5TH GLOBAL WORKSHOP ON PROXIMAL SOIL SENSING

PSS 2019

Program and Proceedings

May 28-31, 2019 / Columbia, Missouri, USA

Editors: K.A. Sudduth, N.R. Kitchen, and K.S. Veum

USDA Agricultural Research Service

International Union of Soil Sciences Working Group Proximal Soil Sensing

University of Missouri

SPONSORS

Veris technologies

soil systems an Open Access Journal by MDPI

SOILOPTIX

GEONICS

Structured Procedure for Selection of Suitable Soil Data Acquisition Techniques

M. Knotters[1], F.M. van Egmond[1]

[1]Wageningen Environmental Research, Wageningen, the Netherlands, fenny.vanegmond@wur.nl

Abstract

Soil mapping is carried out for various purposes and at various scales but is always aimed at answering a defined or undefined (research) question. To answer this question information or data is needed which is then supplied by soil mapping using existing and/or new data. At present the choice which data to collect is often based on expert knowledge of the researchers or stakeholders involved. For the Dutch Key Registration of the Subsurface (BRO) the need was felt to develop a structured procedure to identify the gap between information need and availability and then to select the most suitable data acquisition technique(s) to close that gap. This should address the required accuracy and spatial resolution of the information need and the costs available for data acquisition.

The information need of users of soil information products can be described in terms of the target universe, domain of interest, target variable, target parameter, type of result and a statistical measure for the accuracy such as standard error or confidence interval or the percentage correctly classified in class maps. The question dictates the threshold that is applied to these statistical measures. When this is mapped for the information question as well as the available information, maps and datasets the information gap can be identified. The gap can be addressed by data mining, up- or downscaling maps or acquiring new information. Methods for acquiring new information can be landscape analysis based soil mapping, geostatistical soil sampling or profile description in combination with kriging or machine learning, use of proximal soil sensors, using satellite information or a combination of these. The most effective remedy for supplementing and improving information is the one that answers the question against minimal costs, where the costs are not higher than the gain in terms of improved *value of information*. This is demonstrated for two case studies, fieldscale and a small region, where results show that in more complex cases this can be difficult. Several studies show however that the benefits or gain of soil maps generously outperform the costs associated with making them (costs:benefits 1:46 to 1:123 for 20 to 25 years map lifetime).

Introduction

The Soil Map of the Netherlands, scale 1: 50,000 is among other things included in the Dutch Key Registration of the Subsurface (BRO). This map is consulted to inform all kinds of choices and decisions in agriculture, environment, nature and infrastructure in the Netherlands and therefore needs to be up to date and suited for its purposes. There is a clear need for a structured procedure to identify discrepancies between information needs and information availability since choices become increasingly data-driven. Then a structured procedure is needed as well to select methods for additional data collection with which these discrepancies can be eliminated. The purpose of this study is therefore to answer the following research questions:

1. How do you establish whether area-wide soil information in the BRO meets the requirements that the user(s) sets with regard to accuracy and spatial detail?
2. How do you select the most efficient additional mapping method to allow area-wide soil information in the BRO to meet the requirements the user(s) sets with regard to accuracy and spatial detail?

This research focuses on the quality aspect of accuracy, which is defined as the degree of correspondence between information and reality. It does not answer the question whether the BRO or the user is responsible for additional inventories or to what extend it would be desirable to include such information in national inventories.

Soil information

The different needs and requirements of users of area-wide soil information can be described following the method proposed by De Gruijter e.a. (2006):

1. Detailed description of the information requirement: *target universe* (outer boundaries of the target area and period), *domain (s) of interest* (more precise sub-areas and periods for which information is required), *target variable (s)* (variable (s) for which information is desired, these can be qualitative or quantitative), *target parameter (s)* (the type of statistic that is desired given the type of target variable and domain of interest), *target quantity* (the combination of a domain, target variable and target parameter), *type of result* (qualitative or quantitative, for instance the compliance or degree of compliance).
2. Accuracy measure: for quantitative information this can be for example standard errors or confidence intervals, for qualitative information this is for example a percentage correctly classified (map purity according to user's or producer's accuracy), error rates or chance at a type II error.
3. Accuracy requirement: for quantitative information a threshold for example a maximum for the standard error, for qualitative information for example a threshold like the minimum percentage that is correctly classified.

