U of I NRES White Paper

Department of Natural Resources and Environmental Sciences

Error and Uncertainty Analysis for Ecological Modeling and Simulation

SERDP PROJECT CS1096 FINAL REPORT

by George Gertner

December 2001

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
DEC 2001		**00-00-2001 to 00-00-2001**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Error and Uncertainty Analysis for Ecological Modeling and Simulation	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Illinois Champaign,Department of Natural Resources and Environmental Sciences (NRES),Urbana,IL,91801**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**153**	

Foreword

This study was conducted for the Strategic Environmental Research and Development Program (SERDP) Office, CS-1096, "Error and Uncertainty for Ecological Modeling and Simulation." The technical monitor was Dr. Robert W. Holst, Compliance and Conservation Program Manager, SERDP. Mr. Bradley P. Smith is the Executive Director, SERDP.

The work was performed by the Department of Natural Resources and Environmental Sciences (NRES) at the University of Illinois, Champaign-Urbana. The NRES Principal Investigator was Professor George Gertner.

The study was done in close collaboration with the Ecological Processes Branch (CN-N) of the Installations Division (CN), Construction Engineering Research Laboratory (CERL). The CERL point of contact was Mr. Alan B. Anderson. Mr. Alan Anderson was CERL Principal Investigator for SERDP Project CS-1102, "Improved Units of Measure for Training and Testing Area Carrying Capacity." Much of error and uncertainty work was closely linked to this project. Mr. Steve Hodapp is Chief, CEERD-CN-N, and Dr. John T. Bandy is Chief, CEERD-CN. The associated Technical Director is Dr. William D. Severinghaus, CEERD-TD. The Acting Director of CERL is Mr. William Goran.

CERL is an element of the U.S. Army Engineer Research and Development Center (ERDC), U.S. Army Corps of Engineers. The Director of ERDC is Dr. James R. Houston and the Commander is COL James S. Weller, EN.

Research Team

Alan Anderson	Projects manager at the Engineering Research and Development Center, USACERL, Champaign, IL
Shoufan Fang	Research Programmer
George Gertner	Principal Investigator, Professor of Biometrics in the Department of Natural Resources at the University of Illinois at Champaign-Urbana
Burak Guneralp	Research Assistant
Xianzhong Liu	Research Assistant
Pablo Parysow	Team Leader 1998
Svetlana Shinkareva	Research Assistant
Vivek Singh	Research Assistant
Guangxing Wang	Postdoctoral 1998 and Team Leader 1999-2001
Stephen Wente	Researcher
Xiangyun Xiao	Research Assistant

Collaborators

The collaborators include USACERL (US Army Construction Engineering Research Laboratories), Fort Hood Environmental Office, Geographic Modeling Systems (GMS) Laboratory and the National Center for Supercomputing Applications (NCSA) at the University of Illinois, etc. USACERL provided very strong support of this study for technical discussion, case study area and data sets. The following collaborators were particularly helpful:

- Scott Tweddale (Remote Sensing)
- David Price (Plant ecologist, now at WES)
- Jim Westervelt (Environmental Modeler)
- Chris Rewerts (Ecologist)
- Robert Melton (Statistician)
- Bruce MacAllister (Ecologist)
- William Jackson (Remote Sensing Specialist)
- William Seybold (former USACERL, LCTA Database Manager)
- David Mather (Remote Sensing Specialist)
- Pat Guertin (Natural Resource Analyst)

At Fort Hood, Emmet Gray, Chief of the Environmental Office, and Don Jones, Soil Conservationist in the Environmental Office, were always supportive at our case study area. We had a commitment from Dr. Douglas Johnson, Director of the Geographic Modeling Systems (GMS) Laboratory at the University of Illinois, to utilize his laboratory for the duration of the project. This laboratory is supported by USACERL, University of Illinois, and the National Center for Supercomputing Applications (NCSA). In addition, the NCSA provided time on their super computers located on the University of Illinois campus. Finally, David Tazik, Chief, Natural Resources Division, Waterways Experiment Station, found time in his busy schedule to extend our uncertainty work to threatened and endangered bird species.

Contents

INTRODUCTION

SERDP relevance and project initiative

The Strategic Environmental Research and Development Program (SERDP) was initiated in 1990 to harness the resources of the defense establishment to minimize or remove any negative environmental impacts associated with Department of Defense's (DoD) primary mission of maintaining military readiness for national defense. SERDP is a cooperative program under the DoD in full partnership with the Department of Energy and the Environmental Protection Agency, and with participation by numerous other Federal and non-Federal organizations. SERDP consists of environmental compliance, cleanup, pollution reduction, and conservation programs. Its objectives are to accelerate cost-effective clean up of contaminated defense sites, facilitate full compliance with environmental laws and regulations, enhance training, testing, and operational readiness through prudent conservation measures, and reduce defense industrial waste streams through aggressive pollution prevention. Application of the innovative environmental technologies developed by SERDP should reduce the costs of sustainable environmental and resource management, save the time required to resolve environmental problems, and enhance safety and health.

The conservation program of SERDP focuses on research and development that helps to manage natural and cultural resources for sustained access and uses of land, water, and airspace while protecting wildlife, endangered and threatened species. The objectives are to provide new methods, techniques, and tools to efficiently and effectively inventory, map and manage these resources, including assessment of impacts from military testing and training, design of plans to restore the resources, etc.

