Asian Journal of Water, Environment and Pollution, Vol. 6, No. 3, pp. 73-78.

A Holistic GIS-based Approach for Thematic Extraction of the Soil Erosion Estimates Using Predictors as a Function of Climate, Land Cover, Relief, Soil and Topography

P.S. Senthil Kumar* and S. Aruna Geetha

Department of Soil Science & Agricultural Chemistry, Faculty of Agriculture
Annamalai University, Chidambaram, India
✉ senthil.selvaradjou@gmail.com

Received May 27, 2008; revised and accepted March 3, 2009

Abstract: Present claim for update of existing soil information has taken a heavy toll to fit the needs of the current environmental modelling data demands. The information derived from the age old data of 1960's and 1970's that are being used in most cases at present situation are losing its relevance to represent the reality of now existing soil status. Due to various transformations that have undergone in the land use, crop management practices, intensive cultivation integrated with unscrupulous fertilization (imbalanced fertilization), certain fertile soils of the past have reached a status of degraded lands or unproductive lands. Henceforth, present focus is visualized on developing modelling approaches through exploitation of the new GIS and remote sensing techniques as a feasible option and to cut down the cost factor that would be a certain unaffordable demand through conventional approaches. In this study, "SEIMS network" (Soil and Environment based Mapping System) approach was adopted to update information on the soil loss due to water erosion. Conceptually, this approach is based on the principles of Data Mining and Knowledge Discovery (KDD) method. The spatial relationships among the independent variable related to the soil erosion process (predictors) are accounted to estimate soil erosion through spatial modelling. In this study, about four climatic variables (temperature, rainfall, potential evapotranspiration and rainfall seasonality), one for land cover (derived from MODIS spectral bands), three variables for soil attributes (soil crusting, soil erodibility, top soil organic carbon content) and two terrain parameters (altitude and slope) were chosen as predictors for modelling soil erosion process. The reclassified soil erosion map derived through SEIMS network scheme exhibited a better correlation (r^2 = 0.891) with the published class-based soil erosion map of Tamil Nadu (NBSS & LUP, 1997). Thereby, holistic GIS-based approach was found to be efficient in transforming the useful subjective, qualitative and categorical information into objective and quantitative information serving the present demands of soil information update.

Key words: Soil erosion, Geographical Information System (GIS), spatial modelling, data mining, digital soil mapping.

Introduction

In recent times, the demand for more detailed spatial and attribute information is getting relatively higher for many environmental modelling and land management. The information present in non-generic soil maps like that of soil erosion map are extensively subjective in nature. There is a huge demand for this information as an input for environmental modelling pertaining to global scale assessment studies on climate change, land degradation etc. Developing databases from field observation for soil erosion at regional and continental levels are not practically feasible option and the demand on the intensity of point observation is also highly variable based on the scale involved in the assessment study. Moreover, the information presented in existing soil erosion maps are qualitative and subjective in nature, for being derived from expert assessment (opinion) and/or various

*Corresponding Author

empirical modelling approaches those are location specific and are handicapped for its cross-boundary applicability. Henceforth, there is a strong emphasis for development of new techniques using spatial modelling approaches involving GIS and remote sensing techniques.

In general, data mining refers to a set of tools for computer-automated, exploratory data analysis. With the distinct power of discovering previously unclear knowledge in spatial data, geographic data mining not only improves our spatial data analysis abilities but also helps in knowledge discovery from previously underutilized data sources (i.e. image data, map data, etc.) that provides an alternative to knowledge construction for knowledge-based systems where traditional knowledge acquisition is difficult. The objective of our study is to extract the knowledge of the particular model that a domain expert used to create the 'area-class' map. The knowledge is usually based on the expert's experience. Our goal in the current stage is to recover the subjective expert knowledge from the 'area-class' maps using spatial analysis. We propose a novel innovative approach denominated as "SEIMS network" (Soil and Environment-based Mapping System). Conceptually, this is based on the principles of Data Mining and Knowledge Discovery (KDD) approach, wherein, the spatial relationships among the independent variables related to the soil erosion process (predictors) are accounted to quantify the estimates of soil erosion.

