Expert knowledge in geostatistical inference and prediction

Phuong Ngoc Truong

Thesis committee

Promotor

Prof. Dr P.C. de Ruiter
Personal chair at Biometris
Wageningen University

Co-promotor

Dr G.B.M. Heuvelink
Associate professor, Soil Geography and Landscape Group
Wageningen University

Other members

Dr S. de Bruin, Wageningen University
Prof. Dr C. Kroeze, Wageningen University
Prof. Dr J. Oakley, University of Sheffield, United Kingdom
Prof. Dr F.D. van der Meer, University of Twente, Enschede

This research was conducted under the auspices of the C.T. de Wit Graduate School Production Ecology & Resource Conservation.

Expert knowledge in geostatistical inference and prediction

Phuong Ngoc Truong

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr M.J. Kropff,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Monday 30 June 2014
at 1:30 p.m. in the Aula.

Phuong Ngoc Truong

Expert knowledge in geostatistical inference and prediction,

160 pages.

PhD thesis, Wageningen University, Wageningen, NL (2014)

With references, with summaries in Dutch, Vietnamese and English

ISBN: 978-94-6257-028-3

Contents

Chapter 1

General introduction

1

1.1. Geostatistics and expert knowledge

1.1.1. Geostatistics

Geostatistics is originally the study of the spatial distribution of natural resources in mining and geology (Matheron, 1963), where the statistical modelling of spatial dependence is used for inference of spatial structure and for spatial prediction at unobserved locations from observations (i.e. kriging prediction). These are the two main purposes of geostatistical analysis. It has also founded an important statistical method for uncertainty quantification of mapping spatial phenomena through the kriging variance.

A geostatistical model represents a spatial phenomenon as a regionalised variable whose mean may depend on explanatory environmental variables and whose spatial dependence is modelled by the variogram. When the variation of the spatial phenomenon shows an obvious trend, the geostatistical model is the sum of the spatial trend (i.e. spatial mean) that models the large scale variation and the zero-mean random residual. The spatial trend can be modelled as a (unknown) constant or a linear function of the covariates (i.e. the predictive secondary variables). The zero-mean random residual models the small scale variation (including small-scale, microscale and white-noise variation) and is characterised by the variogram (Cressie, 1991, Section 3.1). The variogram is a mathematical function that plots the semivariance against separation distance, where the semivariance equals half the variance of the differences of the variable at two locations a certain distance apart (Armstrong, 1998; Oliver and Webster, 2014). Geostatistical data have a continuous variation in geographical space, but can be discontinuous in attribute space (Cressie, 1991, Section 1.2.1; Schabenberger and Gotway, 2005, Section 1.2.1).

In this dissertation, geostatistical inference refers to estimation of the variogram parameters and/or the parameters that define the relationship between the spatial variables of interest and the covariates that define the trend. Geostatistical prediction refers to prediction of the spatial variables at unobserved locations. In general, the geostatistical prediction or kriging prediction at an unobserved location is a weighted avarage of the surrounding observations (Cressie, 1990; Stein, 1999). In case there is a spatial trend, the kriging prediction equals the sum of the trend and the weighted average of the trend residuals at the surrounding observed locations. The magnitude of the kriging weights are controlled by the spatial dependence between the unobserved locations and the surrounding observations, and they guarantee unbiasedness and minimise the kriging variance (i.e., provide the 'best' predictor).

Geostatistics has been applied in various disciplines of the Earth and environmental sciences, such as geology, hydrology, soil science, ecology, forestry and climatology. Kriging tools can produce exhaustive maps of the spatial phenomena that are required in many practical cases. For example, in precision agriculture, maps of crop nutrients such as potassium, phosphorus or nitrogen over fields are required for efficient soil fertilising strategies. In environmental pollution monitoring, maps of soil pollutions or ambient air pollutions are needed to assess public exposure to these pollutions that can help prevent public health problems. Recently, mapping of spatial variation of epidemics using geostatistics proves useful in accessing the relationship between disease incidence and environmental, social-demographic factors. There are many more examples from the geostatistical literature that clearly show the scientific and societal value of geostatistics.

1.1.2. The challenges of optimal use of data for geostatistical inference and prediction

Geostatistical inference and prediction are fundamentally dependent on observations (i.e. field measured data). The quantity and quality of the observations determine the quality of the geostatistical inference and prediction. When a spatial variable continuously varies over a certain spatial domain, the observations can be sampled everywhere within this spatial domain for spatial inference. However, very often, the observations used in geostatistics are only a limited sample of locations (point support) or areas (block support). Moreover, the number of sampling locations is often constrained by experimental difficulties, geographical obstacles, budget restrictions, time and environmental impact of sampling. These constraints may lead to unsatisfactory sampling density and unrepresentativeness of the observations that can hinder the effective use of geostatistics in spatial inference and prediction.

Geostatisticians are well aware of the possible drawbacks of using limited observations in geostatistical inference and prediction. Considerable research has studied the magnitude of this effect on the accuracy of geostatistical inference and prediction (e.g. McBratney and Webster, 1983; Webster and Oliver, 1992; Frogbrook, 1999; Oliver and Webster, 2014). Meanwhile, various methods have been developed to increase the accuracy of geostatistical inference and prediction. For example, optimum sampling schemes are recommended to reduce kriging variance (McBratney et al., 1981; van Groenigen et al., 1999; Brus and Heuvelink, 2007; Vasát et al., 2010) and to best use the observations for variogram inference (Warrick and Myers, 1987; Lark, 2002; de Gruijter et al., 2006, Chapter 9; Webster and Lark, 2013, Chapter 9). More efficient statistical algorithms for variogram estimation are recommended such as maximum likelihood (Pardo-Igúzquiza, 1998; Pardo-Igúzquiza et al., 2009) or residual maximum likelihood (REML) (Pardo-Igúzquiza, 1997; Kerry and Oliver, 2007). These inference methods require fewer observations than the method-of-moments (Matheron, 1963) to reach a comparable estimation accuracy.

Geostatisticians have also incorporated different types of data and information in geostatistical models to improve the mapping accuracy. The terms prior information, soft data, secondary information or ancillary data have been used in the geostatistical literature to indicate data or information other than direct (error-free) measurements of the target variable itself (Stein, 1994; Goovaerts, 1997, Chapter 6; Kerry and Oliver, 2003; Oliver et al., 2010b). The use of extra data and information is certainly valuable in many geostatistical applications. For example, optimal sampling design needs prior information about the spatial variation in a certain area before measurements are collected (Kerry and Oliver, 2004). Spatially exhaustive ancillary data can be used to define the trend of the geostatistical model. For example, the correlation between temperature and elevation furnishes the use of elevation as an external drift variable to make a better prediction (Hudson and Wackernagel, 1994). Kriging tools such as regression kriging, cokriging, Bayesian kriging and indicator (co) kriging have been used to incorporate these different sources of data and information (Hoef and Cressie, 1993; Hudson and Wackernagel, 1994; Goovaerts, 1997, Chapter 6; Oberthür et al., 1999; Pardo-Iguzquiza, 1999).

1.1.3. The concept of expert knowledge in geostatistics

While ancillary data and information are often used as an additional source of data and information in modern geostatistics, expert knowledge about spatial phenomena is a huge pool of knowledge that is relatively unnoticed. A study of Stein (1994) gives an early overview of the use of ancillary information as prior information (i.e. information obtained before any field measurement is taken) for spatial sampling and interpolation, and expert knowledge has been mentioned as one option. A large body of expert knowledge about spatial phenomena has been accumulated in various disciplines of the Earth and environmental sciences.

Aforementioned, geostatistics characterises spatial variables by the spatial trend and the variogram. In case of multiple variables, there are also cross-variograms that define the cross-correlations between the target variable and the covariates. Hence, expert knowledge for geostatistical research is essentially about these trends and spatial correlations. For example, experienced pedologists have good knowledge about the relationships between soils and environmental variables such as soil forming factors (parent material, climate, vegetation, rainfall, etc.). A study of Walter et al. (2006) gives an overview of the origin of expert knowledge in pedology. Expert knowledge has been implicitly and informally used in geostatistics to eye-fit the variogram (Webster and Oliver, 2007) and to best guess or 'guesstimate' the magnitude of spatial correlations (Kros et al., 1999). However, systematic use of expert knowledge has been found in only a few studies, e.g. to classify topsoil texture classes of rice fields to be used as soft-information in mapping soil texture (Oberthür et al., 1999), to guide spatial sampling design according to expert judgements about the spatial variation of a certain variable in a certain area (van Groenigen et al., 1999), to supplement sparse observations for spatial inference (Lele and Das, 2000), or to specify the spatial relationship between the target variable and the covariates to develop optimum models for spatial prediction (Lark et al., 2007).

All studies that make use of or refer to expert knowledge show a great potential of using expert knowledge in geostatistics. But these studies also show that expert knowledge has not been formally and systematically used in geostatistical modelling and mapping. The use of expert knowledge has also been criticised or undervalued because expert knowledge that is transformed into expert judgement is considered subjective and intractable (Tversky and Kahneman, 1974; Meyer and Booker, 2001, Chapter 2; O'Hagan et al., 2006, Chapter 3; McKenzie et al., 2008). This might be due to a lack of an efficient and reliable tool to extract knowledge from experts. In all previous studies that use expert knowledge, the description of how expert knowledge is elicited is overlooked.

1.2. Statistical expert elicitation for spatial phenomena

Several common expressions are often encountered in the statistical expert elicitation literature and also in this dissertation: expert, expert knowledge, expert judgement or expert opinion, and expert data. An expert is a person who has qualified knowledge on a subject matter (e.g. scientist, professional or experienced practitioner). Expert knowledge is qualified knowledge that can be expressed in either qualitative or quantitative statements. Expert knowledge is extracted into expert judgement or expert opinion (e.g. a meteorologist's estimate of the difference in average temperature in 2013 between Amsterdam, The Netherlands and Ohio, The United States, an economist's quantification of the unemployment rate in 2014 in The United Kingdom, etc.). There is no distinction between these two terms. Expert data in this dissertation refers to quantitative expert judgements that are used for spatial inference and prediction.

The main scientific objective of statistical expert elicitation research is to provide statistical techniques and formal procedures for eliciting expert judgements about uncertain quantities in a transparent and reliable way. From a statistical perspective, statistical expert elicitation is a systematic process of formulating expert knowledge about uncertain quantities as (joint) probability distributions (Garthwaite et al., 2005). Because statistical expert elicitation is a systematic process, it involves several stages: problem definition, expert recruitment and training, primary elicitation, feedback and revision, documentation and reporting (O'Hagan et al., 2006; Choy et al., 2009; Knol et al., 2010; Kuhnert et al., 2010). Depending on the purpose of the research, statistical expert elicitation techniques for one expert or multiple experts can be applied (O'Hagan et al., 2006). In case of multiple experts, an additional step is required to reach consensus among experts (French, 2011). Statistical expert elicitation is an appropriate approach to capture expert knowledge about the regionalised variables that represent spatial phenomena in geostatistics. To my knowledge, statistical expert elicitation has never been used to elicit expert knowledge to model spatial phenomena in geostatistics. Given its scientific objective and the current advance in statistical expert elicitation research, I assert that expert knowledge can be elicited and used in a responsible and defensible way for geostatistical inference and prediction.

1.3. Research objectives

The research has two main objectives. The first is to identify gaps in geostatistical data and accordingly, to identify the use of expert knowledge in geostatistical inference and prediction. The second is to investigate how to elicit expert knowledge and incorporate expert knowledge in geostatistical models for spatial inference and prediction.

1.4. Research questions and dissertation outline

1.4.1. Main research questions

My research addresses two main research questions, which correspond to the two main research objectives:

1. What is the role of expert knowledge in geostatistical inference and prediction?

2. How to elicit and incoporate expert knowledge in geostatistical inference and prediction?

In order to answer these two questions, I first list all detailed research questions in Section 1.4.2. These need to be answered first. Each of the next four chapters (Chapters 2 to 5) addresses one group of these detailed research questions. The last chapter of this dissertation provides the answers to the two main research questions.

1.4.2. Detailed research questions

The four groups of the detailed research questions that are answered in Chapters 2 to 5 are:

1. How to apply statistical expert elicitation to elicit the variogram from multiple experts' knowledge?

The variogram is the keystone of geostatistics. Almost half of the effort in geostatistical research is spent on estimation of the variogram. Practically, all applied research in geostatistics makes use of the variogram. Chapter 2 of this dissertation gives justification for the demand of the variogram when observations have not been collected yet. Also in this chapter, I show how statistical expert elicitation techniques were applied to elicit from multiple experts probabilistic judgements that can be used to estimate the variogram. The intention of applying statistical expert elicitation techniques for the development of an elicitation tool for the variogram leads to the following research questions:

1.1. Which measure to infer the variogram can be elicited from experts?

1.2. Which statistical expert elicitation technique can be applied to elicit from experts the selected measure?

1.3. How to combine multiple expert judgements?

1.4. Is developing an online statistical expert elicitation tool an effective approach?

2. How to use multiple experts' knowledge to quantify spatial uncertainty?

One important reason for using kriging in spatial mapping is that it provides uncertainty quantification of the spatial prediction. But it is fair to say that kriging is not used in all spatial mapping exercises. For instance, many soil maps have been derived by other ways (e.g. using pedotransfer function, aerial photographs, manual drawing of vegetation maps, etc.) that do not quantify the map uncertainty. However, users of these maps need to know how accurate they are. I tackled this issue in Chapter 3, where I applied the tool for expert elicitation of the variogram developed in Chapter 2. Question 1.4 is again addressed in Chapter 3, together with the following research question:

2.1. How to apply the web-based expert elicitation tool for the variogram to extract multiple experts' knowledge for spatial uncertainty quantification of soil property maps?

3. How to use expert judgements to solve the ill-posed problem in spatial disaggregation using area-to-point kriging?

An important research topic in geostatistics is the change of support problem. Here, the (spatial) support refers to the area or volume over which a measurement or a prediction is made. Geostatistics may be confronted with the problem of spatial disaggregation when the support of the observations is larger than that of the predictions (e.g. using remote sensing imagery or choropleth maps as observations for mapping the continuous variation over a spatial domain). In Chapter 4, this problem is addressed in detail, by answering the following research questions:

3.1. Why is the nugget parameter of the point support variogram often ignored in variogram deconvolution?

3.2. How to incorporate expert judgements in block support data to infer the point support variogram model?

3.3. How to quantify uncertainty propagation from expert judgements (parameter uncertainty) and model uncertainty propagation to spatial disaggregation?

4. How to incorporate expert judgements as observations in geostatistical inference and prediction?

Finally, I address a very conventional issue in geostatistics, which I have also discussed in Section 1.1. This is that in many geostatistical analyses, there is a lack of observations. Chapter 5 addresses the use of expert knowledge as inaccurate observations for geostatistical inference and prediction. The research questions addressed in Chapter 5 are:

4.1. How to measure bias and imprecision of expert probabilistic judgements on the value of a spatial variable at specific locations in comparison to measured data?

4.2. How to incorporate expert judgements as observations to characterise spatial variation using the variogram?

4.3. Which kriging method can be used to incorporate expert data in spatial prediction?

1.5. Scope and expected contributions of the dissertation

1.5.1. Scope of the dissertation

Mapping spatial variation of natural phenomena using expert knowledge is the main focus of my research; the spatio-temporal aspect of natural phenomena in geostatistics is not touched. Four illustrative examples and case studies presented in Chapters 2 to 5 are:

1. mapping of air temperature over The Netherlands;

2. mapping spatial uncertainty of soil water content at field capacity of the East Anglian Chalk area, The United Kingdom;

3. mapping air temperature over the Gelderland province, The Netherlands using remote sensing imagery;

4. mapping a biological soil quality indicator (i.e. nematode structure index) of a study area in the Malpiebeemden nature reserve in the south of The Netherlands.

All experts who were involved in this research are scientists (i.e. professors and senior researchers at universities and research institutes) from different research fields, such as soil science, hydrology and meteorology.

In the next four chapters, expert knowledge was always elicited in probabilistic form (i.e. quartiles of probability distributions). A web-based framework for expert elicitation was employed to build the elicitation tools. A model-based perspective in geostatistics (Diggle and Ribeiro, 2007) was taken as a foundation to develop the models to incorporate expert knowledge in geostatistical inference and prediction.

1.5.2. Expected contributions

The introduction chapter gives an overview and justification of using expert knowledge in geostatistical research and the opportunity to enhance the use of expert knowledge. In each of the next four chapters (Chapters 2 to 5), the specific answers to every detailed research questions provide a view of the use of expert knowledge in the four main focuses of geostatistical research: variogram estimation, spatial uncertainty quantification, spatial disaggregation and spatial interpolation - kriging. The solutions for the elicitation approaches and incorporation methods of expert knowledge in these geostatistical research focuses are provided. Chapter 6 concludes the dissertation by addressing the two main research questions that form the backbone of this dissertation. The general discussion also presents my personal reflections on what I have done and what can be done in the future to advance this research topic. This dissertation as a whole may contribute to the optimum use of data and information, both derived from measurements and from experts, for geostatistical inference and prediction. It may help advance the understanding of the Earth surface and subsurface spatial phenomena.

Chapter 2

Web-based tool for expert elicitation of the variogram

Based on: Truong, P.N., Heuvelink, G.B.M., Gosling, J.P., 2013. Computers & Geosciences 51, 390-399.

2.1. Introduction

Geostatistical interpolation of environmental variables from georeferenced observations requires modelling of spatial variability of these variables. In geostatistics, the spatial variability of environmental variables is characterized by the variogram (Journel and Huijbregts, 1978; Goovaerts, 1997; Chilès and Delfiner, 1999; Webster and Oliver, 2007). Theory about the variogram and kriging is well-described in the geostatistical literature. We only recall that the variogram is commonly modelled from the empirical or sample variogram that is estimated from available observations using the common Matheron method-of-moments (Matheron, 1963). A dominant factor that controls the accuracy of the variogram estimate is the number of observations. Webster and Oliver (2007) recommend using at least 100-150 observations for estimating the isotropic variogram and at least 250 observations for the anisotropic variogram.

Collecting and analysing sufficient data for variogram estimation are often expensive and time-consuming. There have been attempts to increase the accuracy of the variogram estimate for a given number of observations by using different statistical inference methods. For example, Pardo-Igúzquiza (1997, 1998) and Marchant and Lark (2007) use maximum likelihood or REML estimation that requires fewer observations. Kerry and Oliver (2007) confirm that maximum likelihood and REML provide an alternative to the method-of-moments when there are fewer than 100 observations. Cui et al. (1995) and Pardo-Igúzquiza (1999) use Bayesian statistical inference for estimation of the variogram parameters and their uncertainty by combining hard measurements with soft data from available prior information.

Environmental scientists are increasingly aware of the use of prior information of spatial variation in cost-effective sampling design for both variogram estimation (Cui et al., 1995; Lark, 2002; Kerry and Oliver, 2007) and optimum spatial interpolation (McBratney and Pringle, 1999; Kerry and Oliver 2003, 2004; Brus and Heuvelink, 2007). In addition, Bayesian inference of environmental variables making use of prior information of spatial variation has also become popular in mapping spatial variables with small samples, e.g. mapping hydrodynamic variables or petroleum reservoirs (Cui et al., 1995; Pardo-Igúzquiza, 1999; Diggle and Ribeiro, 2002, 2007). Previous research on the use of objective prior information for variogram estimation is the use of a (average) variogram derived from a similar study area (Cui et al., 1995; McBratney and Pringle, 1999; Kerry and Oliver, 2004) or using a variogram derived from ancillary data (Kerry and Oliver, 2003). These approaches rely on the similarity of spatial variation between similar areas and situations, which may not always be realistic or available. Alternatively, the value of subjective prior knowledge when available data are scarce or unreliable has been acknowledged recently in landscape ecology, geosciences and geographical research (Denham and Mengersen, 2007; James et al., 2010; Curtis, 2012; Perera et al., 2012a).

There are obvious demands for prior information about the spatial variation of environmental variables in geostatistical inference. As mentioned before, it is required to guide optimum sampling designs for costly measured and analysed variables. The increasing use of Bayesian geostatistics requires an informative prior to gain more information for inference when data are limited due to budget constraints or physical and temporal obstacles (Pardo-Igúzquiza, 1999; Diggle and Ribeiro, 2002; Curtis, 2012). In such cases, experts can be an important source of information because experts can be very knowledgeable about the spatial variability of a variable of interest. Expert knowledge is also important when no data are available to predict the future variation in spatial pattern (e.g. patterns of temperature or ozone concentration over a region ahead of time). We therefore suggest that consulting experts may be sensible to get the a priori variogram. The question is how the variogram can be derived from expert knowledge in a responsible way. In previous research (Kros et al., 1999), the spatial correlations of continuous variables are simply 'guestimated' by deriving a spatial correlation structure from direct consultation of experts. In this study, we argue that this approach is deficient and instead requires the application of formal rules from statistical expert elicitation.

From a statistical perspective, statistical expert elicitation is the process of formulating a person's knowledge and beliefs about uncertain quantities into (joint) probability distributions (Garthwaite et al., 2005). This person must have qualified knowledge about some aspects of the problem that the analysts want to elicit (Meyer and Booker, 2001; Garthwaite et al., 2005). Examples of typical cases that need expert assessment are estimation of new, rare, complex or poorly understood phenomena, future forecasts for particular events, interpretation of existing data, group decision making or extracting the current state of knowledge about certain phenomena (Meyer and Booker, 2001). The ultimate purpose of statistical expert elicitation is to reliably and consistently encode a person's knowledge or belief about an uncertain variable as a probability distribution (in general, expert elicitation may not necessarily need to encode expert knowledge using a probability distribution).

2

Formal statistical expert elicitation procedure involves a systematic process with several stages (O'Hagan et al., 2006; Choy et al., 2009; Knol et al., 2010; Kuhnert et al., 2010). The first stage is to set up the problem: starting with identifying the target variables, next preparing background or briefing document, then recruiting experts. Before conducting the elicitation, experts should be motivated and trained through a dry-run. Execution of the elicitation process can be done in a workshop of a group of experts with support of computer software and must be facilitated by the analysts. It can also be an individual elicitation by means of face-to-face interviews, online or telecom interviews. The analysts play an important role in this stage: they have the responsibility of facilitating, designing or choosing elicitation protocols and supporting tools. Expert judgement is encoded into probability distributions by either parametric or nonparametric fitting (Garthwaite et al., 2005). The next stage is giving feedback to experts, commonly in graphical forms and letting experts revise their judgements if needed. Elicited information from multiple experts can be combined by a mathematical pooling approach (O'Hagan et al., 2006). Bringing experts together in group elicitation is another way of obtaining consensual judgments, in this case using the so-called behavioural approach (O'Hagan et al., 2006). Heuristics and biases in expert cognition may result in inaccurate probability judgements (Kynn, 2008). The structured elicitation protocol has been designed in an attempt to minimize all contaminations to the process of eliciting reliable expert judgement.

Statistical expert elicitation functions as a statistical tool to extract knowledge from experts about real-world phenomena. In practice, statistical expert elicitation procedure needs computer assistance to effectively, conveniently and routinely capture and encode expert judgement (O'Hagan, 1998). In response to this, an increasing number of software and web-based tools have been built. Examples of web-based tools for the elicitation of univariate discrete and continuous probability distributions of uncertain variables are the MATCH Uncertainty elicitation tool (Morris et al., 2014) and The Elicitator (UncertWeb - The Elicitator[1], assessed 29/02/2012); examples of software are SHELF (Oakley and O'Hagan, 2010) - the elicitation framework for single and multiple experts, Elicitator (James et al., 2010; Low-Choy, 2012) for elicitation of regression models in ecology, and ElicitN (Fisher et al., 2012) for elicitation of species richness.

[1] http://elicitator.uncertweb.org

However, so far statistical expert elicitation has not been used to characterise spatial variation and elicit the variogram from experts. In this chapter, we aimed at applying statistical expert elicitation to geostatistical research domain, particularly to elicit the variogram from expert knowledge. We developed a novel and generic statistical expert elicitation protocol and built a web-based tool to facilitate statistical expert elicitation for the variogram of an isotropic second-order stationary multivariate normal or lognormal spatial random field.

In Section 2.2, we present the statistical expert elicitation protocol. Section 2.3 describes the web-based tool, its architecture and functionality. In Section 2.4, we present the results from a simple case study to test the protocol and web-based tool. Finally, we discuss the adequacy, the functionality and potential for routine application of the tool and avenues for further research in Section 2.5.

2.2. Developing a statistical expert elicitation protocol

For the sake of coherence, we present the definition of the variogram and its characteristics under the second-order stationary assumption. From this definition, one of its estimators that forms the basis for the developed protocol is presented.

The variogram is defined as the variance of the first increment $[Z(\mathbf{s}_1) - Z(\mathbf{s}_2)]$ (Journel and Huijbregts, 1978) where Z is the random function that characterizes the environmental variable of interest at two locations defined by coordinate vectors $\mathbf{s}_1$ and $\mathbf{s}_2$ in geographical space: $2\gamma(\mathbf{s}_1, \mathbf{s}_2) = \mathrm{Var}[Z(\mathbf{s}_1) - Z(\mathbf{s}_2)]$.

Assuming that Z is an isotropic second-order stationary random function on the Euclidean plane, its first increment $[Z(\mathbf{s}_1) - Z(\mathbf{s}_2)]$ is a random variable that satisfies:

- The expectation is equal to zero: $\mathrm{E}[Z(\mathbf{s}_1) - Z(\mathbf{s}_2)] = 0$,

- The variance is finite and depends only on the Euclidean distance: $\mathrm{h} = |\mathbf{h}|$, with $\mathbf{h}$ the separation vector, not on the locations defined by the coordinate vectors $\mathbf{s}$ and $\mathbf{s}$+h: $2\gamma(\mathrm{h}) = \mathrm{Var}[Z(\mathbf{s}+\mathrm{h}) - Z(\mathbf{s})]$.

Dowd (1984) introduced a robust estimator of the variogram, which was derived from Cressie and Hawkins (1980), based on the median of the absolute value of the first increment:

$$2\hat{\gamma}(\mathrm{h}) = 2.198[\mathrm{Med}\,|\,Z(\mathbf{s}+\mathrm{h}) - Z(\mathbf{s})\,|\,]^2 \qquad (2.1)$$

Eq. 2.1 provides a mechanism for eliciting the variogram from expert knowledge through a probability elicitation of the median of the absolute first increment: $V(h) = |Z(\mathbf{s}+h) - Z(\mathbf{s})|$ for various discrete lags h. The rationale behind this is that psychological research and practical expert elicitation exercises have shown that the median is a quantity that experts can assess reasonably well, and that this results in the most precise and reliable outcomes (Peterson and Miller, 1964; Kadane and Wolfson, 1998; O'Hagan et al., 2006). Using this approach, the variogram can be inferred at multiple lags. The variogram model is then derived by fitting a valid variogram model through these estimates in the usual way. In addition, the marginal probability distribution function (mpdf) that together with the variogram defines the full probability distribution of the random process Z over a geographical plane is also elicited.

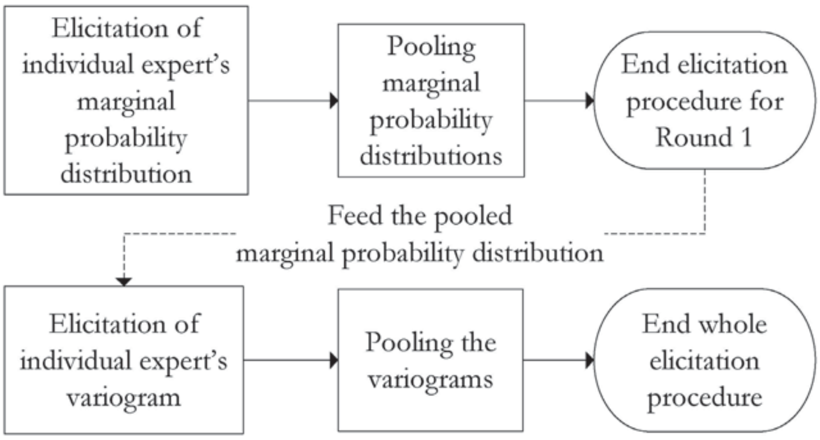

Figure 2.1: Components of elicitation protocol

Based on this concept, the statistical expert elicitation protocol was designed for multiple-expert elicitation with two main rounds. Round 1 is the elicitation for the mpdf and Round 2 is the elicitation for the variogram. Fig. 2.1 outlines the process of the whole elicitation procedure.

2.2.1. Expert elicitation for the marginal probability distribution

To elicit the mpdf, we used the bisection method for unbounded probability distributions (Oakley, 2010). This method was chosen because it only requires basic knowledge from experts on probability (Garthwaite and Dickey, 1985). Fig. 2.2 outlines the process of Round 1. In this method, the ordered range of possible values of the random variable $Z(\mathbf{s})$ is divided into areas of equal probability (note that the location **s** is immaterial because Z is second-order stationary). Each expert is first asked to judge on the minimum and the maximum and next the three quartiles sequentially. The questionnaires were slightly modified from those of the SHELF framework by using a numerical expression of probability. These questions are generic and that can be adapted to various problems. The answers are in the form of a quantitative response that is required to be precise to no more than three digits.

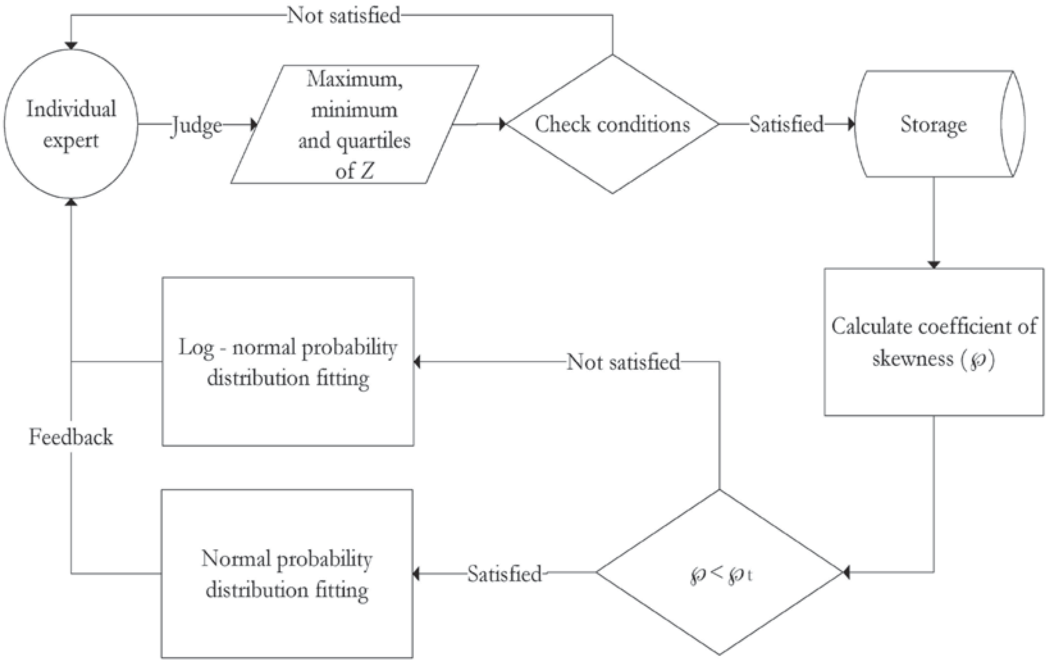

Figure 2.2: Round 1 of elicitation procedure

To avoid complexity, it is reasonable to assume that expert's belief about the mpdf of the random environmental variables can be represented as a normal or log-normal distribution. The decision between normality and lognormality is based on a diagnosis of the Bowley coefficient of skewness $\wp$ (Bowley, 1920):

$$\wp = (Z_{0.75} + Z_{0.25} - 2Z_{med})/(Z_{0.75} - Z_{0.25}) \tag{2.2}$$

where $Z_{0.25}$, Z_{med}, $Z_{0.75}$ are the lower quartile, the median and the upper quartile. We defined a threshold $\wp_t$ equal to 0.05. When $|\wp| \le \wp_t$, the distribution is assumed normally distributed. When $|\wp| > \wp_\tau$, the lognormal distribution is assumed.

Fitting the mpdf is a parametric fitting to an underlying nor-

2

mal or lognormal distribution using ordinary least squares. The mean (μ) and the variance (σ^2) of the mpdf are chosen by numerically minimizing: $[F(Z_{0.25}; \mu, \sigma^2) - 0.25]^2 + [F(Z_{med}; \mu, \sigma^2) - 0.5]^2 + [F(Z_{0.75}; \mu, \sigma^2) - 0.75]^2$, where F is the normal or lognormal cumulative distribution function.

The fitted mpdf is reported back to the expert. Each expert can reflect on their fitted mpdf and can revise their judgements until they are satisfied that their beliefs are correctly conveyed in the given feedback. Because the mpdf of the random function influences the variogram elicitation round, experts must reach a consensus about the mpdf before proceeding to the next round. Section 2.2.3 details how consensus amongst experts can be obtained.

2.2.2. Expert elicitation for the variogram

To model the variogram function, the variogram values for various lags need to be estimated. For kriging, modelling the variogram at small lags is more important than at larger lags because the nearer locations give more weight in the kriging prediction (Myers, 1991; Webster and Oliver, 1992, 2007). Choosing more small lags is therefore preferred for elicitation. The lags are defined by first establishing the maximum distance, which is defined as half of the diagonal (D) of the research area: $D = \text{sqrt}[(\mathbf{x}_{max} - \mathbf{x}_{min})^2 + (\mathbf{y}_{max} - \mathbf{y}_{min})^2]$, where $\mathbf{x}_{max}$, $\mathbf{x}_{min}$, $\mathbf{y}_{max}$, $\mathbf{y}_{min}$ define the extent of the study area (Lark, 2002; Webster and Oliver, 2007). Smaller lags are sequentially defined as half of the immediately preceding larger lag after this larger lag is rounded to the nearest number of type $k \times 10^a$ with k=1, 2 or 5 and an integer *a* (e.g. if the initial distance is 5×10^3 then the next is 2×10^3, if it is 1×10^3 then the next is 5×10^2, etc.). We chose to elicit from experts the median of no more than seven lags because experience has shown that experts cannot give proper judgements for more than seven values in a single session (Meyer and Booker, 2001). Note also that 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.

Depending on the result of the first round, that is, whether the mpdf is a normal or lognormal distribution, the next step of eliciting the variogram will be different. Fig. 2.3 outlines the procedure of Round 2.

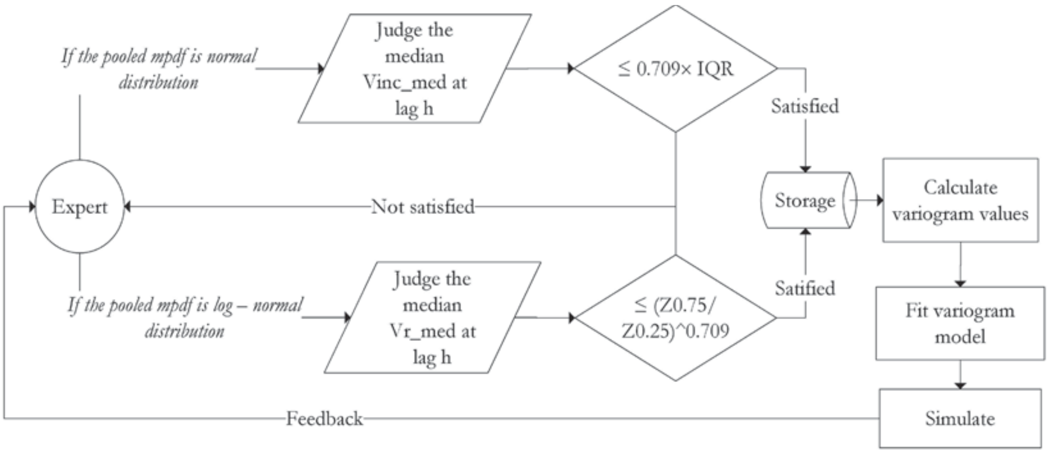

Figure 2.3: Round 2 of elicitation procedure

Variogram elicitation in case of the normal distribution

When the consensus mpdf is a normal distribution, Z is a second-order stationary Gaussian random field. In this case, each expert is asked to judge the medians of the absolute first increments: $V(h) = |Z(\mathbf{s} + h) - Z(\mathbf{s})|$, called V_{inc_med} for each of the seven lags. The medians elicited from each expert are used to calculate the variograms using Eq. 2.1. The variogram values cannot be larger than the variance of the pooled mpdf, which is the consensual mpdf from the Round 1 (Barnes, 1991). Thereby, it is easy to derive that the medians judged by experts must satisfy the following condition: $V_{inc\_med} \leq 0.709(Z_{0.75} - Z_{0.25})$. This condition is checked during Round 2.

The procedure continues with fitting a variogram model using ordinary least squares to the variogram values for the seven lags. The fitted variogram model is used to sample from the spatial distribution using unconditional sequential Gaussian simulation (Goovaerts, 1997; Pebesma, 2004) along an arbitrary transect within the study area. The variation in simulated values along the transect is shown to the experts. Note that, experts can only see the outcomes from their own judgements. Several simulations are generated and experts can toggle between these to get an impression of the whole range of possible realities. Experts can reconsider whether the spatial structure shown along the transect conveys what they think it should be. If not, they can revise their judgements about the medians at lags and the variogram elicitation is reiterated. Note that at this stage, they can no longer change the judged values of the mpdf because they already reached a consensual mpdf after finishing the Round 1.

2

2

Variogram elicitation in case of the lognormal distribution

When the mpdf is lognormal, each expert is asked to judge the median of the absolute ratio of change V_r, called V_{r_med}, between two locations at distance h: $V_r = |Z(\mathbf{s}+h)/Z(\mathbf{s})|$, assuming that $|Z(\mathbf{s}+h)| \geq |Z(\mathbf{s})|$. The median of the log-transformed difference is calculated by: $\text{Med}\{|\log(Z(\mathbf{s}+h))-\log(Z(\mathbf{s}))|\} = \text{Med}(|\log(Z(\mathbf{s}+h)/Z(\mathbf{s}))|) = \log(\text{Med}|Z(\mathbf{s}+h)/Z(\mathbf{s})|)$. This is satisfied as always $|Z(\mathbf{s}+h)/Z(\mathbf{s})| \geq 1$. The check on the sill not increasing the variance yields the condition: $V_{r\_med} \leq (Z_{0.75}/Z_{0.25})^{0.709}$.

The variograms are estimated by Eq. 2.1 using the median of the absolute increment of log-transformed values. The remaining steps are the same as in the case of normal mpdf. Note that, because the simulated values in this case are taken from the logarithm of the random function, back-transformation is required before simulated values along a transect are shown to experts.

2.2.3. Pooling experts' judgements

The multiple judgments from experts are combined using the mathematical combination method, also known as opinion pooling (O'Hagan et al., 2006). We follow the linear opinion pooling method in which all experts' judgements are combined by applying an equal weighted average. Fig. 2.1 indicates that pooling is required both in the elicitation for the mpdf and for the variogram.

The pooling of the mpdf is done by applying probabilistic averaging of many quantiles (i.e. 500) generated from the fitted probability distributions of all experts. The minimum value from all experts is taken as the minimum, likewise for the maximum. The coefficient of skewness (Eq. 2.2) of the combined distribution is calculated again to diagnose normality. The unique average values are fitted again to the chosen probability distribution (i.e. the normal or lognormal). The consensual mpdf is reported back to each expert, giving them a chance to compare it with their own mpdf and revise their judgements. The process continues until all experts are satisfied with the final consensual result and stop changing their own judgements. At this point, the Round 1 is ended. In practice, it may be sensible to allow just one revision turn.

To pool the variogram, the medians elicited from all experts for the seven lags are averaged and a variogram is fitted as described in Section 2.2.2. The whole elicitation procedure is ended when all experts stop revising their own judged values for the medians.

2.3. Description of the web-based tool

General structure of the web-based tool has three main components: web interface, database management and statistical computation (Fig. 2.4). Their specifications and design are discussed in detail hereafter.

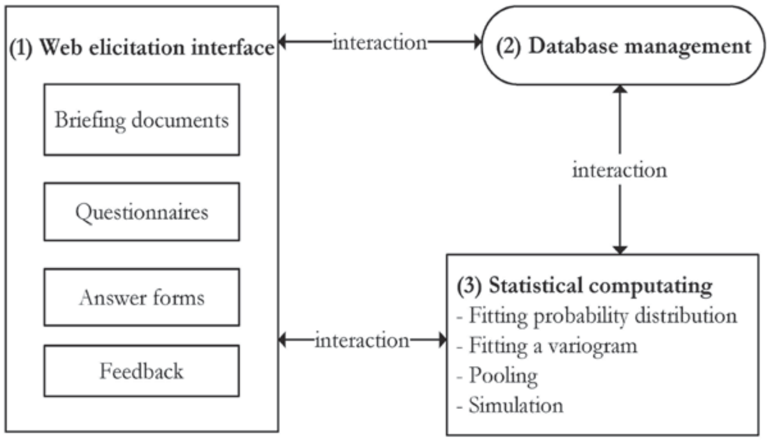

Figure 2.4: Three main components of web-based tool

2

2.3.1. Web elicitation interface

The web interface was built around Symfony, which is an Open Source PHP Web application development framework (Symfony, 2012). It facilitates interaction of individual expert with the tool to automatically proceed through the elicitation procedure. The web interface is mainly designed to present the briefing document, the question and answer forms and the graphical feedbacks. Google Maps is embedded to provide flexible view on geographical attributes of the study site. Simple buttons were designed in each webpage with different functions such as saving experts' judgments, executing statistical computation, rendering graphical feedbacks and navigating.

To access the tool, experts login at URL: http://www.variogramelicitation.org using given username and password. An example screenshot of the first page of the web interface is given in Fig. 2.5. Form validations based on the defined conditions were set up to handle submission of experts' judgements. Graphical feedbacks provide graphs of the fitted mpdf in the Round 1 and a transect of simulated values in the Round 2 both for individual and pooled outcomes. The graphical feedbacks are rendered using the Flot - Javascript plotting library[2] for jQuery (jQuery[3], accessed 28/02/2012).

2.3.2. Database

Relational database stores the final individually and consensually judged values and all fitted parameter of the mpdf and variogram. Database persistence is maintained by open source database MySQL. Symfony integrated with Doctrine provides an object oriented query language DQL (Doctrine Query Language[4]) to interact with the MySQL database to retrieve and store data.

