ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/225844493

Homosoil, a Methodology for Quantitative Extrapolation of Soil Information Across the Globe

Chapter · January 2010

DOI: 10.1007/978-90-481-8863-5_12

CITATIONS
41

READS
305

3 authors, including:

Budiman Minasny
The University of Sydney

523 PUBLICATIONS **18,237** CITATIONS

SEE PROFILE

Alex B Mcbratney
The University of Sydney

690 PUBLICATIONS **34,273** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Soil carbon View project

Spatial sampling View project

All content following this page was uploaded by Budiman Minasny on 15 May 2014.

The user has requested enhancement of the downloaded file.

Chapter 12
Homosoil, a Methodology for Quantitative Extrapolation of Soil Information Across the Globe

B.P. Mallavan, B. Minasny, and A.B. McBratney

Abstract In many places in the world, soil information is difficult to obtain and can be non-existent. When no detailed map or soil observation is available in a region of interest, we have to extrapolate from other parts of the world. This chapter will discuss the Homosoil method, which assumes homology of soil-forming factors between a reference area and the region of interest. This includes: climate, physiography, and parent materials. The approach will involve seeking the smallest taxonomic distance of the *scorpan* factors between the region of interest and other reference areas (with soil data) in the world. Using the digital information of soil climate from the Climate Research Unit (CRU) (solar radiation, rainfall, temperature, and evapo-transpiration), topography from the HYDRO1k (elevation, slope, and compound topographic index), and lithology of the world on a $0.5° \times 0.5°$ grid, we calculated Gower's similarity index between an area of interest and the rest of the world. The rules calibrated in the reference area can be applied in the region of interest realising its limitations and extrapolation consequences.

Keywords Global soil mapping · Homoclime · Climate · Map extrapolation · Soil forming factors

12.1 Introduction

In many places around the world, legacy soil information is difficult to obtain and is practically nonexistent. With limited time and practically no resources to collect new soil samples in these areas, we have to develop a new methodology for estimating soil attributes based on our knowledge of soil forming factors and pedogenesis.

In Australia, at the system level (ASRIS, mapping scale 1:100,000) only about 60% of the intensive land-use zones, and less than 5% of rangelands, are covered by soil information (McKenzie et al., 2005). Countries in South America, South

A.B. McBratney (✉)
Faculty of Agriculture, Food & Natural Resources, The University of Sydney, Sydney, NSW 2006, Australia
e-mail: alex.mcbratney@sydney.edu.au

J.L. Boettinger et al. (eds.), *Digital Soil Mapping*, Progress in Soil Science 2,
DOI 10.1007/978-90-481-8863-5_12, © Springer Science+Business Media B.V. 2010

East Asia, and Africa only have 10–30% of landscape coverage for maps finer than 1:100,000 (Hartemink, 2008). When no detailed maps or soil observations are available in a region of interest, we have to extrapolate from other areas or other parts of the world. If we are only dealing with global modelling at a coarse resolution, we can extrapolate soil observations that are available from other similar areas (that are geographically close) or by using spatial interpolation or spatial soil prediction functions (Schenk and Jackson, 2005). However if we wish to map the soil at a much finer resolution, such as the target resolution of GlobalSoilMap.net (90 m $\times$ 90 m), geographical extrapolation may not work (See Chapter 33). In this case, a new approach is required that allows extrapolation of spatial soil prediction functions, realising that similar areas could be in other parts of the region or other parts of the world.

Given severe prior soil information crisis, we introduce a new method for digital soil mapping called *Homosoil*, which aims to find areas in the world with similar soil-forming factors for the purpose of extrapolation of soil mapping rules.

The idea of extrapolating environmental variables has been explored previously, mainly to identify areas in the world with similar climates for crop production. Prescott (1938) coined the term "homoclime" referring to areas or regions in the world with similar climate. In Australia, studies of homoclimes have been carried out with reference to particular economic crops (Prescott, 1938, 1943). However these early studies simply compared some climatic variables that were considered to be critical and no quantitative similarity index was developed. Russell and Moore (1970) define homoclimates (similarity between weather stations) based on various similarity coefficients. They also noted that the detection of areas of similar climate could be of interest in pedology.

