Chapter 1

Pedometric mapping

"I am a pedomagician"

[by A. McBratney in Pedometron # 14 "Pedometrics in a sentence", available via www.pedometrics.org]

1.1 Soil mapping

Soil mapping or soil survey[1] is *a process of determining the spatial distribution of physical, chemical and descriptive soil properties and presenting it in understandable and interpretable form to various users* (Beckett, 1976; Dent & Young, 1981). It, in general, consists of the following steps:

1. Project planning;
2. Preparation for fieldwork;
3. Photo-interpretation and pre-processing of auxiliary data;
4. Field data collection and laboratory analysis;
5. Data input and organization and
6. Presentation and distribution of soil survey products.

Project planning is especially important step for a success of soil survey project as it includes selection of sampling plan, inspection density, classification system and data organization system. Preparation for fieldwork typically includes literature study and reconnaissance surveys. The end product of a soil mapping project is a soil resource inventory, i.e. a map showing distribution of soils and its properties accompanied by a soil survey report (Avery, 1987; Rossiter, 2001).

In the age of information technologies, the soil resource inventory data is organized into a thematic type of a geoinformation system (GIS) called a **Soil Information System** (SIS), of which the major part is a **Soil Geographical Database** (SGDB) (Burrough, 1991). This is, in most cases, a combination of polygon and point map linked with attribute tables for profile observations, soil mapping units and soil classes. Often the soil mappers extend their expertise to the land use planning and decision making activities, so that a SIS not only offers information on soils but also on their potential (and actual) use, environmental risks involved (e.g. erosion risk) and gives prediction of soil behavior on intended management.

Soil mapping projects differ in the inspection intensity levels, purpose and type of conceptual models used. Considering the intensity level, soil mapping projects typically range from small scale (1:100 K to 1:1 M) surveys to medium (1:50 K) and large scale surveys (1:25 K to 1:5 K or larger). Considering the intended purpose, a soil mapping project can be classified as the special-purpose (commonly referred to as thematic) and general-purpose. The first is completely demand-driven and focuses

---

[1] See the definition of terms (at the beginning of the book) used consistently throughout the thesis.

on a limited set of soil variables or a single soil variable, often ignoring soil boundaries and soil horizons. The general-purpose mapping is more holistic, but also more complex, hence more expensive and often not affordable at large scales. The conceptual models of soils reflect the purpose of the mapping project: (i) special-purpose mapping projects commonly follow the continuous model of spatial variation, thus geostatistical techniques are used to make predictions; (ii) general-purpose mapping projects commonly rely on photo-interpretation and profile descriptions, following the discrete model of spatial variation.

Coping with soil variation has not been an easy task from the beginning of the soil survey. Soil variables vary not only horizontally but also with depth, not only continuously but also abruptly. If compared to vegetation or land use mapping, soil mapping requires much denser field inspections. Moreover, soil horizons and soil types are fuzzy entities, often hard to distinguish or measure. Especially the polygenetic nature of soils has always been a main problem in description and classification of soils (White, 1997). In fact, many pioneer soil geographers have wondered whether we will ever be able to fully describe the patterns of soil cover (Jenny, 1941). The quality and usefulness of the polygon-type soil maps (area partitions) has for decades been an object of argue (Webster & Beckett, 1968). The technological and theoretical advances in the last 20 years, however, have lead to a number of new methodological improvements in the field of soil mapping. Most of these belong to the domain of the new emerging discipline — pedometrics.

1.2 What is Pedometrics?

Pedometrics, a term coined by Alex B. McBratney, is a neologism, which stems from the Greek words $\pi\epsilon\delta o\varsigma$ [soil] and $\mu\epsilon\tau\varrho o\nu$ [measurement]. It is formed and used analogously to other applied statistical fields such as biometrics, psychometrics, econometrics and others (Webster, 1994). The most recent definition of pedometrics, available via the website of the Pedometric society (www.pedometrics.org), is:

> **“the application of mathematical and statistical methods for the quantitative modelling of soils, with the purpose of analysing its distribution, properties and behaviors”**

The domain of pedometrics changed somewhat since its foundation. At the moment, pedometrics is best defined as an interdisciplinary field between soil science, applied statistics/mathematics and geo-information science (Fig. 1.1). This means that it gathers many different scientific fields, ranging from geostatistics to soil microbiology. The domain of pedometrics, however, is not limited to only these three general sciences, as McBratney stated in his first communication: *“It can* *include numerical approaches to classification — ways of dealing with a supposed deterministic variation…the definition is certainly incomplete but as the subject grows its core will become well defined"* (preface of Geoderma, 1994: 62).

