Determining Uncertainty in Finnish Soil Data

Rangsima Sunila[1] and Jukka Krisp[2]

[1&2]Institute of Cartography and Geoinformatics, Department of Surveying

Helsinki University of Technology, P.O. Box 1200, FIN-02015, Espoo, Finland

Email: rangsima.sunila@hut.fi[1] and jukka.krisp@hut.fi[2]

ABSTRACT

Uncertainty issues in geographical information are widely researched. Uncertainty of information may lead to imprecision in analysis and decision-making. The Ministry of Agriculture and Forestry of Finland is funding a project to study uncertainty in geographical information of soil data sets. In order to study the issue in depth, imprecision of soil polygon boundaries was selected as a case study as their boundaries are not crisp in reality. In order to study benefits and the impacts of imprecise soil polygon maps, a survey of expert opinions on uncertainty in the information of soil map is conducted. In this paper, expert opinions about the imprecise soil maps will be explained and evaluated and this information will be used as support information in further research. The study shows the reasons why imprecise soil maps exist and possible weaknesses in the soil mapping process. The paper also explains the importance and advantages of awareness of soil map uncertainty information and how it is involved in decision analysis. The paper provides a basis for discussion of alternatives for future soil maps that contain uncertain information.

1. INTRODUCTION

Geographic information systems (GIS) are often used to aid the decision-making process. With the help of GIS, decision makers may make more accurate and informed decisions through the analysis and exploration of spatial data. Since there is no perfect information that is absolutely accurate in GIS, uncertainty in geographic information is a crucial issue. The impression of certainty usually conveyed by GIS is at odds with the uncertain nature of geographic information, a contradiction that has been acknowledged as an important research for nearly two decades (Duckham, 2002). As GIS analyses are often a basis for decision-making, ignoring uncertainty may at worst lead to wrong decisions and also the undermine the credibility of the system or operator (Fisher, 1996). Therefore, the users should be aware of the uncertainty of information in order to avoid ineffective decision analysis.

1.1 Background

Generally the Finnish soil data can be described as a continuum. It may be spread thinly over bedrock in some places or occurs thickly in others. Soil properties can vary significantly over the space of a few centimetres, due to a discontinuity in the parent material (e.g. edge of

sand deposited by an old delta). In other locations soil properties can have minor transitions over a distance of kilometres, where similar parent material, landscape, and climate combine to create a relatively homogeneous soil. Soil has an important three-dimensional aspect, and characteristics vary with its depth. Identifiable differences are used to divide the soil into layers or soil horizons. Soils are typically intermingled, due to local drainage effects and the distribution of the parent material. Soil characteristics may vary in a predictable way related to its landscape position. A substantial knowledge of topography and landscape effects can be combined with soil survey information to estimate locations of individual soils.

To study uncertainty in geographical information in Finland, research was conducted in 2003 to investigate the imprecision of soil polygon boundaries. Soil polygon boundaries were selected because in reality, they are not crisp and consequently they are a good example of presenting uncertain information. In Finland, soil mapping is conducted by manual interpretation, as it is the only feasible alternative since the country area is relatively large and the survey resources are rather limited. To define soil polygon boundaries, geologists or soil surveyors use aerial photos, geologic maps, topographic maps and knowledge of geomorphology. To classify soil types, drillings are made in order to take soil samples from 1 m. depth to examine their properties. Some samples may be taken to the soil laboratory for detailed testing (Sunila *et al*., 2004). As the country area is about 337, 000 square kilometres, it is not possible to have dense soil surveys and the field observations are, therefore, relatively sparse. As a result, dissimilar methods used by soil surveyors to partition landscapes can lead to different data, and hence decisions based on them could be intrinsically unreliable (Burrough and McDonnell, 1998). In order to study the benefits and impacts of imprecise soil polygon maps, a survey of expert opinions on uncertainty in the information of soil map is conducted.

1.2 The aims of the research

The research is interested in discovering how imprecision of soil polygon boundaries exists in soil databases, the expert opinions about imprecise soil polygon boundaries, the level of certainty of soil maps and the expert opinions considering users' point of views. The results of the research should provide wider visions of imprecision of soil maps concentrating on soil polygon boundaries. They will also give better ideas for future research in terms of development of soil information and mapping process and will help to provide alternative methods of presentation of future soil maps for specific uses.

