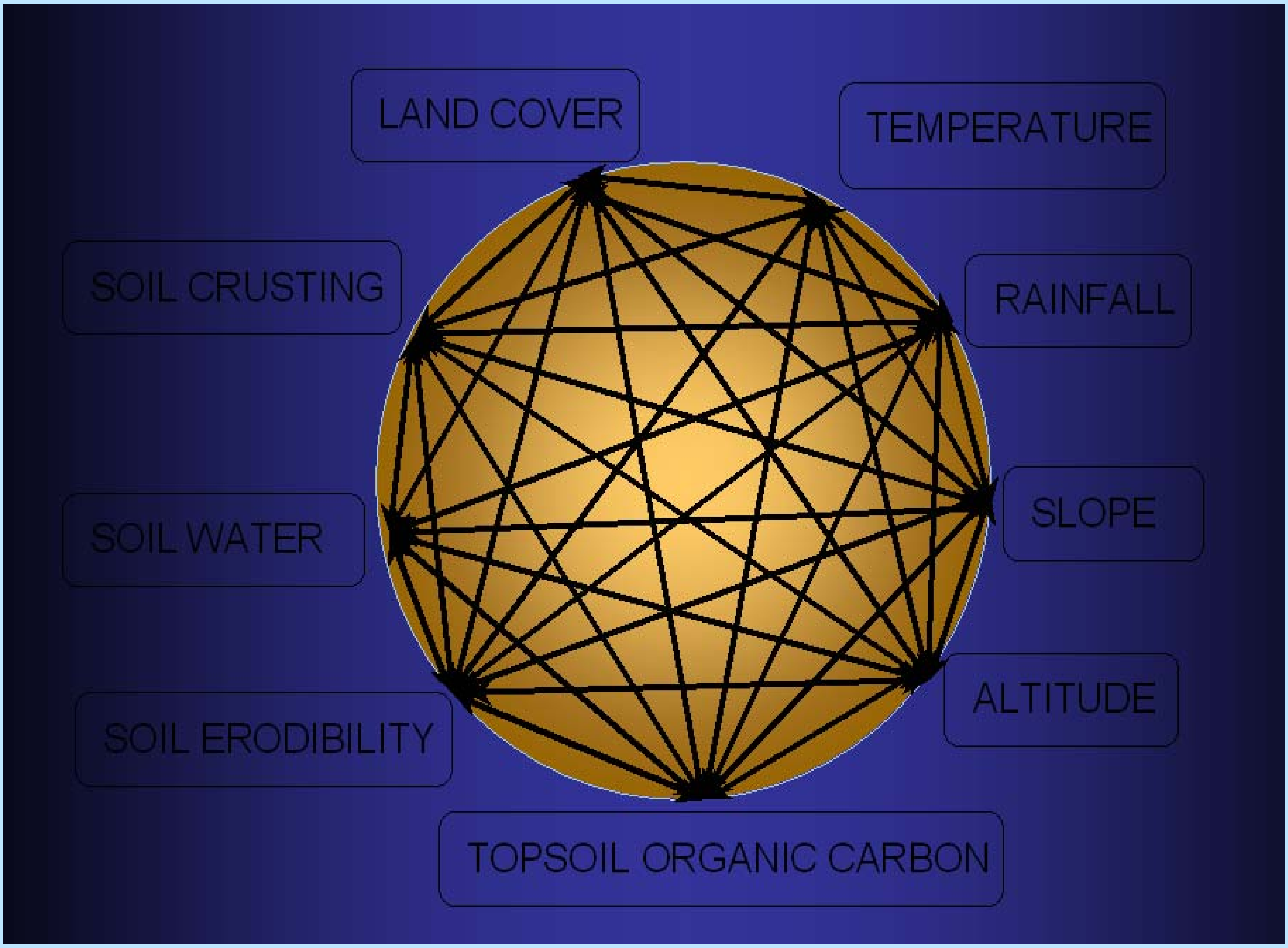
EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

An Innovative Approach for Updating Soil Information based on Digital Soil Mapping Techniques

Senthil-Kumar Selvaradjou, Luca Montanarella, Florence Carre, Arwyn Jones, Panos Panagos, Kaliaperumal Ragunath, Ramalingam Kumaraperumal, and Subramanian Natarajan



EUR 22545 EN 2007

Land Management and Natural Hazards Unit

Soil Data and Information Systems (Action – SOIL)

Institute for Environment and Sustainability

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

An Innovative Approach for Updating Soil Information based on Digital Soil Mapping Techniques

Senthil-Kumar Selvaradjou[1], Luca Montanarella[1], Florence Carre[1], Arwyn Jones[1], Panos Panagos[1], Kaliaperumal Ragunath[2], Ramalingam Kumaraperumal[2], and Subramanian Natarajan[2]

[1]Institute of Environment and Sustainability, Joint Research Centre, Ispra, Italy
[2]Dept. Soil Science & Agrl. Chem., Tamil Nadu Agricultural University, Coimbatore, India

EUR 22545 EN

Land Management and Natural Hazards Unit

Soil Data and Information Systems (Action – SOIL)

MISSION OF THE INSTITUTE FOR ENVIRONMENT AND SUSTAINABILITY

The mission of the Institute for Environment and Sustainability is to provide scientific and technical support to the European Union's policies for protecting the environment and the EU Strategy for Sustainable Development.

European Commission
Directorate-General Joint Research Centre
Institute for Environment and Sustainability
http://ies.jrc.cec.eu.int
http://www.jrc.cec.eu.int

EUR 22545 EN
ISBN 92-79-03878-8
ISSN 1018-5593

Luxembourg: Office for Official Publications of the European Communities

This publication may be cited as:
Selvaradjou, S-K., Montanarella, L., Carre, F., Jones, A., Panagos, P., Ragunath, K-P., Kumaraperumal, R. and S. Natarajan. 2007. An innovative approach for updating soil information based on digital soil mapping techniques. EUR 22545 EN. 34 pp. Office for Official Publications of the European Communities, Luxembourg. ISBN 92-79-03878-8

A PDF version is available at
http://eusoils.jrc.it

CONTENTS

ABSTRACT

In most part of the world the information on the thematic soil maps (soil erosion, soil degradation, soil organic matter content etc.) are developed as a tool for policy and management support. This information are typically derived through expert interpretation or empirical modeling approaches using typically decades old soil information originating from field investigation, laboratory analysis, reports etc. In recent period, there is a strong emphasis to update the existing soil information in a cost-effective and accurate manner. The advancements in the emerging Geographical Information System (GIS) and digital soil mapping techniques are found to be handy to derive tools addressing the above mentioned problem. In this study, we propose a novel innovative approach to address the issues on evaluating the traditional soil maps and updating the existing soil information based on the principles of digital soil mapping i.e. deriving objective soil information by reformulating the relationships between soil and its environment using ancillary and minimal datasets. The new approach is called as "SEIMS network" (Soil and Environment Interaction based Mapping System). The SEIMS network is a Data Mining and Knowledge Discovery (KDD) based method derived by fusing GIS, DSM techniques along with working principle of the Diagnosis and Recommendation Integrated System (DRIS) approach. The SEIMS network based approach provides a scheme to transform and update the less detailed, discrete and subjectively derived soil information into a continuous, non-subjective and quantitative spatial datasets. The study was tested on the soil erosion map of Tamil Nadu region, southern part of Indian Peninsula. The SEIMS network attempts to reformulate the soil erosion map of Tamil Nadu region and redefine the contours by spreading back the knowledge acquired from the relationship among the soil and its environmental variables used as predictors in the system. The variables like temperature, rainfall, potential evapotranspiration, rainfall seasonality, land cover percentage (derived from MODIS spectral bands), soil crusting, soil erodibility, top soil organic carbon content, altitude and slope that are having major influence on soil erosion were chosen as predictors for characterizing the relationship among the variables on the context of soil erosion process. The test on the efficiency of the SEIMS network's capability to extrapolate and derive the target soil information using the ancillary datasets was performed over established soil erosion map of Europe. The erosion index value derived through SEIMS network scheme exhibited a better correlation with PESERA soil erosion estimates (r^2 = 0.81), thereby proving its ability to mimic and characterize the soil erosion process by studying the complex interrelationship among the environmental variables. The weights derived for prediction are mathematically unique, thereby holds scope for further elaboration on its application to derive a tool, addressing the upscaling and downscaling issues in digital soil mapping. The flexibility and reproducibility are the main advantages visualized for this approach. Moreover, the results are objective and easy for interpretation. This study demonstrated that the SEIMS network holds some potential features that can be exploited in digital soil mapping to evaluate and update the existing soil information with minimal ancillary datasets.

.

1. INTRODUCTION

In recent times, the demand for soil information with more detailed soil spatial and attribute information is getting relatively higher for many environmental modeling and land management (Beven and Kirkby, 1979; Burrough, 1996; Corwin et al., 1997; and Jury, 1985; McBratney et al, 2003). Currently, the stocks of available conventional maps (WOSSAC, FAO, EUDASM etc.) are the major source of soil spatial information for these applications. However, these existing maps were not designed to provide the detailed (high-resolution) soil information required and in some cases these maps are not appropriate enough to address various environmental issues (scale, soil content etc.). The conventional soil maps derived by traditional soil mapping are high subjectivity and lacks complete documentation. Therefore there is an implicit need for new methods like Digital Soil Mapping (DSM), which would help us to develop hybrid methods using the present modern spatial information technologies like geographical information systems (GIS), remote sensing and the Global Positioning System (GPS). But, at any case, the information in conventional soil maps derived through expert knowledge is also equally important and should be impregnated on developing DSM tools.

In general, soil mapping is a process that depicts the relationships among soils and some environmental variables through a soil-landscape model. Inspite of the possibility to characterize the environmental conditions in detail with the advancement of new supportive technologies (GIS, GPS, remote sensing etc.), characterization of the soil-environmental relationship still remains to be a challenge for soil mapping (Hudson, 1992; McKenzie et al., 2000) in order to substitute the subjective manual mapping approach (McBratney and Odeh, 1997; Zhu, 1997). Some hybrid approaches fusing the soil scientist's specific knowledge on the soil-environment relationship with DSM based techniques are under evolution and still needs further research for their successful application (Shi, 2004).

In most part of the world, various soil survey activities are being directed towards revision of existing soil information. The existing soil information, which is typically decades old, is originating from field investigation, laboratory analysis, reviews of other surveys, etc. and holds the following limitation on its use: 1) limitations inherent in the manual process preventing total consistency on its application to the existing models; 2) the difficulty to incorporate the new emerging knowledge on soils in a revised soil information database; 3) The lack of information on the soil-landscape model that are implicitly applied the during the soil survey. To solve this problem, spatial data mining focusing on updating existing soil survey information can be of vital use. The knowledge discovery and data mining (KDD) attempts to detect previously unknown relationships among the data (McBratney, 2003). In its favour, a number of variables derived like Digital Elevation Matrix (slope, aspect, planform and profile curvatures, wetness index, and other terrain indices), land cover (remote sensed spectral bands, derived products like NDVI, EVI etc), climate (rainfall, temperature, potential evapo-transpiration, etc.) can be used as a suite of environmental variables for extracting the information on soil-landscape relationships through data mining and knowledge discovery (Burt, *et al*., 2006).

In favor to the high level of demand to evaluate the traditional soil maps and reformulating the relationship between the soil and its environment into it, the present development in digital soil mapping techniques holds potential to achieve it with the available ancillary data and even with minimal datasets. Attempts were made by few research workers to update the hydrological information in the traditional maps on its semantic part using possibility theory (Cazemier et al., 2001) and some geometric studies to redefine the contours of soil maps (Behrens et al., 2006). But till date, relatively not so much work had been done in this concern and little research was focused to deal with both semantic and geometric component on updating the soil information of the existing soil maps.

