Geospatial uncertainties: a focus on intervals and spatial models based on inverse distance weighting

Priscillia Labourg[1][0009−0005−6780−7380], Sébastien Destercke[1][0000−0003−2026−468X], Romain Guillaume[2][0000−0003−4978−630X], Jeremy Rohmer[3][0000−0001−9083−5965], Benjamin Quost[4][0000−0002−0456−9953], and Stéphane Belbèze[3][0000−0001−9072−486X]

[1] CNRS, laboratoire Heudiasyc, UMR 7253
[2] Université de Toulouse II Jean Jaurès, IRIT, UMR 5505
[3] Bureau de Recherches Géologiques et Minières
[4] Université de Technologie de Compiègne, Heudiasyc, UMR 7253

Abstract. Processing geospatial data requires to manage many sources of uncertainties; some appear in classical inference problems, some others are specific to this setting. The goal of this paper is to study the management of these uncertainties via standard intervals and sets when the inference model considered relies on inverse distance weighting. We provide a general discussion with examples, together with a study of the associated optimisation problems induced by different sources of uncertainty. We conclude the paper by an illustration on a semi-synthetic use case, generated according to data recorded via real studies.

Keywords: Geospatial data processing · Uncertainty · Imprecise Probability · Interval data

1 Introduction

We consider the difficult problem of modelling and making predictions from geospatial data, which are subject to many kinds of uncertainty. Over the recent years, a number of papers addressed this issue [4, 14], focusing in turn on the representation [5], modelling [6], or propagation [19, 17] of uncertainty.

Uncertainties are commonly separated into two categories: *aleatoric* (or stochastic) uncertainty is deemed inherent to the modelled phenomenon, whereas *epistemic* uncertainty [10] comes from a lack of information. Yet, distinguishing between them is neither easy nor operational: for instance, while measurement error is usually perceived as stochastic, using its uncertainty model on a single datum makes it epistemic, as this datum value may reasonably be considered as non-aleatoric. This paper does not aim at discussing large, general taxonomies of uncertainty, but rather to focus on classical treatments of geospatial problems, and to discuss the various uncertainties that may occur in the associated pipelines, from measured data to final representations.

For this purpose, and for the sake of simplicity, we will focus on intervals and on a very simple predictive model for geospatial problems, referred to as the Inverse Distance Weighted model (IDW), which dates back to the late 60s [20] but remains nevertheless commonly used (see e.g. the review by [13]). This model is deterministic: it produces precise output predictions when given precise inputs, in opposition to geostatistical models such as kriging [12] that can produce stochastic outputs from precise inputs (i.e., when using their Gaussian process interpretation). We voluntarily choose this simple yet popular model to avoid the inherent additional difficulties that come with the explanation of a complex model. It should also be noted that while we will consider its simplest version, IDW can be made more complex and flexible, and remains the focus of recent research [15].

We introduce the basic ideas of IDW and of interval uncertainty in Section 2. Section 3 discusses the various sources of uncertainty that can arise in standard geospatial studies, their handling in interval IDW, and illustrates them with the help of a simplified running example. Finally, Section 4 presents the results of considering specific uncertainties in the propagation process.

2 Preliminaries

This section presents the basic inverse distance weighting model, and a formalisation of interval- or set-valued uncertainty.

2.1 Geospatial data and IDW model

In basic geospatial problems, we assume that we observe n points of coordinates $s_j \in \mathbb{R}^2$ (within a restricted domain $D \subseteq \mathbb{R}^2$), together with a quantity of interest $Z(s_j) \in \mathbb{R}$. We therefore have n observations $\{s_j, Z(s_j)\}$. The goal is then to estimate from these observations the value $z(s)$ at points s where we do not have information. IDW is a very simple model to achieve this:

$$z(s) = \begin{cases} z(s_j) & \text{if } d = 0 \\ \frac{\sum_{j=1}^{N} w_j \times z(s_j)}{\sum_{j=1}^{N} w_j} & \text{else} \end{cases} \tag{1}$$

with $w_j := d_{s,s_j}^{-u}$, d_{s,s_j} a distance measure (we typically consider the Euclidean distance) between points s and s_j, and $u \in \mathbb{R}^+$ a parameter value which sets how fast weights are decreasing, and can be identified to a smoothing parameter. For $u \to 0$, we get that $z(s)$ tends to the average value, and for $u \to \infty$, we get that $z(s)$ tends to the nearest neighbour value.

