Local-Global Support for Earth Sciences: Economic Analysis

Uyen Hoang Pham[1], Aaron Velasco[2], and Vladik Kreinovich[3]

[1]University of Economics and Law, Ho Chin Minh City, Vietnam
uyenph@uel.edu.vn

[2]Department of Earth, Environmental, and Resource Sciences
University of Texas at El Paso, 500 W. University
El Paso, Texas 79968, USA, aavelasco@utep.edu

[3]Department of Computer Science, University of Texas at El Paso
500 W. University, El Paso, Texas 79968, USA, vladik@utep.edu

1. Financial support is important for science

- 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.

2. Who provides this financial support

- Some scientific research is done by companies.
- However, the vast majority of scientific efforts are eventually supported by taxpayers.
- Taxpayers directly support salaries of faculty at public universities.
- Taxpayers indirectly support faculty at private universities – by supporting funding agencies that distribute research grants.
- To maintain financial support of science, we need to continue convincing taxpayers that this support is beneficial to the society.

3. 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: geological, environmental, biological, atmospheric and meteorological, etc.
- In these studies, many research efforts lead to clear local benefits.
- Geophysical studies of a given region can help better predict the probability of earthquakes of different strength.
- Thus, they help protect the building against possible shocks without overstraining the budget.

4. Local-global aspects of seeking financial support (cont-d)

- 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.
- This helps agriculture and helps better prepare for possible extreme events, etc.

5. 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.

6. In Earth sciences, there is a local support, but this support is often too localized (cont-d)

- 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.

7. In Earth sciences, there is a local support, but this support is often too localized (cont-d)

- An extreme example is that many of El Paso geoscientists interested in El Paso seismicity perform measurements in Kenya.
- Reason: many geological structures there are similar to Rio Grande Rift which goes through El Paso.
- Studies there have contributed (and continue to contribute) to a much better understanding of seismicity around El Paso.

8. 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 new Center for Collective Impact in Earthquake Science (C-CIES).
- This project sponsored by the US National Science Foundation, in which one of us (AV) is the Principle Investigator.

9. 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.
- In this paper, on a simplified model of local-global effects, we show how to find the optimal balance of efforts.
- We hope that our methodology can (and will) be applied to more realistic – and thus, more complex – local-global models as well.

10. 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: $y = A(f(x_1), f(x_2), \ldots)$.

11. 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.
- 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 $f(x_i)$.

12. Need to take uncertainty into account (cont-d)

- If the resulting accuracy of y is not sufficient, we need to perform additional measurements:
  - to get more accurate estimates for the values $f(x_i)$ and
  - thus, get a more accurate estimate for y.

13. 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.
- Indeed, 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

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

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

14. Possibility of linearization (cont-d)

- Since we have $|\Delta f(x_i)| \ll |f(x_i)|$, we can do what physicists usually do in such situations:
  - 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 $\Delta f(x_i)$:

$$\Delta y = \sum_i c_i \cdot \Delta f(x_i), \text{ for some } c_i.$$

- 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 $|c_i|$.

15. 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 above formula, the value Δy is a linear combination of a large number of similar-size independent variables.
- According to the Central Limit Theorem, 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*).

16. 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 of 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 is also 0.
- Thus, the only characteristic of the uncertainty Δy is its variable V.

17. The formula for the variable V

- Since the variables $\Delta f(x_i)$ are independent, we have $V = \sum\limits_i c_i^2 \cdot V_i$, 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, 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 $n_1 + n_2 + \ldots = N$.
- Then, we have $V_i = V_0/n_i$ and thus, the above formula takes the form

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

18. 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 for which $\sum n_i = N$, to select the values that minimize the variance V.

19. Let us use Lagrange multiplier method

- The usual way to solve constraint optimization problems is to use the *Lagrange multiplier method.*
- According to this method:
  - 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.
- This parameter is called *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:

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

20. Let us use Lagrange multiplier method (cont-d)

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

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

- Thus, $n_i = c \cdot |c_i|$ where we denoted $c \stackrel{\text{def}}{=} \sqrt{\dfrac{V_0}{\lambda}}$.
- The value c can be determined from the condition $\sum n_i = N$ that takes the form $c \cdot \sum\limits_j |c_j| = N$, hence $c = \dfrac{N}{\sum\limits_j |c_j|}$.
- For this value c, the optimal solution takes the following form.

21. Optimal measurement strategy

- Our goal is get the smallest possible estimation error.
- For this purpose, at each location i, we need to perform $n_i = \dfrac{N \cdot |c_i|}{\sum\limits_j |c_j|}$ measurements.
- Here:
  - 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 $f(x_i)$.

22. 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 $|c_i|$.
- 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.

23. What is the resulting accuracy of estimating y

- Substituting the optimal values n_i into the formula for V, we conclude that:

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

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

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

- So, we get the following formula:

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

24. Acknowledgments

This work was supported in part by:

- National Science Foundation grants 1623190, HRD-1834620, HRD-2034030, and EAR-2225395;
- AT&T Fellowship in Information Technology;
- program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and
- a grant from the Hungarian National Research, Development and Innovation Office (NRDI).