2 METHODOLOGY

To collect expert opinions, questionnaire-based research is selected. Questionnaires are defensible and furthermore an effective method of collecting data from human participants.

2.1 The survey of expert opinions

The questionnaire is conducted in order to collect the knowledge from geologists to build up the knowledge-based system. The questionnaire forms were completed by five expert geologists from Geological Survey of Finland and the Finnish Defence Forces. The questions were formed in both open format and closed format. McNamara (1999) suggested that closed format questions offer many advantages in time and money. Also by restricting the answer set, it is easy to calculate percentages and other hard statistical data for the whole group or any subgroup of participants. He also pointed out that an obvious advantage of an

open format questionnaire is that the variety of responses should be wider and more accurately reflects the opinions of the respondents. The first section is closed format concentrating on soil maps in relation to imprecise soil polygon boundaries. The second section is open format so as not to restrict the opinions of respondents.

2.2 Closed format section

The experts were asked to indicate their opinions by indicating their agreements with the statements. The level of agreement is ranked by 1=Strongly Disagree, 2= Disagree, 3= Neither Agree nor Disagree, 4=Agree, 5=Strongly Agree. The statements are shown together with the results in table 1.

Table 1. The results of the closed questions in average.

Question	Result	
The soil polygon boundaries are imprecise because of generalisation method.	4	[Agree]
The soil polygon boundaries are imprecise because of ineffective equipment using in the survey.	2.4	[Disagree]
The soil polygon boundaries are imprecise because of lacking of supportive information.	3	[Neither agree nor disagree]
The geologists use their experiences to judge how to draw the lines of soil polygon boundaries.	5	[Strongly agree]
Different geologists give different opinions on how to define or draw the soil polygon boundaries.	3.6	[Agree]
Out of date information from topographic maps and aerial photos lead to imprecision of soil polygon boundaries.	2.4	[Disagree]
Do you think users are well aware of imprecision of soil maps when using them for spatial analyses?	3.4	[Neither agree nor disagree]
In your opinion, imprecision of soil maps may lead to wrong decision-making.	2.6	[Neither agree nor disagree]
Do you agree that creating a new soil data layer that presents imprecision of soil maps will be useful data for users?	3.8	[Agree]
Do you agree that imprecision of soil maps should be included in soil databases?	4	[Agree]
Please indicate level of certainty of soil maps (in general). (Range 0 to 1)	0.84	

An addition, question asks whether or not the imprecise soil maps are useful and all of the experts concur and give the same positive answer.

2.3 Open format section

The open format is used for giving open opinions and examples of real cases. The questions and answers in this section are:

2.3.1 *Who are the main users of imprecise soil maps?*

Most of the experts provide quite similar answers that visualization of imprecision is useful for all users. However, the main users are anyone who refers to soil maps when making other data from imageries, satellites and aerial photos. Other users can be the persons who use soil maps as a reference or data source in multidata analyses like terrain mobility analysis.

2.3.2 *What benefits will users get from imprecise soil maps?*

The benefit of imprecise soil maps is the knowledge of data reliability for data makers and end users so that they can handle the quality of work and data utility, especially in the decision making process.

2.3.3 *Example cases of using imprecise soil maps*

The expert' opinions reveal that maps presenting imprecise soil polygon boundaries will probably be most useful when used with data classifications for vegetation purposes, for instance, agriculture, forestry and natural protection. This information is useful for effective planning, especially when the soil types change gradually in both horizontal and vertical directions.

2.3.4 *Additional Information*

The experts also provide additional statements that the whole problem of imprecision may depend on the means like paper maps and geographical data and how the mapping information is given to the users. The more relevant question is how the users understand the data for example, how it has been collected, the issues of fieldwork versus interpretation, what the assumptions are, understanding the mapping in depth and the mapping classifications. The experts also state that people tend to think that mapping data is very precise so presenting the imprecise data will make users aware of the reliability and quality of data when using the data for analyses.