2.2 Interval and set-valued uncertainty

When it comes to model uncertainties about a quantity X taking values in $\mathcal{X}$, the mathematical tools that are probabilities and (convex) sets are arguably two extremes of the representation spectrum: probabilities amount to assign a unique value $P(A) \in [0,1]$ to any event $A \subseteq \mathcal{X}$, while providing a set $S \subseteq \mathcal{X}$ of possible values only allows to know whether an event is necessarily true ($S \subseteq A$), i.e. implied by S; necessarily false ($S \cap A = \emptyset$), i.e. inconsistent with S; or totally possible yet not certain ($S \cap A \neq \emptyset$ and $S \cap A^c \neq \emptyset$). This can be formalised by the two following binary measures:

$$\underline{P}_S(A) = \begin{cases} 1 & \text{if } S \subseteq A, \\ 0 & \text{else;} \end{cases} \quad \text{and} \quad \overline{P}_S(A) = \begin{cases} 1 & \text{if } S \cap A \neq \emptyset, \\ 0 & \text{else.} \end{cases}$$

In this paper, we mostly consider convex subsets of $\mathbb{R}$, that is, intervals $[\underline{x}, \overline{x}]$; or convex subsets of $\mathbb{R}^2$, in which case we will restrict ourselves to rectangles $[\underline{x}^1, \overline{x}^1] \times [\underline{x}^2, \overline{x}^2]$, where x^j denotes the projection of x onto the jth dimension of the space.

Note that a set S can be interpreted in two ways: either as expressing the imprecise measurement of a fixed, yet unknown quantity, or as being associated to the set of all probabilities that has S for support, i.e. $\mathcal{P}_S = \{P : P(S) = 1\}$. While in general this does not change how one will compute from them, this shows that the mathematical model and its interpretation are two different things, as intervals can be perceived as modelling pure epistemic uncertainty, or as modelling an unknown probability defined only by its support, possibly modelling an underlying ill-known aleatoric uncertainty.

3 Uncertainties in geospatial problems: a discussion

In this paper, we consider the simple pipeline described in Figure 1. We discuss various sources of uncertainties within this pipeline, and how they impact the handling of the IDW model. Note also that we will not deal explicitly with the problem of model estimation from data, and will only mention it as a natural way to build uncertainty in the model (whether in a geostatistical approach or for deterministic models). Indeed, while model estimation is an important step of data-driven approach, it is not intrinsically linked to uncertainty (as one could well pick a single estimate).

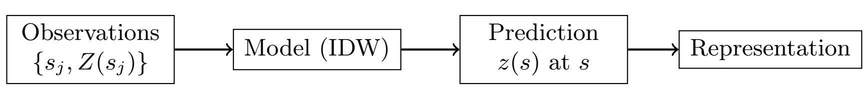


Fig. 1: Classical geospatial treatment pipeline

Two features distinguish this problem from classical uncertainty quantification: first, values s_i are spatial, which should be accounted for in their treatment (e.g., by modelling spatial dependence); furthermore, the end result should not be only a prediction (as would be the case in a standard classification or regression setting), but a representation as a map. This latter user-oriented step is particularly important, as the conveyed information may be of a critical nature (e.g., pollution concentrations in soil as a support for urban planning). Figure 2 provides an illustration of a use case considered in Section 4, where each dot represents a sampled location, and its colour indicates the pollution level.