Many models have been developed and are being widely used to predict the state of natural and cultural resources. These models are used to formally describe and scientifically understand the underlying mechanisms and spatial relationships that produce the state of a resource and, therefore, provide a basis for extrapolation. Thus, it is possible to use these models to predict the behavior of a system under a wide range of scenarios including scenarios that have never occurred. This characteristic allows us to analyze the potential effect of individual as well as the cumulative effects of a combination of factors on the behavior of the systems under consideration. Natural and cultural resource models are also

being extensively used to provide management guidelines, and thus, are becoming powerful decision-making tools as well.

Additionally in the past ten years, Geographic Information Systems (GIS) have become powerful tools for natural resource management. Using GIS, decisions can be made from digital maps on which spatial patterns, distributions, processes and relationships are clearly visualized and easily updated. (This contrasts with the more traditional approach in which decisions are made from spatially aggregated and infrequently updated information.) Likewise, remotely sensed data such as aerial photos and satellite images has become more important as a method of generating and updating natural resource maps.

If these maps are considered to be results of interpolation from sample data and prediction by traditional models, the maps can be regarded as site-specific spatial models with the traditional models as their core. For example, a map of soil erosion can be generated by interpolation from soil loss estimates at sample field plots with estimates calculated as a product of empirical models related to rainfall-runoff, soil properties, slope steepness, slope length, vegetation cover and management, and management practice factor. Thus, the empirical (traditional) models are essential to the spatial model of soil erosion.

Model and map users often implicitly assume that the values that characterize model entities are true or error-free. This is usually known as the deterministic assumption. However, most values employed in traditional and spatial simulation modeling are estimates of the true parameters and, therefore, have an associated uncertainty. This uncertainty can be due to non-sampling errors such as measurement errors, sampling errors, prediction errors, expert knowledge uncertainty, etc. Obviously, when there is uncertainty in the inputs to a system there must be uncertainty in the predictions as well (Figure 1). Moreover, the sensitivity of predictions to these uncertainties can vary considerable in both time and space.

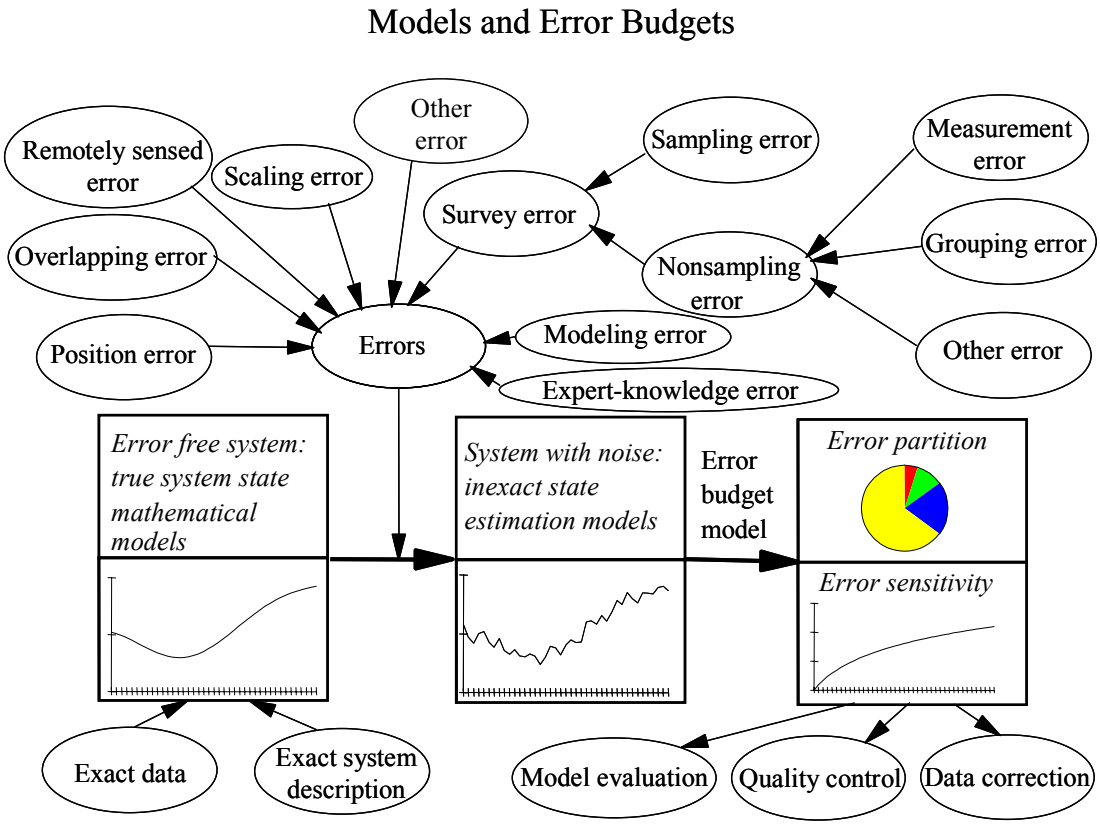

Figure 1. An example of a modeling system with error.

Assessing the quality of simulation systems is a difficult task. This is particularly true for multi-component systems, whose prediction quality is determined not only by its components, but also by the interactions of those components and by the inputs from the monitoring system. Because the components are linked together, interactions between them will produce properties that did not previously exist. In the simulation system, the outputs from one component are used as the inputs for other components. Errors from individual components propagate and accumulate throughout the entire simulation system. The effects of such errors will be evident in the final outputs.

