Scale-Invariance and Fuzzy Techniques Explain the Empirical Success of Inverse Distance Weighting and of Dual Inverse Distance Weighting in Geosciences

Laxman Bokati[1], Aaron Velasco[2], and Vladik Kreinovich[1,3]

[1]Computational Science Program
[2]Department of Geological Sciences
[3]Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, Texas 79968, USA
lbokati@miners.utep.edu, aavelasco@utep.edu
vladik@utep.edu

1. Need for Interpolation of Spatial Data

- Often, we are interested in the value of a certain physical quantity at different spatial locations.
- In geosciences, we may be interested in how depths of diff. geological layers depend of the spatial location.
- In environmental sciences, we may be interested in the concentration of substances in the atmosphere, etc.
- In principle, at each location, we can measure – directly or indirectly – the value of the corresponding quantity.
- However, we can only perform the measurement at a finite number of locations.
- But we are interested in the values of the quantity at all possible locations.

2. Need for Interpolation (cont-d)

- So, we need to estimate these values based on the measurement results – *interpolate* and *extrapolate*.
- In precise terms:
  - We know the values $q_i = q(x_i)$ of the quantity of interest q at several locations x_i, $i = 1, 2, \ldots, n$.
  - We would like to estimate the value $q(x)$ of this quantity at a given location x.

3. Inverse Distance Weighting

- A reasonable estimate q for $q(x)$ is a weighted average of the known values $q(x_i)$: $q = \sum\limits_{i=1}^{n} w_i \cdot q_i$, with $\sum\limits_{i=1}^{n} w_i = 1$.
- Naturally, the closer is the point x to the point x_i, the larger should be the weight w_i.
- So, the weight w_i with which we take the value q_i should decrease with the distance.
- Empirically, the best interpolation is attained when $w_i \sim (d(x, x_i))^{-p}$ for some $p > 0$.
- Since the weights have to add up to 1, we thus get

$$w_i = \frac{(d(x, x_i))^{-p}}{\sum\limits_{j=1}^{n} (d(x, x_j))^{-p}}.$$

- This method is known as *inverse distance weighting*.

4. First Challenge: Why Inverse Distance Weighting?

- In general, the fact that some algorithm is empirically the best means that:
  - we tried many other algorithms, and
  - this particular algorithm worked better than everything else we tried.
- In practice, we cannot try all possible algorithms, we can only try finitely many different algorithms.
- So, in principle, there could be an algorithm:
  - that we did not try and
  - that will work better than the one which is currently empirically the best.

5. First Challenge (cont-d)

- Because of this:
  - every time we have some empirically best alternative,
  - it is desirable to come up with a theoretical explanation of why this alternative is indeed the best.
- And if such an explanation cannot be found, maybe it this alternative is actually not the best? Thus:
  - the empirical success of inverse distance weighting prompts a natural question:
  - is this indeed the best method?
- This is the first challenge that we will deal with in this talk.

6. Limitations of Inverse Distance Weighting

- This method works well when we have a reasonably uniformly distributed spatial data.
- The problem is that in many practical cases, we have more measurements in some areas and fewer in others.
- For example, when we measure meteorological quantities such as temperature, humidity, wind speed:
  - we usually have plenty of sensors (and thus, plenty of measurement results) in cities,
  - but much fewer measurements in not so densely populated areas – e.g., in the deserts.
- Let us provide a simple example explaining why this may lead to a problem.
- Suppose that we have two locations A and B at which we perform measurements.

7. Limitations (cont-d)

- Location A is densely populated, so we have two measurement results q_A and $q_{A'}$ from this area.
- Location B is a desert, so we have only one measurement result q_B from this location.
- Since locations A and A' are very close, the corresponding values are also very close.
- So we can safely assume that they are equal: $q_A = q_{A'}$.
- Suppose that we want to predict the value of the quantity x at a midpoint C between A and B.
- Intuitively, we should combine the values q_A and q_B with equal weights, i.e., take $q_C = \dfrac{q_A + q_B}{2}$.

