A Tool for Uncertainty Analysis of Fuzzy-Probabilistic Models in GIS: Conceptualization & Implementation

J. F. Rios[1]

[1]University at Buffalo, Dept. of Geography, 105 Wilkeson Quad, University at Buffalo, Buffalo NY.
Telephone: (716) 645-0471
Email: jfernand@buffalo.edu

Abstract

Uncertainty quantification is not often performed in spatial analysis and modeling applications. One reason is the lack of tools which can handle various kinds of uncertainty associated with spatial data. In particular, positional uncertainty in objects is often ignored. There has been much research in regards to the modeling of different kinds of positional uncertainty (e.g., measurement error, indeterminacy). However, a conceptualization which considers several kinds while remaining simple enough for everyday use has not been developed. This work presents such a conceptualization along with a Python-based implementation which handles positional and non-positional uncertainty. A real world example is used to show how the tool is able to answer the question "what is the effect of positional uncertainty in my model?".

Keywords: positional uncertainty, fuzzy / probabilistic methods, error propagation, geographic objects.

1. Introduction

The GIScience community has identified the need for methodologies and tools for uncertainty analysis (UA) which are easy to use but also flexible and objective (see Aerts, et al. 2003, Zhang and Goodchild 2002). Although there has been much theoretical work regarding uncertainty in geographical information (Zhang and Goodchild 2002), practical tools, particularly those which incorporate positional uncertainty in objects, are few. This work presents an implementation of a combination of existing methodologies, resulting in a fuzzy-probabilistic UA framework incorporating positional uncertainty, suitable for use in a geographic information system (GIS). The tool is demonstrated using a simple GIS-based groundwater contamination example.

2. Conceptual Framework

In addressing positional uncertainty, the focus is on the object-based view of geography. The conceptual framework (fig. 1) aims to capture the advantages of probabilistic and fuzzy representations of positional uncertainty in geographic objects by categorizing them as "rigid", "deformable", "vertex-defined" or "edge-defined" (Brimicombe 1998, Heuvelink et al. 2007, Rios 2014). A rigid object can be used when there is uncertainty in its location/orientation but not its shape. A deformable object is suitable when shape is uncertain. In both cases, the object possesses a crisply defined boundary and is modeled probabilistically. If the object has a vague or indeterminate boundary, a fuzzy approach is more appropriate. A vertex-defined object is one where uncertainty in each constituent vertex is modeled as a fuzzy point consisting of fuzzy numbers representing Cartesian coordinates. In an edge-defined object, the uncertainty is represented by a perpendicular displacement of the object's edges (line segment connecting two vertices).

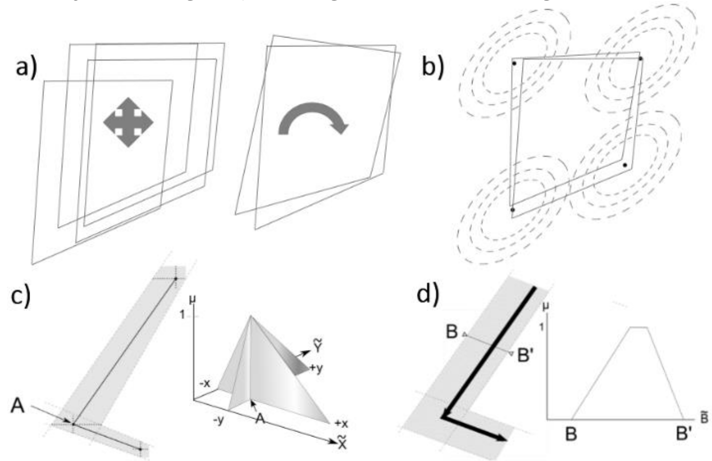

Figure 1: Uncertain objects. a) Rigid polygon, b) deformable polygon, c) vertex-defined polyline, d) edge-defined polyline. Adapted from (Rios 2014).

The choice between the four categories depends on the use case. For instance, the footprint of a building may be well represented by a rigid object since it has a well-defined shape while an edge-defined object might be better for a wetland since its boundary is vague.