2.3.5 *Evaluation of the results*

The survey shows a relationship between the experience of experts and level of certainty of soil maps. Figure 1 indicates that with an increase of work experience the certainty about the information in soil maps decrease. A number of reasons can influence this as technical developments and the use of complex analysis methods continue to change the way of producing soil maps. Figure 2 illustrates the results of the experts individually in order to compare and evaluate their opinions. All experts agree that the generalisation method creates imprecise soil polygon boundaries. It is interesting to mention that in the least experienced geologist's opinion the ineffective use of equipment in the soil survey does not affect the precision of soil polygon boundaries, but the most experienced geologists strongly agree that ineffective equipment is one of the reasons for imprecision. This is because the development and new technology of soil survey equipment are nowadays modern and more reliable than in the past but still imprecision is evident. Another interesting point is that all experts strongly agree on geologists using their experience to judge how to define the boundaries of soil polygons. This issue is evaluated later on and it was found out that it is one of the most important reasons for identifying imprecision of soil polygon boundaries. The out of date information is not considered a big issue for medium-level experienced experts, as information like aerial photos and topographic data are likely to be up to date and not difficult to obtain. The experts give varied levels of agreement when it comes to the users' point of

view questions because some experts are main users of soil maps and some are only map producers.

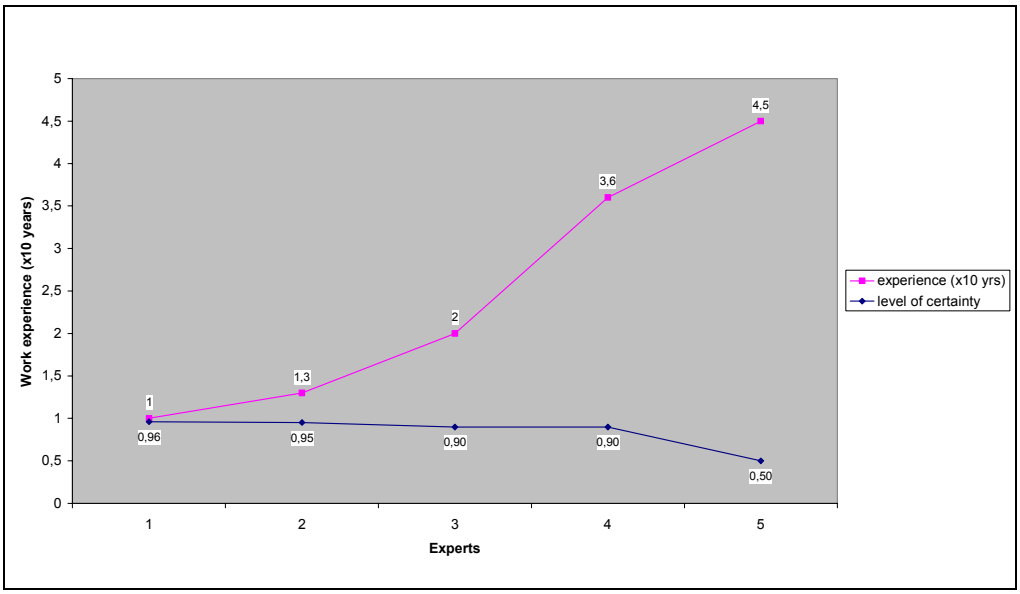

Figure 1. Experience (x10 years) of experts and Level of certainty of soil maps.

