Summary

Hengl, T. 2003. Pedometric mapping: bridging the gaps between the conventional and pedometric approaches. PhD thesis, Wageningen University.

In recent years, digital soil mapping has faced rapid development of new and economic methods, mainly due to the increasing sources of auxiliary maps. The main objective of this research was to develop a methodology for pedometric mapping that can be used to bridge gaps between the mechanistic pedometric and conventional techniques. The thesis covers seven methodological aspects of soil mapping: sampling, pre-processing, photo-interpretation, interpolation, visualisation, organisation and quality control.

SAMPLING: This chapter evaluates spreading of observations in feature and geographical spaces as a key to sampling optimisation for spatial prediction by correlation with auxiliary maps. Although auxiliary data are commonly used for mapping soil variables, problems associated with the design of sampling strategies are rarely examined. When generalized least squares estimation is used, the overall prediction error depends upon spreading of points in both feature and geographical space. Allocation of points uniformly over the feature space range proportionally to the distribution of predictor (equal range stratification or ER design) is suggested as a prudent sampling strategy when the regression model between the soil and auxiliary variables is unknown. An existing 100-observation sample from a 50×50 km soil survey in central Croatia was used to illustrate these concepts. It was re-sampled to 25-point datasets using different experimental designs: ER and two response surface designs (minmax and D2). The designs were compared for their performance in predicting soil organic matter from elevation (univariate example) using the overall prediction error as an evaluation criterion. The ER design gave similar overall prediction error as the minmax design, suggesting that it is a good compromise between accurate model estimation and minimisation of spatial autocorrelation of residuals. In addition, the ER design was extended to the multivariate case. Four predictors (elevation, temperature, wetness index and NDVI) were transformed to standardised principal components. The sampling points were then assigned to the components in proportion to the variance explained by a principal component analysis and following the ER design.

PRE-PROCESSING: Quality of DEMs and DEM-derived products directly affects the quality of terrain analysis applications. Three approaches to the reduction of errors in

DEM and DEM-derived products have been described: (a) by using empirical knowledge, e.g. to adjust elevations using medial axes or stream networks; (b) by applying filtering operations and (c) by error propagation. Filtering operations are used to replace erratic values or reduce outliers using the spatial dependence structure and probability of exceeding a value estimated from the neighbours. In the case of error propagation, the errors are reduced by calculating the average value of multiple realisations. The methods were tested using a 3.8×3.8 km sample area covering two distinct landscapes: hilland and plain with terraces. The contour data was interpolated using the linear interpolation. The proportion of artefacts (padi terraces) in the unfiltered DEM was 17.3%. After the addition of medial axes, filtering of outliers and adjustment of elevation for streams, the proportion of padi terraces was reduced to 2.2%. Remaining errors in terrain parameters such as undefined pixels and local outliers were reduced using filtering with iterations and by error propagation. The proportion of outliers in all terrain parameters did not exceed 2% of the total area. Both the filtering approach and error propagation give somewhat smoother maps of terrain parameters. The advantage of filtering of outliers is that it employs the structure of the spatial dependence. The advantage of error propagation is that it can be easier automated. The reduction of errors improved the mapping of landform facets (classification) and solum thickness (regression). The classification accuracy increased from 51.3% to 72% and the R^2 of the regression model for the prediction of the solum thickness increased from 0.27 to 0.40.

PHOTO-INTERPRETATION: A method to enhance manual landform delineation using photo-interpretation to map a larger area is described. Conventional aerial photo-interpretation (API) maps using a geo-pedological legend of 21 classes were prepared for six sample areas totaling 111 km^2 in Baranja region, eastern Croatia. Nine terrain parameters extracted from a digital elevation model (ground water depth, slope, plan curvature, profile curvature, viewshed, accumulation flow, wetness index, sediment transport index and the distance to nearest watercourse) were used to extrapolate photo-interpretation over the entire survey area (1062 km^2). The classification accuracy was assessed using the error matrix, calculated by comparing both the whole API maps and point samples, with the results of classification. The first results, using a maximum-likelihood classifier, were 58.2% (hill land), 39.1% (plain), and 45.3% (entire area) reproducibility of the training set. Six classes in the plain were responsible for a large proportion of the misclassifications, due to an insufficiently detailed digital elevation model and the complex nature of landforms (point bar complexes, levees, active channel banks), which can not be explained with the terrain parameters only. Reproducibility for a simplified legend of 15 classes over the study area was improved to 65.8% (plain), 58.2% (hill land) and 63.4% (entire area) using the whole-API training set. After the simplification of legend (15) and with the iterative (3) selection of point-sample training set, classification was able to reproduce 97.6% (hill land), 86.7% (plain), and 90.2% (entire area) of the training set. The supervised classification showed fine details not achieved by photo-interpretation. The number of manual photo-interpretations that had to be prepared was reduced from 84 to 6.