Materials and Methods

Location of the Study Area

The Tamil Nadu region located in the southern part of peninsular India (latitude 8°5'N-13°35'N and longitude 76°15'E-8°20'E) was chosen as the study area with a geographical coverage area of 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 22,933 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 Handbook, 2005).

Modelling Framework

The SEIMS model outlined by Selvaradjou et al. (2007) has been adopted in this study as a data mining tool for knowledge discovery from the existing soil erosion map. The scheme of the SEIMS network workflow has been depicted in Figure 1. The approach involves data preparation, data preprocessing, pattern extraction, knowledge examination and interpretation.

The data preparation step is to construct a GIS database containing relevant environmental variables and spatial relationships.

First step involves selection of input variables, i.e. identification of the predictor variables. The careful selection of covariates (predictors) plays a key role towards the success of the system. The set of variables assembled to represent the state factors of soil erosion were used as predictors to develop the model. 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. Common data layers used in this study to describe climate includes rainfall, rainfall seasonality, temperature, potential evapotranspiration (PET) (Hijman et al., 2006); for land use include the land cover derived from the remote sensed MODIS data (Townshend et al., 2001) such as Normalized Difference Vegetation Index (NDVI); for relief includes elevation and slope derived from SRTM version 4 (Void-filled seamless SRTM data V3, 2006) and for soil includes crusting, erodibility and top soil organic carbon derived from the FAO-UNESCO soil map (FAO, 1988). The variables for the area are characterized at 90-m resolution based on the digital elevation model recently produced by SRTM version 4. The variables of the climate at 1000-m resolution, land cover at 500-m resolution and the soil at 1:5 million scale were resampled to 90-m resolution.

Secondly, the soil erosion map of Tamil Nadu (NBSS & LUP, 1997) was used as referential spatial domain for training the datasets. The third step involves the establishment of the optimized variable ratio thresholds of the predictors. The optimized ratio thresholds which are named as "weights" are obtained from the domain with better equilibrium of the system (e.g. unaffected areas or areas under none or slight risk of erosion). 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, e.g. areas of none to slight risk of soil erosion) and the second map representing the space that lacks equilibrium in the system (e.g. areas of severe soil erosion risk).

The weights are defined after deriving the relationship between the covariate pairs and their respective standard deviations or coefficient of variation. In general, there are two possible expressions of the ratio among the covariate pairs. 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 (e.g. area under none to