Jones et al. (2005) revisited the homoclime approach and defined the *homologue* approach to determine which crops could be grown at a specific site in the world. The principle is that crops will perform in a similar fashion in similar environments wherever they happen in space and time. The algorithm calculates the similarity between monthly climate variables and soil data at the place of interest and elsewhere in the tropical world.

Defining areas with similar soil-forming factors is not as straightforward as homoclime estimation. While climate defines the weathering regime of the soil, the parent materials and the age of the soil play a more important role in defining the type of soil that is formed. Topography and landuse also influence the distribution of soil. Nevertheless, results from spatial data mining from the Australian-wide mapping project ASRIS (Bui et al., 2006) showed that the state factors of soil formation form a hierarchy of interacting variables, with climate being the most important factor at a continental scale and different climatic variables dominating in different regions. It was shown that lithology is almost equally important to climate in defining broad-scale spatial patterns of soil properties, whereas shorter-range variability in soil properties appears to be driven more by terrain variables.

This paper will investigate methods to find areas in the world with similar soil-forming factors as expressed by environmental covariates. We will demonstrate an algorithm for calculating the taxonomic distance of the environmental variables between the region of interest and other reference areas with good soil data coverage.

12.2 Conceptual Framework

The homosoil approach involves finding areas in the world with similar soil-forming factors. One can start by identifying areas where reliable and adequate soil maps or soil data are available, then use them to search for the homologous areas in environmental space. The methodology of this approach is represented in Fig. 12.1 and summarised as follows:

(1) Mapping areas in the world for soil donor and recipient sites. The first step is to delineate areas in the world where sufficient soil information is available. There are many areas in the world where soil information is not available at a resolution finer than 1:100,000. The level of soil information coverage in areas of the developing world is only available as estimates of percent coverage in the country based on questionnaires (Hartemink, 2008; Zinck, 1990). With the development of GlobalSoilMap.net consortium, we can now access global soil information and identify areas in the world where adequate soil information is available. These areas are

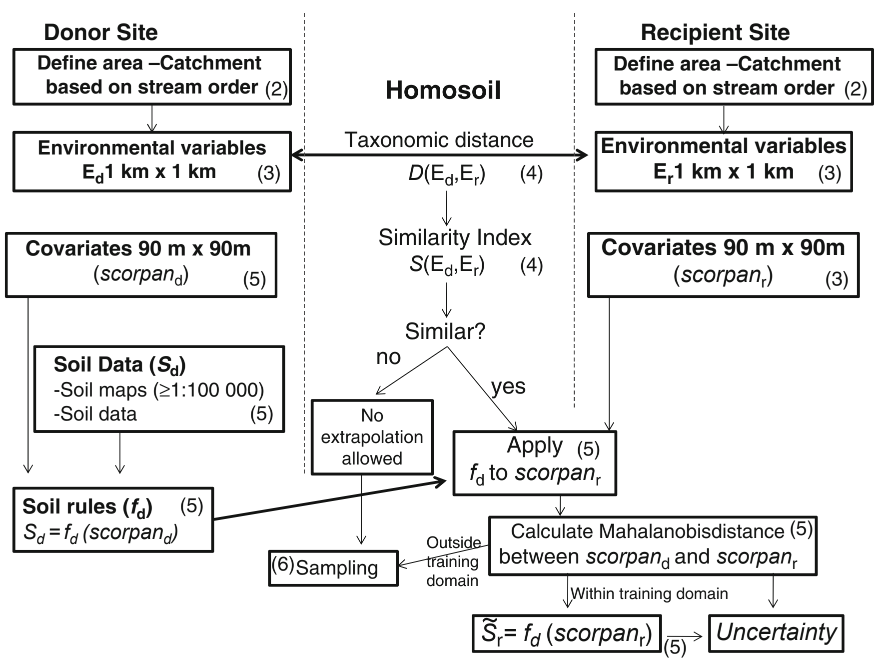


Fig. 12.1 A framework for developing Homosoil. *Numbers* in *brackets* represent the steps in the methods as outlined in the proposal

called “donor sites”, where potentially, the soil mapping rules developed can be exported to other areas. Areas with inadequate soil information are called “recipient sites”, where soil rules from the donor sites may be applied.