2.3.3. Statistical computation

Several statistical functions are required to calculate the lags, to fit the mpdf and the variogram, to combine multiple experts' judgements and to simulate realisations of the random variable along a transect. The functions for fitting a probability density function were adopted from the SHELF framework. For fitting the variogram, we adapted the autofitVariogram function of the R package automap (Hiemstra et al., 2009). The chosen variogram models for fitting include the Nugget, Exponential, Spherical, Gaussian and Matérn models. No nested model is fitted, except for the combination of the Nugget model with each of the others. The initial parameters of the variogram model are defined as the default in the autofitVariogram function. Other statistical functions were originally built around the R package gstat (Pebesma, 2004). All statistical functions were assembled in a standardized format of an R package, named eeVariogram.

Executions of these statistical functions are initialized by experts after they have given their judgements. Experts' judgements are first stored in the database and then fed into the statistical functions. The fitting process is automatic so that experts cannot interfere. PHP executes and passes arguments to R scripts which invoke R functions from the eeVariogram package. The outputs are returned to PHP for rendering by Flot.

[2] http://www.flotcharts.org/
[3] http://jquery.com
[4] http://www.doctrine-project.org/

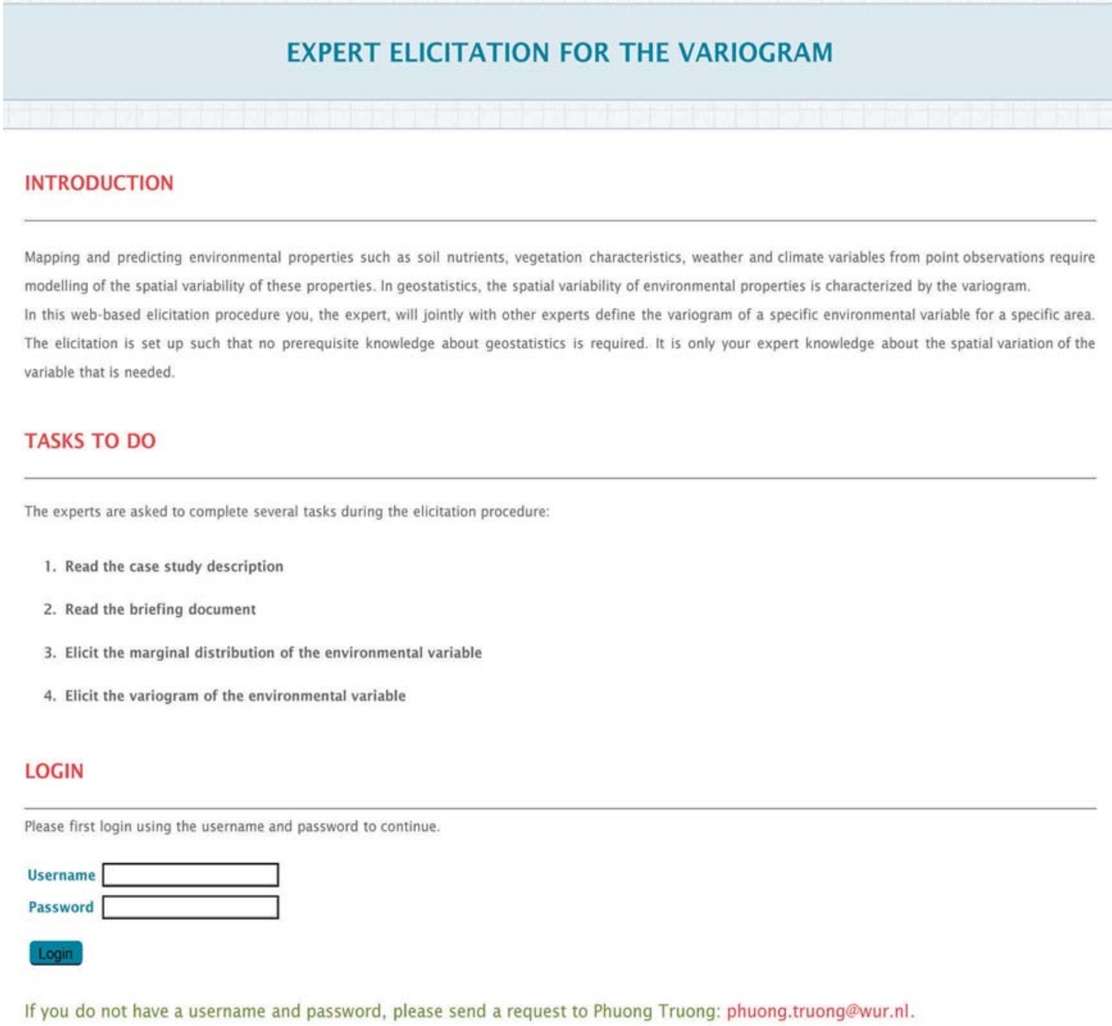

Figure 2.5: Screenshot of first page of web interface

2.4. Illustrative example

To test the adequacy of the protocol and the functionality of the web-based tool, we set up a simple case study on elicitation of the spatial variability of the maximum temperature over The Netherlands on April 1^{st}, 2020. Historical data from KNMI-Royal Netherlands Meteorological Institute were used to compare with the elicitation outcomes. The data are the measured maximum temperature on April 1^{st} from 1993 to 2012 at 35 stations over The Netherlands. Five partners from UncertWeb project[5] were invited to join the case study as experts. It should be noted that this simple example was only chosen to test the tool, and that the experts are not climatologists of The Netherlands. Hence, a variable was chosen that, with the right background

[5] http://www.uncertweb.org

information provided by the tool, each participant could form an opinion on.

The web-based tool started with information about the geographical attributes, a link to the KNMI website where the experts could obtain information about the weather of the Netherlands and a Google map of The Netherlands. Although the protocol requires little statistical knowledge from experts, a briefing document was designed for explanation of probabilistic summaries such as quartiles. In the briefing document, the causes of biased judgements including cognitive bias due to limitations in human information processing and motivational bias due to human subjectivity (Meyer et al., 1990) were also explained to point the experts to possibly major causes of bias in their judgements. The experts were asked to carefully read the introduction and briefing document first and then spend no more than 30 minutes to finish each of the two rounds. Figs. 2.6 and 2.7 show a screenshot of the question forms of Round 1 and 2.

We present the results from one expert as an example. The expert's judged values for the mpdf are a maximum of 30ºC, minimum 7ºC, lower quartile 12ºC, median 14ºC and upper quartile 18ºC. Fig. 2.8 is the feedback graph of the mpdf fitted to the judgements of this expert. The graph shows a normal distribution with lower quartile 11.7ºC, median 14.6ºC and upper quartile17.5ºC. Fig. 2.9 shows the feedback of the pooled mpdf from the five experts. Comparison with a histogram of data over the past twenty years at an arbitrary selected station shows a fair degree of agreement (Fig. 2.10). Apparently, the experts have a fair idea of the variations that occurred in the past and projected this to assess uncertainty about the maximum temperature on April 1st, 2020.

Because the pooled mpdf was normal, the quantity to be judged in the Round 2 was the V_{inc_med}. Fig. 2.11 shows the variograms computed from the elicited medians from all experts and the pooled medians at the seven lags (ranging from 2 to 200 km). Note that this figure was not presented to the experts because interpretation requires knowledge about geostatistics that was not presumed. The pooled variogram model is the Matérn model with nugget = 0.02ºC^2, partial sill = 4.56ºC^2, range = 27.6 km and smoothness (kappa) = 0.7.

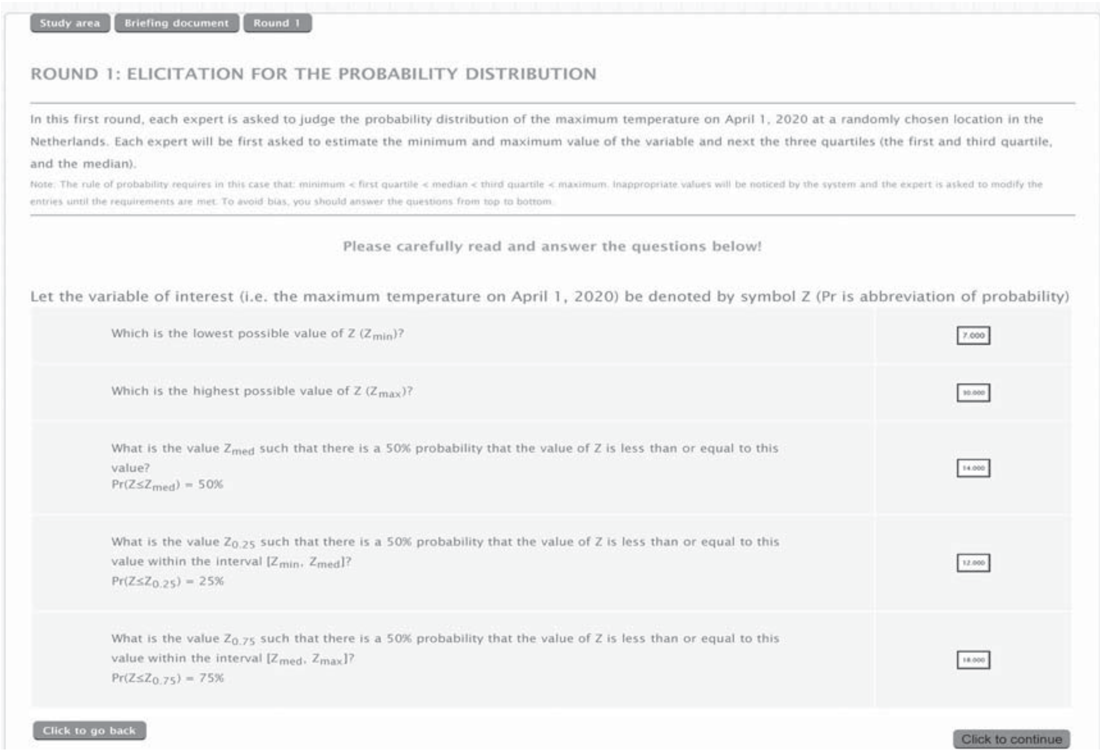

Figure 2.6: Screenshot of question form of Round 1

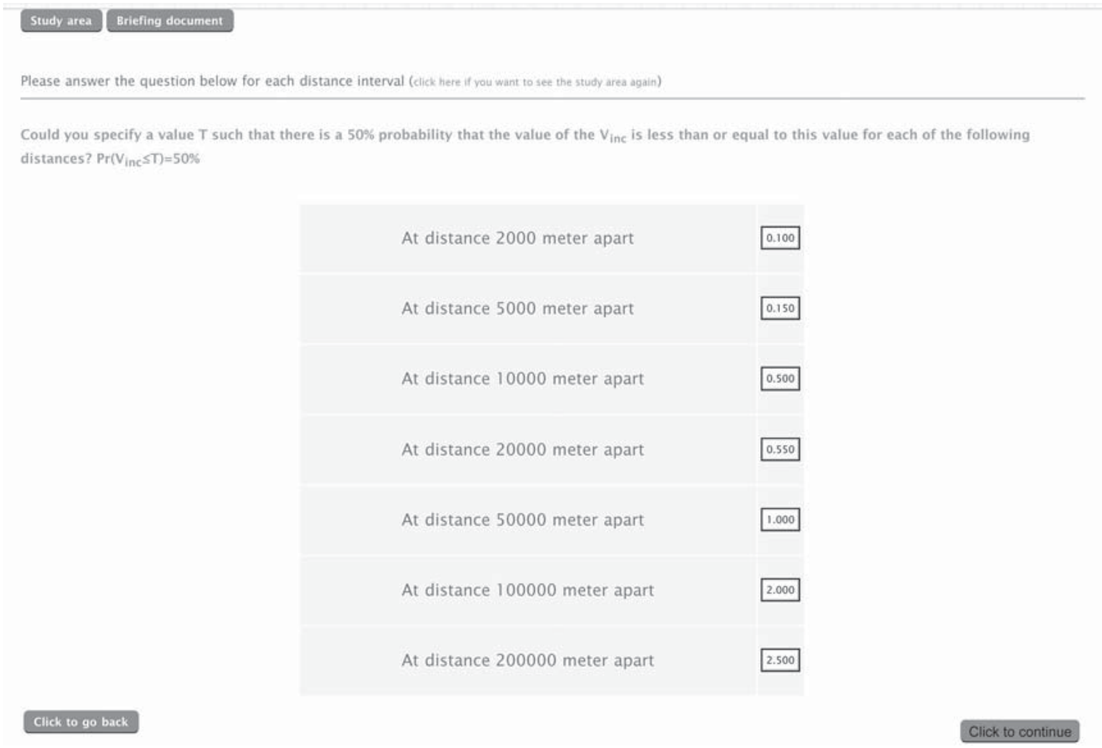

Figure 2.7: Screenshot of question form of Round 2

2

Fig. 2.12 shows an example of simulation transects that are the actual feedbacks to the experts. The pooled transect shows a substantial degree of short-distance variation; this is in agreement with the pooled variogram model (Fig. 2.11). The variogram that was derived from the data over the past twenty years at 35 stations was calculated by first pooling the twenty empirical variograms and next fitting a variogram model (Fig. 2.11). Although there are data from only 35 stations that make variogram estimation inaccurate, part of this inaccuracy is taken away by pooling over twenty years. By comparison with the variogram model from the data, the elicited variograms from the experts show that the experts tend to overestimate the spatial variability, especially at short distances. This difference may partly be explained by the fact that the experts are indeed not experts in climatology of the Netherlands, partly because the variogram was derived from data of only 35 stations, and partly because the expert elicitation addressed the future temperature.

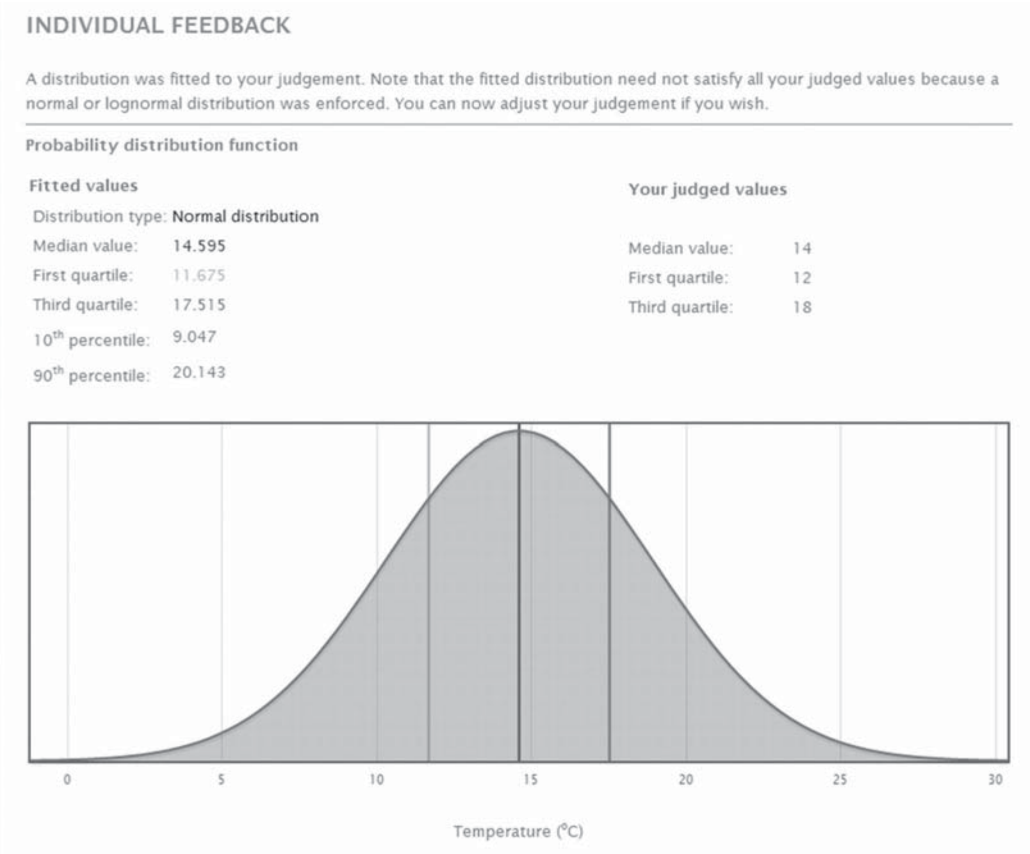

Figure 2.8: Screenshot of graphical feedback for individual expert's marginal probability distribution

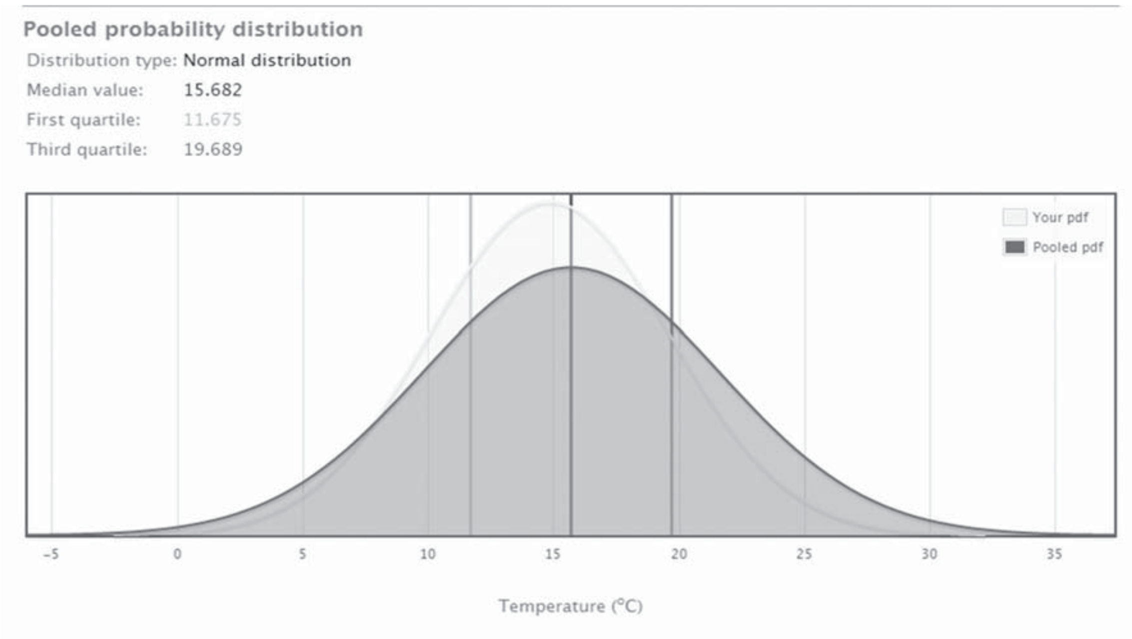

Figure 2.9: Screenshot of graphical feedback for pooled marginal probability distribution

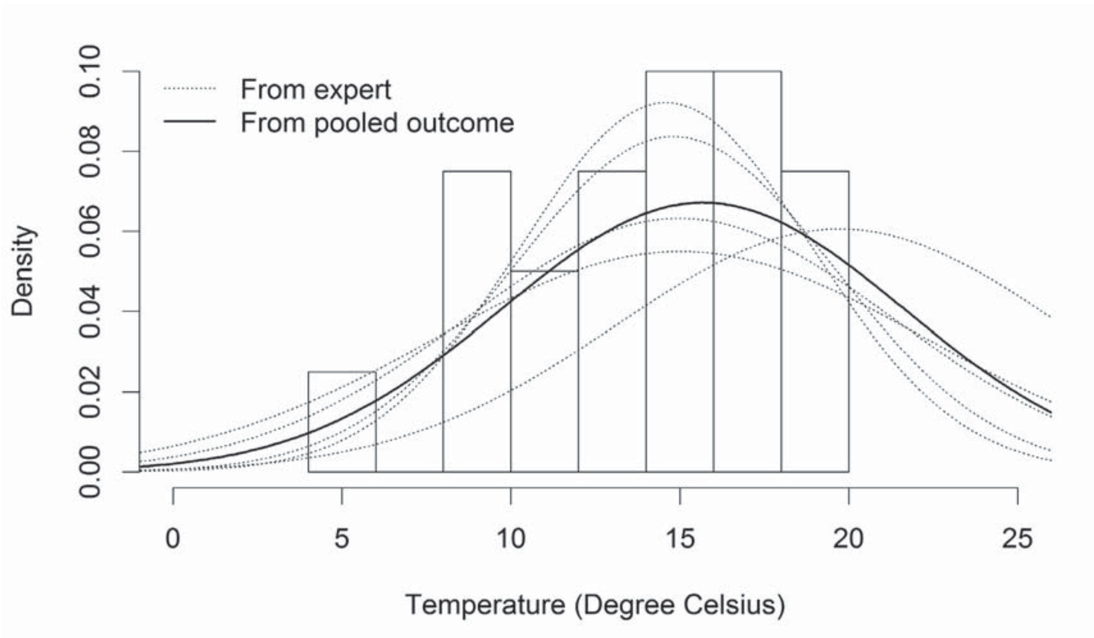

Figure 2.10: Histogram of maximum temperature from one station on April 1st over past twenty years compared with experts' (dotted lines) and pooled (solid line) marginal probability distribution function

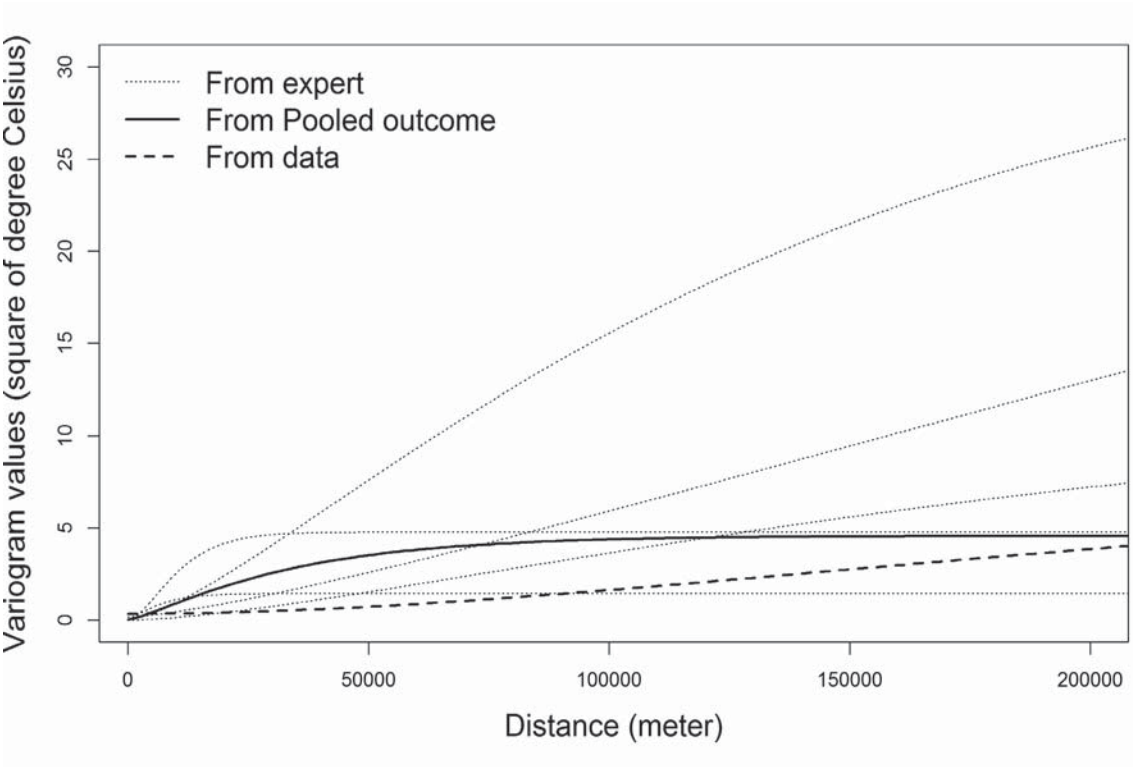

Figure 2.11: Elicited variogram models from experts (dotted lines) and from pooling (solid line). The dashed line represents the variogram model derived from historical data

2.5. Discussion and Conclusions

The results from the test case study show that we have fulfilled our objective of designing an elicitation protocol for the variogram. The variogram elicitation procedure works and the elicited variogram captures the experts' knowledge of spatial variability.

The web-based tool functions well and satisfies the intentions that it is simple in use, has an interactive interface, provides immediate graphical feedback, is remotely accessible, provides a database management and automatically performs the mathematical and statistical computations in the background. It is easily adaptable to different case studies. The protocol requires neither advanced knowledge of probability from experts nor experts to be geostatisticians. It was developed based on the recommended seven steps of designing an expert elicitation protocol and software (Choy et al., 2009; Knol et al., 2010). We used indirect encoding method (Choy et al., 2009) in Round 2 which asks experts about what values they observe in the study area, rather than directly ask them the parameters of the variogram. We believe that non-geostatistical expert knowledge of spatial variability can be easily communicated in this way.

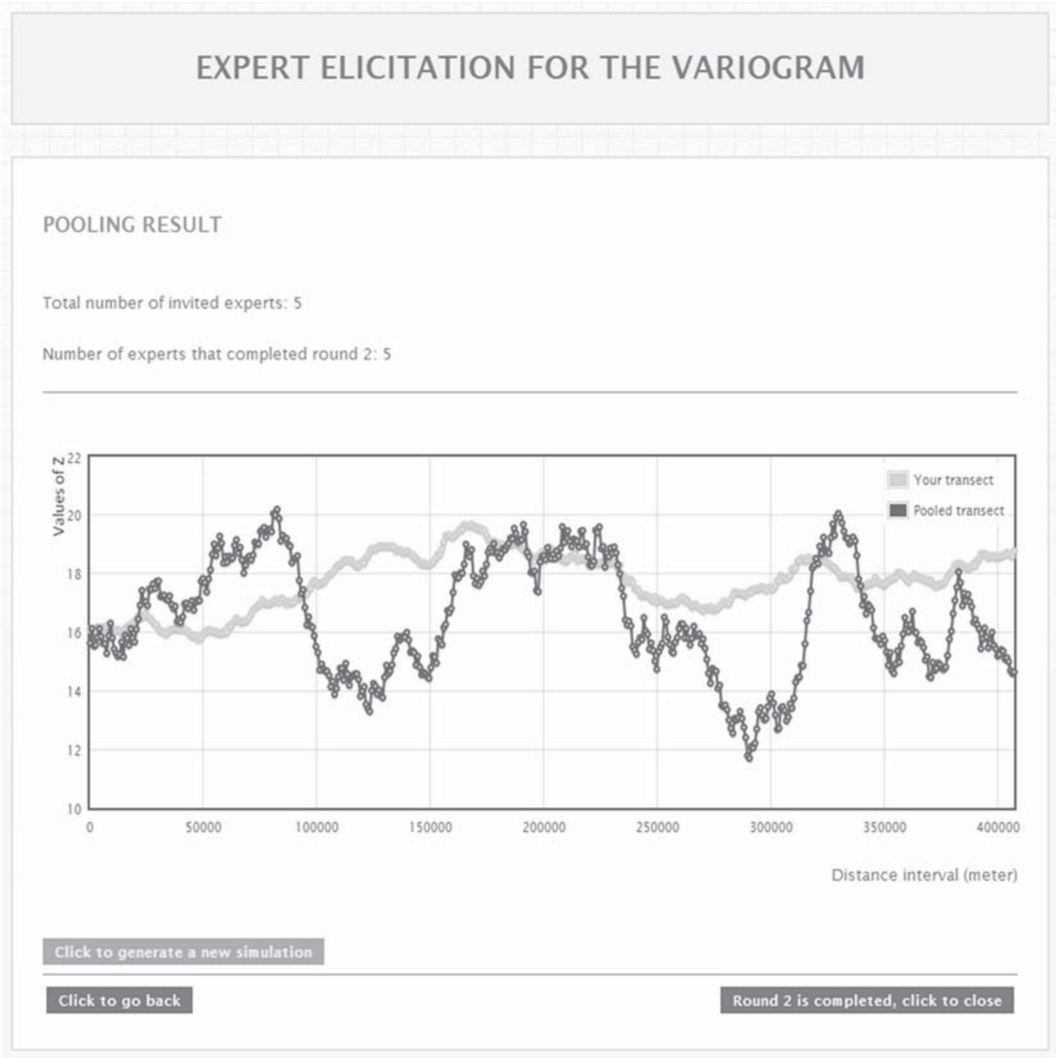

Figure 2.12: Feedback of simulated transects from one expert (grey line) and pooling (dark line)

The web-based elicitation tool only incorporates mathematical opinion pooling. Behavioural pooling methods are unappealing because these require a physical gathering of experts that cannot be remedied using video-conferencing or similar meeting formats. Amongst available mathematical combination methods, the average pooling method is the simplest but is generally found to perform as well as more complex approaches (Clemen and Winkler, 1999).

To minimize common biases in expert judgments, especially cognitive bias which occurs more often in a web-based elicitation process, we followed the guidelines in Meyer et al. (1990) and Choy et al. (2009) in the design of the protocol. The briefing document provides explanations about commonly potential biases that experts can encounter. Well-documented information about the study area and related information of the target quantity provides useful information to prevent availability bias. The questionnaire is ordered such that the risk of anchoring bias is reduced as much as possible, while revision of judgements is allowed to further reduce anchoring biased. The questionnaire can be adapted to each new case study using expert's terminology to prevent misunderstandings. Question forms of both rounds have no more than seven questions to prevent inconsistency bias caused by limited memory capacity. Motivational bias is limited in real time, e.g. experts are not informed about other experts and their individual judgements. The indirect encoding method used does not directly show a link between the experts' answers and the encoded variogram parameters, which takes the advantage of pure observation of quantities without subjective interpretations.

The web-based tool presented in this work is a research prototype that has several limitations. Firstly, the feedback of the simulated transect in Round 2 was found to be difficult to understand for some experts. An alternative is providing feedback of multiple simulated maps over whole study area, although this can slow down the process. Secondly, automatic fitting of the variogram can result in a model that does not exactly represent expert's belief about spatial smoothness due to its limited fitting accuracy. This can frustrate experts to spend much time on the revision. However, experts can always revisit and revise, which is in fact quite efficient and fast using the web-based tool (Choy et al., 2009). Adapting the tool to different problems requires some IT skills that might limit the easy reuse of the tool. Implementation of a user interface to conveniently create different case studies (for example, see The Elicitator[6]) is recommended for future development.

In spite of the limitations, the tool functions appropriately and is ready to be used for real-world case studies. These real-world case studies need to be carried out with domain experts to more exhaustively evaluate the protocol and tool. The potential routine use of the tool is promising because of its simple principle. Despite the

[6] http://elicitator.uncertweb.org/

hesitation of using expert knowledge to infer uncertain environmental variables, the expert elicitation literature has demonstrated quite clearly that when used appropriately, scientific knowledge from experts is much more than just prior knowledge and can be used directly as scientific data for modelling or analysis (Cooke and Goossens, 2000; Meyer and Booker, 2001; O'Hagan et al., 2006). This is also valid in geostatistics, where experts can be an important source of information about the spatial variability of a phenomenon, particularly when data are scarce or completely lacking. Expert information should not be discarded because it is supposedly subjective. We need proper tools to extract information from experts in a responsible way. With this work, we have provided such a tool and we hope that it may encourage further deployment of expert knowledge in geosciences.

Chapter 3

Uncertainty quantification of soil property maps with statistical expert elicitation

Based on: Truong, P.N., Heuvelink, G.B.M., 2013. Geoderma 202-203, 142-152.

3

3.1. Introduction

Errors in mapped soil properties are inevitable because our knowledge about the soil is always limited (Webster, 2000; Heuvelink et al., 2007). Mapping soil properties using a geostatistical prediction framework has the advantage that these errors are automatically quantified with the kriging prediction error (Goovaerts, 1999). The kriging prediction error characterises the uncertainty about the unknown true value at a prediction location and is represented by a probability distribution, centered around the predicted value. In a kriging based approach, the magnitude of the uncertainty is quantified by the variance of the kriging prediction error, while in a simulation based approach, it is quantified by the variation in the conditional realisations, but these are effectively the same quantifications of uncertainty because the variance of a large number of simulations equals the kriging variance (Goovaerts, 2001). The kriging variance map can be taken as a summary measure of the accuracy of the predictions because it characterises how close the predictions on average are to the unknown true values.

Although much work has been done on soil property mapping using the geostatistical framework, many maps of soil physical and chemical properties are not produced using this framework and are often not accompanied by accuracy measures. Hence, there is in general a lack of information about the quality of soil property maps. This prohibits a sensible assessment of the usability and validity of soil maps for decision making and prevents uncertainty propagation analyses that trace the propagation of errors and uncertainties through environmental models (Heuvelink, 2006).

Since many soil property maps are not produced using a geostatistical framework and the uncertainty quantification of soil maps produced by a geostatistical framework is valid only under the assumptions made in the geostatistical model, Brus et al. (2011) propose a design-based validation of soil maps using independent field measurements collected by probability sampling to obtain a model-free quantification of the soil map accuracy. Depending on the probability sampling technique used, this validation approach often requires a fairly large number of field measurements to gain the required minimum precision for accuracy assessment. For example, using stratified simple random sampling, Brus et al. (2011) obtained a relative precision of 7% of the estimated overall purity of the categorical soil map of the province of Drenthe, The Netherlands with 150 validation observations. Malone et al. (2011) combine model-based and design-based approaches to introduce two new measures of the quality of digital soil maps, one of which allows to evaluate the validity of the uncertainty quantification such as derived under the geostatistical framework. However, the estimation of these measures also requires an independent validation dataset. The requirement of many field measurements makes design-based validation approach less appealing for quality assessment of costly measured and analysed soil properties. Moreover, independent validation only provides summary measures of the map accuracy and does not yield a full spatial-probabilistic description of the uncertainty as required by uncertainty propagation analyses.

Instead of carrying out additional and, in many cases, extensive field measurements to validate the soil maps, soil experts can be invited to assess the quality of the soil maps based on their experience and knowledge. In such cases, when independent data are not available, the use of expert knowledge can be an adequate alternative source of information to fill the information gap (Choy et al., 2009; Krueger et al., 2012; Perera et al., 2012a). Experts, and the knowledge they provide, can be valuable in quantifying different kinds of uncertainty in modelling processes (Booker et al., 2001; Krueger et al., 2012).

In this chapter, we aimed at applying an existing statistical expert elicitation framework to quantitatively, probabilistically and spatially characterise the error in the mapped volumetric soil water content at field capacity (SWFC) over the East Anglian Chalk area of The United Kingdom. The SWFC map is part of the National Soil Map of England and Wales (NATMAP). The SWFC map can be prone to uncertainty due to measurement and mapping errors of the covariate data and errors in the pedotransfer function used to create the map (Minasny et al., 1999; McBratney et al., 2002; Minasny and McBratney, 2002). The SWFC map is used as one of the main inputs in a chain of models that predict regional future crop yield for the East Anglia Chalk area (UncertWeb, 2010). The uncertainty about the SWFC can lead to uncertainty in crop yield and hence, uncertainty propagation can cause bias and lack of precision in the yield prediction outcomes.

The results from this chapter are meant to serve as a demonstration of how expert knowledge can be used to quantify spatial uncertainty in maps of soil properties. Also, the presented work is meant to show the use of the elicitation protocol and web-based tool for expert elicitation of the variogram as proposed in Chapter 2. In Section 3.2, we elaborate on the materials and methods used in this work. Section 3.3 presents the results of the case study. In Section 3.4, we discuss the results, draw conclusions and give recommendations for future research.

3.2. Materials and Methods

3.2.1. Description of the study area

The study area is located in East Anglia in the southeast of The United Kingdom (Fig. 3.1). The mainly arable region was formed on a narrow continuation of the chalk ridge that runs from southwest to northeast across southern England. The region is about 839 km^2 in size, spanning about 69 km along its longest dimension; its width ranges from about 10 km to 20 km. The altitude of the region gradually increases from about 0 meters in the northeast to about 167 meters above sea level in the south-west. According to historical data from the Met Office[7], the East Anglia region is drier than other regions in The United Kingdom: it has a low annual rainfall (less than 700 mm per year) with much more even distribution of rainfall throughout the year than most other parts of The United Kingdom. The mean annual temperature is about 9-10ºC; the difference in temperature between winter and summer is about 10-15ºC. According to NATMAP - soil map of the East Anglia region, the main soil types over the study area are loam over chalk, sandy loam, deep clay and shallow silty over chalk.

3.2.2. Map of volume metric soil water content at field capacity

In this study, the SWFC is the volumetric water content at 10kPa suction. The SWFC map for the East Anglian Chalk area (Fig. 3.2) is part of the NATMAP[8]. The NATMAP is a vector map with national coverage for England and Wales. It was produced by the National Soil Resources Institute at Cranfield University, The United Kingdom. The map has a scale of 1:250,000.

The values of the SWFC map for the East Anglian Chalk area are expressed in a percentage (i.e. the volume fraction of soil water multiplied by 100%). The SWFC values are ‘representative’ for the dominant soil series associated with the polygons in the map. The values are computed by applying a pedo-transfer function that predicts the SWFC from basic soil properties (clay, silt, organic carbon and bulk density) using a multiple linear regression derived from the soil survey of England and Wales (Hall et al., 1977).

[7] http://www.metoffice.gov.uk
[8] http://www.landis.org.uk.

The values of the basic soil properties per soil series used in the regression are based on observations taken across England and Wales of which the mean values per soil series of these properties are computed and assigned to corresponding polygons on the map.

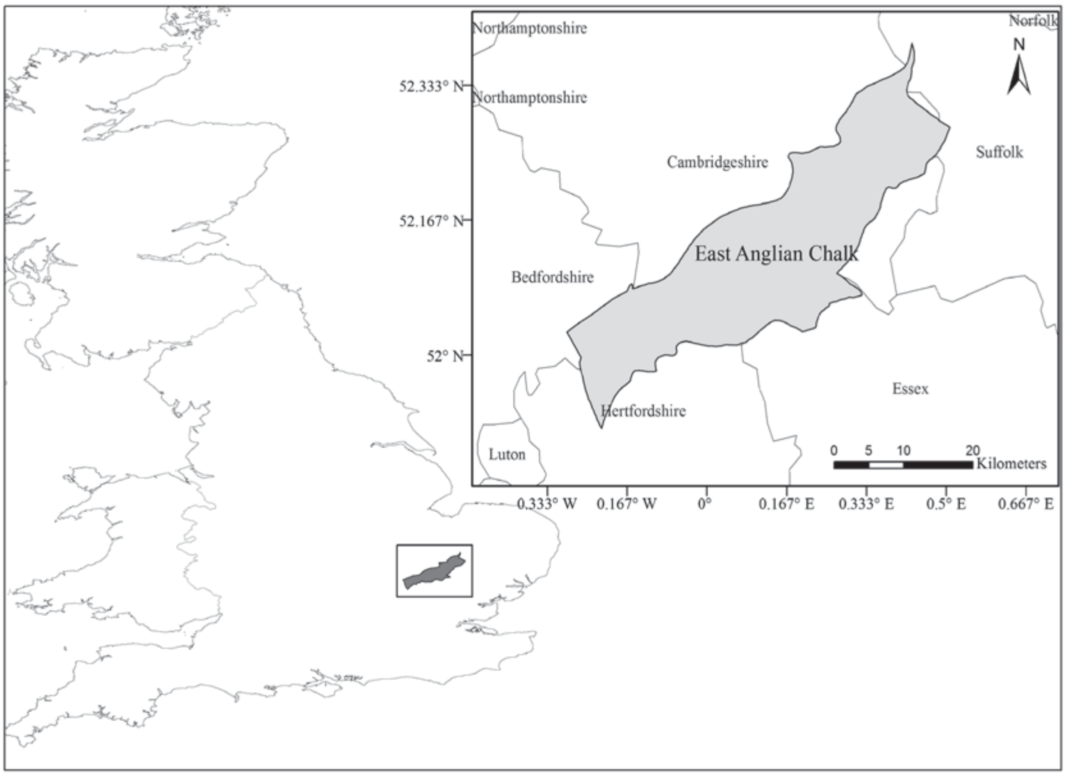

Figure 3.1: East Anglian Chalk area

In addition, maps of soil type, land cover, texture, bulk density, geology, elevation and climatic information for the study area were used as ancillary information and provided to experts. These maps were also extracted from the NATMAP vector data.

3

3.2.3. Design of expert elicitation procedure

We assumed that the error in the SWFC map is a realisation of a spatial random field because in reality, the exact value of the error at any location in the study area is unknown. Hence, we treated it as an uncertain quantity that can only be characterised by a spatial probability distribution function, defined over the study area (Webster, 2000; Heuvelink et al., 2007). In order to facilitate experts to characterise this full spatial probability distribution, we used a formal expert elicitation framework following the seven step procedure recommended by Knol et al. (2010), with some modifications. Particularly, we used the web-based tool that was designed for expert elicitation for the variogram (Chapter 2). By using this tool, we implicitly assumed that the random function model of the error is either normally or log-normally distributed, and is second-order stationary. We now describe the steps of the framework.