Therefore, through this study we attempt to deal with updating both semantic and geometric components of existing soil information i.e. finding back hidden rules for establishing the target soil information and re-delineating the contours for ancillary data. In this case, we tried to exploit the conceptual principle behind the "Diagnosis and Recommendation Integrated System" (DRIS) approach capable of analyzing the equilibrium condition of a system. The DRIS approach developed by Beaufils (1957; 1971; 1973) for foliar nutrient diagnosis has been expanded for its application on soil nutrient involving the essential and non-essential nutrients (Beaufils and Sumner, 1976). Later, this method had been extensively applied around different parts of the world on different crops, specifically to monitor the nutrient balance in the crop and soil. Assuming the ability of the DRIS approach to assess the balance behavior or equilibrium state of a system can be exploited for knowledge discovery and prediction of target map, we developed an innovative approach named ""Soil and Environment Interaction Mapping System" (SEIMS) in a form of network coupling the data mining capability of GIS techniques and the knowledge discovery and prediction capability of DRIS approach. Therefore, the SEIMS network approach is the Data Mining and Knowledge Discovery (KDD) based method derived by fusing GIS, DSM techniques along with DRIS approach. The KDD based component in the SEIMS network helps with developing the semantic part and later the application of the discovered knowledge through KDD helps with developing or updating the geometric part of the target soil information.

Generally, the SEIMS network approach is designed to acquire knowledge through a learning process and by finding optimum thresholds among the predictor variable ratio pair expressions. Later through this designed ability to learn from the given input, it develops a knowledgebase and trains the inputs to predict and deliver better output. This model has a strong potential application to characterize large scale complex problems of any system.

2. THEORETICAL BACKGROUND OF SEIMS network

The SEIMS network, holds three compartments as follows: 1) set of input units, *z*, representing the input variables that are used as predictors, and the *EIS (Equilibrium Index of the System)*, set of output unit, representing the target variable which is a soil map in this study, interconnected by a set of intermediary units, *r*, f(r), and *I*, representing the optimized ratio pair expressions, functions of the optimized ratio pair expressions and the indices of the predictors, respectively, that will be described in detail later. The schematic representation of the workflow of the SEIMS model is presented in Fig 1. The first compartment is representing the input data layer from available sources.

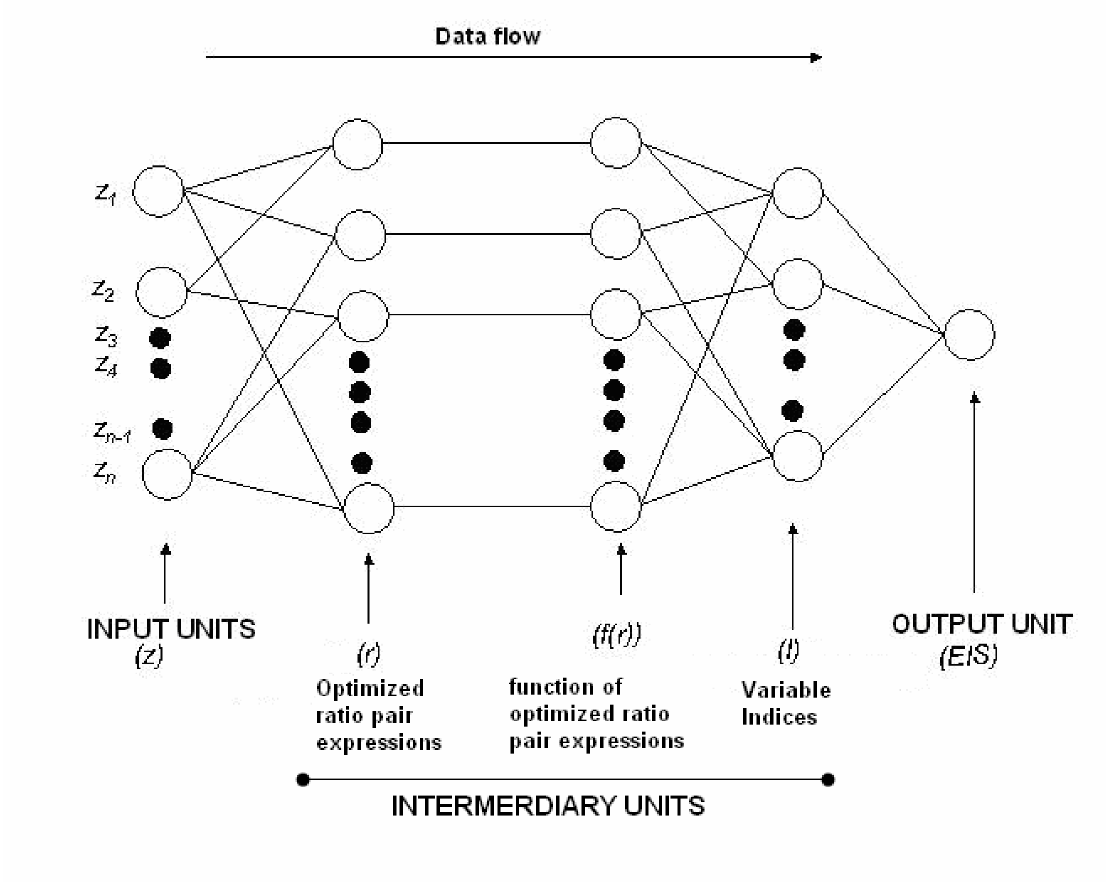


Fig. 1. Schematic representation of a SEIMS model structure

The second compartment of the SEIMS network, involves the optimization of indices, the calculation of functions of the various ratio pair expressions and the index for each input variables using the DRIS approach. We attempt to elaborate the potential application of the concept behind DRIS approach for assessing the soil-environment relationship. The soil-environment relationship is extracted through optimization of the ratio pair expressions among the variables introduced as predictors and expressed through a variable index in a continuous scale. Thus, each variable index describes its effect on the balance of the system and its role on the complex interaction of the process.

The indices of the predictors are expressed by positive or negative values. The positive or negative values indicate the degree of relative negative or positive influence of the concerned variable over the equilibrium of the system. Relative closeness of the indices towards zero for all the variables, the system tends to be reaching a equilibrium condition leading to environmental stability.

The approach assumes that the ratios of the input variables are frequently a better indicator of the relationship among the variables to derive the indices than individual variables. It is also assumed that maximum equilibrium of the system is reached only when ratios of the variables with dominant influence for that spatial location are in vicinity to the ideal or optimum conditions. The reference spatial locations are identified from input soil information that are introduced for updating process In the identified reference space that are used to derive the optimized ratio pair expressions, it is assumed that the variance of the variable ratios having strong influence on the system is smaller than the non-reference locations. Henceforth, the relationship between the variances of the reference and non-reference space can be used as an appropriate indicator for identifying the effectiveness of the predictors. The indices can be computed individually, for each variable, using the mean variable ratio deviation obtained from the comparison with the thresholds fixed for each optimized ratio pair expressions of a given variable ratio. The ideal value of the variable index for each predictors should be zero, which means that the system is in better balance/ equilibrium state.

Building training, validation and test GIS database

The first step for implementation of the SEIMS model is the identification of the predictor variables.

Second step is the establishment of the optimized variable ratio thresholds of the predictors. The optimized ratio thresholds which are named as “weights” are always obtained from the space domain with better equilibrium of the system (e.g. unaffected areas or areas under none or slight risk of erosion, areas with high top soil organic carbon content etc.).

In order to have a better applicability and precision for the generalization of the weights over diverse environmental conditions and ecosystems, it is implicit to include all possible spatial variability in order to have a representative spatial domain for deriving the thresholds of the optimized ratio pair expressions.

The size of the database or extent of the spatial domain might not be directly related to quality of the derived weights unless the necessary diversity of the environment and ecosystem in the driven system is introduced in the training spatial domain.
The weights derived from a smaller spatial domain or plot measurements holding the variability conditions would be more representative and efficient for improving the prediction capability of the model than thresholds derived from a larger GIS database that fails to include the diverse environmental variations in it. Apart from the spatial extent and variability of the system, it is evident that the quality of the input data used as predictors will also have significant impact on the precision of the predicted target output.

The spatial domain to be used for deriving the threshold ratio expression should be subdivided into two single class maps, first map representing the locations where the system is in equilibrium (full membership for the space, eg. areas of none to slight risk of soil erosion) and the second map representing the space that lacks equilibrium in the system (eg. areas of severe soil erosion risk). The GIS database with the above information can be built up from any relevant existing source. It is recommended that the spatial distributions of each of these two categories of the spatial domain are used in the training to derive weights that contain at least 10 % of the overall spatial extent by area in the GIS database.

Deriving the weights or ratio thresholds

The weights are defined after deriving the relation between the variables pairs used for prediction and their respective standard deviations or coefficient of variation. In, general, there are two possible expressions of the ratio among the variable pairs, which are discriminated by choice of only one expression for deriving the ratio threshold values. The choice among the forward and inverse ratio pair expression is made by selecting the one that has a higher variance ratio among the population in the reference spatial domain (eg. area under none to slight risk of soil erosion) and the non-reference spatial domain (eg. areas under severe risk of soil erosion).

Calculation of indices

After deriving the weights, the results are used to calculate the functions for each variable pair ratio ($f(r)$) as shown in eq. 1.

$$f(r) = \left[\left(\frac{r}{\bar{r}} - 1\right) \times \frac{1}{cv}\right] \quad if \;\; r > \bar{r}$$

$$else \quad f(r) = \left[\left(1 - \frac{\bar{r}}{r}\right) \times \frac{1}{cv}\right] \text{......................Eq.1.}$$

where r is the optimized ratios of each variable pair, $\bar{r}$ is the mean thresholds of the optimized variable ratio pair derived over the reference spatial domain, cv is coefficient of variation of the optimized variable ratio pair over the reference spatial domain.

The variable index is the average function of all the ratio pairs containing the corresponding variable and is derived as in eq. 2.

$$I_{(i=(1,n))} = \left(\frac{\sum_{i=1}^{i=n} f(r_i)}{n} \right) \qquad \text{------Eq. 2.}$$

where *n* is the number of variables.

Each variable index value is a function of the coefficient of variation associated with the corresponding ratio pair associated with the concerned variable that reflects the relative influence of the ratio expression on the relative balance of the system. During the variable index calculation, the presence of the corresponding variable in the numerator in the expression of the ratio pair will be summed to the following variable ratio expression and denominator will be subtracted with a set of corresponding variable ratio pairs used for deriving the specific variable index.

The absolute sum values of the variable indices generate an additional index denominated as Equilibrium index of the System (EIS) computed as in Eq. 3.

$$EIS = \sum_{i=1}^{i=n} |I_i| \qquad \text{------Eq. 3.}$$

This index can be useful to identify the relative equilibrium of the system as quantitative value. The higher the EIS value, the larger will be the indication of instability of the system's equilibrium and therefore, higher is the risk on the system. As the position of the variable in the ratio pair expression is taken into account for its summation or negation for deriving the corresponding variable index, the overall sum of all these indices are balanced around equilibrium condition, therefore, the sum of all the variable indices must be zero.