The UA framework is based on fuzzy Monte Carlo simulation (FMCS) (Sadeghi, et al. 2010), which handles both probabilistic and fuzzy uncertainty using a nested approach. Probabilistic variables are sampled in an outer loop, which are then held fixed in an inner fuzzy loop. The resulting set of fuzzy numbers are processed to produce a pair of cumulative distribution functions (CDFs) at each membership level, α. The mean and variance of a particular CDF captures probabilistic uncertainty while the effect of fuzzy uncertainty is represented by the separation of these CDFs. In the absence of either probabilistic or fuzzy uncertainty, FMCS reduces to an ordinary Monte Carlo or fuzzy analysis respectively. FMCS is selected over other hybrid methods (Guyonnet et al. 2003, Baudrit et al. 2006) due to its ease of interpretability, at the expense of a difficulty in selecting an appropriate α-level for decision-making.

3. Implementation

Due to its wide support in many GIS, the conceptualization is implemented using the Python programming language as the package Wiggly (available at http://zoidy.github.io/wiggly). The algorithm (fig. 2) is the following. First the user creates a control script which defines the uncertain variables and shapes contained in the analysis model M. Shapes are added to an ObjectManager which stores the definitions for later use. The uncertainty quantification (UQ) block then iterates over an outer loop which evaluates the fuzzy variables using decomposed fuzzy numbers, using (user-selectable) random sampling or the reduced transformation method (Hanss 2002). These are then held constant in a probabilistic inner loop (loop order reversed for ease of implementation compared to FMCS). In the inner loop, the parameters (which are now all single numbers) are passed to a user-created Test Program script which wraps M. A particular value of M, m*, is output and saved. The process is repeated until all probabilistic runs within all sets of fuzzy samples within all α-cuts are processed. If M contains only probabilistic variables, the result is a single CDF and if only fuzzy variables are present, a membership function is output. If both are present, a pair of CDFs at each α-cut is the result.