Mapping the information requirement using the procedure offered here yields a clear description and set of qualitative or quantitative measures for the information needed to answer the (research, policy or planning) question at hand. When available information such as (parts of) the Soil Map of the Netherlands, more detailed maps, new surveys, ancillary and descriptive information is classified according to this description and measures, the type or specifics and the size of the information gap for a given question becomes apparent.

There are a number of remedies that are conceivable to solve discrepancies that may arise between information requirements and information availability (the information gap): an additional inventory to cover the *target universe* (this is easier in space than in time), (geo-) statistical downscaling of information in the BRO to specific *domains of interest,* an additional inventory of the *target variable* within the domain of interest, statistical processing of the information in the BRO and/or of information from additional inventories to derive the desired *target parameter*, or combinations of these remedies. If qualitative information is required while the BRO contains quantitative information, classification, testing or detection must take place. If quantitative information is required while the BRO only contains qualitative information, a form of downscaling is needed, possibly with the aid of an additional inventory. If the information in the BRO does not meet the accuracy requirement, an additional inventory must take place until it does. If the accuracy of the information in the BRO is not known, it must be determined with a validation study. In addition to carrying out field observations, 'inventory' also means the exploration of existing datasets at local, regional, national or international agencies and organisations.

The most efficient remedy is the one where information is supplemented and improved at the lowest possible cost, and where the costs of supplementing and improving information are not higher than the revenues in terms of increased *value of information*: the gain of supplementation and improvement of information (Morgan et al., 1990).

This means that if a soil map is used for a longer period of time for multiple purposes, as is usually the case with national inventories, then an accurate cost-benefit calculation is complicated. However, various studies have shown that the costs of these maps are far outweighed by the benefits, even if it is not possible to determine all benefits. Klingebiel (1966) argued that the costs of a soil map that is used for several purposes are already payed off in the first year. He estimated the cost-benefit ratio for a 'lifespan' of the soil map up to 25 years is 1:46 for extensively used land such as forests, 1:61 for moderately intensively used areas with mixed agriculture that consist of approximately half of arable land and 1: 123 for intensively used areas that have been urbanized for about a quarter and where built-up areas are growing rapidly. From research results by Giasson et al. (2000) it can be calculated that for a farming region the cost-benefit ratio of a 1: 50,000 soil map is 1:122 when the calculations are made over a 20-year period and an interest rate of 3% is applied. This corresponds to the cost-benefit ratio that Klingebiel (1966) calculated for an intensively used, urbanizing region. This would suggest that the cost:benefit ratio of general soil maps is expected to be more favourable than can be calculated from the main uses that are expected up front. Calculating the value of information for a few main uses is relevant in such cases is still relevant but indicative.

The value of area-wide soil science information is easier to express in monetary values for one specific application. Examples are calculations of crop damage due to groundwater extraction (Knotters and Vroon, 2015) and forecast of crop yields (Bie and Ulph, 1972; Dent and Young, 1981; Giasson et al., 2000). In precision agriculture, research is conducted into the added value of detailed soil science information, for example by Cook and Bramley (2000). As far as is known to the authors, the study by Knotters and Vroon (2015) is the only analysis of the value of information of a soil map that is based on a validation. In this small regional study, damage payments to farmers in an area with groundwater extraction were calculated on the basis of the Soil Map of the Netherlands, scale 1: 50,000 and on the basis of a detailed supplementary mapping, scale 1: 25,000. The validation showed that the costs of detailed mapping compared to the benefits in terms of reduced error in compensation to farmers are 1: 8, assuming an interest of 3% and a payment period of 30 years.

Methods for additional soil data inventories

There are several methods to supplement the information to close the information gap defined by the above methodology:

- Free or *landscape based soil mapping*, where a soil geographer maps soil patterns in the field, based on information from augerings, landscape features that are visible in the field and all kinds of additional information such as covariates. The choices and decisions that a soil geographer makes to draw unit boundaries cannot be reproduced fully by another and depend on the knowledge and insight of the geographer (Ten Cate e.a., 1995a,b,c).
- *Geostatistic surveys* use interpolation algorithms such as kriging with field observations serving as input (Goovaerts, 1997). When the interpolation algorithms have been described, the result of a geostatistical soil mapping can be reproduced. The choices that are made when modelling the spatial correlation or structure (variogram) will vary from expert to expert, which prevents geostatistical soil mapping to be entirely objective (Englund, 1990).
- In *digital soil mapping*, geostatistical techniques and machine learning techniques are used to make soil maps based on soil observations in combination with exhaustive sources of area-wide ancillary information, so-called covariates (McBratney e.a., 2003; Kempen, 2011).
- *Upgrading* an existing soil map means that the current content of the map units on that map is described based on the results of new probability samples performed within those map units (Brus e.a., 1992). The map boundaries remain the same, the (type of) information they contain is updated or supplemented.
- Various sensor techniques have been developed to map soil variables spatially by performing measurements (mostly) above ground (*proximal soil sensing*): the measurement of electrical conductivity using electrical resistance or electromagnetic induction (EC, EMI), VIS / NIR spectroscopy, gamma-ray spectrometry, ground penetrating radar, magnetometry, soil moisture and pH sensors. The choice for a particular technique depends, among other things, on the target variable and terrain characteristics of the area of interest (trafficability, soil profile, available platforms such as walking, driving, drones) (Knotters e.a., 2017, van Evert e.a., 2018).
- *Satellite images* with optical, thermal and radar information can be used as an auxiliary variable in digital soil surveys or the spectral information can be used directly for soil mapping. Multi-spectral satellites have a better availability and have shorter return times. Hyperspectral images provide better accuracies. For both, composite images are made in order to map as much soil as possible and to distinguish soil reflections from reflections of vegetation and clouds.

Example for case "Peat actualisation Utrecht"

For the Province of Utrecht in the Netherlands there is a need for a soil map, scale 1: 50,000, which shows the distribution of peat and organic rich soils in 2018 with a percentage correctly classified of at least 70% per mapped soil property. This requirement is based on a list of tangible and non-tangible judicial, political/governmental, societal, economic risks and risks for the knowledge-system. These are difficult to quantify making a precise calculation of the value of information impossible or incomplete. This does however not prevent a calculation of accuracy and cost options for reaching the information requirement. The current Soil Map of the Netherlands was made over 1959-1995 and therefore does not meet the domain of interest requirement of 2018. A Peat Soil Map of the Province of Utrecht, scale 1: 25,000 (BvU) from 2008 is available. When this is 'plugged in' in the Soil Map of the Netherlands, scale 1: 50,000, then a map purity of 70% for soil type at main group, group and subgroup level are still not met as is evaluated in a validation study by Kempen e.a. (2011). The BvU can therefore only be used as auxiliary information. An additional inventory is therefore necessary to meet the desired purity and timeliness for the soil type characteristic at the main group, group and subgroup level. A mapping pilot is proposed to determine whether electromagnetic (EMI) or ground penetrating radar measurements can contribute to the efficiency of the supplementary inventory and at which observation densities of augerings and measurements a map purity of 70% can be achieved for each soil property. When achievable accuracies and costs of eth GPR and EMI surveys are known (expected in Spring 2019) an evaluation can be made of the expected value of information for EMI, GPR and augerings. This will determine the choice to fill the information gap for this question.

Example for case "field in Peel region"

For a field in the Peel region in the South of the Netherlands, research was carried out into the effectiveness of a measure that is taken (drainage) to nullify the wet damage that could have been caused by a change in the hydrological structure of an adjacent raised bog area. First augerings are carried out to fill the information gap. When this was determined insufficient, measurements with ground penetrating radar (GPR) were carried out to increase the accuracy in determining the effect of the measure and the calculation of the damage. The question is whether the costs of the additional GPR research have been worthwhile compared to the revenues in terms of a more accurate calculation of the damage and a more accurate estimate of the effects of compensatory measures. To answer this question, the following information is required: soil maps and maps with average highest and lowest groundwater levels (or groundwater levels) before and after the intervention, the inventories made with the conventional method (augerings) and with GPR, the costs of the augering inventory, the costs of the inventory with GPR, a risk assessment and considerations for further investigation, the table that provides the translation of soil and groundwater level information to yield depressions; the monetary value of a percentage of revenue depression; the period for which the damage is calculated (for example 30 years) and the internal interest rate that has been used. This information is currently gathered by the stakeholders involved in the case and will be assessed when available.

Conclusions and Recommendations

The following conclusions can be drawn from this study:

1. The proposed method shows in detail and completely where offered information (such as by the BRO) does not match the information requirement and proposes possibilities to close the information gap.
2. The case of the peat update for the province of Utrecht shows that there is no explicit accuracy requirement, it is based on a risk analysis. The value of information in terms of increased accuracy and therefore reduction of risks can therefore not be calculated. At the same time an evaluation of accuracies and costs of possible options to reach the general information requirement can still be performed.
3. The damage calculation case for a field in the Peel region shows that for applications with a single application it is possible to impose an accuracy requirement and to weigh the costs of data collection against the revenues in terms of reduced financial risks.