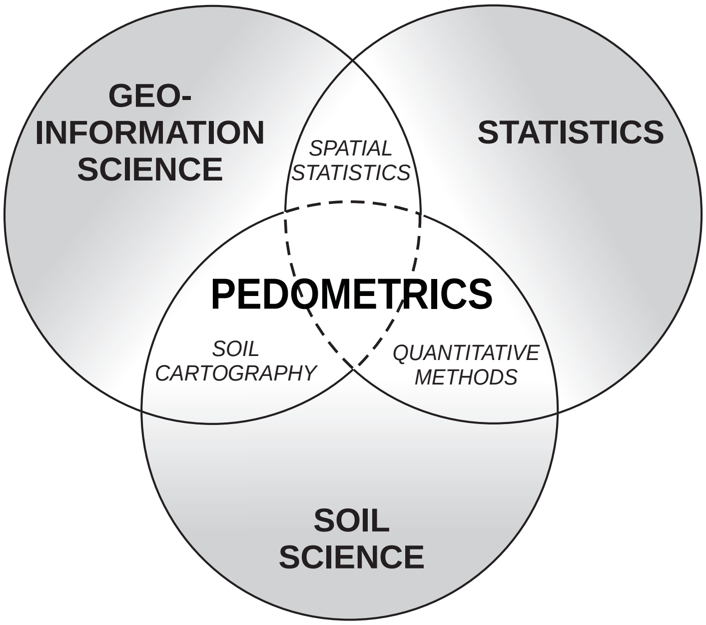

Figure 1.1: Pedometrics can be considered an interdisciplinary science between soil science, applied statistics and geoinformation science.

Another way of looking at pedometrics is to see it as implementation of newly emerging scientific theories, such as wavelets analysis and fuzzy set theory, in soil data modelling applications (Fig. 1.2). The development of pedometrics is also a result of new technological discoveries and improvements, remote and close-range sensing techniques, GPS positioning and computers in general (Burrough *et al.*, 1994). The expansion of new applications in the early 90's has made pedometrics one of the leading sub-disciplines in the area of soil research (Hartemink *et al.*, 2002). Pedometrics is promoted and communicated via publications, conferences and workshops organized by the Pedometrics society, a working group under the International Union of Soil Sciences (IUSS). After a decade of existence and numerous conferences and workshops, this Working Group has been promoted, at the 17th World Congress of Soil Sciences, to become a Commission under the IUSS.

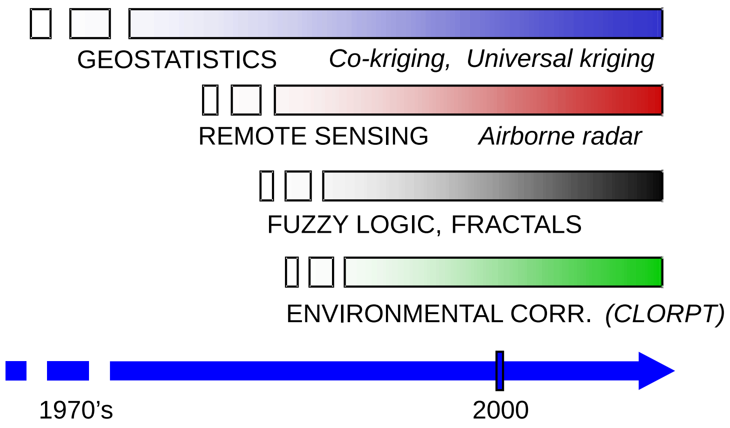

Figure 1.2: Some new emerging scientific fields that can be related to the development of pedometrics in the last decades.