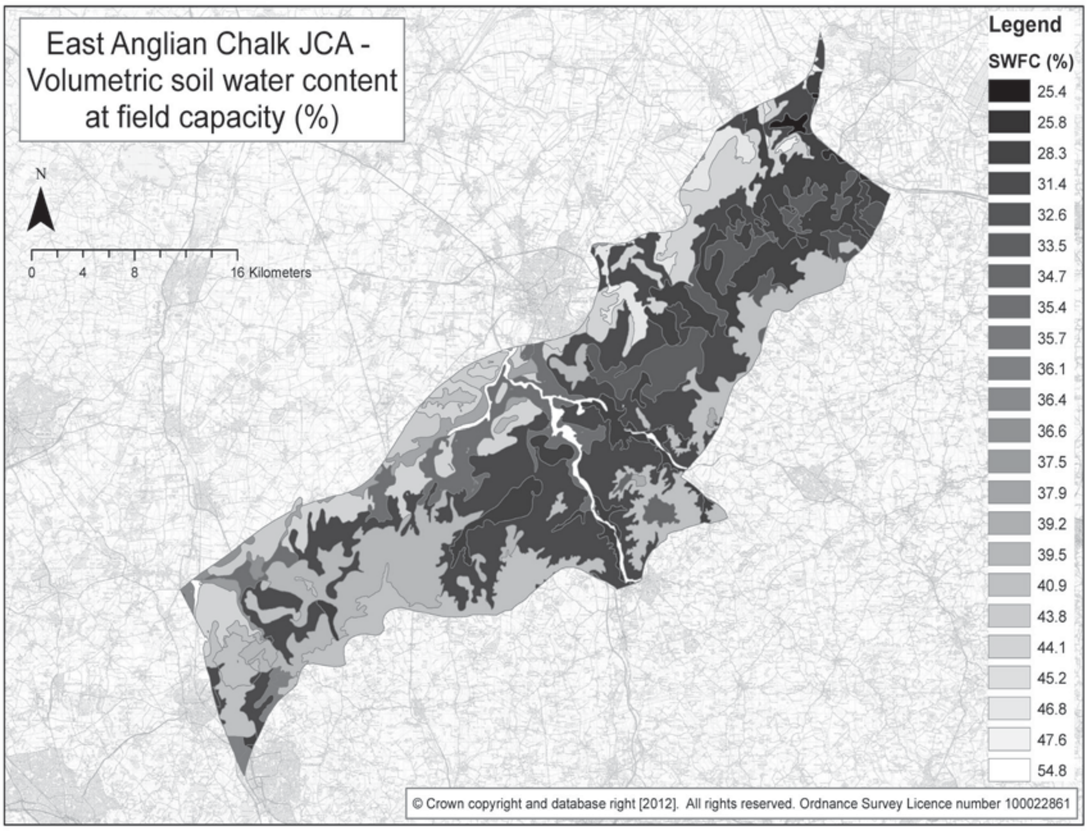

Figure 3.2: Soil water content at field capacity of the East Anglian Chalk area

Step 1: Characterisation of uncertainty

In this step, the uncertain quantity that is the target quantity of the elicitation task was defined. In this case, the target quantity is the error in the SWFC map. The error is defined as the difference between the true value of the SWFC and that given by the map at any location in the study area: $Z(\mathbf{s}) = X(\mathbf{s}) - \hat{X}(\mathbf{s})$, where Z is the error value at a location $\mathbf{s} \in D$, D is the study area, $\hat{X}$ is the SWFC value provided by the map and X is the true value of the SWFC. The true value is the SWFC of a core taken at 25 cm depth. Experts were asked to take all sources of error that cause the map value to differ from the true value into account.

As already mentioned, the error was assumed to have either a normal or log-normal distribution. Further, we assumed that the error is second-order stationary and isotropic. This means that its spatial mean and standard deviation are location-independent and that the spatial correlation of the error at two locations only depends on the scalar distance h = $|\mathbf{h}|$ between these two locations. In this work, the spatial correlation is characterised by the variogram.

The probabilistic model of the error $\{Z(\mathbf{s}), \mathbf{s}\in D\}$ is defined as: $Z(\mathbf{s}) = \mu + \varepsilon(\mathbf{s})$, where $Z(\mathbf{s})$ is a random variable that represents the error at location **s**, μ is the spatial mean error that depicts the bias in the SWFC map, the stochastic error ε is a second-order stationary and isotropic random function with zero mean and variogram function: $\gamma_{\varepsilon}(h) = \gamma_{Z}(h) = ½E[(Z(\mathbf{s}+h) – Z(h))^2]$ (Goovaerts, 1997), where γ stands for the variogram.

Step 2: Selection of experts

We selected experts through a two-step procedure. First, we nominated a list of experts. Based on this list, we studied their CVs to have a better understanding of their expertise, experience and motivation. Expertise was qualified based on their peer-reviewed papers in soil science, particularly in agricultural hydrology. Experience was derived from the time they started their research in soil science. Motivation was defined based on their acknowledgement of the usefulness of expert knowledge. In this way, we selected ten experts from The United Kingdom and The Netherlands with a high level of expertise, at least ten year experience in soil science with a good level of motivation and who are familiar with the study area. All selected experts have more than five peer-reviewed papers in soil science or agricultural hydrology. Some of them have used expert knowledge in their research. After refining the list of selected experts, we sent them by email an invitation letter for the elicitation exercise. Six of the ten selected experts, one from The Netherlands and five from The United Kingdom were willing to participate in the full elicitation procedure. The others could not participate, mainly due to a conflict in time. We decided to execute the elicitation exercise with the six experts as we expected that this number of experts is sufficient to achieve robust results (Knol et al., 2010).

Step 3: Design of the elicitation protocol

In this step, we adopted the web-based protocol for expert elicitation of the variogram (Chapter 2) which allows to elicit from experts the mpdf and the variogram to form the full spatial distribution of the spatial random error under the assumptions of a normal and log-normal distribution. There are two main rounds in the elicitation protocol. Round 1 is the elicitation of the mpdf and Round 2 is the elicitation of the variogram (Fig. 2.1). By using the web-based tool, the tasks in Step 2-Scope and format of the elicitation, Step 5-Preparation of the elicitation session and Step 6-Elicitation of expert judgements in Knol et al. (2010) were incorporated into this step. Implementing this step is the most laborious task among all steps.

After experts were given usernames and passwords, they could access the tool at: http://www.variogramelicitation.org. We designed an introduction page that provided the experts with a detailed description of the study area, information and map of the SWFC together with maps of all auxiliary variables as described in Section 3.2.2. The introduction was provided to familiarise the experts with the context and purpose of the case study.

The definition of the target uncertain quantity, i.e. the error as defined in Step 1 was stated in the briefing document. The following additional information was provided in the briefing document: objective of the elicitation task, outline of the elicitation procedure, techniques used for elicitation, definition of probabilistic terms, requirements from experts and common causes of biased judgments. The experts were requested to carefully read the introduction and the briefing document before starting to answer questionnaires. Table 3.1 presents the questionnaire for both rounds. The question in Round 2 was used for all lag distances. The experts were informed that the outcomes from the elicitation were to be presented in a research paper aiming at students, experts, decision makers and scientists in crop yield modelling and soil science. This information clarified the expected use of the outcomes to properly motivate the experts in giving their judgements.

We scheduled about a week for the experts to finish the first round and about two weeks for the second round. For Round 1, because there were holidays in between, the time lasted from 20th of December, 2011 when the first notice was emailed to the experts until 6th of January, 2012 as the final day. Round 2 lasted two weeks from 23rd of January, 2012 to 5th of February, 2012. We recommended that the actual time that each expert needed to finish each elicitation round was about half an hour. The web-based tool provided a mechanism for minimising bias in experts' judgements (Chapter 2). It also provides graphical feedback and a mechanism (equal weight pooling) for pooling multiple experts' judgements.

The experts were contacted by email to start each of the elicitation rounds and to discuss any questions or problems during the elicitation procedure. We played the role of the facilitator to answer the questions and clarify any unclear statements and issues during the elicitation procedure.

Step 4: Evaluation and report of results

This is the final step of the formal elicitation procedure. To evaluate the elicitation exercise, besides communicating with the experts during the elicitation session, we also asked for feedback from the experts through a questionnaire. The questionnaire covered two main aspects: the practicality of the elicitation exercise and the experts' performances. The questionnaire was sent to the experts by email and required a maximum of ten minutes for each expert to complete.

Documentation of the elicitation procedure and presentation of the results are accomplished in this paper. The results are presented graphically and in the form of summary tables. Judgements by individual expert are reported anonymously.

3.3. Results

For the sake of consistency when reporting results, we define several notations. Because we anonymously report judgements from each expert, we encode the six experts as Ei (i = 1, ..., 6). Table 3.2 presents the experts' affiliations, titles and expertise. The variogram lags are encoded as Lj of which the index j = 1, ..., 7 numbers the lags from the shortest to the longest lag. Recall that Z denotes the isotropic second-order stationary spatial random error in percentages. The results of the two elicitation rounds are presented in turn.

Table 3.1: Questionnaire for Round 1 and Round 2 of the elicitation protocol.

Questions of Round 1
1. What is the lowest possible value of *Z*? (Z_{min})
2. What is the highest possible value of *Z*? (Z_{max})
3. What is the value Z_{med} such that there is a 50% probability that the value of *Z* is less than or equal to this value? $Pr(Z \leq Z_{med}) = 50\%$
4. What is the value $Z_{0.25}$ such that there is a 50% probability that the value of *Z* is less than or equal to this value within the interval $[Z_{min}, Z_{med}]$? $Pr(Z \leq Z_{0.25}) = 25\%$
5. What is the value $Z_{0.75}$ such that there is a 50% probability that the value of *Z* is less than or equal to this value within the interval $[Z_{med}, Z_{max}]$? $Pr(Z \leq Z_{0.75}) = 75\%$
Questions of Round 2
For each of seven variogram lags:
Could you specify a value V_{inc_med} such that there is a 50% probability that the value of the V_{inc} is less than or equal to this value for each of the following distances? $Pr(V_{inc} \leq V_{inc\_med}) = 50\%$

3.3.1. Judgements on the marginal probability distribution

In the first round, the experts were asked to judge the mpdf of the random error *Z* at a random location within the study area. Due to the second-order stationarity assumption, the mpdf is the same at all locations within the study area. Hence, it is immaterial which location it is, and indeed, it was important not to influence experts by pointing them to a particular location. The experts gave judgements for the plausible range of the value of *Z* (the possible minimum and maximum values) and the three quartiles by answering the questionnaire. Fig. 3.3 shows the judged outcomes of each expert for the plausible range and the three quartiles of the mpdf of *Z*.

There were variations in the plausible range of the error in the SWFC map. Expert E3 judged the largest range, from -50% to 50%. Expert E6 gave the smallest plausible range, from -5% to 5%, which is almost comparable to that of expert E5. The three experts E1, E2 and E4 judged nearly the same range. All ranges had overlaps so that all plausible ranges contained the interval of 5% to 5% while two of the judged medians were outside this interval. Experts E2 and E4 judged the median to be 7.5% and 10%, all other experts judged the median to be 0%. All of the judged values of all experts resulted in a (more or less) symmetric distribution around the median values that indicates a normal distribution. The results of fitting a probability density function to the expert's judged values are shown in Fig. 3.4.

Table 3.2: Information about experts participating in the study

Expert	Affiliation	Title	Expertise
E1	- [a]	-	Land surface modelling, soil physical, hydro-meteorological and plant-physiological measurement methods
E2	Department of Geography & Environmental Science, University of Reading	Professor in Soil Science	Soil science, Pedometrics
E3	Computer Science, Aston University	Reader in Computer Science	Environmental Statistics, Geo-Informatics
E4	Department of Geography, Brigham Young University	Associate Lecture	Soil Science
E5	Rothamsted Research	Lawes Trust Senior Fellow	Soil science, Pedometrics
E6	Soil Physics and Land Use Team, Alterra, Wageningen University	Senior Researcher	Soil Sciences, Soil Physics, Land Evaluation

[a]E1 wishes the information to remain anonymous.

3

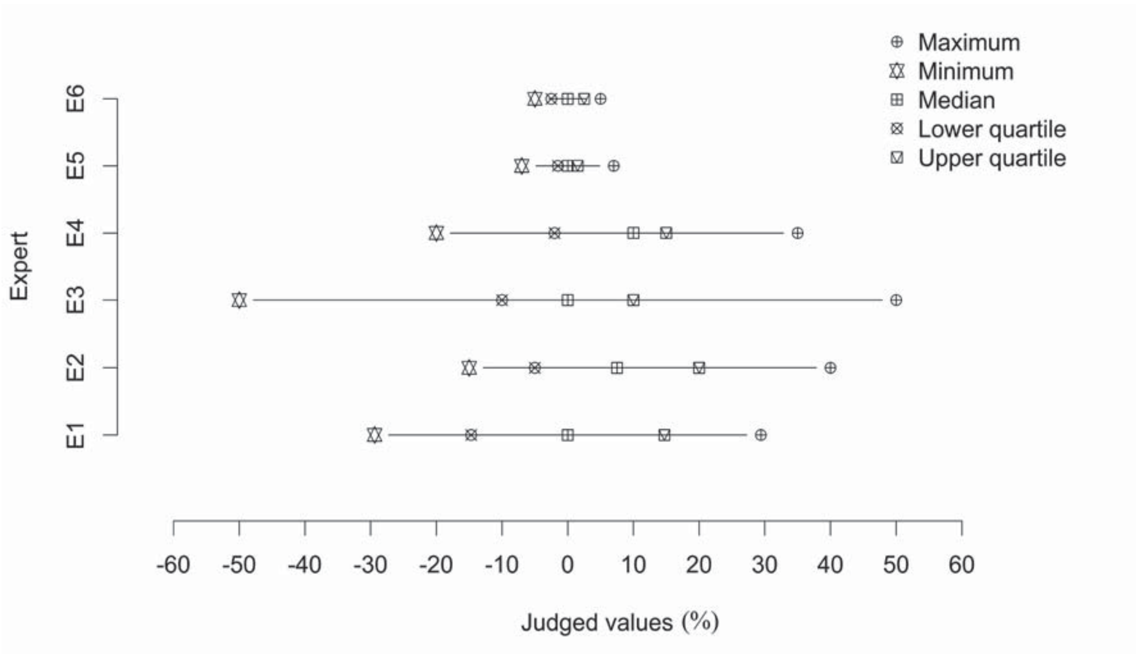

Figure 3.3: Judged values from six experts for the minimum, maximum and the three quartiles of the marginal probability distribution of the error at a randomly chosen location in the study area

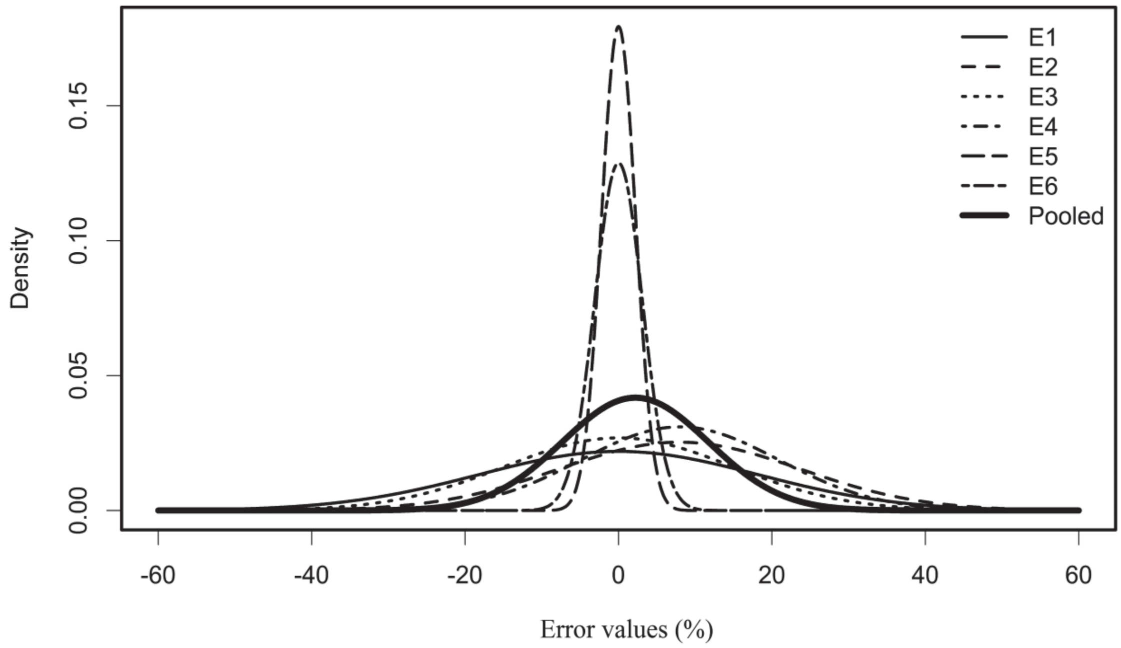

Figure 3.4: Fitted marginal probability density functions to individual expert' judgements and to the pooled opinion

All fitted marginal probability distributions were found to be normal distributions, which are fully characterised by their means and variances. The mean defines systematic error (bias), the variance (or its square root, the standard deviation) defines the random error. Table 3.3 presents the means and standard deviation of the fitted probability density functions to experts' judgements. Experts E2 and E4 judged that there is a positive systematic error in the SWFC map of about 8%. The other experts judged that the map was not biased and only contaminated by random error. The fitted standard deviation measures the absolute degree of variation in the random error. The fitted standard deviation to the judged values of experts E5 and E6 are much smaller than those of all other experts. This is also depicted in Fig. 3.4 by the two narrow probability density graphs for experts E5 and E6.

As also shown in Fig. 3.4, the pooled mpdf that is a probabilistic average of all mpdfs is a normal distribution with a fitted mean of 2.2% and a standard deviation of 9.5%. The pooled mpdf indicates that there is not only a random error but also a positive systematic error of 2.2%.

Table 3.3: Means and standard deviations of the fitted probability distributions to experts' judgements

Expert	Fitted mean value (%)	Fitted standard deviation (%)
E1	0.0	18.2
E2	8.0	15.8
E3	0.0	14.8
E4	8.0	12.9
E5	0.0	2.2
E6	0.0	3.1

3

3.3.2. Judgements on the variogram

Variogram elicitation was started by eliciting the median of the absolute values of the first-order increments: $V_{inc} = |Z(\mathbf{s}) - Z(\mathbf{s}+h)|$ for seven values of the spatial lag distances h (Chapter 2). We denote this median by V_{inc_med}. The seven lags where the values of V_{inc_med} were elicited were estimated based on the extent of the study area. The seven lags Lj range from 0.5km to 50km and are shown on the bottom-right legend in Fig. 3.5. In general, the judged values of V_{inc_med} from all experts increased when the lag distances increased and satisfied Tobler's first law of geography (Miller, 2004). The V_{inc_med} for seven lags were used to calculate the variogram values by the equation: $2\hat{\gamma}(h) = 2.198(V_{inc\_med})^2$ (Cressie and Hawkins, 1980). The judged values must result in estimated variogram values that are smaller than or equal to the variance of the pooled mpdf. This is because the variance of the pooled mpdf represents the maximum possible total variance of the error over the study area. These are two of the coherence conditions incorporated in the web-based tool (Chapter 2). However, the increases of the variogram values with increasing distances were quite different between experts, and this resulted in different scales of spatial dependency. The variations in the values of V_{inc_med} among experts were also different for different lags. There were almost no common values of V_{inc_med} in each lag, except for lag L1 (two experts: E3, E4) and lag L2 (three experts: E3, E4 and E5). There were considerable variations in the judged values of V_{inc_med} at the first three shorter lags compared with the other four lags. This means that there were diverse opinions from the six experts on spatial variability at short distances, which is indeed reflected in the fitted variogram models shown in Fig. 3.6. Five best fitted variogram models are Matérn models with different values of the parameters (E1, E2, E3, E5, E6); one best fitted model is the spherical model (E4).

A first glance on Fig. 3.6 shows that there were diverse opinions amongst the experts about the spatial variability of the error in the SWFC map over the study area. The fitted outcomes (Fig. 3.6 and Table 3.4) show that it appears that according to the judgements of expert E6, spatial variability of the error at short distances is large (largest values of the nugget and partial sill and smallest range value). Its range parameter was smallest, about 2km, meaning that the extent of spatial correlation is short. There was a large disagreement between the best fitted variogram to the judgement of expert E6 compared with those of other experts.

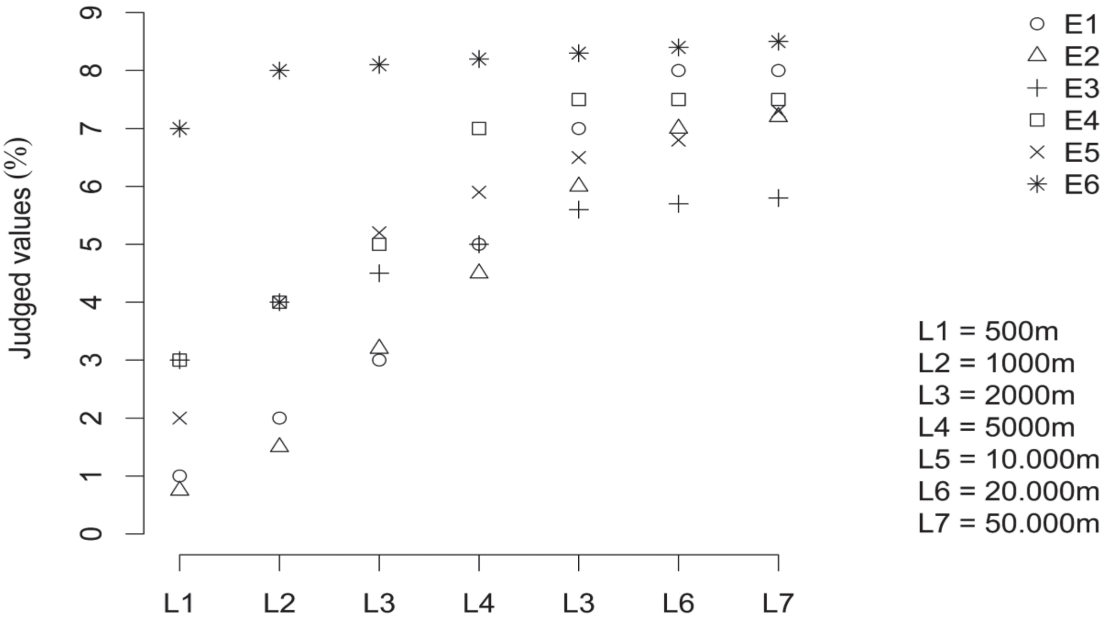

Figure 3.5: Judged values of medians of first increments (V_{inc_med}) at seven spatial lags from six experts

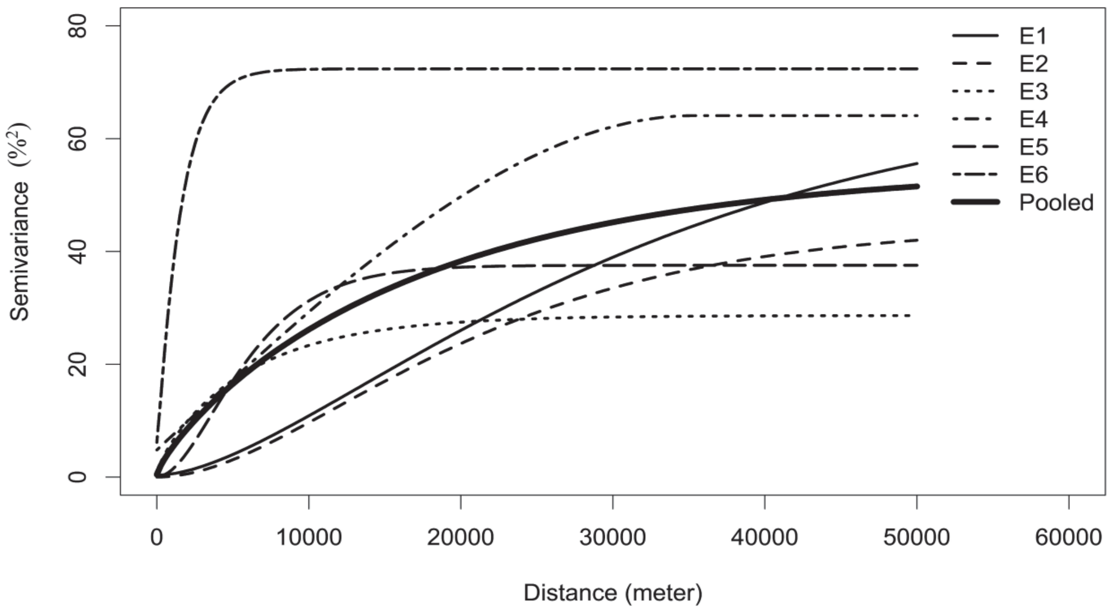

Figure 3.6: Variogram models fitted to expert' judgements and to the pooled one

3

The variogram models of experts E1, E2 and E5 depict almost the same behaviour at short distances: strong spatial correlation resulting in smooth variation (larger values of the kappa parameters of the Matérn models and smaller nugget effect). The range parameter of the model of expert E5 (7.5 km) and that of expert E2 (10 km) were shorter than that of expert E1 (20 km). This means that according to experts E2 and E5, the spatial correlation declined fairly quickly, while expert E1 judged a gradual decrease of spatial correlation with increasing distance. The opinions of experts E3 and E4 can be classified into another group. Although the fitted variogram models to their judgements were also different among each other, the behaviours of the fitted variogram models at short distances up to 10 km were more similar than for those of the other experts. This group of opinion seemed to be 'moderate', while the other two groups of opinion were seen as the two extremes. The fitted variogram to expert E4 was the only spherical model that has a large range, large partial sill and also a large nugget. This model resulted in a combination of a noisy signal and gradually changing values over distances.

Table 3.4: Fitted variogram models to six experts' judgements and their parameters.

Expert	Model	Parameters			
		Range (meter)	Nugget ($\%^2$)	Partial Sill ($\%^2$)	Kappa
E1	Matérn	19,623	0.34	68.88	1.1
E2	Matérn	10,352	0.004	44.55	1.7
E3	Matérn	6,918	0.73	27.92	0.4
E4	Spherical	35,368	4.81	59.28	-
E5	Matérn	7,575	0.09	37.47	2.0
E6	Matérn	2,164	6.16	66.22	0.6

The pooled variogram was obtained by fitting a variogram model to the variogram values that were the average values of all experts' judgements for the seven lags (Chapter 2). The pooled variogram model is the Matérn model with parameter values: nugget = $0.45\%^2$, partial sill = $54.6\%^2$, range = 25,400 meters, kappa = 0.4. A set of simulated maps of the error based on the pooled mpdf and the pooled variogram were generated using unconditional sequential Gaussian simulation (Fig. 3.7). Fig. 3.8 shows the simulated SWFC maps that were generated by adding each of the simulated error maps to the SWFC map (i.e. Fig. 3.2).

3.3.3. Feedback from experts on the elicitation exercise

The feedback from the experts about the elicitation exercise concerned two aspects: feedback on the experts' performances and on the practicality of the elicitation exercise. Table 3.5 provides the summary of the experts' answers on the questionnaire (see Appendix 3.A) that they received.

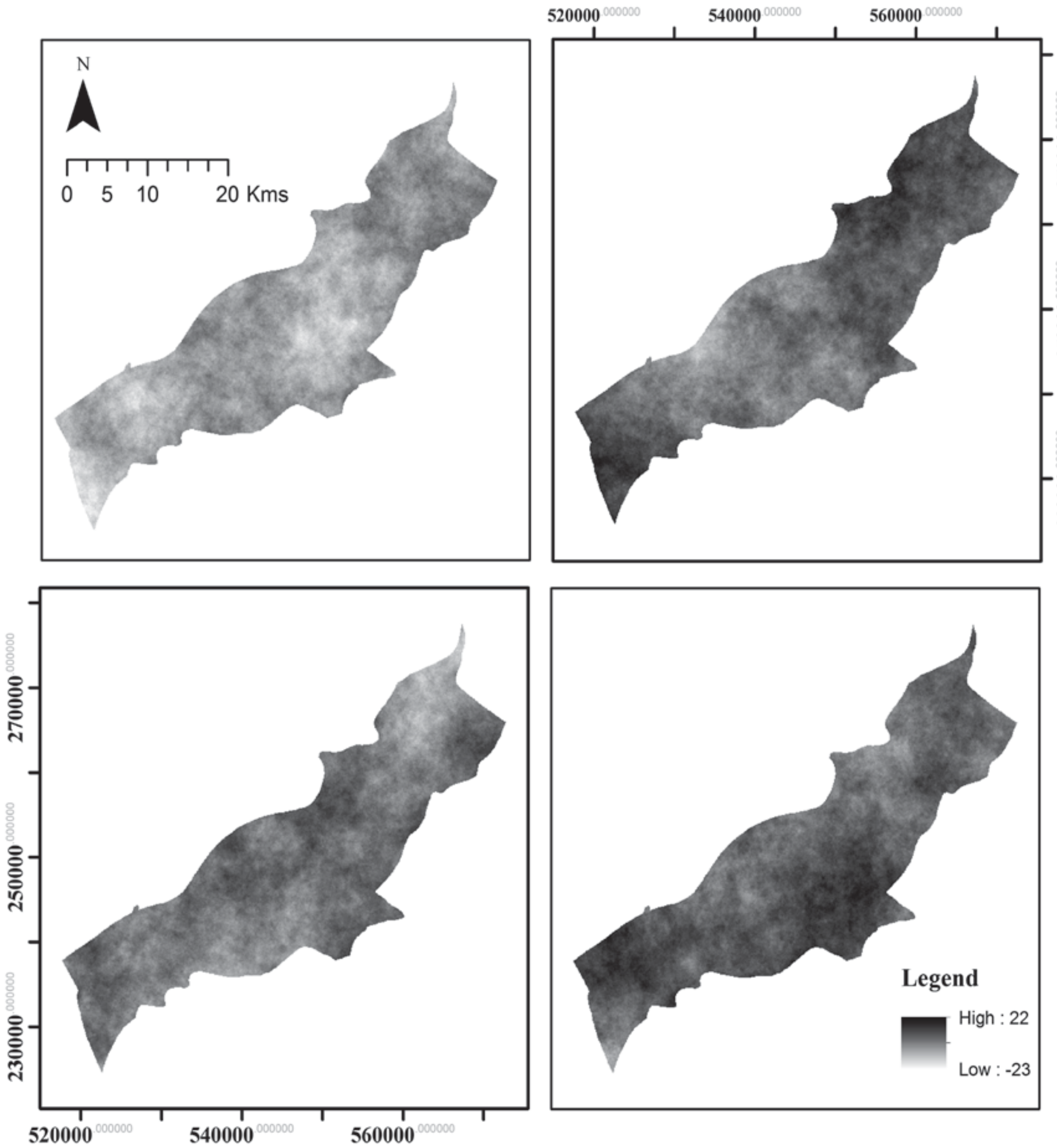

Figure 3.7: Four simulated maps of spatial error of soil water content at field capacity for the East Anglian Chalk area

3

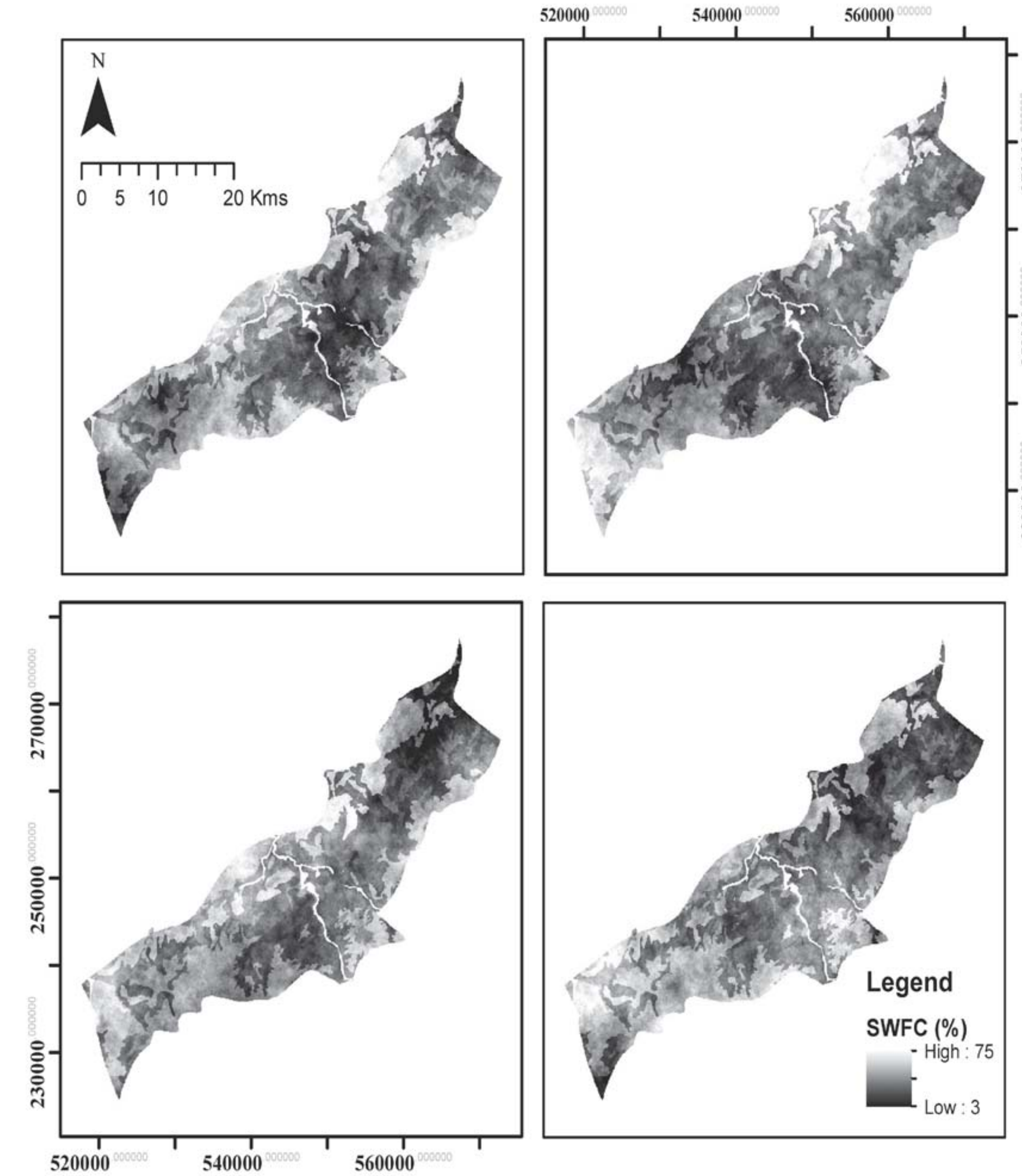

Figure 3.8: Four simulated soil water content at field capacity maps for the East Anglian Chalk area

Table 3.5: Summary of experts' feedbacks on the elicitation exercise

Feedback on expert performances							
Confidence		Very confident		Confident		Less confident	
	Round 1			E2,E5,E6		E1,E3,E4	
	Round 2			E2,E5		E1,E3,E4,E6	
Elicitation time		< 15 mins	15 mins		30 mins		> 30 mins
	Round 1				E2,E5,E6		E1,E3,E4
	Round 2		E2,E4,E5		E1,E3,E6		
Experience on expert elicitation		Yes				No	
		E1				E2,E3,E4,E5,E6	
Field work on study site		Yes				No	
		E4				E1,E2,E3,E5,E6	
Most useful data	E1 Soil texture and structure, land cover	E2 Soil texture and structure, elevation, soil map	E3 Soil map	E4 Soil texture and structure		E5 Soil map	E6 Soil texture and structure
Feedback on the web-based tool							
	Very satisfied		Satisfied			Not satisfied	
Introductory part	E4		E1,E2,E3,E5,E6				
Briefing document			E1,E2,E3,E5,E6			E4	
	Very clear		Clear			Not clear	
Definition of the error			E3,E5,E6			E1,E2,E4	
	Very easy		Easy			Not easy	
Question of Round 1							
Understanding	E3		E2,E6			E1,E4,E5	
Answering			E2,E4,E6			E1,E3,E5	
Question of Round 2							
Understanding			E2,E4,E5,E6			E1,E3	
Answering			E2,E4,E5,E6			E1,E3	

On the performances, most experts were only moderately confident in their judgements. For the results of Round 1, three out of six experts were less confident in their judgements. Two of the three experts who were confident on their judgements also judged quite narrow mpdfs (Fig. 3.4). For the results of Round 2, four out of six experts were less confident. Not every expert was very confident on all judgements. Although most of them expressed less confidence on answering the questions of Round 2, the time they spent on this round was less than that for Round 1. All experts spent at least thirty minutes on Round 1. In Round 2, the two experts who expressed confidence on their judgements spent only 15 minutes, half of the recommended time of 30 minutes while the rest spent the maximum of 30 minutes. This contrasts with our expectation that experts would spend more time carefully examining the SWFC map to answer the questions in Round 2.

Continuing on the performance of the experts, only one expert had previous experience on participating in an elicitation exercise. Only one expert had done field work in the study area, although all were familiar with the soils in this region. We were interested in which information was most important for the experts to base their judgements on. Four experts found the information about soil texture and structure of the study area most useful. The other two experts found the soil map most useful.

Another aspect reflected in the experts' feedback regarded the practicality of the tool. During the elicitation procedure, there was no complaint about the functionality of the tool. All experts were (very) satisfied with the information in the introductory part, although one expert felt somehow distracted by the interesting background information. For the briefing document (considered more important by us and also by the experts), one expert was not satisfied with the information provided in this document without further comments, while other experts were satisfied. Half of the experts found the definition of the error of the SWFC map not easy to understand. This indeed resulted in a misunderstanding about what the error really reflected. Several experts initially interpreted the error as the absolute error, meaning that it could only take non-negative values. This was clarified by email contact in the initial phase of Round 1. Most experts found it easy to answer the questions when they could easily understand them and vice versa; while one expert found the questions of Round 1 very easy to understand but difficult to answer.

3.4. Discussion and Conclusions

In this section, we discuss the results of the expert elicitation process, especially their robustness and the possibility of bias. Robustness here means how closely the six experts' opinions represent the total expert community's. We will also discuss the practicality of the web-based tool and facilitating expert elicitation using the web-based tool.

The spatial probability distribution of the error in the SWFC map has a normal distribution that can be fully specified by the mean of the pooled mpdf and the pooled variogram. Because we used the equal weight pooling method, each expert opinion has equal impact on the captured knowledge about the two moments of the spatial error. The mean of the pooled mpdf with a positive value of 2.2% indicates that the SWFC map was overall underestimated for all soil types (or at every location). Compared to the square root of the sill variance of the pooled variogram (i.e., 7.4%), the systematic error is smaller than the random error. On average at every location the systematic error contributes 22% to the total error (defined as the square root of the sum of the squared systematic error and the variance of the random error), which equals 7.7%. The relative error, defined as the ratio of the total error and the SWFC map, is on average about 20%.

To elicit the uncertainty of the SWFC map, we assumed that the spatial error satisfied the stationarity assumption. This is a strong assumption, but without it the elicitation would have become much more difficult, perhaps stretching it beyond what may reasonably be expected from experts. Second-order stationarity is frequently assumed in geostatistics, but it is important to verify that the resulting model is a plausible description of reality. In our case, we assumed that the error in the SWFC has constant mean and variance, while it may be more realistic to relax this assumption and let it vary with soil type (e.g., larger uncertainty in stony soils). This could be a topic for follow-up research.

3

The equal weight pooling method simply averaged experts' judgements on both the mpdf and the variogram. Unweighted averaging is simplistic but pragmatic and a quite efficient approach amongst alternative mathematical aggregation methods (Clemen and Winkler, 1999; O'Hagan et al., 2006; French, 2011). Alternatively, weighted pooling can be used to give some experts larger weights than others. The weights can be interpreted in a variety of ways (Genest and McConway, 1990), e.g., relative quality of the experts (i.e. relative experts' level of expertise), the combination of the informativeness of experts' judgements and the experts' performance (Cooke, 1991), etc. The weights can be assigned to experts by the analysts or the decision maker (French, 2011) or the experts can weigh each other and/or choose their own weights (i.e. self-assigned weights) (DeGroot, 1974; Genest and McConway, 1990). But there is still controversy about how to adequately assign weights to different experts and in which conditions using weighted average truly improves results compared to unweighted average (O'Hagan et al., 2006; Clemen, 2008). Therefore, we chose the equal weight pooling method.

The pooled outcomes can be interpreted as the average knowledge of six representatives selected from the population of experts who are well-qualified for the investigated case. Based on the recommendations from several publications that serve as guidelines to design and conduct a statistical expert elicitation, six experts should be enough to obtain robust results when considering the trade-off between expenses and informative gain (Meyer and Booker, 2001; Hora, 2004; Knol et al., 2010). We also selected experts from different institutes who have qualified knowledge about the soil properties of the study area. This makes the elicited results well representative for (diverse) opinions on the error of the SWFC map. Concerning the reliability of the elicited outcomes, half of the experts were confident about their judgements in Round 1 while only two out of six experts expressed confidence in their judgements in Round 2. We conclude that the elicited outcomes encapsulate the current knowledge of multiple experts of the error in the SWFC map for the East Anglian Chalk area, The United Kingdom at a less confident level.

The elicitation method we used is a variation of the Delphi method (Ayyub, 2001) where the experts' judgements are anonymously and independently elicited. By examining the elicited outcomes from every expert, we see that the experts' judgements for both the mpdf and the variogram seem to be clustered. The cluster of the judgements might indicate true consensus in a subgroup of experts about the error in the SWFC map. However, it can also indicate a correlation or dependence in experts' knowledge that can bias the equal weight pooled outcomes (Meyer and Booker, 2001). The striking difference in judged values from Round 1 is that between the nonzero (E2 and E4) and zero median (E1, E3, E5 and E6) of the mpdf. Assuming that E2 and E4 are completely dependent, one of the expert judgements would be eliminated from the pooling, then the positive bias would reduce. But, if experts in the second subgroup are completely dependent, only one opinion from the second subgroup can contribute to the pooling, in this case the positive bias increases. It would be interesting to examine the dependence in expert judgements. However, the feedback on experts' performances (Table 3.5) and information about experts given in Table 3.2 are not sufficient to extensively analyse the dependence in experts' opinions, and this is beyond the scope of this study. Moreover, in the context of web-based elicitation, detecting the occurrence of cognitive and motivational biases in the expert judging process (Meyer and Booker, 2001) was difficult or impossible because the performances of the experts while giving judgements could not be observed.

Initially, the outcomes from Round 1 were systematically biased due to misinterpretation of the experts about the definition of the error (i.e., the absolute error was elicited). Thereby, all experts redid the elicitation task for Round 1. This misinterpretation might have been avoided by doing a pre-elicitation training (Knol et al., 2010); but, we did not include it in our four steps (Section 3.2.3). Moreover, although the elicitation session was prepared according to a formalised elicitation protocol, the introduction of the case study and the briefing documents were not given to the experts prior to the elicitation exercise. These documents should ideally have been accessible to the experts at least two weeks in advance (Ayyub, 2001). The lack of a pre-elicitation training also reflected on the experts' performances (Table 3.5). While some experts were familiar with giving probabilistic judgements, other experts found doing this more difficult. It would have been better if they had been trained to give probabilistic judgements prior to their involvement in the elicitation exercise (Hogarth, 1975). These experiences confirm that pre-elicitation training is very useful to familiarize experts to the elicitation exercise and giving probabilistic judgements and to clarify misunderstandings or issues, especially about the target quantity, in the context of web-based statistical expert elicitation.

