^2}} \cdot \widetilde{X}_1$$

to find the corrected ("fused") values x_i. Here, $\sigma_h = 0.5$, $\sigma_{e,1}^2 = 2.5$, so

$$\frac{\sigma_h^2}{\sigma_{e,1}^2} = \frac{0.25}{2.5} = 0.1$$

and therefore, with two digit accuracy,

$$\frac{1}{1+\dfrac{\sigma_h^2}{\sigma_{e,1}^2}} = \frac{1}{1.1} \approx 0.91$$

and

$$\frac{\dfrac{\sigma_h^2}{\sigma_{e,1}^2}}{1+\dfrac{\sigma_h^2}{\sigma_{e,1}^2}} \cdot \widetilde{X}_1 = \frac{0.1}{1.1} \cdot 3.7 \approx 0.34.$$

Therefore,

$$x_1 \approx 0.91 \cdot (2.0-0.3) + 0.34 \approx 1.89;$$

$$x_2 \approx 0.91 \cdot (3.0-0.3) + 0.34 \approx 2.79;$$

$$x_3 \approx 0.91 \cdot (5.0-0.3) + 0.34 \approx 4.62;$$

$$x_4 \approx 0.91 \cdot (6.0-0.3) + 0.34 \approx 5.53.$$

The arithmetic average of these four values is equal to

$$\frac{x_1+x_2+x_3+x_4}{4} \approx \frac{1.89+2.79+4.62+5.53}{4} \approx 3.71,$$

i.e., within the computation accuracy (since all the computations were performed with two digits after the decimal point) coincides with the lower spatial resolution estimate $\widetilde{X}_1 = 3.7$.

$\widetilde{x}_1 \approx 1.89$	$\widetilde{x}_2 \approx 2.79$
$\widetilde{x}_3 \approx 4.62$	$\widetilde{x}_4 \approx 5.53$

Figure 3.3: The result of model fusion: general setting

3.3 Model Fusion: Case of Interval Uncertainty

Main idea. The solution to the model fusion problem is to take into account three different types of approximate equalities:

- Each higher spatial resolution estimate $\widetilde{x}_i$ is approximately equal to the actual value x_i in the corresponding (smaller size) cell i, with the approximation error $x_i - \widetilde{x}_i$ bounded by the known value $\Delta_{h,i}$:

$$\widetilde{x}_i - \Delta_{h,i} \le x_i \le \widetilde{x}_i + \Delta_{h,i}.$$

- Each lower spatial resolution estimate $\widetilde{X}_j$ is approximately equal to the average of values of all the smaller cells $x_{i(1,j)}, \ldots, x_{i(k_j,j)}$ within the corresponding larger size cell, with the estimation error bounded by the known value $\Delta_{l,j}$:

$$\widetilde{X}_j - \Delta_{l,j} \le \sum_i w_{j,i} \cdot x_i \le \widetilde{X}_j + \Delta_{l,j}.$$

- Finally, there are prior bounds $\underline{x}_{pr,i}$ and $\overline{x}_{pr,i}$ on the values x_i, i.e., bounds for which

$$\underline{x}_{pr,i} \le x_i \le \overline{x}_{pr,i}.$$

The objective is to find, for each $k = 1, \ldots, n$, the range $[\underline{x}_k, \overline{x}_k]$ of possible values of x_k.

The estimates lead to a system of linear inequalities for the unknown values $x_1, \ldots, x_n$. Thus, for each k, finding the corresponding endpoints $\underline{x}_k$ and $\overline{x}_k$ means optimizing the values x_k under linear constraints. This is a particular case of a general linear programming problem; see, e.g., [3]. So, Linear Programming can be used to find these bounds:

- the lower bound $\underline{x}_k$ can be obtained by minimizing x_k under the constraints

$$\widetilde{x}_i - \Delta_h \le x_i \le \widetilde{x}_i + \Delta_h, \quad i = 1, \ldots, n;$$

$$\widetilde{X}_j - \Delta_l \le \sum_i w_{j,i} \cdot x_i \le \widetilde{X}_j + \Delta_l; \quad \underline{x}_{pr,i} \le x_i \le \overline{x}_{pr,i}.$$

- the upper bound $\overline{x}_k$ can be obtained by maximizing x_k under the same constraints.

Mathematical comment. For each i, the two constraints $\widetilde{x}_i - \Delta_h \le x_i \le \widetilde{x}_i + \Delta_h$ and $\underline{x}_{pr,i} \le x_i \le \overline{x}_{pr,i}$ can be combined into a single set of constraints:

$$x_i^- \le x_i \le x_i^+,$$

where

$$x_i^- \stackrel{\text{def}}{=} \max(\widetilde{x}_i - \Delta_h, \underline{x}_{pr,i}); \quad x_i^+ \stackrel{\text{def}}{=} \min(\widetilde{x}_i + \Delta_h, \overline{x}_{pr,i}).$$

Description of the simplest case. Consider the simplest case when only a single lower spatial resolution estimate is available $\widetilde{X}_1$. In this case, the linear constraints take the form $x_i^- \le x_i \le x_i^+$ and

$$\widetilde{X}_1 - \Delta_l \le \sum_{i=1}^{n} w_{1,i} \cdot x_i \le \widetilde{X}_1 + \Delta_l.$$

Comment. This general expression also includes the case when some cells are not covered by the estimate $\widetilde{X}_1$. This can be described by setting for $w_{1,i} = 0$ the values $w_{1,i}$ corresponding to these cells.

Derivation for the simplest case. For each variable x_k, $k = 1, \ldots, n$, the objective is to find out which values of x_k are possible.

If the k-th cell is not affected by the estimate $\widetilde{X}_1$, i.e., if $w_{1,k} = 0$, then the only restrictions on x_k come from the prior bounds on x_k and from the higher spatial resolution estimates. Thus, for such a cell, the set of possible values is the interval $[x_k^-, x_k^+]$.

Consider now the case when the k-th cell is affected by the estimate $\widetilde{X}_1$, i.e., when $w_{1,k} > 0$. In this case, a possible value x_k must be within the interval $[x_k^-, x_k^+]$, and for the remaining variables x_i, $i = 1, \ldots, k-1, k+1, \ldots, n$, the resulting system of inequalities $x_i^- \le x_i \le x_i^+$ and

$$\widetilde{X}_1 - \Delta_l - w_{1,k} \cdot x_k \le \sum_{i \ne k} w_{1,i} \cdot x_i \le \widetilde{X}_1 + \Delta_l - w_{1,k} \cdot x_k$$

must be satisfied.