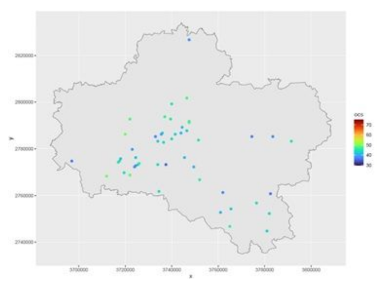
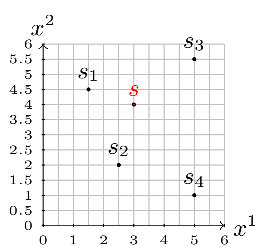


Fig. 2: Synthetic use-case for the Loiret department in France: dots represent the 50 sampled locations, colours indicate the pollution level

In the remaining of this section, we also consider the following running example to illustrate our various points. Though multiple predictors (covariates) can be considered in the prediction task, we only consider here geographical coordinates.

Example 1. Consider the very simple toy problem pictured in Figure 3, where we have four points and want to evaluate the value in a given point. Coordinates are explicitly specified in the figure, and outputs and distances are provided in the associated table. In this case, the resulting estimate with $u = 2$ is $z(s) = 2.7$.



point s_i	s_1	s_2	s_3	s_4
output $Z(s_i)$	3	1	5	2
distance d_{s,s_i}	1.6	2.1	2.7	3.6

Fig. 3: Geospatial toy example

3.1 Uncertainties in a datum

In the following, we refer to any tuple $\{s_i, z(s_i)\}$ as datum. Two kinds of uncertainties can affect this datum, on the measured quantity of interest and on the localisation.

Measurement A first kind of uncertainty concerns the measurement $z(s_i)$ itself. Such uncertainties usually comes from measurement errors, that arise from the sensor or even from the situation—for instance, if one wants to measure a concentration in a soil, the obtained value will vary according to the soil volume considered or extracted (see the discussion in [3]). Censored data, i.e., measurements below below the limit of detection (LoD) $\hat{y}$ of a quantity, also induce uncertainties in the form of intervals $[0, \hat{y}]$. Generally speaking, the measurement then becomes interval-valued, i.e. we have $z(s_i) \in [\underline{z}(s_i), \overline{z}(s_i)]$. Should we be interested in assessing the value $z(s)$ at position s, we would have to compute the interval $[\underline{z}(s), \overline{z}(s)]$ defined by

$$\underline{z}(s) = \frac{\sum_{j=1}^{N} d(s, s_j)^{-u} \times \underline{z}(s_j)}{\sum_{j=1}^{N} d(s, s_j)^{-u}}, \quad \overline{z}(s) = \frac{\sum_{j=1}^{N} d(s, s_j)^{-u} \times \overline{z}(s_j)}{\sum_{j=1}^{N} d(s, s_j)^{-u}}. \tag{2}$$

Example 2. Consider Example 1 with the following intervals:

$$[\underline{z}(s_1), \overline{z}(s_1)] = [2, 4], \qquad [\underline{z}(s_2), \overline{z}(s_2)] = [0, 1.5],$$
$$[\underline{z}(s_3), \overline{z}(s_3)] = [4.5, 5.5], \qquad [\underline{z}(s_4), \overline{z}(s_4)] = [2, 2],$$

with the second being possibly a censored measurement, and the fourth is precise. Setting $u = 2$, we get the interval-valued prediction $[\underline{z}(s), \overline{z}(s)] = [1.86, 3.38]$.

Localisation A second uncertainty may come from an ill-known location modelled here as a box $[s]_i = [\underline{s}_i^1, \overline{s}_i^1] \times [\underline{s}_i^2, \overline{s}_i^2]$. The cause may be a bad registration, or a localisation device with limited accuracy. For instance, measuring air quality in a city through a phone using GPS localisation would cast uncertainty on the exact location of the measurement point.

The precision of the localisation can vary greatly even within the same set of recorded data, depending on the GNSS receiver quality, the kind of area (urban, agricultural, forest , etc), or the data treatment of the localisation process. This uncertainty can range from centimeters to kilometers [21], with the vast majority being within the 500 meter range.