Moreover, the use of digital maps in management expands the sources of errors, while assessing errors has become more complicated. For example, position errors need to be identified and quantified, and their effects on attribute errors have to be assessed. Secondly, errors occur when interpolating sample observations to unknown locations. Because appropriate map unit sizes or spatial resolutions may differ greatly for different system variables, thirdly, the maps of different natural resources have to be inferred from one

resolution to another, which is scaling that results in uncertainty. Additionally, the use of remotely sensed data to improve the accuracy of maps may also lead to new errors due to sensor systems, platforms, weather, geometric errors, etc. Finally, spatial information from nearby locations is usually used to improve predictions at unknown locations, therefore, the data configuration effect needs to be assessed. All these error sources will lead to spatial variability of accuracy and uncertainty. That is, accuracy of a map will vary over space and the main error source will differ from place to place. Therefore, spatial uncertainty analysis has become necessary, which has made it very complicated to assess the quality of simulation systems.

Error budgets can be used to assess the quality of the overall simulation system (Gertner and Guan 1991). An error budget can be considered as a catalog of the different error sources (Gelb et al. 1974) that allows the partitioning of the projection variance and bias according to their origins (Table 1). As a specialized form of sensitivity analysis, an error budget shows the effects of individual errors and groups of errors on the quality of a multi-component model's predictions. The goal in developing the error budget is to account for all major sources of errors that can be expected in a system. By doing this, the sources of errors can be examined and partitioned in different ways. Additionally, an error budget can be generated for different time steps and spatial scales.

Because of the way an error budget is generated, the components that cause the most uncertainty can be readily identified. These components will be the ones that contribute the most toward final prediction variance and/or bias. Additionally, if the model is modified, the newly created uncertainty contributions can be assessed quickly. More important is that accounting for uncertainty has management implications. For example, management decisions can be made after taking into account the uncertainty of the information on which the decision is based.

Taking into account the growing importance of simulation modeling in resource assessment and management, the need for a comprehensive framework for analyzing uncertainty of simulation results is apparent. Although progress has been made in the areas of uncertainty analysis (e.g., Dale et al. 1988; Gardner and O'Neill 1981; Gertner and Guan 1991; Gertner et al. 1995; Hanes et al. 1991; Kremer 1981; McCarthy et al. 1995; O'Neill and Gardner 1979; O'Neill et al. 1980; Rossing et al. 1994a,b; Summers et al. 1993) and error budgets (Gelb et al. 1974; Gertner and Guan 1991; Gertner et al. 1995), it is necessary to develop the statistical and computational tools that will enable model users to jointly assess and quantify the sources and magnitude of input error, develop error budgets, and optimize data collection, modeling and simulation, and management decisions in terms of errors, expense and risks for the array of large scale simulation models employed in resource assessment and management.

Furthermore, the need for spatial error budgets requires maps of estimates for natural resources and their variance maps as well. Traditional methods of creating maps by interpolating sample data to unknown locations, for example supervised and unsupervised classification (Campbell, 1996, Wang et al., 1997) and even various kriging methods (Goovaerts, 1997), may not produce the information necessary for spatial uncertainty analysis. New methods need to be developed that provide population and local unbiased estimates and their variances and co-variances as uncertainty and spatial correlation measures when interest variables are spatially correlated with each other. Therefore, there is a very strong need to develop a systematic methodology and tool to generate unbiased maps with uncertainty measures, and further to make spatial error budgets. Therefore, in 1998 this project 'Error and Uncertainty Analysis for Ecological Modeling and Simulation' was initiated.

Table 1. A schematic representation of an error budget for final prediction variance and bias of a hypothetical multi-component monitoring-simulation system. Both final variance and bias are partitioned according to the sources of errors.

Sources of Errors	Variance of Final Prediction	Bias of Final Prediction
Input Measurement Error variable 1 variable 2 … variable n Subtotal		
Sampling Error Subtotal		
Component Model Error Component 1 equation 1 equation 2 … Component 2 … Component n equation 1 equation 2 … Subtotal		
…		
Digitizing errors Data conversion error Remotely sensed error Interpolation error … Scaling error Subtotal		
Temporal Error Subtotal		
 Grand Total		

Project objectives

This project intends to overcome current significant gaps in the generation and use of models and maps for the assessment and management of natural and cultural resources. Specifically, this study will account for spatial effect of different sources of error on uncertainty of predictions and maps generated through models, and also provide the rationale for efficiently reducing uncertainty and error for data collection and spatial prediction, and further reducing risks of poor management decisions being made. This methodology will be relevant to all users of natural, ecological and environmental modeling systems. The proposed analytical framework will be made available as a user-friendly interactive software package. This package will be fully compatible with the computational environments employed by SERDP members. It is expected that this project will provide users with the means not only to assess but also to exert control over the quality of simulation results. This, in turn, will provide the necessary quality control/quality assurance mechanisms to support decision-making regarding natural and cultural resources. The technical objectives thus include:

a) Providing a rationale to account for spatial effect of different sources of uncertainty in temporal-spatial models and maps employed in the assessment and management of natural and cultural resources.

b) Presenting a theoretical and methodological framework for optimizing sampling design, data collection, spatially modeling, and management in terms of precision (errors) and/or expense as an integral part of the continuous monitoring-simulation process.

c) Developing user-friendly portable software (tool kit) that can be used for spatial uncertainty analysis of simulation modeling systems in general.

d) Illustrating this methodology through a case study in which a soil erosion modeling system is being applied by the military for assessment and/or management of resources at one military installation.