All the weights $w_{1,i}$ are non-negative. Thus, when $x_i \in [x_i^-, x_i^+]$, the smallest possible value $\underline{s}$ of the sum

$$s \stackrel{\text{def}}{=} \sum_{i \ne k} w_{1,i} \cdot x_i$$

is attained when all x_i attain their smallest possible values $x_i = x_i^-$, and the largest possible value $\overline{s}$ of the sum s is attained when all x_i attain their largest possible values $x_i = x_i^+$:

$$\underline{s} = \sum_{i \ne k} w_{1,i} \cdot x_i^-; \quad \overline{s} = \sum_{i \ne k} w_{1,i} \cdot x_i^+.$$

Thus,

$$\sum_{i \ne k} w_{1,i} \cdot x_i^- \le \sum_{i \ne k} w_{1,i} \le \sum_{i \ne k} w_{1,i} \cdot x_i^+.$$

Now, the two interval $[\widetilde{X}_1 - \Delta_l - w_{1,k} \cdot x_k, \widetilde{X}_1 + \Delta_l - w_{1,k} \cdot x_k]$ and $\left[\sum\limits_{i \ne k} w_{1,i} \cdot x_i^-, \sum\limits_{i \ne k} w_{1,i} \cdot x_i^+\right]$ contain the same sum $\sum\limits_{i \ne k} w_{1,i}$. Thus, their intersection must be non-empty, i.e., the lower endpoint of the first interval cannot exceed the upper endpoint of the second interval, and vice versa (one can easily check that if these conditions are satisfied, then the above inequalities are indeed consistent):

$$\widetilde{X}_1 - \Delta_l - w_{1,k} \cdot x_k \le \sum_{i \ne k} w_{1,i} \cdot x_i^+; \quad \sum_{i \ne k} w_{1,i} \cdot x_i^- \le \widetilde{X}_1 + \Delta_l - w_{1,k} \cdot x_k.$$

By moving the term $w_{1,k} \cdot x_k$ to the other side of each of the inequalities and dividing both sides of each resulting inequality by a positive number $w_{1,k}$, it can be concluded that

$$\frac{1}{w_{1,k}} \cdot \left(\widetilde{X_1} - \Delta_l - \sum_{i \neq k} w_{1,i} \cdot x_i^+ \right) \leq x_k \leq \frac{1}{w_{1,k}} \cdot \left(\widetilde{X_1} + \Delta_l - \sum_{i \neq k} w_{1,i} \cdot x_i^- \right).$$

Resulting formulas for the simplest case. For the cells k which are not affected by the estimate $\widetilde{X_1}$, the resulting bounds on x_k are $[\underline{x}_k, \overline{x}_k]$ with $\underline{x}_k = x_k^-$ and $\overline{x}_k = x_k^+$.

For the cells k which are affected by the estimate $\widetilde{X_1}$ (i.e., for which $w_{1,k} > 0$), the resulting range $[\underline{x}_k, \overline{x}_k]$ has the form

$$\underline{x}_k = \frac{1}{w_{1,k}} \cdot \left(\widetilde{X_1} - \Delta_l - \sum_{i \neq k} w_{1,i} \cdot x_i^+ \right); \quad \overline{x}_k = \frac{1}{w_{1,k}} \cdot \left(\widetilde{X_1} + \Delta_l - \sum_{i \neq k} w_{1,i} \cdot x_i^- \right).$$

3.4 Combining Models with Different Types of Uncertainty: A Challenge

In the previous sections, it was described how to combine (fuse) models with probabilistic uncertainty and how to combine models with interval uncertainty. Sometimes, there is a need to combine models with different types of uncertainty, i.e., models with probabilistic uncertainty and models with interval uncertainty.

There may be several models with probabilistic uncertainty, and several models with interval uncertainty. In this case, it is reasonable to first fuse all the models with probabilistic uncertainty into a single fused probabilistic model, and to fuse all the models with interval uncertainty into a single fused interval model. After this procedure, the original task is reduced to the task of merging two models:

- the first is a (combined) model with probabilistic uncertainty and
- the second is a (combined) model with interval uncertainty.

In general, probabilistic models provide a more detailed description of uncertainty than the interval model. Indeed, in the case of probabilistic uncertainty, it is assumed that the mean μ (equal to 0) and the standard deviation σ of the approximation error are known. In this case, for each certainty level p_0, it can be concluded that the actual (unknown) value of the approximation error belongs to the interval $[\mu - k(p_0)\cdot\sigma, \mu + k(p_0)\cdot\sigma]$. For example, for $p_0 = 90\%$, set $k(p_0) = 2$; for $p_0 = 99.9\%$, set $k(p_0) = 3$, etc.

In case of interval uncertainty, only the interval $[-\Delta, \Delta]$ of possible values of approximation error is known. In this case, there is no knowledge of the exact values of μ and σ, other than that the actual (unknown) values of μ and σ satisfy the conditions $-\Delta \le \mu - k(p_0)\cdot\sigma$ and $\mu + k(p_0)\cdot\sigma \le \Delta$. In other words, the second (interval) uncertainty model corresponds to the whole class of possible probabilistic uncertainty models. So, a natural way to combine the probabilistic and the interval models is to consider the combinations of the first probabilistic model with all possible probabilistic models corresponding to the second (interval) model.

For example, as mentioned earlier, fusing the n values $\widetilde{x}^{(i)}$ whose measurement errors are random with mean 0 and known standard deviations $\sigma^{(i)}$, results then in an estimate $x = \dfrac{\sum\limits_{i=1}^{n} \widetilde{x}^{(i)}\cdot(\sigma^{(i)})^{-2}}{\sum\limits_{i=1}^{n}(\sigma^{(i)})^{-2}}$ whose standard deviation is equal to $\sigma^{-2} = \sum\limits_{i=1}^{n}(\sigma^{(i)})^{-2}$. If the value $x^{(n)}$ is only known with interval uncertainty, i.e., if only the bounds $\underline{x}^{(n)}$ and $\overline{x}^{(n)}$ are known for which $\underline{x}^{(n)} \le x^{(n)} \le \overline{x}^{(n)}$, then, in contrast to the probabilistic case, there is no knowledge the exact mean $\widetilde{x}^{(n)}$ and standard deviation $\sigma^{(n)}$ corresponding to the n-th measurement; instead, it is only known that, for an appropriately chosen $k_0 = k(p_0)$, the following inequalities are satisfied: $\underline{x}^{(n)} \le \widetilde{x}^{(n)} - k_0\cdot\sigma^{(n)}$ and $\widetilde{x}^{(n)} + k_0\cdot\sigma^{(n)} \le \overline{x}^{(n)}$. Thus, for the fused result, instead of a single value x, there is now a whole range of values, namely, the set of all possible values of the ratio $x = \dfrac{\sum\limits_{i=1}^{n-1} \widetilde{x}^{(i)}\cdot(\sigma^{(i)})^{-2} + \widetilde{x}^{(n)}\cdot(\sigma^{(n)})^{-2}}{\sum\limits_{i=1}^{n-1}(\sigma^{(i)})^{-2} + (\sigma^{(n)})^{-2}}$ corresponding to all possible values $\widetilde{x}^{(n)}$ and $\sigma^{(n)}$ that satisfy the above two inequalities. A similar idea can be used when there are more data points known with interval uncertainty, and in cases

where there is also a need to take into account spatial resolution.