Note that in this case, how one can get interval bounds from Equation (1) when values s_j become imprecise is not obvious. First, one has to compute resulting lower/upper bounds over $d(s, s_j)$; if d is Euclidean, then this is fairly easy (for $\underline{d}(s, s_j)$: the closest coordinate to s should be considered for each dimension, and conversely the furthest for $\overline{d}(s, s_j)$)—see for example [1, proposition 1].

Second, one has to find for which values of $d(s, s_j) \in [\underline{d}(s, s_j), \overline{d}(s, s_j)]$ the minimum/maximum of Equation (1) is reached, given a value $z(s_j)$. For simplicity, let us consider that $w_j = d(s, s_j)^{-u}$, and note that as all sets are convex and all functions are continuous, optimising over $[s]_j$ or over $[\underline{w}_j, \overline{w}_j]$ is equivalent. We then have to find

$$\underline{z}(s) = \inf_{w_j \in [\underline{w}_j, \overline{w}_j]} \frac{\sum_{j=1}^N w_j \times z(s_j)}{\sum_{j=1}^N w_j}, \quad \overline{z}(s) = \sup_{w_j \in [\underline{w}_j, \overline{w}_j]} \frac{\sum_{j=1}^N w_j \times z(s_j)}{\sum_{j=1}^N w_j}. \quad (3)$$

It can easily be checked that taking the partial derivative of $z(s)$ with respect to a given w_j, we get

$$\frac{\partial z'(s)}{\partial w_j} = \frac{z(s_j) \times \sum_{i=1}^N w_i - \sum_{i=1}^N w_i \times z(s_i)}{(\sum_{i=1}^N w_i)^2} = \frac{\sum_{i=1}^N w_i(z(s_j) - z(s_i))}{(\sum_{i=1}^N w_i)^2}, \quad (4)$$

meaning that $z(s)$ is monotone in w_j, as the partial derivative does not depend on it (as $z(s_j) - z(s_j) = 0$), but can be either positive ($z(s)$ increasing in w_j) or negative (decreasing), depending on the values taken by the uncertain w_i (note that $z(s_j) - z(s_i)$ can be positive or negative, but are known).

This means that the solutions to Equation (3) are obtained on the bounds of intervals $[\underline{w}_j, \overline{w}_j]$: they can reasonably be identified with a search procedure if the number of imprecise locations is limited, but not if this latter is high due to the exponential increase of extreme points. In this latter case, notice also that Equation (1) is a fraction of two linear functions of w_j, with the domain of $[w]_j$ being bounded and $\underline{w}_j \geq 0, \forall j$: hence, one can also use linear fractional programming and the Charnes-Cooper transform [7] to solve this problem efficiently.

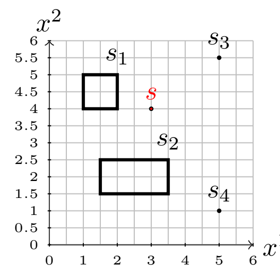


Fig. 4: Localisation uncertainty

Example 3. Consider the two sets in Figure 4 for $[s]_1$ and $[s]_2$: we then get

$$[\underline{d}_{s,s_1}, \overline{d}_{s,s_1}] = [1, 2.23], \quad [\underline{d}_{s,s_2}, \overline{d}_{s,s_2}] = [1.5, 2.92].$$

Setting $u = 2$, we then have that $[\underline{z}(s), \overline{z}(s)] = [2.20, 2.97]$, with the value 2.20 obtained for $d_{s,s_1} = 2.23$ and $d_{s,s_2} = 1.5$, and the value 2.97 for $d_{s,s_1} = 1$ and $d_{s,s_2} = 2.92$.

3.2 Data imperfection

A second kind of data imperfection concerns the population level. These uncertainties mainly come from the fact that places where samples can be taken are often constrained, e.g., by the situation (for instance, one cannot sample soil under a building). As such aspects typically require to work at the population level and to integrate them within the model capabilities, we will only briefly sketch them.

Outliers Outliers are anomalous measurements, or "out-of-distribution" samples that can greatly affect the end-result of the inferred values. In our example, this could correspond to having a measure $Z(s_i) = -10000$.