We can conclude from the case study that the web-based tool, which provides a uniform procedure to characterise the spatial probability distribution of uncertain variables from expert knowledge, functioned well. With the developed elicitation protocol and the web-based tool, we can quantify spatial uncertainty of soil property maps from expert knowledge. Simulated SWFC maps of the study site such as shown in Fig. 3.8 can be used to investigate the propagation of uncertainty from the SWFC map to the output of the regional crop yield model. This study also showed that expert knowledge can be used to derive site-specific variograms of uncertain (map errors in) soil properties. This can overcome the limitations of using an average variogram that is not site-specific (McBratney and Pringle, 1999). We have learnt several lessons from our experiences of facilitating an elicitation exercise with a web-based tool:

1. The facilitators play a crucial role in the success of the elicitation exercise, also for the web-based elicitation methods where a self-elicitation process is expected.

2. Motivation is a very important criterion when choosing experts for the success of the elicitation exercise and reliability of the elicited outcomes.

3. Differences in experts' opinions are legitimate (Morgan and Henrion, 1990); but reliable elicitation protocols are those that do not exaggerate these inherent differences.

4. To determine whether experts' judgements are dependent, an extensive investigation on the data relating to experts' problem-solving process is required (Meyer and Booker, 2001).

5. Choosing a suitable elicitation technique is not easy; while analysing expert judgements is even more difficult. However, a reliable elicitation protocol can positively ascertain the generalisation of the elicitation results.

6. Computer tools are uniform, supportive and reusable mechanisms for eliciting expert knowledge, but they have the disadvantage compared to physical expert elicitation meetings that experts' performances cannot be monitored for the possibility of bias occurrence.

7. Precision in elicited outcomes from multiple experts might indicate a poor elicitation protocol, while imprecision does not necessarily represent inaccuracy in experts' knowledge.

This study showed that statistical expert elicitation is a promising method to characterise spatial uncertainty of soil property maps using expert knowledge when data-based validation methods are not affordable or feasible. The value of expert knowledge in soil science was acknowledged as a valuable informative prior, especially when there are no alternative useful sources of information (Stein, 1994). Exploring, developing and applying reliable methods to extract knowledge from experts, e.g. using statistical expert elicitation for the variogram elicitation as done in this study, should be stimulated among soil scientists to effectively and reliably extract information from experts in soil research.

Appendix 3.A. Questionnaire for elicitation exercise evaluation

Dear Expert,

The whole elicitation exercise has now ended. Below is a questionnaire to help us evaluate the elicitation exercise. Please do not spend more than 10 minutes to finish this questionnaire. All the results from the elicitation exercise and the answers to this questionnaire will be presented in a research paper (you will have access to the paper when it is ready to be published).

Thank you very much for your contribution to the elicitation exercise.

The elicitation team.

Please choose only one option for each answer to every question

1. How confident are you about your judged values?

Confidence levels	Round 1	Round 2
Very confident	☐	☐
Confident	☐	☐
Less confident	☐	☐

2. Are you satisfied with the overall information provided on the website: that is, the “Introduction” part and the “Briefing document” part?

Satisfaction levels	Introduction	Briefing document
Very satisfied	☐	☐
Satisfied	☐	☐
Not satisfied	☐	☐

3

3. What information did you find the most useful information when judging the error in the mapped soil water content?

Information	Your choice
Soil texture and structure	☐
Temperature	☐
Soil map	☐
Land cover	☐
Annual precipitation	☐
Geology map	☐
Elevation map	☐

Please specify any others:

4. Was the description of the error (Z) in the mapped soil water content at field capacity clear to you?

Clearness	Your choice
Very clear	☐
Clear	☐
Not clear	☐

5. Are the questions easy to understand?

Easiness	Round 1	Round 2
Very easy	☐	☐
Easy	☐	☐
Not easy	☐	☐

6. Are the questions easy to answer?

Easiness	Round 1	Round 2
Very easy	☐	☐
Easy	☐	☐
Not easy	☐	☐

7. Have you had any field work before on the study site, i.e. in the East Anglian Chalk region of The United Kingdom, particularly for the soil water content at filed capacity?

Yes	☐
No	☐

8. How much time did you spend for each round of the elicitation exercise?

Time	Round 1	Round 2
Less than 15 minutes	☐	☐
15 minutes	☐	☐
30 minutes	☐	☐
More than 30 minutes	☐	☐

9. Have you ever participated in an elicitation exercise before?

Yes	☐
No	☐

10. Do you have other comments on the elicitation exercise?

Thank you for taking the time to complete this evaluation questionnaire!

Chapter 4

Bayesian area-to-point kriging using expert knowledge as informative priors

Based on: Truong, P.N., Heuvelink, G.B.M., Pebesma, E., 2014. International Journal of Applied Earth Observation and Geoinformation 30, 128-138.

4

4.1. Introduction

Spatial disaggregation (downscaling) is becoming more important in a world where the demand for data transformation from global to local scales is rapidly increasing. In climate research, for example, regional or local climate models may require data of spatial climate attributes (e.g. precipitation, air temperature or atmospheric vapour) at finer resolution than those measured using remote sensing instruments or predicted using global climate models. Here, spatial resolution or pixel size stands for the spatial support, i.e. the geometrical size, shape and spatial orientation of a spatial unit of an observation or a prediction. Changing the spatial support of a variable changes its statistical and spatial properties (Schabenberger and Gotway, 2005). This is the well-known change of support problem (Cressie, 1996; Gotway and Young, 2002; Schabenberger and Gotway, 2005).

Spatial support and change of support problem have been acknowledged as an important source of uncertainty in remote sensing analyses due to aggregation and zoning effects (Marceau and Hay, 1999; Dungan, 2006). Spatial disaggregation of remotely sensed imagery through interpolation shows an important application of geostatistics to remote sensing analysis (Van der Meer, 2012). Well-known geostatistical techniques for downscaling remotely sensed imagery of continuous variables are Area-to-point (ATP) kriging and multivariate ATP kriging (Atkinson, 2013).

In this study, we focused on ATP kriging (Kyriakidis, 2004) for spatial disaggregation of a Gaussian random field. ATP kriging follows the principle of classical kriging and makes predictions of an attribute at point support (PoS) from block support observations (BSO) of the same attribute. It also quantifies the uncertainty about the disaggregated predictions by means of the ATP kriging variance. ATP kriging satisfies the condition that the arithmetic average of the predictions (and simulations) at all point locations within a block equals the value of this block when the same number of BSO are used as conditioning data (Goovaerts, 2008). Hence, to use ATP kriging, BSO must be (assumed to be) the arithmetic average of PoS data within the blocks.

Let z be the variable of interest that is assumed to be a realisation of a second-order stationary Gaussian random function Z and let $\overline{z}(B_i) = {}^{1}\!/\!_{|B_i|} \int_{s \in B_i} z(\mathbf{s})d\mathbf{s}$ be the value of z at block support, where z($\mathbf{s}$) is the value of z at point location $\mathbf{s}$ and $|B_i|$ is the area of a block B indexed by i. Because the arithmetic averaging is linear in its argument, the random process at block support is also a Gaussian process.

Let $\boldsymbol{Z}_{\mathrm{p}} = (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_{\mathrm{M}}))^{\mathrm{T}}$ and $\bar{\boldsymbol{Z}}_{\mathrm{B}} = \left(\bar{Z}(\mathrm{B}_1),\ldots,\bar{Z}(\mathrm{B}_{\mathrm{N}})\right)^{\mathrm{T}}$ denote vectors of Z at point and block support, then their joint probability distribution is jointly Gaussian:

$$\begin{bmatrix} \boldsymbol{Z}_{\mathrm{p}} \\ \bar{\boldsymbol{Z}}_{\mathrm{B}} \end{bmatrix} \sim \mathrm{N}\left(\mu \begin{bmatrix} \mathbf{1}_{\mathrm{M}} \\ \mathbf{1}_{\mathrm{N}} \end{bmatrix}, \begin{bmatrix} \mathbf{C}_{\mathrm{pp}} & \mathbf{C}_{\mathrm{pB}} \\ \mathbf{C}_{\mathrm{Bp}} & \mathbf{C}_{\mathrm{BB}} \end{bmatrix} \right) \tag{4.1}$$

where μ is the constant spatial mean of Z, $\mathbf{1}_{\mathrm{M}}$ and $\mathbf{1}_{\mathrm{N}}$ are M and N vectors of ones, $\mathbf{C}_{\mathrm{pp}}$ is the M×M variance - covariance matrix of $\boldsymbol{Z}_{\mathrm{p}}$, $\mathbf{C}_{\mathrm{BB}}$ is the N×N variance-covariance matrix of $\bar{\boldsymbol{Z}}_{\mathrm{B}}$, $\mathbf{C}_{\mathrm{pB}}$ and $\mathbf{C}_{\mathrm{Bp}}$ are the variance-covariance matrix between $\boldsymbol{Z}_{\mathrm{p}}$ and $\bar{\boldsymbol{Z}}_{\mathrm{B}}$ and vice versa. Because their joint distribution is normal, the optimal predictor of $\boldsymbol{Z}_{\mathrm{p}}$ given $\bar{\boldsymbol{Z}}_{\mathrm{B}}$ is a linear combination of the BSO (Chilès and Delfiner, 1999, Section 3.3.4):

$$\hat{Z}_{\mathrm{p}} = \mu \mathbf{1}_{\mathrm{M}} + \mathbf{C}_{\mathrm{pB}} \mathbf{C}_{\mathrm{BB}}^{-1} \left(\bar{\boldsymbol{Z}}_{\mathrm{B}} - \mu \mathbf{1}_{\mathrm{N}} \right) \tag{4.2}$$

The variance-covariance matrix of the prediction error, called $\mathbf{C}(Z_{\mathrm{p}} - \hat{Z}_{\mathrm{p}})$, is given by:

$$\mathbf{C}\left(Z_{\mathrm{p}} - \hat{Z}_{\mathrm{p}} \right) = \mathbf{C}_{\mathrm{pp}} - \mathbf{C}_{\mathrm{pB}} \mathbf{C}_{\mathrm{BB}}^{-1} \mathbf{C}_{\mathrm{pB}}^{\mathrm{T}} \tag{4.3}$$

This shows that ATP kriging is straightforward and very similar to common kriging, but its main difficulty is that it requires the PoS variogram (Kyriakidis, 2004) to calculate the point-point, point-block and block-block variance-covariance matrices in Eq. 4.1, where the latter two require a regularisation (Journel and Huijbregts, 1978, Section II.D.4). Estimation of the PoS variogram from BSO is usually done using deregularisation or deconvolution (Journel and Huijbregts, 1978, Section II.D.4). Pardo-Igúzquiza and Atkinson (2007) introduce an iterative numerical deconvolution method to derive the PoS variogram from regular BSO (i.e. satellite imagery). In their study, the types of models included in the nested PoS variogram model are defined based on the nested variogram model fitted to the BSO. The optimisation condition is that the derived PoS variogram is the one minimising the difference between the theoretically regularised variogram model and the model fitted to the BSO. Goovaerts (2008) extends the method of Pardo-Igúzquiza and Atkinson (2007) to derive the PoS variogram from both regular and irregular (i.e. different size and shape) BSO. Gotway and Young (2007) present an iterative generalised estimation approach to estimate the parameters of the PoS covariance function and the trend surface using irregular BSO. Nagle et al. (2011) use maximum likelihood estimation for the PoS covariance function using BSO. Gelfand et al. (2001) address Bayesian estimation of PoS variogram parameters from BSO of a spatial-temporal process. Their study focuses on developing objective Bayesian inference methods, where the priors of the PoS variogram model parameters are given as noninformative priors. This is one of few studies that addresses PoS variogram estimation from BSO using a Bayesian approach.

In all aforementioned methods for deriving the PoS variogram, the nugget component of the PoS variogram was dismissed and assumed to be zero. There was surprisingly little attention on resolving the issue of inferring the nugget parameter from BSO, despite the material impact of the nugget variance on the ATP prediction and associated uncertainty (Kyriakidis, 2004). From the performance assessment of the iterative numerical deconvolution method using irregular BSO, Goovaerts (2008) concludes that the behaviour at the origin of the PoS variogram model (i.e. the nugget effect and within-block semivariance) could not be characterised with only BSO. Recently, Nagle et al. (2011) points out that the BSO retain little information to infer the nugget component of the PoS variogram and recommends using prior knowledge to overcome this problem.

The advantage of using a Bayesian approach is that the Bayesian estimator can quantify the uncertainty about the inference of the PoS variogram parameters. It is also the only formalised method to combine prior knowledge with BSO. However, extracting expert knowledge as informative priors is a delicate process in order to obtain reliable information. Much research has been done recently on using statistical expert elicitation to extract expert knowledge to use as informative priors for Bayesian statistical models, e.g. in Bayesian environmental and ecological modelling (Choy et al., 2009; Kuhnert et al., 2010; Kuhnert, 2011) and Bayesian geological modelling (Wood and Curtis, 2004). Formal statistical expert elicitation (Garthwaite et al., 2005; O'Hagan et al., 2006) provides transparent and reliable techniques to elicit from expert knowledge the probability distributions of the PoS variogram parameters to use as informative priors (Chapters 2 and 3). The statistical expert elicitation procedure comprises several structured stages: starting from defining the issues that require expert knowledge, finding experts, choosing an elicitation approach and doing the real elicitation task with experts to post-processing and using the statistical expert elicitation outcomes. There is increasing literature presenting detailed guidelines of developing and using statistical expert elicitation methods, e.g. Hahn (2006), Knol et al. (2010), Kuhnert et al. (2010), O'Hagan (2012) to name a few. This promises to be a sufficient solution for the issue of lacking information from BSO to infer the nugget component of the PoS variogram.

Our aim in this study was twofold. Firstly, we wanted to resolve the issue of poor estimation of the nugget effect from BSO by using a Bayesian approach that incorporates knowledge of multiple experts. Secondly, we wanted to quantify the propagation of PoS variogram parameters and ATP kriging model uncertainty to the disaggregated outcomes using Bayesian ATP conditional simulation. We illustrate the method with an example on disaggregating MODIS air temperature data measured on a coarse grid of 5 km resolution to a finer grid of 1 km resolution. To this end, the remainder of this paper has three main sections. Section 4.2 presents the statistical methods and a description of the example. Section 4.3 presents the main results of the study and a discussion. Section 4.4 provides the conclusions and recommendations for further research.

4.2. Materials and methods

Figure 4.1 shows the three main steps of the method.

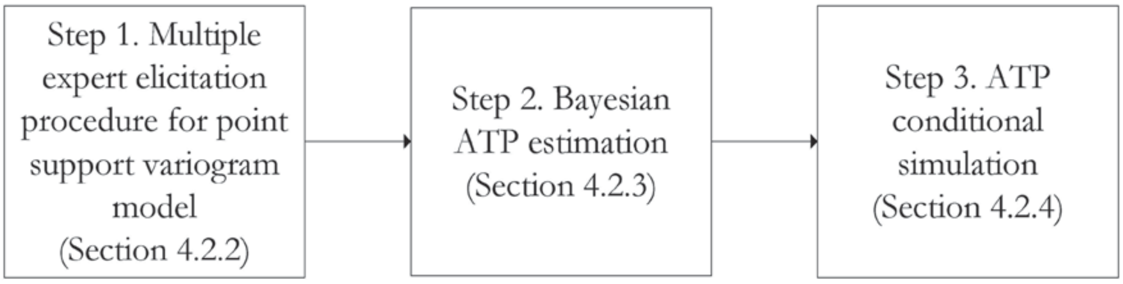

Figure 4.1: Three main steps of Bayesian area-to-point kriging method

4.2.1. Data

Spaceborne thermal imagery is becoming important in climate modelling, soil moisture assessment, irrigation management, etc. (Kuenzer et al., 2013a; Ha et al., 2013). Products of daily spaceborne thermal imagery often have lower spatial resolution (e.g., MODIS: 1 km to 5 km, NOAA-AVHRR: 1 km, Sentinel 3-ESA future mission: 1 km), whereas higher spatial resolution at several tenths of meters is often required, e.g. in precision agriculture or irrigation management at field level or in assessing urban heat effect (Kuenzer et al., 2013b). For these reasons and for illustration purposes, we used a dataset from the MODIS atmospheric temperature profile retrieval product (MOD07_L2) as BSO of a continuous variable. The study area was the Gelderland province, The Netherlands (Fig. 4.2). It is located in the east of The Netherlands with a total area of 5,137 km^2. The MOD07_L2 dataset (cloud-free, Collection 5) of August 1^{st}, 2012 at 10:05 a.m. UT (i.e. 12 a.m. CEST) at 5 km resolution was obtained from LAADS Web - Level 1 and Atmosphere Archive and Distribution System[9]. A single layer of the temperature profile at atmospheric pressure (100 kPa) was extracted and subset to the area of the Gelderland province (Fig. 4.3). The BSO of the study area had a mean of 21.6ºC and a standard deviation of 0.85ºC.

4.2.2. Statistical expert elicitation for prior distributions

Fig. 4.4 shows the five main steps of the formal statistical expert elicitation procedure to extract multiple expert knowledge for the prior distributions of the PoS variogram parameters.

In step 1, the Matérn variogram model (Eq. 4.4) (Stein, 1999) was chosen to characterise the spatial structure of the air temperature at PoS. We defined its four parameters (i.e. sill, nugget, range and smoothness parameters) as the target variables of the elicitation task. The smoothness parameter of the Matérn variogram model allows great flexibility to model the behaviour of the PoS variogram near the origin. The Matérn variogram model γ_p is given by:

$$\gamma_p(h) = c_0 + c_1 \{1 - [(2^{1-\nu}/\Gamma(\nu))(h/r)^{\nu} K_{\nu}(h/r)]\} \quad (4.4)$$

where h is the Euclidean distance, Γ is the gamma function, K_ν is the modified Bessel function of the second kind of order ν, c_0 is the nugget variance, c_1 is the partial sill variance, r is the range parameter, and ν is the smoothness parameter.

In step 2, we identified researchers from the meteorological domain who have knowledge about the temperature variation within the Gelderland province. We also required that experts have a basic understanding of geostatistics. The three Dutch researchers (two from Wageningen University and one from KNMI - Royal Netherlands Meteorological Institute) were recruited into the online elicitation task lasting for two weeks from January 3^{rd}, 2013 to January 17^{th}, 2013.

[9] http://ladsweb.nascom.nasa.gov/data/search.html.

Figure 4.2: Gelderland province, The Netherlands

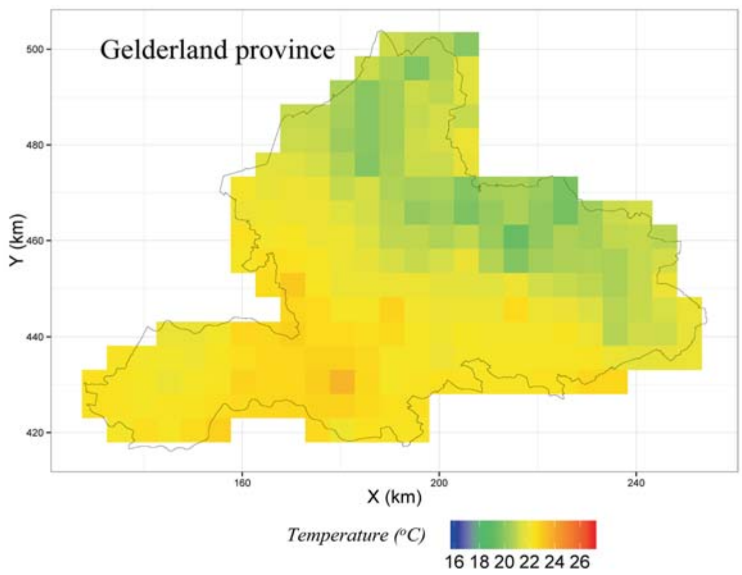

Figure 4.3: MODIS dataset (MOD07_L2) of Gelderland province

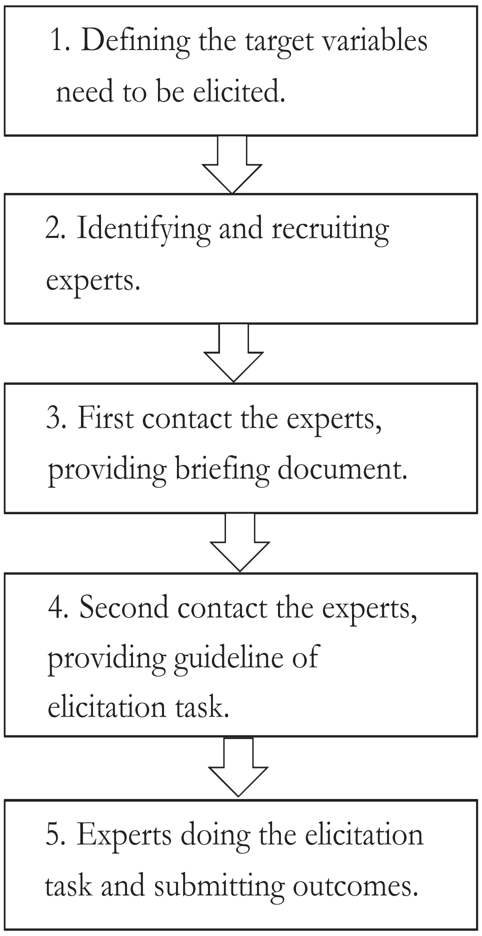

Figure 4.4: Formal statistical expert elicitation procedure

Step 3: we prepared a briefing document that explained the experts: (1) the study and all assumptions related to the dataset and the Matérn variogram model, (2) the statistical expert elicitation approach, (3) the requirements for the experts to elicit the vairogram parameters, and (4) the common causes of biased judgements. Because the experts have expertise in geostatistics, there was no explanation provided on statistical terminology. The experts were provided with the briefing document one week in advance. They were informed of the possibility to gather all possible sources of knowledge including experience, literature, field work, etc. to prepare for the elicitation task.

Step 4: we set up the elicitation task where the experts were facilitated by the MATCH Uncertainty elicitation online tool[10] (accessed 21/03/2013) to independently and individually judge the probability distributions of the PoS Matérn variogram

[10] http://optics.eee.nottingham.ac.uk/match/uncertainty.php

model parameters. Feedback on the elicitation outcomes was provided with the fitted probability density function (pdf) of each parameter (available on the MATCH tool) and the variogram generated from the pdfs of the four parameters. A guideline of how to proceed with the elicitation task using the MATCH tool was given. The experts were asked to spend time on self-training in giving probabilistic judgements with the MATCH tool. Thereby, we provided a week for the experts in total, of which only 45 minutes was recommended to complete the elicitation task for all four parameters.

Step 5: The final outcomes from the elicitation task were the summary statistics (maximum, minimum, mean and standard deviation) of the pdfs that best conveyed the experts' beliefs about the probability distributions of the parameters.

Three elicited pdfs from the three experts for each parameter were probabilistically averaged (i.e. using the equal weight pooling method) (Clemen and Winkler, 1999; O'Hagan et al., 2006; French, 2011) to derive a single prior probability density function for each parameter. We chose the equal weight pooling method because of its advantage of being simple, pragmatic and able to equally combine all uncertainty in all experts' knowledge. These combined pdfs were used as informative priors in the Bayesian ATP estimator.

4.2.3. Bayesian ATP estimator using Markov Chain Monte Carlo

Assuming that the spatial random process characterising air temperature at PoS, called Z, over the study area is a Gaussian isotropic second-order stationary random process, it is fully characterised by the mean and variogram: $\boldsymbol{Z}_p \sim \mathrm{N}(\mu, \gamma_p)$. The variogram $\gamma_P = \gamma_P(\mathrm{h}, \boldsymbol{\theta})$ is a function of the Euclidean distance h and the vector $\boldsymbol{\theta}$ of four parameters of the Matérn variogram. Block averages: $\bar{\boldsymbol{Z}}_B \sim \mathrm{N}(\mu, \gamma_B)$ have the same spatial mean (Gotway and Young, 2002) but a different spatial structure. γ_B is the block support variogram and a function of the distance h and the vector of parameters $\boldsymbol{\theta}$.

In Bayesian estimation, the joint posterior distribution of μ and $\boldsymbol{\theta}$ is related to the joint prior and the likelihood function via Bayes' theorem as in Eq. 4.5:

$$\mathrm{p}(\mu, \boldsymbol{\theta}|\cdot) \propto \pi(\mu, \boldsymbol{\theta})\mathrm{L}(\mu, \boldsymbol{\theta}|\cdot) \tag{4.5}$$

where $\mathrm{p}(\cdot)$ is the joint posterior distribution, $\pi(\cdot)$ is the joint prior distribution and $\mathrm{L}(\cdot)$ is the likelihood. The joint prior distribtion of $\boldsymbol{\theta}$ was derived from fitting a multivariate kernel density to the pooled pdfs of all four variogram parameters.

The likelihood of μ and $\boldsymbol{\theta}$ conditioning on BSO is given by:

$$L\left(\mu,\theta\middle|\bar{\mathbf{Z}}_B\right)=2\pi^{-N/2}\left|\mathbf{C}_{BB}\right|^{-1/2}\exp\left\{\left(-1/2\right)\left[\left(\bar{\mathbf{Z}}_B-\mu\mathbf{1}_N\right)^T\mathbf{C}_{BB}^{-1}\left(\bar{\mathbf{Z}}_B-\mu\mathbf{1}_N\right)\right]\right\}$$

Approximately numerical calculation of each item of $\mathbf{C}_{BB}$ is:

$$\mathbf{C}_{BB}\left[i,j\right]=\begin{cases} c_1/K+\left(2/K(K-1)\right)\sum_{k=1}^{K}\sum_{l>k}^{K}\left(c_1+c_0-\gamma_p\left(\mathbf{s}_k-\mathbf{s}_l\right)\right), i=j \\ \left(1/K^2\right)\sum_{k=1}^{K}\sum_{l=1}^{K}\left(c_1+c_0-\gamma_p\left(\mathbf{s}_k-\mathbf{s}_l\right)\right), i\neq j \end{cases}$$

where the i, j index the BSO (i, j = 1, …, N), the k and l index the discretised points within blocks and K is the number of discretisation points per block. Blocks were regularly discretised with equal spacing between discretisation points.

The 'Metropolis within Gibbs' or hybrid MCMC algorithm (Chib and Greenberg, 1995; Robert and Casella, 1999, Chapters 6 and 7; Albert, 2009, Chapter 6) that simultaneously uses both Gibbs sampling steps and Metropolis-Hastings steps was used to simulate the joint posterior distribution in Eq. 4.5. We used the Gibbs sampler to iteratively and alternatingly sample $\boldsymbol{\theta}$ from its full distribution conditional on μ and sample μ from its full distribution conditional on $\boldsymbol{\theta}$. To simulate a set of $\boldsymbol{\theta}$ from the full conditional distribution on μ, the Metropolis-Hastings algorithm was used because their joint conditional distribution cannot be sampled directly. The same approach was used for μ. We used Geweke's convergence diagnostic test (Geweke, 1992) to test the convergence of the MCMC chains to stationary distributions.

The hybrid MCMC was implemented in R (R Core Team, 2013), gstat package (Pebesma, 2004) for implementing ATP kriging, and the geweke.plot function of the CODA package for convergence diagnostics (Plummer et al., 2006). These computations resulted in successive sets of $\boldsymbol{\theta}$ and μ that were thinned to obtain a reasonably independent set to use as input to ATP conditional simulation (Section 4.2.4).

4.2.4. ATP conditional simulation

ATP conditional simulation enables to generate realisations of Z at PoS conditional on BSO $\bar{Z}_B$. We applied the approach of stochastic conditional simulation by generating unconditional simulation first and next conditioning these to observations using kriging of the differences between the BSO and simulated block arithmetic averages (Defouquet, 1994; Journel and Huijbregts, 1978, Section VII.A.1). Defining Z_{pcs} as the

ATP conditional simulation at PoS, it can be shown that:

$$Z_{\text{pcs}} = Z_{\text{ps}} + \Delta\hat{Z}_{\text{pk}} \tag{4.6}$$

where Z_{ps} is the PoS unconditional simulation and $\Delta\hat{Z}_{\text{pk}}$ is ATP simple kriging of the difference between BSO and the simulated block arithmetic average $\bar{Z}_{\text{Bs}}$.

Practical implementation of this approach comprised four steps:

Step 1. Unconditional simulation at PoS prediction locations: unconditional sequential Gaussian simulation was used to generate realisations on a regular 1 km grid for each set of PoS variogram parameters derived from the Bayesian ATP estimation (Section 4.2.3).

Step 2. Aggregating each PoS unconditional simulation to block support to obtain $\bar{Z}_{\text{Bs}}$ and calculating the difference: $\Delta\bar{Z} = \bar{Z}_{\text{B}} - \bar{Z}_{\text{Bs}}$.

Step 3. ATP simple kriging of $\Delta\bar{Z}$ at PoS prediction locations to obtain $\Delta\hat{Z}_{\text{pk}}$ (Eq. 4.2).

Step 4. Summing Z_{ps} and $\Delta\hat{Z}_{\text{pk}}$ (Eq. 4.6) to obtain ATP conditional simulation final outcomes Z_{pcs}.

All four steps were implemented in R (R Core Team, 2013).

4.3. Results and Discussion

4.3.1. Informative priors from multiple expert knowledge

We anonymously report the elicitation outcomes from the three experts, named Expert 1, Expert 2 and Expert 3. Figs. 4.5 - 4.8 show the pdfs of the four parameters of the PoS Matérn variogram model that, according to the experts, best convey their opinion about the probability distributions of these parameters. Tables 4.1 provides summary statistics of these pdfs.

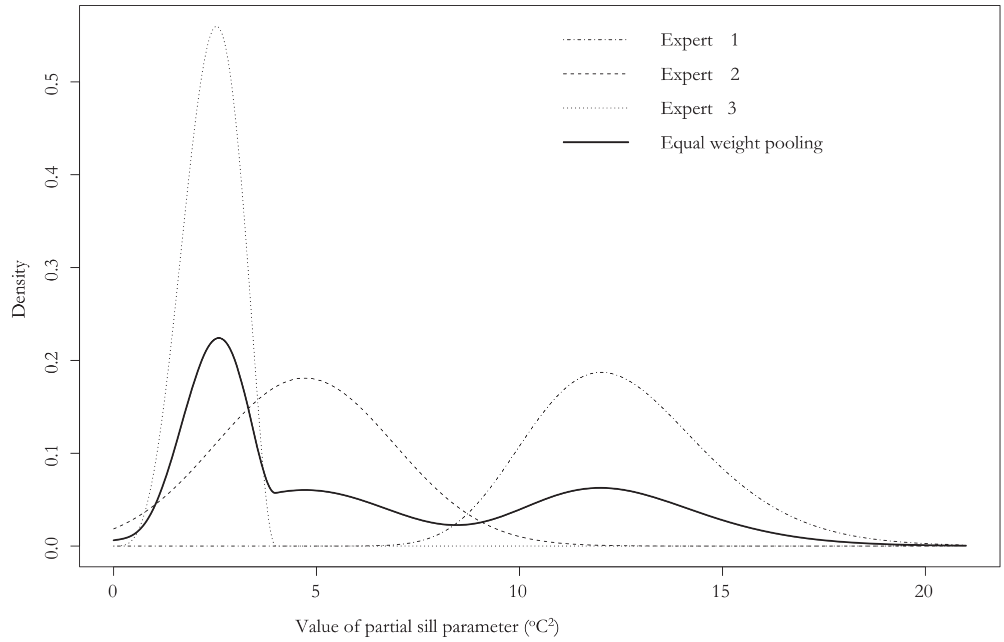

Figure 4.5: Probability density functions of partial sill parameter from three experts and equal weight pooling

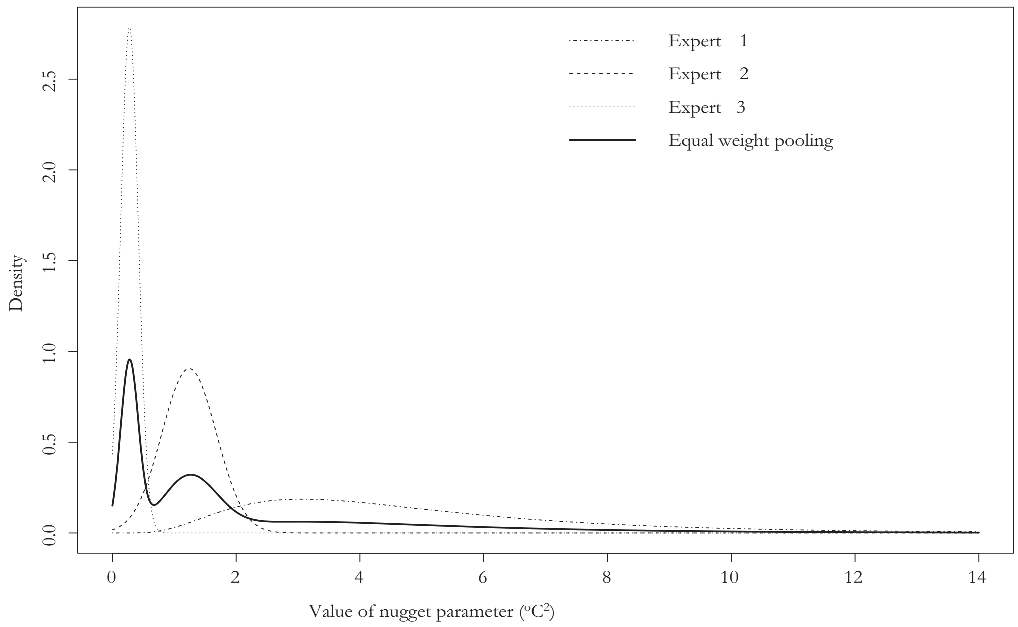

Figure 4.6: Probability density function of nugget parameter from three experts and equal weight pooling

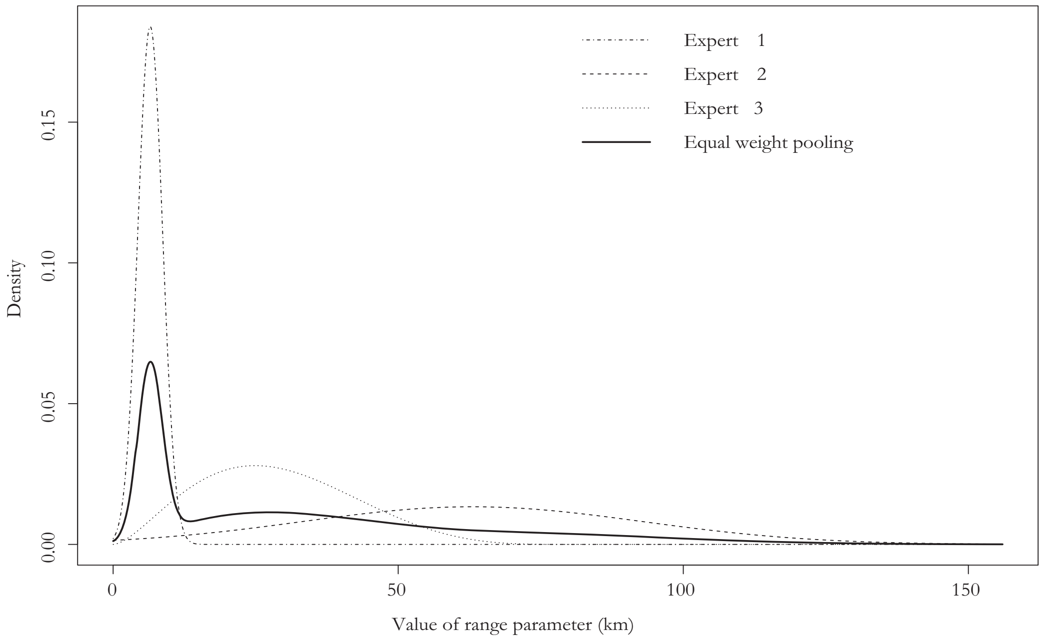

Figure 4.7: Probability density function of range parameter from three experts and equal weight pooling

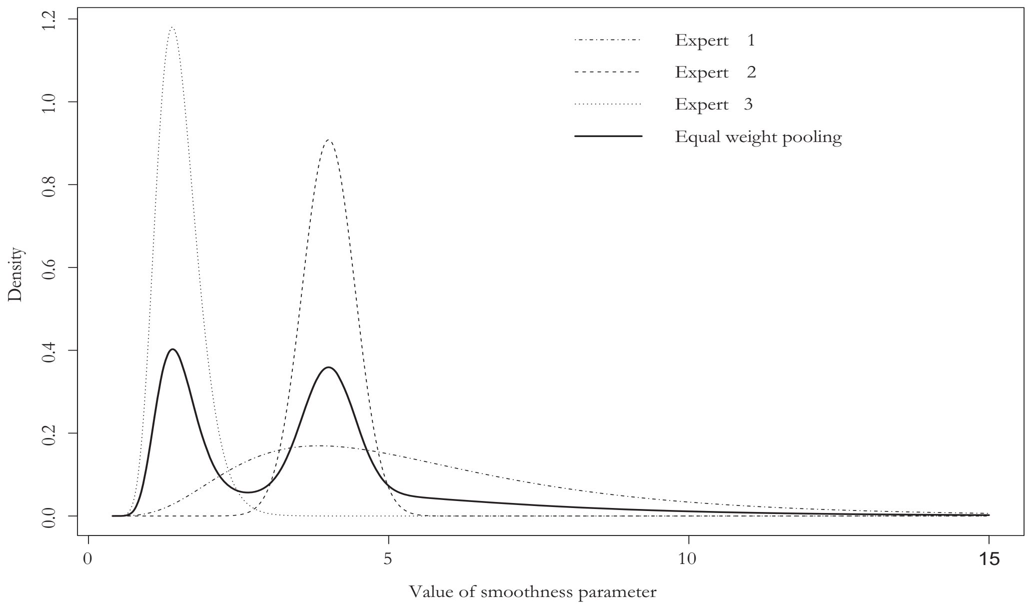

Figure 4.8: Probability density function of kappa parameter from three experts and equal weight pooling

Table 4.1: Elicited outcomes from experts

Parameters	Pdf	Minimum	Maximum	Mean	Standard deviation
Expert 1					
Partial sill (°C^2)	Log-normal	9.0	16.0	2.5	0.17
Nugget (°C^2)	Log-normal	0.5	9.0	1.5	0.59
Range (km)	Normal	3.0	10.0	6.5	2.17
Kappa	Log-normal	1.0	10.0	1.6	0.53
Expert 2					
Partial sill (°C^2)	Normal	1.0	10.0	4.7	2.20
Nugget (°C^2)	Normal	0.0	3.0	1.2	0.44
Range (km)	Normal	25.0	151.0	63.0	29.84
Kappa	Normal	3.0	5.0	4.0	0.44
Expert 3					
Partial sill (°C^2)	Scaled-beta	0.0	4.0	4.7	3.14
Nugget (°C^2)	Normal	0.0	1.0	0.3	0.14
Range (km)	Scaled-beta	0.0	80.0	2.7	4.69
Kappa	Log-normal	0.4	4.0	0.4	0.24

The experts had quite different opinions about the central value and uncertainty of the parameters of the PoS variogram. In particular, there was little agreement among the experts about the values of the sill and the nugget parameters. Their pdfs have modes located at different values. The different widths of the pdfs indicated different level of experts' uncertainty about these parameters. The variograms in Fig. 4.9 show a better agreement among the experts for the effective range (i.e. the distance where the variogram reaches 95% of its total sill), although this agreement is not apparent from Figs. 4.7 and 4.8 for the range and smoothness parameters. The reason is that there is a correlation between the smoothness parameter and the range parameter: a large range parameter corresponds to a small smoothness parameter and vice versa, which when combined, results in similar variograms. Although the effective ranges are similar, the variograms from Experts 1 and 2 approach the origin differently with different combinations of smoothness and range. Compared to the extent of the study area, the judgements of Expert 2 resulted in a very smooth spatial process, while those of Expert 1 resulted in a much more rough spatial process.

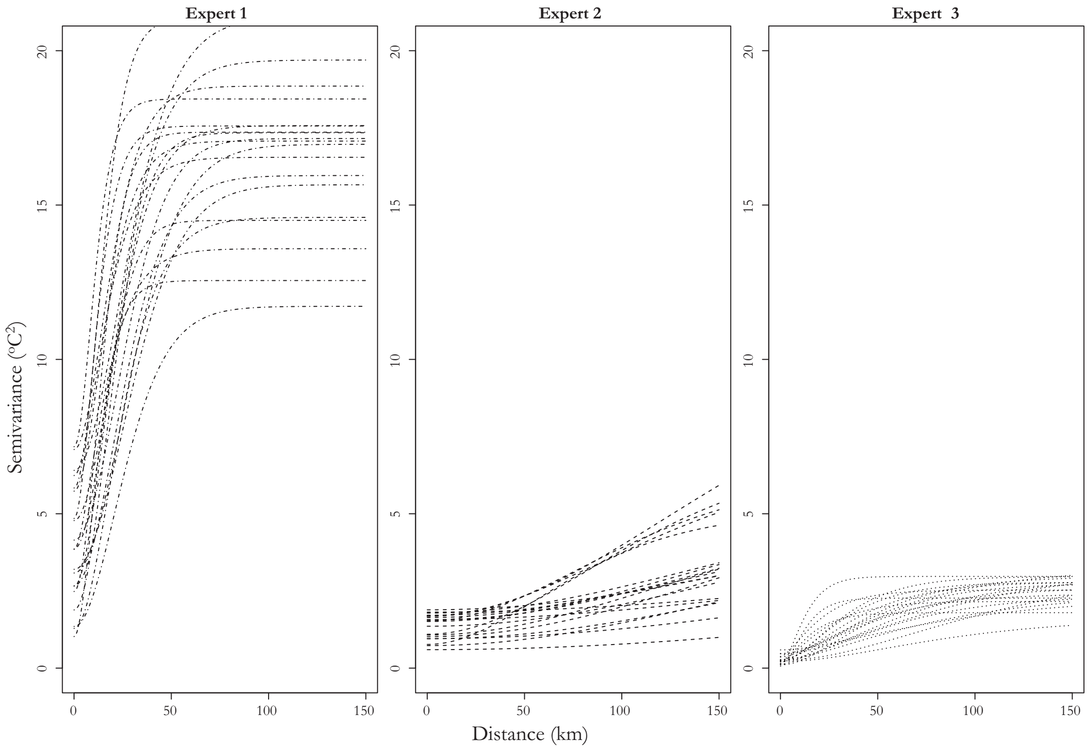

Figure 4.9: Experts' prior variograms

The elicitation procedure had the form of a Delphi elicitation process for multiple experts (Kuhnert et al., 2010), of which the feedback of the pdfs and the variogram provided a mechanism for calibrating experts' judgements regarding the best expression of their opinions. The difference from the traditional Delphi method is that the outcomes from each expert were not given as feedback to other experts. This kept maximum diversity in expert opinions. To validate the experts' judgements, we had measured data from only three KNMI meteorological stations located across the Gelderland province. The data were too few to make an accurate estimation of the PoS variogram for validation. Using KNMI station data of the whole of The Netherlands, the PoS variogram derived from these data (see Salet, 2009) agreed well with those of Expert 2 and Expert 3. However, this comparison is not adequate in terms of spatial and temporal extent. In fact, we consulted the experts because of insufficient point support data available to map the air temperature at PoS for the Gelderland province.