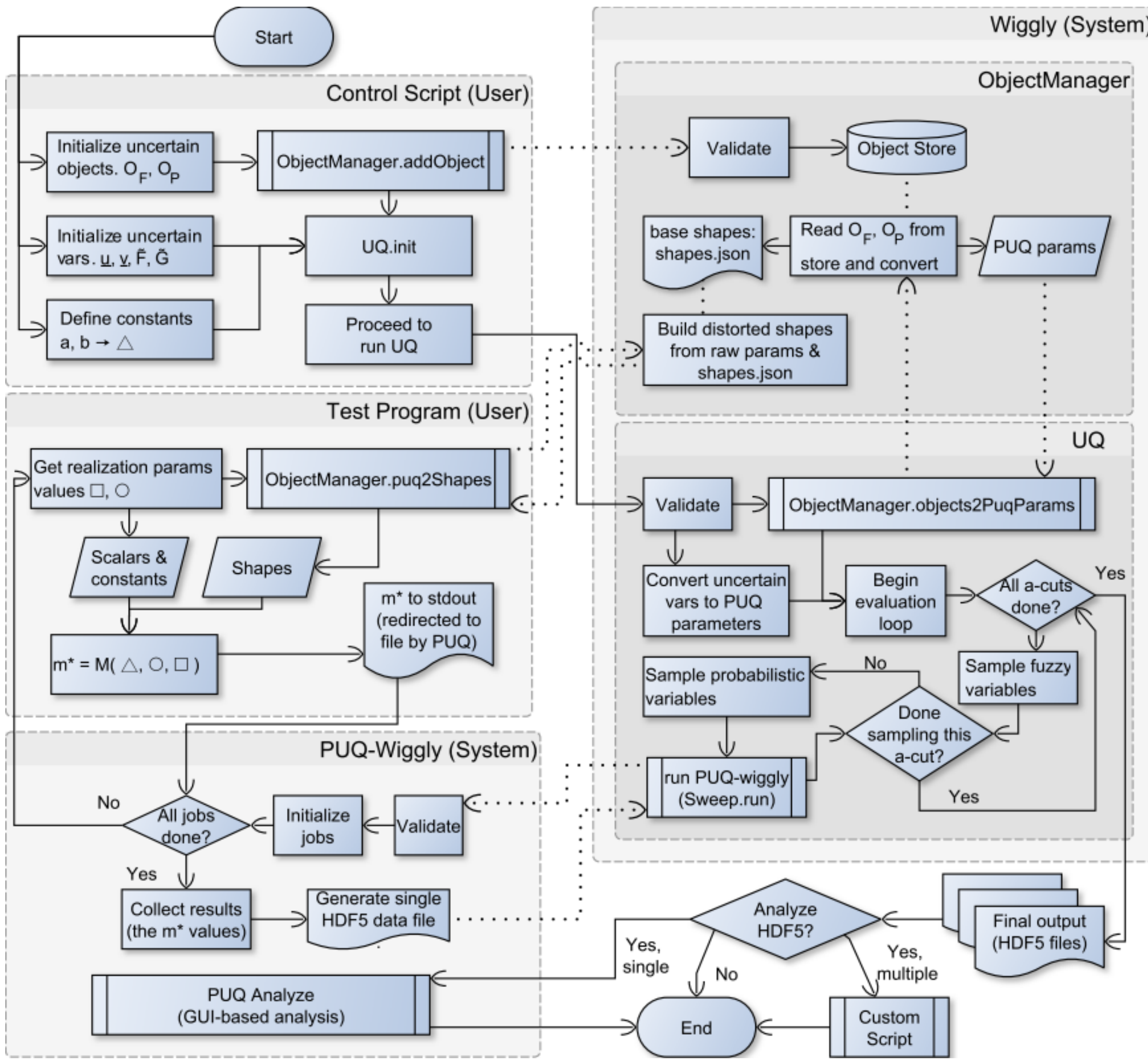

Figure 2: Implementation of the uncertainty conceptualization. The solid arrows indicate primary logic flow. Dotted arrows indicate transfer of flow to (or from) a subroutine. If the dotted arrows do not return, primary logic flow may continue without waiting for a return value. The dotted lines with no arrows indicates interaction between components, without transfer of flow.

To enable GIS interaction, an interface to the QGIS software is made available which allows the use of Wiggly as a loosely-coupled tool to analyze GIS-based models.

4. Comparison to Other Tools

Although general software such as MATLAB or R is capable of UA, purpose-built tools save much effort and are generally easier to use. The DUE (Brown and Heuvelink 2007) is one of the only tools which includes positional uncertainty in the analysis of spatial data. Its main strength is the capability of guiding the user through the modeling process in a structured way and managing certain aspects of uncertainty modeling. It is not capable of executing a model or analyzing results. In comparison, Wiggly's focus is on providing a flexible way of including mixed positional uncertainties in GIS models, managing UA execution, and providing analysis tools in the form of plots. Table 1 compares the main features of the two tools.

	Wiggly	**DUE**
Uncertainty Representations	Probabilistic, fuzzy, mixed	Probabilistic
Positional Uncertainty	Vector	Vector, raster
Attribute Uncertainty	Yes	Yes
Correlation	Yes, manually defined	Yes, aided by an included tool
Programmable	Yes	Yes
GIS coupling	Yes, via included library	No
Graphical interface	No (future work)	Yes
Execute an UA	Yes, models are black-boxes	No, generate realizations only
Process output	Yes, plot results automatically	No

Table 1: Comparison of DUE and Wiggly

5. Example

We wish to determine the nitrate load to a river, due to nearby septic systems (fig. 3), given that the locations of the septic system drainfields and (a portion of) the boundary of the river are uncertain. Contaminant transport is modeled in one-dimension using (Bear 1972, p. 631):

$$C(x,t) = \frac{C_0}{2} f_x(x,t), \tag{1}$$

$$f_x(x,t) = \left(\exp\left\{\frac{x}{2\alpha_x}[1-\phi]\right\} \times erfc\left\{\frac{x - vt\phi}{2(\alpha_x vt)^{\frac{1}{2}}}\right\}\right) + \left(\exp\left\{\frac{x}{2\alpha_x}[1+\phi]\right\} \times erfc\left\{\frac{x + vt\phi}{2(\alpha_x vt)^{\frac{1}{2}}}\right\}\right)$$

$$\phi = \left(1 + \frac{4k\alpha_x}{v}\right)^{1/2}$$

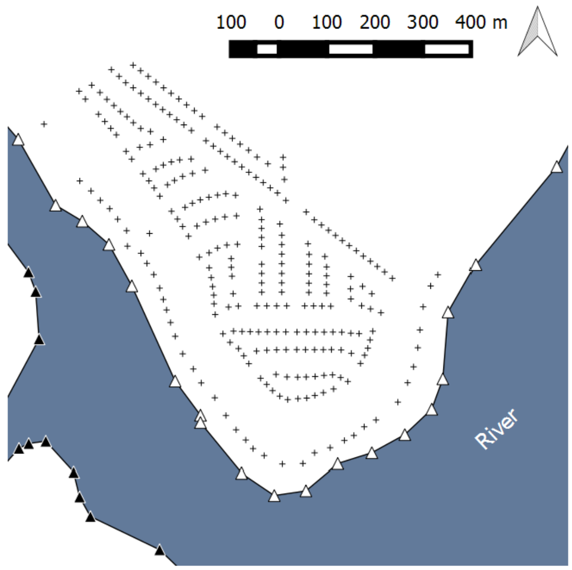

Figure 3: Example problem. + = Drainfields. River polygon vertices: ▲ = certain, Δ= uncertain.