3. UPDATING THE SOIL EROSION MAP OF SOUTHERN PENSINSULAR INDIA (TAMIL NADU REGION) USING SEIMS NETWORK

3.1*.* Data Mining and Knowledge Discovery from existing Soil Erosion Map

In this case study, the SEIMS network was tested in Tamil Nadu region, the southern part of peninsular India (Latitude 8°5'N-13°35'N and Longitude 76°15'E-80°20'E). The Tamil Nadu study area is about 13 Mha of which nearly 5.9 Mha are under agriculture activities and, 3.0 Mha are under irrigated lands. Forest accounts to about 22933 ha. The temperature ranges from 12 to 42.8 °C in the plains and 4.6 to 33.5 °C in the hill stations. The mean annual rainfall is around 970-1078 mm. The population is about 62.4 million with a population density of 480/sq km (Statistical hand book, 2005)

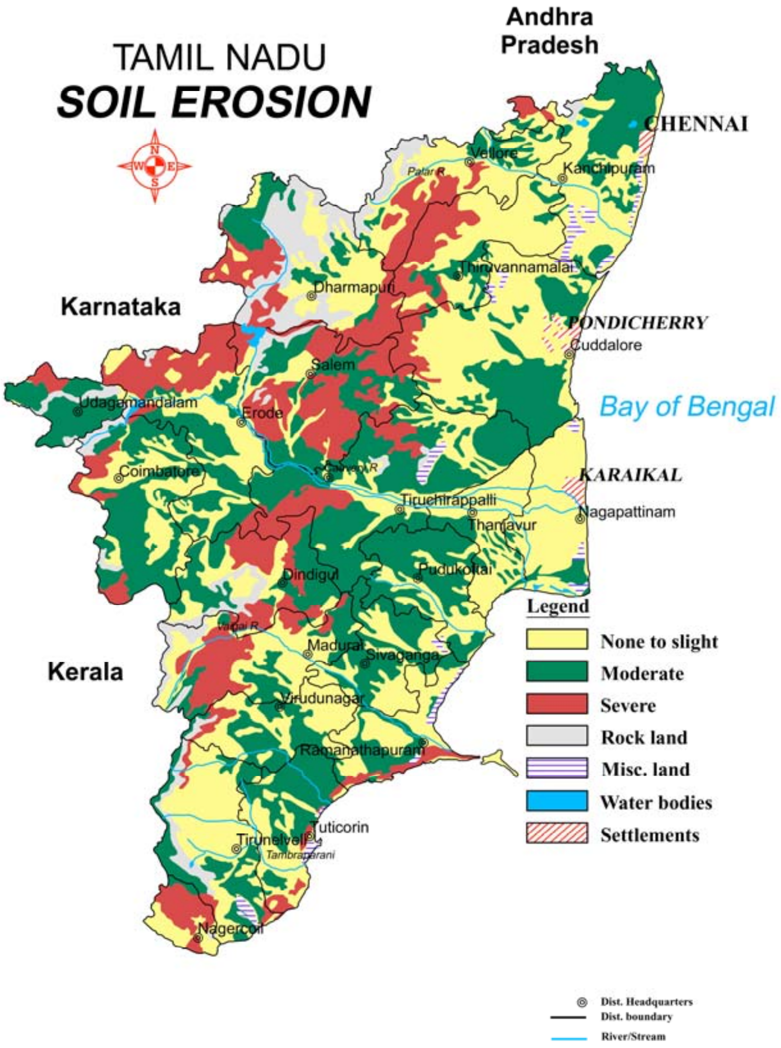


Fig. 2. Soil Map of erosion of Tamil Nadu region (NBSS & LUP, 1997)

Demarcation of spatial domains specific for knowledge discovery

To address this stage of the SEIMS model, we have to define the areas in which the system under study is in equilibrium and disequilibrium. The printed soil erosion map of Tamil Nadu region (NBSS & LUP, 1997), located in southern Peninsular India was selected for investigation to test-validate the SEIMS model's practical applicability. The description of the map is given below:

The soil erosion map was scanned at high resolution and was converted to grid using IMAGEGRID command in ArcInfo 9.1. The resulting output was a multiband raster. Individual bands were geo-referenced and reclassified into specific classes of soil erosion as indicated in the source map based on the appropriate membership of the pixel value. Different single class maps membership grade or membership value ranging from 0 (no membership in the class) to 1 (full membership) were developed for different erosion classes. Pixel values over the study area with multiple memberships yielding a fuzzy class map were discarded for this study.

These single class maps representing individual erosion classes of interest as an input for knowledge discovery to build a knowledgebase were merged assigning proper classes as in the source map. The raster generalization tools were used to fill in the areas with fuzzy membership originating from the source map in relation to pixels representing location names, boundaries etc. A final georeferenced raster as a digital soil erosion map for Tamil Nadu region as in Fig. 3.

Establishment of knowledgebase using the DRIS module

The areas with none to slight erosion were used as a workspace representing the areas in which the system is in better equilibrium state and the areas with moderate to severe erosion were used as representative for the workspace with the state of instability in the equilibrium condition of the system.

Here we use the same principle segregating into two spatial domains for studying the relationship among the independent variables influencing soil erosion (predictors) in a complex pattern depending on the geographical location.

Choice of predictors

The variables used to characterize the soil erosion process are decided based on the major factors acting as driving forces and the data availability of the variables influencing the process (McBratney, 2003; Vrieling 2005:). The model has the flexibility to adjust for the variables used depending on the data availability and the differences in the soil erosion process over different areas. Common data layers used in this study to describe climate includes rainfall, rainfall seasonality, temperature, potential evapotranspiration(PET); for land use include the land cover derived from the remote sensed data such as Normalized Difference Vegetation Index (NDVI); for relief includes elevation and slope derived from SRTM version 4 and for soil includes crusting, erodibility and top soil organic carbon derived from the FAO-UNESCO soil map. The variables for the area are characterized at 90-m resolution based on the digital elevation model recently produced by SRTM version 4 (Table 1). The variables of the climate at 1000-m resolution (Table 1), land cover at 500-m resolution and the soil at 1:5 million scale were resampled to 90-m resolution. Finally, the environmental variables used are mean annual rainfall, seasonality of rainfall (coefficient of variation), mean annual temperature, mean annual PET, land cover percentage, soil crusting, erodibility, top soil organic carbon content, altitude and slope percentage.

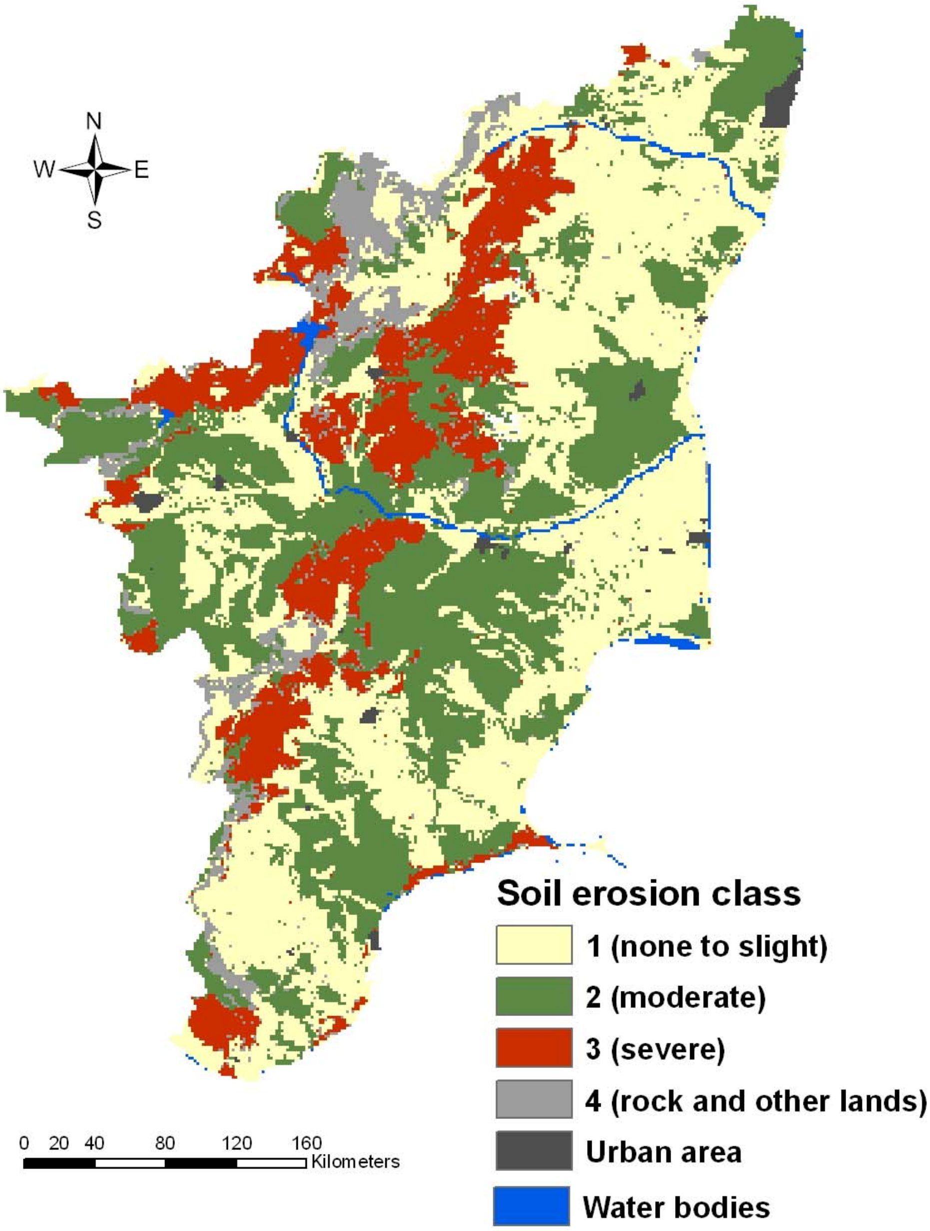


Fig. 3. Soil Map of erosion of Tamil Nadu region

Table 1. Details on the Variables used as predictors, the data sources and methodology used.

S.No.	***Predictors***	***Source***	***Reference***
1.	Mean annual temperature	WORLDCLIM 1.4	Hijman et al., (2006) *Methodology:* Hijman et al., (2005)
2.	Mean annual rainfall	WORLDCLIM 1.4	Hijman et al., (2006) *Methodology:* Hijman et al., (2005)
3.	Mean annual potential Evapotranspiration	WORLDCLIM 1.4	Hijman et al., (2006) *Methodology:* Hijman et al., (2005)
4.	Seasonality of rainfall (variation of rainfall over the year)	WORLDCLIM 1.4	Hijman et al., (2006) *Methodology:* Hijman et al., (2005)
5.	Soil crusting	Derived from 1:5 million scale FAO-UNESCO soil map	*FAO, 1988*
6.	Soil erodibility	Derived from 1:5 million scale FAO-UNESCO soil map	*FAO, 1988*
7.	Soil organic carbon content	Derived from 1:5 million scale FAO-UNESCO soil map	*FAO, 1988*
8.	Land cover %	Derived from MODIS spectral bands provided by The Global Land Cover Facility, University of Maryland; Department of Geography	Townshend et al., (2001)
9.	Altitude	Void-filled seamless SRTM data V3, 2006, International Centre for Tropical Agriculture (CIAT), available from the CGIAR-CSI SRTM 90m Database: http://srtm.csi.cgiar.org	Void-filled seamless SRTM data V3, (2006)
10.	Slope	Derived from the DEM of SRTM version 3	Void-filled seamless SRTM data V3, (2006)
11.	Soil Map	NBSS & LUP	NBSS & LUP (1997)

Deriving optimized ratio pair expressions

Knowledge acquisition based on all possible ratio pair expressions of the variables employed for prediction was done over two workspaces established. The climate-landuse-relief-soil (CLRS) relationships were extracted through the optimized ratio pair expression of each variable over the other variables were derived based on the higher variance of the possible ratio pair expressions between the areas that designated as lower in soil erosion risk and areas with high soil erosion risk. These derived weights are stored in the knowledgebase.

The knowledgebase is composed of optimized ratio combination, position of the variable (either in numerator or denominator during ratio computation in the optimized ratio pair expression) and the coefficient of variation of the optimized ratio pair expression in the area where the system is assumed to be in equilibrium i.e. the areas where soil erosion risk is more or less of lower concern or negligible.