INTERPOLATION: A methodological framework for spatial prediction based on regression-kriging is described and compared versus ordinary kriging and plain regression. The data are first transformed using logit transformation for target variables and factor analysis for continuous predictors (auxiliary maps). The target variables are then fitted using step-wise regression and residuals interpolated using kriging. A generic visualisation method is used to simultaneously display predictions and associated uncertainty. The framework was tested using 135 profile observations from the national survey in Croatia, divided into interpolation (100) and validation sets (35). Three target variables: organic matter, pH in topsoil and topsoil thickness were predicted from six relief parameters and nine soil mapping units. Prediction efficiency was evaluated using the mean error and root mean square error (RMSE) of prediction at validation points. The results show that the proposed framework improves efficiency of predictions. Moreover, it ensured normality of residuals and enforced prediction values to be within the physical range of a variable. For organic matter, it achieved lower relative RMSE than ordinary kriging (53.3% versus 66.5%). For topsoil thickness, it achieved a lower relative RMSE (66.5% versus 83.3%) and a lower bias than ordinary kriging (0.15 versus 0.69 cm). The prediction of pH in topsoil was difficult with all three methods. This framework opens a possibility to develop a bundle algorithm that can be implemented in a GIS to interpolate soil profile data from existing datasets.

VISUALISATION: A method to visualise multiple membership maps, called *"Colour mixture"* (CM) is described and compared to alternative techniques: defuzzification and Pixel mixture. Six landform parameters were used to derive the landform classes using supervised fuzzy k-means classification. The continuous categorical map is derived by GIS calculations with colours, where colour values are considered to represent the taxonomic space spanned by the attribute variables. Coordinates of the 9 class centres (landform facets) were first transformed from multivariate to two-dimensional attribute space, and then projected on the Hue Saturation Intensity (HSI) colour-wheel. The taxonomic value was coded with the Hue and confusion with Saturation. To improve visual impression, saturation was replaced with whiteness. Classes that were closer in attribute space were merged into similar generic colours. The CM technique limits the derived mixed-colour map to seven generic hues independently of the total number of classes, which provides basis for automated generalisation. Saturation derived from the mixed-colour map was used to derive primary boundaries and to locate areas of higher taxonomic confusion.

ORGANIZATION: The key concepts, operations and organizational structure of a grid-based Soil Information System (SIS) are compared to a conventional polygon-based SIS and illustrated with a case study of a 3.8×3.8 km area in eastern Croatia. The key spatial entity in this system is a grid cell and all GIS layers were brought to the same grid resolution (25 m in this case). The soil variables were modelled using the mixed model of spatial variation, so that both discrete and continuous transitions were possible. The SIS, in this case study, included 21 predictor maps (photo-interpretation map, terrain parameters and remote sensing images), six maps of soil variables (solum thickness, occurence of the mollic, calcic and gleyic horizon, toposil thickness and topsoil silt content) and six derived maps of soil types. Each soil variable was interpolated using a hybrid interpolation technique (regression-kriging). The interpolated maps were then classified using a continuous classifier (fuzzy k-means) to produce membership maps. These were then used to derive land suitability for wheat production on a continuous scale (0–1), as an example of interpretation that can be derived from the SIS. The photo-interpretation map was shown to be a somewhat better predictor of the listed soil variables than the terrain and remote sensing maps. Comparison of goodness of fit and thematic confusion showed that the grid-based SIS gives in general better fit to the original data, higher level of detail and more reliable predictions than the conventional (polygon-based) SIS. The advantages of the proposed SIS, compared to a conventional survey, are: (1) it offers a map of soil types rather than of the soil-mapping units; (2) all variables are mapped as continuous spatial fields at fine grain of detail; (3) it offers a measure of uncertainty for both input and derived maps; (4) both discrete and continuous transitions are possible and (5) the original soil observations and interpolation/classification parameters are stored in tables as a part of the SIS, so that derived maps can be updated. The disadvantages are: (1) it is computationally demanding and requires a large amount of storage; (2) it is more costly (collection and pre-processing of auxiliary variables) and (3) SIS is sensitive on the quality of the input data.

QUALITY CONTROL: Methodology to assess the quality and adequacy of a national soil resource inventory and to evaluate its usability is described. Six 1:50 K map sheets (of 185 total), three control surveys (each of size 4×4 km) and ten full profile descriptions in the main landscape regions of Croatia were used to estimate the effective map scale, accuracy of map legends and thematic accuracy of profile observations. In addition, the existing digital data sets (soil map of Croatia at scale 1:300 K and database with 2198 profiles) were evaluated for thematic purity and contrast for clay content, pH and organic matter. New methods were developed and tested to assess the spatial accuracy of soil boundaries and the thematic overlap among map units. In the case study, the average polygon size and the positional accuracy of primary soil boundaries (about ±40 m) correspond to the 1:150 K scale, while the inspection density corresponds to the 1:250 K scale. Mapping units are heterogeneous with an average relative variation of 17% within units and a mean thematic overlap of 66% among geographically-adjacent units. There is a large difference between the original legend and the validation sample when considered as taxonomic classes, but much less so when classes are grouped by similarity. The inventory is adequate for small-scale applications but not in general at detailed scales. The major usability problems are compound map units, lack of specific interpretations corresponding to user needs, and lack of legal clarity on ownership and use.

The general conclusion is that the proposed pedometric mapping methodology enhances the practice of soil mapping making the soil maps more objective, detailed and more compatible for integration with other environmental geo-data. There is no need to use the concept of soil mapping units or use double-crisp soil maps anymore. On the other hand, instead of abandoning photo-interpretation, soil classification or empirical knowledge on soils, these methods can be successfully integrated with pedometric techniques.