Clearly, a data can be deemed an outlier only with respect to the others, hence the need to consider the whole population (or a representative sample thereof). Classical approaches to address this issue either detect the outliers [8] and possibly remove them, or use statistical approaches robust to outliers [11], for example replacing a weighted average in Equation (1) by a (weighted) median. Note that outliers do not really induce additional uncertainty, but they induce errors (in terms of higher variance or systematic bias) which should be avoided.

Data clustering and sparsity A simple look at Figure 2 shows that some regions have a high density of sampled values, as opposed to others.

As for outliers, regions of clustered samples do not really induce additional uncertainties, but their influence should also be mitigated, as their importance could easily be overestimated, for instance in the IDW model. Sparse regions, on the other hand, do induce some kind of uncertainty, as predictions are made in these regions from scarce, less reliable information, at least from a distance-to-sample perspective. Geostatistical approaches such as kriging [12] typically try to address this issue by providing confidence intervals and proposing specific (variogram) estimation strategies; one could also think of other strategies such that using a local density estimation procedure, akin to inverse sampling-intensity weighting [9].

3.3 Model and prediction uncertainties

Another source of uncertainty comes from the inference process performed when computing $z(s)$, even when the data and their amount is deemed sufficient to have reliable inferences.

Model parameters When the model depends on parameters, for instance the exponent u in the IDW model, it is clear that different parameter choices lead to different inferences. It seems then natural to allow the parameter to be uncertain or to vary, so that a sensitivity analysis can be conducted. In our case, this would amount to consider an interval $[\underline{u}, \overline{u}]$ and to compute

$$\underline{z}(s) = \inf_{u \in [\underline{u}, \overline{u}]} \frac{\sum_{j=1}^{N} w_j \times z(s_j)}{\sum_{j=1}^{N} w_j}, \quad \overline{z}(s) = \sup_{u \in [\underline{u}, \overline{u}]} \frac{\sum_{j=1}^{N} w_j \times z(s_j)}{\sum_{j=1}^{N} w_j}, \qquad (5)$$

with $w_j = d_{s,s_j}^{-u}$. This optimisation problem is trickier to solve, yet u being here one-dimensional means that a grid-search is possible.

Prediction uncertainty It is clear that any propagation of uncertainty in the data or in the model to $z(s)$ induces an uncertainty on the predictions made—see Equations (2), (3), and (5)), although prediction uncertainty may arise from other sources such as the assumption that the relation between the input s and the observed output $z(s)$ is non-deterministic: this may for instance be due to error measurements, or more generally to $Z(s)$ being a random variable—note that this assumption within geostatistics has been discussed in [16], as commonly used models remain precise for sampled data). Statistical approaches can then be used to derive confidence or credibility intervals on the predictions, including for the IDW model [2].

3.4 Representation

As stated above, the final representation is often of a critical importance in geospatial applications, since produced maps will often be used to make important decisions such as where to position new constructions and more generally for urban planning. It often happens that final representations consist of regions rather than point-wise estimates, as end-users may be interested in such regions (e.g. aggregation of the results over some administrative districts). Such representation choices being purely subjective, accounting for the uncertainty resulting from these choices can be done in various ways.

In the case where one is only interested in the values for regions, all other things being precise, the problem corresponds to considering a set-valued region $[s]$, e.g., a square or a region corresponding to some administrative entity, in which case we can define the bounds