Table 2. Thresholds derived for the knowledgebase in SEIMS network.

Factor ratio expressions	Equilibrium domain		Non equilibrium domain		$\sigma^2_{eq}/\sigma^2_{neq}$	Choice of expression
	$\bar{r}$	cv	$\bar{r}$	cv		
z_{mtemp}/z_{meanrf}	0.06	0.229	0.07	0.258	0.728	1
z_{mpet}/z_{mtemp}	1.61	0.012	1.60	0.019	0.440	-1
$z_{rfseason}/z_{mtemp}$	1.40	0.132	1.38	0.135	0.980	-1
z_{mtemp}/z_{crust}	20.1	0.184	19.3	0.185	1.071	1
z_{erod}/z_{mtemp}	0.07	0.122	0.07	0.144	0.679	-1
z_{mtemp}/z_{oc}	48.2	0.316	43.4	0.432	0.657	1
z_{lcover}/z_{mtemp}	0.80	0.224	0.84	0.271	0.631	-1
z_{mtemp}/z_{alt}	2.18	2.596	1.06	2.982	3.229	1
z_{mtemp}/z_{slope}	63.3	0.914	46.2	1.130	1.226	1
$z_{mtemp}/z_{smclass}$	57.6	0.021	25.6	0.180	0.071	1
z_{mpet}/z_{meanrf}	0.10	0.227	0.11	0.262	0.700	-1
$z_{rfseason}/z_{meanrf}$	0.09	0.306	0.09	0.326	0.827	-1
z_{crust}/z_{meanrf}	0.003	0.308	0.004	0.303	0.889	-1
z_{erod}/z_{meanrf}	0.004	0.284	0.005	0.295	0.815	-1
z_{meanrf}/z_{oc}	776	0.347	672	0.474	0.713	1
z_{lcover}/z_{meanrf}	0.050	0.265	0.05	0.261	0.926	-1
z_{meanrf}/z_{alt}	40.8	2.886	18.0	3.426	3.624	1
z_{meanrf}/z_{slope}	1041	0.965	694	1.174	1.521	1
$z_{meanrf}/z_{smclass}$	946	0.215	416	0.385	1.611	1
$z_{mpet}/z_{rfseason}$	1.17	0.125	1.18	0.125	0.984	1
z_{mpet}/z_{crust}	32.5	0.186	30.9	0.184	1.129	1
z_{erod}/z_{mpet}	0.042	0.124	0.04	0.143	0.699	-1
z_{mpet}/z_{oc}	77.7	0.316	69.9	0.436	0.650	1

Table 2. (Continued…)

Factor ratio expressions	Equilibrium domain		Non equilibrium domain		$\sigma^2_{eq} / \sigma^2_{neq}$	Choice of expression
	$\bar{r}$	cv	$\bar{r}$	cv		
z_{lcover}/z_{mpet}	0.50	0.228	0.52	0.285	0.577	-1
z_{mpet}/z_{alt}	3.54	2.595	1.72	2.988	3.210	1
z_{mpet}/z_{slope}	103	0.916	74.8	1.134	1.225	1
$z_{mpet}/z_{smclass}$	93.0	0.030	41.1	0.189	0.132	1
$z_{rfseason}/z_{crust}$	28.2	0.256	26.7	0.231	1.373	1
$z_{erod}/z_{rfseason}$	0.05	0.187	0.052	0.192	0.890	-1
$z_{rfseason}/z_{oc}$	66.8	0.321	60.0	0.448	0.635	1
$z_{lcover}/z_{rfseason}$	0.59	0.269	0.62	0.325	0.615	-1
$z_{rfseason}/z_{alt}$	3.46	2.787	1.66	3.212	3.284	1
$z_{rfseason}/z_{slope}$	89.8	0.948	65.5	1.175	1.226	1
$z_{rfseason}/z_{smclass}$	80.6	0.137	35.5	0.239	1.709	1
z_{erod}/z_{crust}	1.35	0.050	1.32	0.065	0.622	-1
z_{crust}/z_{oc}	2.50	0.372	2.31	0.453	0.792	1
z_{lcover}/z_{crust}	16.4	0.334	16.3	0.354	0.901	-1
z_{crust}/z_{alt}	0.09	2.570	0.05	2.934	2.363	1
z_{crust}/z_{slope}	3.20	0.928	2.48	1.161	1.055	1
$z_{crust}/z_{smclass}$	2.94	0.157	1.37	0.255	1.762	1
z_{erod}/z_{oc}	3.33	0.355	3.06	0.447	0.747	1
z_{lcover}/z_{erod}	12.1	0.299	12.4	0.391	0.557	-1
z_{erod}/z_{alt}	0.13	2.575	0.07	2.943	2.517	1
z_{erod}/z_{slope}	4.28	0.919	3.28	1.150	1.089	1
$z_{erod}/z_{smclass}$	3.94	0.121	1.80	0.241	1.205	1
z_{lcover}/z_{oc}	38.1	0.358	34.5	0.461	0.735	-1
z_{oc}/z_{alt}	0.06	3.028	0.03	3.493	3.505	1
z_{oc}/z_{slope}	1.50	1.116	1.13	1.282	1.352	1
$z_{oc}/z_{smclass}$	1.44	0.574	0.79	0.626	2.810	1
z_{lcover}/z_{alt}	1.67	2.466	0.78	2.841	3.503	1
z_{lcover}/z_{slope}	49.8	0.958	34.7	1.145	1.442	1
$z_{lcover}/z_{smclass}$	46.2	0.217	21.0	0.258	3.445	1
z_{slope}/z_{alt}	0.05	2.506	0.03	2.397	2.203	-1
$z_{alt}/z_{smclass}$	155	1.323	137	1.273	1.372	1
$z_{slope}/z_{smclass}$	2.97	2.651	3.10	1.868	1.848	1

z_{mtemp} - mean annual temperature, z_{meanrf} – mean annual rainfall, z_{mpet} – mean annual PET,
$z_{rfseason}$ – seasonality of rainfall, z_{crust}– soil crusting, z_{erod} – soil erodibility, z_{oc}– topsoil organic carbon
z_{lcover} – percentage land cover, z_{alt}– Altitute, z_{slope} – slope percentage, $z_{smclasS}$ – soil map class value

3.2. Deriving the index for predictor variables

The optimized ratio expressions derived are used as thresholds (weights) for predicting the equilibrium of the system i.e. the level of soil erosion and to derive an output in a continuous scale or as a quantitative soil erosion map.

Data characterizing the soil erosion are stored in a GIS database. An inference engine is constructed linking the knowledgebase created through Eq. 1. and the GIS database to derive the variable indices. Based on the rule set in the previous intermediate unit of the SEIMS network flow, the function of the optimized ratios and individual variable index values are computed.

In general, for pixel (i,j), the inference engine takes the data from the GIS database of the corresponding variables based on the optimized variable relationship from the knowledgebase to derive the variable indices.

The results from these case studies are discussed here to provide an assessment of the effectiveness of the SEIMS in deriving continuous soil spatial information for soil erosion estimates. The assessment will be conducted through the comparison of the indices derived from the SEIMS model with the discrete type of raster dataset derived from conventional existing soil erosion map.

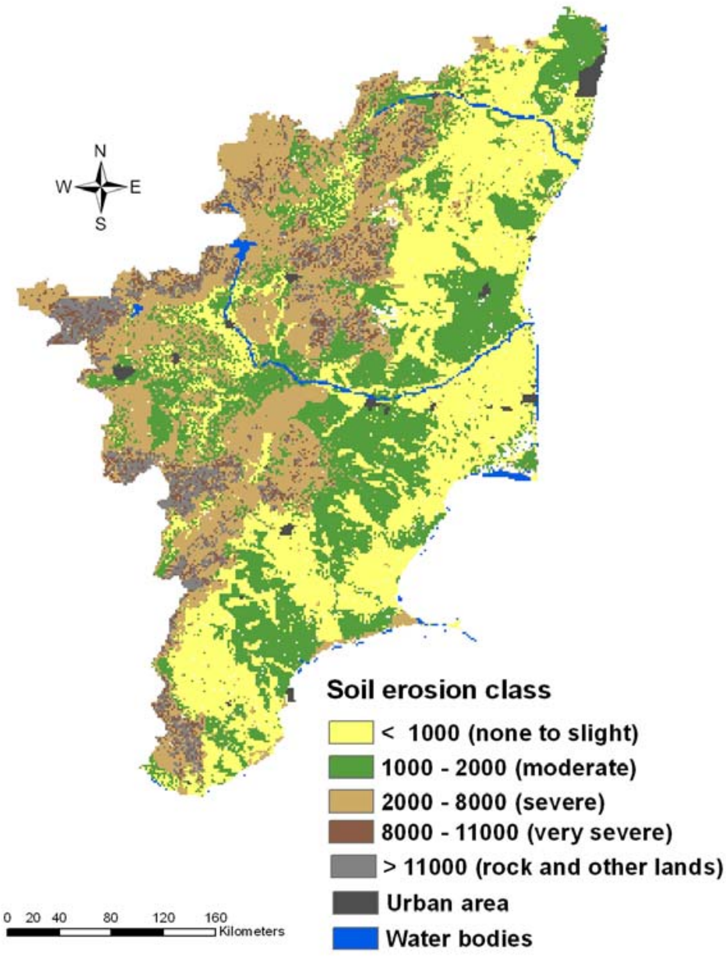


Fig 4. EIS values classified into different soil erosion classes corresponding to the source map.

The soil erosion map was derived based on EIS values (Fig 4.) using the computation of the index value of the variables used for prediction. The results were analyzed thoroughly and the reclassification of different class groups for the membership of the spatial entities among the group were performed on establishing the thresholds for the Equilibrium Index of the System, which is to be called as "Erosion Index" in the context of this study. The thresholds indicated that the erosion index values were identified as follows: < 1000 as none to slight erosion, 1000 to 2000 as moderate, 2000 to 8000 as severe, 8000 to 110000 as very severe and the index values above 11,000 are considered as the spatial locations that could correspond to bare rocks, miscellaneous lands excluded from the erosion studies. The results revealed that the spatial locations under the category of none to slight erosion level (Fig 4.) were in accordance with the source map (Fig. 2 & 3). Whereas, in the case of moderate and severe soil erosion classes, there were overlapping membership of the spatial entities between the source and the output. Further, subcategories among the severe soil erosion levels were established and a more precise soil erosion map with various categories representing the soil loss rate in t/ha/yr was derived (Fig. 5).

