4.2. SOTER Methods

4.2.1. Introduction

The standard, GIS-based approach to regional analysis is to prepare geographic layers for which unique sets of driving variables are derived, including land use, climate and soils (Paustian et al., 1997; Falloon et al., 2002; Batjes, 2004a). These GIS layers, with their associated attribute data, can then be fed into a range of models.

Common methodologies for collating and formatting regional data sets on land use, climate, and soils were adopted for the GEFSOC project. This permitted development of a uniform protocol for handling the various input for the GEFSOC Soil Carbon Modelling System (Easter et al., 2005).

This chapter focuses on the collation, screening, consolidation, and application of national-scale soil data sets using so-called "SOTER methods." It also indicates how the work described here helped to fulfil the project's research objectives.

4.2.2. Compilation of Primary SOTER Databases

SOTER Methodology

The GEFSOC project opted for the internationally endorsed SOTER methodology for compiling primary data on soil and terrain conditions. The SOTER approach has been developed by ISRIC, FAO and UNEP, under the aegis of the International Union of Soil Sciences (IUSS), in the context of the ongoing update of information on world soil resources (Oldeman et al., 1993). Ultimately, SOTER at a scale of 1:5M is to supersede the Soil Map of the World (FAO-Unesco, 1971-1981; FAO, 1995).

In many aspects, the SOTER methodology resembles physiographic or land systems mapping. It allows mapping and characterization of areas of land with a distinctive, often repetitive, pattern of landform, lithology, surface form, slope, parent material, and soils. The SOTER approach involves no new ground surveys, being based upon available data (van Engelen et al., 1995).

Each SOTER database is comprised of two main elements, a geographic component and an attribute data component. The *geographic database* holds information on the location, extent, and topology of each SOTER unit – this information is managed using a geographic information system (GIS). The *attribute database* describes the characteristics of the spatial unit and comprises both area data and point data – this information is handled using a relational database management system (RDBMS).

Each soil component within a SOTER unit is characterized by a *typical* profile (Figure 4.2.), identified as being regionally representative by national soil experts. Being derived from available soil survey reports, complete and uniform sets of soil analytical data were seldom available for all profiles. Such data gaps preclude the direct use of *primary* SOTER databases in environmental assessments and modelling. These were therefore filled using a system of taxotransfer and expert rules (Section 4.2.3).

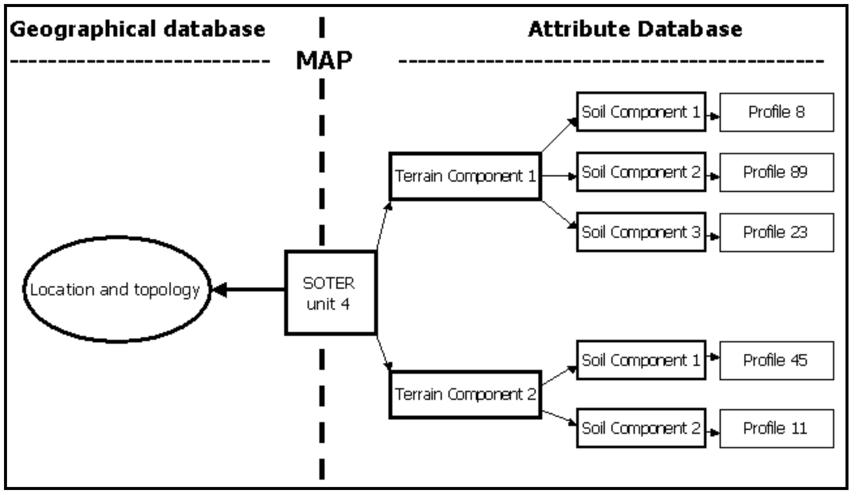

Figure 4.2. Schematic representation of a SOTER map unit with its geographical and attribute data.

Data compilation

The scale at which data were compiled for the national scale SOTER databases was generally determined by the wishes of the host countries. Consequently, the SOTER databases considered in the GEFSOC project have different scales. These range from 1:500 000 in the case of Jordan to 1:5M for Amazon-Brazil. The level of detail, both in terms of soil geographical and attribute data presented, can also vary depending on the base materials available in the four countries.