Table 2 shows the explanation and values of the parameters used for C(x,t) and table 3 shows the uncertain object specifications. Each fuzzy point was sampled using the reduced transformation method at 11 α-cuts. For the probabilistic loop, 200 realizations were computed. Flow paths were calculated from each drainfield (hydraulic gradient obtained from an external model) using a script in QGIS. Eq. (1) was applied to each path and the load calculated using an estimate of cross-sectional area at the river boundary of 2 m^2 for each path.

Parameter	Meaning	Value
t	Time	99999 days
α_x	Longitudinal dispersivity	10.88 m
v	Groundwater velocity	0.9 m/day
k	1^{st} order decay coefficient	0.08 1/day
C_0	Initial concentration	40 mg/l
x	Position	L (flow path length. Varies from path to path.)

Table 3: Values for eq. 1

Object	Class	Determination of Uncertainty	Specification (all units are m)
River	Deformable object (probabilistic)	Estimate from source metadata	Each vertex: N(0, 0.612)
Drainfields	Vertex-defined object (fuzzy)	Estimate from expert judgement	Each point: trapezoidal fuzzy number. Core = [-2, 2], support = [-3.5, 3.5]

Table 2: Specification of uncertain objects

An interpretation of the results is as follows. Fig. 4a shows there is little probabilistic variability in the load estimate, indicated by the near vertical CDFs. On the other hand, uncertainty due to fuzziness appears to be large, indicated by the wide spread of the CDFs. Using fig. 4a, fig. 4b shows that an optimistic estimate ($\boldsymbol{\alpha} = \mathbf{1}$) of the 80% quantile of the load is between 160 and 200 mg/day. Visualizing the results in QGIS (fig. 5) shows that the path length (parameter 'x' in eq. 1) is mainly influenced by uncertainty in the river boundary while the point of intersection with the river depends mainly on the uncertain drainfield location.

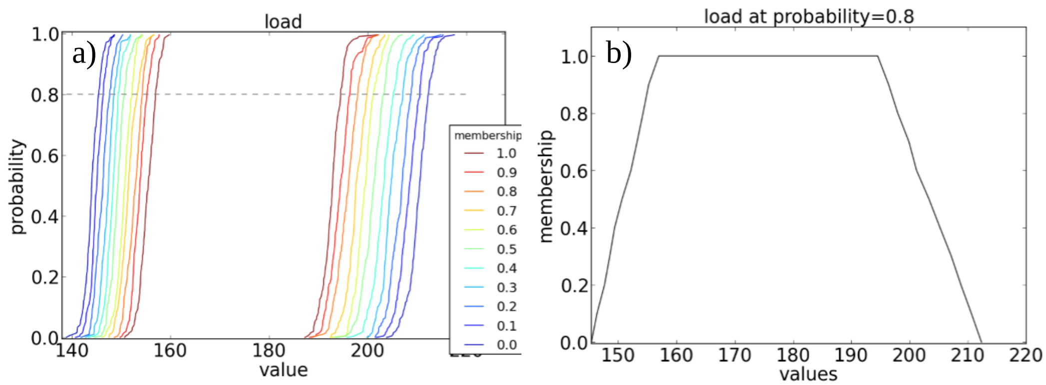

Figure 4: a) Uncertainty in the load estimate [mg/day]. b) Membership of the 80th quantile.