(2) Definition of donor and recipient sites. We need to define the appropriate scale for the donor and recipient sites, which can be represented as catchments. While calculation of the similarity index can done by a pixel-by-pixel comparison; for soil landscape analysis, it is more realistic to define the areas or sites as a catchment based on stream order. Mourier et al. (2008) found that the Strahler stream order is a crucial hierarchical descriptor of the distribution pattern of soil. Our initial investigation suggests that the 4th and 5th order catchment maybe appropriate for comparison of the soil-forming factors. This will need further investigation.

(3) Building a spatial database for environmental variables. A spatial database that serves as a proxy to represent soil-forming factors and will allow us to define areas of homosoil. This spatial database will contain information on environmental variables over large areas (e.g. a continent) with a resolution of 1 km $\times$ 1 km. The important environmental variables governing soil formation are: climate, topography, lithology, and age. At the global scale, only coarse climate and lithology maps are available at a resolution of 0.5° $\times$ 0.5°. The global climate data are available from the ERA-40 reanalysis and the Climate Research Unit (CRU) dataset (New et al., 1999) and the lithology data are available from a global digital map (Durr et al., 2005). However, for global soil map prediction, a finer resolution is required to capture the local environmental heterogeneity. Therefore we shall build a database of environmental variables (climate, lithology, topography, age) at a resolution of 1 km $\times$ 1 km.

We can derive global topographic information from elevation data provided by the SRTM. Elevation, slope, and compound topographic index (CTI) are important topographic variables. However regional-scale indices representing physiography, erosional and depositional areas may play more important roles in this broad-scale landscape analysis. Lithology must reflect the soil parent materials, and the age of the land surface is also an important factor in defining the type of soils. Age of the land can be defined in a number of ways, e.g. the time since the last glaciation (Adams, 1997).

(4) Calculating the taxonomic distances of environmental variables. The homology of soil-forming factors requires aggregation of the various components: climatic factors (homoclime), lithology (homolith), topography (homotop), and age (homochron).

Similarity Indices

The monthly climate data need to be converted to climatic indices (Booth et al., 1987) so that the taxonomic distance of the climate variables can be calculated. This also attempts to standardise climatic events across various parts of the world, such as the differences arising between the Northern and Southern hemispheres. For the climatic variables (e.g. rainfall, temperature, solar radiation and evapotranspiration) we can calculate various indicators, such as: annual mean, mean of the driest period, mean of the wettest period, annual range, driest quarter mean, wettest quarter mean, coldest quarter mean, hottest quarter mean, seasonality, etc. Thus we obtain an abundance of climatic information. Using these climatic variables, we can calculate the climatic similarity (homoclime) between two areas either by Mahalanobis distance or by a Gower similarity measure (Booth et al., 1987; Gower, 1971). The Gower distance measures the similarity S_{ij} between sites i and j, each with p number of variables, standardised by the range of the variable k:

$$S_{ij} = \frac{1}{p} \sum_{k=1}^{p} \left(1 - \frac{\left| x_{ik} - x_{jk} \right|}{\text{range } k} \right)$$

where p denotes the number of climatic variables; $\left| x_i - x_j \right|$ represents the absolute difference of climate variables between site i and j. The similarity index has a value between 0 and 1. Initial investigations suggest that the Gower's similarity works better than the Mahalanobis distance, but further investigation is needed. All studies of homoclimes are based on comparison between two pixels on a grid, whereas for homosoil, the comparison will be based on two catchments.

Similarity Index for Soil-Forming Factors
Considering the scale and the resolution of this study and the available global and regional data, the climatic factor is probably the most important and reliable soil-forming factor (Bui et al., 2006). However other factors are still equally important, therefore we need to define our *Homosoil* based on *homoclimes*, areas with similar lithology (*homolith*), similar age (*homochron*), and similar topography (*homotop*). One approach could be to apply a hierarchical approach, i.e. by first calculating a similarity index based on the most important factor (e.g. climate or age) and then refining the selection of homologous areas based on the second most important factor (e.g. lithology), and so on. However this approach requires an assumption of the hierarchical nature of the soil-forming factors, which may not work in every area. Another approach includes assigning different weight to each of the factors. We will make a thorough comparison of the various approaches