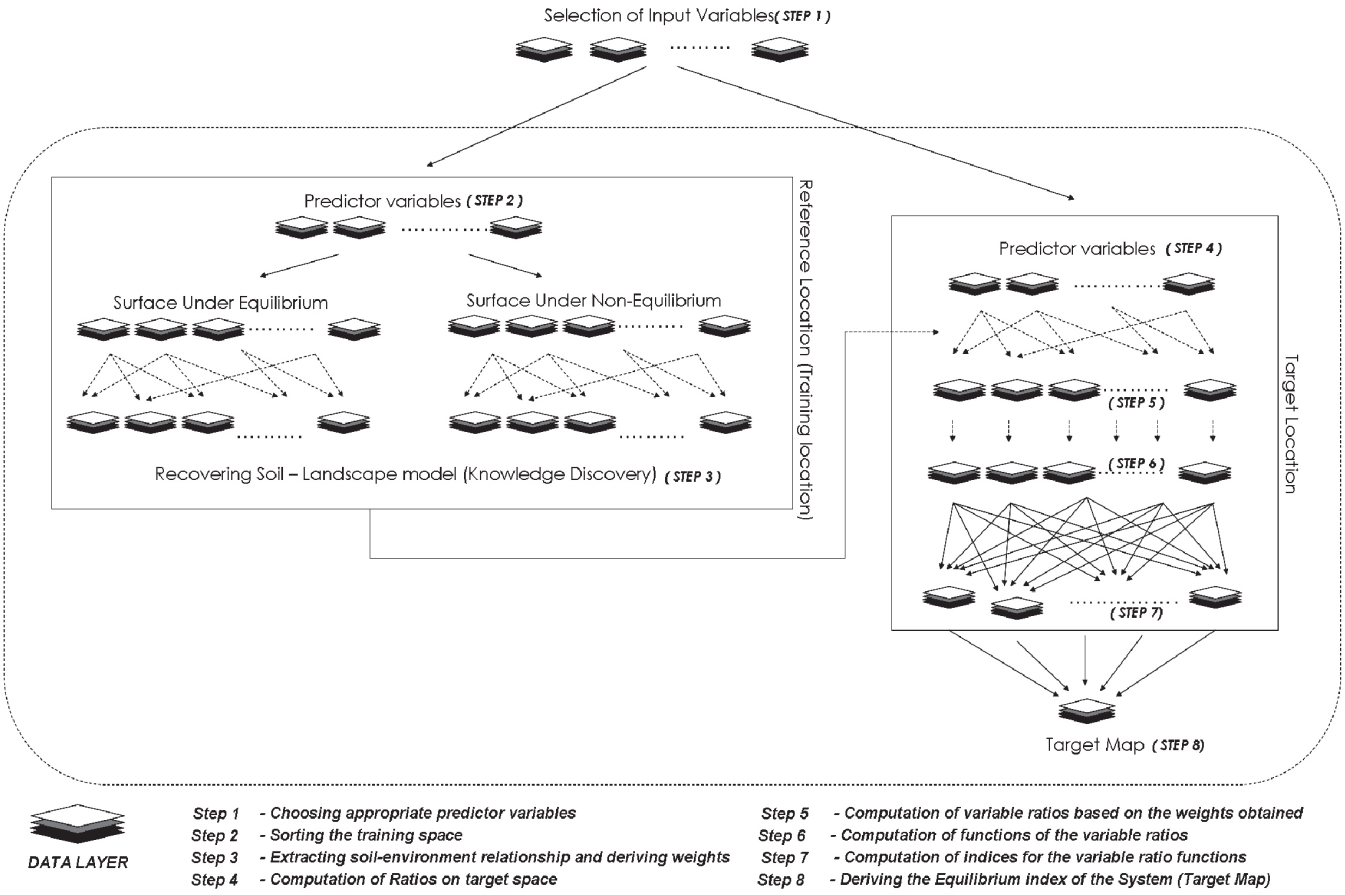

Figure 1: Scheme of the SEIMS network.

slight risk of soil erosion) and the non-reference spatial domain (e.g. areas under severe risk of soil erosion). Later, the rule on the choice of optimum ratio pairs was imposed on the target space.

After deriving the weights, function of the variable ratios ($f(r)$) can be computed as shown in Equation (1).

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

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

where r is the optimized ratios of each variable pair, $\bar{r}$ is the mean 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.

Computation of Indices for the Variable

The variable index is the average function of all the ratio pairs containing the corresponding variable and can be derived as in Equation (2).

$$I(i-(1,n)) = \left(\frac{\sum_{i-1}^{i-n} f(r_i)}{n}\right) \tag{2}$$

where n is the number of variables.

Deriving the Equilibrium Index of the System

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

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

Results and Discusssion

Data characterizing the soil erosion and its state factors are stored in a GIS database. The covariates used as predictors under the SEIMS network approach constitute the input data. Following this, a knowledge base is developed through spatial analysis on the data

components present in the GIS database. The optimized ratio expression of the covariates and its statistical derivates constitutes the knowledge base (Table 1) and are used as thresholds (weights) for predicting the soil erosion. The approach assumes that the ratios of the covariates are frequently a better indicator of the