8. Limitations (cont-d)

- From the commonsense viewpoint, it should not matter:
  - whether we made a single measurement at the location A
  - or we made two different measurements.
- However, the inverse distance weighting leads to

$$q_C = \frac{q_A + q_{A'} + q_B}{3} = \frac{2}{3} \cdot q_A + \frac{1}{3} \cdot q_B.$$

9. Dual Inverse Distance Weighting

- To overcome the above limitation, a recent paper proposed a new method.
- This method is empirically better than all previously proposed attempts to overcome this limitation.
- In this method, we give more weight to the points which are more distant from others:

$$w_i \sim (d(x, x_i))^{-p} \cdot \left(\sum_{j \neq i} (d(x_i, x_j))^{p_2} \right), \text{ for some } p_2 > 0.$$

- Let us show, on an example, that this idea indeed helps overcome the above limitation.

10. Dual Inverse Distance Weighting (cont-d)

- Indeed, in the above example we get the following expressions for the additional factors $f_i = \sum\limits_{j \neq i} (d(x_i, x_j))^{p_2}$:

$$f_A = (d(A, A'))^{p_2} + (d(A, B))^{p_2} \approx (d(A, B))^{p_2},$$

$$f_{A'} = (d(A', A))^{p_2} + (d(A', B))^{p_2} \approx (d(A, B))^{p_2},$$

$$f_B = (d(B, A))^{p_2} + (d(B, A'))^{p_2} \approx 2(d(A, B))^{p_2}.$$

- So, the weights w_A and $w_{A'}$ with which we take the values q_A and $q_{A'}$ are proportional to

$$w_A \approx w_{A'} \sim (d(A, C))^{-p} \cdot f_A \approx (d(A, C))^{-p} \cdot (d(A, B))^{p_2}.$$

- Meanwhile,

$$w_B \approx w_B \sim (d(B, C))^{-p} \cdot f_2 \approx (d(A, C))^{-p} \cdot 2(d(A, B))^{p_2}.$$

- The weight w_B is thus twice larger than the weights w_A and $w_{A'}$: $w_B = 2w_A = 2w_{A'}$.

11. Dual Inverse Distance Weighting (cont-d)

- So the interpolated value of q_C is equal to

$$q_C = \frac{w_A \cdot q_A + w_{A'} \cdot q_{A'} + w_B \cdot q_B}{w_A + w_{A'} + w_B} =$$

$$\frac{w_A \cdot q_A + w_A \cdot q_{A'} + 2w_A \cdot q_A}{w_A + w_{A'} + 2w_A}.$$

- Let us divide both numerator and denominator by $2w_A$ and take into account that $q_{A'} = q_A$.
- We conclude that $q_C = \dfrac{q_A + q_B}{2}$, i.e., exactly the value that we wanted.

12. Second Challenge: Why Dual Inverse Distance Weighting?

- In view of the above, it is desirable to come up with a theoretical explanation for this method.
- This is the second challenge that we take on in this talk.

13. What Is Scale Invariance

- When we process the values of physical quantities, we process real numbers.
- The numerical value of each quantity depends on the measuring unit.
- For example, suppose that we measure the distance in kilometers and get a numerical value d such as 2 km.
- Alternatively, we could use meters instead of kilometers.
- In this case, the exact same distance will be described by a different number: 2000 m.

14. What Is Scale Invariance (cont-d)

- In general:
  - if we replace the original measuring unit with a new one which is λ times smaller,
  - all numerical values will be multiplied by λ:

$$x \to \lambda \cdot x.$$

- Scale-invariance means that the result of interpolation should not change if we change the measuring unit.
- Let us analyze how this natural requirement affects interpolation.

15. General Case of Distance-Dependent Interpolation

- Let us consider the general case, when the further the point, the smaller the weight.
- In precise terms, the weight w_i is proportional to $f(d(x, x_i))$ for some decreasing $f(z)$: $w_i \sim f(d(x, x_i))$.
- Since the weights should add up to 1, we conclude that:

$$w_i = \frac{f(d(x, x_i))}{\sum\limits_j f(d(x, x_j))}, \text{ so } q = \sum_{i=1}^{n} \frac{f(d(x, x_i))}{\sum\limits_j f(d(x, x_j))} \cdot q_i.$$

- In this case, scale-invariance means that:

$$\sum_{i=1}^{n} \frac{f(\lambda \cdot d(x, x_i))}{\sum\limits_j f(\lambda \cdot d(x, x_j))} \cdot q_i = \sum_{i=1}^{n} \frac{f(d(x, x_i))}{\sum\limits_j f(d(x, x_j))} \cdot q_i.$$