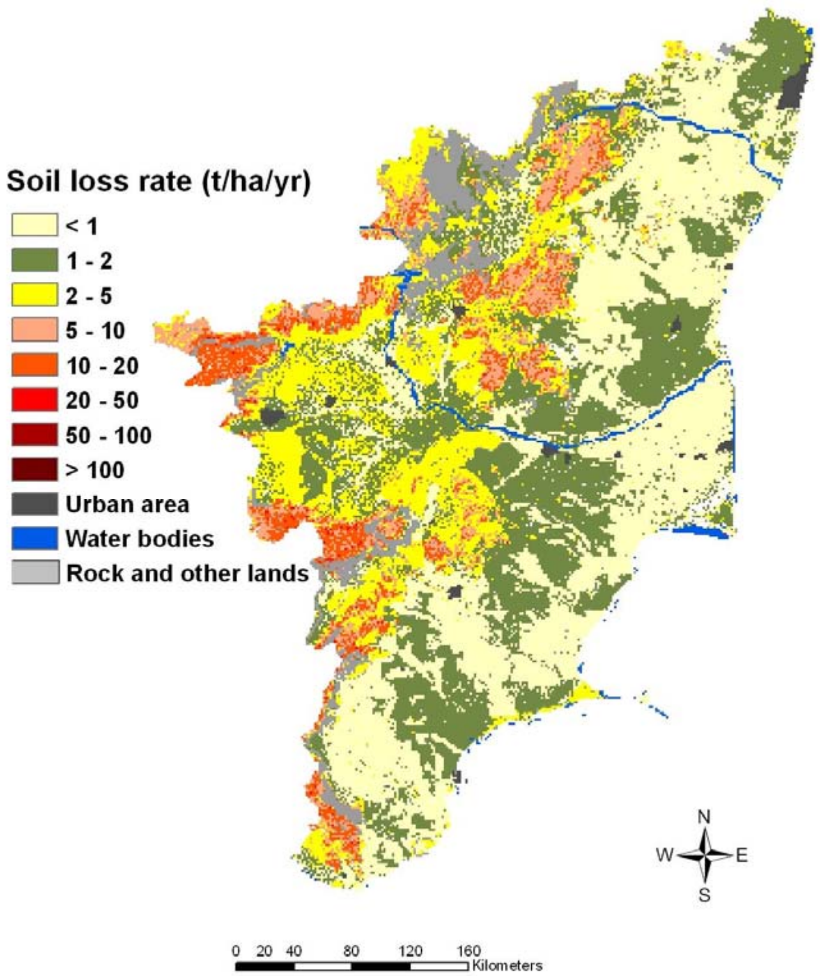


Fig. 5. New quantitative Soil erosion map of Tamil Nadu region

The published soil erosion map (NBSS & LUP) used for the study was prepared at of 1: 2.5 million scale and was presented at 1:5 million scale. Even though, the map was delivered in a degraded information level, it still holds valuable information on the soil erosion that could be used as soil information for various environmental modeling purposes. But the challenging task is how far can we find a way to transform this valuable information an into operational one. The SEIMS network based Erosion index values can be used as a solution to find ways and means to overcome the drawback and transforming the information into a quantitative form applicable for various environmental modeling. In this study, using the existing information as a base and with the environmental variables (predictors) at 90-m resolution in coherence with the DEM (high resolution dataset available and used in this study), characterization of the relationship among the soil-enviromental variables and their interaction among themselves in relation to soil erosion process was achieved through the creation of a new soil erosion index map.

4. TESTING THE EFFICIENCY OF SEIMS NETWORK

Validation of any modeling approach is an implicit requirement for its applicability and use. Since 1980, various approaches were undertaken for assessing soil erosion risk in European scale as given in Table 3. The USLE approach proved to be unsuitable for European conditions, as underlined by various research studies (Gobin et al., 2003; 2004). At present, even though not exactly accurate, PESERA is being claimed to be the best model available providing quantitative estimates for assessing soil erosion risk in Europe (Gobin et al., 2003). Henceforth, PESERA soil erosion estimates and their data input sources were chosen for validation.

Table 3. Overview of soil erosion assessment approaches at European scale

Model	Type	Reference
De Ploey	Expert based	De Ploey et al., 1989
CORINE	Factorial scores	CORINE, 1992
IMAGE	Factorial scores	RIVM, 1992
GLASOD	Expert based	Van Lynden, 1994
Hot Spots	Expert based	EEA, 2000
USLE	Quantitative	Van der Knijff et al., 2000
INRA	Factorial scores	INRA, 2001
PESERA	Quantitative	Gobin and Govers, 2003

Source: Van Rompaey, (2003)

The SEIMS network was applied over the soil erosion estimation results of PESERA model, to check the appropriateness of the capacity to reproduce or mimics the system behavior based on the relationship among various environmental variables used as predictors and also its capability to extract the hidden information that had not been directly captured by the physical process model.

Creating the platform for establishing the knowledgebase

The soil erosion result as shown in the fig. 6. are taken as an input for segregating the training workspace in the SEIMS network to develop the so called weights, the optimized ratio pair combination of the variables and defining its thresholds using the DRIS approach. The details on the data processing and model details are made available in reports of the PESERA project. The reports on the details of the model functionalities can be obtained from the web portal of the European Commission (http://eusoils.jrc.it/ESDB_Archive/pesera/pesera_cd/index.htm). The documentation clearly describes the structure, execution and validation process carried over and its applicability. The details on the description of the input data source (Table 4.) ard its processing details are described in the report by Gobin et al., (2003).

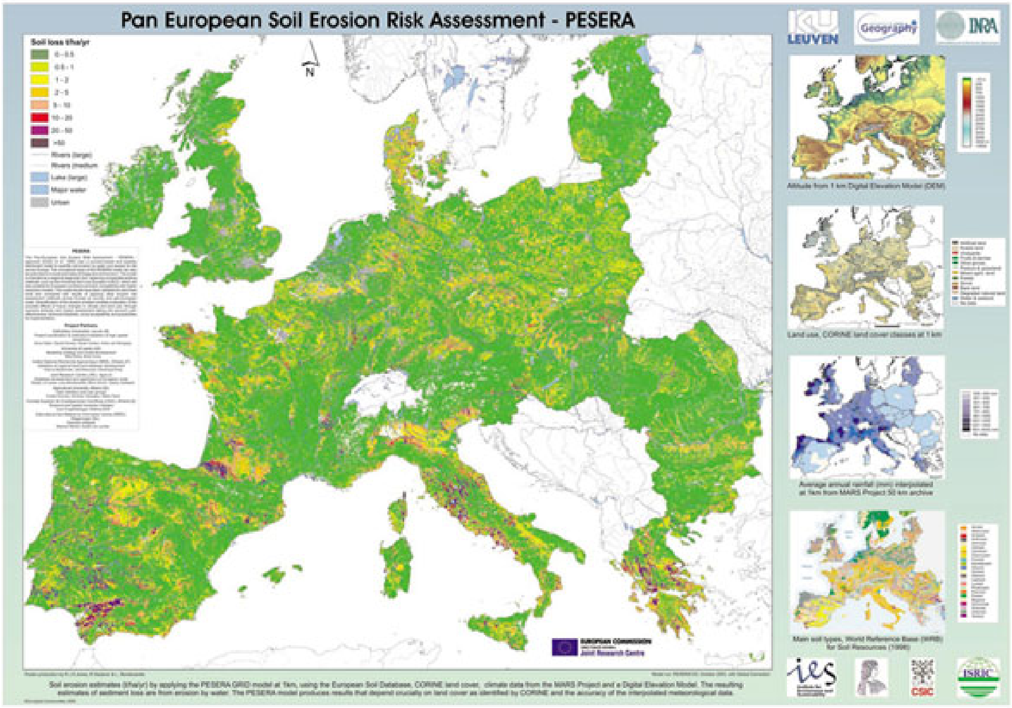


Fig. 6. The soil erosion estimates to identify the risk area due to soil water erosion in Europe.

Source: Kirkby et al., (2004) available for download in http://euroils.jrc.ec.europa.eu/

Table 4. Details on the data sources used in PESERA model.

S.No.	**Input variables/ factors**	**Source**
1.	Soil	Soil Geographical Database of Eurasia at scale of 1:1 000 000 scale, European Soil Bureau, JRC, Ispra, Italy.
2.	Meteorology	50 km grid daily meteorological database for 25 years, MARS database, JRC, Ispra, Italy
3.	Relief	GTOPO30 30 arc seconds DEM (~ 1 km), USGS HYDRO1K database, except France: 250 m DEM, Institut Géographique National, and Italy: 250 m DEM
4.	Land use & agricultural land use	CORINE land cover, 250 m resolution raster; Farm Structure Survey and NUTS regions from Eurostat

Table 5. Information details of the variables selected as predictors in SEIMS network.

S.No.	Predictors	Min	Max	Mean	Unit	Source	Reference
1.	Mean annual temperature	-12.5	19	10.1	°C	Derived from the data source of the PESERA project funded by European Commission	Kirkby et al., (2004)
2.	Mean annual rainfall	17.0	154	54.6	mm	Derived from the data source of the PESERA project funded by European Commission	Kirkby et al., (2004)
3.	Mean annual potential Evapotranspiration	17.0	123	62.3	mm	Derived from the data source of the PESERA project funded by European Commission	Kirkby et al., (2004)
4.	Seasonality of rainfall (mean annual coefficient of variation of rainfall)	-0.99	10.9	1.79	-	Derived from the data source of the PESERA project funded by European Commission	Kirkby et al., (2004)
5.	Land cover %	0	100	66.7	%	Derived from the data source of the PESERA project funded by European Commission	Kirkby et al., (2004)
6.	Soil crusting	0	5	3.31	class	European Soil Database	European Soil Data Centre (ESDAC)
7.	Soil erodibility	0	5	3.44	class	European Soil Database	European Soil Data Centre (ESDAC)
8.	Soil water storage capacity	0	183.4	100.6	-	European Soil Database	European Soil Data Centre (ESDAC)
9.	Altitude	-164	4536	375.6	Meter	Global 30-Arc-Second (GTOPO30) digital elevation model (DEM	U.S. Geological Survey - http://edc.usgs.gov/products/elevation/gtopo30/gtopo30.html
10.	Slope %	0	89.3	3.76	%	Derived from Global 30-Arc-Second (GTOPO30) digital elevation model (DEM	U.S. Geological Survey - http://edc.usgs.gov/products/elevation/gtopo30/gtopo30.html
11.	Soil organic carbon content	1	6	2.34	class	European Soil Database	Jones, et al., (2004)
12.	Soil erosion map of Europe	0	430	0.67	t/ha/yr	Data source of the PESERA project funded by European Commission	Kirkby et al., (2004)

Among the input variables used for the PESERA model and from other data sources of Europe, the choice of the influencing variables as predictors in SEIMS network was done as in Table. 5.

In order to confirm the efficiency of the SEIMS network on its prediction capability and accuracy, we decided to maintain the same input datasets that had been used for PESERA model execution to derive quantitative estimates of the soil erosion. Assuming the ability of the SEIMS network to replicate or mimic the system in synchronization with the PESERA results using the set of chosen predictors (influencing variables) as the success of the SEIMS networks's efficiency its exhibit its capacity to extrapolate in an accurate mnner, the following test was executed.

The geographical area representing Europe with approximately 324 Mha, composed of about 199 Mha in agricultural areas with 106 Mha under arable lands and 134.8 Mha under forest and semiarid natural vegetation areas was chosen. The details of the input variables are given in Table. 5. Based on the soil erosion estimates of the PESERA model results, partitioning of two workspaces were done as follows: The areas with soil erosion estimate of less than 1 t/ha/yr are considered as the space where the system is under equilibrium state and those areas with more than 1 t/ha/yr as those are under the state of instability (areas subjected to erosion) based on the threshold set up by EEA (1999) assuming that any soil loss of more than 1 t/ha/yr could be possibly an irreversible phenomenon within a time span of 50-100 years. These two spatial domains segregated based on the above threshold of 1t/ha/yr were used for characterizing the relationship among the soil-environmental variables using the DRIS approach in the SEIMS network.

Building the knowledgebase

The relationships among the predictor variables are characterized on the spatial locations in the two different workspaces. The optimized ratio pair expressions are therefore established. Using these established optimized ratio pair expressions, the so called weights, the thresholds for the ratio pair expressions were derived and stored into the knowledgebase for predicting the locations that are subjected to erosion (non-equilibrium condition of the system) in a quantitative manner. The derived optimized ratio pair expressions and their threshold values, their variation (CV), and their conditional position for the computation of the ratios are given in table 6. These were stored in the knowledgebase.

Table 6. Thresholds derived for the knowledgebase using DRIS approach in SEIMS network.