(5) How to import soil mapping rules. The principle of the method for building soil mapping rules is based on the *scorpan* spatial soil prediction functions (McBratney et al., 2003). The function or mapping rules f is usually calibrated in an area where soil information is available. This function is then applied to the area of interest for interpolation (and limited extrapolation) of soil properties. In this project, we will extend the application of f into geographic areas other than the one used for calibration, in this sense f is used for extrapolation in other areas in the world. The requirement for this extrapolation is that both areas (the donor and recipient) should have the same *scorpan* factors, which cover both areas at a resolution of $90\,\text{m} \times 90\,\text{m}$. The *scorpan* factors define soil distribution at a local catchment scale. This is in contrast with the broader district scale environmental factors ($1\,\text{km} \times 1\,\text{km}$) for defining soil-forming factors. The commonly used function

for f is the classification or regression-tree approach (Bui and Moran, 2001). In the Bui and Moran (2001) approach, the decision-trees have two levels: one is used to capture mapping rules from the training area, and another one is used to define the domain over which those rules can be extrapolated. We can also evaluate the appropriate domain of the mapping rules f using distance metrics of the *scorpan* factors (Tranter et al., 2009). The Mahalanobis distances will be used to determine the distance between the mean of the donor's *scorpan* factors and the recipient's *scorpan* factors. Distances exceeding a designated cutoff limit are deemed distinct from the calibration data and as such unsuitable for function application.

12.3 Examples

We illustrate the application of the homosoil approach using coarse resolution global climate and elevation data. This is the first attempt to demonstrate the homosoil concept.

12.3.1 Global Data

We compiled 3 data sources for the global climate, topography, and lithology data, all on a regular $0.5° \times 0.5°$ grid as matrix of 360×720 pixels. In total, there are 62 254 pixels with all data available. The data were kindly provided by Prof. Marc Bierkens and Dr. Rens van Beek from Utrecht University.

We used the long-term mean monthly temperature, rainfall, solar radiation and evapotranspiration data to represent the climate. We also use the DEM representing topography, and lithology, which gives broad information on the parent material. The climate data come from the ERA-40 reanalysis and CRU dataset. Maps represent the monthly average course over the year. The period corresponded to the datasets are 1961–1990 for the CRU (temperature, rainfall and solar radiation), and September 1957–August 2002 for the ERA-40 analysis (evapotranspiration). Temperature data is in Celsius degree, and we have the monthly value for the minimum temperature, mean temperature and maximum temperature. Rainfall value is in meter per day. Evapotranspiration is in meter per day and it is calculated with Penman formula. More details on the datasets are available on the website http://www.ipcc-data/obs/cru_climatologies.html.

The DEM is from the HYDRO1k dataset, which includes the mean elevation, slope, and compound topographic index (CTI). Details of the dataset are available at: http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/gtopo30/hydro (accessed 21 April 2010).

The lithology is from a global digital map (Durr et al., 2005) with 7 values which represent the different broad groups of parent materials. The lithology classes are: non- or semi-consolidated sediments, mixed consolidated sediments, silic-clastic sediments, acid volcanic rocks, basic volcanic rocks, complex of metamorphic and igneous rocks, and complex lithology.

12.3.2 Climatic Indices

The monthly climate data was first converted to climate indices (Booth et al., 1987) so that the taxonomic distance of the climate variables can be calculated. This also attempts to standardize the climatic events that can vary in various parts of the world such as the difference in the Northern and Southern hemisphere.

For each of the 4 variables climatic (rainfall, temperature, solar radiation and evapotranspiration), we calculated 13 indicators: annual mean, mean for the driest month, mean at the wettest month, annual range, driest quarter mean, wettest quarter mean, coldest quarter mean, hottest quarter mean, lowest ET quarter mean, highest ET quarter mean, darkest quarter mean, lightest quarter mean, and seasonality. Thus we obtain 52 climate variables.