16. Let Us Show That Scale-Invariance Leads to Inverse Distance Weighting

- Indeed, let us consider the case when we have only two measurement results:
  - at the point x_1, we got the value $q_1 = 1$, and
  - at point x_2, we got the value $q_2 = 0$.
- Then, for any point x, if we use the original distance values $d_1 \stackrel{\text{def}}{=} d(x, x_1)$ and $d_2 \stackrel{\text{def}}{=} d(x, x_2)$, we get:
$$q = \frac{f(d_1)}{f(d_1) + f(d_2)}.$$
- So, scale invariance implies
$$\frac{f(\lambda \cdot d_1)}{f(\lambda \cdot d_1) + f(\lambda \cdot d_2)} = \frac{f(d_1)}{f(d_1) + f(d_2)}.$$
- If we take the inverse of both sides, we get:
$$\frac{f(\lambda \cdot d_1) + f(\lambda \cdot d_2)}{f(\lambda \cdot d_1)} = \frac{f(d_1) + f(d_2)}{f(d_1)}.$$

17. Scale-Invariance Proof (cont-d)

- Subtracting number 1 from both sides, we get:

$$\frac{f(\lambda \cdot d_2)}{f(\lambda \cdot d_1)} = \frac{f(d_2)}{f(d_1)}.$$

- If we divide both sides by $f(d_2)$ and multiply by $f(\lambda \cdot d_1)$, we separate d_1 and d_2:

$$\frac{f(\lambda \cdot d_2)}{f(d_2)} = \frac{f(\lambda \cdot d_1)}{f(d_1)}.$$

- The left-hand side does not depend on d_1; thus, the right-hand side does not depend on d_1 either.
- It must thus depend only on λ; let us denote it by $c(\lambda)$.
- Then, from $\dfrac{f(\lambda \cdot d_1)}{f(d_1)} = c(\lambda)$, we conclude that

$$f(\lambda \cdot d_1) = c(\lambda) \cdot f(d_1).$$

18. Scale-Invariance Proof (cont-d)

- It is known that for decreasing functions $f(z)$, the only solutions to this functional equation are:

$$f(z) = c \cdot z^{-p} \text{ for some } p > 0.$$

- For this function $f(z)$, the extrapolated value has the form $\sum w_i' \cdot q_i$, with

$$w_i' = \frac{c \cdot (d(x, x_i))^{-p}}{\sum\limits_{j=1}^{n} c \cdot (d(x, x_j))^{-p}}.$$

- If we divide both numerator and denominator by c, we get exactly the inverse distance weighting formula.

19. Comment

- The equation $f(\lambda \cdot d_1) = c(\lambda) \cdot f(d_1)$ is easy to solve for smooth function $f(x)$.
- Indeed, differentiating both sides by λ and taking $\lambda = 1$, we get $f'(d_1) \cdot d_1 = \alpha \cdot f(d_1)$, where $\alpha \stackrel{\text{def}}{=} c'(1)$.
- So, $\dfrac{df}{dd_1} = \alpha \cdot f$.
- If we divide both sides by f and multiply by dd_1, we separate d_1 and f: $\dfrac{df}{f} = \alpha \cdot \dfrac{dd_1}{d_1}$.
- Integrating both sides, we get $\ln(f) = \alpha \cdot \ln(d_1) + C$, where C is the integration constant.
- Applying $\exp(z)$ to both sides, we get $f(d_1) = c \cdot d_1^\alpha$, where $c \stackrel{\text{def}}{=} \exp(C)$.
- Since the function $f(z)$ is decreasing, we should have $\alpha < 0$, i.e., $\alpha = -p$ for some $p > 0$. Q.E.D.