Factor ratio expressions	Equilibrium domain		Non equilibrium domain		$\sigma^2_{eq} / \sigma^2_{neq}$	Choice of expression
	$\bar{r}$	cv	$\bar{r}$	cv		
z_{meanrf}/z_{mtemp}	6.67	104	4.98	0.39	124740	-1
z_{mpet}/z_{mtemp}	7.11	73.9	5.78	0.21	193734	-1
z_{mcvrf}/z_{mtemp}	0.33	300	0.22	5.181	7964	-1
z_{lcover}/z_{mtemp}	8.78	52.2	3.21	0.400	127854	-1
z_{crust}/z_{mtemp}	0.39	79.2	0.30	0.425	60419	-1
z_{erod}/z_{mtemp}	0.40	74.6	0.37	0.368	50707	-1
z_{swsc}/z_{mtemp}	26.7	303	15.2	4.39	14771	-1
$z_{\%slope}/z_{mtemp}$	0.71	251	0.27	1.596	173557	-1
z_{alt}/z_{mtemp}	67.5	306	28.3	1.097	440927	-1
z_{oc}/z_{mtemp}	0.31	113	0.17	0.738	77910	-1
z_{meanrf}/z_{mpet}	0.9	0.43	0.89	0.411	1.20	1
z_{mcvrf}/z_{meanrf}	0.03	0.36	0.03	0.347	1.22	-1
z_{lcover}/z_{meanrf}	1.35	0.51	0.69	0.414	5.65	-1
z_{meanrf}/z_{crust}	18.0	0.55	17.8	0.359	2.40	1
z_{meanrf}/z_{erod}	17.6	0.56	14.4	0.390	3.14	1
z_{swsc}/z_{meanrf}	1.94	0.51	1.48	0.687	0.98	-1

Table 6. continued…

Factor ratio expressions	Equilibrium domain		Non equilibrium domain		$\sigma^2_{eq} / \sigma^2_{neq}$	Choice of expression
	$\bar{r}$	cv	$\bar{r}$	cv		
$z_{\%slope} / z_{meanrf}$	117	2.50	70.5	2.334	3.15	1
z_{meanrf} / z_{alt}	0.57	9.28	0.76	5.034	1.93	1
z_{oc} / z_{meanrf}	0.05	0.61	0.04	0.603	1.72	-1
z_{mcvrf} / z_{mpet}	0.03	0.23	0.02	0.250	1.03	-1
z_{lcover} / z_{mpet}	1.21	0.54	0.57	0.427	7.47	-1
z_{crust} / z_{mpet}	0.06	0.38	0.05	0.416	0.94	-1
z_{mpet} / z_{erod}	19.8	0.41	17.4	0.349	1.84	1
z_{swsc} / z_{mpet}	7.45	0.53	6.13	0.616	1.08	-1
$z_{mpet} / z_{\%slope}$	127	2.43	79.3	2.141	3.34	1
z_{alt} / z_{mpet}	6.11	1.06	4.64	0.860	2.62	-1
z_{oc} / z_{mpet}	0.04	0.68	0.03	0.659	2.14	-1
z_{lcover} / z_{mcvrf}	49.0	0.48	25.3	0.341	7.33	-1
z_{mcvrf} / z_{crust}	0.48	0.34	0.46	0.286	1.53	1
z_{mcvrf} / z_{erod}	0.46	0.37	0.36	0.226	4.21	1
z_{swsc} / z_{mcvrf}	72.7	0.35	66.4	0.392	0.97	-1
$z_{mcvrf} / z_{\%slope}$	3.24	2.41	1.74	2.149	4.34	1
z_{mcvrf} / z_{alt}	0.02	7.72	0.02	4.608	2.24	1
z_{oc} / z_{mcvrf}	1.66	0.59	1.31	0.57	1.78	-1
z_{lcover} / z_{crust}	23.2	0.63	11.5	0.44	8.28	1
z_{lcover} / z_{erod}	22.5	0.64	9.06	0.38	17.0	1
z_{swsc} / z_{lcover}	0.68	1.02	0.46	1.114	15.3	-1
$z_{\%slope} / z_{lcover}$	137	2.56	45.9	2.13	13.0	1
z_{lcover} / z_{al}	0.67	9.64	0.47	4.66	8.59	1
z_{lcover} / z_{oc}	39.0	0.71	25.0	0.54	4.20	1
z_{crust} / z_{erod}	1.02	0.30	0.84	0.29	1.50	1
z_{swsc} / z_{crust}	32.9	0.51	28.7	0.400	2.14	-1
$z_{crust} / z_{\%slope}$	7.39	2.53	3.99	2.10	5.00	1
z_{crust} / z_{alt}	0.04	8.45	0.04	4.41	2.84	1
z_{oc} / z_{crust}	0.84	0.94	0.60	0.62	4.47	-1
z_{swsc} / z_{erod}	31.5	0.49	24.1	0.342	3.59	-1
$z_{erod} / z_{\%slope}$	7.91	2.44	5.28	2.19	2.79	1
$z_{\%alt} / z_{erod}$	126	1.14	83.4	0.98	3.15	-1
z_{oc} / z_{erod}	0.82	0.96	0.47	0.58	7.93	-1
$z_{swsc} / z_{\%slope}$	237	2.48	135	2.386	3.35	1
z_{swsc} / z_{alt}	1.14	8.84	1.26	4.924	2.60	1
z_{swsc} / z_{oc}	0.03	0.88	0.03	0.869	1.20	1
$z_{alt} / z_{\%slope}$	280	2.26	223	1.92	2.19	-1
$z_{oc} / z_{\%slope}$	6.14	3.12	2.17	2.45	12.9	-1
z_{oc} / z_{alt}	-0.003	-114	0.003	38.9	9.05	-1

z_{mtemp} - mean annual temperature, z_{meanrf} – mean annual rainfall, z_{mpet} – mean annual PET,
z_{cvrf} – seasonality of rainfall, z_{crust}– soil crusting, z_{erod} – soil erodibility, z_{oc}– topsoil organic carbon
z_{lcover} – percentage land cover, z_{alt}– Altitute, z_{slope} – slope percentage, z_{swsc} – soil water storage capacity

Deriving the Soil Erosion Index – EIS of SEIMS network

On linking the knowledgebase established with the GIS database, the variable ratio pair expressions were computed and subsequently based information from the knowledgebase on the corresponding coefficient of variation values and the position of the variable in its combination with other variables, the function of variable pair expressions are derived. These derived function of variable pair expressions were used to compute the individual index value of the predictor variables. As described above in the previous case study on Tamil Nadu region, these variable index values are the indicators describing the extent and intensity of influence of the variable on the equilibrium state of a system. The more they are oriented towards the zero, the more balance is the system i.e. in equilibrium state. The cumulative index derived by summation of all the variable indices irrespective of the negative or positive value. The derived cumulative index value is the Equilibrium Index of the System (EIS). This EIS can be used as the Erosion Index, since the system under investigation in this study is focused on soil erosion. The results of the Erosion Index are presented in Fig. 7. The results had a strong similarity with the soil erosion map of PESERA and its results. As the Erosion index is represented in continuous values in a similar form to the results of PESRA model, it provides the possibility for further statistical comparing among the results of the model and to validate the efficiency of the SEIMS network.

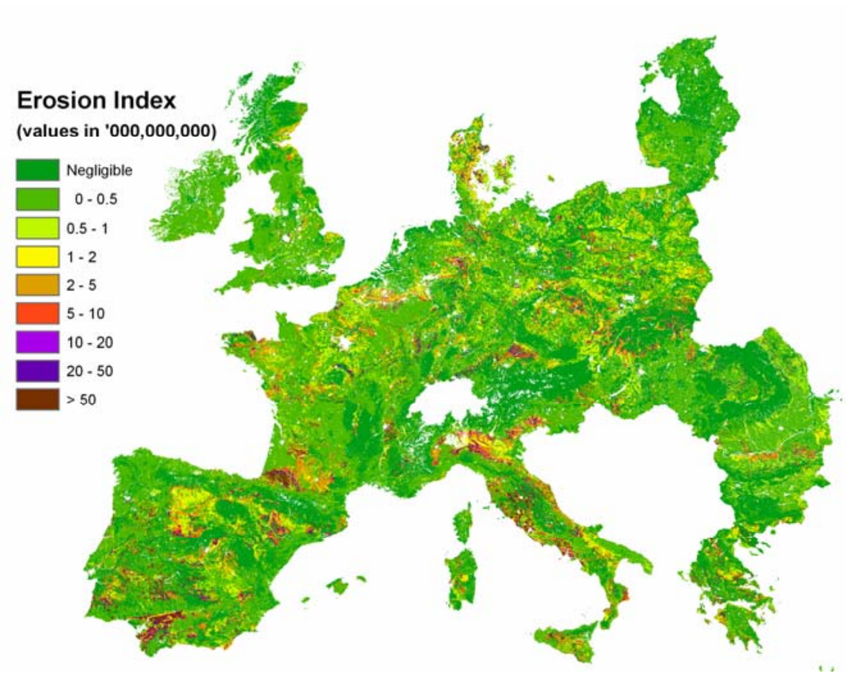


Fig.7. Soil erosion map of Europe based on SEIMS erosion index

5. INTERPRETATION OF THE MODEL OUTPUTS

The erosion index values derived using the SEIMS network exhibited a higher positive correlation (r^2 = 0.813) with the PESERA estimates and especially in those areas where the erosion estimates were more than 1 t/ha/yr, the correlation coefficient (r^2) was found to be still stronger with a value of 0.891. Based on above results and detailed analysis of the relationship between the erosion index values and the PESERA based soil erosion estimates, the index values are reclassified into similar classes with a direct transformation of values to the same class groups of PESERA as shown in the fig. xxx.

The similar results of the erosion index of SEIMS network with the PESERA estimates which is confirmed by the statistical analysis, reveals the validity and applicability of this SEIMS network as a tool for effective data mining and knowledge discovery. Using the individual variable index value of the predictors, various phenomenon of influence under the complex interacting system can be understood, which in turn would help to develop indicators for the specific process in the system and choose appropriate mitigation practices to prevent the imbalance in the equilibrium of the system i.e. soil erosion risk in the context of the study.

The index values were reclassified both for PESERA and SEIMS erosion index estimates as shown in fig.7, Based on the membership of the spatial location under different classes specified, a comparison was done by calculating the ratio between the established PESERA soil erosion classes and the SEIMS erosion index as shown in fig.8.

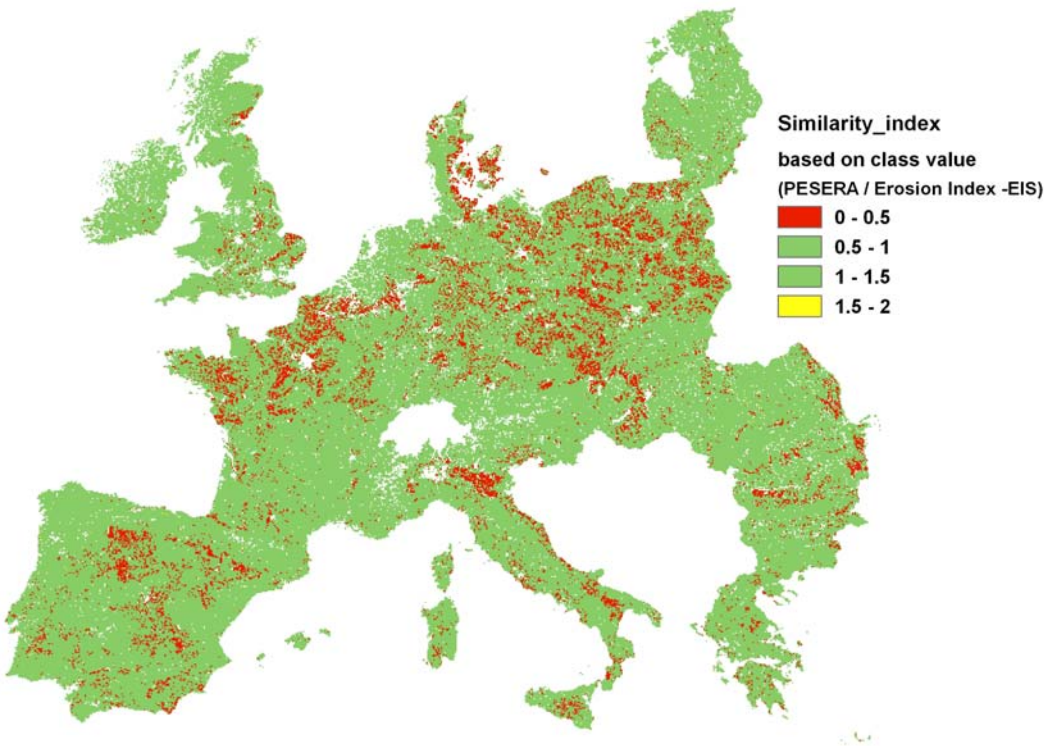


Fig. 8. Similarity index map for comparing SEIMS erosion index and PESERA soil estimates.