Most recent topics covered by pedometrics include: multiscale data integration; use of wavelets transforms to analyse complex variation; soil-landscape modelling using digital terrain analysis; quantification of uncertainty and fuzziness of information and evaluation criteria; soil genesis simulation; soil pattern analysis; design and evaluation of sampling schemes; incorporation of exhaustively sampled information (remote sensing) in spatial interpolation; precision agriculture applications and others. A major topic of pedometric research is the development of models and tools that can deal with the spatio-temporal variation of soils (McBratney *et al.*, 2000). These tools and methods can then be implemented to improve or replace conventional soil mapping.

1.3 Pedometric mapping

Pedometric mapping is generally characterised as a quantitative, (geo)statistical production of soil geoinformation. It usually finishes with a fine-grain raster map and a measure (map) of uncertainty. Pedometric mapping is also referred to as digital soil mapping, as it heavily depends on the use of information technologies, although pedometric mapping specifically means that primarily quantitative methods are used in the production of soil geoinformation.

In recent years, digital soil mapping has faced rapid development of new and economic methods, mainly due to the increasing sources of auxiliary maps. Here, two main groups have played a key role: terrain parameters and remote sensing images (Dobos *et al.*, 2000). The terrain parameters are DEM-derived products that can be used to quantify (geo)morphology of the terrain, i.e. accumulation and deposition potential or adjust influence of climatic factors on the local terrain, while the remote sensing images reflect surface roughness, colour, moisture content and other surface characteristics of soils.

Although it was originally expected that remote sensing would revolutionize soil mapping, as it had done for vegetation mapping, direct derivation of soil properties from the remote sensing data is still limited to areas of low vegetation cover, such as grasslands, semi-deserts or agricultural plots in fallow. Apart from some specific cases, such as using radar images to map soil moisture content (Hu *et al.*, 1997), it has not yet proved possible to directly use images of visible and infrared part of spectra to map soils in all parts of the study area. This is due to the complex illumination structure caused by terrain, cloud interference and atmospheric attenuation, or reflectance of vegetation (Skidmore *et al.*, 1997; Moran *et al.*, 2002). However, compound indices such as NDVI that generally reflects biomass status, have been shown to correlate well with the distribution of the organic matter or epipedon thickness (McKenzie & Ryan, 1999). Even the coarse (1x1 km) AVHRR data have shown to be useful for mapping the clay content, CEC, EC or pH (Odeh & McBratney, 2000). A logical further development was to combine DEM-derived and remote sensing data to improve prediction models (Dobos *et al.*, 2000). The use of terrain data and remote sensing imagery has been especially interesting for medium scale-surveys (grid resolutions from 20–200 m), although there is an increasing number of field-site (precision agriculture) studies also (Fig. 1.3).

Pedometric approach to soil mapping is fairly different from the conventional approach. For a long time, the term *pedometrics* has been used as a challenge and contradiction of soil taxonomies, i.e. traditional systems. The key differences between the two approaches are summarized in Table 1.1. The conventional soil survey relies on photo-interpretation and prediction of soil types, while the pedometric techniques are (still) primarily focused on soil properties, produced using some (geo)statistical technique. The conventional survey typically leads to a polygon-based soil map and products of pedometric techniques are fine grain maps of soil properties.

1.4 Motives for the research

In recent years, there have been strong moves towards quantifying soil data: *"there has been corresponding increase in the demand for quantitative information at finer and finer resolutions"* (McBratney *et al.*, 2000). Even in USA, surveyors expect a full transition to a quantitative (pedometric) survey in the 21st century (Indorante *et al.*, 1996). Regardless of the many appeals to abandon the conventional