12.3.3 Similarity Index

The climatic similarity between the indices values of two points of the grid is calculated by a Gower similarity measure (Booth et al., 1987; Gower, 1971). Considering the scale and the resolution of this study and the available global data (0.5° × 0.5°), the climatic factor is probably the most important and reliable soil forming factor (Bui et al., 2006). Thus as our first attempt, we define our *Homosoil* based on three steps. First is to identify the *homoclime* around the world. The next step, within the *homoclime*, we find areas with similar lithology (*homolith*) and similar topography (*homotop*).This is summarized as a decision tree in Fig. 12.2. For the lithology

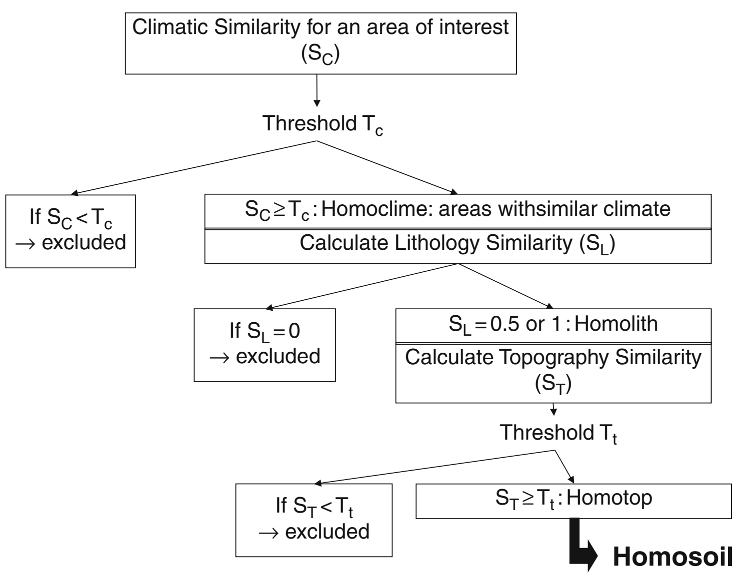


Fig. 12.2 A simple decision tree for homosoil

similarity index, a value of 1 is given when the same lithological class is encountered, 0.5 if the complex lithology class is encountered, and otherwise a value of 0 is given.

12.4 Results

12.4.1 Homoclime

To assess the method we calculate the similarity for a sample of 42 known locations, and see the coherence of the homoclime results. The locations are chosen following the climate classification of the ecozones (Schultz, 2005). Figure 12.3 shows the *homoclime* results for 4 different places in the world with variation in seasonal climate:

– A Mediterranean climate (Montpellier) where the rainfall and the temperature are seasonal
– Sydney where the temperature is seasonal but not the rainfall
– Harare where the climate is tropical with a seasonality of rainfall
– Manaus with an equatorial climate without seasonality.

The results showed that the Gower similarity index gives areas that are in accordance with the ecozones of the world (Schultz, 2005). It seems to be an appropriate method to the homoclime analysis.