As it can be seen from this discussion, even in the simplest case of data fusion, to combine probabilistic and interval uncertainty, there is a need to solve a complex non-linear optimization problem. Thus, combining interval and probabilistic uncertainty remains an important computational challenge.

Chapter 4

Additional Challenge: Need to Fuse Continuous and Discrete Data

An additional challenge is that in many application areas, there is a need to fuse continuous and discrete models of the same phenomena. For example, in geophysics, there are two main models for describing how the sound velocity changes with location and depth: a discrete gravity-based model, in which there are 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 these models are simply fused, the resulting 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.

Computational problem related to the need for fusing discrete and continuous models. Traditionally, seismic models are *continuous* in the sense that in these models, the velocity smoothly changes as there is a change the location and/or depth. In contrast, the gravity models are *discrete*: in these models, there are 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, by simply combining

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.

To avoid the misleading double-transition models, it is desirable, before fusing the models, to first fuse the corresponding transition locations. In this chapter, an algorithm is provided for such location fusion.

Specifically, first, a formulation of the problem in the probabilistic terms is described (see, e.g., [19]), and an algorithm is provided that solves the corresponding probabilistic problem and produces the most probable transition location. Then, it is shown that the result of the probabilistic location algorithm is in good accordance with common sense – which reaffirms that this location is reasonable. It is also shown how the commonsense intuition can be reformulated in fuzzy terms (see, e.g., [5, 8]).

4.1 Available Information

For each location, in the discrete model, there is an exact depth z_d at which there is transition between the two layers. In contrast, for the continuous model, there is no abrupt transition; instead, there are velocity values $v(z)$ at different depths. Therefore, the goal is to extract the corresponding transition value z_c from the velocity values.

To be more precise, there are values $v_1, v_2, \ldots, v_i, \ldots, v_n$ corresponding to different depths. There is a need to find i for which the transition occurs between the depths i and $i + 1$.

4.2 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., [19]. The corresponding probability density is equal to

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

It is assumed that the 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

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

Determining the location using the Maximum Likelihood Approach. In the probabilistic approach, a set of parameters is selected for which the likelihood of the observed data is the largest; see, e.g., [19]. In other words, in this Maximum Likelihood Approach, the selected values of the parameters are for the ones which the likelihood L attains the largest possible value.

Determining the optimal location i_0. Due to the fact that $\exp(a)\cdot\exp(b)=\exp(a+b)$, the expression L_i can be represented as

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

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:

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

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, find the value i for which the following expression is the smallest:

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

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:

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

This sum can be represented as

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

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.

Resulting location. As the most probable location of the transition point, 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.

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.

4.3 The Results of the Probabilistic Approach are in Good Accordance with Common Sense

Common sense description of the problem. Intuitively, for each depth i, there is a high degree of 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 confidence that this is the actual transition depth, and
- the larger the difference, the more confidence that this is the actual transition depth.

The results of the probabilistic model are in accordance with common sense. Thus, the fact that in the probabilistic model, the location selected was the one with the largest possible value $|\Delta v_i|$ shows that the probabilistic model is in good accordance with common sense – increasing the level of confidence in this result.

Formulating the common sense description in fuzzy terms. Fuzzy logic (see, e.g., [5, 8]) is known to be a useful way to formalize imprecise commonsense reasoning. To apply fuzzy logic to this situation, it is necessary to determine the degree of confidence d_i that the transition occurs at the depth i. The above commonsense idea means that this degree of confidence d_i is equal to $f(|\Delta v_i|)$, for some monotonically increasing function $f(z)$.

If there is a need to select a single location i_0, it is reasonable to select a value for which the degree of confidence is the largest

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

Since the function $f(z)$ is strictly increasing, this function attains its largest possible values when z is the largest. Thus, to find the optimal location i_0, determine the value i for which the expression $|\Delta v_i|$ is the largest possible.

Of course, to come up with this conclusion, there is no need to use fuzzy logic, this conclusion can already be deduced from the above commonsense description. However, this description may be useful in conjunction with other expert information.

4.4 How Accurate Is This Location Estimate?

Formulation of the problem. The estimate for the location i obtained in the previous section is approximate. How accurate is this location estimate?

Which approach should be used to solve this problem? It was shown known that for location, both probabilistic and fuzzy models lead to the same result. It is therefore reasonable to use only one of the two models when estimating accuracy of the location estimate.

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

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

It was 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, σ can be found from the requirement that the expression

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

attains the largest possible value over all possible i and σ. It is known that with respect to i, the largest value is attained when i is equal to the above estimate i_0; so, the problem is reduced to maximizing the expression

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

over σ.

Since the function $-\ln(z)$ is strictly increasing, this maximization 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, it is possible to 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 results in

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

Resulting probability distribution. Now, that values of all the parameters are known, 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 $\exp(a-b) = \exp(a)\cdot\exp(b)$, it is possible to 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 the following notations are used:

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

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 at one of the locations i. Since the transition is only at one of the locations, this means that the probabilities P_j of having transitions at different locations should add up to one: $\sum\limits_{j=1}^{n-1} P_j = 1$. 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 mean square deviation σ_0^2 of the actual (unknown) transition depth from the 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, leads to

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

Resulting algorithm. First, 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. This algorithm was applied to seismic model of El Paso area derived in [1] (see also [18]). In this case, the transition is between the lower and the upper crust; see, e.g., [20].

For this map, the result was $\sigma_0 \approx 1.5$ km. This result is in good accordance with the fact that the empirical difference between the border depth estimates coming from the seismic data and from the gravity data is of the same order of magnitude (1–2 km).

4.5 How to Fuse the Estimates of the Transition Depth and the Corresponding Models

Available estimates for the transition depth. Now, there are 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, the usual statistical techniques can be used to estimate the standard deviation σ_d of this estimate.