Project methodology summary

The methodology proposed in this work is a continuation and improvement of a research program initiated by George Gertner more than a decade ago (e.g., Gertner 1987, 1991; and Gertner et al. 1996). The overall goal of this study is to account for the sources and the effect of spatial uncertainty in simulation modeling. Thus, we plan to employ some of the analytical

tools developed so far, and also build upon the previous work to meet the goals established in this proposal.

We have developed a GIS-based methodology to make spatial and temporal predictions, analyze uncertainty, and build error budgets. This methodology is based on modeling spatial variability of variables and spatial cross variability between them. The geostatistical methods – various sequential simulation and co-simulation algorithms are developed and used for generating prediction, variance and co-variance maps from sample data sets. Various and co-located available auxiliary data including digital elevation models and remotely sensed images are introduced into the algorithms to improve spatial simulation accuracy. The algorithms result in a grid-based database containing various maps of natural resources and their uncertainty measures. The spatial and temporal predictions are made at different optimal operational scales. Based on the maps of estimates, variances and co-variances, spatial uncertainty analyses and error budgets can be produced using the uncertainty analysis methods obtained by improvement of the existing methods that include Taylor series, Fourier Amplitude Sensitivity Test (FAST), regression modeling, etc. That is, the error budget is developed on the basis of pixel by pixel in addition to populations and homogeneous areas. Moreover, the variables themselves, the interactions between these variables, and the effect of spatial information from neighbors are taken into account in the error budgets.

As a case study, we applied the proposed methodology to a soil erosion prediction system – Revised Universal Soil Loss Equation (RUSLE) (Renard et al., 1997) employed by the military for assessment and/or management of land capacity with training activities at one military installation – Fort Hood, Texas. The case study was done in parallel with the methodology development above.

Project performance and achievement summary

This project started in Jan. 1998 and ended in Dec. 2001. The project performance can be divided into four stages corresponding to four research years. The performance stages, research years, and corresponding tasks follow:

The first stage - Year 1998:

SELECTED A MONITORING-MODELING SYSTEM – THE REVISED UNIVERSAL SOIL LOSS EQUATION (RUSLE) AS A CASE STUDY, AND THE INSTALLATION – FORT HOOD, TEXAS, AS THE CASE STUDY AREA.

a) Carried out relevant literature and study review, and started development of methodological and theoretical foundation for sampling design, spatial modeling and simulation, identification and definition of errors, uncertainty assessment, and rational of reducing errors by evaluating existing methods and developing new approaches.

b) Reviewed the existing database for the case study and complemented sampling and ground data collection.

The second stage - Year 1999:

a) Finished the calibration and improvement of existing models for the case study, and completed new models.

b) Completed the design of methodological framework for sampling design, spatial modeling and simulation, identification and definition of errors, uncertainty assessment, and rational of reducing errors.

c) Applied the methods to the case study for generating soil erosion factor maps including rainfall-runoff erosivity, soil erodibility, slope steepness, slope length, vegetation cover and management, and support practice (These factors will be described in the next chapter).

d) Identified and defined all possible source errors, and started spatial and temporal uncertainty analysis in the case study area.

The third stage - Year 2000:

a) Completed the methodological framework and its details, and continued the applications of the methods to the case study for spatial and temporal modeling, map generation, and error budgets of soil erosion at different scales in both space and time.

b) Started designing the computer software for realizing and generalizing the methodology.

The fourth stage - Year 2001:

a) Completed the case study for applications of the methodology, and generated declaration of quality for the monitoring-modeling system.

b) Defined general quality control/quality assurance standards for data collection, spatial modeling and simulation, and resource management, and suggested guidelines for error management.

c) Finished the software programming.

d) Documented the methodology, its application results to the case study, and computer software.

Main achievements:

a) A general methodology consisting of the methods to optimize sampling design and data collection, to spatially and temporally model and predict natural resources, that is, to generate maps and their time series, to define and identify various errors, and to do spatial error budgets.

b) A user-friendly software consisting of programs that can be used to carry out error budgets at different levels such as populations, homogeneous areas, and pixel by pixel.

c) A rational to account for spatial effect of different sources of uncertainty in temporal-spatial models and maps employed in the assessment and management of natural resources.

d) One project report, a software user manual, more than 20 peer-reviewed journal articles, and more than 15 conference and technique reports.

e) Many technical breakthroughs, and interesting and important findings, for example, development of new methods and improvement of existing methods to determine appropriate plot size and spatial resolution, model loss of spatial information due to scaling, jointly map multiple variables that are spatially correlated with each other, generate error budgets considering interactions among multiple variables and effect of spatial information from neighbors.

CASE STUDY

ATTACC and ELVS

We applied the proposed methodology to the Army Training and Testing Area Carrying Capacity model (ATTACC) (Anderson et al., 1996) at one military installation as a case study. The military uses this model for the assessment and management of natural and cultural resources. Specifically, ATTACC is an analytical tool used to determine training carrying capacity and evaluate the impact of alternative training exercise scenarios based on the Evaluation of Land Value Study (ELVS) methodology (Siegel et al., 1996). The case study was done in parallel with the uncertainty analysis methodology development.