The similarity index described in the fig. 8. Indicates that there is a strong similarity in the results in the area between two methods when the similarity index value is between 0.5 to 1.5, and the similarity index is below 0.5, indicates the presence of overestimation or overemphasis of the soil erosion risk by SEIMS network or underestimation of the soil erosion estimates by PESERA model and the areas with the similarity index above 1.5 indicates the underestimation on the risk intensity by SEIMS network or overestimation of soil erosion estimates by PESERA model. Therefore, SEIMS network helps to identify the uncertainty or error that could probably propagate into the results from the input source data or some phenomenal changes that are not captured by the predictors chosen for this study and might have arised from some other influencing factors oriented to land use or socio-economic variables which had not been dealt at present in this study.

6. EXTENDED UTILITIES IN SEIMS NETWORK

The threshold value of the ratio pair expression of the variable in the equilibrium domain is higher than those of the non equilibrium domain of the system (Table 6.). The relatively lower coefficient of variation (*cv)* of the variable ratio pair expressions of the equilibrium domain (areas not affected by soil erosion) are better predictors for predicting the equilibrium state of the system on the unknown area. Based on the combination of *cv* of the variable forming the ratio pair expressions, matrix was developed and the order of influence of the variable on the prediction of the system is established for Tamil Nadu case study as given below:

$Z_{temp} > Z_{mpet} > Z_{erod} > Z_{crust} > Z_{mapclass} > Z_{rfseason} > Z_{cover} > Z_{meanrf} > Z_{oc} > Z_{slope} > Z_{alt}$

Similarly, for Europe the order of influence or the efficiency of predictors are given below:

$Z_{oc} > Z_{cover} > Z_{mpet} > Z_{erod} > Z_{crust} > Z_{meanrf} > Z_{alt} > Z_{slope} > Z_{rfseason} > Z_{swsc} > Z_{temp}$

Based on the order of influence of the variable on prediction, it is evident that soil erosion is lower in the higher temperature regimes in Tamil Nadu region representing the sub-tropic and tropic ecosystem. Whereas, the erosion is significantly lower in lower temperature regimes in Europe. Thereby, this indicates the effective prediction capacity of SEIMS network on characterizing the environmental variables and its relationship in an accurate manner relative to its ecosystem behavior or influence.

The presence of the variable in the numerator and denominator described by the choice of expression in the knowledgebase also provides useful information on the influential behavior of the variable on the system. The presence of the variables with higher degree in the numerator for the ratio pair expression indicates that the variable has relatively higher value in the equilibrium state of the system and the predominant presence of the variable in the denominator indicates relative lower value of these variables in the spatial domains with equilibrium condition.

In the case of the Tamil Nadu region, from the results it was observed that the rainfall, altitude, slope and the map classes used as input variable were found in the denominator in majority of the cases, which indicates that these variables had lower value in the areas where the soil erosion level is slight or negligible The temperature, erodibility and coefficient of rainfall variables had partial mixed stands in combination with other variables to form the ratio pair expression. The potential evapotranspiration, soil crusting, top soil organic carbon content and per cent land cover were found in the numerator forming the ratio pair expression in majority of the cases indicating relatively the occurrence of higher values of the variable in the areas with lower soil erosion risk areas.

In the case of soil erosion assessment using SEIMS network on Europe, the results revealed that the temperature, rainfall, potential evapotranspiration, erodibility, slope and altitude were found in the denominator and land cover, soil crusting, soil water available capacity and top organic carbon content in the numerator.

Moreover, the presence of the variable in the indicator in the numerator indicates also that soil erosion relatively increases with the relative increase in the negativeness of the derived index value of the variable, and the presence in the denominator indicates the relative increase in soil erosion with the relative increase in the positiveness of the index value of the variable.

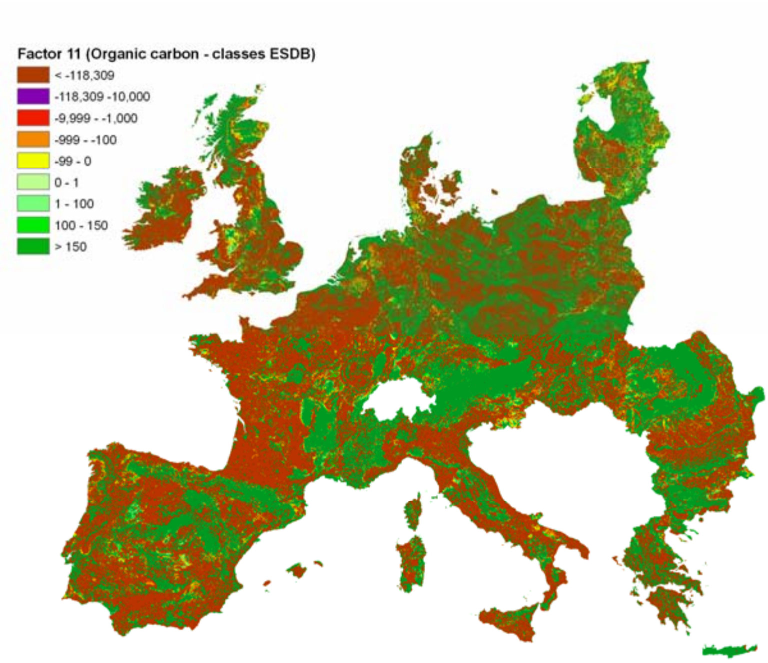


Fig. 9. Topsoil organic carbon index derived based on SEIMS network

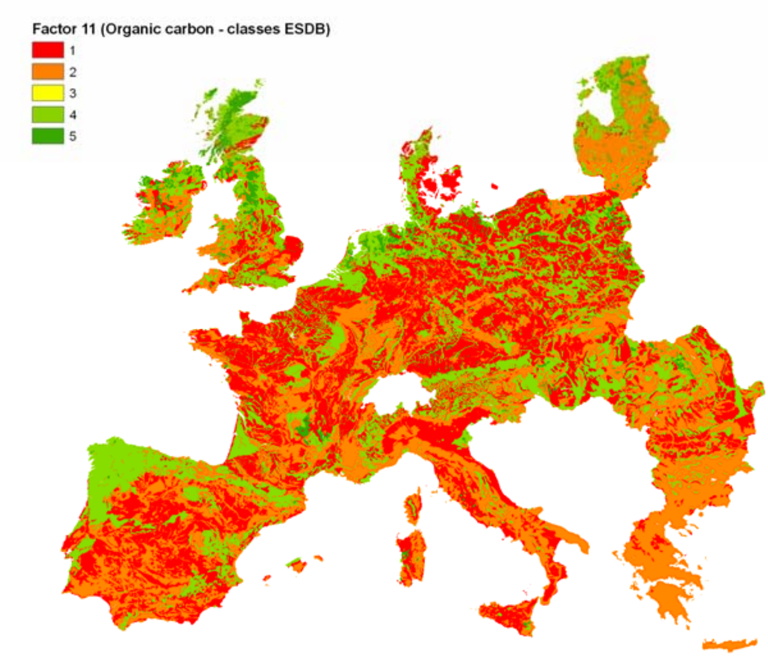


Fig. 10. Topsoil organic carbon content represented in classes in European soil database

The individual index values of the soil variables used as predictors can provide additional information on the influence of soil erosion on those variables and vice versa. As an Example, here we present the use of information from top soil organic content index. The Fig. 9. indicates the areas where the soil erosion would have strong influence on the organic carbon content of the soil. It is evident from fig. 9 & 10 that the spatial pattern of the index value are quite different from the categorical top soil organic carbon map available in the soil database. Thereby, the index value of the organic carbon can be used for various degradation assessment studies or to establish indicators to assess the impact of soil erosion on soil organic carbon which is of primary importance in various global studies like global warming, desertification etc. The quantitative information derived through this method could be more useful and appropriate to meet the needs of various environmental modeling schemes than the estimates obtained from empirical modeling sources ignoring the environmental interaction among variables or only involving certain interaction pattern between selected variables.

Based on these indicators, it is possible to assess the spatial extent of the variable influence over the geographical area. The results indicate that the soil organic carbon content is affected in 56 % of the geographic area in Europe due to soil erosion or vice versa. The soil erosion occurs in nearly 42.8 % of areas due to the poor soil water storage capacity which can be attributed by different soil type, management practices, physical properties of the soil etc.

7. FEATURES AND FORESEEN USES OF THE SEIMS NETWORK ON SOIL MAPPING

The SEIMS approach combines the knowledge/information of the local existing maps with GIS techniques under multivariate ratio-based prediction system to derive soil functional maps with required spatial and attribute information for many environmental modeling and land management applications. Thereby, SEIMS system gains the potential to overcome the limitations on the usage of the existing information on the conventional soil maps.

Using this new innovative approach, with the existing map and on extracting the complex relationship of the ancillary variables derived through remote sensing techniques, one can derive spatially continuous soil erosion map for an area.

The advantages of this approach are as follows:

1. Helps to convert the categorical soil information available in discrete soil map into a continuous scale to meet the present clamour for detailed soil information in various environmental modeling approaches.

2. The index of the variables allows characterize the influence of the variable on the system (soil) over the spatial domain.

3. Helps to identify the balance in the equilibrium of the system, even under the condition that none of the variables are dominating to impart strong influence on the system.

4. Allows identifying the error propagation or uncertainty in the existing soil information.

5. Provides scope through the thresholds of the optimized ratio expressions derived from the relationship of the variables, for better and consistent upscaling and downscaling of soil information.

6. The relationship between the variance of the variables among the reference and non-reference space (areas under risk and not under risk) helps to develop appropriate indicator for various processes involved under environmental studies.

7. Holds relative less influence by the noises propagated through the various sources (input datasets) than the other existing soil mapping techniques.

8. Model is flexible for its functionality on different scales (field level, national and global).

9. Helps to transfer and handle soil information over plot to regional scale, eg. the thresholds for the ratio pair expressions of the variables derived using the georeferenced experimental plot scale can be transferred and optimized for its applicability over the large scale datasets (watersheds, national, continental and global extent).

10. Characterizes the soil-environment relationship in simpler and easy way for interpretation and helps to gather better knowledge on the complex interaction of the system.

11. The outputs are not subjective in nature as in the case of manual soil mapping.

12. The thresholds derived for the variable pair ratio expressions are mathematically unique and does not need multiple simulation runs as in case of other robust models (ANN).

Limitations

1. The choice of the variables used for

prediction which are assumed to be the driving forces of the system under investigation should be taken care. The exclusion of the some variable that has a strong influence on the system would certainly affect the desired output. Better understanding on the system, through expert knowledge on the phenomenon behind the process driving the system would favour the successful use of the model.