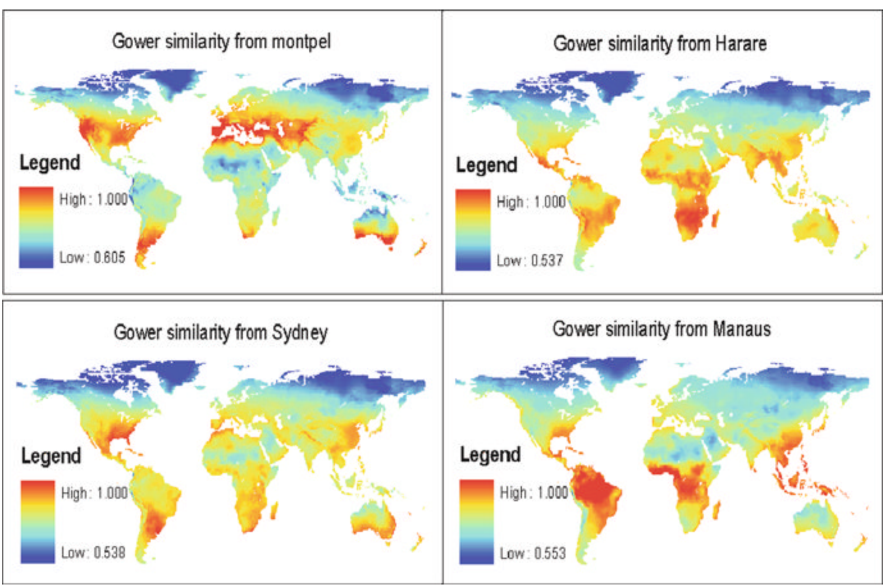


Fig. 12.3 Homoclime degree for 4 locations around the world

12.4.2 Homosoil

Figure 12.4 shows the similarity index for homoclime, homolith, and homotop for Harare. A combination for these factors gives the homosoil similarity index for the area. This is illustrated in Fig. 12.5 for defining homosoil for an area in the South of Somalia (1.5°N 42°E). The first step is to calculate the climate similarity index for Somalia; the second step is to select a threshold for the climate similarity index for homoclime. The selection of similarity based on a threshold of 1-std. dev. of the similarity index. The second step is within the homoclime, to select areas with lithology similarity $\geq$ 0.5. This is homolith. The final step is, within the area of homoclime and homolith, to select areas with similar topography. We select areas with the topography similarity index $>$ 1 – std. dev. The final selection represents area of homosoil.

Homosoil only identifies area with similar soil forming factors, given an area with no soil data, we find other homosoil areas. Within the homosoil areas, we identify are where soil data is available. The soil landscape rules or spatial prediction function from the known area then can be transferred to the area with no soil data. Figure 12.6 shows an illustration of the homosoil concept. Soil information in the Democratic Republic of Congo (2° N, 20.5° E) is not available, so we calculated homosoil areas in the world. It identifies an area in Peru (12.5° S, 69.5° W) as homosoil, and in this area soil information are available. Therefore the soil-landscape rules in Peru can be exported to areas in Congo.

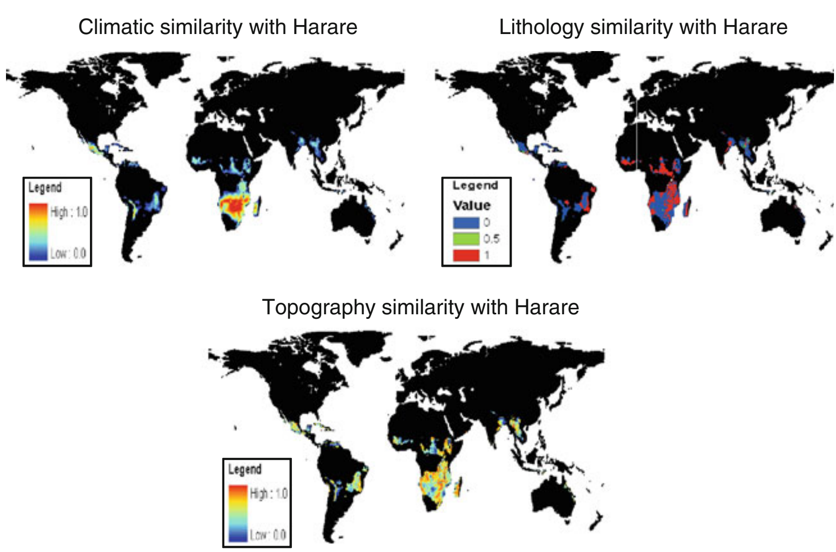
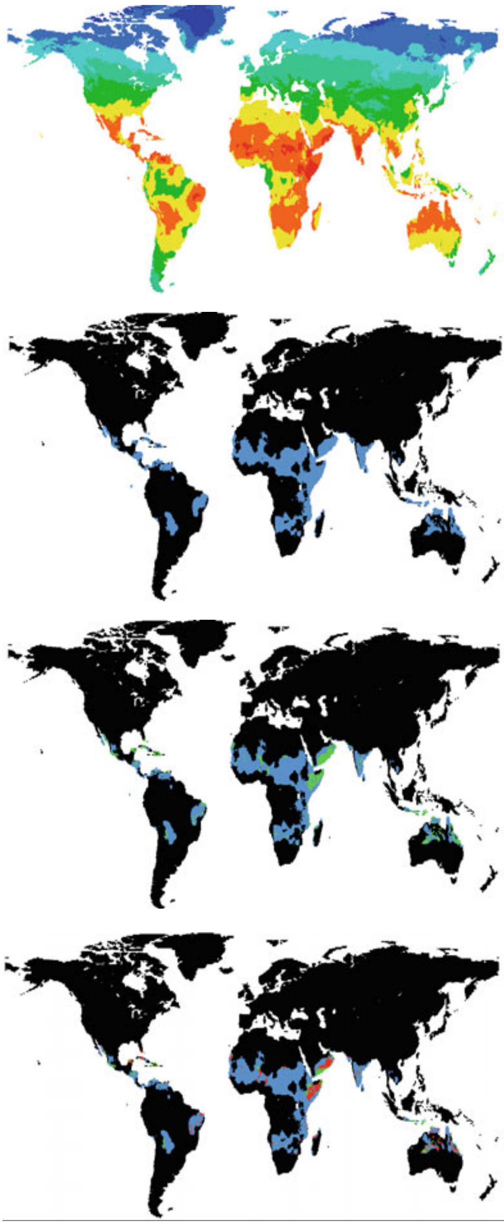