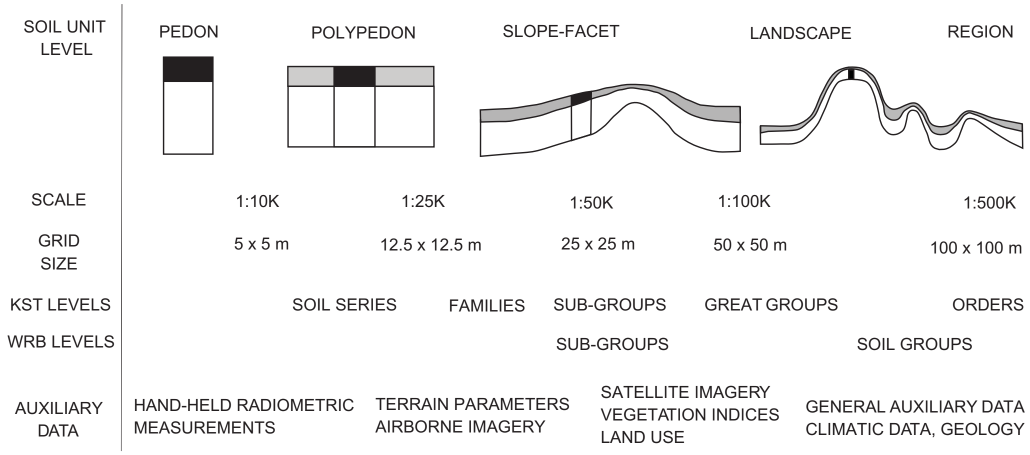

Figure 1.3: Relationship between the level of soil objects, scale, grid resolution and auxiliary maps used. Corresponding classification levels for Keys to Soil Taxonomy (KST) and World Reference Base (WRB) are also given.

approach to mapping with mapping units, this approach is still more popular for most soil survey agencies. There are two most probable reason why the pedometric techniques, in the major of the World, are still in the process of testing. First, as emphasized by Burrough (1993a): *"In spite of a huge research literature, knowledge about soil variability is still dispersed and not well organized. There is a need to organize and systematize our knowledge on soil variability in such a way that users of soil information unskilled in geostatistics and chaos theory can make the best possible decisions under conditions of uncertainty."* Second, pedometric techniques are still inappropriate to model specific soil features such as irregular soil stratigraphy, buried horizons, abrupt transitions between soils, fossil or karstic soils. These soil features and processes are still much easier to map (and generalize) using a mental model and photo-interpretation rather than geostatistics or auxiliary variables. The conventional soil mapping and classification have proven to be successful and popular, especially in the U.S. and Canada, where even the local farmers recognize different soil series. To experienced surveyors, it really seems that there is no reason to change these systems. What is clearly needed is a compromise between the new methods (pedometric approach) and experienced soil survey teams (conventional approach).

Yet, integration of pedometric and conventional methods for operational surveys has not been considered by many. De Bruin (2000) emphasized the importance of

Table 1.1: Comparison of pedometric and (analogue) conventional approach to soil survey.

	PEDOMETRIC APPROACH	CONVENTIONAL APPROACH
Preparation and project planning	Identification of key soil environmental variables (predictors)	Identification of key soil-forming factors (e.g. Catena concept)
Production of auxiliary data (pre-processing)	Remote sensing images; terrain parameters derived from a DEM; geological data etc.	Photo-interpretation; reconnaissance survey
Sampling design	Design-based (random sample, stratified random sample) or model-based (equal area stratification) sampling	Free survey
Field data collection and laboratory analysis	Navigation to points using a mobile-GIS (GPS receiver attached to a palm PC)	Navigation to points using aerial photos
Data input and organization	Data analysis and interpolation using some (geo)statistical technique	Designation of soil mapping units and theirs composition
Presentation and distribution of soil survey products	Fine grain maps of soil variables with estimate of uncertainty (thematic mapping)	Polygon map with attributed soil properties (averaged)

combining these techniques, noting that there are *"disciplinary gaps between the different techniques."* Even within the pedometric approach, there are somewhat isolated techniques that need to be combined. A good example is the gap that still exists between the CLORPT techniques and geostatistics. As de Gruijter stated in the preface of the Pedometrics '97 — International Conference held in Wisconsin, USA: *"...the second major theme of the Conference focused on spatial prediction methods. It was clear that there were two (somewhat) distinct approaches... The first is the geostatistical...the second is what Alex McBratny called 'clorp(t) approach, named from Jenny's equation or environmental regression... The synthesis of these two approaches was not really discussed. This will be an area for much further research in Pedometrics"*. Both approaches have its advantages and disadvantages (Table 1.2). A disadvantage of ordinary kriging, for example, is that it ignores spatial variation of environmental factors, e.g. relief. Moreover, conventional geostatistical techniques have shown to be inefficient at smaller scales (Yost *et al.*, 1982). A discouraging aspect of the conventional geostatistical techniques is that predictions are often non-unique. A kriging-based prediction map depends on numerous parameters, such as lag spacing, variogram function, kriging method etc. All these depend on the geostatistician's opinion rather than the properties of the data (Goovaerts, 1997). A drawback of the plain CLORPT techniques, on the other hand, is that they ignore spatial location of points and spatial autocorrelation of residuals.