While Jordan, Kenya and Brazil already had a national scale SOTER database (KSS, 1995; ACSAD, 1996; NSMLUP, 1996; FAO et al., 1998), this was not yet the case for the Indo-Gangetic Plains (IGP) of India. Following a SOTER training in Nagpur, staff at NBSS&LUP/ICAR compiled a SOTER compatible data set for IGP-India (Bhattacharyya et al., 2004). All these *primary* SOTER materials were screened, consolidated, and re-formatted during the GEFSOC project (Section 4.2.3).

At the request of the GEFSOC consortium, a small SOTER sample database was also compiled to accompany the user instructions for the GEFSOC Soil Carbon Modelling System (Easter et al., 2005).

4.2.3. Preparation of Secondary SOTER Data Sets

Procedure

Data collated in *primary* SOTER attribute tables can be linked to GIS, permitting a wide range of assessments ranging from land evaluation to soil carbon sequestration. In previous studies, however, gaps in the *primary* data had to be filled using tailor-made solutions. Therefore, a consistent procedure for filling gaps in the existing *primary* SOTER databases was developed during the GEFSOC project, resulting in new, so-called, *secondary* data sets. Prior to applying this procedure, however, all existing *primary* data sets were screened for possible inconsistencies and re-formatted where necessary.

Special attention was paid to the inputs required for the spatial runs of the two organic carbon models considered in the GEFSOC Modelling System — RothC and Century. These include: location and relative extent of soil type, soil drainage status (i.e. hydricity), content of clay, sand and silt, content of organic carbon, and bulk density per depth layer (Paustian et al., 1997;

Falloon et al., 1998). This limited set has been expanded to include 18 soil parameters commonly required in studies of agro-ecological zoning, food productivity, soil gaseous emissions/sinks, and environmental change to permit a wide range of applications of the *secondary* SOTER data sets.

The gap-filling procedure involves three stages (Batjes, 2003), the desirability of which decreases from 1 to lowest 3:

- Stage 1: Collating additional soil geographic and attribute data where these exist, in the uniform SOTER format
- Stage 2: Using expert estimates and common sense to fill selected gaps in the measured data in a secondary data set
- Stage 3: Using taxotransfer rules (TTR) to derive soil parameter estimates for similar FAO soil units, clustered by textural class and depth range, complemented with a system of expert rules.

By their nature, stages 1 and 2 were the primary responsibility of the case study partners, while ISRIC's work focussed on methodology development (3) and the elaboration of the *secondary* SOTER data sets. The most appropriate option(s) varied from country to country, depending largely on the overall accessibility to and quality of the available soil data (Table 4.2.). In the case of Brazil, for example, there was no direct need to collate additional soil profile data in SOTER format during stage 1 (see Batjes et al., 2004b). Alternatively, the opposite was true for Jordan but no new profiles could be accessed/supplied by the case study partners. In the case of Kenya, however, some 50 new profiles from so far underrepresented regions were supplied by Kenya Soil Survey for further processing by ISRIC. Similarly, a completely new data set was submitted for IGP-India (Bhattacharyya et al., 2004). Thereafter, during stage 2, a number of synthetic and virtual soil profiles had to be created for Jordan (9) and Kenya (47), while this was not necessary for the Brazilian and Indian data sets. Finally (stage 3), the scheme of taxotransfer- and expert-rules was applied to all databases resulting in four new, consistent, *secondary* SOTER data sets.

Table 4.2. Overview of data consolidation procedures used in the case study areas.

Stage	Case study area			
	Amazon-Brazil	IGP-India	Jordan	Kenya
1	-	X[a]	-	X
2	-	-	X	X
3	X	X	X	X

[a] The underlying soil data were collated specifically for the GEFSOC project (Bhattacharyya et al., 2004); see text for details.

The status of the data sets, screened and consolidated at ISRIC during the GEFSOC project, is summarized in Table 4.3. Country-specific details may be found in the reports for Jordan (Batjes et al., 2003), Kenya (Batjes et al., 2004a), Brazil (Batjes et al., 2004b), and Indo-Gangetic Plains of India (Batjes et al., 2004c).

Table 4.3. Main characteristics of SOTER databases consolidated for the GEFSOC project.

