Proceeding of the 10th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences
Florianopolis-SC, Brazil, July 10-13, 2012.

# Expert elicitation for the variogram

*Phuong N. Truong[1], Gerard B. M. Heuvelink[1] and John Paul Gosling[2]*

[1] Department of Environmental Sciences, Wageningen University, P.O. Box 47, 6700 AA, Wageningen, The Netherlands
phuong.truong@wur.nl, gerard.heuvelink@wur.nl
[2] Department of Statistics, University of Leeds, Leeds, LS2 9JT, UK
j.p.gosling@leeds.ac.uk

## Abstract

*The variogram is the keystone of Kriging. Much research has been devoted to modelling the variogram from structural analysis of observations. However, there are many instances when the variogram is needed and there are no observations to base on. This can be due to budget constraints, physical or temporal obstacles, or a demand for a priori variogram in Bayesian geostatistics and spatial sampling design. Using expert knowledge that is elicited with a formal statistical expert elicitation procedure is suggested. In this study, we designed a protocol for a well-structured elicitation procedure to elicit the variogram from expert knowledge. The protocol has two main stages: elicitation of the marginal probability distribution and elicitation of the variogram. We built a web-based tool to facilitate the procedure. A case study was carried out to quantify spatial accuracy of a legacy map of the volumetric soil water content at field capacity. Expert elicitation returns a variogram that has a Matérn model shape with nugget = 0.45, partial sill = 54.57, range = 25410 meters and kappa = 0.40. The results show that the online elicitation tool satisfactorily captured expert opinion although, currently, it is only a prototype that needs to be further developed.*

**Keywords**: Spatial variability, Variogram, Expert knowledge, Soil property map**,** Uncertainty quantification.

## 1. Introduction

Geostatistics provides an efficient mechanism, i.e. Kriging, for interpolating spatial data and assessing the interpolation accuracy. The variogram is the keystone of Kriging. The spatial variability of environmental variables is characterized by the variogram (Journel and Huijbregts, 1978). Much research has been devoted to modelling the variogram from a structural analysis of available observations. The common Matheron method of moments (Matheron, 1963) is commonly used to model the variogram from the empirical or sample variogram that is estimated from available observations. However, there are many instances when the variogram of a spatial variable is needed and there are no observations to base them on. This can be due to budget constraints, physical or temporal obstacles, or a demand for an a priori variogram in Bayesian geostatistics and spatial sampling design optimization, to name a few. In such cases, experts can be an important source of information

because experts can be very knowledgeable about the spatial variability of some variable.

Formal statistical expert elicitation (SEE) has gradually become a mature research field and has proven to be able to extract reliable information from experts to form a sound scientific database. SEE theory has been developed based on the combination of mathematical probability theory, cognitive psychology and decision theory (O'Hagan et al., 2006). From a statistical perspective, SEE is the process of formulating a person's knowledge and beliefs about uncertain quantities into (joint) probability distributions (Garthwaite et al., 2005).

In this work, we attempted to derive reliable information from experts to model the variogram of a spatial variable using formal expert elicitation techniques. We developed a novel and generic SEE protocol and built a web-based tool to facilitate SEE for the variogram of an isotropic second-order stationary normal or lognormal spatial random field. To evaluate the tool, we carried out a case study where the experts used the tool to quantify the spatial accuracy of the mapped volumetric soil water content at field capacity (SWFC) for the East Anglian Chalk area of the United Kingdom (UK). In section 2, we present the elicitation protocol and the web-based tool. In section 3, preliminary results from the case study are presented. Discussion and conclusions are presented in section 4.

## 2. Elicitation protocol for the variogram and web-based tool

We designed a protocol of a well-structured elicitation procedure to elicit the variogram of a stationary spatial random field. There are two main stages in the elicitation procedure: elicitation of the marginal probability distribution (mpdf) and elicitation of the variogram. The elicitation of the mpdf was based on the bisectional elicitation method (Oakley, 2010) where the three quartiles (the median, the lower and upper quartiles) are elicited from experts through formal questioning. These elicited values are fitted to an underlying normal or lognormal distribution using ordinary least squares. The decision between normality and lognormality was based on a diagnosis of the Bowley coefficient of skewness (Bowley, 1920).

Elicitation of the variogram focused on the difference of the variable's values at two locations at a given lag apart in the case where the marginal probability distribution is normal, or the ratio of the variable's values at two locations a given lag apart in the case of a log-normal marginal probability distribution. The lags are estimated based on the size of the study area. The ratio of the largest and smallest lag is at least 100 which ensures that the smallest lag is small compared to the extent of the study area. The elicited values from the second stage are used to calculate semivariances at lags and a variogram model is fitted to these judgements. The elicitation process is an iterative process where feedbacks are given at each elicitation stage and experts can revise their judgments. An equal weight pooling method (O'Hagan et al., 2006).is used to pool all expert judgements.