For the continuous estimate i_0, it is already known how to compute its standard deviation σ_0.

How to fuse estimates of the transition depth. A natural idea is 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, results in the following.

Based on the estimates i_d and i_0 for the transition depth, as the optimal estimate i_f for the actual transition depth, 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 the best estimate for the transition depth, i.e., for the border between the lower and upper zones which are, in this case, lower and upper crust.

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 can be simply fuse the values from both models – e.g., similarly to how the estimates for the transition depth were fused.

For intermediate depths, there is a need to adjust the models, by replacing the values corresponding to the wrong zone by the nearest value from the correct zone.

How to fuse the actual maps: resulting procedure.

- First, adjust both models so that they both have a transition at depth i_f.
- Second, for each depth i, merge the values v_i' and v_i'' corresponding to the adjusted models.

The following describes this fusion in more detail.

Adjusting the discrete model. Adjusting the discrete model is (relatively) easy: 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, there are now 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 was described earlier, the fused value $\widetilde{v}_i$ can be computed as follows:

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

Chapter 5

Auxiliary Problem: Estimating Accuracy of Different Models

Model fusion techniques described in the previous chapters assume that the accuracies σ_j of different fused models are already known. How can these accuracies be estimated? In this chapter, it is shown that the traditional methods cannot be directly used to estimate these accuracies, and a new method for such estimation is proposed.

5.1 Traditional Methods of Estimating Accuracy are Sometimes Inadequate for Estimating Accuracy of Geophysical Models

This section describes the traditional methods of estimating accuracy (see, e.g., [17]) and explains that sometimes, these methods can not be directly applied to estimating accuracy of the geophysical models.

First method: calibration. The first method of estimating accuracy is to *calibrate* the corresponding measuring instrument. Calibration is possible when there is a "standard" measuring instrument which is several times more accurate than the instrument which is being calibrated, and for which, therefore, each measurement result $x_{i,\text{st}}$ obtained by using this instrument is practically equal to the actual value x_i.

The calibration procedure consists of repeatedly measuring the same quantity by using

both the calibrated measuring instrument and the standard one. Since the result $x_{i,\text{st}}$ of using the standard instrument is practically equal to the actual value x_i, the measurement error $\Delta x_{ij} = x_{ij} - x_i$ is well approximated by the difference $\Delta x_{ij} \approx x_{ij} - x_{i,\text{st}}$ between the measurement results.

Since all the measurements x_{ij}, $i = 1 \ldots, n$, are performed by the same measuring instrument j, all these measurements have the same standard deviation σ_j. In this case, the likelihood formula takes the simplified form

$$L = \frac{1}{(\sqrt{2\pi})^n \cdot \sigma_j^n} \cdot \exp\left(-\sum_{i=1}^{n} \frac{(x_{ij} - x_i)^2}{2\sigma_j^2}\right). \tag{1}$$

A natural way to estimate value σ_j is to use the Maximum Likelihood method, i.e., to find σ_j for which the likelihood L attains the largest possible value. Maximizing L is equivalent to minimizing

$$-\ln(L) = \text{const} + n \cdot \ln(\sigma_j) + \sum_{i=1}^{n} \frac{(x_{ij} - x_i)^2}{2\sigma_j^2}.$$

Differentiating this sum w.r.t. σ_j and equating the derivative to 0, results in the usual estimate

$$\sigma_j^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_{ij} - x_i)^2. \tag{2}$$

Since the approximate values of $x_{ij} - x_i$ are known, σ_j can thus be estimated.

In geophysical applications, it is often not possible to directly use calibration. For calibration to work, there needs to be a measuring instrument which is several times more accurate than the one that is currently used. In geophysics, however, seismic (and other) methods are state-of-the-art, no method leads to more accurate determination of the densities. As a result, calibration techniques cannot be directly applied to estimating approximation errors in the geophysics problems.

Second method: using several similar instruments. In some practical situations, when there is no standard measuring instrument, an alternative is to compare the results x_{i1} and x_{i2} of measuring the same quantity x_i by two measuring instruments of the same

type. The two instruments are independent and have the same accuracy σ, so the likelihood function has the form

$$L = \frac{1}{(\sqrt{2\pi})^n \cdot \sigma^n} \cdot \exp\left(-\sum_{i=1}^{n} \frac{(x_{i1} - x_i)^2}{2\sigma^2}\right) \cdot \frac{1}{(\sqrt{2\pi})^n \cdot \sigma^n} \cdot \exp\left(-\sum_{i=1}^{n} \frac{(x_{i2} - x_i)^2}{2\sigma^2}\right).$$

In this case, neither the accuracy σ nor the actual values $x_1, \ldots, x_m$ are known; in the spirit of the Maximum Likelihood method, it is reasonable to select the values of all these parameters for which the likelihood attains the largest possible value. Maximizing L is equivalent to minimizing

$$-\ln(L) = \text{const} + 2n \cdot \ln(\sigma) + \sum_{i=1}^{n} \frac{(x_{i1} - x_i)^2}{2\sigma^2} + \sum_{i=1}^{m} \frac{(x_{i2} - x_i)^2}{2\sigma^2}. \tag{3}$$

Minimizing with respect to x_i leads to $x_i = \dfrac{x_{i1} + x_{i2}}{2}$. Substituting these values x_i into the formula (4) and minimizing the resulting expression with respect to σ, results in

$$\sigma^2 = \frac{1}{2n} \cdot \sum_{i=1}^{n} (x_{i1} - x_{i2})^2. \tag{4}$$

In geophysical applications, it is often not possible to directly use this method either. In usual measurements, when estimating the accuracy of measurements performed by a measuring instrument, it is possible to produce two similar measuring instruments and compare their results. Our objective is to estimate the accuracy of a geophysical model, e.g., a seismic model, a gravity-based model, etc. In this situation, it is not possible to come up with two similar applications of the same model, so the second method cannot be directly applied either.

Moreover, Maximum Likelihood approach cannot be applied to estimate model accuracy. Consider now the most general situation:

- there are several quantities with (unknown) actual values $x_1, \ldots, x_i, \ldots, x_n$,
- there are several measuring instruments (or geophysical methods) with (unknown) accuracies $\sigma_1, \ldots, \sigma_j, \ldots, \sigma_m$, and

- the results x_{ij} of measuring the i-th quantity by using the j-th measuring instrument are known.