The ELVS was designed to develop and demonstrate a methodology to estimate and analyze resource requirements for training land management, and to provide operation and support costs of land rehabilitation and management (LRAM) accounting for environmental, training, and economic factors. In the ELVS methodology, soil erosion status is used as a quantitative measure of land condition and training land carrying capacity. Training land carrying capacity refers to the ability of specific land parcels to accommodate training and mission activities. Since soil erosion is the primary effect of using land for training, soil erosion status is assumed to be a good indicator of land condition. Erosion incorporates most of the factors that influence land condition and is directly related to vegetation cover, indirectly to habitat for threatened and endangered species and therefore ultimately, to biodiversity. The ELVS methodology is realized by building relationships between soil erosion status and training land carrying capacity. The model used to predict soil erosion status is the Universal Soil Loss Equation (USLE) (Wischmeier and Smith, 1978) and Revised USLE (RUSLE) (Renard et al., 1997). The monitoring system employed is the Army Land Condition Trend Analysis (LCTA) (Tazik et al., 1992).

USLE or RUSLE and uncertainty

In the USA, soil erosion is usually predicted using the Universal Soil Loss Equation (USLE) (Wischmeier and Smith, 1978) or the Revised USLE (RUSLE) (Renard et al., 1997). In both

equations, soil loss (A) is a function of six input factors including rainfall-runoff erosivity (R), soil erodibility (K), slope length (L), slope steepness (S), vegetation cover and management (C), and support practice (P):

$$A = R \times K \times L \times S \times C \times P \tag{2.1}$$

Soil loss (A) is a computed spatial-temporal average soil loss per unit of area and can be expressed in the units selected for factors R and K, for example, in a unit of ton / ha, year. The SI metric unit can be converted to US customary unit, i.e., ton / (acre • year) by multiplying by $\frac{1}{2.242}$. Generally, soil loss is most sensitive to the topographical factor LS (a product of slope steepness S and slope length L), and then C factor (Benkobi et al., 1994; Biesemans et al., 2000; Renard and Ferreira 1993; Risse et al. 1993). Erosion increases as slope length and steepness increases, and it increases more rapidly with slope steepness than slope length. The higher the ground and vegetation cover, the less the potential soil loss. Soil loss is also proportional to the R factor when other factors are held constant.

For each specific soil, furthermore, a tolerance value indicating a maximum soil erosion level for sustainable soil productivity has been derived for agricultural management. The ratio of estimated soil loss (A) to its tolerance (T) is called the erosion status (ES) (dimensionless) of the soil.

$$ES = A/T \tag{2.2}$$

Four levels of erosion status are defined: ES < 1.0; 1.0 ¹ ES < 1.5; 1.5 ¹ ES < 2; and ES Ä 2.0. Higher ES values reflect a poorer land condition (e.g., ES greater than 2.0), whereas lower ES values reflect a better land condition (e.g., ES less than 1.0).

Since training results in vegetative cover disturbance that increases soil loss, training carrying capacity is limited by soil loss tolerance according to the following relationship:

Predicted land condition = Current land condition + Change in land condition due to training load - Change in land condition due to land recovery

This relationship can be expressed in the notation of Eqs. 2.1 and 2.2 as:

$$ES=A/T=(R*K*LS*P*((C_t-C_u)*IA/TA-(C_t-C_u)/M+C))/T \quad (2.3)$$

where

C_t = vegetation cover and management factor after disturbance

C_u = vegetation cover and management factor before disturbance

IA = Impact area

TA = total area suitable for training.

M = time required for the land to naturally recover.

Once the relationship between intensity of military training and disturbance of vegetation cover is derived, Eq. 2.3 is used to predict spatial and temporal average soil erosion status for a given area after military training. Additionally by selecting a maximum allowable soil loss (e.g. ES = 1), the maximum allowable disturbance of vegetation cover and thus, the maximum allowable intensity of training, can be calculated.

Rainfall-runoff factor R

The rainfall-runoff erosivity factor R is the rainfall erosion index plus a factor for any significant runoff from snowmelt. Rainfall and runoff normally lead to soil loss. This factor is highly correlated with the product of the total storm energy and the maximum 30-minute intensity. A rainfall erosion index was derived from data by Wischmeier (1959), and Wischmeier and Smith (1958). The annual R is a sum of erosivity index values for all rain-showers in one year and is usually expressed in unit MJ Ê mm / ha Ê h Ê y, converted to US customary unit - hundreds of foot • tonf • inch / acre • h • y by multiplying by $\frac{1}{17.02}$. The larger the R factor, the higher the potential annual soil loss.

Isoerodent maps have been developed by Wischmeier (1959), and Wischmeier and Smith (1958, 1978), and widely used to obtain the R factor for a specific area by linear interpolation. This method implies the rainfall-runoff erosivity R factor is linear over space and constant over time. As suggested by McGregor et al. (1980), however, these assumptions may not be true. Although a variable R factor over space can be derived by linear interpolation, a constant value for a specific area is usually implied. This may result in a smoothed spatial prediction and leave this source of uncertainty unaccounted. The uncertainty of the R factor values estimated from the isoerodent maps is unknown. Therefore, new maps with uncertainty measures were developed as part of this project.