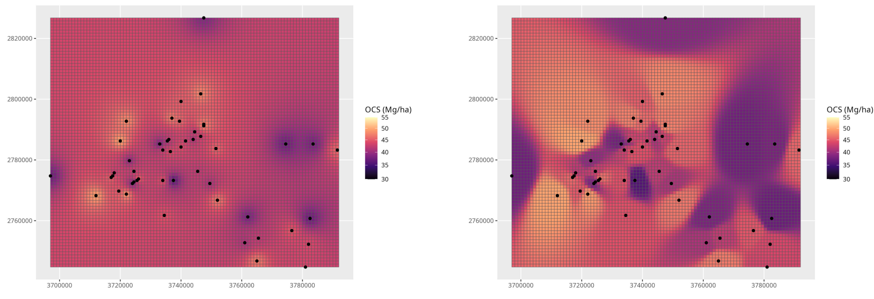
$$\underline{z}([s]) = \min_{s\in[s]} z(s), \quad \overline{z}([s]) = \max_{s\in[s]} z(s) \tag{6}$$

which again may be more or less easy to obtain depending on the nature of the region and of the model. For instance, a typical way to represent maps is through raster, i.e., using a grid of regular squares whose size can vary. In general, the idea is that discretising the representation by computing a partition may have a significant effect on the final representation.

4 An illustrative semi-synthetic use-case

The following use-case focuses on the spatial prediction of soil organic carbon stock (denoted OCS, expressed in Mg/ha) in the 0–30 cm layer. The data is synthetic, and is meant to reproduce the main difficulties encountered in practice for spatial interpolation (clustering of data, limited number of data, etc.). For this purpose, we sub-sample a set of data from the soilgrid dataset (version 2) in [18] by considering the Loiret department in France. In our study, we use a sub-sample of 50 data points displayed in Figure 2 which corresponds to a case of sparsely distributed points.

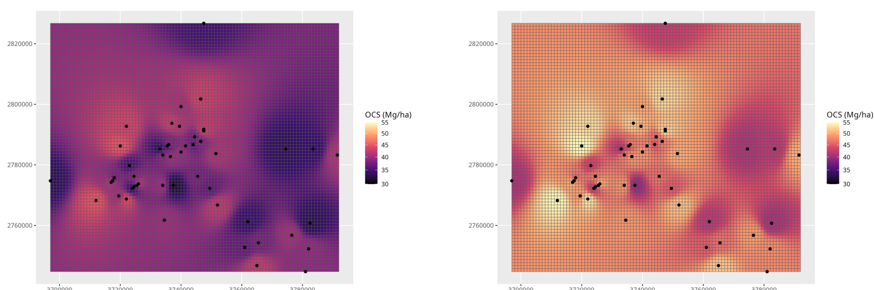
The data is interpolated using IDW and represented through a regular square grid with cells of 1km sides. Figure 5 shows two different values of the u parameter. A classical choice is $u = 2$; using larger values results in the model getting close to a Voronoi diagram—as suggested by the case $u = 8$ displayed in Figure 5. In our case, we will use $u = 4$, as we have seen that $u = 8$ is too strong and $u = 2$ too weak.



(a) u = 2 (b) u = 8

Fig. 5: IDW Interpolation on Loiret dataset with varying u parameter value

We first start with measurement uncertainty: measurements are intervals $[\underline{z(s)}, \overline{z(s)}]$, with $\underline{z(s)} = Z(s) - Z(s) \times \beta$ and $\overline{z(s)} = Z(s) + Z(s) \times \beta$. We assume the level of uncertainty to be homogeneous over our dataset: we set $\beta = 0.1$ for this application.



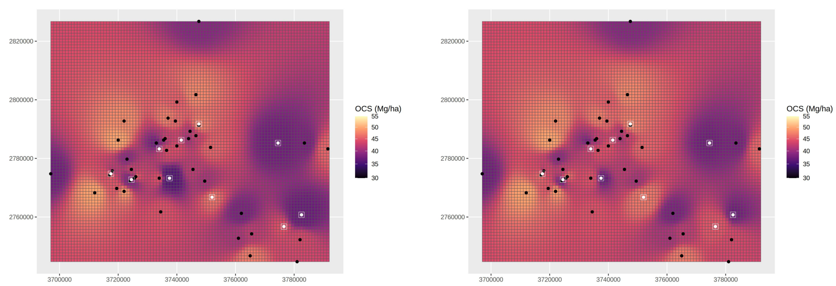
(a) Lower Bound (b) Upper Bound

Fig. 6: Measurement uncertainty for the Loiret application

In our dataset, the measurements span the interval $[35, 50]$. As a consequence, the measurement uncertainty is limited to the $[3.5, 5]$ range, i.e. the differences between upper and lower measurement bounds are close.