At first glance, a reasonable idea is to find all the unknown quantities – i.e., the actual values x_i and the σ_j – from the Maximum Likelihood method. In this case, the likelihood takes the form

$$L = \prod_{i=1}^{n} \prod_{j=1}^{m} \frac{1}{\sqrt{2\pi} \cdot \sigma_j} \cdot \exp\left(-\frac{(x_{ij} - x_i)^2}{2\sigma_j^2}\right). \tag{5}$$

The problem with this approach is that, in contrast to the previous cases, this expression does not attain a finite maximum, it can reach values which are as large as possible. Namely, if some j_0 is chosen, then for $x_i = x_{ij_0} + \varepsilon$ and $\sigma_{j_0} = \varepsilon$, it can be concluded that $\frac{(x_{ij_0} - x_i)^2}{2\sigma_{j_0}^2} = \frac{1}{2}$, so the corresponding exponential factor is equal to $\exp\left(-\frac{1}{2}\right)$; all other factors are also finite (and positive) in the limit $\varepsilon \to 0$ except for the terms $\frac{1}{\sqrt{2\pi} \cdot \sigma_{j_0}}$ which tends to infinity.

It can be checked that if all the values σ_j are positive, then the above likelihood expression attains finite values. Thus, the largest possible – infinite – value is attained when one of the standard deviations σ_{j_0} is equal to 0. In this case, in accordance with the formula for model fusion, results in $x_i = x_{ij_0}$. In other words, for this problem, the Maximum Likelihood method leads to a counterintuitive conclusion that one of the measurements was absolutely accurate. This is not physically reasonable, so Maximum Likelihood method cannot be directly used to estimate random errors.

5.2 Proposed Idea for Estimating Model Accuracy

Analysis of the problem. It is known that $x_{ij} = x_i + \Delta x_{ij}$, where approximation errors $\Delta x_{ij} = x_{ij} - x_i$ are independent normally distributed random variables with 0 mean and (unknown) standard deviations σ_j^2. For every two estimation methods (e.g., measuring instruments) j and k, the difference $x_{ij} - x_{ik}$ between the results of estimating the same

quantity x_i by these two methods has the form

$$x_{ij} - x_{ik} = (x_i + \Delta x_{ij}) - (x_i + \Delta x_{ik}) = \Delta x_{ij} - \Delta x_{ik}.$$

Derivation of the resulting formula. The difference between two independent normally distributed random variables Δx_{ij} and Δx_{ik} is also normally distributed. The mean of the difference is equal to the difference of the means, i.e., to $0 - 0 = 0$, and the variance of the difference is equal to the sum of the variances, i.e., to $\sigma_j^2 + \sigma_k^2$.

Thus, the difference $x_{ij} - x_{ik} = \Delta x_{ij} - \Delta x_{ik}$ is normally distributed with 0 mean and variance $\sigma_j^2 + \sigma_k^2$. For each j and k, there are n values $x_{1j} - x_{1k}$, ..., $x_{nj} - x_{nk}$ from this distribution. Based on this sample, apply the usual formula (2) to estimate the standard deviation $\sigma_j^2 + \sigma_k^2$ as $\sigma_j^2 + \sigma_k^2 \approx A_{jk}$, where

$$A_{jk} \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} (x_{ij} - x_{ik})^2. \tag{6}$$

In particular, for every three different measuring instruments, with unknown accuracies σ_1^2, σ_2^2, and σ_3^2, results in the equations

$$\sigma_1^2 + \sigma_2^2 \approx A_{12}, \quad \sigma_1^2 + \sigma_3^2 \approx A_{13}, \quad \sigma_2^2 + \sigma_3^2 \approx A_{23}. \tag{7}$$

By adding all three equalities (7) and dividing the result by two, we can conclude that

$$\sigma_1^2 + \sigma_2^2 + \sigma_3^2 = \frac{A_{12} + A_{13} + A_{23}}{2}. \tag{8}$$

Resulting formulas. Subtracting, from (8), each of the equalities (7), results in $\sigma_j^2 \approx \widetilde{V}_j$, where

$$\widetilde{V}_1 = \frac{A_{12} + A_{13} - A_{23}}{2}; \quad \widetilde{V}_2 = \frac{A_{12} + A_{23} - A_{13}}{2}; \quad \widetilde{V}_3 = \frac{A_{13} + A_{23} - A_{12}}{2}. \tag{9}$$

In general, when there are M different models, this procedure leads to $\dfrac{M \cdot (M-1)}{2}$ different equations $\sigma_j^2 + \sigma_k^2 \approx A_{jk}$ to determine N unknowns σ_j^2. When $M > 3$, there are

more equations than unknowns, so the Least Squares method can be used to estimate the desired values σ_j^2.

Need for refining the above formulas. The formulas $\sigma_i^2 \approx \widetilde{V}_i$ are approximate. If an estimate $\widetilde{V}_j$ is used for σ_j^2, this may result in physically meaningless negative values for the corresponding variances.

It is therefore necessary to modify the formulas (9) so as to avoid negative values.

An idea of how to deal with the above issue. The negativity challenge is caused by the fact that the estimates in (9) are approximate. So, to come up with the desired modification, it is reasonable to first estimate the accuracy of each of the formulas (9), i.e., the standard deviation Δ_j for the difference $\Delta V_j \stackrel{\text{def}}{=} \widetilde{V}_j - \sigma_j^2$.

For large n, the difference ΔV_j between the actual value of σ_j^2 and its statistical estimate is asymptotically normally distributed, with asymptotically 0 mean; see, e.g., [19]. In the next section, an estimate the standard deviation Δ_j will be generated for this difference. Thus, it can be concluded that the actual value $\sigma_j^2 = \widetilde{V}_j - \Delta V_j$ is normally distributed with mean V_j and standard deviation Δ_j. Additionally $\sigma_j^2 \geq 0$ is known. As an estimate for σ_j^2, it is therefore reasonable to use a conditional expected value $E\left(\widetilde{V}_j - \Delta V_j \,\middle|\, \widetilde{V}_j - \Delta V_j \geq 0\right)$. This new estimate is an expected value of a non-negative number and thus, cannot be negative. In the next section, it will be shown how to compute this new estimate.

