Local-Global Support for Earth Sciences: Economic Analysis

Uyen Hoang Pham, Aaron Velasco, and Vladik Kreinovich

Abstract Most funding for science comes from taxpayers. So, it is very important to be able to convince taxpayers that this funding is potentially beneficial for them. This task is easier in Earth sciences, e.g., in meteorology, where there are clear local benefits. The problem is that while many people support local studies focused on their region, they do not always have a good understanding of the fact that effective local benefits require also studying surrounding areas – and what should be the optimal balance between local and (more) global studies. In this paper, on a (somewhat) simplified model of the situation, we explain what is the appropriate balance. We hope that the corresponding methodology can (and will) be applied to more realistic – and thus, more complex – local-global models as well.

1 Formulation of the Problem

Financial support is important for science. While there are few disciplines like theoretical mathematics or theoretical physics that mainly need brain power, most other sciences need funding. Even theoretical physics indirectly needs funding: it is great to come up with new theories and ideas, but their experimental testing requires funding – and sometimes very serious funding.

Who provides this financial support. Some scientific research is done by companies, but the vast majority of scientific efforts are eventually supported by taxpayers:

Uyen Hoang Pham
University of Economics and Law, Ho Chin Minh City, Vietnam, e-mail: uyenph@uel.edu.vn

Aaron Velasco
Department of Earth, Evironmental, and Resource Sciences,University of Texas at El Paso
500 W. University, El Paso, Texas 79968, USA, e-mail: aavelasco@utep.edu

Vladik Kreinovich
Department of Computer Science,University of Texas at El Paso
500 W. University, El Paso, Texas 79968, USA, e-mail: vladik@utep.edu

- taxpayers directly support salaries of faculty at public universities,
- taxpayers indirectly support faculty at private universities – by supporting funding agencies that distribute research grants.

How to maintain this support. To maintain financial support of science, we need to continue convincing taxpayers that this support is beneficial to the society.

Local-global aspects of seeking financial support. In some areas – e.g., in fundamental physics – research promises to benefit humanity as a whole, irrespective of location. For example, when researchers come up with new ways to design computer chips that could lead to faster and/or less energy-consuming computers, everyone will benefit.

The situation is different with Earth science, be it geological sciences, environmental science, biological research, atmospheric and meteorological studies, etc. In these studies, many research efforts lead to clear local benefits; for example:

- geophysical studies of a given region can help better predict the probability of earthquakes of different strength, and thus, help protect the building against possible shocks without overstraining the budget;
- geophysical studies can help find underground water reservoirs in desert areas, they can find mineral deposts – which can boost local economies;
- meteorological studies help make short- and long-term weather predictions, which helps agriculture and helps better prepare for possible extreme events, etc.

In Earth sciences, there is a local support, but this support is often too localized. Because of the importance of the corresponding research efforts, people often support local research efforts. For example, in El Paso, Texas, where two of us live, recently, a doctoral student – now Doctor – Solymar Ayala performed a study of seismicity around the city. In many public and private places, including local schools, churches, and synagogues, he was very welcome to dig in corresponding equipment and perform measurements, even when this interfered with the usual activities. Local media usually emphasizes the importance of such activity, and thus, helps get community support for this research.

In short, population usually appreciates and supports local research efforts that have a potential local benefit. But there is not that much understanding of the fact that to make local predictions, we cannot restrict ourselves only the local area, we also need to perform studies in nearby – and sometimes faraway – areas. For example, to predict tomorrow's weather in El Paso, it is not sufficient to know today's temperature, wind, and humidity in the city itself, we need to also use measurements in the nearby locations that tell us whether, e.g., a cold front is coming. An extreme example is that many of El Paso geoscientists interested in El Paso seismicity perform measurements in Kenya – because many geological structures there are similar to Rio Grande Rift which goes through El Paso, and studies there have contributed (and continue to contribute) to a much better understanding of seismicity around El Paso.

We are working on solving this problem. There is a clear need to help local communities get a better understanding that local benefits come not only from strictly local research, but also from research in nearby areas. This community communications are one of the main objectives of the recently funded US-based Center for Collective Impact in Earthquake Science (C-CIES), a project sponsored by the US National Science Foundation, in which one of us (AV) is the Principle Investigator; see, e.g., [5].

How should we balance local vs. global support: economic analysis is needed. It is not enough just to convince people that support is needed. Resources are bounded, so we need to better understand what should be the balance between local and no-so-local research efforts that will maximize the research's impact on the local community.

What we do in this paper. In this paper, on a simplified model of local-global effects, we show how to find the optimal balance of efforts. We hope that the corresponding methodology can (and will) be applied to more realistic – and thus, more complex – local-global models as well.