Table 1.2: Comparison of some aspects of the conventional geostatistical and plain regression spatial prediction approaches.

GEOSTATISTICAL APPROACH	CLORPT APPROACH
Requires spatial dependence	Requires significant correlation with the auxiliary data
Higher sampling density desirable	Lower sampling density desirable
Data-driven	Knowledge-driven
Stratification desirable	One model over entire area
Deals with geographical space	Deals with feature space
Aims at spatially correlated random part of variation	Aims at structural part of variation (drift or trend)
Requires stationarity	Requires non-stationarity
Kriging variance reflects a geometry of the point locations while ignoring environmental patterns	Prediction error reflects the 'distance' of the point locations in the feature space while ignoring theirs spatial location
Numerous input parameters such as lag spacing, variogram function model, limiting distance, interpolation method, anisotropy model etc. are required (the predictions are non-unique for the same data set)	For linear regression, in general, no input parameters are required (predictions are unique for the same data set); however, functional relationship between the auxiliary maps and soil variables is unknown and might differ for similar datasets

Another conceptual gap in soil mapping is that between the human perception of soil types and true nature of soils. One solution to the hidden and 'fuzzy' nature of soils is to use conceptual models that are more general: *"In order to bridge the gap, soil distribution modelling should be based on a new classification paradigm: that of a fuzzy set theory"* (de Gruijter *et al.*, 1997). How to operationalize such a system for routine survey and is universal method that can handle any type of soil data possible? McKenzie & Ryan (1999) think that, considering the natural complexity of soils and soil properties, *"the development of models for spatial prediction that are quantitative, mechanistic and mathematical is almost an impossible task in routine survey."* At this level of technology and knowledge, a development of hybrid or semi-automated, semi-subjective expert systems that integrate empirical surveyor's knowledge on soils and GIS tools is only feasible solution. Indeed, *"Solving the full system of multivariate equations needed to describe the products of soil genesis in individual regions, let alone globally, remains one of the biggest challenges for pedometricians"* (Webster, 1994). This thesis is an attempt to bridge the gaps between the empirical and mechanistic methods and improve the practice of soil mapping.

There are also practical motives to develop a flexible mapping methodology that can adopt existing data sets. In Croatia, there are about 10 K profiles described, analysed and classified during the 70's, 80's and 90's (National soil inventory). This data is still not used spatially for soil prediction. In recent years, there has been lots of intention to increase the effective scale of the Basic soil map of the Croatia to a regional level, which is in this case a county level. There are 20 counties in Croatia and if the methodology proves to be successful, soil geoinformation could be improved in detail and brought to the 1:100 K effective scale or even less i.e. field resolution of 20–50 m. Similarly, a large amount of high quality soil field data in the World exists, which could be improved if the methodology proves to be successful.

1.5 Objectives

The main objective of this research was to develop a methodology for pedometric mapping that can be used to bridge the gaps between the mechanistic pedometric and conventional techniques and that can be used for operational soil mapping. Specific objectives, addressed more closely in each chapter, are:

- To develop methodology for optimal point allocation in both feature and geographical space and recommend sampling strategies for the general-purpose survey;
- To develop a systematic methodology to remove artefacts and inaccuracies in the terrain parameters used for soil-landscape modelling;
- To enhance the use of terrain analysis for photo-interpretation in soil survey;
- To develop and test generic interpolation algorithms that optimally employ both correlation with auxiliary maps and spatial dependence and can be used in a user-friendly manner;
- To provide a basis for integration of soil expertise (soil classification, photo-interpretation) and pedometric methods (regression-kriging, terrain analyis, pedo-transfer functions);

- To suggest methods to derive suitable grid resolution and investigate issues of combining multi-scale data sources;
- To develop methodology to visualise fuzziness and uncertainty of soil information and enhance production of the continuous soil maps;
- To develop methodology to assess the adequacy of soil maps and investigate the problems related to the usability of soil maps;