2. Imprecise information on the reference space used for training the model and to build the knowledgebase would affect the target units. Thereby instead of helping to identify the imbalance in the equilibrium of the system, would serve as a tool for mapping the uncertainty or error in the input units.

3. Inclusion of all possible variables that have any relationship on the functioning of the system would improve the precision of the prediction and reduce the level of approximation of the results.

4. The optimized thresholds or weights for the variable ratios derived from a reference spatial domain will be more appropriate for generalization over diverse ecosystems, if the reference space used for training the hidden units include all the possible environmental diversity of the ecosystems.

5. In contract to the empirical modeling procedures, SEIMS network needs a rough output of the target variable to delineating the different equilibrium status of the system, i.e. for example like affected and unaffected areas as in the case of soil erosion.

Conventional soil erosion are either based on the survey through questionnaires or interpolation or extrapolation of the few point data measurements over the space assumed to have behave similarly in other similar landscape/environment or based on results of some models that are not always valid globally. The final information available through maps are more generalized based on the on the purpose of map, scale of the target map, issue under focus etc. Mostly, the available soil information on the maps created before the digital soil mapping era, lacks complete set of semantic information and have lost information due to spatial generalization due to conditional filtering depending on the scale of the target.

The existing less detailed maps that lacks much of application to the day-to-day demands of various environmental modeling and land management applications, but, through DRIS based digital mapping approach we can convert them into functional maps by enhancing the existing information on spatial variation in discrete form into a form represented as a continuum.

8. CONCLUSION

The SEIMS network is proposed through this study as a novel innovative approach by fusing the GIS, DSM techniques along with conceptual working principle of DRIS. This method had demonstrated its capability to explore and extract the hidden rules for establishing the target soil information and redelineating the contours using the ancillary data (predictors). The concept of data mining and knowledge discovery (KDD) impregnated in the network flow helps initially with the semantic part and later application of the KDD for prediction helps with the geometric part. This approach is found to be of vital use to evaluate tradition soil maps and reformulate relationships between the soil and environment. The flexibility and reproducibility are the main advantages of this approach. Moreover, the outputs of SEIMS network are less subjective and easy to interpret. The weights derived for prediction are mathematically unique, thereby holds scope for further elaboration on its application to derive a tool, addressing the upscaling and downscaling issues for digital soil mapping. This hybrid approach is useful for transforming the existing discrete soil information into continuous spatial information by extracting the complex soil-landscape relations from ancillary variables. The accuracy test on the extrapolating capacity of the SEIMS network in comparison with the PESERA model based results had proved its potential scope for being developed as an alternate tool to predict soil erosion with minimal datasets.

9. REFERENCE

Beaufils, E.R. 1957. Researc for rational exploitation of Hevea using a physiological diagnosis based on the mineral analysis of various parts of the plants. Fertilite, 3: 27-38 pp.

Beaufils, E.R. 1971. Physiological diagnosis: a guide for improving maize production based on principles developed for rubber trees. Fertilizer Society of South Africa Journal. 1: 1-28.

Beaufils, E.R. 1973. Diagnosis and Recommendation Integrated System (DRIS). Soil Science Bulletin. 1: 1-132.

Beaufils, E.R. and Sumner, M.E. 1976. Application of the DRIS approach for calibrating soil, plant yield and plant uality factors of sugarcane. Proceedings of the South Africa Sugar Technology Association. 50: 118-124.

Behrens et al., 2006. Removing uncertainities in soil class maps and nominal environmental covariates. 2nd Global Workshop on Digital Soil Mapping. Rio de Janeiro, Brazil. July 4-7.

Beven, K.J., and Kirkby, M.J. 1979. A physically-based, variable contributing area model of basin hydrology. Hydrol. Sci. Bull. 24: 43-69.

Burrough, P.A. 1996. Opportunities and limitations of GIS-based modeling of solute transport at the regional scale. p. 19–38. In D.L. Corwin and K. Loague (ed.) Applications of GIS to the Modeling of Non-point Source Pollutants in the Vadose Zone. SSSA Spec. Publ. No. 48. SSSA, Madison, WI.

Burt, J.E., Wang. R., Zhu A., Meyer. T., and Hempel. J. 2006. Spatial data mining for soil survey updates. New frontiers in soil resource assessment. The 18th World Congress of Soil Science. July 9-15, 2006.

Cazemier, D.R., Lagacherie, P. and Martin-Clouaire, R. 2001. A possibility theory approach for estimating available water capacity from imprecise information contained in soil database. Geoderma. 103: 113-132.

CORINE. 1992. CORINE - Soil erosion risk and important land resources in the southern regions of the European Community, final report EUR 13233-EN. Office for Official Publication of the European Communities.

Corwin, D.L., Vaughan, P.J., and Loague, K. 1997. Modeling nonpoint source pollutants in the vadose zone with GIS. Environ. Sci. Technol. 31:2157–2175.

De Ploey, J. 1989. Erosional systems and perspectives for erosion control in European löss areas. (in collaboration with: Auzet, A.V., Bork, H.-R., Misopolinos, N., Rodolfi, G., Sala, M. and Silleos, N.G.). Soil Technology Series, 1: 93-102.

EEA . 2000. Down to Earth: soil degradation and sustainable development in Europe. Environmental Issues Series, European Environment Agency. No. 16, 32pp.

EEA. 1999. Environment in the European Union at the Turn of the Century. Environmental assessment report 2.

FAO, 1988. FAO/Unesco Soil Map of the World, Revised legend, with corrections and updates. World Soil Resources Report 60, FAO, Rome. Reprinted with updates as Technical Paper 20, ISRIC, Wageningen, Netherlands, 1997. ISBN 90-6672-057-3.

Gobin, A., and Govers, G. 2003. Pan European Soil Erosion Risk Assessment – 3th Annual Report – EU 5th Framework Programme, Project No. QLK5-CT-1999-01323. 144 pp. Available at: http://pesera.jrc.it/

Gobin, A., Daroussin. J., and Jones, R.J.A.. 2003. Pan-European Soil Erosion Risk Assessment. Input data preparation. Deliverable 17. (Contract no. QLK5-CT-1999-01323; Key Action no. 1.1.1.-5.4.).

Gobin, A., Jones, R., Kirkby, M., Campling, P., Gover, G., Kosmas, C., and Gentile, A.R. 2004. Indicators for pan-European assessment and monitoring of soil erosion by water. Environmental Soil Science & Policy. 7: 25-38.

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis, 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965-1978.

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis. 2006. WorldClim (http://www.worldclim.org/) Robert J. Hijmans, Susan Cameron, and Juan Parra, at the Museum of Vertebrate Zoology, University of California, Berkeley, in collaboration with Peter Jones and Andrew

Jarvis (CIAT), and with Karen Richardson (Rainforest CRC).

Hudson, B.D. 1992. The soil survey as paradigm-based science. Soil Sci. Soc. Am. J. 56: 836-841

Jones, R. J.A.., Hiederer, R., Rusco, E., Loveland, P.J. and Montanarella, L. 2004. THE MAP OF ORGANIC CARBON IN TOPSOILS IN EUROPE: VERSION 1.2 - SEPTEMBER 2003 Explanation of: Special Publication Ispra 2004 No.72 S.P.I.04.72 . EUR 21209 EN. Publication of the European Communities, Luxembourg. European Commission.

Jury, W.A. 1985. Spatial variability of soil properties. p. 245–269. In S.C. Hern and S.M. Melancon (ed.) Vadose Zone Modeling of Organic Pollutants. Lewis Publication, Chelsea, MI.

Kirkby, M.J., Jones, R.J.A., Irvine, B., Gobin, A, Govers, G., Cerdan, O., Van Rompaey, A.J.J., Le Bissonnais, Y., Daroussin, J., King, D., Montanarella, L., Grimm, M., Vieillefont, V., Puigdefabregas, J., Boer, M., Kosmas, C., Yassoglou, N., Tsara, M., Mantel, S., Van Lynden, G.J. and Huting, J.(2004) Pan-European Soil Erosion Risk Assessment: The PESERA Map, Version 1 October 2003. Explanation of Special Publication Ispra 2004 No.73 (S.P.I.04.73)." by. European Soil Bureau Research Report No.16, EUR 21176, 18pp. and 1 map in ISO B1 format. Office for Official Publications of the European Communities, Luxembourg.

McBratney, A.B., and Odeh, I.O.A. 1997. Application of fuzzy sets in soil science: Fuzzy logic, fuzzy measurements and fuzzy decisions. Geoderma. 77: 85-113.

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

McKenzie, N.J., Gessler, P.E., Ryan, P.J. and O;Connelll, D.A. 2000. The role of terrain analysis in soil mapping 245-265 pp. In Wilson, J.P. and Gallant, J.C. (ed.) Terrain analysis: Principles and applications. John Wiley and Son, Newyork.

NBSS & LUP, 1997. Soil Resources of Tamil Nadu for Land-Use Planning. NBSS PUBL, 46. National Bureau of Soil Survey and Land Use Planning, Indian Council of Agricultural research.

RIVM (1992). The environment in Europe: a global perspective. Report 481505001, RIVM, Bilthoven. The Netherlands.

Shi, X., Zhu, A-X., Burt, J.E., Qi, F. and Simonson, D. 2004. A case-based reasoning approach to fuzzy soil mapping. Soil Sci. Soc. Am. J. 68: 885-894.

Statistical Hand book 2005. Government of Tamil Nadu, Department of Economics and Statistics. http://www.tn.gov.in/deptst/

Townshend, J.R.; DeFries, R.; Hansen, M.; Sohlberg, R.; Carroll, M.; DiMiceli, C (2001). MODIS 32-Day Composites, (Feb.2-Mar.5, Mar.6-Apr.6, Apr.7-May 8, May 9-Jun.9, Jun.10-Jul.11, Jul.12-Aug.12, Aug.13-Sep.13, Sep.14-Oct.15, Oct.16-Nov.16, Nov.17-Dec.26, Dec.27-31). College Park, Maryland: The Global Land Cover Facility.

Van der Knijff, J.M., Jones, R.J.A., Montanarella, L. (2000). Soil erosion risk assessment in Europe, EUR 19044 EN, 44pp. Office for Official Publications of the European Communities, Luxembourg.

Van Lynden, G.W.J. (1994). The European soil resource: current status of soil degradation in Europe: causes, impacts and need for action. ISRIC, Wageningen. Council of Europe. Strassbourg.

Van Rompaey, R.J.J., Vieillefont, V., Jones, R.J.A., Montarnarella, L., Verstraeten, G., Bazzoffi, P., Dostal, T., Krasa, J., de Vente, J., and Poesen, J. 2003. PESERA-Deliverable 7B: Model validation at the catchment scale. 1-27 (Contract no. QLK5-CT-1999-01323; Key Action no. 1.1.1.-5.4.).

Void-filled seamless SRTM data V3, 2006, International Centre for Tropical Agriculture (CIAT), available from the CGIAR-CSI SRTM 90m Database: http://srtm.csi.cgiar.org.

Vrieling, A. 2005. Satellite remote sensisng for water erosion assessment: A review. Catena. 65(1): 2-18.

Zhu, A.X. 1997. A similarity model for representing soil spatial information. Geoderma. 77: 217-242.

European Commission

EUR 22545 EN – DG Joint Research Centre, Institute for Environment and Sustainability

Title: An innovative approach for updating soil information based on digital soil mapping techniques
Authors: Selvaradjou, S-K., Montanarella, L., Carre, F., Jones, A., Panagos, P., Ragunath, K-P., Kumaraperumal, R. and S. Natarajan.

Luxembourg: Office for Official Publications of the European Communities
2007 – 34 pp.

EUR - Scientific and Technical Research series; ISSN 1018-5593
ISBN 92-79-03878-8

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

LB- NA- 22545- EN- C

Publications Office
Publications.eu.int

ISBN 92-79-03878-8