2 Model of Local-Global Effects and the Resulting Optimization Problem

Main task. We want to predict – or determine – the value y of some quantity at our location x_0. This may be tomorrow's temperature in our city, this may be the amount of water in an underground natural aquifer, etc. To estimate this value, we use our estimates of the values $f(x)$ of related quantities f at different geographical locations x around x_0. For this estimation, we have an algorithm A that transforms estimates for the values $f(x)$ into an estimate for y: $y = A(f(x_1), f(x_2), \ldots)$.

Need to take uncertainty into account. Our information about the values $f(x)$ usually comes from measurements: either directly from measurements, or from processing results of measuring related quantities. Measurements are never absolutely accurate; see, e.g., [2]. Thus, the resulting estimates $\widetilde{f}(x)$ are, in general, different from the actual (unknown) values $f(x)$. Because of this, the estimate $\widetilde{y} = A(\widetilde{f}(x_1), \widetilde{f}(x_2), \ldots)$ that we obtain by using the estimates is, in general, different from the (almost) exact value $y = A(f(x_1), f(x_2), \ldots)$ that we would have gotten if we knew the exact values $f(x_i)$.

The purpose of additional measurements is to get more accurate estimates for the values $f(x_i)$ and thus, get a more accurate estimate for y.

Possibility of linearization. The estimates $\widetilde{f}(x)$ are usually reasonable, in the sense that the difference $\Delta f(x) \stackrel{\text{def}}{=} \widetilde{f}(x) - f(x)$ between each estimate $\widetilde{f}(x)$ and the corresponding actual value $f(x)$ is significantly smaller than the value $f(x)$ itself: if this difference was not smaller, it would have been a wild guess, not an estimate.

Thus, we have $f(x_i) = \widetilde{f}(x_i) - \Delta f(x_i)$, where $|\Delta f(x_i)| \ll |f(x_i)|$. So, the resulting estimation error $\Delta y \stackrel{\text{def}}{=} \widetilde{y} - y$ can be described as

$$\Delta y = A(\widetilde{f}(x_1), \widetilde{f}(x_2), \ldots) - A(f(x_1), f(x_2), \ldots) =$$

$$A(\widetilde{f}(x_1), \widetilde{f}(x_2), \ldots) - A(\widetilde{f}(x_1) - \Delta f(x_1), \widetilde{f}(x_2) - \Delta f(x_2), \ldots). \quad (1)$$

Since we have $|\Delta f(x_i)| \ll |f(x_i)|$, we can do what physicists usually do in such situations (see, e.g., [1, 4]): as the first approximation, expand the dependence on $\Delta f(x_i)$ in Taylor series and keep only linear terms in this expansion. In this approximation, Δy becomes a linear function of the values $\Delta f(x_i)$:

$$\Delta y = \sum_i c_i \cdot \Delta f(x_i), \quad (2)$$

for some coefficients c_i.

In general, the closer the point x_i to the location x_0, the more the corresponding term $f(x_i)$ affects the value y, i.e., the larger the value $|c_i|$.

It is reasonable to assume that Δy is normally distributed with 0 mean. Measurements at different locations are independent, so the corresponding uncertainties $\Delta f(x_i)$ are independent random variables. According to the formula (2), the value Δy is a linear combination of a large number of similar-size independent variables. According to the Central Limit Theorem (see, e.g., [3]), the distribution of such a linear combination is close to Gaussian (normal).

A normal distribution is characterized by two parameters: mean μ and variance V (instead of the variance V, we can consider its square root $\sigma \stackrel{\text{def}}{=} \sqrt{V}$ known as *standard deviation*).

The mean value is usually 0. Measuring instruments are usually calibrated, so that if there is a *bias* – i.e., if the testing shows that the mean value of the measurement error is different from 0 – we subtract this mean value from all the measurement results and thus, conclude that the mean value if each variable $\Delta f(x_i)$ is 0.

In general, the mean value of a linear combination is equal to the linear combination of the corresponding mean values. Since the mean value of each term $\Delta f(x_i)$ is 0, we conclude that the men value μ of their linear combination (2) is also 0.

Thus, the only characteristic of the uncertainty Δy is its variable V.

The formula for the variable V. Since the variables $\Delta f(x_i)$ are independent, we have

$$V = \sum_i c_i^2 \cdot V_i, \quad (3)$$

where V_i denotes the variance of the measurement error $\Delta f(x_i)$.

Let V_0 be the variance of a single measurement. According to statistics (see, e.g., [3]), if we repeat a measurement k times, the variance of the resulting estimate becomes k times smaller. Let N denote the overall number of measurements that we can afford to make. Let n_i be the number of measurements performed in the i-th areas, so that

$$n_1 + n_2 + \ldots = N \quad (4)$$

Then, we have $V_i = V_0/n_i$ and thus, the formula (3) takes the form

$$V = \sum_i c_i^2 \cdot \frac{V_0}{n_i}. \tag{5}$$

Resulting optimization problem. We want to estimate y with the smallest possible uncertainty. As we have mentioned, the uncertainty of y can be gauged by its variance V. Thus, our objective is, among the values n_i that satisfy the equality (4), to select the values that minimize the variance (5).