1.6 Outline of the thesis

This thesis was produced as a compilation of seven research papers, all written by myself as the principal author. All papers have been submitted to international peer-reviewed journals and have either been accepted for publication or are in a review process. Although the content of the thesis chapters and submitted papers does not differ in its essence, I have made some minor changes in the text so the thesis would make a coherent harmony. I have also reduced some sections in the original papers to avoid a thematic overlap and repetition of phrases and statements. The list of the seven research topics can be seen in Fig. 1.4. Note that all these are primarily methodological, not dependent on a specific study or scale. The research chapters are preceded by a definition of terms and concepts used and general introduction to soil mapping and pedometric techniques. The readers are suggested to refer to the definition of terms at the beginning of the book to avoid terminological confusion.

CHAPTER 2: SAMPLING This chapter gives a comparison of possible sampling strategies for the purpose of spatial prediction by correlation with auxiliary maps. This extends the existing sampling optimisation methodology to the issue of spreading in the feature space. The chapter demonstrates how allocation of points in the feature space influences the efficiency of prediction (overall prediction error). It suggests how to represent spatial multivariate soil forming environment; how to optimise sampling design for environmental correlation and which sampling strategies should be used for a general soil survey purposes. The concepts are illustrated using a 50×50 km study area in Central Croatia, four predictors (elevation, temperature, NDVI and CTI) and one target variable (organic matter in the top-soil).

CHAPTER 3: PRE-PROCESSING Because the pedometric mapping heavily relies on auxiliary maps, their quality plays an important role for the success of mapping. How do the inaccuracies and artefacts in auxiliary variables affect the prediction process and how to reduce these problems? In this chapter, systematic methods for reduction of errors (artefacts and outliers) in digital

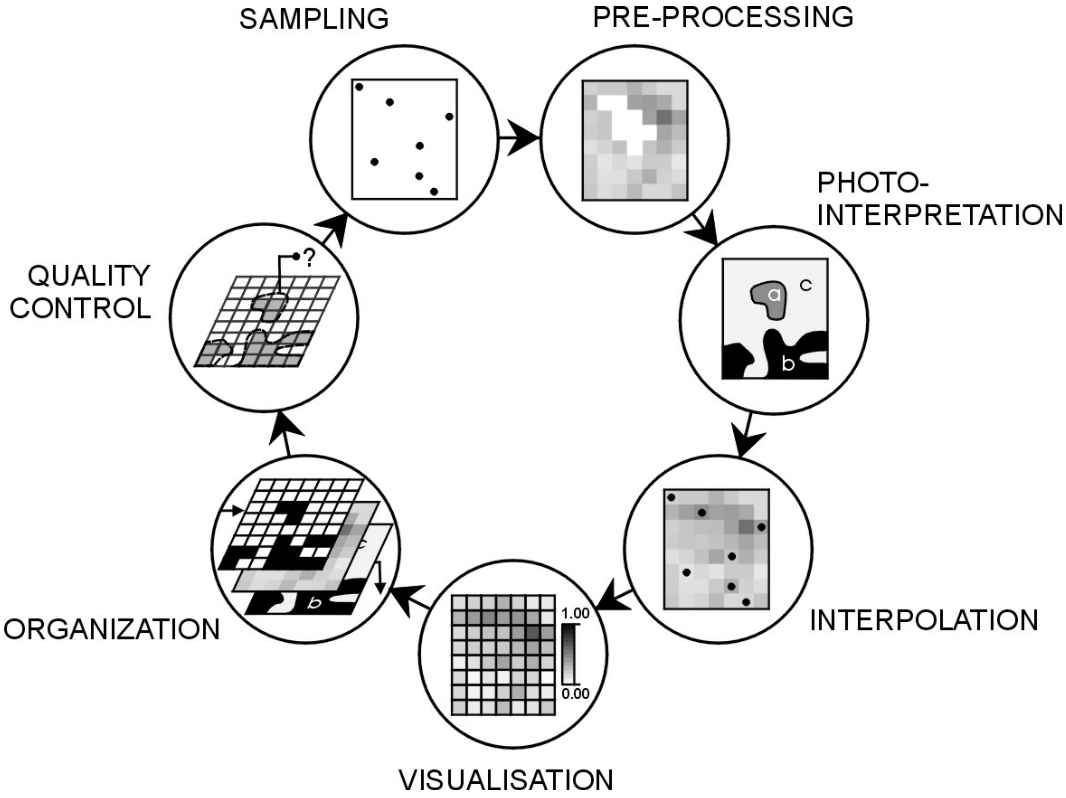

Figure 1.4: Schematic outline of the topics discussed in the thesis.