Based on the developed protocol, we built a transparent, web-based tool to facilitate the expert elicitation procedure. The tool has three main components: (1) a web interface for elicitation and feedback, (2) statistical calculation and mathematical pooling to merge the opinions of multiple experts and (3) database management. The web interface was built around Symfony, which is an Open Source PHP Web application development framework (Symfony, 2012). The experts can inter-

act with the tool online through the URL: http://www.variogramelicitation.org using their username and password. Figure 1 gives an example screenshot of the first page of the web interface. Database persistence is maintained by the open source database MySQL. Statistical computation components were built around the R programming language: R-gstat package (Pebesma, 2004), R-Automap package (Hiemstra et al., 2009) and partly reused the fitting probability density algorithm from the SHELF framework (Oakley and O'Hagan, 2010).

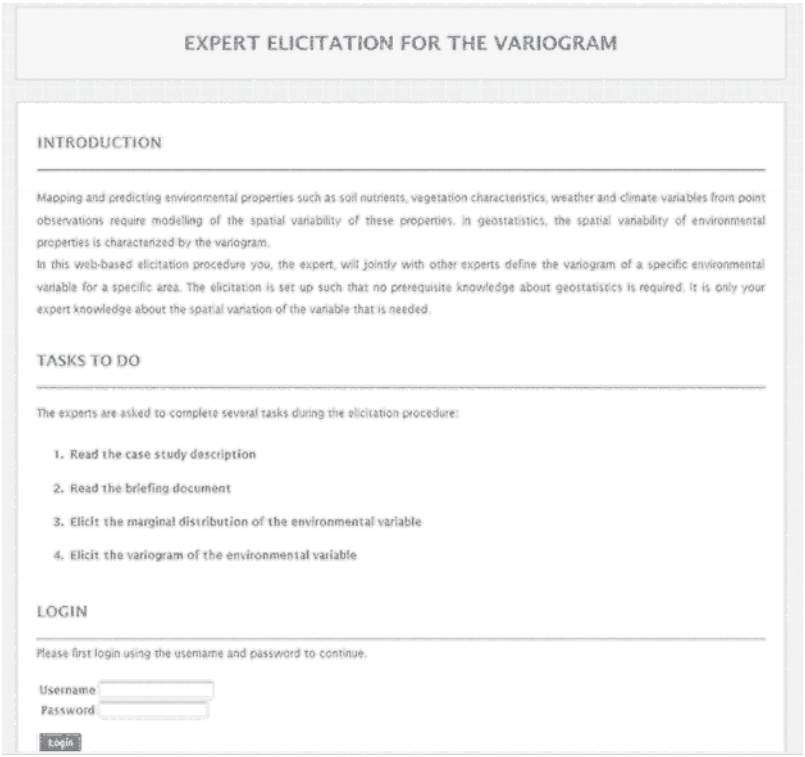


Figure 1: Screenshot of first page of the web interface.

## 3. Case study results

The case study was set up to quantify the accuracy of the SWFC map of East Anglia Chalk area. The study area is located at the East Anglia in the southeast of the UK. The SWFC map was produced from the subset of the National Soil Map of England and Wales (NATMAP, http://www.landis.org.uk) for the East Anglian Chalk area by assigning to each soil mapping unit (polygon) a volumetric water content value (%) at 10 kPa tension for the top soil horizon. Six soil experts from the UK and the Netherlands were invited to join the case study. The target quantity being elicited is the error of the SWFC in the map. The error is defined as the difference between the true value of the SWFC and that given by the map at a location. We assumed that the error is an isotropic second-order stationary spatial ran-

dom variable of which its spatial distribution is characterized by a constant mean and the variogram.

The final results from the SEE of the experts' judgements of the spatial variation of the error are given as the pooled mpdf and the pooled variogram. The pooled variogram reflects the combination of all experts' knowledge, with each expert's opinion having an equal impact on the captured knowledge. The mean of the pooled mpdf represents the combination of multiple experts' knowledge about the global mean of the error of the SWFC map over the study area. The elicited variogram characterises the spatial correlation structure of the error of the SWFC map.

Figure 2 shows the results of eliciting the mpdf of the error of the SWFC map at a random location in the study area. The pooled mpdf that is an probabilistic average of all mpdf's by individual expert is a normal distribution with a fitted mean of 2.1% and a standard deviation of 9.5%. Figure 3 shows the elicited variogram of the error of the SWFC map. The pooled variogram model is the Matérn model with the parameters: nugget = 0.45, partial sill = 54.57, range = 25410 meters, kappa = 0.40.