Although Expert 1 had a remarkably different opinion about the sill and nugget

4

parameters compared with other experts, we still find the judgements informative because of the reasoning behind the judgements of this expert. Disagreements in experts' judgements that result from differences in experts' perception or in weighing and combining various sources of knowledge and information can bring more information but can also imply large uncertainty in the current experts' knowledge. Even though the experts have expertise in geostatistics, prior knowledge elicitation of the parameters of a statistical model (e.g. elicit multiple quartiles of model parameters) is a more difficult task than elicitation of an observable object (Lele and Das, 2000). This can partly explain the large uncertainty and diversity in the experts' judgments.

4.3.2. Bayesian ATP estimation of point support Matérn variogram model

We first present and examine results of the MCMC simulations. Recall that the prior distribution for $\boldsymbol{\theta}$ was the joint distribution of the pooled pdf of all parameters (Section 4.3.1). The prior for μ was a wide, uninformative uniform distribution ranging between 10°C and 40°C. Trace plots in Fig. 4.10 show well-mixing outcomes with an acceptance rate of 0.5 for μ and 0.25 for $\boldsymbol{\theta}$ that satisfy the recommended acceptance rate for one-dimensional and multi-dimensional parameters (Chib and Greenberg, 1995). Geweke's diagnostic test results (Fig. 4.11) show that the MCMC chain converged using 30,000 runs. Because the main focus of this work is on the PoS variogram model's parameters, there will be no more elaboration on μ.

The pdfs of the PoS variogram parameters were plotted together with their informative prior distributions in Figs. 4.12 and 4.13. The posterior distribution of the nugget parameter is almost the same as its prior distribution, except some erratic behaviour at the tail due to its very heavy tail (Fig. 4.12). This result confirms that the BSO did not provide information to identify the nugget component of the PoS variogram and corroborates the findings of Goovaerts (2008) and Nagle et al. (2011). In case of the nugget variance, qualified expert knowledge becomes the only source of information. Even though, there was considerable uncertainty in the experts' opinions about the nugget parameter, the elicited informative prior brought valuable information to the Bayesian ATP estimator.

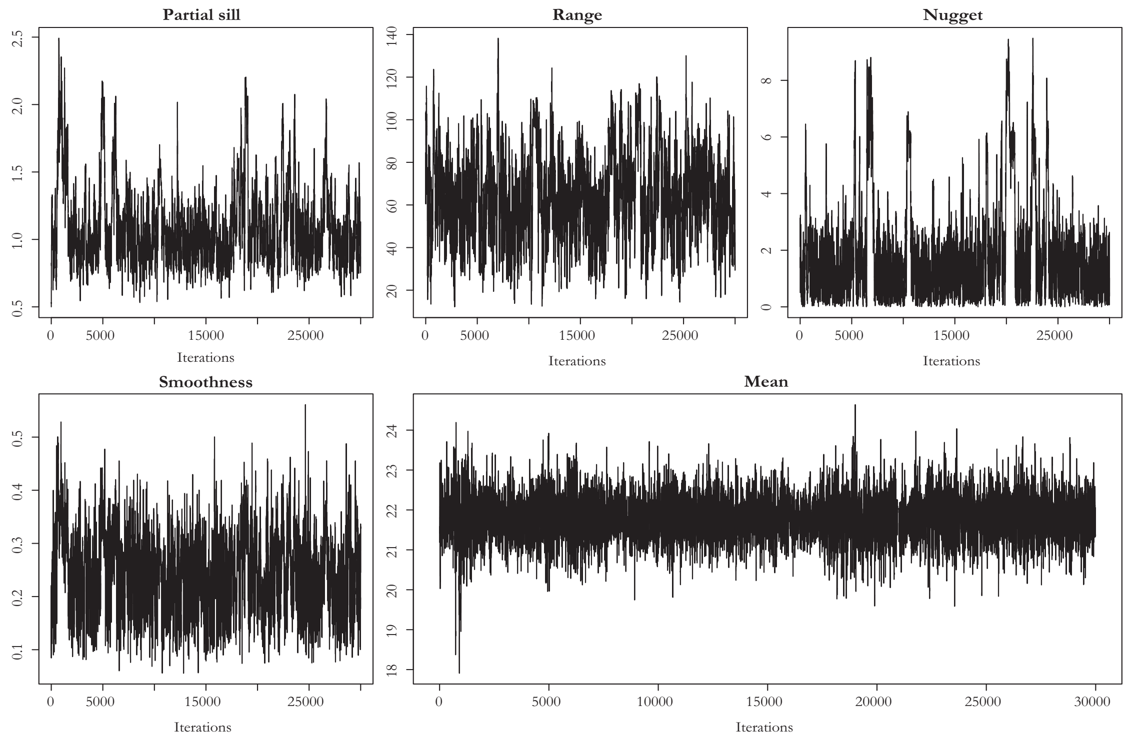

Figure 4.10: Trace plots for point support Matérn variogram model parameters and spatial mean

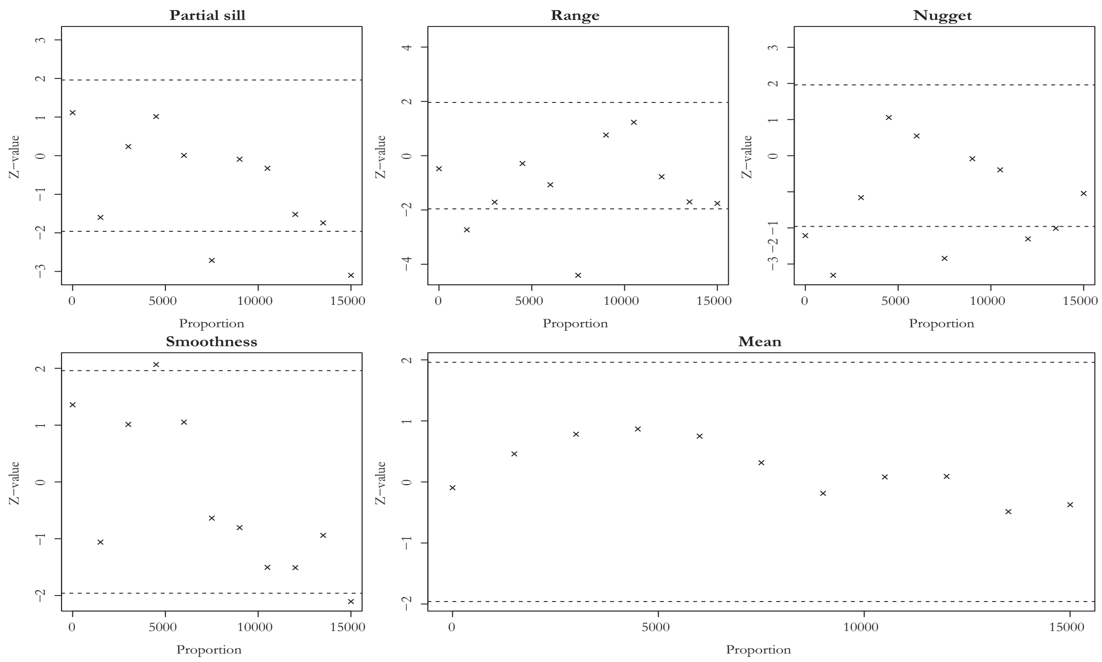

Figure 4.11: Results of Geweke's convergence diagnostic test

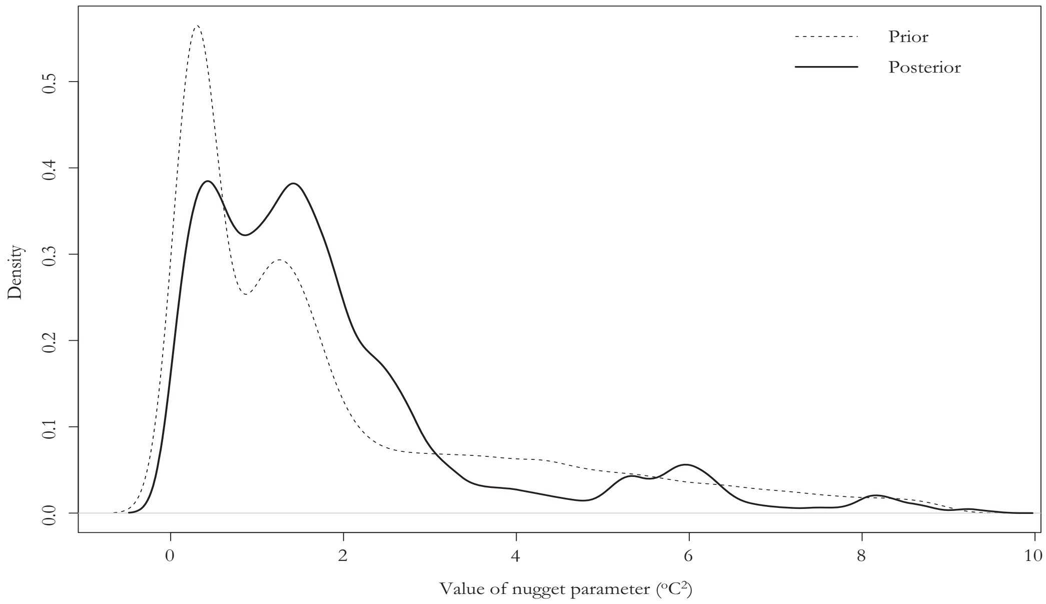

Figure 4.12: Informative prior and posterior probability distribution of nugget parameter

The posterior distributions of the other PoS variogram model parameters deviated largely from their priors and were indeed much narrower. This implies that estimates of the other parameters were strongly driven by BSO. Fig. 4.14 shows the posterior variograms that were generated by several random sets of the four parameters. The total sill of the posterior PoS variogram is much larger than the total variance of BSO (0.73°C²), which agrees with the fact that block support data are smoother and have smaller variance than point support data (Gotway and Young, 2002). The effective range was about 50 km in average. Fig. 4.15 shows a negative correlation between the range and smoothness parameter and high positive correlation between the partial sill and the range parameter.

4.3.3. ATP conditional simulations

Fig. 4.16 shows maps of the outcomes from 1,000 ATP conditional simulations with 1,000 inputs of PoS variogram parameters. These inputs were obtained from the MCMC chains by thinning every 30th iteration to obtain a reasonably uncorrelated set. The mean of the conditional simulations ranged from 18.8°C to 24°C. All simulations satisfied the coherence (or mass-preserving) property of ATP kriging. The maps of the 5th (lower limit) and 95th quantile values (upper limit) at each simulation node (Fig. 4.16) present the 90% symmetric prediction interval of the simulations at PoS. The number of simulations was not enough to obtain a stable estimate of the variance of the ATP simulations. However, the variance of the ATP simulations was larger than the ATP kriging variance with a fixed PoS variogram that used the modes of the joint posterior of the variogram parameters (Fig. 4.17). This shows that uncertainty about the PoS variogram parameters can make a substantial contribution to the uncertainty about the PoS values at prediction nodes.

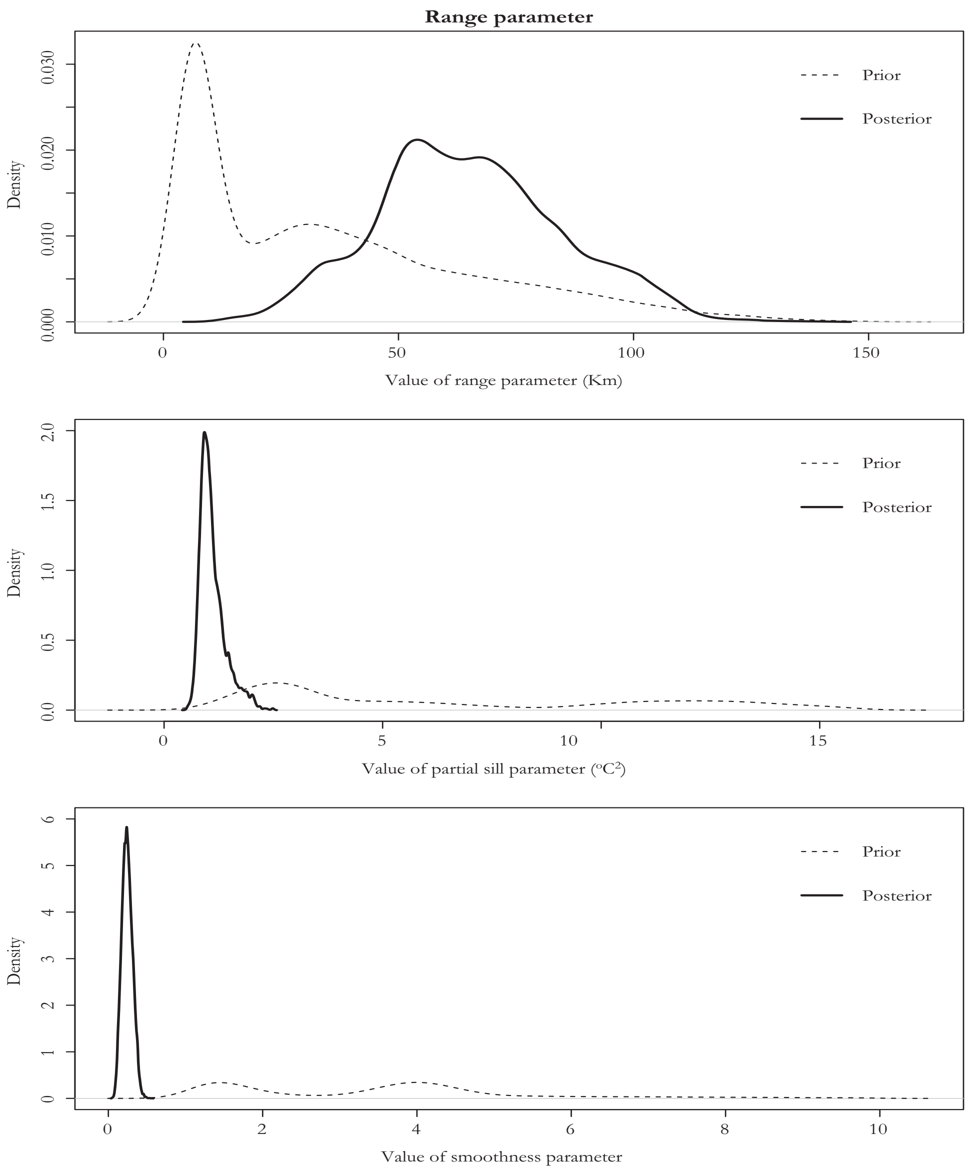

Figure 4.13: Informative prior and posterior distributions of partial sill, range and smoothness parameters

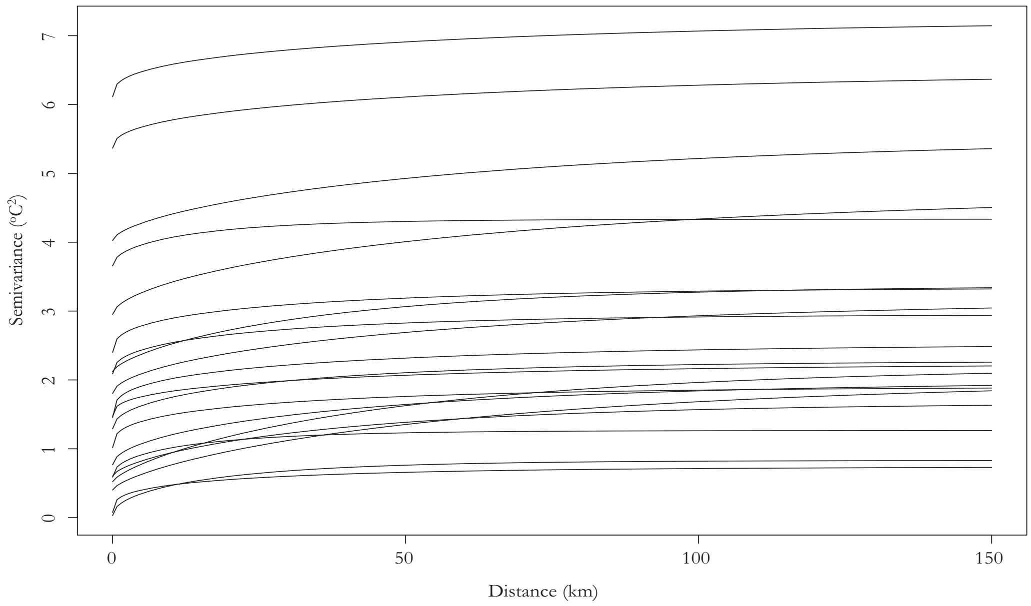

Figure 4.14: Posterior point support variograms

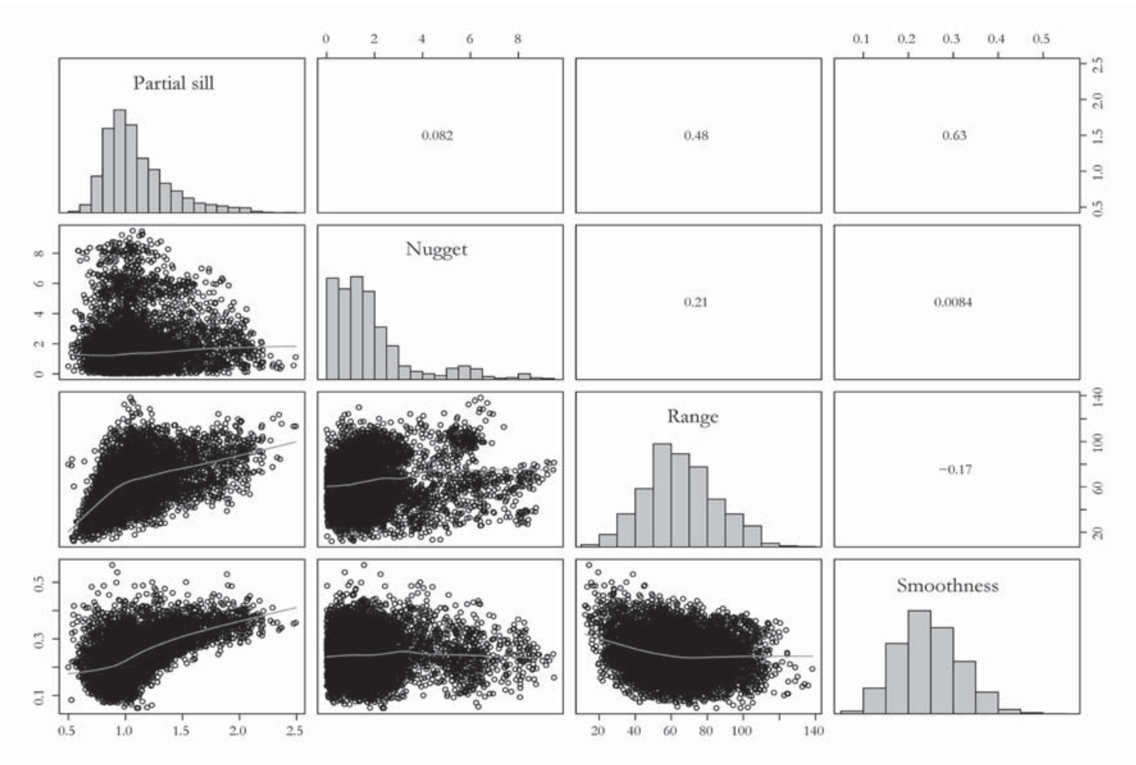

Figure 4.15: Correlation between point support Matérn variogram model parameters

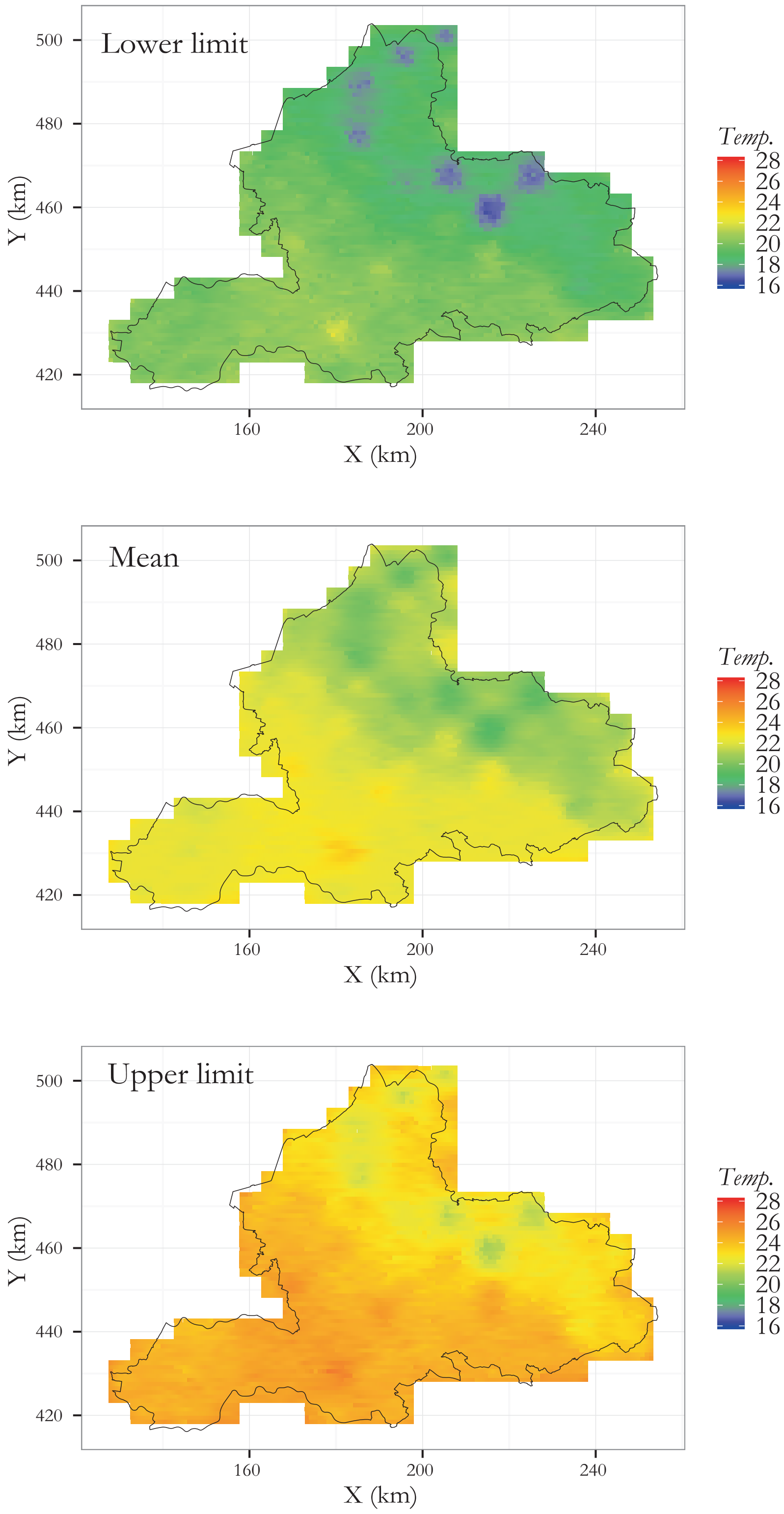

Figure 4.16: Bayesian ATP conditional simulation outcomes (°C)

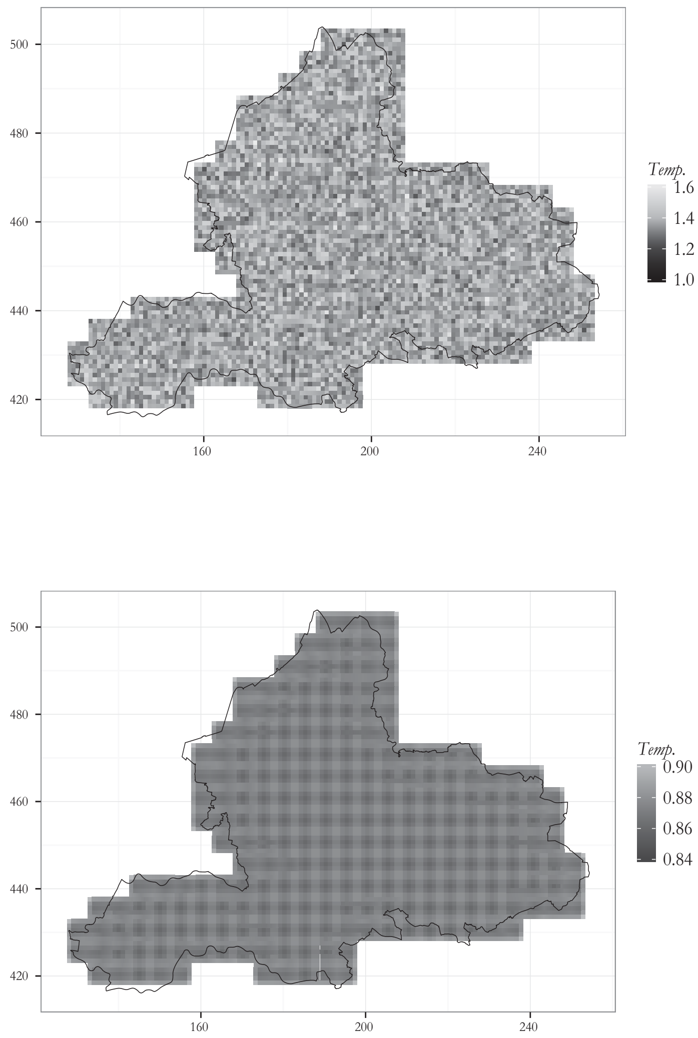

Figure 4.17: Standard deviation of ATP conditional simulations (top) and that of ATP Kriging with modal variogram (bottom) (°C)

4.4. Conclusions and Recommendations

ATP kriging provides a methodological solution to allow maximum use of available information in BSO to derive predictions at finer resolution or spatial support, but it does not introduce new sources of information (Atkinson, 2013). In this study, we introduced a new source of information, i.e. by using expert knowledge via informative priors of the PoS variogram parameters. By using the Bayesian ATP estimator, we have shown that the nugget of the PoS variogram cannot be estimated by only BSO, this ascertains similar findings from previous studies. Without (sufficient) observations at PoS, expert knowledge is the best or perhaps only source of information available about the nugget effect at PoS. For the other parameters, the posterior distributions are narrower than the priors because the BSO did provide real information, which also affected the correlations between these parameters. The example also showed that uncertainty about the PoS variogram parameters can substantially contribute to the overall ATP kriging prediction uncertainty.

The example illustrated that our proposed approach worked well, both in theory and practice. By this, we were able to derive an appropriate estimator of the PoS variogram parameters to perform spatial disaggregation with ATP kriging. It is worth mentioning that the prior knowledge derived by statistical expert elicitation in the form of probability distributions is a delicate task and to keep in mind that the elicitation outcomes are always imprecise (O'Hagan and Oakley, 2004). In this study, we took the view that the best statistical expert elicitation task is the one that can elicit the true opinions of the experts. Although our direct elicitation approach for the PoS variogram parameters was structured and transparent, we noticed that it was a fairly inconvenient elicitation task because the variogram model parameter can be quite abstract to experts. Moreover, biases in expert opinion can directly distort the elicitation outcomes; this impact might be less when using indirect elicitation. We recommend that indirect elicitation approaches, such as those proposed in Chapter 2, should be explored and extended to derive the prior probability distribution of the PoS variogram parameters in future research. ATP conditional simulation was time-consuming because all BSO and already simulated nodes at PoS were used as conditioning data in ATP conditional simulation (i.e. we used global ATP conditional simuation). Implementing ATP conditional simulation with a local neighbourhood or using the circulant embedding method (Dietrich and Newsam, 1993) will help to speed up the simulation.

Chapter 5

Incorporating expert knowledge as observations in mapping biological soil quality indicators with regression cokriging

5.1. Introduction

Geostatistics furnishes a statistical tool - kriging - for spatial prediction at unobserved locations by interpolation from observations. Under specific assumptions, kriging has the smallest prediction error variance among all unbiased spatial interpolation methods (Armstrong, 1998, Chaper 7; Schabenberger and Gotway, 2005, Chapter 5). In geostatistics, reality is treated as if it were a realisation of a random process that is characterised by a probability distribution, which includes the spatial dependence structure (i.e. the variogram). The variogram is estimated from the observations using the methods-of-moments (Matheron, 1963; Cressie, 1985) or maximum likelihood estimation (Pardo-Igúzquiza, 1997; Pardo-Igúzquiza, 1998). Very often, the same observations are used both for modelling the spatial dependence and kriging.

Statistical inference of variogram parameters is problematic when there are few observations because the sampling variance is often large and little information about short distance spatial variation is available (Webster and Oliver, 2007, Section 6.1.2). Also, prediction by kriging with few observations typically results in large prediction error variances because of the low sampling density (Frogbrook, 1999). In short, using few observations for kriging produces inaccurate maps. Unfortunately, in many practical cases, only few observations may be available for mapping because of, for instance, poorly accessible areas or laboriously and expensively measured variables. An alternative to collecting extra measurements of the target variable to improve mapping is using ancillary data and/or soft data that are more easily or more cheaply obtained. The ancillary data could be point observations of auxiliary variables that are spatially correlated with the variable of interest as used in cokriging (Goovaerts and Kerry, 2010) or spatially exhaustive covariate layers used in regression kriging or kriging with external drift (Goovaerts, 1997, Chapter 6; Hengl et al., 2007) while soft data are typically associated with the observations of the target variable that have (larger) measurement error. Such soft data could be measurements with a cheaper and less accurate instrument (e.g. Hamzehpour et al., 2013), but these could also be qualified guesses (Journel, 1986; Goovaerts, 1997, Chapter 6). Soft data can be incorporated in kriging using cokriging and Bayesian kriging (Omre,1987; Goovaerts, 1997, Chapter 6).

There have been few studies that use expert knowledge as soft data to supplement direct measurements in kriging. Journel (1986) suggests the use of qualified guesses of plausible ranges of values in indicator cokriging, but without further investigation on the actual use of expert knowledge. Omre (1987) incorporates qualified guesses of the expected surface of a spatial process with uncertainty as prior information into so-called Bayesian kriging. It suggests a method to incorporate qualified expert judgements but does not provide a description of how expert judgements are obtained nor an evaluation of the use of expert knowledge for kriging. Lele and Das (2000) and Lele and Allen (2006) use expert knowledge as exact estimates of the target variable. The influence of the correlation between expert estimates and true observations on the estimation of the spatial structure (i.e. the trend and the spatial correlation) is investigated, but the method used to extract expert judgements is not elaborated. These studies show the potential of using expert knowledge as supplementary data to direct measurements of the target spatial variable in mapping, but they lack a clear and formalised way to extract expert knowledge, and the important fact that expert judgements are uncertain is often ignored.

In this chapter, we propose to use expert knowledge as probabilistic soft data to augment direct measurements to map spatial variables by cokriging. To our knowledge, the use of expert knowledge in this way has not been investigated before. Probabilistic soft data are soft data recorded in the form of a probability distribution. We use soft data given by experts to improve variogram estimation and kriging. The rationale behind this is that expert knowledge can provide a valuable source of information about spatial variability (Stein, 1994; Chapters 2 and 3). Statistical expert elicitation provides reliable techniques to extract and formulate expert knowledge in a probabilistic form (Garthwaite et al., 2005; O'Hagan et al., 2006). However, data derived from expert knowledge cannot simply be treated as additional observations, and their different characteristics must be recognised in geostatistical inference and prediction. The quality of the expert data was assessed through a comparison to the mean and variance of direct measurements. The efficiency of the use of expert knowledge was evaluated using a case study of mapping biological soil quality indicators in a semi-natural grassland in The Netherlands.

During the last decades, biological soil indicators have been increasingly used not only to indicate soil threats such as soil degradation and contamination but also to indicate the ability of the soil to provide valuable ecosystem services such as the production of food and fibre, climate change mitigation by carbon sequestration and habitat forming for species rich natural ecosystems (Wall, 2004). Nematodes are the most frequently used biological soil quality indicators (Wilson and Kakouli-Duarte, 2009). In the case study, we selected two biological soil quality indicators: the structure index (SI) and enrichment index (EI). Nematodes occur mostly in high abundances and are positioned in the soil food web at all different trophic levels: herbivores, microbivores, omnivores and predators (Yeates, 2003). Differences in life-history traits of nematode species form the basis of the SI and EI (Ferris et al., 2001). Nematodes can be grouped into (rapid growing) enrichment opportunists and general opportunists and (slow growing) persisters. The ratio between enrichment opportunists and general opportunists is the EI, and the ratio between the persisters and general opportunists is the SI (Ferris et al., 2001). A high EI indicates high soil fertility, i.e. high nutrient availability or eutrophication, and a low SI indicates low soil quality due to environmental stress or disturbance (de Goede et al., 1993; Ferris et al., 2001). Although the conceptual framework of these indices is well developed and current ways to assess the EI and SI are technically adequate, direct measurements are still expensive and time-consuming. Moreover, soil biological properties like nematode assemblages may vary strongly over space (Ettema and Wardle, 2002). Therefore, the EI and SI are relevant variables for which the value of incorporating expert knowledge may be investigated.

Section 5.2 describes the geostatistical model used for incorporating expert knowledge in spatial inference and prediction. Section 5.3 presents the case study and explains how expert knowledge was extracted at specific locations in probabilistic form by statistical expert elicitation. The main findings of this work are presented and discussed in Section 5.4.

5.2. Methods

5.2.1. Model definitions

We consider the case where the variable of interest is modelled by a spatial Gaussian random field, for which we have two different datasets: measurement data collected in the field and expert data collected from expert judgements. Furthermore, we assume that the measured data are error-free, while the expert data can be biased and imprecise and have different error variances that measure the expert uncertainty about the values of the target variable at specific locations. Let the vector $\mathbf{Z}_1$ represent the measured data at n locations $\mathbf{s}_i$, i = 1, …, n within a spatial domain D and let $\mathbf{Z}_2$ represent the expert data at m locations $\mathbf{s}_j$, j = 1, …, m. Note that, in this study, the $\mathbf{s}_i$ are a subset of the $\mathbf{s}_j$; hence, m>n.

We define the model for the measured data as $Z_1(\mathbf{s}) = \mathbf{F}(\mathbf{s})\boldsymbol{\beta}_1 + e_1(\mathbf{s})$, where: $\boldsymbol{\beta}_1 = \{\beta_{1k}, k = 0, \ldots, p\}$ is a (p+1) column vector of the unknown regression coefficients, $\mathbf{F}(\mathbf{s}) = \begin{Bmatrix} f_0(\mathbf{s}) = 0, k = 0 \\ f_k(\mathbf{s}) \quad , k = 1, \ldots, p \end{Bmatrix}$ is a (p+1) vector of covariates at any location s in D, and e_1 is a zero-mean second-order stationary Gaussian random field. We denote the n×n variance-covariance matrix of the $\mathbf{Z}_1$ at the measurement locations by $\mathbf{C}_{11}$.

The model for the expert data is defined as $Z_2(\mathbf{s}_j) = Z_1(\mathbf{s}_j) + m_\varepsilon(\mathbf{s}_j) + \varepsilon(\mathbf{s}_j) - \eta e_1(\mathbf{s}_j)$, where the zero-mean Gaussian random variables $\varepsilon(\mathbf{s}_j)$ represent the random errors of the expert data. It is assumed that the $\varepsilon(\mathbf{s}_j)$ are mutually independent and independent of $e_1(\mathbf{s}_j)$. The $m_\varepsilon(\mathbf{s}_j)$ represents the systematic error (bias) of the expert data, which is taken as a sum of a constant and a linear combination of the covariates, to capture systematic differences in interpretation by the expert of the influence of the covariates on the target (dependent) variable. The term $-\eta e_1(\mathbf{s}_j)$ measures the conditional bias toward the mean (i.e. the smoothing effect of the expert judgements), with η a coefficient between 0 and 1. Z_1 and Z_2 have the same set of covariates because they model the same variable of interest. Substituting the model of Z_1 in that of Z_2 gives $Z_2(\mathbf{s}_j) = \mathbf{F}(\mathbf{s}_j)\boldsymbol{\beta}_2 + e_2(\mathbf{s}_j)$, where $\boldsymbol{\beta}_2 = \{\beta_{2k}, k = 0, \ldots, p\}$ is a (p+1) column vector of unknown regression coefficients and $e_2(\mathbf{s}) = \varepsilon(\mathbf{s}) + (1 - \eta)e_1(\mathbf{s}) \sim N(0, \mathbf{C}_{22})$, with $\mathbf{C}_{22}$ a variance-covariance matrix derived from the covariance structures of e_1 and ε. The possible systematic bias of the expert data is, as specified, a linear function of the covariates: $m_\varepsilon(\mathbf{s}) = \mathbf{F}(\mathbf{s})[\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1]$.

5.2.2. REML estimation

To estimate the parameters of the models defined in Section 5.2.1, we used the REML method (Pardo-Igúzquiza, 1997; Webster and Oliver, 2007, Section 9.2). In case of a non-stationary spatial trend, REML estimates of covariance parameters of a Gaussian random field are unbiased and require fewer data to reach a given accuracy level compared to maximum likelihood estimation and the method-of-moments (Schabenberger and Gotway, 2005, Section 5.5; Lark et al., 2006). Recall that the locations of expert data include the locations of the measured data. Because of this, the trend and the sill of the cross-covariance function can be estimated (Chilès and Delfiner, 1999, Section 5.4; Wackernagel, 2003, Section 23).

In order to apply REML, we first need to remove the non-stationary trend. Hence, we transform $\mathbf{Z}_1$ to $\mathbf{Z}_1^* = \mathbf{A}_1\mathbf{Z}_1$, where $\mathbf{A}_1$ is an n−(p+1)×n matrix formed by removing the last p+1 rows of $\mathbf{I}_n - \mathbf{F}_1\left(\mathbf{F}_1^T\mathbf{F}_1\right)^{-1}\mathbf{F}_1^T$, $\mathbf{Z}_1^*$ is an n−(p+1) column vector, $\mathbf{F}_1 = \mathbf{F}(\mathbf{s}_i)$ is the n×(p+1) design matrix of $\mathbf{Z}_1$, $\mathbf{I}_n$ is the n×n identity matrix. Similarly, for $\mathbf{Z}_2$, we use the m−(p+1)×m projection matrix $\mathbf{A}_2$ by removing the last p+1 rows of $\mathbf{I}_m - \mathbf{F}_2\left(\mathbf{F}_2^T\mathbf{F}_2\right)^{-1}\mathbf{F}_2^T$, $\mathbf{Z}_2^* = \mathbf{A}_2\mathbf{Z}_2$ is an m−(p+1) column vector, $\mathbf{F}_2 = \mathbf{F}(\mathbf{s}_j)$ is the m×(p+1) design matrix and $\mathbf{I}_m$ is the m×m identity matrix.

Because $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are linearly transformed, $\mathbf{Z}_1^*$ and $\mathbf{Z}_2^*$ are also jointly normally distributed and both have zero mean due to the trend removal transformation. The likelihood is derived from the multivariate Gaussian probability density function:

$$\mathrm{L}\left(\boldsymbol{\theta},\eta\middle|\mathbf{Z}_1^*,\mathbf{Z}_2^*\right) = 2\pi^{-N/2}\left|\boldsymbol{\Sigma}^*\right|^{-1/2}\exp\left\{\left(-1/2\right)\left[\mathbf{Z}^{*T}\boldsymbol{\Sigma}^{*-1}\mathbf{Z}^*\right]\right\}$$

The log-likelihood is:

$$\log\mathrm{L}\left(\boldsymbol{\theta},\eta\middle|\mathbf{Z}_1^*,\mathbf{Z}_2^*\right) = \left(-1/2\right)\left[\mathrm{N}\log\left(2\pi\right)+\log\left|\boldsymbol{\Sigma}^*\right|+\mathbf{Z}^{*T}\boldsymbol{\Sigma}^{*-1}\mathbf{Z}^*\right] \tag{5.1}$$

where: N=n+m−2(p + 1), $\mathbf{Z}^* = \begin{bmatrix}\mathbf{Z}_1^*\\ \mathbf{Z}_2^*\end{bmatrix} = \begin{bmatrix}\mathbf{A}_1\mathbf{Z}_1\\ \mathbf{A}_2\mathbf{Z}_2\end{bmatrix}, \Sigma^* = \begin{bmatrix}\mathbf{C}_{11}^* & \mathbf{C}_{12}^*\\ \mathbf{C}_{21}^* & \mathbf{C}_{22}^*\end{bmatrix}$ with $\mathbf{C}_{11}^* = \mathbf{A}_1\mathbf{C}_{11}\mathbf{A}_1^T$.

The variance-covariance matrix of $\mathbf{Z}_1$ is

$\mathbf{C}_{11}\left[\mathbf{s}_i,\mathbf{s}_i+\mathrm{h}\right] = \begin{cases}\tau^2+\sigma^2, \mathrm{h}=0\\ \mathrm{c}\left(\sigma^2,\phi,\mathbf{s}_i+\mathrm{h}\right), \mathrm{h}\neq 0\end{cases}$; h is Euclidean distance in the spatial domain D, $\mathbf{C}_{22}^* = \mathbf{A}_2\mathbf{C}_{22}\mathbf{A}_2^T$ with the variance-covariance matrix of $\mathbf{Z}_2$ given by

$\mathbf{C}_{22}\left[\mathbf{s}_j,\mathbf{s}_j+\mathrm{h}\right] = \begin{cases}\sigma_\varepsilon^2\left(\mathbf{s}_j\right)+\left(1-\eta\right)^2\left(\tau^2+\sigma^2\right), \mathrm{h}=0\\ \left(1-\eta\right)^2\mathrm{c}\left(\sigma^2,\phi,\mathbf{s}_j+\mathrm{h}\right), \mathrm{h}\neq 0\end{cases}$, $\mathbf{C}_{12}^* = \mathbf{A}_1\mathbf{C}_{12}\mathbf{A}_2^T$ and $\mathbf{C}_{21}^* = \mathbf{A}_2\mathbf{C}_{21}\mathbf{A}_1^T$

with the cross covariance matrices between $\mathbf{Z}_1$ and $\mathbf{Z}_2$ and vice versa:

$\mathbf{C}_{12}\left[\mathbf{s}_i,\mathbf{s}_j\right] = \mathbf{C}_{12}\left[\mathbf{s}_i,\mathbf{s}_j\right]^T = \begin{cases}\left(1-\eta\right)\sigma^2, \mathbf{s}_i\equiv\mathbf{s}_j\\ \left(1-\eta\right)\mathrm{c}\left(\sigma^2,\phi,\mathbf{s}_i-\mathbf{s}_j\right), \mathbf{s}_i\neq\mathbf{s}_j\end{cases}$; i=1,…,n;j=1,…,m. $\sigma_\varepsilon^2(\mathbf{s})$ is the known variance of expert data dependent on location $\mathbf{s}$, c(·) is the covariance function of the parameter vector $\boldsymbol{\theta} = \{\sigma^2, \phi, \tau^2\}$, with τ^2 the nugget parameter, σ^2 the partial sill parameter and ϕ the range parameter.