Fig. 12.4 Homoclime, homotop, and homolith for Harare



(a) Climate similarity index for an area in the South of Somalia 1.5°N 42°E with values scaled between 0 and 1.

(b) Homoclime. The first selection of similarity based on a threshold of 1-std dev of the similarity index.

(c) Homolith
The second step is to select similar lithology ≥ 0.5.

(d) Homotop, Homosoil.
The final step, within the area of homoclime and homolith, to select areas with similar topography. The areas in red represent the selection with value > 1- std dev of similarity index.

Fig. 12.5 Decision tree for selecting homosoil for an area in Southern Somalia

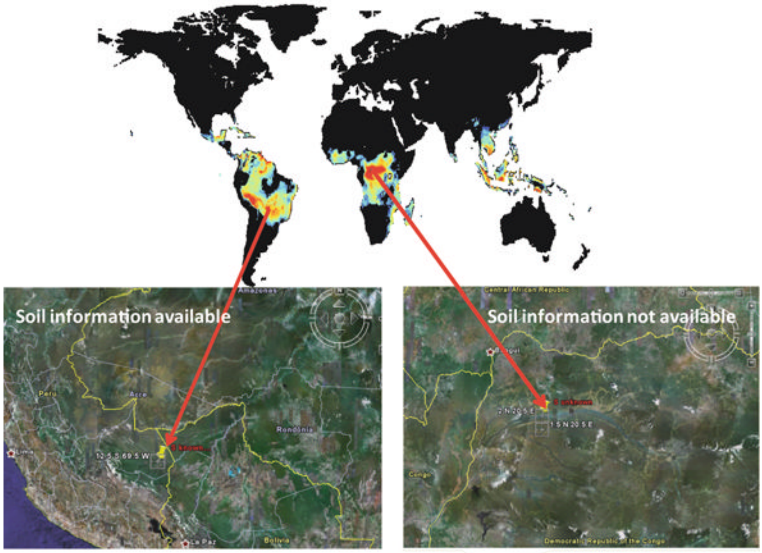


Fig. 12.6 An illustration of the homosoil concept. Soil information in the Democratic Republic of Congo (1.5°N, 20.5°E) is not available, meanwhile areas in Peru (12.5°S, 69.5°W) soil information is available. Therefore the soil-landscape rules in Peru can be exported to areas in Congo

12.5 Discussion and Conclusions

The homosoil approach allows us to export soil-landscape rules from one area to unmapped area. The first approach is based on the hierarchical nature of soil forming factors: climate, lithology and topography. In reality the interaction within these three factors and also biological factors maybe the same. We choose this approach as the first step considering the scale we are working, and at the moment, we can only validate the approach to well known global climate pattern. We have attempted to define homosoil areas with equal weighting for climate, lithology, and topography and the result is not promising, as the result showed patches of local homosoil near the location of interest.

Biological factors and local topography are important, but this factor hopefully will be taken care of when we use the local rules which are exported from other areas in the world. Of course, extreme landuse will influence soil properties, thus homosoil results with soil-landscape rules from an area with homosoil but completely different landuse may not work. An important factor of the age is not yet taken into account in this example (See Chapter 4 for the development of the age factor).