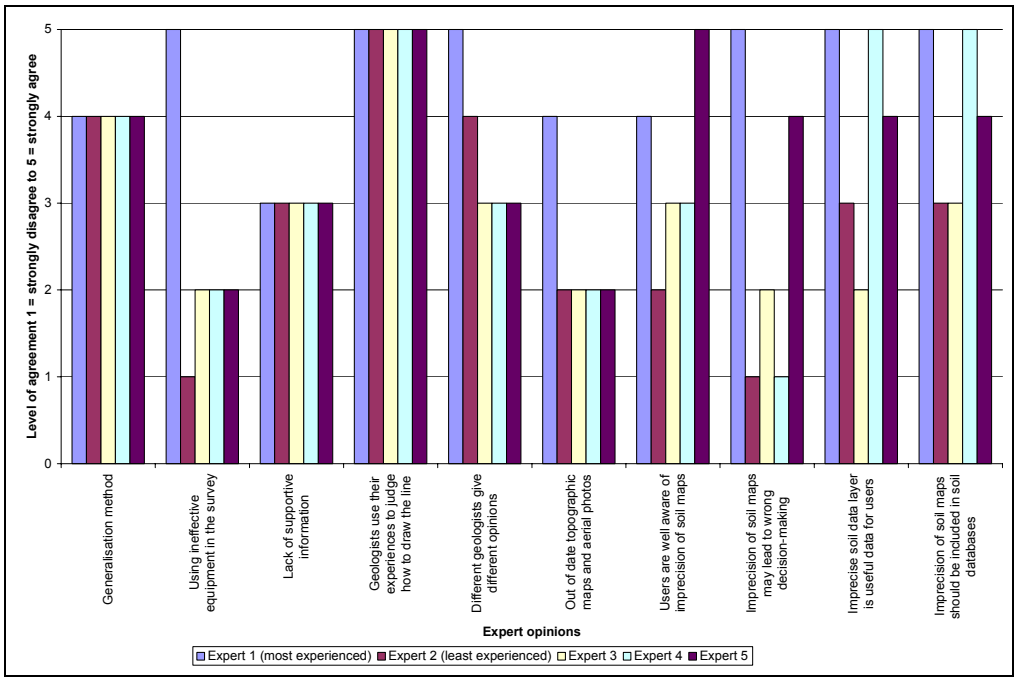

Figure 2. Expert opinions.

3. DISCUSSION AND CONCLUSION

Soil data is an important factor in the decision making process. Soil datasets are used for a great variety of applications. These include land use planning (like agriculture and forestry), agricultural production, pesticide registration, engineering (mainly foundation stability), site selection for radio towers (as electrical conductivity can affect radio wave

transmission), weather modelling and soil degradation risk assessment (water / wind erosion). Different applications have different needs concerning the accuracy of the datasets. Therefore an indication of the precision is an important part and the information can be integrated as proposed in several metadata standards, like ISO19115 (2003) or the CSDGM (Content Standard for Digital Geospatial Metadata 1998).

Analysis and decisions are false if they are based on incorrect or inaccurate information. In other words the data that is used in the analytical process needs to be as accurate and correct as required for the specific field of application (different decisions require particular levels of detail). Decision makers need to be aware of the quality of the data that they use, and how it will affect the result of their analysis. In addition to the way the soil data has been recorded and assembled, there is a growing need for documentation regarding the quality of the information, especially when the data is being accumulated in various national, regional and local spatial databases. Generally soil databases are becoming accessible to larger numbers of people so the data available may be used for a purpose other than the one for which it was originally collected. Decisions based on this data inherent "assumptions" that the ways of obtaining the information have "imprinted" on the dataset. Consequently it may be unsuitable for a specific analysis. There is a danger that the users of the "publicly available" spatial soil dataset pay less attention to the details of data, the process of data collection and processes that historically were a basis for understanding data quality.

The results present that imprecision of soil polygon boundaries is an important issue and it should be taken into consideration when using it for spatial analyses. The results also give better ideas for future research in terms of development of soil information and mapping process. The presentation of imprecise soil data on map sheet and database is an interesting topic to be researched in the near future.

5 ACKNOWLEDGEMENTS

The authors would like to thank the Ministry of Agriculture and Forestry of Finland for funding the project, Professor Kirsi Virrantaus for her comments and support, Dr. Jukka-Pekka Palmu, Ms. Maija Haavisto-Hyvärinen, Mr. Jukka Ojalainen, Ms. Pirkko Oranne from Geological Survey of Finland and Mr. Rami Immonen from the Finnish Defence Forces for their expert opinions and GISIDEAS 2004 conference committee for giving a chance to present the paper.

6 REFERENCES

Burrough, P. and McDonnell, R.A., 1998. *Principles of Geographical Information Systems*, Oxford University Press, Oxford, 265-291.

Duckham, M., 2002. Uncertainty and geographic information: *computational and critical convergence*. In: Representation in a digital geography. John Wiley, New York.

Fisher, P.F., 1999. *Models of uncertainty in spatial data*. In: Longley *et al,* Geographical Information Systems, Vol. 1, Principles and Technical Issues. John Wiley, New York.

McNamara, C., 1999. *Basics of Questionnaires*, Used by The Management Assistance Programme for Nonprofits.

Sunila, R., Laine, E., and Kremenova, O., 2004. *Fuzzy model and Kriging for Imprecise Soil Polygon Boundaries,* Geoinformatics 2004 proceedings, Sweden, and Advances in Mineral Resources Management and Environmental Geotechnology proceedings, Greece.