5.3 Derivation of the Corresponding Formulas

Estimating accuracies Δ_j of the estimates $\overline{V}_j$ for σ_j^2. Our first task is to estimate the accuracy Δ_j of $\widetilde{V}_j$, i.e., the expected value $\Delta_j^2 = E\left[\left(\widetilde{V}_j - \sigma_j^2\right)^2\right]$. According to (9), $\widetilde{V}_j$ is computed based on the values

$$A_{jk} = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_{ij} - x_{ik})^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (\Delta x_{ij} - \Delta x_{ik})^2.$$

To simplify notations, denote $a_i \stackrel{\text{def}}{=} \Delta x_{ij}$, $b_i \stackrel{\text{def}}{=} \Delta x_{ik}$, and $c_i \stackrel{\text{def}}{=} \Delta x_{i\ell}$; then, concludes that

$$\widetilde{V}_j = \frac{1}{2} \cdot \left[\frac{1}{n} \cdot \sum_{i=1}^{n} (a_i - b_i)^2 + \frac{1}{n} \cdot \sum_{i=1}^{n} (a_i - c_i)^2 - \frac{1}{n} \cdot \sum_{i=1}^{n} (b_i - c_i)^2 \right],$$

i.e.,

$$\widetilde{V}_j = \frac{1}{2n} \cdot \sum_{i=1}^{n} \left[(a_i - b_i)^2 + (a_i - c_i)^2 - (b_i - c_i)^2 \right]. \tag{10}$$

Opening parentheses inside the sum results in

$$(a_i - b_i)^2 + (a_i - c_i)^2 - (b_i - c_i)^2 =$$

$$a_i^2 - 2a_i \cdot b_i + b_i^2 + a_i^2 - 2a_i \cdot c_i + c_i^2 - b_i^2 + 2b_i \cdot c_i - c_i^2.$$

Thus, the formula (10) takes the form

$$\widetilde{V}_j = \frac{1}{n} \cdot \sum_{i=1}^{n} (a_i^2 - a_i \cdot b_i - a_i \cdot c_i + b_i \cdot c_i).$$

Therefore,

$$\Delta_j^2 = E\left[\left(\widetilde{V}_j - \sigma_j^2 \right) \right] = E\left[\left(\widetilde{V}_j \right)^2 - 2\widetilde{V}_j \cdot \sigma_j^2 + \sigma_j^4 \right] = E_1 - 2\sigma_1^2 \cdot E_2 + \sigma_1^4, \tag{11}$$

where

$$E_1 \stackrel{\text{def}}{=} E\left[\left(\widetilde{V}_j \right)^2 \right] = E\left[\left(\frac{1}{n} \cdot \sum_{i=1}^{n} (a_i^2 - a_i \cdot b_i - a_i \cdot c_i + b_i \cdot c_i) \right)^2 \right], \tag{12}$$

$$E_2 \stackrel{\text{def}}{=} E\left[\widetilde{V}_j \right] = E\left[\frac{1}{n} \cdot \sum_{i=1}^{n} (a_i^2 - a_i \cdot b_i - a_i \cdot c_i + b_i \cdot c_i) \right].$$

The expected value E_2 is equal to linear combination of the expected values of the expressions a_i^2, $a_i \cdot b_i$, $a_i \cdot c_i$, and $b_i \cdot c_i$:

$$E_2 = \frac{1}{n} \cdot \sum_{i=1}^{n} \left(E[a_i^2] - E[a_i \cdot b_i] - E[a_i \cdot c_i] + E[b_i \cdot c_i] \right). \tag{13}$$

All variables a_i, b_i, and c_i are independent and normally distributed with 0 mean and the corresponding variances $V_j = \sigma_j^2$. Due to independence, $E[a_i \cdot b_i] = E[a_i] \cdot E[b_i] = 0 \cdot 0 = 0$; similarly $E[a_i \cdot c_i] = E[b_i \cdot c_i] = 0$, and the only non-zero term is $E[a_i^2] = \sigma_j^2$. Thus, in the sum in E_2, only n terms $a_1^2, \ldots, a_n^2$ lead to non-zero expected value σ_j^2, hence $E_2 = \frac{1}{n} \cdot n \cdot \sigma_j^2 = \sigma_j^2$.

Now that E_2 is computed, to estimate Δ_j based on the formula (11), it is necessary to compute E_1. In general, the square of a sum can be represented as

$$\left(\sum_i z_i\right)^2 = \sum_i z_i^2 + \sum_{i \neq i'} z_i \cdot z_{i'}.$$

In the case of E_1, $z_i = a_i^2 - a_i \cdot b_i - a_i \cdot c_i + b_i \cdot c_i$. Thus, the expected value E_1 can be presented as

$$E_1 = \frac{1}{n^2} \cdot \sum_{i=1}^{n} E[z_i^2] + \frac{1}{n^2} \cdot \sum_{i \neq i'} E[z_i \cdot z_{i'}]. \tag{14}$$

Here, the expression $z_i^2 = \left(a_i^2 - a_i \cdot b_i - a_i \cdot c_i + b_i \cdot c_i\right)^2$ takes the form

$$z_i^2 = a_i^4 + a_i^2 \cdot b_i^2 + a_i^2 \cdot c_i^2 + b_i^2 \cdot c_i^2 + \text{ terms which are odd in } a_i, b_i, \text{ or } c_i.$$

Due to independence and to the fact that all normally distributed variables a_i, b_i, and c_i have 0 mean and thus, 0 odd moments, the expected values of odd terms like $a_i^3 \cdot b_i$ is zero: e.g., $E[a_i^3 \cdot b_i] = E[a_i^3] \cdot E[b_i] = 0$. Thus,

$$E[z_i^2] = E[a_i^4] + E[a_i^2 \cdot b_i^2] + E[a_i^2 \cdot c_i^2] + E[b_i^2 \cdot c_i^2].$$

For the normal distribution, $E[a_i^4] = 3\sigma_j^4$; due to independence, $E[a_i^2 \cdot b_i^2] = E[a_i^2] \cdot E[b_i^2] = \sigma_j^2 \cdot \sigma_k^2$. Thus,

$$E[z_i^2] = 3\sigma_j^4 + \sigma_j^2 \cdot \sigma_k^2 + \sigma_j^2 \cdot \sigma_\ell^2 + \sigma_k^2 \cdot \sigma_\ell^2,$$

and

$$\frac{1}{n^2} \cdot \sum_{i=1}^{n} E[z_i^2] = \frac{1}{n} \cdot (3\sigma_j^4 + \sigma_j^2 \cdot \sigma_k^2 + \sigma_j^2 \cdot \sigma_\ell^2 + \sigma_k^2 \cdot \sigma_\ell^2). \tag{15}$$