3 Solution to the Optimization Problem

Let us use Lagrange multiplier method. The usual way to solve constraint optimization problems is to use the *Lagrange multiplier method*, according to which optimizing a function $F(z)$ under constraint $G(z) = 0$ can be reduced to an unconstrained optimization of an auxiliary function $F(z) + \lambda \cdot G(z)$ for some parameter λ known as *Lagrange multiplier*. Once we find the solutions corresponding to different λ, we select the value λ for which the resulting solution satisfies the constraint $G(z) = 0$.

In our case, this method means that we maximize the following auxiliary function:

$$\sum_i c_i^2 \cdot \frac{V_0}{n_i} + \lambda \cdot \left(\sum_i n_i - N \right). \tag{6}$$

Differentiating this expression with respect to n_i and equating the derivative to 0, we conclude that

$$-c_i^2 \cdot \frac{V_0}{n_i^2} + \lambda = 0,$$

hence

$$n_i^2 = c_i^2 \cdot \frac{V_0}{\lambda}$$

and thus,

$$n_i = c \cdot |c_i|, \tag{7}$$

where we denoted

$$c_0 \stackrel{\text{def}}{=} \sqrt{\frac{V_0}{\lambda}}.$$

The value c can be determined from the condition (4) that takes the form

$$c \cdot \sum_j |c_j| = N,$$

hence

$$c = \frac{N}{\sum_j |c_j|}.$$

For this value c, the optimal solution takes the following form.

Optimal measurement strategy. To get the smallest possible estimation error, we need to perform

$$n_i = \frac{N \cdot |c_i|}{\sum_j |c_j|} \quad (8)$$

measurements at each location i, where:

- the integer N is the overall number of measurements that we can afford, and
- the value c_i describes the degree to which y depends on the estimate $f(x_i)$.

Discussion. In line with common sense, we need to perform not only local measurements, but also measurement at nearby points. The further away we are from our location x_0, the smaller the value $|c_i|$ and thus, the smaller amount of effort we should allocate to measuring the corresponding value $f(x_i)$.

Comment. This qualitative conclusion we could, of course, make without doing any calculations. We need calculations to determine how exactly to distribute the measurements.

What is the resulting accuracy of estimating y: derivation. Substituting the values (8) into the formula (5), we conclude that

$$V = \sum_i c_i^2 \cdot \frac{V_0}{n_i} = V_0 \cdot \sum_i \frac{c_i^2}{n_i} = \frac{V_0}{N} \cdot \left(\sum_j |c_j|\right) \cdot \sum_i \frac{c_i^2}{|c_i|}. \quad (9)$$

Here, $c_i^2 = |c_i|^2$, thus

$$\sum_i \frac{c_i^2}{|c_i|} = \sum_i \frac{|c_i|^2}{|c_i|} = \sum_i |c_i|,$$

and we get the following formula.

What is the resulting accuracy V of estimating y: the resulting formula.

$$V = \frac{V_0}{N} \cdot \left(\sum_i |c_i|\right)^2. \quad (10)$$

Acknowledgments

This work was supported in part by the National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in

Computer Science), HRD-1834620 and HRD-2034030 (CAHSI Includes), EAR-2225395 (Center for Collective Impact in Earthquake Science C-CIES), and by the AT&T Fellowship in Information Technology.

It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and by a grant from the Hungarian National Research, Development and Innovation Office (NRDI) and by the Center of Excellence in Econometrics, Faculty of Economics, Chiang Mai University, Thailand.

The authors are thankful to Hung T. Nguyen for his encouragement.

References

1. R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Addison Wesley, Boston, Massachusetts, 2005.
2. S. G. Rabinovich, *Measurement Errors and Uncertainty: Theory and Practice*, Springer Verlag, New York, 2005.
3. D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.
4. K. S. Thorne and R. D. Blandford, *Modern Classical Physics: Optics, Fluids, Plasmas, Elasticity, Relativity, and Statistical Physics*, Princeton University Press, Princeton, New Jersey, 2021.
5. A. A. Velasco, J. Weidner, M. Karplus, S. Bilek, M. Brudzinski, D. Chandrasekhar, J. Ebel, T. Hobbs, J. M. Hurtado, S. Jaume, E. Jones, A. Kafka, V. Kreinovich, Y. Lin, A.-M. Nuñez, K. Pankow, Z. Peng, A. Savvaidis, I. Tien, C. Bolton Valencius, E. Vanacore, and M. Alvillar, "Center for Collective Impact in Earthquake Science (C-CIES): building inclusive excellence, diversity, equity, and community into earthquake science", *Abstracts of the 48th Annual Natural Hazards Research and Applications Workshop*, Boulder, Colorado, July 9–13, 2023.