This could have been anticipated from Equation (2), since setting the difference between lower and upper bounds to a fixed value (which is almost the case here) results in distances between their weighted averages being the same. This is hinted by Figure 6: the maps of lower and upper bounds are very similar to each other, with the main difference being the values.

Location uncertainty can also be treated as an interval. Since we consider 2D locations, as mentioned in Section 3, the area representing the possible actual positions for a given observation is assumed to be a square. Note that we consider not all observations to be subject to location uncertainty: we randomly select a number of them, so as to visually exemplify the effect of location uncertainty in different parts of the map. Finally, even though the area in our use case is quite large and contains different kinds of environments, we consider location uncertainty to be homogeneous over the affected observations. Due to the size of the area, the uncertainty on position will be expressed as a percentage of the total length of the map.



(a) Minimized values (b) Maximized values

Fig. 7: Pessimistic and optimistic concentration measurements for the Loiret use-case with observations with uncertain locations

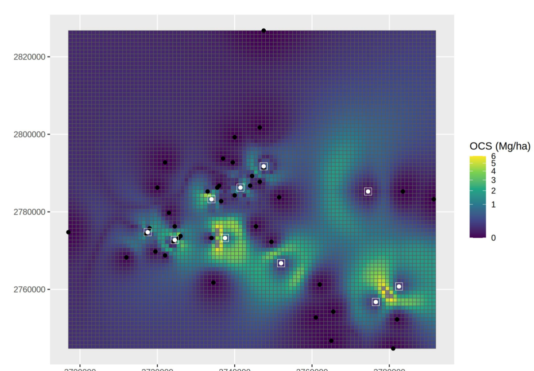


Fig. 8: Measurement uncertainty for the Loiret use-case with observations with uncertain locations

Figure 8 displays the differences between the two maps in Figure 7. These differences span the interval $[0, 6]$, with the majority of them being less than 1. The map also points out where location uncertainty is prominent. It becomes obvious that location uncertainty mitigates the impact of an observation on the interpolation.

5 Conclusion and perspectives

This paper discusses the sources of uncertainties frequently encountered in geospatial studies, focusing on the representation of uncertainty through intervals, and using the inverse distance weighting model, which remains very popular in the geospatial literature. More precisely, we show how intervals can be used to represent these kinds of uncertainty, and provide several examples which give an intuition of the effects of these uncertainties on the inference made by the model.

This preliminary study, however, can be extended in various directions. We may investigate how intervals compare to probabilities or more general uncertainty models. Another important problem is that of estimating parameters from data pervaded with uncertainty. Last, when uncertainty is propagated throughout the model (which therefore corresponds to sets of interpolation functions), how the quality of its predictions can be fairly assessed remains open.

Acknowledgments. We acknowledge financial funding by ANR-HOUSES (grant number: ANR-22-CE56-0006).