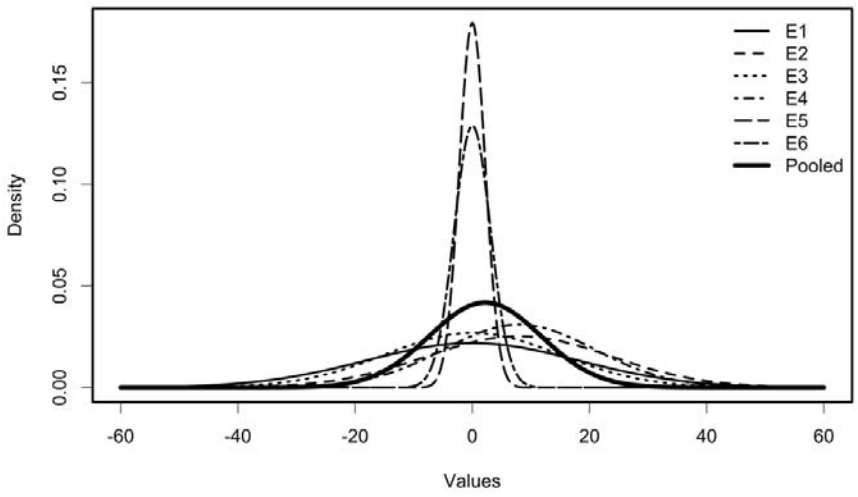


Figure 2: Fitted probability density functions to individual expert' judgements and to the pooled outcome for the marginal probability distribution.

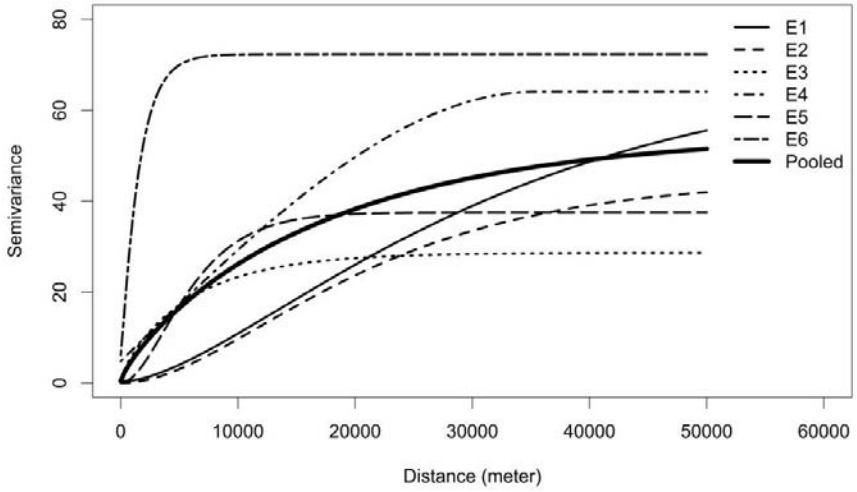


Figure 3: Fitted variogram models to every expert' judgments and pooled outcome.

## 3. Discussion and Conclusions

The pooled results capture the experts' knowledge about the spatial distribution of the error of the legacy SWFC map. According to the pooled outcome, there can be a systematic error on the map with a positive deviation of 2.1 % that is small compared to its random error. The variogram measures the variation in the error over the study area.

The elicitation protocol was designed such that it can minimize the possible causes of bias in expert judgement. This results in reliable elicitation outcome.

The results from the case study show that the online elicitation tool satisfactorily captures expert opinion although, currently, it is only a prototype that needs to be further developed.

## Acknowledgments

The study was funded by the UncertWeb (http://www.uncertweb.org) EC FP7 project (248488). We are grateful to the experts for their participation in the elicitation procedure. We thank Mathew Williams, Hannes Reuter, Jill Johnson and Sarah Knight for help and advice.

## References

Bowley, A.L. (1920), *Elements of Statistics*, London, U.K, 463p.

Garthwaite, P. H., Kadane, J. B., O'Hagan, A. (2005), "Statistical methods for eliciting probability distributions". *Journal of the American Statistical Association*, Vol. 100: 680-700.

Hiemstra, P.H., Pebesma, E.J., Twenhöfel, C.J.W., Heuvelink, G.B.M. (2009), "Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network". *Computers & Geosciences,* Vol. 35: 1711-1721.

Journel, A. G., Huijbregts, C. J. (1978), *Mining geostatistics*, Academic Press, London, U.K., 600p.

Matheron, G. (1963), "Principles of geostatistics". *Economic Geology*, Vol. 58: 1246-1266.

O'Hagan, A., Buck, C., Daneshkhah, A., Eiser, J., Garthwaite, P., Jenkinson, D., Oakley, J., Rakow, T. (2006), *Uncertain Judgements : Eliciting Experts' Probabilities*, John Wiley & Sons, Ltd., Chichester, 321p.

Oakley, J. E. (2010), "Eliciting univariate probability distributions". *In*: Böcker, K. (Ed.): *Rethinking Risk Measurement and Reporting: Volume I*, Chapter 6, Risk Books, London.

Oakley, J. E., O'Hagan, A. (2010), "SHELF: the Sheffield Elicitation Framework (version 2.0)". School of Mathematics and Statistics, University of Sheffield, Sheffield, http://www.tonyohagan.co.uk/shelf/, [Accessed February 28, 2012].

Pebesma, E.J. (2004), "Multivariable geostatistics in S: the gstat package". *Computers & Geosciences*, Vol. 30: 683-691.

Symphony (2012), *A gentle introduction to symphony: Build professional websites faster with PHP and symphony, symfony 1.3 & 1.4, License: GFDL, Version: gentle-introduction-1.4-en-2012-02-26*, SensioLabs, 397p.