The following recommendations follow from this study:

1. If there is no explicit accuracy requirement, such as for soil maps that are used for various purposes, it is recommended to aim for a quality criterion such as the historically grown 70% map purity, provided that this criterion is precisely defined. The method of data collection must then be selected in such a way that this quality criterion can be achieved at the lowest possible cost.
2. It is recommended to accurately define the historically developed quality criterion of 70% map purity for soil maps. Currently, the following, different, interpretations of this criterion are possible:

a. 70% strict map purity: on 70% of the map all soil properties are correctly classified. This is seldom achieved in practice, see Marsman and De Gruijter (1986);
b. Each soil property separately is correctly classified on 70% of the map;
c. On average, all soil properties are classified correctly on 70% of the map. This is in line with findings from Marsman and De Gruijter (1986), but may mean that some features are much less accurate than others.

We recommend to adopt b.

3. It is recommended to substantiate the quality criterion of 70% map purity with risk analyses and adjust if necessary.

Acknowledgements

This study was funded by the BRO program in the Netherlands assigned by the Ministry of Agriculture Nature and Environment and the Ministry of Internal Affairs.

References

Bie, S.W. & Ulph, A. (1972). The economic value of soil survey information. *Journal of Agricultural Economics*, 23, 285-297.

Cate, J.A.M. ten, van Holst, A., Kleijer, H., & Stolp, J. (1995a). Handleiding bodemgeografisch onderzoek; deel A, Bodem. Wageningen, DLO-Staring Centrum, Technisch Document 19A.

Cate, J.A.M. ten, van Holst, A., Kleijer, H., & Stolp, J. (1995b) Handleiding bodemgeografisch onderzoek; deel B, Grondwater. Wageningen, DLO-Staring Centrum, Technisch Document 19B.

Cate, J.A.M. ten, van Holst, A., Kleijer, H., & Stolp, J. (1995c) Handleiding bodemgeografisch onderzoek; deel C, Kaart tekenen, rapporteren en samenstellen digitale bestanden. Wageningen, DLO-Staring Centrum, Technisch Document 19C.

Cook, S.E. & Bramley, R.G.V. (2000). Coping with variability in agricultural production –implications for soil testing and fertiliser management. *Communications in Soil Science and Plant Analysis,* 31, 11-14, 1531-1551, doi: 10.1080/00103620009370524.

Dent, D. & Young, A. (1981). Soil survey and land evaluation. George Allen & Unwin, London.

Englund, E.J., (1990). A variance of geostatisticians. *Mathematical Geology,* 22(4), 417-454.

Evert, F.K. van, Wolters, S., van Boheemen, K., van Dijk, C. (2018). Final report on research results, research projects and industry solutions. SmartAKIS, Report D1.6.

Giasson, E., van Es, C., van Wambeke, A., Bryant, R.B. (2000). Assessing the economic value of soil information using decision analysis techniques. *Soil Science,* 165, 971-978.

Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. Applied Geostatistics Series. New York: Oxford University Press.

Gruijter, J.J. de, Brus, D.J., Bierkens, M.F.P., Knotters, M. (2006). *Sampling for Natural Resource Monitoring*. Berlin: Springer, 332 pp.

Kempen, B., (2011) Updating soil information with digital soil mapping. Wageningen, Proefschrift Wageningen Universiteit, 483 pp.

Klingebiel, A., (1966) Costs and returns of soil surveys. *Soil Conservation,* 32, 3-6.

Knotters, M., & Vroon, H.R.J. (2015). The economic value of detailed soil survey in a drinking water collection area in the Netherlands. *Geoderma Regional,* 5, 44-53.

Knotters, M., van Egmond, F., Bakker, G., Walvoort, D.J.J., & Brouwer, F. (2017). A selection of sensing techniques for mapping soil hydraulic properties. Wageningen, Wageningen Environmental Research, Report 2853.

Marsman, B.A., & de Gruijter, J.J. (1986). Quality of soil maps. Wageningen, Soil Survey Papers 15.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B. (2003). On digital soil mapping. *Geoderma,* 117, 3-52.

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