Case study Area	Scale	Area (x1000 km^2)	Number of				Profile Density (per 10^3 km^2)
			Polygons	Unique SUIDs	SCIDs per SUID[a]	Profiles	
Brazil-Amazon	1:5M	5100	571	299	1-5	331	0.06
Kenya	1:1M	583	3261	397	1-4	495[b]	0.8
IGP-India	1:1M	480	497	36	1	36	0.08
Jordan	1:0.5M	89	47	27	1-4	48[b]	0.5

[a] SUID= SOTER unit; SCID= Soil component (see Figure 4.2.). [b] Includes a number of synthetic and virtual profiles as detailed in the country reports.

Development of taxotransfer rules

Gaps in the measured data were filled using taxotransfer rules. A *taxotransfer* function is a means of estimating soil parameters based on modal soil characteristics of soil units from a combination of their classification name — which by definition implies a certain range for various soil attributes —, expert knowledge and empirical rules, and statistical analysis of a large number of soil profiles belonging to the same taxon (Batjes et al., 1997). The elaboration of taxotransfer rules thus requires the availability of large, auxiliary soil profile databases, such as WISE (Batjes et al., 1994).

The current work expanded on ISRIC's taxotransfer-related work with FAO and IIASA (Batjes et al., 1997) and a follow up study for IFPRI (Batjes, 2002), which focussed on applications of the 1:5M Soil Map of the World (FAO, 1995). During the GEFSOC project, an updated taxotransfer approach was developed for use with *primary* SOTER databases. It considers the Revised FAO Legend (FAO, 1988) — in accordance with SOTER requirements valid at the start of the GEFSOC project — and uses a more detailed procedure for aggregating the soil profile data. Data for a given soil unit were clustered according to five textural classes and five depth ranges (0-20, 20-40, 40-60, 60-80 and 80-100 cm). Alternatively, only two depth classes (0-30 and 30-100 cm) and 3 *topsoil* textural classes were used in the preceding TTR-work for applications with the Soil Map of the World, which considered the original Legend (FAO-Unesco, 1974).

All taxotransfer- and expert-rules have been flagged in the *secondary* databases, to provide an indication of the inferred, confidence in the soil parameter estimates presented for the four case study areas (see Batjes et al., 2003; Batjes et al., 2004b; Batjes et al., 2004c; Batjes et al., 2004a).

Linkage to GIS

The soil parameter estimates generated for the constituent soil components of a given SOTER unit, as characterized by the typical profiles (see Fig. 4.2.1), were linked to the national scale SOTER GIS-files using the unique SOTER unit identifiers. These files provided consistent, geo-referenced soil data sets for the subsequent assessment of soil carbon stocks and changes at national scale, using empirical SOTER-methods (see Section 4.2.4) and the dynamic GEFSOC Soil Carbon Modelling System© (see Chapters 5-8).

4.2.4. Application of Secondary SOTER Sets

Stage 3 of the GEFSOC project (Figure 4.1.) included a comparison of estimates of national SOC stocks, for “current” conditions, computed with the GEFSOC Soil Carbon Modelling System with independent estimates obtained using so-called “existing techniques.” The latter generally involve combining soil mapping units and soil point data.

Whereas estimates of regional scale SOC stocks were already available for Amazon-Brazil (Chapter 5) and IGP-India (Chapter 6), this was not so for Kenya and Jordan. Therefore, new SOTER-methods were developed to compute national-scale SOC stocks using data held in the *secondary* SOTER databases.

The data set for Kenya was used for methodology development. Four different methods were compared (Batjes, 2004b):

(a) the carbon content to 0-30 cm and 0-100 cm computed for each representative profile, which was then linked to the spatial information held on the GIS map annexe database
(b) as above, but using the *average* content of carbon computed per FAO soil unit
(c) as above, but using the *median* carbon content
(d) through simulation of *phenoforms*, using the typical profile as the *genoform* (Droogers et al., 1997; Bouma et al., 1998).