For $z_i \cdot z_{i'}$ with $i \neq i'$, similarly,

$$z_i \cdot z_{i'} = (a_i^2 - a_i \cdot b_i - a_i \cdot c_i + b_i \cdot c_i) \cdot (a_{i'}^2 - a_{i'} \cdot b_{i'} - a_{i'} \cdot c_{i'} + b_{i'} \cdot c_{i'}) =$$

$$a_i^2 \cdot a_{i'}^2 + \text{ odd terms with 0 mean.}$$

Thus, $E[z_i \cdot z_{i'}] = E[a_i^2 \cdot a_{i'}^2] = E[a_i^2] \cdot E[a_{i'}^2] = \sigma_j^2 \cdot \sigma_j^2 = \sigma_j^4$ and so, after adding over all $n^2 - n$ pairs (i, i') with $i \neq i'$, it can be concluded that

$$\frac{1}{n^2} \cdot \sum_{i \neq i'}^{n} E[z_i \cdot z_{i'}] = \frac{n^2 - n}{n^2} \cdot \sigma_j^4 = \left(1 - \frac{1}{n}\right) \cdot \sigma_j^4. \tag{16}$$

Substituting the expressions (15) and (16) into the formula (14) results in

$$E_1 = \frac{1}{n} \cdot (3\sigma_j^4 + \sigma_j^2 \cdot \sigma_k^2 + \sigma_j^2 \cdot \sigma_\ell^2 + \sigma_k^2 \cdot \sigma_\ell^2) + \left(1 - \frac{1}{n}\right) \cdot \sigma_j^4.$$

Substituting this expression for E_1 and the formula $E_2 = \sigma_j^2$ into the formula (11) results in

$$\Delta_j^2 = \frac{1}{n} \cdot (3\sigma_j^4 + \sigma_j^2 \cdot \sigma_k^2 + \sigma_j^2 \cdot \sigma_\ell^2 + \sigma_k^2 \cdot \sigma_\ell^2) + \left(1 - \frac{1}{n}\right) \cdot \sigma_j^4 - 2\sigma_j^4 + \sigma_j^4,$$

i.e.,

$$\Delta_j^2 = \frac{1}{n} \cdot (2\sigma_j^4 + \sigma_j^2 \cdot \sigma_k^2 + \sigma_j^2 \cdot \sigma_\ell^2 + \sigma_k^2 \cdot \sigma_\ell^2). \tag{2}$$

The exact values σ_j^2 are not known, but the estimates $\widetilde{V}_j$ for these values are known; thus, Δ_j can be estimated as follows:

$$\Delta_j^2 \approx \frac{1}{n} \cdot \left(\left(\widetilde{V}_j\right)^2 + \widetilde{V}_j \cdot \widetilde{V}_k + \widetilde{V}_j \cdot \widetilde{V}_\ell + \widetilde{V}_k \cdot \widetilde{V}_\ell\right). \tag{21}$$

From estimating Δ_j to a non-negative estimate for σ_j^2. In the previous text, an estimate $\widetilde{V}_j$ for σ_j^2 has been described (as defined by the formula (9)). It is known that the difference $\Delta V_j = \widetilde{V}_j - \sigma_j^2$ is normally distributed with 0 mean and the known standard deviation Δ_j. Since, as was mentioned in the previous section, the original estimate $\widetilde{V}_j$ may be negative, it is desirable to use a new estimate $E\left(\widetilde{V}_j - \Delta V_j \,\middle|\, \widetilde{V}_j - \Delta V_j \ge 0\right)$.

The Gaussian variable ΔV_j has 0 mean and standard deviation Δ_j; thus, it can be represented as $t \cdot \Delta_j$, where t is a Gaussian random variable with 0 and standard deviation 1. In terms of the new variable t, the non-negativity condition $\widetilde{V}_j - \Delta V_j \ge 0$ takes the form $\widetilde{V}_j - \Delta_j \cdot t \ge 0$, i.e., $t \le \delta_j \stackrel{\text{def}}{=} \dfrac{\widetilde{V}_j}{\Delta_j}$. Thus, the desired conditional mean is equal to

$$E\left(\widetilde{V}_j - \Delta_j \cdot t \,\middle|\, t \le \delta_j\right) = E\left(\widetilde{V}_j \,\middle|\, t \le \delta_j\right) - \Delta_j \cdot E\left(t \,|\, t \le \delta_j\right) = \widetilde{V}_j - \Delta_j \cdot E\left(t \,|\, t \le \delta_j\right). \tag{22}$$

So, to compute the desired estimate, it is sufficient to be able to compute the value $E\left(t \,|\, t \le \delta_j\right)$ for the standard Gaussian variable t, with the probability density function $\rho(t) = \dfrac{1}{\sqrt{2\pi}} \cdot \exp\left(-\dfrac{t^2}{2}\right)$. By definition, this conditional mean is equal to the ratio

$E\,(t\mid t\le\delta_j\,)=\dfrac{N_j}{D_j}$, where

$$N_j=\int_{-\infty}^{\delta_j} t\cdot\rho(t)\,dt;\quad D_j=\int_{-\infty}^{\delta_j}\rho(t)\,dt. \tag{23}$$

The denominator D_j is equal to $\Phi(\delta_j)\stackrel{\text{def}}{=}\text{Prob}(t\le\delta_j)$. The numerator N_j of this formula is equal to

$$N_j=\int_{-\infty}^{\delta_j} t\cdot\frac{1}{\sqrt{2\pi}}\cdot\exp\left(-\frac{t^2}{2}\right)\,dt. \tag{24}$$

By introducing a new variable $s=\dfrac{t^2}{2}$ for which $ds=t\cdot dt$, the formula (24) can be reduced to

$$N_j=\frac{1}{\sqrt{2\pi}}\cdot\int_{\infty}^{\delta_j^2/2}\exp(-s)\,ds.$$

This integral can be explicitly computed, resulting in

$$N_j=-\frac{1}{\sqrt{2\pi}}\cdot\exp\left(-\frac{\delta_j^2}{2}\right)$$

and thus,

$$E\,(t\mid t\le\delta_j\,)=-\frac{1}{\sqrt{2\pi}}\cdot\frac{\exp\left(-\dfrac{\delta_j^2}{2}\right)}{\Phi(\delta_j)}.$$

So,

$$E\left(\widetilde{V}_j-\Delta_j\cdot t\,\middle|\,t\le\delta_j\right)=\widetilde{V}_j-\Delta_j\cdot E\,(t\mid t\le\delta_j\,)=\widetilde{V}_j+\frac{\Delta_j}{\sqrt{2\pi}}\cdot\frac{\exp\left(-\dfrac{\delta_j^2}{2}\right)}{\Phi(\delta_j)}.$$

5.4 Resulting Algorithm

It is assumed that for each value x_i $(i=1,\ldots,n)$, there are three estimates x_{i1}, x_{i2}, and x_{i3} corresponding to three different models. The objective is to estimate the accuracies σ_j^2 of these three models.