Where rain gauge data are available, the values of the rainfall-runoff erosivity R factor can be calculated for each rainfall station. If a rainstorm implies that there is a period of 6 hours with less than 1.27 cm of rain, a rainfall erosion index (EI_{30}) of the rainstorm is obtained by multiplying total storm energy (E) with the maximum 30-minute intensity (I_{30}) (Wischmeier, 1959; Wischmeier and Smith, 1958, 1978). Different empirical equations have been developed and used to calculate the unit energy contained in the volume of rain (brown and Foster, 1987; Foster et al., 1981). In this project, we used the following equation developed by a research team headed by Steven Hollinger at the Illinois State Water Survey, Atmospheric Environmental Section.

$$e = 0.29\left[1 - 0.72\exp(-0.082i)\right] \tag{2.4}$$

where e is the kinetic energy (MJ ha^{-1} mm^{-1}) and i is the shower intensity (mm h^{-1}). The annual R factor is the sum of the erosion index values for all rainstorms in one year. In an N year period, the R factor (MJ mm $ha^{-1}h^{-1}y^{-1}$) is calculated as follows:

$$R = \frac{\sum_{i}^{j}(EI_{30})_i}{N} \tag{2.5}$$

where $(EI_{30})_i$ is the erosion index EI_{30} for storm i, and j is the number of storms in the N year period. In addition to the annual R factor, seasonal and half-month average values of the rainfall-runoff erosivity R factor can be computed.

Soil erodibility factor K

The soil erodibility factor (K) is the soil loss rate per erosion index unit for a specific soil as measured on a standard plot defined as a 22.1 m or 72.6 ft length of uniform 9 % slope in continuous clean-tilled fallow. It is expressed in SI metric unit t、ha、h / ha、MJ、mm, and can be converted to US customary unit ton • acre • hour / hundreds of acre • foot • tonf • inch by multiplying by $\frac{1}{0.1317}$.

The soil erodibility factor (K) measures the contribution of soil intrinsic properties to soil erosion. For major soil types and soil texture classes in the United States, the values of soil erodibility factor (K) have been published and can be obtained from the USDA- Natural Resources Conservation Service (NRCS) (SWCS, 1995; Wischmeier and Smith, 1978). Each soil type corresponds with a published soil erodibility value. The published values from USDA-NRCS are the average values within the soil types when the data were collected and are assumed to be constant over time. However, heterogeneity of soil in time and in space

tends to support the concept that soil erodibility depends dynamically and spatially on the properties of a specific soil.

The main factors considered in the practical calculation of soil erodibility include soil sand %, silt %, organic matter %, structure, and permeability. By sampling, collecting and measuring soil samples, the soil erodibility factor (K) values of soil samples can be calculated using the following formula (Wischmeier and Smith, 1978):

$$K = \frac{2.1 \bullet 10^{-4} (12 - OM) \bullet M^{1.14} + 3.25 \bullet (S - 2) + 2.5 \bullet (P-3)}{7.59 \bullet 100} \quad (2.6)$$

where OM is soil organic matter, M is (%silt + %very fine sand) (100 -%clay), S is soil structure code and P is permeability class. If soil organic matter content is greater or equal to 4%, OM is considered constant at 4%. Moreover, the influence of rock fragments on soil loss is accounted for by a subsurface component in the soil erodibility K factor (Renard et al. 1997). The soil profile descriptions with permeability classes for all the soil samples in this study included the effect of rock fragments on permeability. The soil erodibility (K) factor and the subsurface component for effect of rock fragments were explained via an adjustment for permeability classes.

Because of the underlying forces shaping soils, soil properties vary with time and space and are affected by climate, organisms, topography and parent materials interacting with time (Jenny, 1941). Climate factors (temperature and rainfall) affect soils as well as the plants growing on those soils. Plant community succession due to the change of the soil physical environment is well observed and change in plant composition in turn affects the soil properties. The soil properties vary also in space because of the variation of soil formation factors. Thus, a soil erodibility value for a specific soil may vary temporally and spatially. Using the soil erodibility values obtained previously from an extensive database for a specific area may lead to uncertainty. Therefore, it is necessary to include the uncertainty associated with soil erodibility into the overall uncertainty analysis of soil loss and to improve methods for mapping the soil loss.

Topographical factor LS

Slope length factor (L) is the ratio between soil loss from the field slope length and soil loss from a slope that has a length of 22.13 meters or 72.6 ft, where all other conditions are the same. Slope steepness factor (S) is the ratio of soil loss from the field slope gradient to soil loss from a 9% slope under otherwise identical conditions. The product of slope length (L) and steepness (S), called topographical factor (LS) (dimensionless), accounts for the effect of topography on erosion in both USLE and RUSLE. Among all input factors, soil erosion is

most sensitive to the topographical factor (LS), and more sensitive to slope steepness than slope length (Benkobi et al. 1994, Renard and Ferreira 1993, Risse et al. 1993).

The slope steepness factor (S) is defined as a function of the slope angle measured in degrees and the slope length factor (L) as the function of slope length value in meters. A lot of studies have been done to derive equations for calculating factors S and L. Table 2.1 lists two sets of empirical models involved in the USLE and RUSLE, respectively, which can be used to calculate the slope length factor (L) and steepness factor (S) with the field measurements of slope length λ in meters and slope angle β in degrees (Foster et al. 1977, Moore and Wilson 1992, Renard et al. 1997, Wischmeier and Smith 1978).

Table 2.1 Empirical models for calculation of slope steepness factor (S) and slope length factor (L).