20. What We Want

- We want to overcome the challenge.
- For this, we multiply the previous weights $f(d(x, x_i)) = (d(x, x_i))^{-p}$ by an additional factor f_i.
- This factor should depend on how far away is location x_i from other locations.
- The further away the location x_i from other locations, the higher the factor f_i shall be.
- So, the factor f_i should be larger or smaller depending on our degree of confidence in the following statement:

$$d(x_i, x_1) \text{ is large and } \ldots d(x_i, x_n) \text{ is large.}$$

21. Let Us Use Fuzzy Techniques

- We need to translate the above informal statement into precise terms.
- A reasonable idea is to use fuzzy techniques – techniques specifically designed for such a translation.
- To each basic statement – like "d is large" – we assign a degree to which this statement is true.
- This degree is usually denoted by $\mu(d)$, so:
  - the degree to which $d(x_i, x_1)$ is large is $\mu(d(x_i, x_1))$;
  - the degree to which $d(x_i, x_2)$ is large is $\mu(d(x_i, x_2))$; etc.
- We need estimate the degree to which the above "and"-statement is satisfied.

22. Let Us Use Fuzzy Techniques (cont-d)

- In fuzzy techniques, we combine the above degrees by an appropriate "and"-operation $f_\&(a,b)$):

$$f_\&(\mu(d(x_i,x_1)),\ldots,\mu(d(x_i,x_n))).$$

- It is known that for any "and"-operation and for any $\varepsilon>0$, there exists an ε-close "and"-operation:

$$f_\&(a,b)=g^{-1}(g(a)+g(b)) \text{ for some monotonic } g(a).$$

- The approximation error ε can be arbitrarily small.
- So, for all practical purposes, we can safely assume that the actual "and"-operation has this g-based form.
- So, f_i should monotonically depend on the expression

$$g^{-1}(g(\mu(d(x_i,x_1)))+\ldots+g(\mu(d(x_i,x_n)))).$$

23. Let Us Use Fuzzy Techniques (cont-d)

- Since the function g^{-1} is monotonic, this means that f_i is a monotonic function of the expression

$$G(d(x_i,x_1))+\ldots+G(d(x_i,x_n))), \text{ where } G(d) \stackrel{\text{def}}{=} g(\mu(d)).$$

- In other words, for some monotonic function $F(z)$:

$$f_i = F(G(d(x_i,x_1)) + \ldots + G(d(x_i,x_n))).$$

24. Let's Recall the Motivation for the Factors f_i

- The main motivation for introducing the factors f_i is to make sure that:
  - for the midpoint C between A and B,
  - we will have the estimate $\dfrac{q_A + q_B}{2}$.
- Let us consider the case when:
  - we have m measurement locations $A_1, \ldots, A_m$ in the close vicinity of the location A and
  - we have one measurement result at location B.
- Let $d \stackrel{\text{def}}{=} d(A, B)$. Then, $d(A_i, C) = d(B, C) = d/2$, so the formula for q becomes
$$q = \frac{m \cdot f_A \cdot q_A + f_B \cdot q_B}{m \cdot f_A + f_B}.$$
- We want to make sure that this value is equal to the arithmetic average $\dfrac{q_A + q_B}{2}$.

25. Motivation for the Factors f_i (cont-d)

- Thus, the coefficient at q_A in the formula for q should be equal to $1/2$:
$$\frac{m \cdot f_A}{m \cdot f_A + f_B} = \frac{1}{2}.$$
- If we multiply both side by their denominators and subtract $m \cdot f_A$ from both sides, we get $m \cdot f_A = f_B$.
- Since $f_i = F(G(d(x_i, x_1)) + \ldots + G(d(x_i, x_n)))$, this means $m \cdot F(G(d) + (m-1) \cdot G(0)) = F(m \cdot G(d))$.
- In the limit $d = 0$, this formula becomes
$$m \cdot F(m \cdot G(0)) = F(m \cdot G(0)), \text{ so } F(m \cdot G(0)) = 0.$$
- Since the function $F(z)$ is monotonic, we cannot have $G(0) \neq 0$, since then we would have $F(z) = 0$ for all z.
- Thus, $G(0) = 0$, $F(G(0)) = F(0) = 0$, and the above formula takes the form $F(m \cdot G(d)) = m \cdot F(G(d))$.