terrain parameters are suggested. These methods ensure more natural and more complete representation of the terrain morphology, which then reflects on the success of spatial prediction also. The effects of errors in the terrain parameters on mapping landform facets and predicting the thickness of the solum are demonstrated using the Baranja Hill study area (3.8×3.8 km square).

CHAPTER 4: PHOTO-INTERPRETATION Delineation of landform facets through the photo-interpretation is the key step to determine soil boundaries in a conventional soil survey. It relies on subjective impression of the terrain shapes and mapper's experience with the specific study area. Can the subjective delineation of landform facets be improved with the help of terrain analysis? Moreover, should we aim at replacing photo-interpretation or search for a compromise solution? This chapter suggests a semi-automated method to extrapolate photo-interpretation from limited number of study sub-areas to the whole area. The intention was to enhance and not to replace mapper's knowledge and expertise. The map of landform facets was produced using nine terrain parameters for Baranja region (1062 km^2) in Eastern Croatia.

CHAPTER 5: INTERPOLATION This chapter considers development of a flexible statistical framework for spatial prediction, which should be able to adopt both continuous and categorical soil variables. It suggests methods to deal with non-normality of input data and multicollinearity of predictors. The logit transformation is suggested as step to prevent predictions outside the physical limits. How well does this framework performs in real case studies and does it really improves the prediction efficiency? The framework was evaluated using the 135 profile observations of organic matter, pH and topsoil thickness from a 50×50 km study area in Central Croatia.

CHAPTER 6: VISUALISATION In conventional soil mapping, colours in the choropleth maps are typically selected following the human perception of soils. Continuous classification of soil classes, e.g. by using the fuzzy k-means, has shown to have numerous advantages for mapping soil bodies. The result of continuous classification, however, is a set of membership maps that can be hard to visualise and manipulate at once. In this chapter, an algorithm to visualize multiple memberships and analyse geographical and thematic confusion is suggested. Multiple memberships are visualized using the Hue-Saturation-Intensity model and GIS calculations on colours. This colour mixing was demonstrated using the landform classification of nine landform facets in the Baranja hill study area (3.8×3.8 km square).

CHAPTER 7: ORGANIZATION This chapter brings together methods from the chapters 2, 5 and 6. It answers how to select a suitable grid size, how to aggregate and disaggregate soil information and what are the advantages and disadvantages of a grid-based SIS. Concepts, operations and organizational structure of a hybrid grid-based soil information system (SIS) are first described. The prediction maps are then made using both photo-interpretation and auxiliary maps, which ensures both continuous and crisp transitions. The grid-based SIS was produced using a soil survey data (59 profile observations) of Baranja hill and compared with a SIS produced using the conventional methodology.

CHAPTER 8: QUALITY CONTROL In this chapter, systematic steps to assess the effective scale, accuracy of soil boundaries, accuracy of map legends, thematic purity of mapped entities and overlap among the adjacent entities are suggested. This assessment was based on number of control surveys including the control profile observations and photo-interpretations. The adequacy and usability of soil resource inventories was assessed for the extensive National soil inventory in Croatia. This was done by:

- examining the average delineation area of six map sheets;
- comparing soil data from ten control profile observations with the original profile observations;
- examining the thematic overlap between the adjacent mapping units using the data from 2198 profile observations and
- evaluating the accuracy of soil boundaries and map legends using the three control survey sub-areas.

CHAPTER 9: CONCLUSIONS AND DISCUSSION In the last chapter general conclusions related to question posed-above are given. This extends to a discussion on limitations of this research, unexpected and conflicting findings. I finally give some recommendations for a further research and emphasize research problems in the area of pedometric mapping that still need to be tackled.