Model	S		L	
USLE	$S=65.4Sin^2\beta+4.56Sin\beta+0.0654$		$L>(m/22.13)^{0.5}$	when $Tan\beta>0.05$
			$L>(m/22.13)^{0.4}$	$0.03<Tan\beta<=0.05$
			$L>(m/22.13)^{0.3}$	$0.01<Tan\beta<=0.03$
			$L>(m/22.13)^{0.2}$	$Tan\beta\leq 0.01$
RUSLE	$S=10.8Sin\beta+0.03$	when $Tan\beta<0.09$	$L>(m/22.13)^{(F/(1+F))}$	
	$S=16.8Sin\beta-0.50$	$Tan\beta>=0.09$	where $F=(Sin\beta/0.0896)/(3Sin^{0.8}\beta+0.56)$ (assuming a moderate rill / interrill ratio);	
	$S=3Sin^{0.8}\beta+0.56$	$\lambda<=4m$		
	$S=(Sin\beta/0.0896)^{0.6}$	Thawing soils with $Tan\beta>=0.09$	or F=0 when there is deposition when $\lambda=4m$ to $\lambda<=4m$.	

When soil loss is estimated using a geographic information system (GIS) for large areas with converging and diverging terrain, the empirical models above cannot differentiate between those areas experiencing net erosion and net deposition. A physically based topographical factor (LS) equation has thus been developed based on a digital elevation model (DEM) (Moore and Burch, 1986; Moore and Wilson, 1992) as follows:

$$LS=\left[\frac{\text{Up\_area}}{22.13}\right]^{m}\left[\frac{\sin\beta}{0.0896}\right]^{n} \tag{2.7}$$

where m and n are constants equal to 0.6 and 1.3 respectively. β is the land surface slope in degrees, Up_area is the up-slope contributing area per unit width of cell spacing [m^2m^{-1}] from which the water flows into a given grid cell. The area Up_area for a given grid cell is calculated as follows (Mitášová et al., 1996):

$$\text{Up\_area}=\frac{n\times\mu\times a}{b} \tag{2.8}$$

where a is the area of a grid cell; n is the number of cells draining into the cell; μ is a weight depending on the runoff generation mechanism and infiltration rates; and b is the spatial resolution. If rainfall and infiltration are assumed to be uniform across the study area, the weight μ can be assumed to be one (Mitášová et al., 1996). Because a is constant for a specific resolution, $a=b\times b$. Thus $Up\_area=n\times b$. In practice, Up_area can be approximated by multiplying the down-slope flow-line density with the DEM spatial resolution. However, the precision for predicting the LS factor is related to the DEM accuracy, spatial and vertical resolution, and the methods to derive topographical variables related to LS. For example, Mitášová et al. (1996) investigated this approach by interpolating

DEMs to finer spatial resolutions and suggested that the commonly used 30m-spacing USGS DEMs are insufficient.

Vegetation cover and management factor C

The vegetation cover and management factor (C) is the ratio between soil loss from an area with specified cover and management and soil loss from an identical area in tilled continuous fallow. The C factor represents the effect of cropping and management practices in agricultural management, and the effect of ground, tree and grass covers on reducing soil loss in non-agricultural situations. Higher ground and vegetation covers result in less potential soil erosion, and vice versa. According to Benkobi et al. (1994) and Biesemans et al. (2000), the vegetation cover factor is one of the three factors (the others being slope steepness and length) to which soil loss is most sensitive.

In RUSLE (Renard et al., 1997), the C factor value for an area where conditions change rapidly over time is derived by weighting the soil loss ratio values for a given conditions by rainfall erosion index values. That is, an entire time period is divided into n time periods and for each of the n periods a soil loss ratio is calculated. Then, the soil loss ratio values are weighted by corresponding rainfall erosion index values. The soil loss ratio for the given conditions is a product of five sub-factors including the prior land use sub-factor, canopy cover sub-factor, surface cover sub-factor, surface roughness sub-factor, and soil moisture sub-factor. Each of the sub-factors contains cropping and management variables that affect soil erosion. Each sub-factor is an empirical function of one or more variables such as residue cover, canopy cover, canopy height, surface roughness, below ground biomass, prior cropping, soil moisture and time. The calculation of the C factor, thus, is very complicated.

In this project, we used the USLE method to calculate C factor. That is, the vegetation cover C factor is derived based on empirical diagrams that explain the relationship of the C factor with measurements of ground cover, aerial cover and minimum drip height (Wischmeier and Smith, 1978). Often the measurements of these variables are obtained by sampling subplots along transect lines. The average ground cover, aerial cover and minimum drip vegetation height are calculated for each plot (transect). However, because it would be difficult to perform automatic calculations with these empirical diagrams, we used the empirical equations developed by Bill Seybold of the U.S. Army Construction Engineering Research Laboratory (USACERL) to calculate C factor. These empirical equations (Table 2.2) describe the C factor as a function of ground cover, aerial cover and minimum drip height measurements under different ground and canopy cover conditions.

Table 2.2 Empirical models for calculating vegetation cover factor C (GC – ground cover, CC – canopy cover, VH – minimum drip vegetation height, EVH – effect of vegetation height, ECC – effect of canopy cover, C1 – effect of vegetation height and canopy cover, C2 – effect of ground cover).