The estimates of the unknown parameters τ^2, σ^2, ϕ and η that maximize the log-likelihood (Eq. 5.1) were derived numerically and used to estimate the $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ by generalized least squares: $\hat{\boldsymbol{\beta}}_1 = \left(\mathbf{F}_1^{\mathrm{T}}\hat{\mathbf{C}}_{11}^{-1}\mathbf{F}_1\right)^{-1}\mathbf{F}_1^{\mathrm{T}}\hat{\mathbf{C}}_{11}^{-1}\mathbf{Z}_1$ and $\hat{\boldsymbol{\beta}}_2 = \left(\mathbf{F}_2^{\mathrm{T}}\hat{\mathbf{C}}_{22}^{-1}\mathbf{F}_2\right)^{-1}\mathbf{F}_2^{\mathrm{T}}\hat{\mathbf{C}}_{22}^{-1}\mathbf{Z}_2$.

5.2.3. Plug-in regression cokriging

We used the REML estimated parameters in regression cokriging (RCK) (Stein and Corsten, 1991; Lark et al., 2006). The best linear unbiased predictor of the variable of interest at an unobserved location $\mathbf{s}_0$ is: $\hat{}_{\mathrm{RCK}}\left(\mathbf{s}_0\right) = \mathbf{F}_0\hat{\boldsymbol{\beta}}_1 + \hat{\boldsymbol{\Sigma}}_0^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{\ 1}\begin{bmatrix}\mathbf{Z}_1 & \mathbf{F}_1\hat{}_1\\ \mathbf{Z}_2 & \mathbf{F}_2\hat{}_2\end{bmatrix}$ where:

$\mathbf{F}_0 = \mathbf{F}(\mathbf{s}_0)$ is a (p+1) row vector of the covariates at location $\mathbf{s}_0$, $\hat{\boldsymbol{\Sigma}} = \begin{bmatrix}\hat{\mathbf{C}}_{11} & \hat{\mathbf{C}}_{12}\\ \hat{\mathbf{C}}_{21} & \hat{\mathbf{C}}_{22}\end{bmatrix}$ is the (n+m)×(n+m) estimated variance-covariance matrix of the expert data and the measured data, and $\hat{\boldsymbol{\Sigma}}_0 = \begin{bmatrix}\hat{\mathbf{C}}_{11}\left(\mathbf{s}_i, \mathbf{s}_0\right)\\ \hat{\mathbf{C}}_{21}\left(\mathbf{s}_j, \mathbf{s}_0\right)\end{bmatrix}$ the (n+m)×1 variance-covariance matrix between the expert data, measured data and the target variable at $\mathbf{s}_0$. The regression cokriging variance is given by: $\hat{\sigma}_{\mathrm{RCK}}^2\left(\mathbf{s}_0\right) = \hat{\mathbf{C}}_{11}\left(\mathbf{s}_0, \mathbf{s}_0\right) - \hat{\boldsymbol{\Sigma}}_0^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Sigma}}_0 + \left(\mathbf{F}_0 - \mathbf{F}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Sigma}}_0\right)^{\mathrm{T}}\left(\mathbf{F}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{F}\right)^{-1}\left(\mathbf{F}_0 - \mathbf{F}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Sigma}}_0\right)$ with $\mathbf{F} = \begin{bmatrix}\mathbf{F}_1\\ \mathbf{F}_2\end{bmatrix}$.

All estimation and prediction methods were implemented in R (R Core Team, 2013). In particular, the DEoptim package (Mullen et al., 2011), the gstat package (Pebesma, 2004) and the geoR package (Diggle and Ribeiro, 2007) were used.

5.3. Case study

To illustrate the methods presented in Section 5.2, we used a case study on mapping the spatial variation of the biological soil quality indicators by nematodes. All free-living nematode taxa are allocated to functional guilds according to their feeding behaviour and coloniser-persister (cp) classification (Ferris et al., 2001). To calculate the SI and EI, nematodes are allocated to three components:

- Basal component (b) comprising bacterivores of cp2 and fungivores of cp2;

- Enrichment component (e) including all cp1 nematodes (bacterivores) and fungivores of cp2;

- Structure component (s) representing all cp3-5 nematodes as well as predators of cp2.

A weight is assigned to each functional guild based on the hypothesis of constant connectance in the community food web. Weights are 0.8 for cp2, 1.8 for cp3, 3.2 for cp1 and cp4, and 5.0 for cp5 (Ferris et al. 2001). Weighed combinations of functional guilds are used to infer the SI and EI. EI is calculated as e/(b+e)×100% and SI is calculated as s/(b+s)×100%.

The study area (Fig. 5.1) is about 23 ha and located in the Malpiebeemden nature reserve in the south of The Netherlands. The main vegetation type is grassland with some patches of forest. The patches with forest were excluded. The area is used for extensive cattle grazing. At the east, the area borders a small river (river Dommel) that was heavily polluted with heavy metals until recently. Due to flooding, part of the grassland area became polluted with e.g. cadmium and zinc.

Figure 5.1: Study area located in the Malpiebeemden nature reserve

5.3.1. Soil sampling and nematode identification

Soil samples were collected using stratified simple random sampling. Probability sampling was used because the measurement data were also used for validation (see Section 5.3.4), and it can improve the accuracy of validation measures compared to simple random or systematic sampling schemes (de Gruijter et al., 2006, Section 7.2.4). The study area was divided into eight strata with equal surface using the k-means clustering algorithm as provided in the R-spcosa package (Walvoort et al., 2010). In each stratum, two locations were randomly selected. This yielded in total 16 locations to be used as measured data (i.e., $\mathbf{Z}_1$ as defined in Section 5.2.1). In addition, eight other locations were randomly selected within each stratum for validation (see Section 5.3.4). Fig. 5.2 shows the strata and the sampling locations where in total 80 measurements were collected.

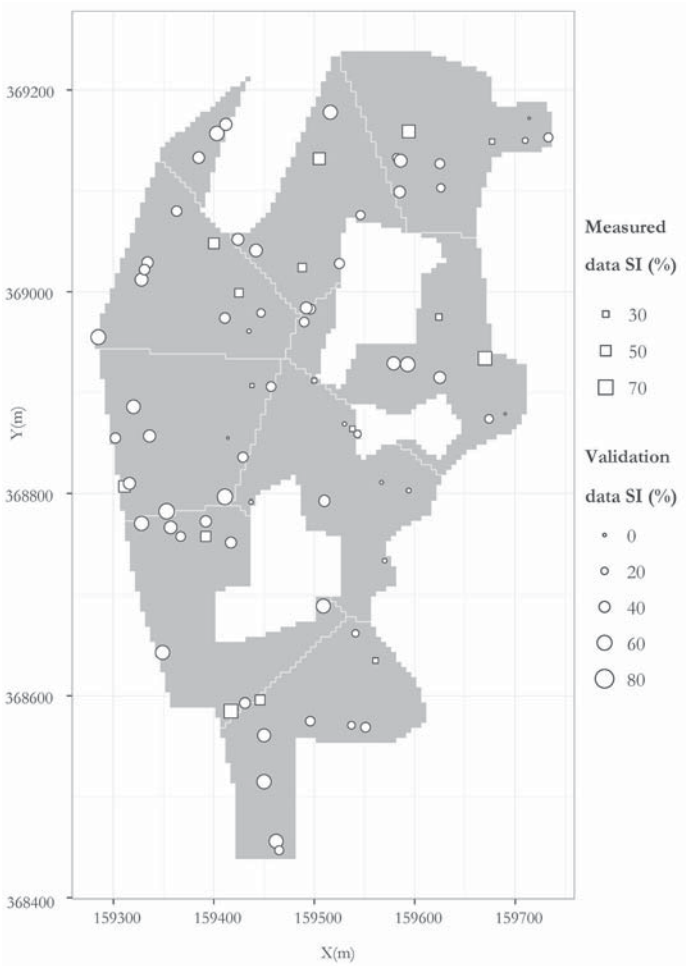

Figure 5.2: Soil sampling scheme for measured data (as defined in Section 5.2.1) and for validation data with SI values

Soil samples were collected on March 20th, 2013 at 0-20 cm depth using a 4 cm diameter soil corer. The nematodes in each of the 80 soil samples were extracted using an Oostenbrink elutriator (Oostenbrink, 1960), within three weeks after sampling. Nematodes were identified to genus or family based on morphological features and identification keys (see Ikoyi (2013) for more details).

5.3.2. Sampling scheme for expert elicitation

For expert elicitation, we added 34 locations to the available 16 locations where measurements were taken to make in total 50 locations. The 34 additional locations were uniformly spread over the study area as well as at short distances to better infer short distance variation in all directions (we assumed isotropic variation because there is no reason for anisotropic variation of the regression residual). For this purpose, we first used spatial coverage sampling (Walvoort et al., 2010) to spread as uniformly as possible 19 sampling locations, conditional to the available 16 measurement locations. The remaining 15 locations were added at short distances (about 15 meters) from measurement locations in all directions. The final number of 66 observations (16 measured data and 50 expert data) is accepted as an appropriate sample size for variogram inference using REML (Kerry and Oliver, 2007). Fig. 5.3 shows all locations for expert elicitation within the study area.

5.3.3. Expert elicitation procedure

A web-based tool[11] was built to facilitate the expert elicitation procedure. Fig. 5.4 shows the main page of the web-based tool. We recruited one expert who is a soil ecologist with hands-on experience in nematode biomonitoring, including data interpretation. The expert has never visited the study area and received an allowance for the job. The expert was given access to the web-based tool where information about the study area, the sampling scheme, soil conditions of the study area (maps and tables, see Appendix 5.A, with soil properties: moisture content, organic matter content, pH-water, clay content, total concentrations of cadmium and zinc, elevation, and distance to river), and explanation of the probability expert elicitation procedure was given.

The expert was asked to quantify the most probable range of the indicator values (EI and SI) at each of the 50 selected locations, using an elicitation technique that directly elicits the upper and lower quartiles of the probability distribution of the

[11] http://variogram.isric.org/

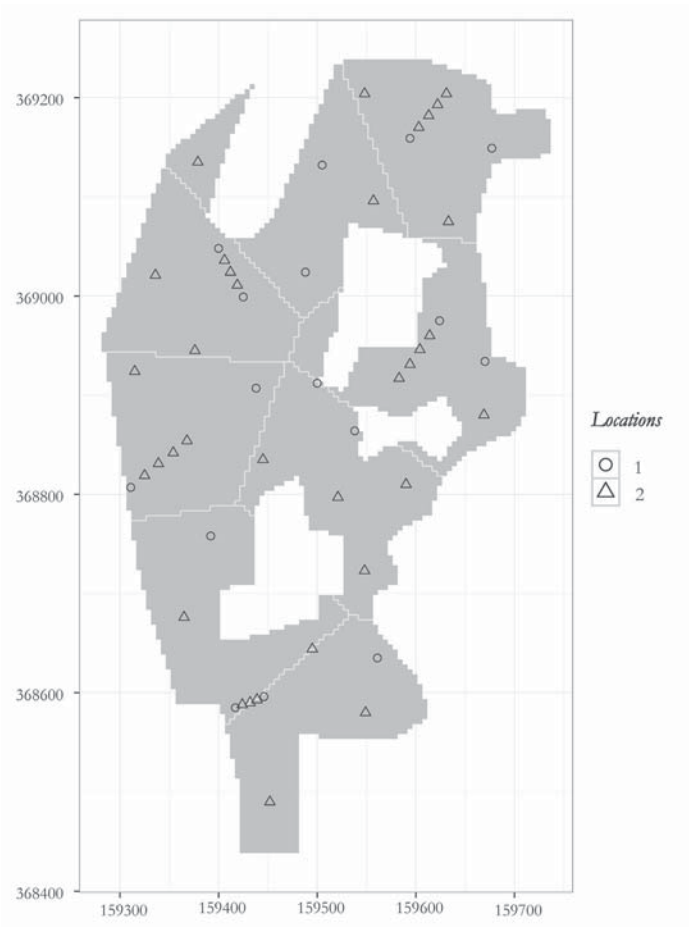

Figure 5.3: Sampling scheme for expert elicitation of nematode community indices: 1. Locations with measured and expert data (16 locations), 2. Locations with only expert data (34 locations)

indicator values (O'Hagan, 1998; Bedford and Walls, 2010). The expert was asked to work on the exercise during one working day including getting familiar with the tool, the soil conditions in the study area, the statistical background on estimating the values of the two nematode indices, and quantifying the upper and lower quartiles of SI and EI. The elicitation exercise was finished on December 1st, 2013.

In Section 5.4, we report and analyse only the results of SI from the elicitation exercise because the EI did not show spatial correlation from the direct measurements. The soil properties that have a high correlation with SI from the direct measured data are pH-water, soil moisture (%) and organic matter content (%). Therefore, the spatial trend of SI was modelled as a linear function of these three covariates:

pH-water (f_1), soil moisture (f_2) and organic matter content (f_3). The exhaustive maps of the three covariates over the study area (Fig. 5.5) to be used for regression were interpolated by ordinary kriging on nodes of a 5m-grid map covering the study area from 100 soil samples that were collected and analysed in 2008.

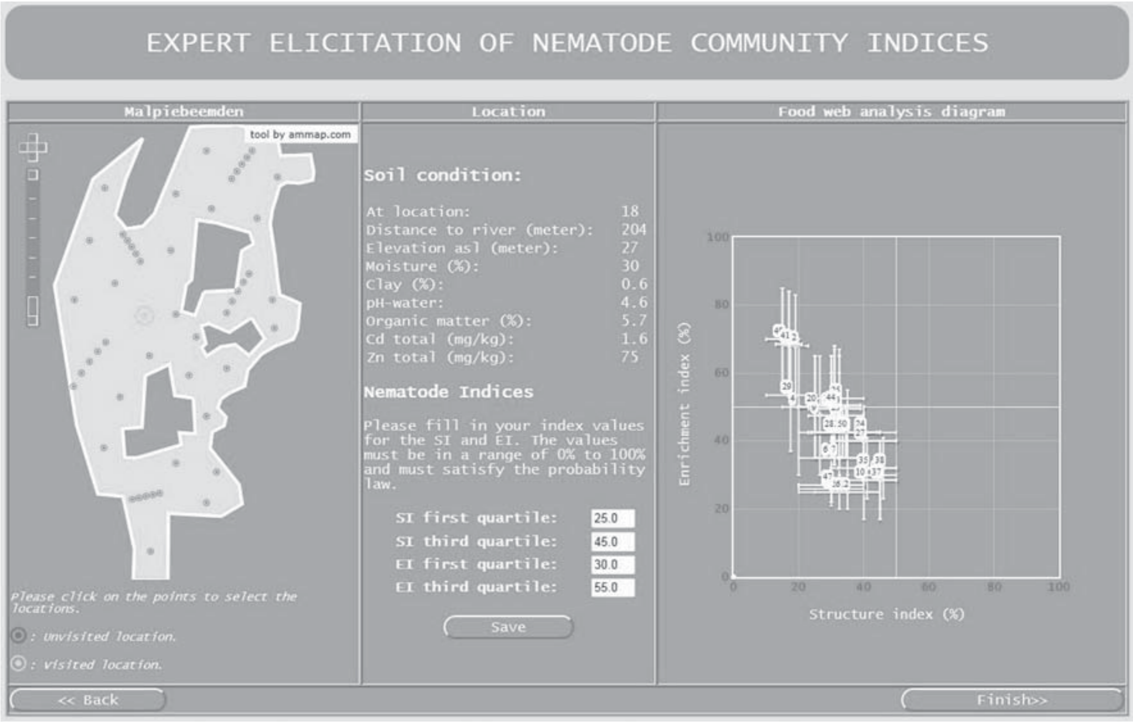

Figure 5.4: Main page of web-based tool for expert elicitation of nematode structure and enrichment indices

5.3.4. Accuracy assessment

Prediction accuracy was quantified by the kriging standard deviation and by validation. The two validation measures used for accuracy quantification of the prediction are the root mean squared error (RMSE) and the spatial mean of the kriging standard deviation ($\overline{\sigma}$). To estimate the RMSE and its confidence interval, we used the design-based statistical inference of stratified simple random sampling (de Gruijter et al., 2006, Section 7.2.4).

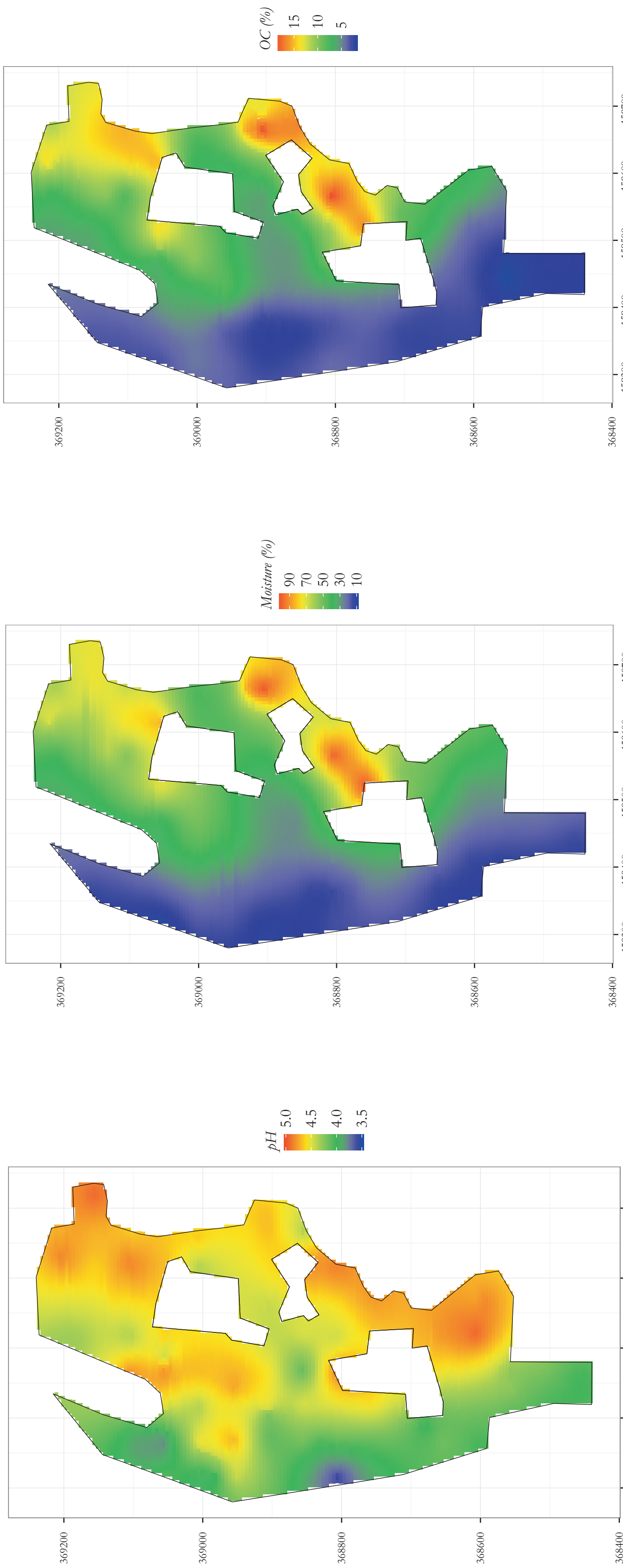

Figure 5.5: Exhaustive maps of SI covariates: pH-water (left), soil moisture (middle) and organic matter content (right)

The accuracy of the RCK prediction (Section 5.2.3) was quantified with the two measures: $\bar{\sigma}_{RCK} = \frac{1}{64}\sum_{i=1}^{64}\sigma_{RCK}(\mathbf{s}_v)$ and $RMSE_{RCK} = \sqrt{\frac{1}{64}\sum_{i=1}^{64}\left(\hat{Z}_{RCK}(s_v) - Z(s_v)\right)^2}$, where $\mathbf{s}_v$, v = 1, …, 64 are the validation locations. For the purpose of comparison, we also did regression kriging (RK) using only the measured data with the same three covariates (Section 5.3.3) and quantified the accuracy of the RK prediction using the same measures. The 95% confidence intervals of the $RMSE_{RCK}$ and $RMSE_{RK}$ were also calculated to analyse if the performances of RCK and RK are significantly different: $RMSE_{RCK(RK)} \pm 1.998\sqrt{var\left(RMSE_{RCK(RK)}\right)}$, with 1.998 is the 0.975 quantile of the Student distribution with 63 degrees of freedom.

5

5.4. Results and Discussion

5.4.1. Expert data of SI

Fig. 5.6 shows the mean and standard deviation of SI that were calculated from the elicited upper and lower quartiles of the SI (see Appendix 5.A) under a normal distribution assumption. The mean value of the coefficient of variation over the study areas equals to 0.38 (minimum of 0.13 and maximum of 0.64). This is a relatively high variation in the expert's quantification, which can be caused by the expert's lack of knowledge on the values of SI at specific locations, given the fact that the expert has never done fieldwork in this area. The information about soil conditions at specific locations were given to help the expert distinguish the spatial differences, particularly between nearby locations, but this demanded the expert the ability to link and weigh all information in the judgements. The number of locations for which the expert gave judgements is quite large; as a result, the expert had to quantify in total 100 values during the exercise. These two factors can also contribute to high uncertainty in expert judgements. Over all locations, the interquartile ranges are more or less equal, which might reflect the overall uncertainty of the expert about the SI in the study area, rather than only at any specific location.

Fig. 5.6 (left) shows that nearby locations tend to have similar mean values of SI. The magnitude of the spatial autocorrelation is presented in Section 5.4.2. The mean values of the SI also show a slight trend with higher values in the west and lower values in the east. There is a slightly positive linear correlation (Pearson correlation coefficient equals 0.35, p-value < 0.05) between the mean and standard deviation of

SI. This means that the expert is more uncertain at locations where the expert thought that SI is large. At co-located locations, the mean values of the expert data are on average 15.5% smaller than the measured data. When the means were used as $\mathbf{Z}_2$ in the model (Section 5.2), the magnitude of this systematic bias could be quantified. The Pearson correlation coefficient between the mean values of expert and measured data at co-located locations is 0.11 (p-value = 0.347). This indicates a positive linear relationship between the measured data and the expected values of the expert data, but the correlation is not strong.

Figure 5.6: Nematode structure index estimated from expert judgements: mean value (left) and standard deviation (right)

The weak relationship between the mean values of SI from expert data and the measured data at co-located locations and high variation in SI values of expert data would influence the efficiency of using the expert data in RCK. Specifying the mean values as $\mathbf{Z}_2$ and the standard deviation as σ_ε is legitimate but apparently has consequences. We discuss this in Section 5.4.3.

5.4.2. REML estimation

Numerical optimization of the log-likelihood (Eq. 5.1) yielded REML estimates of the covariance parameters $\boldsymbol{\theta}$ and the coefficient η. The measure of the conditional bias of the mean of the expert data towards the spatial mean compared to that of the measured data is $\hat{\eta} = 0.95$. This is a high value but a reasonable value as can be seen in Fig. 5.7, which shows less spatial variation in the mean of SI from expert judgements than the measured data. The reason can be that the expert tends to condition quantification on the overall mean value of SI over study area. Note that the η coefficient is the comparative quantification of the conditional bias of the mean of the expert data at every location compared to that of the measured data. This must be distinguished (as internal bias) from the external bias m_ε of the difference between the two expected surfaces of $\mathbf{Z}_1$ and $\mathbf{Z}_2$.

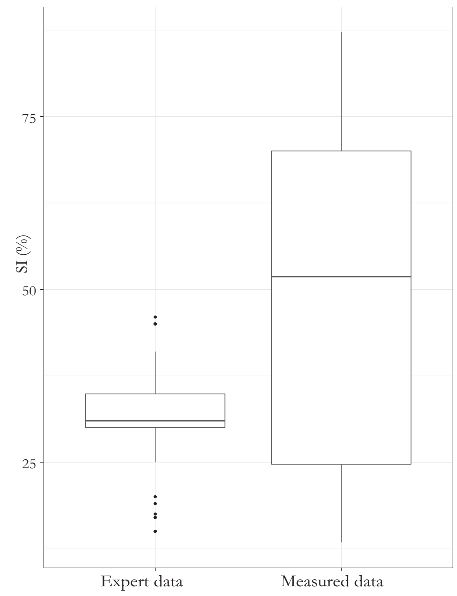

Figure 5.7: Boxplot of expert data versus that of measured data

The covariance function of the residual of $\mathbf{Z}_1$ is an exponential function with the REML estimated parameters: $\hat{\sigma}^2$=630.1(%^{2}), $\hat{\phi}$ =62.3(m) and $\hat{\tau}^2$ =0(%^{2}). Fig. 5.8 shows the covariance functions of the residual of $\mathbf{Z}_1$ (c11), $\mathbf{Z}_2$ (c22) and the cross-covariance function between the residual of $\mathbf{Z}_1$ and $\mathbf{Z}_2$ (c12). Notice that the covariance fucntion of the $\mathbf{Z}_1$ residual has no nugget effect. The covariance function of the $\mathbf{Z}_2$ residual is almost a pure nugget because the uncertainty in the expert data was modelled as white noise. The nugget value in the plot c22 is the mean of $\sigma_\varepsilon^2(\mathbf{s})$, just for the purpose of visualization. The general least squares estimated regression coefficient of $\mathbf{Z}_1$ is $\hat{\boldsymbol{\beta}}_1 = \left\{\hat{\beta}_{10} = 389.1, \hat{\beta}_{11} = -85.2, \hat{\beta}_{12} = 1.9, \hat{\beta}_{13} = -4.6\right\}$ and that of $\mathbf{Z}_2$ is $\hat{\boldsymbol{\beta}}_2 = \left\{\hat{\beta}_{20} = 89.8, \hat{\beta}_{21} = -11.6, \hat{\beta}_{22} = 0.5, \hat{\beta}_{23} = -3.75\right\}$. Hence, the systematic bias of the expert data is $m_\varepsilon(\mathbf{s}) = -299.3 + 73.6 f_1(\mathbf{s}) - 1.4 f_2(\mathbf{s}) + 0.85 f_3(\mathbf{s})$.

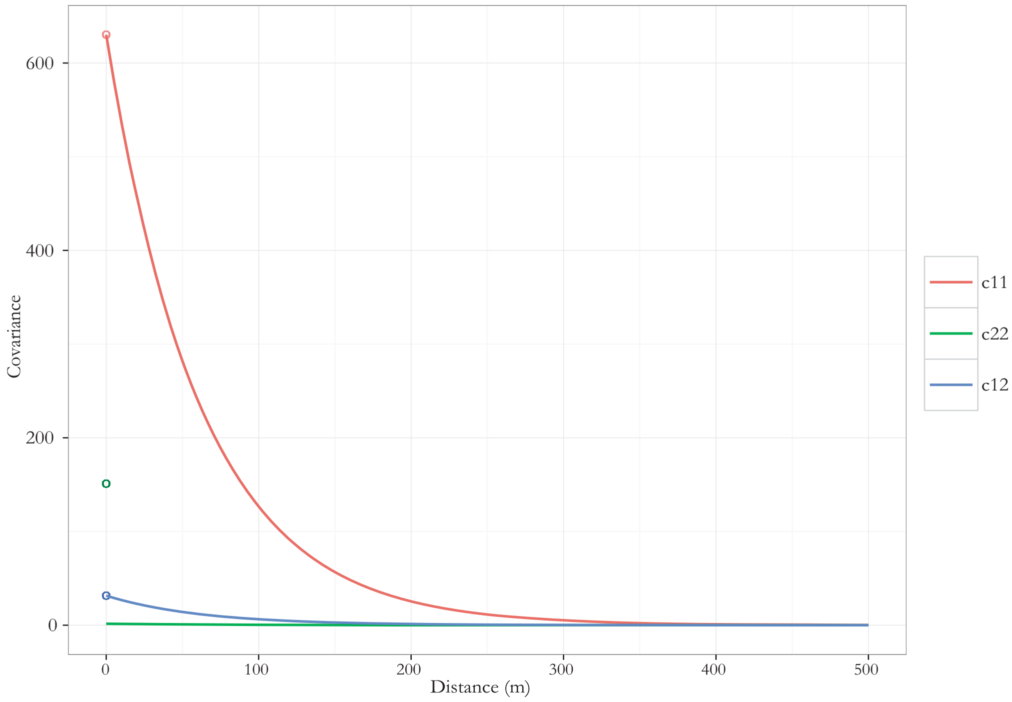

Figure 5.8: REML estimated covariance functions (c11, c22) and cross-covariance function (c12)

Fig. 5.9 shows the trends of SI from the two models on an arbitrary transect through the study area from west to east. The values were sorted from smallest to largest according to the mean values of Z_1. It clearly shows that the expert data are systematically biased. The systematic error has unequal magnitude. The mean value of the expert data for the SI has been smoothed over the study area: the high mean values (about > 35(%)) are underestimated and the low mean values (about < 35(%)) are overestimated. This smoothing effect is confirmed by the high value of the η coefficient. The systematic bias in the expert data was corrected by the unbiased condition of the RCK predictor (Journel and Huijbregts, 1978, Section V.A.4; Chilès and Delfiner, 1999, Section 3.7). Hence, the systematic bias did not affect the RCK prediction.

5

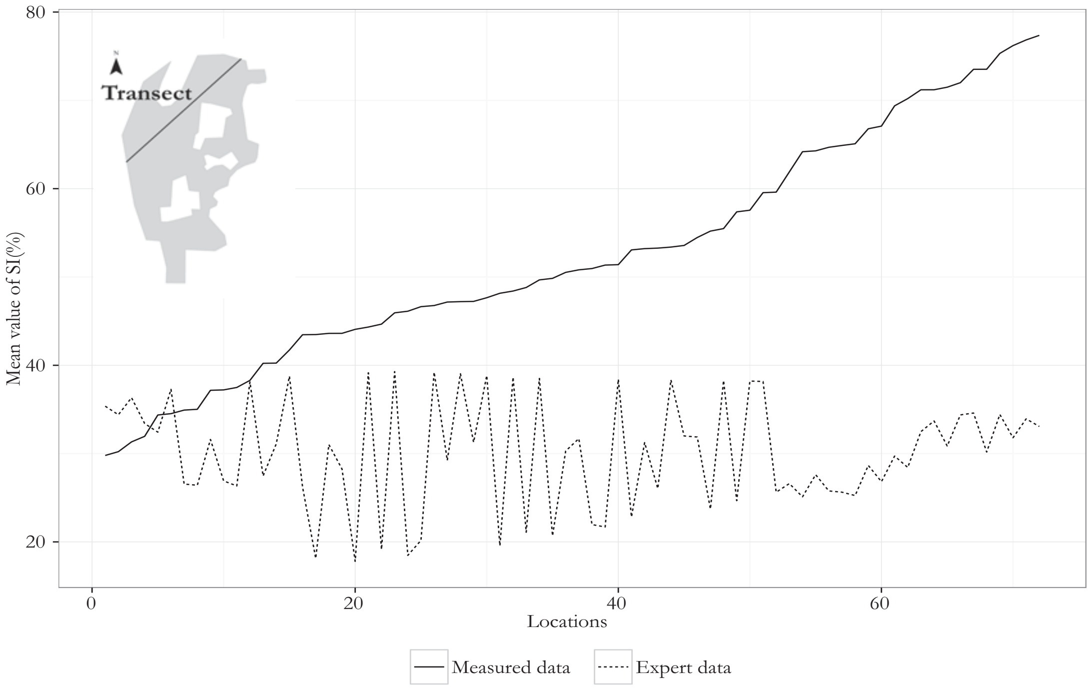

Figure 5.9: Difference in mean value (spatial trend) of SI from expert data versus measured data over an arbitrary transect from east to west in the study area

5.4.3. RCK prediction and accuracy

Recall that the data used for RCK included 16 error-free measurements and 50 soft data. Fig. 5.10 shows the RCK prediction and standard deviation maps. The RCK predictions of SI range from 7.9(%) to about 100(%), while the RCK standard deviation ranges from 0(%) to 25.2(%). At measurement locations, the RCK standard deviation equals zero because RCK is an exact interpolator (Stein and Corsten, 1991). However, the expert data were not reproduced exactly because these had error.

For the case of using only the measured data, RK predictions over the study area have a wider range of value of the standard deviation than that of RCK (i.e. RK standard deviation ranged from 0(%) at measurement locations to 63.5(%)). This confirmed that using expert data as soft measurements in RCK can decrease a large amount of the kriging variance at locations farther away from exact measurements. At the 64 validation locations, $\overline{\sigma}_{RCK}$= 21.1(%) and $RMSE_{RCK}$ = 24.8 (%), while for the case of using only measured data, $\overline{\sigma}_{RK}$ = 23.8 (%) and $RMSE_{RK}$ = 24.3 (%^{2}). The 95% confidence interval of the $RMSE_{RCK}$ is 24.8 ± 17.65 (%) and that of

$RMSE_{RK}$ is 24.3 ± 17.38 (%). These measures show that there is no significant difference in the performance of RCK compared to that of RK. Fig. 5.11 shows a scatter plot of the prediction error of RCK versus that of RK at the validation locations.

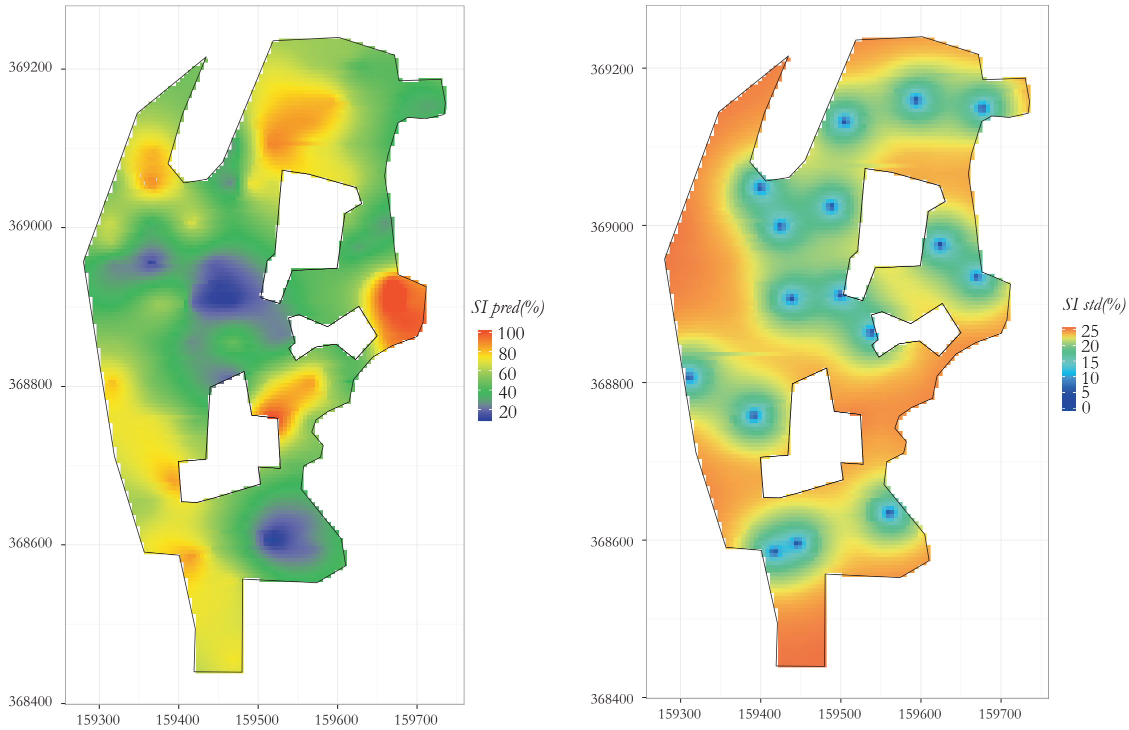

Figure 5.10: Regression cokriging of SI using expert data and measured data: kriging prediction (left) and kriging standard deviation (right)

The validation results showed no improvement in the prediction accuracy when RCK was used. This inefficiency can be explained by the weak co-located cross-correlation between expert and measured data and the contrast in the nugget effect between the residual covariance functions of the two datasets. As a result, the RCK weights of the expert data were very small compared to that of the measured data (Asli and Marcotte, 1995; Goovaerts, 1997, Section 6.2). The RCK system assigned different weights corresponding to the quality of the data. The reduction in the variance of the estimated trend at locations without direct measurements is the main reason that caused a large reduction of the RCK standard deviation. Similar results have been observed in the study of Stein and Corsten (1991).

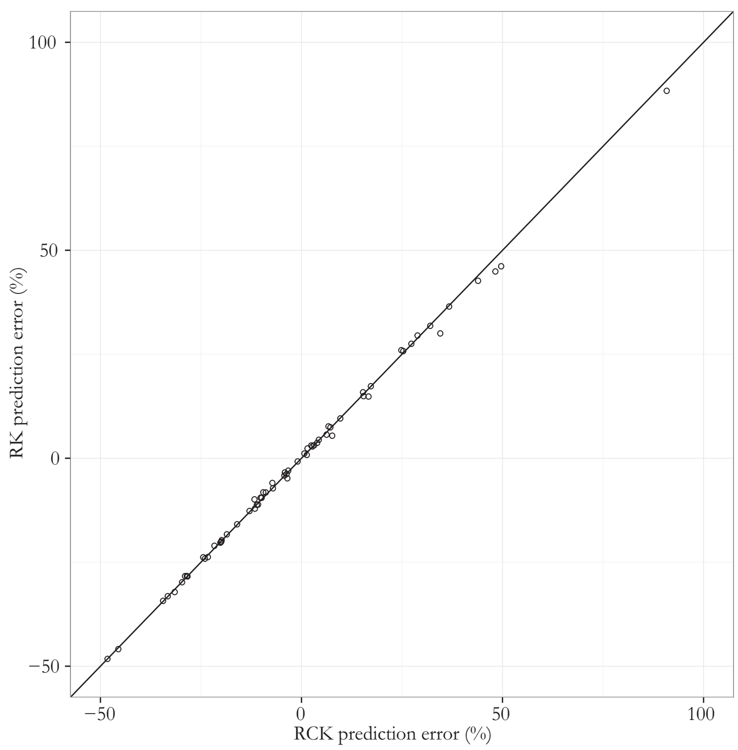

Figure 5.11: Regression cokriging prediction error versus regression kriging prediction error at validation locations

5.5. Conclusions

In this study, expert judgements based on expert knowledge about the values of a spatial variable at specific locations were used as soft data for supplementing direct measurements in geostatistical prediction. This has not been the only work that attempted to make use of information from experts for geostatistical prediction (other examples are Omre (1987), Okx et al.(1991), Stein (1994), Chapters 2 - 3), but this is the first work that treated expert quantitative judgements as soft data where expert knowledge was extracted by a formal statistical expert elicitation procedure. By using state of the art statistical expert elicitation techniques, the elicited quantities (i.e. the quartiles) were recommended to optimally capture the current stage of knowledge of experts in a probabilistic form.

The geostatistical models were built not only for incorporating expert knowledge in kriging but also for quantitative validation of expert data. The models hence provided the tool to validate expert judgements by measures of the systematic bias, the conditional bias and random error. The case study indicated that when expert data are treated as soft data that have large random errors, inclusion of expert data has no added value. Given the model specification, uncertain knowledge of only one expert cannot improve the prediction accuracy. However, this conclusion must be adhered to the model specification and the number of experts who contributed knowledge in this study. Future research may investigate different ways of specifying the model for the expert data. One may test the influence of aggregated expert knowledge (i.e. increasing the number of experts sharing knowledge) on the prediction accuracy. One may also test different spatial configurations of multiple expert data.