References

1. Alarcon, Y.C.C., Destercke, S.: Imprecise gaussian discriminant classification. Pattern Recognition **112**, 107739 (2021)
2. Babak, O., Deutsch, C.V.: Statistical approach to inverse distance interpolation. Stochastic Environmental Research and Risk Assessment **23**, 543–553 (2009)
3. Belbèze, S., Rohmer, J., Négrel, P., Guyonnet, D.: Defining urban soil geochemical backgrounds: A review for application to the french context. Journal of Geochemical Exploration p. 107298 (2023)
4. Bielecka, E., Burek, E.: Spatial data quality and uncertainty publication patterns and trends by bibliometric analysis. Open Geosciences **11**(1), 219–235 (Jan 2019). https://doi.org/10.1515/geo-2019-0018
5. Bulgin, C.E., Thomas, C.M., Waller, J.A., Woolliams, E.R.: Representation Uncertainty in the Earth Sciences. Earth and Space Science **9**(6), e2021EA002129 (2022). https://doi.org/10.1029/2021EA002129
6. Chapman, M.A., Alesheikh, A., Karimi, H.: Error Modeling and Management for Data in Geospatial Information Systems. In: Green, D.R., King, S.D. (eds.) Coastal and Marine Geo-Information Systems: Applying the Technology to the Environment, pp. 355–369. Coastal Systems and Continental Margins, Springer Netherlands, Dordrecht (2003). https://doi.org/10.1007/0-306-48002-6_25
7. Charnes, A., Cooper, W.W.: Programming with linear fractional functionals. Naval Research logistics quarterly **9**(3-4), 181–186 (1962)
8. Chen, D., Lu, C.T., Kou, Y., Chen, F.: On detecting spatial outliers. Geoinformatica **12**, 455–475 (2008)
9. De Bruin, S., Brus, D.J., Heuvelink, G.B., van Ebbenhorst Tengbergen, T., Wadoux, A.M.C.: Dealing with clustered samples for assessing map accuracy by cross-validation. Ecological Informatics **69**, 101665 (2022)

10. Hoffman, F.O., Hammonds, J.S.: Propagation of uncertainty in risk assessments: the need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability. Risk analysis **14**(5), 707–712 (1994)
11. Huber, P.J.: Robust statistics, vol. 523. John Wiley & Sons (2004)
12. Li, J.: A Review of Spatial Interpolation Methods for Environmental Scientists (Jan 2008)
13. Li, J., Heap, A.D.: A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. Ecological Informatics **6**(3), 228–241 (Jul 2011). https://doi.org/10.1016/j.ecoinf.2010.12.003
14. Li, Y., Chen, J., Feng, L.: Dealing with Uncertainty: A Survey of Theories and Practices. IEEE Transactions on Knowledge and Data Engineering **25**(11), 2463–2482 (Nov 2013). https://doi.org/10.1109/TKDE.2012.179
15. Li, Z., Zhang, X., Zhu, R., Zhang, Z., Weng, Z.: Integrating data-to-data correlation into inverse distance weighting. Computational Geosciences **24**(1), 203–216 (Feb 2020). https://doi.org/10.1007/s10596-019-09913-9
16. Loquin, K., Dubois, D.: Kriging and epistemic uncertainty : A critical discussion. In: Jeansoulin, R., papini, O., Prade, H., Schockaert, S. (eds.) Methods for Handling Imperfect Spatial Information, Studies in Fuzziness and Soft Computing Book Series (STUDFUZZ), vol. 256, pp. 269–305. Springer (2010). https://doi.org/10.1007/978-3-642-14755-5_11
17. Phillips, D.L., Marks, D.G.: Spatial uncertainty analysis: Propagation of interpolation errors in spatially distributed models. Ecological Modelling **91**(1), 213–229 (Nov 1996). https://doi.org/10.1016/0304-3800(95)00191-3
18. Poggio, L., De Sousa, L.M., Batjes, N.H., Heuvelink, G.B., Kempen, B., Ribeiro, E., Rossiter, D.: Soilgrids 2.0: producing soil information for the globe with quantified spatial uncertainty. Soil **7**(1), 217–240 (2021)
19. Quelch, J., Cameron, I.T.: Uncertainty representation and propagation in quantified risk assessment using fuzzy sets. Journal of Loss Prevention in the Process Industries **7**(6), 463–473 (Nov 1994). https://doi.org/10.1016/0950-4230(94)80004-9
20. Shepard, D.: A two-dimensional interpolation function for irregularly-spaced data. In: Proceedings of the 1968 23rd ACM National Conference. pp. 517–524. ACM '68, Association for Computing Machinery, New York, NY, USA (Jan 1968). https://doi.org/10.1145/800186.810616
21. Yoo, E.H., Roberts, J.E., Eum, Y., Shi, Y.: Quality of hybrid location data drawn from GPS-enabled mobile phones: Does it matter? Transactions in GIS **24**(2), 462–482 (2020). https://doi.org/10.1111/tgis.12612