The decision-tree approach has other limitations, such as it is an aggregated result with no uncertainty estimation. The assumption is that it is not possible to have a similar area if the lithology group is different. This condition may restrictive considering the rough scale of the lithology and lithology groups. A better taxonomic distance for the lithological groups is required.

There are other important issues and questions that still need to be addressed:

- Which are the most important soil forming-factors? This of course partly depends on the scale and resolution of the global dataset, and their accuracy.
- How do we define the threshold that quantifies similarity for homoclime, homolith, and homotop? We need empirical studies to find the appropriate values.
- Uncertainty estimation, which will involve the uncertainty of climate similarity, lithology and topography. At the local scale, the uncertainty due to landuse and local topography may be more important.

References

Adams, J.M., 1997. Global land environments since the last interglacial. Oak Ridge National Laboratory, TN, USA, http://www.esd.ornl.gov/ern/qen/nerc.html

Booth, T.H., Nix, H.A., Hutchinson, M.F., and Busby, J.R., 1987. Grid matching: a new method for homoclime analysis. Agricultural and Forest Meteorology 39:241–255.

Bui, E.N., Henderson, B.L., and Viergever, K., 2006. Knowledge discovery from models of soil properties developed through data mining. Ecological Modelling 191:431–446.

Bui, E.N. Moran, C.J., 2001. Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. Geoderma 103:79–94.

Durr, H.H., Meybeck, M., and Durr, S.H., 2005. Lithologic composition of the Earth's continental surfaces derived from a new digital map emphasizing riverine material transfer. Global Biogeochemical Cycles 19, GB4S10.

Gower, J.C., 1971. A general coefficient of similarity and some of its properties. Biometrics 27:857–871.

Hartemink, A.E., 2008. Soil map density and a nation's wealth and income. In: Hartemink, A.E., McBratney, A.B., and Mendonca-Santos, M.L. (eds.), Digital Soil Mapping with Limited Data. Springer, Dordrecht.

Jenny, H., 1941. Factors of Soil Formation. McGraw-Hill, New York, NY.

Jones, P.G., Diaz, W., and Cock, J.H., 2005. Homologue, a computer system for identifying similar environments throughout the tropical world, version Beta a.0. Centro Internacional de Agricultura Tropical (CIAT), Cali, Colombia.

McBratney, A.B., Mendonça Santos, M.L., and Minasny, B., 2003. On digital soil mapping. Geoderma 117:3–52.

McKenzie, N.J., Jacquier, D.W., Maschmedt, D.J., Griffin, E.A., and Brough, D.M., 2005. ASRIS Australian Soil Resource Information System. Technical specifications. Version 1.5. October 2005. Australian Collaborative Evaluation Program.

Mourier, B., Walter, C., Merot, P., 2008. Soil distribution in valleys according to stream order. Catena 72:395–404.

New, M., Hulme, M., and Jones, P.D., 1999. Representing twentieth century space-time climate variability. Part 1: Development of a 1961–90 mean monthly terrestrial climatology. Journal of Climate 12:829–856.

Prescott, J.A., 1938. The climate of tropical Australia in relation to possible agricultural occupation. Transactions of the Royal Society of South Australia 62:229–240.

Prescott, J.A., 1943. The Australian homoclime of the zone of natural occurrence of *Parthenium argentatum*. Transactions of the Royal Society of South Australia 67:312–318.

Russell, J.S., and Moore, A.W., 1970. Detection of homoclimates by numerical analysis with reference to the brigalow region (eastern Australia). Agricultural Meteorology 7:455–479.

Schenk, H.H., and Jackson, R.B., 2005. Mapping the global distribution of deep roots in relation to climate and soil characteristics. Geoderma 126:129–140.

Schultz, J., 2005. The Ecozones of the World: the Ecological Divisions of the Geosphere. Springer-Verlag, Stuttgart, Germany.

Tranter, G., McBratney, A.B., and Minasny, B., 2009. Using distance metrics to determine the appropriate domain of pedotransfer function predictions. Geoderma 149:421–425.

Zinck, J.A., 1990. Soil Survey: Epistemology of a Vital Discipline. ITC, Enschede, the Netherlands.