Table 1: Thresholds derived for the knowledge base in SEIMS network

Factor ratio expressions	*Equilibrium domain*		*Non-equilibrium domain*		$\sigma^2_{mg} / \sigma^2_{nmg}$	*Choice of expression*
	$\bar{r}$	*cv*	$\bar{r}$	*cv*		
mtemp/meanrf	0.06	0.229	0.07	0.258	0.728	1
mpet / mtemp	1.61	0.012	1.60	0.019	0.440	-1
rfseason / mtemp	1.40	0.132	1.38	0.135	0.980	-1
mtemp / crust	20.1	0.184	19.3	0.185	1.071	1
erod / mtemp	0.07	0.122	0.07	0.144	0.679	-1
mtemp / oc	48.2	0.316	43.4	0.432	0.657	1
lcover / mtemp	0.80	0.224	0.84	0.271	0.631	-1
mtemp / alt	2.18	2.596	1.06	2.982	3.229	1
mtemp / slope	63.3	0.914	46.2	1.130	1.226	1
mpet / meanrf	0.10	0.227	0.11	0.262	0.700	-1
rfseason / meanrf	0.09	0.306	0.09	0.326	0.827	-1
crust / meanrf	0.003	0.308	0.004	0.303	0.889	-1
erod / meanrf	0.004	0.284	0.005	0.295	0.815	-1
meanrf /oc	776	0.347	672	0.474	0.713	1
lcover /meanrf	0.050	0.265	0.05	0.261	0.926	-1
meanrf /alt	40.8	2.886	18.0	3.426	3.624	1
meanrf /slope	1041	0.965	694	1.174	1.521	1
mpet / rfseason	1.17	0.125	1.18	0.125	0.984	1
Mpet / crust	32.5	0.186	30.9	0.184	1.129	1
erod /mpet	0.042	0.124	0.04	0.143	0.699	-1
mpet /oc	77.7	0.316	69.9	0.436	0.650	1
lcover /mpet	0.50	0.228	0.52	0.285	0.577	-1
Mpet /alt	3.54	2.595	1.72	2.988	3.210	1
Mpet /slope	103	0.916	74.8	1.134	1.225	1
rfseason / crust	28.2	0.256	26.7	0.231	1.373	1
erod / rfseason	0.05	0.187	0.052	0.192	0.890	-1
rfseason / oc	66.8	0.321	60.0	0.448	0.635	1
lcover / rfseason	0.59	0.269	0.62	0.325	0.615	-1
rfseason / alt	3.46	2.787	1.66	3.212	3.284	1
rfseason / slope	89.8	0.948	65.5	1.175	1.226	1
erod / crust	1.35	0.050	1.32	0.065	0.622	-1
crust / oc	2.50	0.372	2.31	0.453	0.792	1
lcover / crust	16.4	0.334	16.3	0.354	0.901	-1
crust / alt	0.09	2.570	0.05	2.934	2.363	1
crust / slope	3.20	0.928	2.48	1.161	1.055	1
erod / oc	3.33	0.355	3.06	0.447	0.747	1
lcover / erod	12.1	0.299	12.4	0.391	0.557	-1
erod / alt	0.13	2.575	0.07	2.943	2.517	1
erod / slope	4.28	0.919	3.28	1.150	1.089	1
lcover / oc	38.1	0.358	34.5	0.461	0.735	-1
oc / alt	0.06	3.028	0.03	3.493	3.505	1
oc / slope	1.50	1.116	1.13	1.282	1.352	1
lcover / alt	1.67	2.466	0.78	2.841	3.503	1
lcover / slope	49.8	0.958	34.7	1.145	1.442	1
slope / alt	0.05	2.506	0.03	2.397	2.203	-1

mtemp – mean annual temperature, meanrf – mean annual rainfall, mpet – mean annual PET, rfseason – seasonality of rainfall, crust– soil crusting, erod – soil erodibility, boc – topsoil organic carbon lcover – percentage land cover, alt – Altitute, slope – slope percentage.

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 covariates with dominant influence for a specific spatial location are in vicinity to the ideal or optimum conditions.

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 is introduced in the training spatial domain. 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. In the identified reference space that is 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 that of 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.

An inference engine is constructed linking the knowledgebase and the GIS database to derive the variable indices through the function of the optimized ratios. 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. Each variable index describes its effect on the balance of the system and 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 denotes that the system tends to be reaching an equilibrium condition leading to state of balance.