Appendix 5.A. Soil condition and expert judgements at 50 locations

Loc	pH	Clay	OC	Moist	Disri	Elev	Cdtot	Zntot	LQ	UQ
1	4.09	0.56	2.30	18.45	136	27.29	0.26	8	20	40
2	4.10	0.64	3.54	17.22	289	27.43	0.34	12	30	50
3	4.42	1.39	6.66	39.39	143	26.50	2.10	81	20	40
4	4.75	11.15	14.11	72.78	13	26.38	73.64	1,326	15	25
5	4.79	0.88	4.98	36.24	63	27.00	2.83	116	25	40
6	4.78	1.69	7.15	44.84	18	26.84	8.08	214	20	40
7	4.72	1.74	5.38	37.19	73	27.16	5.17	140	25	40
8	4.18	0.56	2.65	14.95	167	27.54	0.41	15	40	50
9	4.72	4.80	10.85	65.54	21	26.64	55.44	753	23	30
10	4.08	1.10	2.71	17.57	197	27.33	0.15	10	30	50
11	4.34	1.07	4.00	27.38	178	26.88	0.55	31	20	40
12	4.46	0.92	6.12	31.50	154	27.15	2.78	81	25	45
13	4.49	3.49	12.60	72.45	69	26.69	13.11	313	25	35
14	4.75	8.82	15.55	80.74	33	26.57	30.27	561	20	30
15	4.46	1.62	7.87	45.41	105	26.63	4.15	81	25	35
16	4.48	0.57	5.69	32.29	161	26.63	1.66	51	25	35
17	4.48	1.37	6.60	44.39	100	26.56	6.46	105	23	44
18	4.55	0.60	5.74	30.22	204	26.88	1.55	75	25	45
19	4.55	8.63	16.76	87.67	24	26.46	29.27	509	20	30
20	4.61	10.09	14.55	79.62	20	26.48	30.28	661	20	30
21	4.17	1.63	2.38	16.60	231	26.90	0.05	6	20	40
22	4.30	1.04	2.96	13.64	310	27.20	0.42	29	30	50
23	4.28	0.68	5.48	30.42	268	27.05	0.77	30	25	40
24	4.58	0.85	3.56	24.58	276	26.39	0.53	31	30	50

Loc	pH	Clay	OC	Moist	Disri	Elev	Cdtot	Zntot	LQ	UQ
25	4.63	1.14	10.65	55.22	180	26.49	4.11	185	25	40
26	3.60	1.88	3.92	15.62	271	27.95	0.31	13	20	45
27	4.09	0.57	4.11	17.30	331	27.25	0.56	22	30	50
28	4.57	2.54	8.58	57.66	111	26.43	11.74	190	20	41
29	4.72	11.15	14.95	76.45	35	26.35	66.18	1,023	10	25
30	4.45	1.47	6.85	44.21	164	26.45	1.99	68	25	40
31	4.82	8.68	13.56	68.20	71	26.31	98.35	1,743	10	20
32	4.61	3.34	11.20	60.20	87	26.30	23.05	476	15	23
33	4.46	0.57	7.13	43.57	243	26.62	0.84	56	25	40
34	4.47	2.14	7.24	44.01	49	26.49	7.39	126	23	44
35	4.35	0.66	2.81	20.14	136	27.45	0.63	24	32	50
36	4.26	0.59	2.70	17.20	152	27.45	0.50	19	40	50
37	4.22	0.57	2.67	15.95	159	27.47	0.45	17	40	50
38	4.30	0.62	2.74	18.62	144	27.46	0.56	21	42	50
39	4.70	5.15	11.98	61.64	81	26.35	49.75	877	13	21
40	4.78	6.54	12.63	64.35	82	26.33	70.66	1,259	10	20
41	4.64	4.14	11.54	60.10	83	26.33	35.13	639	13	21
42	4.40	0.65	6.25	37.61	255	26.69	0.78	49	23	39
43	4.37	0.70	5.93	33.97	261	26.74	0.84	44	23	39
44	4.39	0.58	6.52	41.26	249	26.67	0.65	47	23	39
45	3.87	1.48	3.02	13.73	249	27.19	0.22	13	20	40
46	3.72	1.69	3.52	14.60	260	27.80	0.30	14	20	40
47	4.01	1.42	2.61	14.08	240	27.04	0.11	10	20	40
48	4.49	1.49	7.08	45.39	72	26.50	5.11	96	25	45
49	4.48	1.52	7.01	43.95	59	26.53	5.03	94	24	45
50	4.49	1.49	6.96	45.64	87	26.56	5.71	102	24	45

Loc = locations, pH = pH-water, Clay (%), OC = Organic matter content (%), Moist = Moisture (%), Disri = Distance to river (meter), Elev = Elevation asl (meter), Cdtot = Cd total concentration (mg/kg), Zntot = Zn total concentration (mg/kg), LQ = Lower quartile of SI (%), UP = Upper quartile of SI (%).

Appendix 5.B. **Locations of sampling scheme for expert elicitation**

Chapter 6

General discussion

6.1. Introduction

Geostatistics is increasingly used in numerous disciplines of the Earth and environmental sciences and in practical problem-solving, along with the increased urge to advance our understanding of the Earth surface and subsurface spatial phenomena. To accomplish this, geostatisticians either refine and extend their geostatistical models or collect more (direct and ancillary) data and information. In this dissertation, I have gone for the second approach that seeks for more data and information. In my case, the source of more data and information is expert knowledge. Due to this, I also had to modify the geostatistical models that incorporate expert knowledge because using this type of data and information requires different methods than those used for incorporating just field measurements and covariate layers. The overall motivation for my research originates from a combination of three factors: the request in geostatistical research for data and information that cannot be sensibly obtained using physical measurement systems, the rapid developments in expert elicitation research and the availability of increasingly knowledgeable experts working in geostatistical research and other disciplines within the Earth and environmental sciences. These factors not only motivate but also make my research justified and achievable.

This chapter synthesises the main accomplishments of the work that I have done and that were presented in the previous chapters: to identify the role of expert knowledge in geostatistical research and to offer methods to elicit and incorporate expert knowledge in geostatistical inference and prediction. While the foregoing chapters answer all detailed research questions identified in Section 1.4.2, this chapter gives answers to the two main research questions presented in Section 1.4.1: 1. What is the role of expert knowledge in geostatistical inference and prediction?, 2. How to elicit and incorporate expert knowledge in geostatistical inference and prediction? The answer to the first main research question is given in Section 6.2, with discussion on the advantages and difficulties of using expert knowledge in geostatistical inference and prediction. Section 6.3 provides the answer to the second main research question, with my reflection on the methods used for extracting and incorporating expert knowledge in geostatistical models. Section 6.4 provides my personal view and discussion about how to advance this research topic. Finally, concluding remarks are given in Section 6.5.

6.2. What is the role of expert knowledge in geostatistical inference and prediction?

6.2.1. The role of expert knowledge in geostatistical inference and prediction

The geostatistical literature shows that expert knowledge has been marginally and informally used so far for geostatistical inference and prediction. When geostatistical data and information are unattainable (promptly and conveniently), the only remaining option is to ask for expert opinions. Several situations where expert knowledge as probabilistic judgements has added value were identified in this research: spatial structure inference (i.e. variogram estimation described in Chapters 2, 3, 4 and 5, and spatial trend estimation addressed in Chapter 5), spatial interpolation (Chapters 4 and 5), spatial uncertainty quantification (Chapter 2) and spatial data disaggregation (Chapter 4). These situations can be categorised into cases of: 1. inaccessible or unattainable data, 2. data available at inappropriate level of detail, and 3. insufficient data. The data and information gaps recognised in this research are: information about the spatial structure to use as prior information or forecast (Chapter 2), data for spatial uncertainty quantification of (legacy) unknown-quality-maps (Chapter 3), prior information about the spatial structure at a different spatial support (Chapter 4) and spatial gaps in the measurements for the variogram and spatial trend inference and kriging (Chapter 5).

Throughout this research, expert knowledge has been used not only as prior information but also as (soft) data. The formal roles of expert knowledge in geostatistical inference and prediction are classified according to the categories defined by Drescher et al. (2013): as *surrogate* for absent geostatistical data to infer the variogram and perform spatial prediction (Chapter 2), and quantify the spatial uncertainty that is poorly captured by conventional mapping methods (Chapter 3), as *complementary* data (i.e. a dataset given by experts to completely fill one gap in geostatistical data together with the measured data that fill another gap) to predict at the smaller spatial support (Chapter 4), and as *supplement* to measured data to enhance inference and prediction (Chapter 5).

Expert knowledge was used in this research only as probabilistic quantitative judgement. This is the type of data and information that geostatisticians need for parameter estimation and prediction because geostatistics is a quantitative analysis of geostatistical data, and this analysis is subject to uncertainty. The role of expert knowledge as a probabilistic quantitative judgement in geostatistics is similar to that in other research disciplines such as geology (Wood and Curtis, 2004; Curtis, 2012), landscape ecology (Perera et al., 2012b; Drescher et al., 2013), environmental health impact assessment (Knol et al., 2010), and environmental modelling and management (Krueger et al., 2012). The marginal and informal use of expert knowledge in geostatistics has been replaced in this research by a formal and systematic use. This formal use enhances the added value of expert knowledge because the expert elicitation procedure is formalised and explicitly reported. Hence, the validity of expert judgement can be monitored and assessed.

6.2.2. The difficulties of using expert knowledge in geostatistics

The use of expert knowledge in a probabilistic quantitative form for mapping has a dependence on location or area, i.e. the knowledge used is about the spatial variability in a certain geographical area. To characterise the spatial dependence, it is requested to have knowledge that can be used to infer the spatial correlation structure of a spatial variable. This is a distinctive feature, but it is also the main cause of difficulty in using expert knowledge in geostatistical inference and prediction, besides other common difficulties that have been listed in the expert elicitation literature (e.g. Chesley, 1975; Meyer and Booker, 2001). As mentioned in Chapter 2, there are limited algorithms that can be used to estimate the variogram, and they require a measure of the second-order moment (i.e. the variance) (Cressie, 1991, Section 2.3.1; Oliver et al., 2010a). Expert elicitation literature indicates that experts are subject to cognitive issues to directly quantify the variance (Kadane and Wolfson, 1998). Nevertheless, Chapter 2 succeeded in finding a solution for the elicitation of the variogram.

Judging spatially-dependent attributes always requires experts to process information in two dimensions: spatial space and attribute space. As a result, it is more difficult for experts than in cases of decision making, policy making or risk assessment where only the judgement for an attribute is required. Vagueness about the variables of which judgements are asked for can cause difficulties for the experts. The experiences reported in Chapter 3 where experts were asked to judge the spatial structure of the error provide a good example. The error is a quantity that experts cannot directly observe at a location or area; hence, judging it requires more abstract cognition (Kadane and Wolfson, 1998). Another aspect is that local expert knowledge might out-weigh general expert knowledge in a geostatistical context because local and detailed knowledge are often required given the aforementioned distinctive characteristic of the data used in geostatistics. Besides the case shown in Chapter 4, another clear example is given in Chapter 5 where the expert was asked to judge the value at specific locations within a given area. The high uncertainty in the expert judgements is good news in terms of risk of overconfidence bias (McKenzie et al., 2008), but the bad news is that the required knowledge in this case might be too specific for the expert to be able to give a more accurate quantification.

My research also showed that expert knowledge cannot be helpful in all cases, as also stated in Kuhnert et al. (2010) and Kuhnert (2011). Its usefulness is determined by both the quality of expert judgements and the wise and delicate use of expert knowledge. The former is determined by all factors that build the elicitation procedure (e.g. see Knol et al., 2010), while the latter might be dependent on the domain expert (not the elicitation expert). Chapter 5 gives an example of the unsuccessful use of expert knowledge to improve spatial inference and prediction. The reasons for this can be blamed on the elicitation skills and techniques that cannot extract 'better' knowledge from the expert. Natural scientists might encounter this obstacle when they work with experts because they may lack social and psychological skills. This important point will be further discussed in Section 6.3. It is also suggested to use a better approach than directly eliciting the values of the variable at multiple locations, because this long and intensive knowledge extraction procedure can be tiring (Miller, 1956). Another reason for the disappointing results in Chapter 5 might be that single expert knowledge is not satisfactory in all cases.

Indeed, using knowledge from a single expert can be risky and is only recommended in the case of testing new elicitation methods. However, in practice, there might not be many experts, particularly in geostatistics where the experts are expected to have field experience in the study area. The case study in Chapter 5 is an example where there were no experts with fieldwork experience in the study area, except for one of the co-authors. As emphasised in the expert elicitation literature (e.g. O'Hagan, 2012; Drescher et al., 2013), this research corroborates that uncertainty in expert knowledge and judgement must be taken into account. Otherwise, the use of expert knowledge is not rigorous. This can be done by either quantifying uncertainty in each expert judgement, which was done in Chapter 5. Another way is to consult multiple experts where all of the differences in expert knowledge are taken into account (Chapters 2, 3 and 4). When expert data are integrated with measured data at the same spatial support (i.e. the expert knowledge plays a supplement role), there can be a (complete) overlap between measured data and expert knowledge if the knowledge has been accumulated based on available measured data published elsewhere. This issue has not yet been discussed in the scientific literature.

In summary, the difficulties of using expert knowledge in geostatistics can be due to its geographical dependence, the two dimensional information processing (spatial and attribute space), the requirement of good elicitation skills, a lack of multiple experts and uncertain expert knowledge. Note that the discussion in this section focused on the role of expert knowledge as probabilistic quantitative judgements only. The formal use of expert knowledge as qualitative judgements and expertise (i.e. using expert knowledge to identify relevant models, data sources, types of knowledge, etc. (Booker et al., 2001)) has not been investigated in this research. In spite of the difficulties associated with using expert knowledge in geostatistics, it was also clear that expert knowledge can be very meaningful. Having recognised this, the challenges then lie in how to benefit from this knowledge in geostatistical analysis. Section 6.3 discusses some mechanisms to tackle these challenges.

6.3. How to elicit and incorporate expert knowledge in geostatistical inference and prediction?

In this research, I have been working on geostatistical analysis of only spatial datasets of natural phenomena, and they are collected from measurements and expert knowledge. The aims of using geostatistical analyses are to statistically analyse the spatial dependence and use the quantified spatial dependence in spatial interpolation. These scientific objectives of geostatistics drive the use of expert knowledge to overcome the issues of insufficient or unavailable geostatistical data and information.

6.3.1. Statistical expert elicitation in a geostatistical context

Throughout this research, expert elicitation techniques for univariate probability distributions were used to extract expert knowledge, but expert judgements were used for statistical inference of the distributions of spatial random fields (i.e. a collection of random variables indexed by spatial coordinates (Schabenberger and Gotway, 2005, Section 2.1)). This is the challenge that requires a new elicitation tool in geostatistical settings. The model-based geostatistical approach (Diggle and Ribeiro, 2007) was applied throughout this research by explicitly assuming the underlying process of the spatial phenomena as normal or log-normal multivariate distributions. Hence, the required elicitation techniques turned out to be the ones for a multivariate probability distribution. Although there are elicitation methods for the multivariate probability distribution, these are far too complicated and inappropriate to be applicable to the geostatistical setting. Nevertheless, the elicitation method in this research was developed based on the same principle that is used in the expert elicitation literature for multivariate distributions (e.g. Al-Awadhi and Garthwaite, 1998; O'Hagan et al., 2006, Section 5.3). The method includes two components (Chapters 2 and 3): the elicitation of the marginal probability distribution and that of the spatial dependence (i.e. equivalent to the elicitation of the association in the expert elicitation literature).

A systematic and formal procedure was used to develop the expert elicitation tasks in each of the four case studies (Section 1.5.1) of which the core component of the elicitation facilitator is a web-based tool. This formal procedure (e.g. as used in Chapter 3) completely followed recommended procedures recently published in the expert elicitation literature (e.g. Choy et al., 2009; Knol et al., 2010; Kuhnert et al., 2010), although the web-based tools were newly built in Chapters 2, 3 and 5. One of the most difficult tasks in this formal procedure is to motivate and retain experts (e.g. as discussed in Section 3.4). Accomplishing this task requires not only domain expertise of the facilitator but also social skills (interpersonal skills) that can be a main obstacle for a natural scientist. Some previous investigations on this task also share the same experience, e.g. as reported in Drescher et al. (2013). The most laborious task is designing the elicitation tool, and this task is distinct among different research disciplines. The other tasks more or less stay unchanged as in the recommended procedure when applied in geostatistical setting.

The only elicitation technique that was applied in all web-based tools is the bisection method or quartile method that asks experts to specify, in a fixed order, the minimum, the quartiles and the maximum (Oakley, 2010). This is the direct expert elicitation technique, while the target quantity of elicitation can be an indirect measure (i.e. the first-order increment in Chapters 2 and 3), a direct measure of the model parameters (i.e. the parameters of the Matérn variogram model in Chapter 4) and direct values of the spatial variables at specific locations (Chapter 5). In the Chapters

2 to 5, I have classified the elicitation methods based on the character of the quantity being elicited: the indirect method was used in Chapters 2 and 3 and the direct method in Chapters 4 and 5. As recommended in the expert elicitation literature (e.g. Choy et al., 2009; Kuhnert et al., 2010), eliciting indirect quantities is also preferable in this research to lessen bias, but this also results in a more difficult task for the domain expert. The domain experts in this case are demanded to use their expertise to identify the proper indirect quantity, while defining the direct quantity is straightforward.

The web-based elicitation tools were developed for both one and multiple experts' elicitation. In case of multiple expert elicitation, the Delphi method was used where the experts were anonymous to each other, and only the combined expert judgement was revealed to all experts (Cooke, 1991, Section 3.3.1; Kuhnert et al., 2010). The purpose was to seek for a consensus among multiple experts, except in Chapter 4. To derive a single judgement, the mathematical pooling method that equally weighs the knowledge of the experts was used to combine multiple experts' judgements. The use of the equal weight pooling in this research was mainly done for a pragmatic reason (i.e. simple use and debate, taking into account all expert opinions equally and also their uncertainty) rather than reaching a true consensus in expert opinion. One of the main reasons is that I am not convinced by other more complicated approaches (i.e. the Bayesian method, weighted pooling, etc.) in terms of reaching a better representation of expert consensus. This sceptical point of view also appears in other studies (e.g. Clemen and Winkler, 1999; Clemen, 2008; Kuhnert et al., 2010). The technical solution for implementation of a web-based tool is limited, especially when many investigations are required if e.g. the weighted pooling or Cooke's classical methods (Cooke, 1991) were used. Another reason for mathematical pooling with equal weights is due to my intention to take an impartial opinion from all experts (Chapters 2 and 3) and to use the maximum diversity of the expert opinions to consider all experts' uncertainty (Chapter 4).

When developing the elicitation tools, I paid much attention to controlling the (cognitive) biases in expert judgements because it is a severe problem in using expert knowledge. My work was aimed at not only raising awareness of the experts themselves about the biases (e.g. providing information in the briefing document) but also minimising the biases at some stages of the elicitation procedure. For example, the order of elicitation questions was fixed to control anchoring biases and the Delphi method with anonymous experts was used to prevent motivational biases. This is also one of the most difficult tasks to accomplish, which requires interdisciplinary social knowledge from psychology and decision theory. Regardless of all the effort, the self-de-biased capability of experts might play a decisive role, particularly in the context of individual self-elicitation using web-based tools.

The role of the facilitator is always important (e.g. as proven Section 3.4), regardless which elicitation approach used (i.e. interview, workshop, web-based, etc.). The people who directly control the elicitation can be classified into the domain expert and the elicitation facilitator or elicitator. The responsibility of the domain expert is on the domain knowledge (i.e. geostatistical knowledge in this research), and that of the elicitation expert is on the expert elicitation skills. Consequently, playing the role of both the domain expert and the elicitator (as done in my case) demands a double effort; and hence, drawbacks are inevitable (e.g. the revision of expert judgement in Chapter 3, a large uncertainty in the elicitation outcome of one expert in Chapter 5). To apply expert elicitation theory in a geostatistical context, geostatistical background is a 'must' condition, while expert elicitation knowledge is an 'enough' condition to obtain a successful task; and again all frustrations that may turn up are due to the fact that working with human knowledge is a more complex and totally different situation from the conventional field measurement. A close cooperation between geostatisticians and the elicitation experts is recommended.

In this research, I have developed practically rigorous expert elicitation approaches that satisfy two criteria: incorporating a measure of uncertainty for the elicited expert knowledge and including an assessment of internal or external validity of findings (Drescher et al., 2013). Chapter 5 might be the best illustration of this among other chapters, where expert uncertainty was taken into account, and the internal and external biases of expert judgements were measured; hence, validity of the kriging map was assessed. This is an effort that should be stimulated, even though the results of this chapter are provisionally frustrating.

6.3.2. Methods of expert knowledge incorporation in geostatistics

To incorporate expert knowledge in geostatistical inference and prediction, two paradigms were leading this research: expert knowledge as prior information under a Bayesian perspective (i.e. Bayesian area-to-point kriging in Chapter 4) and as (soft) data under a geostatistical (frequentist) perspective (e.g. regression cokriging in Chapter 5). The important consideration for incorporating expert knowledge in geostatistical models is about the inexact or uncertain nature of judgements. The Bayesian approach is increasingly well-known and influential in geostatistics due to the ability to include uncertainty about the model parameters in kriging (Pilz and Spock, 2008) and combine data from different spatial supports (Gotway and Young, 2002). When expert data are integrated with measured data (either at the same or a different spatial support), properly weighing the contribution of expert knowledge in inference and prediction compared to that of the measured data may be of most concern. This weighing system should be controlled by the size of the dataset, the overlap between the data, the quality of the data, and in the case of a spatial variable, the relative distances. In this regard, both paradigms can be satisfactory.

Geostatistical models are usually well-prepared to integrate incompatible datasets (see e.g. Goovaerts, 1997, Chapter 6; Gotway and Young, 2002 for an overview). This statement is made based on the fact that all geostatistical models that were used to incorporate expert knowledge in this research are available in the literature. The important modification of the geostatistical models if any lies in the possible use of the models to calibrate expert judgments, in order to benefit from the measured data that are integrated (e.g. Chapter 5). However, this can be a biased statement because in the case studies used in this research, the required knowledge from experts was actually specified with the intention of direct use in available geostatistical algorithms. Hence, when the elicitation tasks succeed in extracting the required knowledge from experts, less work is demanded on 'reinventing' geostatistical models. A better comprehensive approach can be to ask the expert what knowledge they can best provide and then modify the geostatistical models to best use these knowledge. The evolution of a hierarchical perspective in spatial-temporal geostatistics (Cressie and Wikle, 2011) can provide a more flexible approach for incorporating, and hence, enhancing a flourishing use of expert knowledge in geostatistics.

In summary, while elicitation of expert knowledge in geostatistics may be accomplished by following an existing formal elicitation procedure with a new elicitation web-based tool for the spatial random fields, incorporation of expert knowledge in geostatistical inference and prediction may make use of the Bayesian and the geostatistical approaches. Note that all the comparisons made so far are based on the literature from other research disciplines because no previous research on expert elicitation in geostatistics has been published.

6.4. Insight and Implications

6.4.1. Expert knowledge through the eye of geostatisticians

Geostatisticians might look at expert knowledge of a spatial variable as a (linear) regression prediction (with error) because they might expect that experts can relate their judgements to the factors in the environment that affect the distribution of the spatial variable. They might even think that experts make a simple extrapolation in their judgements when they judge the value at unobserved locations. When knowledge of geostatistical model parameters is required, they might think that they can extract these only from the geostatistical specialists. These preliminary views are my personal, and the geostatisticians that I have in mind are geostatistical practitioners. This view may help experts understand the expectation when they are involved in a geostatistical elicitation task.

6.4.2. The interaction between geostatistical models and expert knowledge

There are certainly mutual influences between expert knowledge that can be used in geostatistics and the geostatistical models that can incorporate expert knowledge. On the one hand, the former can be noticeable in this research. The required knowledge from experts was shaped in one way or another by geostatistical assumptions. For example, the marginal distribution of a random spatial variable was limited to be either normal or log-normal (Chapters 2 and 3), and the intrinsic and isotropic variogram model was elicited (Chapter 4). All of the basic assumptions of geostatistical models (i.e. (log-) normal probability distributions, (second-order or intrinsic) stationary, isotropy) were either explicitly or implicitly imposed on the expert judgements. To a certain extent, these imposed assumptions on the experts and their judgements are considered as a kind of bias (it can be classified into motivational biases); but it is fair to say that geostatisticians have formed their knowledge about spatial phenomena in this way, and hence, the same expectation holds for the knowledge of other experts of applied geostatistics.

On the other hand, the influence of the use of expert judgements on the geostatistical models was not considerable in this research. Some limitations that were required are that the method-of-moments for estimating the variogram may not be appropriate any more (Chapter 2), and that kriging should consider the error in the data (Chapter 5). I argue that the better expert knowledge is the one that is not limited by geostatistical assumptions; otherwise, besides the possible (motivational) bias, the overlap between expert knowledge and published data and information can be large, which lessens the added value of expert knowledge. It is interesting but not trivial to investigate what knowledge humans actually hold about a spatial phenomenon. Human knowledge is much more than a geostatistical sample. A different geostatistical model is thus need to sensibly make better use of the data and information given by the expert judgements.

6.4.3. The possibility of expert judgement calibration and validation by geostatistical models

The expert elicitation literature clearly differentiates between the quality of the elicitation procedure and the quality of expert judgment (e.g. see O'Hagan et al., 2006, Chapter 8). They are not necessarily related. The former has been discussed in the previous four chapters, in the implementation of the elicitation procedures. In this section, only the calibration and validation of expert judgements are of concern. This is the assessment of an agreement between expert judgments and reality in a geostatistical context (O'Hagan et al., 2006).

I argue that expert judgments of a spatial variable should be better calibrated and validated by (the outcome from) a geostatistical model. As clarified in Section 6.3, while the experts judge the univariate probability distributions, their probabilistic judgements are used to infer the spatial random field. Hence, calibration and validation of the individual univariate probability distribution are inappropriate because the (auto) correlation of the judgements matters. In case the expert judgements are the model parameters directly, the outcomes of the geostatistical models should be used. The case studies presented in this research demonstrate the possibility of using geostatistical models for expert judgement calibration and validation.

6.4.4. Toward an efficient but cheap elicitation framework for spatial phenomena

In Chapters 2 and 3, I have argued for the necessary assistance of computer software in statistical expert elicitation. Henceforth, all elicitation tasks in this research were assisted by computer software, and they all are web-based tools. Numerical optimization when fitting the probability density function or graphical feedback of the elicitation outcomes, for example is a favourite aid of computer software. The visual aids are even more appealing in geostatistics when feedback might require to be presented in the form of a spatial map; and web-mapping technology is already that much advanced that it can be greatly advantageous.

Besides the requirements of IT solutions, the elicitation exercise can be expensive due to allowance and relocating cost in case recruited experts are geographically scattered. As a result, it can be costly (or even impossible) to gather multiple (busy) experts at once. In such cases, the web-based tool can be an efficient and economical solution. It is efficient because all core components of a formal elicitation procedure can be web-based operationalised (e.g. Morris et al., 2014). The remote elicitation exercise also provides more convenient and time-flexibility for busy experts to complete their tasks. Hence, web-based tools can be greatly beneficial when the project budget is modest. Although much work can be initially required to build the web-based tool, when it is established, it provides a long term use.

6.4.5. Future research challenge

The important challenge for future researchers on this topic lies in how to maximise the added value of expert knowledge in geostatistics both in qualitative and quantitative terms. To tackle this challenge, researchers must be well-trained and equipped with interdisciplinary knowledge and skills. The challenge is worth the effort because humans interact with natural and artificial spatial phenomena in everyday life, and their knowledge about them is often of high value.

6.5. Conclusions

1. The added value of expert knowledge in geostatistical inference and prediction is real.

2. The use of statistical expert elicitation to extract knowledge from experts is essential to enhance the use of expert knowledge in geostatistics.

3. Expert knowledge should ideally be elicited freely from geostatistical assumptions, and the geostatistical model should be well-prepared for integrating elicited knowledge.

References

Al-Awadhi, S.A., Garthwaite, P.H., 1998. An elicitation method for multivariate normal distributions. Commun. Stat. Theory Methods 27, 1123 - 1142.

Albert, J., 2009. Bayesian computation with R. Springer, New York.

Armstrong, M., 1998. Basic linear geostatistics. Springer, Berlin.

Asli, M., Marcotte, D., 1995. Comparison of approaches to spatial estimation in a bivariate context. Math. Geol. 27, 641-658.

Atkinson, P.M., 2013. Downscaling in remote sensing. Int. J. Appl. Earth Obs. Geoinf. 22, 106-114.

Ayyub, B.M., 2001. Elicitation of expert opinions for uncertainty and risk. CRC Press, Boca Raton, Florida.

Barnes, R., 1991. The variogram sill and the sample variance. Math. Geol. 23, 673-678.

Bedford, T., Walls, L., 2010. Evaluation of elicitation methods to quantify Bayes linear models. Proc. Inst. Mech. Eng. O J. Risk Reliab. 224, 322-332.

Booker, J.M., Anderson, M.C., Meyer, M.A., 2001. The role of expert knowledge in uncertainty quantification (Are we adding more uncertainty or more understanding?) Seventh army conference on applied statistics, 155-161.

Bowley, A.L., 1920. Elements of statistics. P.S. King & Son Limited, London.

Brus, D.J., Heuvelink, G.B.M., 2007. Optimization of sample patterns for universal kriging of environmental variables. Geoderma 138, 86-95.

Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. Eur. J. Soil. Sci. 62, 394-407.

Chesley, G.R., 1975. Elicitation of subjective probabilities: a review. Account. Rev. 50, 325-337.

Chib, S., Greenberg, E., 1995. Understanding the Metropolis-Hastings algorithm. Am. Stat. 49, 327-335.

Chilès, J.P., Delfiner, P., 1999. Geostatistics: modelling spatial uncertainty. Wiley series in probability and statistics. Wiley, New York.

Choy, S.L., O'Leary, R., Mengersen, K., 2009. Elicitation by design in ecology: using expert opinion to inform priors for Bayesian statistical models. Ecology 90, 265-277.

Clemen, R., Winkler, R., 1999. Combining probability distributions from experts in risk analysis. Risk Anal. 19, 187 - 203.

Clemen, R.T., 2008. Comment on Cooke's classical method. Reliability Engineering & System Safety 93, 760-765.

Cooke, R.M., 1991. Experts in uncertainty: opinion and subjective probability in science. Environmental ethics and science policy series. Oxford University Press, New York.

Cooke, R.M., Goossens, L.H.J., 2000. Procedures guide for structured expert judgement in accident consequence modelling. Radiat. Prot. Dosimetry 90, 303-309.

Cressie, N., 1985. Fitting variogram models by weighted least-squares. Math. Geol. 17, 563-586.

Cressie, N., 1990. The origins of kriging. Math. Geol. 22, 239-252.

Cressie, N., 1991. Statistics for spatial data. Wiley series in probability and mathematical statistics. Wiley, New York.

Cressie, N., 1996. Change of support and the modifiable areal unit problem. J. Geogr. Syst. 3, 159-180.

Cressie, N., Hawkins, D., 1980. Robust estimation of the variogram: I. Math. Geol.

12, 115-125.

Cressie, N., Wikle, C.K., 2011. Statistics for spatio-temporal data. Wiley series in probability and statistics. Wiley, Oxford.

Cui, H., Stein, A., Myers, D. E., 1995. Extension of spatial information, Bayesian kriging and updating of prior variogram parameters. Environmetrics 6, 373-384.

Curtis, A., 2012. The science of subjectivity. Geology 40, 95-96.

De Goede, R.G.M., Bongers, T., Ettema, C.H., 1993. Graphical presentation and interpretation of nematode community structure: c-p triangles. Med. Fac. Landbouww. Univ. Gent 58/2b: 743-750.

De Gruijter, J.J., Bierkens, M.F.P., Brus, D.J., Knotters, M., 2006. Sampling for natural resource monitoring. Springer-Verlag Berlin Heidelberg: Berlin, Heidelberg.

Defouquet, C., 1994. Reminders on the conditioning kriging, in: Armstrong, M., Dowd, P.A. (Eds.), Geostatistical simulations: Proceedings of the Geostatistical simulation workshop 7, Fontainebleau, pp. 131-145.

DeGroot, M.H., 1974. Reaching a consensus. J. Am. Stat. Assoc. 69, 118-121.

Denham, R., Mengersen, K., 2007. Geographically assisted elicitation of expert opinion for regression models. Bayesian Anal. 2, 99-136.

Dietrich, C.R., Newsam, G.N., 1993. A fast and exact method for multidimensional gaussian stochastic simulations. Water. Resour. Res. 29, 2861-2869.

Diggle, P.J., Ribeiro, Jr.P.J., 2002. Bayesian inference in Gaussian model-based geostatistics. Geographical and Environmental Modelling 6, 129-146.

Diggle, P.J., Ribeiro, P.J., 2007. Model based geostatistics. Springer Series in Statistics, Springer, New York.

Dowd, P.A., 1984. The variogram and kriging: robust and resistant estimators, in: Verly, G., David, M., Journel, A.G. (Eds.), Geostatistics for natural resources characterisation (Part 1), pp. 91-106.

Drescher, M., Perera, A.H., Johnson, C.J., Buse, L.J., Drew, C.A., Burgman, M.A., 2013. Toward rigorous use of expert knowledge in ecological research. Ecosphere 4, art83.

Dungan J.L., 2006. Toward a comprehensive view of uncertainty in remote sensing analysis, in: Foody, G.M., Atkinson, P.M. (Eds), Uncertainty in remote sensing and GIS, John Wiley & Sons, Ltd, pp. 25-35.

Ettema, C.H., Wardle, D.A., 2002. Spatial soil ecology. Trends Ecol. Evol. 17, 177-183

Ferris, H., Bongers, T., de Goede, R.G.M., 2001. A framework for soil food web diagnostics: extension of the nematode faunal analysis concept. Appl. Soil Ecol. 18, 13-29.

Fisher, R., O'Leary, R.A., Low-Choy, S., Mengersen, K., Caley, M. J., 2012. A software tool for elicitation of expert knowledge about species richness or similar counts. Environ. Model. Softw. 30, 1-14.

French, S., 2011. Aggregating expert judgement. Serie A. Matematicas 105, 181-206.

Frogbrook, Z. L., 1999. The effect of sampling intensity on the reliability of predictions and maps of soil properties. Precision agriculture '99, Part 1 and Part 2. 2nd European conference on Precision agriculture, Odense, Denmark, 11-15 July 1999.

Garthwaite, P.H., Dickey, J.M., 1985. Double- and single-bisection methods for subjective probability assessment in a location-scale family. J. Econom. 29, 149-163.

Garthwaite, P.H., Kadane, J.B., O'Hagan, A., 2005. Statistical methods for eliciting probability distributions. J. Am. Stat. Assoc. 100, 680-700.

Gelfand, A.E., Zhu, L., Carlin, B.P., 2001. On the change of support problem for spatio-temporal data. Biostatistics 2, 31-45.

Genest, C., Mcconway, K.J., 1990. Allocating the weights in the linear opinion pool. J. Forecast. 9, 53-73.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments, in: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds), Bayesian statistics 4. Oxford University Press, pp. 169-193.

Goovaerts, P., 1997. Geostatistics for natural resources evaluation. Oxford University Press, New York.

Goovaerts, P., 1999. Geostatistics in soil science: state-of-the-art and perspectives. Geoderma 89, 1-45.

Goovaerts, P., 2001. Geostatistical modelling of uncertainty in soil science. Geoderma 103, 3-26.

Goovaerts, P., 2008. Kriging and semivariogram deconvolution in the presence of irregular geographical units. Math. Geosci. 40, 101-128.

Goovaerts, P., Kerry, R., 2010. Using ancillary data to improve prediction of soil and crop attributes in precision agriculture, in: Oliver, M.A. (Ed), Geostatistical applications for precision agriculture. Springer Netherlands, Dordrecht, pp. 167-194.

Gotway, C.A., Young, L.J., 2002. Combining incompatible spatial data. J. Am. Stat. Assoc. 97, 632-648.

Gotway, C.A., Young, L.J., 2007. A geostatistical approach to linking geographically aggregated data from different sources. J. Comput. Graph. Stat. 16, 115-135.

Ha, W., Gowda, P., Howell, T., 2013. A review of downscaling methods for remote sensing-based irrigation management: part I. Irrigation. Sci. 31, 831-850.

Hahn, E.D., 2006. Re-examining informative prior elicitation through the lens of Markov chain Monte Carlo methods. J. R. Stat. Soc. Ser. A Stat. Soc. 169, 37-48.

Hall, D.G.M., Reeve, M.J., Thomasson, A.J., 1977. Water retention, porosity and density of field soils. Soil survey technical monograph - no. 9. Rothamsted Experimental Station, Harpenden.

Hamzehpour, N., Eghbal, M.K., Bogaert, P., Toomanian, N., Sokouti, R.S., 2013. Spatial prediction of soil salinity using kriging with measurement errors and probabilistic soft data. Arid Land Research and Management 27, 128-139.

Hengl, T., Heuvelink, G.B.M., Rossiter, D.G., 2007. About regression-kriging: From equations to case studies. Comput. Geosci. 33, 1301-1315.

Heuvelink, G.B.M., 2006. Analysing uncertainty propagation in GIS: Why is it not that simple?, in: Foody, G.M., Atkinson, P.M. (Eds), Uncertainty in remote sensing and GIS. John Wiley & Sons, Ltd, pp. 155-165.

Heuvelink, G.B.M., Brown, J.D., van Loon, E.E., 2007. A probabilistic framework for representing and simulating uncertain environmental variables. Int. J. Geogr. Inf. Sci. 21, 497-513.

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. Comput. Geosci. 35, 1711-1721.

Hoef, J., Cressie, N., 1993. Multivariable spatial prediction. Math. Geol. 25, 219-240.

Hogarth, R.M., 1975. Cognitive processes and the assessment of subjective probability distributions. J. Am. Stat. Assoc. 70, 271-289.

Hora, S.C., 2004. Probability judgments for continuous quantities: linear combinations and calibration. Manage. Sci. 50, 597-604.

Hudson, G., Wackernagel, H., 1994. Mapping temperature using kriging with external drift: theory and an example from Scotland. Int. J. Climatol. 14, 77-91.

Ikoyi, I.O., 2013. Spatial distribution of nematodes in a heavy metal-contaminated nature reserve, Thesis report, Ghent University, Belgium. http://lib.ugent.be/fulltxt/RUG01/002/061/224/RUG01-002061224_2013_0001_AC.pdf.

James, A., Choy, S.L., Mengersen, K., 2010. Elicitator: an expert elicitation tool for regression in ecology. Environ. Model. Softw. 25, 129-145.

Journel, A.G., Huijbregts, C.J., 1978. Mining geostatistics. Academic Press, London.

Journel, A.G., 1986. Constrained interpolation and qualitative information: the soft kriging approach. Math. Geol. 18, 269-286.

Kadane, J.B., Wolfson, L.J., 1998. Experiences in elicitation. Journal of the Royal Statistical Society. Series D (The Statistician) 47, 3-19.

Kerry, R., Oliver, M.A., 2003. Variograms of ancillary data to aid sampling for soil surveys. Precision Agriculture 4, 261-278.

Kerry, R., Oliver, M.A., 2004. Average variograms to guide soil sampling. Int. J. Appl. Earth Obs. Geoinf. 5, 307-325.

Kerry, R., Oliver, M.A., 2007. Comparing sampling needs for variograms of soil properties computed by the method of moments and residual maximum likelihood. Geoderma 140, 383-396.

Knol, A., Slottje, P., van der Sluijs, J., Lebret, E., 2010. The use of expert elicitation in environmental health impact assessment: a seven step procedure. Environmental Health 9, 19.

Kros, J., Pebesma, E.J., Reinds, G.J., Finke, P.A., 1999. Uncertainty assessment in modelling soil acidification at the European scale: a case study. J. Environ. Qual. 28, 366-377.

Krueger, T., Page, T., Hubacek, K., Smith, L., Hiscock, K., 2012. The role of expert opinion in environmental modelling. Environ. Model. Softw. 36, 4–18.

Kuenzer, C., Dech, S., Guo, H., Ottinger, M., Zhang, J., Dech, S., 2013a. Space borne thermal infrared observation: an overview of most frequently used sensors for applied research, in: Kuenzer, C., Dech, S., (Eds), Thermal infrared remote sensing. Remote sensing and digital image processing. Springer Netherlands, Dordrecht, pp. 131-148.

Kuenzer, C., Dech, S., Sobrino, J., Frate, F., Drusch, M., Jiménez-Muñoz, J., Manunta, P., 2013b. Review of high resolution thermal infrared applications and requirements: the Fuegosat synthesis study, in: Kuenzer, C., Dech, S., (Eds), Thermal infrared remote sensing. Remote sensing and digital image processing. Springer Netherlands, Dordrecht, pp. 197-214

Kuhnert, P.M., 2011. Four case studies in using expert opinion to inform priors. Environmetrics 22, 662-674.

Kuhnert, P.M., Martin, T.G., Griffiths, S.P., 2010. A guide to eliciting and using expert knowledge in Bayesian ecological models. Ecol. Lett. 13, 900-914.

Kynn, M., 2008. The 'heuristics and biases' bias in expert elicitation. J. R. Stat. Soc. Ser. A Stat. Soc. 171, 239-264.

Kyriakidis, P.C., 2004. A geostatistical framework for area-to-point spatial interpolation. Geogr. Anal. 36, 259-289.

Lark, R. M., 2002. Optimized spatial sampling of soil for estimation of the variogram by maximum likelihood. Geoderma 105, 49-80.

Lark, R.M., Bishop, T.F.A., Webster, R., 2007. Using expert knowledge with control of false discovery rate to select regressors for prediction of soil properties. Geoderma 138, 65-78.

Lark, R.M., Cullis, B.R., Welham, S.J., 2006. On spatial prediction of soil properties in the presence of a spatial trend: the empirical best linear unbiased predictor (E-BLUP) with REML. Eur. J. Soil. Sci. 57, 787-799.

Lele, S.R., Allen, K.L., 2006. On using expert opinion in ecological analyses: a frequentist approach. Environmetrics 17, 683-704.

Lele, S.R., Das, A., 2000. Elicited data and incorporation of expert opinion for statistical inference in spatial studies. Math. Geol. 32, 465-487.

Low-Choy, S., James, A., Murray, J., Mengersen, K., 2012. Elicitator: a user-friendly, interactive tool to support the elicitation of expert knowledge, in: Perera, A.H., Drew, C.A., Johnson, C.J. (Eds.), Expert knowledge and its applications in landscape ecology. Springer, New York, pp. 39-67.

Malone, B., de Gruijter, J.J., McBratney, A.B., Minasny, B., Brus, D.J., 2011. Using additional criteria for measuring the quality of predictions and their uncertainties in a digital soil mapping framework. Soil Sci. Soc. Am. J. 75, 1032-1043.

Marceau, D.J., Hay, G.J., 1999. Remote sensing contributions to the scale issue. Can.

Aeronaut. Space J. 25, 357-366.

Marchant, B.P., Lark, R.M., 2007. Robust estimation of the variogram by residual maximum likelihood. Geoderma 140, 62-72.

Matheron, G., 1963. Principles of geostatistics. Economic Geology 58, 1246-1266.

McBratney, A.B., Webster, R., Burgess, T.M., 1981. The design of optimal sampling schemes for local estimation and mapping of regionalised variables .1. Theory and method. Comput. Geosci. 7, 331-334.

McBratney, A.B., Webster, R., 1983. How many observations are needed for regional estimation of soil properties? Soil Sci. 135, 177-183.

McBratney, A.B., Pringle, M.J., 1999. Estimating average and proportional variograms of soil properties and their potential use in precision agriculture. Precision Agriculture 1, 125-152.

McBratney, A.B., Minasny, B., Cattle, S.R., Vervoort, R.W., 2002. From pedotransfer functions to soil inference systems. Geoderma 109, 41-73.

McKenzie, C.R.M., Liersch, M.J., Yaniv, I., 2008. Overconfidence in interval estimates: what does expertise buy you? Organ. Behav. Hum. Decis. Process. 107, 179-191.