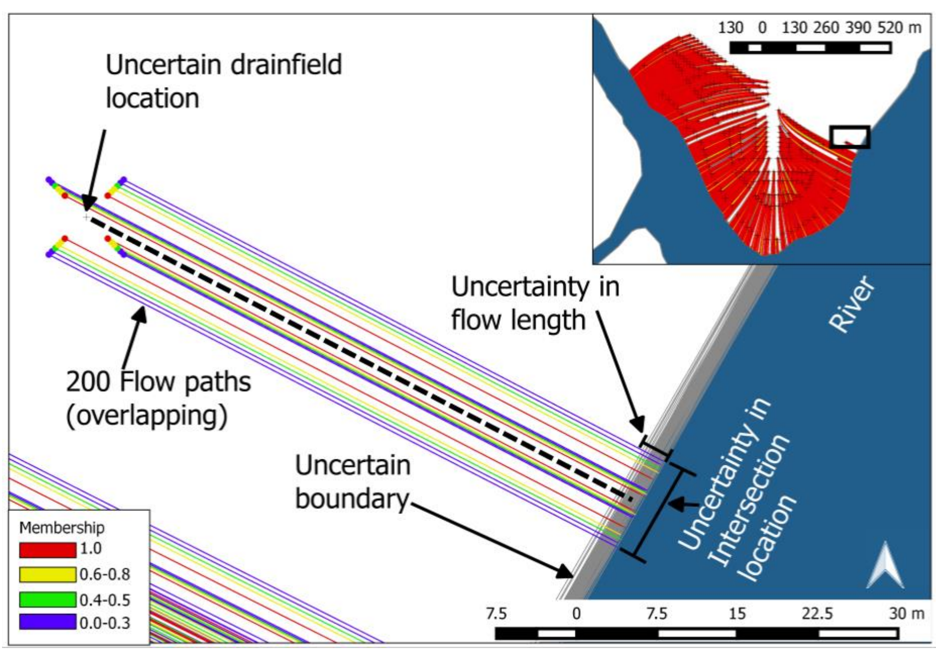

Figure 5: Effect of uncertainty on the flow path. Dashed line is the path without uncertainty. Note the 'X' shaped sampling of the drainfield location resulting from the reduced transformation method.

6. Conclusions

A tool suitable for analyzing GIS-based models with object-based positional uncertainty is presented. Compared to other tools, it allows for the treatment of positional and non-positional mixed uncertainties and treats models as black boxes, thereby increasing the tool's flexibility. A demonstration using a simple GIS model shows how the tool may be applied.

A drawback to the tool in comparison to others is the lack of a graphical interface. Such an interface, embedded within a GIS would ease the use of the software by allowing for such things as visual selection of uncertain shapes/vertices and allowing for an easier definition of correlation structures. Such an interface is planned. Also planned is the integration of a systematic sensitivity analysis methodology.

7. Acknowledgements

The author acknowledges the support from the Natural Science and Engineering Research Council of Canada postgraduate scholarship.

8. References

Aerts, J. C., Goodchild, M. F. & Heuvelink, G., 2003. Accounting for spatial uncertainty in optimization with spatial decision support systems. *Transactions in GIS*, 7(2), pp. 211-230.

Baudrit, C., Dubois, D. & Guyonnet, D., 2006. Joint propagation and exploitation of probabilistic and possibilistic information in risk assessment. *Fuzzy Systems, IEEE Transactions on*, 14(5), pp. 593-608.

Bear, J., 1972. *Dynamics of Fluids in Porous Media*. New York: American Elsevier Pub. Co.

Brimicombe, A., 1998. A fuzzy co-ordinate system for locational uncertainty in space and time. In *Innovations in GIS*, 5th ed., S. Carver, Ed. London: Taylor & Francis, 143-152.

Brown, J. D. & Heuvelink, G., 2007. The Data Uncertainty Engine (DUE): A software tool for assessing and simulating uncertain environmental variables. *Computers and Geosciences*, 33(2), pp. 172-190.

Guyonnet, D. et al., 2003. Hybrid approach for addressing uncertainty in risk assessments. *Journal of Environmental Engineering*, 129(1), pp. 68-78.

Heuvelink, G. B. M., Brown, J. D., and van Loon, E. E., 2007. A probabilistic framework for representing and simulating uncertain environmental variables. *International Journal of Geographical Information Science* 21, 5, 497–513

Hanss, M., 2002. The transformation method for the simulation and analysis of systems with uncertain parameters. *Fuzzy Sets and Systems*, 130(3), pp. 277-289.

Rios, J. F., 2014. Conceptualizing locational uncertainty in geographic objects: towards a more balanced approach. In: *Proceedings of the 11th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, July 8-11, East Lansing, Michigan, pp. 66-70.

Sadeghi, N., Fayek, A. R. & Pedrycz, W., 2010. Fuzzy Monte Carlo simulation and risk assessment in construction. *Computer-Aided Civil and Infrastructure Engineering*, 25(4), pp. 238-252.

Zhang, J. & Goodchild, M. F., 2002. *Uncertainty in geographical information*. Taylor & Francis.