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 predictor should be zero, which means that the system is in better balance/equilibrium state.

Soil Erosion Index (Equilibrium Index of the System)

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. The EIS value was used to derive the soil erosion map (Figure 2).

The thresholds for the Equilibrium Index of the System, which is to be called as "Erosion Index" in the context of this study were ascertained based on its match with the broader classes of the erosion map provided by NBSS & LUP. 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 11,000 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 were in accordance with the source map. 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

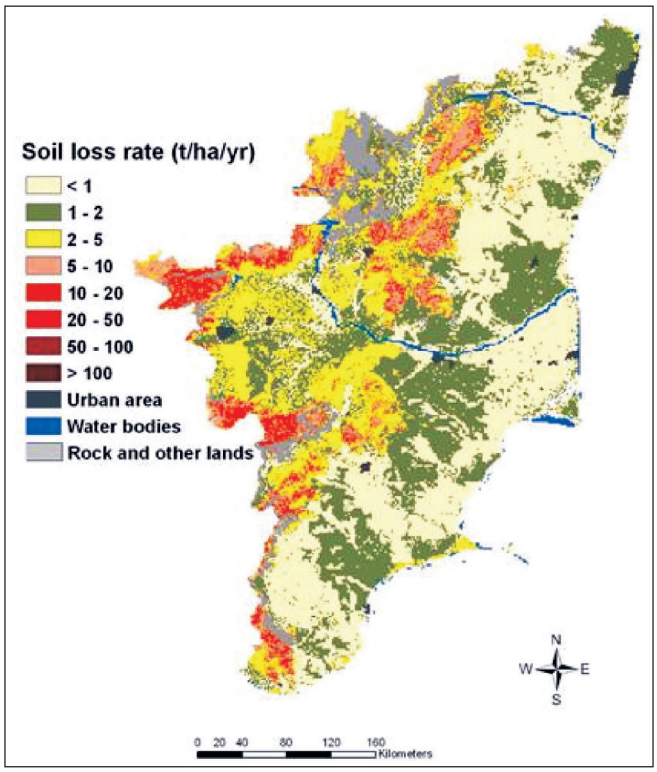

Figure 2: Soil erosion map of Tamil Nadu region based on EIS values of SEIMS network modelling approach.

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 (Figure 2).

The published soil erosion map (NBSS & LUP) used for the study was prepared at 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 modelling purposes. But the challenging task is how far can we find a way to transform this valuable information 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 modelling. 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-environmental variables and their interaction among themselves in relation to soil erosion process was achieved through the creation of a new soil erosion index map. This method also provides more insight on the order in which variables play significant role based on its relative influence on process of soil erosion. We feel that knowledge discovery from data mining of inherent subjective information in the soil erosion map has been demonstrated substantively.

References

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.

Hijman, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. and A. Javis (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 Javis (CIAT and with Karen Richardson (Rainforest).

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.

Palmer, W.C. and A.V. Havens (1958). A graphical technique for determining evapotranspiration by the Thornthwaite method. *Monthly Weather Review*, **86:** 123-128.

Penman, H.L. (1948). Natural evaporation from open water, bare soil and grass. *Proc. Roy. Soc. London*, **A(194):** S. 120-145.

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. EUR22545EN. Office for Official Publications of the European Communities, Luxembourg. ISBN 92-79-03878-8.

Soil Survey Staff (1999). Soil Taxonomy: A Basic System of Soil Classification for Making and Interpreting Soil Surveys. U.S. Natural Resources Conservation Service. USDA Handbook #436. Washington D.C.

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

Thornthwaite, C.W. (1948). An approach toward a rational classification of climate. *Geographic Review*, **38:** 55-94.

Townshend, J.R., DeFries, R., Hansen, M., Sohlberg, R., Carroll, M. and C. DiMiceli (2001). MODIS 32-Day Composites. College Park, Maryland: The Global Land Cover Facility.

UEA/CRU Report. October 1990. An Empirically Derived Adjustment Factor for Annual Thornthwaite PET Estimates Supplied Under Phase II.

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/