Meyer, M.A., Booker, J.M., Bradshaw, J.M., 1990. A flexible six-step program for defining and handling bias in knowledge elicitation, in: Wielinga, B., Boose, J.H., Gaines, B.R., Schreiber, G., Van Someren, M. (Eds.), Current trends in knowledge acquisition, Amsterdam: IOS Press, pp. 237-256.

Meyer, M.A., Booker, J.M., 2001. Eliciting and analysing expert judgment: a practical guide. Society for industrial and applied mathematics and American statistical association, Philadelphia, Pennsylvania.

Miller, G.A., 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol. Rev. 63, 81-97.

Miller, H.J., 2004. Tobler's first law and spatial analysis. Ann. Assoc. Am. Geogr. 94, 284-289.

Minasny, B., McBratney, A.B., 2002. Uncertainty analysis for pedotransfer functions. Eur. J. Soil Sci. 53, 417-429.

Minasny, B., McBratney, A.B., Bristow, K.L., 1999. Comparison of different approaches to the development of pedotransfer functions for water-retention curves. Geoderma 93, 225-253.

Morgan, M., Henrion, M., 1990. Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis, Cambridge University Press, Cambridge.

Morris, D.E., Oakley, J.E., Crowe, J.A., 2014. A web-based tool for eliciting probability distributions from experts. Environ. Model. Softw. 52, 1-4.

Mullen, K.M., Ardia, D., Gil, D.L., Windover, D., Cline, J., 2011. DEoptim: an R package for global optimization by differential evolution. J. Stat. Softw. 40, 1-26.

Myers, D.E., 1991. On variogram estimation, in: Öztürk, A., van der Meulen, E.C., Dudewicz, E.J., Nelson, P.R. (Eds), Conference proceedings on the frontiers of statistical scientific theory and industrial applications (Vol. II). American Sciences Press, Çesme, Izmir, Turkey.

Nagle, N.N., Sweeney, S.H., Kyriakidis, P.C., 2011. A geostatistical linear regression model for small area data. Geogr. Anal. 43, 38-60.

O'Hagan, A., 1998. Eliciting expert beliefs in substantial practical applications. Journal of the royal statistical society. Series D (The Statistician) 47, 21-35.

O'Hagan, A., Buck, C., Daneshkhah, A., Eiser, J., Garthwaite, P., Jenkinson, D., Oakley, J., Rakow, T., 2006. Uncertain judgements : eliciting experts' probabilities. Statistics in practice. John Wiley & Sons, Ltd., Chichester.

O'Hagan, A., Oakley, J.E., 2004. Probability is perfect, but we can't elicit it perfectly. Reliab. Eng. Syst. Safe. 85, 239-248.

O'Hagan, A., 2012. Probabilistic uncertainty specification: overview, elaboration techniques and their application to a mechanistic model of carbon flux. Environ. Model. Softw. 36, 35-48.

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/.

Oberthür, T., Goovaerts, P., Dobermann, A., 1999. Mapping soil texture classes using field texturing, particle size distribution and local knowledge by both conventional and geostatisical methods. Eur. J. Soil Sci. 50, 457-479.

Okx, J.P., Heuvelink, G.B.M., Grinwis, A.W., 1991. Expert knowledge and (geo)statistical methods: complementary tools in soil pollution research. Contaminated soil '90. Third KfK/TNO conference, Karlsruhe, 1990. Vol. I, 729-736.

Oliver, M.A., Fischer, M.M., Getis, A., 2010a. The variogram and kriging. Handbook of applied spatial analysis. Springer Berlin, Heidelberg, pp. 319-352.

Oliver, M.A., Goovaerts, P., Kerry, R., 2010b. Using ancillary data to improve prediction of soil and crop attributes in precision agriculture, in: Oliver, M.A. (Eds), Geostatistical applications for precision agriculture. Springer Netherlands, Dordrecht, pp. 167-194.

Oliver, M.A., Webster, R., 2014. A tutorial guide to geostatistics: computing and modelling variograms and kriging. Catena 113, 56-69.

Omre, H., 1987. Bayesian kriging-merging observations and qualified guesses in kriging. Math. Geol. 19, 25-39.

Oostenbrink, M., 1960. Estimating nematode populations by some selected methods., in: Sasser, J.N., Jenkins, W.R. (Eds), Nematology: fundamentals and recent advances with emphasis on plant parasitic and soil forms. The University of North Carolina Press, Chapel Hill, pp. 85-102.

Pardo-Igúzquiza, E., 1997. MLREML: A computer program for the inference of spatial covariance parameters by maximum likelihood and restricted maximum likelihood. Comput. Geosci. 23, 153-162.

Pardo-Igúzquiza, E., 1998. Maximum likelihood estimation of spatial covariance parameters. Math. Geol. 30, 95-108.

Pardo-lgúzquiza, E., 1999. Bayesian inference of spatial covariance parameters. Math. Geol. 31, 47-65.

Pardo-Igúzquiza, E., Atkinson, P.M., 2007. Modelling the semivariograms and cross-semivariograms required in downscaling cokriging by numerical convolution-deconvolution. Comput. Geosci. 33, 1273-1284.

Pardo-Igúzquiza, E., Mardia, K.V., Chica-Olmo, M., 2009. MLMATERN: A computer program for maximum likelihood inference with the spatial Matérn covariance model. Comput. Geosci. 35, 1139-1150.

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Comput. Geosci. 30, 683-691.

Perera, A.H., Drew, C.A., Johnson, C.J., 2012a. Expert knowledge and its application in landscape ecology. Springer New York, New York, NY.

Perera, A.H., Drew, C.A., Johnson, C.J., Perera, A., Drew, C.A., Johnson, C., 2012b. Experts, expert knowledge, and their roles in landscape ecological applications, in: Perera, A.H., Drew, C.A., Johnson, C.J., (Eds), Expert knowledge and its application in landscape ecology. Springer New York, pp. 1-10.

Peterson, C., Miller, A., 1964. Mode, median, and mean as optimal strategies. J. Exp. Psychol. 68, 363-367.

Pilz, J., Spock, G., 2008. Why do we need and how should we implement Bayesian kriging methods. Stoch. Environ. Res. Risk Assess. 22, 621-632.

Plummer, M., Best, N., Cowles, K., Vines, K., 2006. CODA: Convergence diagnosis and output analysis for MCMC, R News. 6, 7-11.

R Core Team, 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Robert, C.P., Casella, G., 1999. Monte Carlo statistical methods. Springer texts in statistics. Springer, New York.

Salet, F.W.J., 2009. Het interpoleren van temperatuurgegevens. Report of a research internship at KNMI. KNMI Publicatie, De Bilt. http://www.knmi.nl/bibliotheek/stageversl.html.

Schabenberger, O., Gotway, C.A., 2005. Statistical methods for spatial data analysis. Texts in statistical science. Chapman & Hall/CRC, Boca Raton, Florida.

Stein, A., 1994. The use of prior information in spatial statistics. Geoderma 62, 199-216.

Stein, A., Corsten, L.C.A., 1991. Universal kriging and cokriging as a regression procedure. Biometrics 47, 575-587.

Stein, M.L., 1999. Interpolation of spatial data: some theory for kriging. Springer series in statistics. Springer, New York.

Symfony, 2012. A gentle introduction to Symfony: build professional websites faster with PHP and Symfony, Symfony 1.3 & 1.4, License: GFDL, Version: gentle-introduction-1.4-en-2012-02-26, SensioLabs.

Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. Science 185, 1124-1131.

UncertWeb, 2010. Uncertainty enabled model web – Work package 5, Deliverable D5.1, http://www.uncertweb.org/documents/deliverables.

Van de Meer, F., 2012. Remote-sensing image analysis and geostatistics. Int. J. Remote. Sens. 33, 5644-5676.

Van Groenigen, J.W., Siderius, W., Stein, A., 1999. Constrained optimisation of soil sampling for minimisation of the kriging variance. Geoderma 87, 239-259.

Vasát, R., Heuvelink, G.B.M., Boruvka, L., 2010. Sampling design optimization for multivariate soil mapping. Geoderma 155, 147-153.

Wackernagel, H., 2003. Multivariate geostatistics: an introduction with applications. Springer, Berlin.

Wall, D.H., 2004. Sustaining biodiversity and ecosystem services in soils and sediments. SCOPE-64. Island, Washington, DC.

Walter, C., Lagacherie, P., Follain, S., 2006. Integrating pedological knowledge into digital soil mapping, in: Lagacherie, P., McBratney, A.B., Voltz, M. (Eds), Developments in soil science 31, Chapter 22. Elsevier, pp. 281-615.

Walvoort, D.J.J., Brus, D.J., de Gruijter, J.J., 2010. An R package for spatial coverage sampling and random sampling from compact geographical strata by k-means. Comput. Geosci. 36, 1261-1267.

Warrick, A.W., Myers, D.E., 1987. Optimization of sampling locations for variogram calculations. Water Resour. Res. 23, 496-500.

Webster, R., 2000. Is soil variation random? Geoderma 97, 149 163.

Webster, R., Oliver, M.A., 1992. Sample adequately to estimate variograms of soil properties. Eur. J. Soil. Sci. 43, 177-192.

Webster, R., Oliver, M.A., 2007. Geostatistics for environmental scientists, Wiley, Chichester.

Webster, R., Lark, R.M., 2013. Field sampling for environmental science and management. Routledge, London.

Wilson, M.J., Kakouli-Duarte, T., 2009. Nematodes as environmental indicators. CAB International, Wallingford, 326 pp.

Wood, R., Curtis, A., 2004. Geological prior information, and its application to geoscientific problems, in: Curtis, A., Wood, R. (Eds.), Geological prior information: informing science and engineering, Geological Society. 239, pp. 1-15.

Yeates, G.W., 2003. Nematodes as soil indicators: functional and biodiversity aspects. Biol. Fertil. Soils 37, 199-210.

Summary

Geostatistics is commonly applied to predict the values of a spatial variable at unobserved locations in a study area. For this purpose, geostatistics first quantifies spatial dependence, and then based on this quantification, predicts the values at unobserved locations. The optimum use of data for spatial inference and prediction is a fundamental challenge in geostatistics because this determines the quality of model parameter estimation and prediction. Geostatisticians succeeded in improving the mapping quality by refining or extending geostatistical models and using ancillary data, of which expert knowledge is often considered as prior information and soft data. Although expert knowledge can be a valuable source of data and information in geostatistics, expert knowledge (such as about the mean and (auto)correlation of spatial variables) has so far been marginally and informally used in geostatistical research.

Meanwhile, an increasing amount of literature shows great benefit of using expert scientific knowledge in the Earth and environmental sciences, in particular when the knowledge is extracted from experts using a formal expert elicitation procedure. Statistical expert elicitation is a research field that aims at developing statistical techniques and formal procedures to elicit expert judgements on uncertain quantities in a transparent and reliable way. The rapid advance in statistical expert elicitation research can provide a sensible mechanism for the use of expert knowledge in geostatistical inference and prediction.

Besides the justification of the value of expert elicitation for geostatistics, Chapter 1 of this dissertation presents the two main objectives of this research: 1. to identify gaps in geostatistical data and accordingly, to identify the use of expert knowledge in geostatistical inference and prediction; 2. to investigate how to elicit expert knowledge and incorporate expert knowledge in geostatistics. In Chapters 2 to 5, the use of expert knowledge is investigated for four main focuses of geostatistical research: variogram estimation, spatial uncertainty quantification, spatial disaggregation and kriging prediction.

In Chapter 2, expert knowledge is used to help infer the variogram to guide optimum sampling design, in a case when observations cannot easily or cheaply be collected. An expert elicitation protocol for variogram estimation was developed, based on the bisection elicitation method and a formal elicitation procedure that is recently published in the expert elicitation literature. The protocol has two main rounds: elicitation of the marginal probability distribution, which was assumed either normally or log-normally distributed, and elicitation of the variogram. Multiple experts' knowledge were combined by equal weight pooling. The protocol was implemented as a web-based tool with three main components: a web interface, a statistical computation and mathematical pooling, and a database management. The results of mapping air temperature over The Netherlands showed that the protocol is adequate to capture expert knowledge about spatial variation, and that the online elicitation tool functions satisfactorily.

There is a lack of information about the accuracy of soil property maps that are produced by conventional, deterministic mapping methods. The common approach is to do additional fieldwork to collect validation data. However, this approach can be labour- and cost-intensive, and the validation outcome does not provide a full probabilistic description of the spatial uncertainty. In Chapter 3, a formal statistical expert elicitation procedure and the web-based tool for variogram elicitation developed in Chapter 2 are applied to extract multiple experts' knowledge about the probabilistic model description of the error in a soil property map. The spatial uncertainty about the volumetric soil water content at field capacity of the East Anglian Chalk area, The United Kingdom can be characterised by a normally distributed spatial random field, which was inferred using the experts' judgements. The web-based tool for the variogram elicitation again functioned satisfactorily in a real-world case study.

In Chapter 4, the point support variogram, of which the nugget parameter cannot be inferred from only block support observations was elicited from multiple experts' knowledge using the MATCH uncertainty elicitation tool. It was used for spatial disaggregation with area-to-point kriging. The experts' judgements were combined with block support observations in a Bayesian estimator to infer and quantify uncertainty of the point support variogram parameters. The estimation of the Matérn variogram model parameters of the air temperature over the Gelderland province confirmed that expert knowledge brought new information to infer the nugget parameter of the point support variogram, while block support observations (i.e. a 5 km resolution MODIS satellite image) only brought new information to infer the other variogram parameters. Bayesian area-to-point conditional simulation provided a satisfactory way to predict air temperature at point support and model uncertainty propagation through spatial disaggregation.

Chapter 5 addresses the issue of insufficient observations for geostatistical inference and prediction by using expert judgements as probabilistic soft data. A geostatistical model was developed to model the expert data and integrate these as additional data in cokriging. The geostatistical model used includes measures of the bias, imprecision and smoothing effects of expert judgements. The model parameters were estimated from both expert data and measured data by maximum likelihood. A case study to map the nematode structure index in a 23 ha nature area in the south of The Netherlands showed that the expert was quite uncertain about the values of the nematode structure index. While the use of expert data could largely decrease the cokriging prediction variance at areas farther away from measurement locations, the results of validation showed that the overall accuracy of the map is not improved compared to using only the measured data. More investigation on the models, the sampling design and the number of experts sharing knowledge were hence recommended in future research.

Chapter 6 concludes the research by first answering the two main research questions: 1. What is the role of expert knowledge in geostatistical inference and prediction? 2. How to elicit and incorporate expert knowledge in geostatistical inference and prediction? The formal role of expert knowledge in geostatistical inference and prediction is similar to that in other research disciplines, such as the role of surrogate, complementary or supplement data and information. Besides all difficulties that are encountered in statistical expert elicitation, the geographical dependence of spatial variables is a distinctive character that might cause difficulty in using expert knowledge in geostatistics. Elicitation of expert geostatistical knowledge can be accomplished by applying a formal expert elicitation procedure published in the literature, but a new approach for the elicitation of spatial random fields is required. Expert knowledge can be incorporated in geostatistical models (with observations) by Bayesian or geostatistical approaches (e.g. maximum likelihood, cokriging).

This dissertation showed that the added value of expert knowledge in geosta-

tistics is real. The informal and marginal use of expert knowledge in geostatistics has been replaced by a formal and systematic use, with the application of statistical expert elicitation to extract the knowledge. In so doing, the added value of expert knowledge in geostatistics can be enhanced. This dissertation also showed that existing statistical expert elicitation techniques need to be extended for application in geostatistics, as this has been done in this research. The challenge of future research lies in how to qualitatively and quantitatively maximise the added values of expert knowledge in geostatistics.

Samenvatting

Geostatistiek wordt veelvuldig gebruikt om de waarden van een ruimtelijke variabele op onbemeten locaties in een studiegebied te voorspellen. Hiertoe maken geostatistici gebruik van een kwantitatieve maat voor de ruimtelijke correlatie, het 'variogram'. Het variogram wordt eerst geschat op basis van puntgegevens uit het gebied en vervolgens gebruikt voor de ruimtelijke interpolatie. Het optimale gebruik van gegevens voor modellering van de ruimtelijke correlatiestructuur en voor ruimtelijke interpolatie is een belangrijke uitdaging in de geostatistiek, omdat deze de kwaliteit van de schatting van de modelparameters en van de interpolatie bepaalt. Geostatistici kunnen de nauwkeurigheid van de geïnterpoleerde kaarten vergroten door verbetering van het geostatistisch model en door gebruik van aanvullende kennis en gegevens, waaronder ook expertkennis. Hoewel expertkennis een waardevolle bron van informatie in de geostatistiek kan zijn, is deze kennis tot nu toe slechts marginaal en informeel gebruikt.

Ondertussen toont een groeiende hoeveelheid wetenschappelijke literatuur het grote voordeel van het gebruik van expertkennis in de aard- en milieuwetenschappen aan, in het bijzonder wanneer de kennis van experts wordt verzameld door middel van een formele 'expertelicitatie'-procedure. Statistische expertelicitatie is een onderzoeksveld dat tot doel heeft statistische technieken en formele procedures te ontwikkelen, waarmee expertkennis op een transparante en betrouwbare wijze beschikbaar wordt gemaakt. De snelle vooruitgang in het onderzoek naar statistische expertelicitatie kan ook van groot nut zijn in de geostatistiek.

Naast het aannemelijk maken van de toegevoegde waarde van statistische expertelicitatie voor de geostatistiek, presenteert Hoofdstuk 1 van dit proefschrift de twee hoofddoelen van dit onderzoek: 1. identificatie van tekortkomingen in ruimtelijke gegevens voor geostatistische analyses en hoe deze tekortkomingen met expertkennis te verhelpen; 2. ontwikkeling van methoden voor elicitatie van expertkennis voor geostatistische doeleinden. In hoofdstukken 2 tot en met 5 wordt het gebruik van expertkennis onderzocht voor vier kerntaken van geostatistisch onderzoek:

variogramschatting, ruimtelijke onzekerheidskwantificering, ruimtelijke desaggregatie en kriging-interpolatie.

In Hoofdstuk 2 wordt expertkennis gebruikt om afleiding van het variogram voor meetnetoptimalisatie te ondersteunen, in een situatie waarbij waarnemingen duur of moeilijk te verkrijgen zijn. Het hoofdstuk presenteert een protocol voor expertelicitatie van het variogram dat gebruik maakt van de zogeheten bisectie-elicitatiemethode. Het protocol bestaat uit twee ronden: elicitatie van de marginale waarschijnlijkheidsverdeling (die normaal dan wel log-normaal verdeeld wordt verondersteld), en elicitatie van het variogram. Kennis van meerdere experts werd gecombineerd door berekening van het rekenkundig gemiddelde. Het protocol werd geïmplementeerd als een webapplicatie met drie hoofdcomponenten: een webinterface, een implementatie van statistische en wiskundige berekeningen, en een databasebeheer. De resultaten van het karteren van de luchttemperatuur in Nederland toonde aan dat het protocol geschikt is om expertkennis over ruimtelijke variatie vast te leggen, en dat de online elicitatie-applicatie naar behoren functioneert.

Conventionele bodemkaarten geven vaak geen informatie over de nauwkeurigheid van de weergegeven bodemeigenschappen. Met behulp van aanvullend veldwerk kan validatiedata verzameld worden waarmee de nauwkeurigheid kan worden vastgesteld, maar deze aanpak kan arbeids- en kostenintensief zijn. Daarnaast geeft de validatie geen volledige probabilistische beschrijving van de ruimtelijke onzekerheid. In Hoofdstuk 3 zijn de formele statistische expertelicitatie-procedure en de webapplicatie voor variogramelicitatie uit Hoofdstuk 2 toegepast om kennis van meerdere experts over de probabilistische modelbeschrijving van de fout in een bodemeigenschappenkaart te extraheren. De ruimtelijke onzekerheid over het volumetrisch bodemvochtgehalte op veldcapaciteit van het East Anglian krijtgebied (Verenigd Koninkrijk) is gekarakteriseerd door een normaal-verdeeld stochastisch veld, dat werd afgeleid met expertkennis. De webapplicatie voor de elicitatie van het variogram functioneerde opnieuw naar wens.

In Hoofdstuk 4 werd het point support-variogram, dat refereert aan de ruimtelijke correlatie van puntmetingen, en waarvan de nugget-parameter niet kan worden afgeleid van alleen block support -waarnemingen, geëliciteerd met kennis van meerdere experts. De expertkennis werd met behulp van Bayesiaanse statistiek gecombineerd met block support-waarnemingen voor optimale schatting van het point support-variogram. De schatting van de parameters van het variogram van de luchttemperatuur van de provincie Gelderland bevestigde dat expertkennis waardevolle informatie bevat over de nugget-parameter van het point support-variogram, terwijl block support-waarnemingen (in dit geval een MODIS-satellietbeeld met een ruimtelijke resolutie van vijf bij vijf kilometer) slechts informatie levert over de andere variogramparameters. Na afleiding van het point support-variogram werd vervolgens met behulp van Bayesiaanse area-to-point-geostatistische interpolatie een hoge-resolutiekaart van de luchttemperatuur gemaakt.

Hoofdstuk 5 draagt een oplossing aan voor het probleem van onvoldoende waarnemingen voor geostatistische modellering en interpolatie door gebruik van expertkennis als additionele 'waarnemingen'. Er werd een geostatistisch model ontwikkeld om expertkennis als additionele gegevens in de ruimtelijke interpolatie mee te nemen. Het gebruikte geostatistische model maakt onderscheid tussen werkelijke waarnemingen en waarnemingen afgeleid uit expertkennis, door rekening te houden met systematische en toevallige fouten en het afvlakkende effect van expertoordelen. Als case study werd de structuurindex van aaltjes in een 23 ha groot natuurgebied in het zuiden van Nederland met geostatistiek gekarteerd. In deze toepassing bleek dat de geraadpleegde expert tamelijk onzeker was over de ruimtelijke verspreiding van de structuurindex in het gebied. Dit had tot gevolg dat de meerwaarde van de expertkennis in dit geval gering was zodat de nauwkeurigheid van de resulterende kaart niet verbeterde, hetgeen werd bevestigd door een onafhankelijke validatie. Meer onderzoek naar het gebruikte model, het meetnetontwerp en het aantal benodigde experts is nodig om tot daadwerkelijke verbetering te komen.

Hoofdstuk 6 rondt het onderzoek af door eerst de twee hoofdonderzoeksvragen te beantwoorden. De formele rol van expertkennis in geostatistische modellering en interpolatie komt overeen met die in andere onderzoeksdisciplines, en vervult de rol van alternatieve, complementaire of aanvullende data en informatie. Naast de gebruikelijke moeilijkheden bij toepassing van expertelicitatie brengt toepassing in de geostatistiek diverse additionele problemen met zich mee. Veel van deze problemen worden opgelost door gebruik van formele expertelicitatie-procedures, maar daarnaast zijn ook nieuwe technieken nodig zoals ontwikkeld in dit proefschrift. Expertkennis kan worden ingebed in geostatistische modellen door gebruik van Bayesiaanse technieken.

Dit proefschrift heeft aangetoond dat expertkennis een toegevoegde waarde heeft in de geostatistiek. Het informele en incidentele gebruik van expertkennis in geostatistiek is vervangen door een formele en systematische aanpak. Aldus is de toegevoegde waarde van expertkennis in de geostatistiek verhoogd. Dit proefschrift heeft ook laten zien dat bestaande statistische expertelicitatie-technieken dienen te worden aangepast voor toepassing in de geostatistiek. Hoe de toegevoegde waarde van expertkennis in geostatistiek kwalitatief en kwantitatief te optimaliseren is de uitdaging voor toekomstig onderzoek.

Tóm tắt

Địa thống kê (geostatistics) được áp dụng phổ biến trong nội suy giá trị của các biến môi trường mà biến đổi liên tục trên bề mặt Trái đất như nhiệt độ, nồng độ các chất gây ô nhiễm không khí, tính chất pH của đất, ... Thuật toán nội suy trong địa thống kê, hay còn được gọi là thuật toán nội suy kriging, bao gồm hai bước chính: 1. Phân tích định lượng sự tương quan về giá trị của một biến nào đó ở các khoảng cách khác nhau trên bề mặt không gian, còn được gọi là phân tích định lượng hàm cấu trúc hay biểu đồ phương sai theo khoảng cách giữa các điểm (variogram); 2. Dựa trên phân tích định lượng này, ước lượng giá trị của biến tại các vị trí không có dữ liệu đo đạc. Độ chính xác của kết quả phân tích định lượng hàm cấu trúc variogram và các giá trị nội suy kriging phụ thuộc vào khả năng sử dụng tối ưu các dữ liệu và thông tin sẵn có về sự biến đổi giá trị của biến môi trường trong khu vực khảo sát. Bằng cách cải tiến hoặc mở rộng các mô hình địa thống kê và sử dụng thêm các nguồn dữ liệu có liên quan khác, các nhà địa thống kê đã thành công trong việc nâng cao chất lượng của các sản phẩm bản đồ từ nội suy kriging. Trong đó, kiến thức khoa học của các chuyên gia, ví dụ như kiến thức về giá trị trung bình bề mặt và mối tương quan không gian của các biến môi trường, là nguồn dữ liệu và thông tin quý báu, nhưng chưa được sự quan tâm và sử dụng một cách có hiệu quả trong các nghiên cứu của địa thống kê.

Trong khi đó, một số lượng lớn các nghiên cứu được xuất bản gần đây cho thấy lợi ích to lớn của việc sử dụng kiến thức khoa học từ các chuyên gia trong lĩnh vực khoa học Trái đất và khoa học môi trường, đặc biệt là khi các kiến thức khoa học này được thu thập theo một quy trình có phương pháp và có hệ thống. Một lĩnh vực nghiên cứu khá mới (expert elicitation) phục vụ cho việc thu thập kiến thức khoa học từ các chuyên gia ra đời nhằm mục đích phát triển các phương pháp thống kê và các quy trình thu thập kiến thức có hệ thống. Sự phát triển nhanh chóng của lĩnh vực nghiên cứu này hứa hẹn cho khả năng thu thập và sử dụng có hiệu quả các kiến thức khoa học của các chuyên gia trong lĩnh vực nghiên cứu về địa thống kê.

Ngoài việc trình bày ý nghĩa khoa học và ý nghĩa thực tiễn của luận án này, Chương 1 còn trình bày hai mục tiêu nghiên cứu chính của luận án: 1. Xác định các dữ liệu cần và còn thiếu, từ đó đề xuất việc sử dụng kiến thức khoa học của các chuyên gia thay thế; 2.

Nghiên cứu việc áp dụng các phương pháp thống kê và quy trình có hệ thống trong việc thu thập kiến thức chuyên gia và sử dụng các kiến thức này trong các mô hình địa thống kê. Chương 2 đến Chương 5 trình bày việc sử dụng kiến thức chuyên gia trong các trọng tâm nghiên cứu của địa thống kê, bao gồm: phân tích định lượng hàm cấu trúc, phân tích định lượng tính bất định (uncertainty) của các kết quả nội suy kriging, sử dụng trong thuật toán chi tiết hóa (hay làm tăng tỷ lệ bản đồ) và sử dụng trực tiếp cho nội suy kriging.

Ở Chương 2, kiến thức chuyên gia được sử dụng để phân tích định lượng hàm cấu trúc dùng trong tối ưu hóa việc xác định các vị trí lấy mẫu, nhằm tăng hiệu quả của việc thu thập dữ liệu đo đạc trong trường hợp chi phí thu thập mẫu cao. Công cụ thu thập thông tin từ các chuyên gia được thiết lập dưới hình thức trang web dựa trên phương pháp bisection và quy trình được đề xuất trên các tài liệu nghiên cứu đã được xuất bản. Quy trình gồm có hai bước: thu thập kiến thức chuyên gia về hàm mật độ xác xuất biên, với giả thiết phân phối chuẩn hoặc phân phối loga chuẩn, và hàm cấu trúc. Kiến thức từ nhiều chuyên gia được tổng hợp dựa trên phương pháp lấy trung bình không có trọng số. Quy trình này được xây dựng dưới hình thức công cụ web, có ba thành phần chính: giao diện web, bộ phận tính toán thống kê, và cơ sở dữ liệu. Kết quả nghiên cứu cho việc lập bản đồ nhiệt độ không khí ở Hà Lan chứng minh quy trình được thiết lập là phù hợp cho việc thu thập kiến thức từ các chuyên gia cho các biến trên bề mặt không gian hai chiều, và công cụ web hỗ trợ cho quy trình này hoạt động tốt.

Đo đạc thêm dữ liệu kiểm nghiệm là phương pháp thường được sử dụng để đánh giá tính chuẩn xác của các bản đồ tính chất đất mà được xây dựng bằng các phương pháp tất định truyền thống. Tuy nhiên, việc thu thập thêm dữ liệu kiểm nghiệm có thể tốn kém nhiều chi phí và nhân công; bên cạnh đó, kết quả đánh giá tính chính xác bằng cách sử dụng dữ liệu đo đạc kiểm nghiệm không thể xác lập được dưới dạng hàm mật độ xác xuất. Trong Chương 3, quy trình và công cụ web giới thiệu ở Chương 2 được áp dụng để thu thập kiến thức từ các chuyên gia nhằm thiết lập mô hình xác xuất cho sai số có thể của các bản đồ tính chất đất. Dựa trên kiến thức từ các chuyên gia, mức độ sai số của bản đồ về khả năng giữ ẩm tự nhiên của đất cho khu vực East Anglian Chalk, Vương quốc Anh có thể được mô phỏng bằng hàm ngẫu nhiên phân bố chuẩn. Công cụ web đã hoạt động tốt khi được ứng dụng vào tình huống nghiên cứu thực tế này.

Trong Chương 4, công cụ web MATCH Uncertainty Elicitation Tool được sử dụng để thu thập kiến thức từ nhiều chuyên gia cho việc phân tích định lượng hàm cấu trúc của biến môi trường ở tỷ lệ lớn (điểm) hơn tỷ lệ thu thập dữ liệu đo đạc (khối) vì thông số nugget của hàm cấu trúc điểm không thể ước lượng được chỉ từ dữ liệu đo đạc trung bình khối. Hàm cấu trúc này sau đó được sử dụng trong thuật toán nội suy kriging ATP. Kiến thức của các chuyên gia được kết hợp với dữ liệu đo đạc trung bình khối thông qua ước lượng Bayes để định lượng các thông số của hàm cấu trúc điểm và mức độ chuẩn xác của nó. Kết quả nghiên cứu sự biến đổi nhiệt độ không khí trên khu vực tỉnh Gelderland, Hà Lan cho thấy kiến thức từ các chuyên gia cung cấp thông tin cần thiết cho việc ước lượng giá trị của thông số nugget của hàm cấu trúc Matérn ở tỷ lệ điểm; trong khi đó, giá trị đo đạc trung bình khối từ ảnh vệ tinh MODIS với độ phân giải 5 km chỉ có thể sử dụng để định lượng các thông số còn lại của hàm cấu trúc Matérn. Mô phỏng xác xuất có điều kiện ATP Bayes là phương pháp hữu hiệu trong việc dự đoán nhiệt độ ở tỷ lệ điểm và phân tích tính bất định của các kết quả dự đoán này.

Chương 5 trình bày giải pháp cho vấn đề thiếu dữ liệu đo đạc cho các ước lượng và dự đoán trong địa thống kê bằng cách sử dụng kiến thức từ các chuyên gia để bổ sung cho lượng dữ liệu đo đạc ít ỏi hiện có thông qua thuật toán nội suy cokriging. Mô hình địa thống kê được xây dựng cho việc ước lượng các thông số hàm cấu trúc và nội suy cokriging, bao gồm cả việc xác định độ chênh lệch, độ không chuẩn xác và hiệu ứng san bằng giá trị của các đánh giá từ chuyên gia. Các thông số của mô hình được ước lượng thông qua thuật toán hợp lý cực đại, sử dụng dữ liệu đo đạc kết hợp với kiến thức thu thập từ chuyên gia. Việc sử dụng kiến thức của chuyên gia trong việc lập bản đồ phân bố chỉ số giun tròn trong đất (nematode SI) ở khu bảo tồn thiên nhiên, miền nam Hà Lan cho thấy kiến thức của chuyên gia có độ bất định cao. Việc sử dụng kiến thức này có khả năng làm giảm đáng kể phương sai của thuật toán nội suy cokriging ở những khu vực xa vị trí lấy mẫu, nhưng độ chuẩn xác chung của toàn bản đồ không được cải thiện đáng kể. Kiến nghị cho các nghiên cứu tiếp theo là cần có sự điều chỉnh về sự phân bố các vị trí trong khu vực khảo sát mà chuyên gia cần đánh giá và số lượng các chuyên gia tham gia trong tình huống nghiên cứu này.

Chương 6 trình bày phần kết luận, bao gồm câu trả lời cho hai câu hỏi nghiên cứu chính của luận án: 1. Kiến thức khoa học của chuyên gia đóng vai trò gì trong địa thống kê? 2. Kiến thức khoa học của chuyên gia có thể thu thập và sử dụng trong địa thống kê như thế nào? Vai trò của kiến thức từ chuyên gia trong địa thống kê cũng tương tự như trong các lĩnh vực nghiên khác, đó là vai trò thay thế, bổ sung toàn phần hay bổ sung một phần dữ liệu và thông tin cần thiết. Ngoài những khó khăn thường gặp trong việc thu thập kiến thức khoa học từ chuyên gia, sự tương quan về mặt địa lý của giá trị các biến môi trường là một đặc tính quan trọng và có thể gây khó khăn cho việc sử dụng kiến thức của chuyên gia trong địa thống kê. Quy trình thu thập kiến thức từ các chuyên gia đã được xuất bản có thể áp dụng trong các nghiên cứu của địa thống kê, nhưng đòi hỏi phải có các kỹ thuật thu thập mới phù hợp cho các biến không gian. Kiến thức của chuyên gia sau khi được thu thập có thể kết hợp với các dữ liệu đo đạc có sẵn trong các mô hình địa thống kê thông qua thuật toán Bayes hay các thuật toán địa thống kê (ví dụ như phương pháp hợp lý cực đại, nội suy cokriging).

Tóm lại, kết quả nghiên cứu của luận án này đã cho thấy giá trị thực tiễn của việc sử dụng kiến thức của chuyên gia trong địa thống kê. Trong luận án này, nguồn kiến thức quý báu từ các chuyên gia đã được sử dụng một cách chính quy và có hệ thống, thay thế cho việc sử dụng không đáng kể trước đây. Bằng cách này có thể nâng cao được giá trị thực tiễn của các kiến thức khoa học được sử dụng trong địa thống kê. Luận án này cũng chỉ ra rằng cần xây dựng các phương pháp thu thập kiến thức phù hợp cho việc ứng dụng trong các nghiên cứu địa thống kê. Thách thức cho các nghiên cứu sắp tới trong lĩnh vực này là làm thế nào để tối đa hóa việc sử dụng cũng như giá trị sử dụng của nguồn kiến thức khoa học từ các chuyên gia trong địa thống kê.

Publications

Journal articles

Truong, P.N., Heuvelink, G.B.M., Gosling, J.P., 2013. Web-based tool for expert elicitation of the variogram. Comput. Geosci. 51, 390-399.

Truong, P.N., Heuvelink, G.B.M., 2013. Uncertainty quantification of soil property maps with statistical expert elicitation. Geoderma 202-203, 142-152.

Truong, P.N., Heuvelink, G.B.M., Pebesma, E., 2014. Bayesian area-to-point kriging using expert knowledge as informative priors. Int. J. Appl. Earth Obs. GeoInf. 30, 128-138.

Conference contributions

Skoien, J., **Truong**, P.N., Dubois, G., Cornford, D., Heuvelink, G.B.M., Geller, G., 2011. Uncertainty propagation in the model web - A case study with eHabitat, in: Proceedings of the 34th International symposium on remote sensing of environment, Sydney, Australia, 10-15 April 2011.

Bastin, L., Williams, M., Gosling, J.P., **Truong**, P.N., Cornford, D., Heuvelink, G.B.M., Achard, F., 2011. Web-based expert elicitation of uncertainties in environmental model inputs. Geophysical Research Abstracts 13, 5384.

Skoien, J., Dubois, G., **Truong**, P.N., Cornford, D., Shulz, M., 2011. Uncertainty propagation in assessing ecosystems similarities - an example application using eHabitat. Geophysical Research Abstracts 13, 11580.

Truong, P.N., Heuvelink, G.B.M., Gosling, J.P., 2012. Expert elicitation for the variogram, in: Proceedings of the 10th International symposium on spatial accuracy assessment in natural resources and environmental sciences, 10-13 July 2012, Florianopolis, Brazil.

Truong, P.N., Heuvelink, G.B.M., Pebesma, E., 2012. Influence of point-support variogram on spatial disaggregation with ATP kriging, in: GI Zeitgeist 2012: Young researchers forum on geographic information science, Münster, Germany, 16-17 March 2012.

Truong, P.N., Heuvelink, G.B.M., 2013. Bayesian area-to-point kriging using expert knowledge as informative priors. Geophysical Research Abstracts 15, 2291.

Professional publications

Heuvelink, G.B.M., Bastin, L., Cornford, D., Gossling, J.P., **Truong**, P.N., Williams, M., 2011. Expert elicitation of input and model uncertainties. The uncertainty enabled model web, Deliverable 3.1, pp 61.

Stasch, C., **Truong**, P.N., Pebesma, E., Heuvelink, G.B.M., 2011. Spatio-temporal aggregation of uncertainties. The uncertainty enabled model web, Deliverable 3.2, pp 52.

Gerharz, L., Senaratne, H., Autermann, C., **Truong**, P.N., Heuvelink, G.B.M., Williams, M., Pebesma, E., Stasch, C., Cornford, D., 2012. Tools for communicating and visualising uncertainties. The uncertainty enabled model web, Deliverable 3.3, pp 40.

PE&RC Training and Education Statement

With the training and education activities listed below, the PhD candidate has complied with the requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities)

Review of literature (6 ECTS)

- Uncertainty analysis in ecosystem services modelling (2010)

Writing of project proposal (4.5 ECTS)

- Uncertainty analysis in ecosystem services modelling (2010)

Post-graduate courses (7.2 ECTS)

- Bayesian statistics; PE&RC (2010)
- Geostatistics; PE&RC (2010)
- Sampling in space and time for survey and monitoring of natural resources; PE&RC (2012)
- 7th Summer school on sensitivity analysis; JRC in Ispra, Italy (2012)
- ABS12-2012 Applied Bayesian statistics school: stochastic modelling for systems biology; IMATI CNR & Università Cattolica, Milano, Italy (2012)
- Multivariate analysis; PE&RC (2013)

Laboratory training and working visits (2.4 ECTS)

- UncertWeb coding marathon workshop; Aston University (2010)
- Expert elicitation working visit; Food and Environment research Agency (FERA) (2010)
- Uncertainty visualization working visit; Muenster University (2010)
- UncertWeb workshop; Muenster University (2012)

Invited review of (unpublished) journal manuscript (1 ECTS)

- Geophysical Journal International: Eliciting spatial statistics from geological experts using genetic algorithms (2013)

Competence strengthening/skills courses (3.2 ECTS)

- Techniques for writing and presenting a scientific paper; PE&RC (2010)
- PhD Competence assessment; PE&RC (2010)
- Scientific writing; WUR (2011)

PE&RC Annual meetings, seminars and the PE&RC weekend (2.1 ECTS)

- PE&RC Introduction weekend (2010)
- PE&RC Day: innovation for sustainability (2011)
- Scientific publishing workshop (2012)
- PE&RC Weekend for candidates in their last years (2013)

Discussion groups/local seminars/other scientific meetings (4.5 ECTS)

- UncertWeb project meetings (2010-2012)

International symposia, workshops and conferences (9 ECTS)

- Uncertainty in computer models conference: managing uncertainty in complex models (2010)
- Spatial statistics: mapping global change (2011)
- GI Zeitgeist: young researches forum on geographic information science (2012)
- Spatial accuracy (2012)
- EGU General Assembly (2013)

Lecturing/supervision of practical's/tutorials (3 ECTS)

- Environmental data collection and analysis (2011, 2012)

Acknowledgements

On the accomplishment of this dissertation, I am grateful to:

My (co) promotors: Gerard Heuvelink and Peter de Ruiter;

The co-authors of journal articles: John Paul Gosling and Edzer Pebesma;

The co-workers and contributors: Ron de Goede, Jorge Mendes de Jesus, Israel Ikoyi, Jill Johnson, Sarah Knight, Hannes Reuter and Mathew Williams;

The experts who participated in the case studies;

Colleagues from the UncertWeb project (www.uncertweb.org);

Colleagues from Soil Geography and Landscape group, Wageningen University;

My friends in Wageningen;

Thesis proofreader: Christos Kissoudis;

Dutch translators for the thesis summary: Paul Link and Gerard Heuvelink;

Proofreader for the Vietnamese translation of the thesis summary: Thu-Phuong Nguyen;

Paranymphs: Paul Link and Thu-Phuong Nguyen;

Homeland affiliation: Can Tho University, Viet Nam.

My gratitude for my parents, brother, sister-in-law and little nephew in Viet Nam:

Cám on Ba Mẹ, Anh Hai, Chị Hai và cháu Phúc Hưng!

About the author

Phuong Ngoc Truong (in Vietnamese: Trương Ngọc Phương) was born in 20 June, 1981 in Can Tho city, Viet Nam. She grew up and finished her undergraduate study in the same city. In 2004, she got a Bachelor's degree in Environmental Engineering in Can Tho University, Viet Nam. After that, she became a junior lecturer at Can Tho University. In 2007, she got a scholarship from The Netherlands Fellowship Programmes to attend a Master programme in Wageningen University, The Netherlands. In 2009, she obtained a Master's degree in Geo-information Science. Her M.Sc. thesis was awarded a Hissinkprijs prize of 2009 by The Dutch Soil Association (NBV). In 2010, she joined the Land Dynamics group (now Soil Geography and Landscape (SGL) group), Wageningen University to take part in the UncertWeb project (www.uncertweb.org), of which SGL was one of the project partners, and conduct her PhD. In 2012, she contributed to the establishment of the R Users discussion group in Wageningen University and remained active in the group till 2013. Her hobbies during the time in Wageningen were cycling in the Dutch national parks in summer time, playing squash and travelling. Her favourite place in The Netherlands is the Texel island.

The research described in this thesis was financially supported by the European Union Seventh Framework Programme [FP7/2007-2013] under grant agreement no. 248488 from 2010 to 2013.

Financial support from Wageningen University for 2013-2014 and printing this thesis is gratefully acknowledged.

Cover photo © NLshop/Fotolia.com