First, for each $j\ne k$, compute $A_{jk}=\dfrac{1}{n}\cdot\sum_{i=1}^{n}(x_{ij}-x_{ik})^2$. Then, compute

$$\widetilde{V}_1=\frac{A_{12}+A_{13}-A_{23}}{2};\quad \widetilde{V}_2=\frac{A_{12}+A_{23}-A_{13}}{2};\quad \widetilde{V}_3=\frac{A_{13}+A_{23}-A_{12}}{2}.$$

After that, for each j, compute

$$\Delta_j^2 = \frac{1}{n} \cdot \left(\left(\widetilde{V}_j \right)^2 + \widetilde{V}_j \cdot \widetilde{V}_k + \widetilde{V}_j \cdot \widetilde{V}_\ell + \widetilde{V}_k \cdot \widetilde{V}_\ell \right).$$

Once the preliminary estimates $\widetilde{V}_j$ and their accuracies Δ_j have been computed, the auxiliary ratios $\delta_j = \dfrac{\widetilde{V}_j}{\Delta_j}$ can be computed. Finally, the following estimate $\widetilde{\sigma_j^2}$ for σ_j^2 is computed and returned:

$$\widetilde{\sigma_j^2} = \widetilde{V}_j + \frac{\Delta_j}{\sqrt{2\pi}} \cdot \frac{\exp\left(-\dfrac{\delta_j^2}{2} \right)}{\Phi(\delta_j)}.$$

Chapter 6

Concluding Remarks

In many practical situations, there is a need to combine (fuse) data from different datasets. An ideal approach would be to perform *joint inversion*, i.e., to use all the data from all the datasets to find the best model of the corresponding phenomenon. In many practical situations, however, joint inversion techniques are still being developed, and in situations when the corresponding algorithms are already available, these preliminary algorithms often require an impractical amount of computational resources (such as computation time and/or number of processors).

The Model Fusion approach described in this thesis is a new technique that can serve as a fast practical alternative to joint inversion. The main idea of this approach is to first process each dataset separately (by using known efficient algorithms), and then by fusing resulting models. Specifically, in this thesis, algorithms are proposed that can fuse models with different accuracy, different coverage, different spatial resolution, and different types of uncertainty (probabilistic or interval). In addition, this thesis addresses the additional challenge of fusing discrete models with continuous models and the auxiliary challenge of estimating the accuracy of different models.

The work described in this thesis can help geophysicists better understand the interior structure of the earth by combining the complementary results of all their current techniques.

References

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

[2] Averill, M.G., Miller, K.C., Keller, G.R., Kreinovich, V., Araiza, R. and Starks, S.A. (2007) 'Using expert knowledge in solving the seismic inverse problem', *International Journal of Approximate Reasoning*, Vol. 45, No. 3, pp.564–587.

[3] Cormen, C.E., Leiserson, R., Rivest, R.L. and Stein, C. (2009) *Introduction to Algorithms*, MIT Press, Boston, Massachusetts.

[4] Hole, J.A. (1992) 'Nonlinear high-resolution three-dimensional seismic travel time tomography', *Journal of Geophysical Research*, Vol. 97, pp.6553–6562.

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

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

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

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

[9] Novitskii, P.V., and Zograph, I.A. (1991) *Estimating the Measurement Errors*, Energoatomizdat, Leningrad (in Russian).

[10] Ochoa, O. (2009) 'Towards a fast practical alternative to joint inversion of multiple datasets: model fusion', *Abstracts of the 2009 Annual Conference of the Computing Alliance of Hispanic-Serving Institutions CAHSI*, Mountain View, California, January 15–18, 2009.

[11] Ochoa, O., Velasco, A.A., Kreinovich, V. and Servin, C. (2008) '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.

[12] Ochoa, O., Velasco, A.A., and Kreinovich, V., 'Fusing Continuous and Discrete Data, on the Example of Merging Seismic and Gravity Models in Geophysics', *Proceedings of the 30th Annual Conference of the North American Fuzzy Information Processing Society NAFIPS'2011*, El Paso, Texas, March 18-20, 2011, pp.1–5.

[13] Ochoa, O., Velasco, A.A., and and Servin, C., 'Towards Model Fusion in Geophysics: How to Estimate Accuracy of Different Models', *Journal of Uncertain Systems*, 2013, Vol.7, No.3, pp.190–197.

[14] Ochoa, O., Velasco, A.A., Servin, C., and Kreinovich, V., 'Model Fusion under Probabilistic and Interval Uncertainty, with Application to Earth Sciences', *International Journal of Reliability and Safety*, 2012, Vol. 6, No. 1–3, pp. 167–187.

[15] Orlov, A.I. (1991) 'How often are the observations normal?', *Industrial Laboratory*, 1991, Vol. 57. No. 7, pp.770–772.

[16] Pinheiro da Silva, P., Velasco, A., Ceberio, M., Servin, C., Averill, M., Del Rio, N., Longpré, L. and Kreinovich, V. (2008) 'Propagation and provenance of probabilistic and interval uncertainty in cyberinfrastructure-related data processing and data fusion', In: Muhanna, R.L. and Mullen, R.L. (eds.), *Proceedings of the International Workshop on Reliable Engineering Computing REC'08*, Savannah, Georgia, February 20–22, 2008, pp.199–234.

[17] Rabinovich, S. (2005) *Measurement Errors and Uncertainties: Theory and Practice*, Springer Verlag, New York.

[18] Servin, C., Ochoa, O. and Velasco, A.A. (2008) 'Probabilistic and interval uncertainty of the results of data fusion, with application to geosciences, *Abstracts of 13th International Symposium on Scientific Computing, Computer Arithmetic, and Verified Numerical Computations SCAN'2008*, El Paso, Texas, September 29 – October 3, 2008, p.128.

[19] Sheskin, D.G. (2011) *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall/CRC Press, Boca Raton, Florida.

[20] Tarbuck, E.J., Lutgens, F.K., and Tasa, D. (2010) *Earth: An Introduction to Physical Geology*, Prentice Hall, Upper Saddle River, New Jersey.

Curriculum Vitae

Omar Ochoa holds a Bachelor of Science in Computer Science and is a graduate student in the Department of Computer Science at the University of Texas at El Paso. He is a Research Associate at the Cyber-ShARE Center of Excellence. His areas of research are software engineering, formal specifications, and model fusion with application to geophysical tomography. He is a recipient of the Murchison Graduate Engineering Scholarship, an Alliance for Graduate Education and Professoriate Fellowship and a past recipient of an IBM Fellowship. He is an IEEE Certified Software Development Professional and a member of IEEE, ACM and Upsilon Pi Epsilon

Permanent address: 3337 Monroe
El Paso, Texas 79930