26. Motivation for the Factors f_i (cont-d)

- This is true for any value $z = G(d)$, so we have

$$F(m \cdot z) = m \cdot F(z) \text{ for all } m \text{ and } z.$$

- For $z = 1$, we get $F(m) = c \cdot m$, where $c \stackrel{\text{def}}{=} F(1)$.
- For $z = 1/m$, we have $F(1) = c = m \cdot F(1/m)$, hence

$$F(1/m) = c \cdot (1/m).$$

- Similarly, we get

$$F\left(\frac{p}{q}\right) = F\left(p \cdot \frac{1}{q}\right) = p \cdot F\left(\frac{1}{q}\right) = p \cdot \left(c \cdot \frac{1}{q}\right) = c \cdot \frac{p}{q}.$$

- So, for all rational values $z = p/q$, we get $F(z) = c \cdot z$.
- Since the function $F(z)$ is monotonic, the formula $F(z) = c \cdot z$ is true for all values z.

27. Motivation for the Factors f_i (cont-d)

- So, $f_i = c \cdot (G(d(x_i, x_1)) + \ldots + G(d(x_i, x_n))$.
- Dividing both the numerator and the denominator of the formula for q by the coefficient c, we conclude that

$$q = \sum_{i=1}^{n} \frac{F_i \cdot (d(x, x_i))^{-p} \cdot q_i}{\sum\limits_{j=1}^{n} F_j \cdot (d(x, x_j))^{-p}}, \text{ where } F_i \stackrel{\text{def}}{=} \sum_j G(d(x_i, x_j)).$$

28. Let Us Now Use Scale-Invariance

- We want to make sure that the estimate for q does not change after re-scaling $d(x,y) \to d'(x,y) = \lambda \cdot d(x,y)$.
- So, $q = q'$ where

$$q = \sum_{i=1}^{n} \frac{F_i \cdot (d(x,x_i))^{-p} \cdot q_i}{\sum\limits_{j=1}^{n} F_j \cdot (d(x,x_j))^{-p}}, \text{ where } F_i \stackrel{\text{def}}{=} \sum_j G(d(x_i,x_j)),$$

$$q' = \sum_{i=1}^{n} \frac{F'_i \cdot (d'(x,x_i))^{-p} \cdot q_i}{\sum\limits_{j=1}^{n} F'_j \cdot (d'(x,x_j))^{-p}}, \text{ where } F'_i \stackrel{\text{def}}{=} \sum_j G(d'(x_i,x_j)).$$

- Here, $(d'(x,x_i))^{-p} = \lambda^{-p} \cdot (d(x,x_i))^{-p}$.

29. Let Us Use Scale-Invariance (cont-d)

- Dividing both the numerator and the denominator of q'-formula by λ^{-p}, we get

$$q' = \sum_{i=1}^{n} \frac{F_i' \cdot (d(x, x_i))^{-p} \cdot q_i}{\sum\limits_{j=1}^{n} F_j' \cdot (d(x, x_j))^{-p}}.$$

- The two expressions q and q' are linear in q_i.
- Thus, their equality implies that coefficients at each q_i must be the same.
- In particular, this means that the ratios of the coefficients at q_1 and q_2 must be equal, i.e., we must have

$$\frac{F_1 \cdot (d(x, x_1))^{-p}}{F_2 \cdot (d(x, x_2))^{-p}} = \frac{F_1' \cdot (d(x, x_1))^{-p}}{F_2' \cdot (d(x, x_2))^{-p}}, \text{ i.e., } \frac{F_1}{F_2} = \frac{F_1'}{F_2'}.$$

30. Let Us Use Scale-Invariance (cont-d)

- For the case when we have three points with $d(x_1, x_2) = d(x_1, x_3) = d$ and $d(x_2, x_3) = D$, we get:

$$\frac{2G(d))}{G(d) + G(D)} = \frac{2G(\lambda \cdot d))}{G(\lambda \cdot d) + G(\lambda \cdot D)}.$$

- Inverting both sides, multiplying both sides by 2 and subtracting 1 from both sides, we conclude that

$$\frac{G(D)}{G(d)} = \frac{G(\lambda \cdot D)}{G(\lambda \cdot d)} \text{ for all } \lambda, d, \text{ and } D.$$

- We already know – from the first proof – that this implies that $G(d) = c \cdot d^{p_2}$ for some c and p_2.
- By dividing both numerator and denominator by c, we can get $c = 1$.
- Thus, we indeed get a justification for the dual inverse distance weighting.

31. 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-1242122 (Cyber-ShARE Center of Excellence).