Method *d* was found to be the most useful, because it allows defining 95% confidence intervals for median soil carbon stocks at national scale, as opposed to the *single* estimates obtained with methods *a, b* and *c.* So it was recommended for use with *secondary* SOTER data sets (Table 4.4.). In the case of Amazon-Brazil, similar gross results were obtained by other researchers using different methods – this despite the different spatial patterns mapped by these methods –, providing a ‘validation’ of method *d* (Batjes, 2005b).

Table 4.4. Stocks of organic carbon estimated using SOTER-methods.

Study area	Area (x1000 km^2)	Depth (cm)	Organic carbon (Tg C) [a]
Amazon-Brazil	5100	0–30	23943 – 24151
		0–100	42343 – 43814
IGP-India	480	0–30	572 – 587
		0–100	1163 – 1184
Jordan	89	0–30	76 – 78
		0–100	136 – 139
Kenya	582	0–30	1892 – 1911
		0-100	3669 – 3715

[a] Data shown are 95% confidence intervals for the median, using simulation of *phenoforms.* For methodological details see Batjes (2004b). 1 Tg C = 10^{12} g C.

Complementary to the direct goals of the GEFSOC project (Chapter 1), the *secondary* SOTER data sets for Kenya and Jordan, in combination with auxiliary data sets on climate and land cover, were also used to: (1) calculate the stocks of organic (SOC) and inorganic (SIC) or carbonate carbon per agro-ecological region, and (2) to project changes in soil organic carbon stocks — for defined changes in land use and management —, using an empirical approach that included a physical land evaluation (Batjes, 2004b; Batjes, 2005a). Similar to Global Agro-Ecological Zoning (Fischer et al., 2002) procedures, physical land evaluation (FAO, 1976) allows the user to filter-out areas considered biophysically unsuited for the proposed land use/management types (scenarios). Such criteria should be included in future versions of the GEFSOC Soil Carbon Modelling System.

A spin-off of the taxotransfer-related work has been the creation of a 5 x 5 arcminutes Harmonized Global Harmonized Soil Resources Database, using all continental scale SOTER data sets compiled to date (Van Engelen et al., 2005).

4.2.5. Conclusions

The secondary SOTER data sets for Brazil, IGP-India, Jordan, and Kenya are appropriate for a wide range of environmental applications at national scale. These include agro-ecological zoning, land evaluation, modelling of soil carbon stocks and changes, and studies of soil vulnerability to pollution.

Linkage between the geographic component and soil attribute data in SOTER required generalisation of measured profile data. This involved transformation of variables that show a marked spatial and temporal variation and that have been determined in a range of laboratories using various analytical methods, and over a number of years.

A uniform procedure was developed to fill gaps in measured (*primary*) data, resulting in consistent *secondary* soil data sets for the four case study countries. These new data sets provided the soil parameters estimates and soil geographic input for the dynamic modelling phase of the GEFSOC project (Chapters 5 to 8).

Although based on the best available soil geographic and attribute data, as well as an elaborate scheme of taxotransfer- and expert-rules, various sources of uncertainty will remain in the *secondary* SOTER sets. When applying these data in modelling studies at national scale, these sources of uncertainty must be understood and accepted. Similarly, various types of uncertainty will be attached to the models themselves (Burrough, 1986; Goodchild, 1994; Bouwman et al., 1999; Raupach et al., 2005).

Ideally, the national soil survey organizations should continually update and expand their national soil information systems, thus allowing periodic updates of the existing *primary* SOTER databases. Unfortunately, however, systematic soil surveys have been abandoned in many countries. Contrary to what has been the case so far, the delineation of SOTER units can now be standardized and refined using Digital Elevation Models (Dobos et al., 2002; King et al., 2002). Once such revisions have been carried out, new *secondary* data sets can easily be generated for the regions under consideration using the SOTER-methods described here.

Summarizing, the work described here contributed to delivering the soil-related outputs listed under GEFSOC Research Objectives 2 and 3 (see Project Outcomes 1 and 2). Besides being suited for modelling soil organic carbon stocks and changes at national- and sub-national scale (see Chapters 5-8), the secondary SOTER data sets can be used for a wide range of other natural resource assessments at national scale. All *secondary* SOTER data sets produced during the project will be made available in the public domain, thereby ensuring their free-access to a wide range of potential users. Finally, the present SOTER-methods can be used for other regions of the globe as new *primary* SOTER data sets become available.