Empirical equation	Conditions
Vegetation height and canopy effect	
EVH = exp(4.574 – (0.056*ln(VH)) + (0.366*VH))	VH >= 0.1
EVH = exp(4.574 – (0.056*ln(0.1)) + (0.366*0.1))	0 < VH < 0.1
EVH = exp(0.000001)	VH < 0
EVH = -1	VH = 0
ECC = CC – (CC * GC / 100)	GC > 0 and CC => 0
ECC = CC	GC = 0 and CC => 0
ECC = -1	Otherwise
C1 = 1 – (ECC / EVH)	ECC >=0 and EVH >0
C1 = -1	Otherwise
Ground cover effect	
C2 = 0.734 – (0.0139*GC) + (0.0000665*(GC^2))	GC = 90
C2 = 0.625 – (0.0124*GC) + (0.0000635*(GC^2))	80 <= GC < 90
C2 = 0.312 – (0.0049*GC) + (0.0000187*(GC^2))	51 <= GC < 80
C2 = 0.362 - (0.00745*GC) + (0.0000492*(GC^2))	41 <= GC < 51
C2 = 0.313 – (0.00431*GC)	30 <= GC < 41
C2 = 0.358 - (0.0058*GC)	20 <= GC < 30
C2 = 0.45 - (0.0151*GC) + (0.000234*(GC^2))	0 <= GC < 20
C2 = 0	Otherwise
C factor	
C = C1*C2	C1 >= 0 and C2 >= 0
C = -1	Otherwise

The values of the C factor at the non-sample locations are usually estimated by spatial interpolation of the C factor values at the sampling locations. In order to provide accurate maps of soil loss, it is important to create a reliable map of vegetation cover and management factor C. The traditional method widely used for the spatial interpolation of the C factor is the so called point-in-polygon or point-in-stratum (Warren and Bagley, 1992). Within each polygon or stratum the cells are assumed to be homogeneous and an average is calculated and

assigned to each cell. The polygons or strata are derived by supervised or unsupervised classification of all pixels using remote sensing data and the C factor values at measured locations. Siegel (1996) and Wheeler (1990) used the procedure to map C factor for the USLE. This method is based on correlation of the C factor and remote sensing data. The shortcomings, however, are that the C factor is indirectly mapped through vegetation classification, and the classification errors are thus introduced into the C factor map. Using average C factor value for each vegetation type leads to smoothing of estimates and disappearance of spatial heterogeneity and variability.

Support practice

The support practice factor P is the ratio between soil loss with a support practice such as contouring, strip cropping, terracing, etc. and soil loss with straight row farming up and down the slope. Here P is assumed to be one unit because no support practices are being applied to the study area. Vegetation restoration plans are not considered in this study.

LCTA plot inventory field methods

The U.S. Army Land Condition Trend Analysis (LCTA) program was developed at the U.S. Army Construction Engineering Research Laboratory (USACERL) under the sponsorship of the U.S. Army Engineering and Housing Support Center (USAEHSC) as a means to inventory and monitor natural resources on military installations. LCTA uses standard methods to collect, analyze and report natural resources data (Anderson et al., 1995a, 1995b, 1996; Diersing et al., 1992; Tazik et al., 1992), and is the Army's standard for land inventory and monitoring (Technical Note 420-74-3 1990). Over 50 military installations and training areas in the United States and Germany have begun or plan to implement LCTA. LCTA data is available for over three-quarters of the Army’s 12 million acre land base (Shaw and Kowalski, 1996).

The LCTA standard methods are designed to sample, collect, and maintain a permanent database on the condition of Army land resources. The methods include the required data collection equipment and detailed procedures (sampling and establishing permanent field plots, measuring topographical variables, collecting soil samples and plant specimens, recording ground and canopy cover, inventorying wildlife populations, and maintaining the data bases) for periodic short- and long-term monitoring of the field plots.

Plots were located using a stratified random sampling scheme based on soil and land cover types (derived from satellite imagery). Stratified random sampling allows statistical inferences to be made, while ensuring that all of the largest strata are represented in the

sample. Within the Geographic Resources Analysis Support System (GRASS) (GRASS, 1993), satellite images in green, red and near infrared wavelength bands are first used to perform an unsupervised classification allowing the selection of up to 20 land cover categories. The resulting land cover data layer is superimposed on a digital soil survey of the area. The occurrence of each land cover / soil combination of more than 2 ha (called a polygon or stratum) is identified. Then plot locations are selected by randomly assigning plots within polygons with the number of plots in each polygon proportional to it's area, which resulted in a random stratification by soil and land cover type. The total number of plots is calculated based on one plot per 200 ha and with a maximum of 200 plots.

Each field plot is 100 m in length by 6 m in width (600 m^2). A 100 m line transect is oriented lengthwise down the center of each plot. The plot data obtained can be used to analyze land use, ground cover, surface disturbance, allowable use and carrying capacity, tactical concealment, soil erosion, land rehabilitation effectiveness, plant community composition, wildlife habitats, etc. Because the field plots are located with Global Position System (GPS), the data can be readily used with a geographic information system and with satellite imagery data.

Slope length in meters and gradient (steepness) in percent are measured at the zero, 50, and 100m points along the 100 m line transect. Slope length is defined as the straight-line distance runoff travels across each sample point and estimated by pacing the distance between point of origin and point of deposition. Slope gradient is measured with a clinometer to the nearest half percent. Aspect is determined by standing at the 50 m point and estimating the general direction that water would flow across the site. Using a compass, aspect is estimated to the nearest octant. If the average slope is less than 5 percent, aspect is considered unimportant and 'level' is recorded.

Soil depth is estimated for each LCTA plot by driving steel rods into the soil. A composite soil sample and five small samples are taken approximately 1 m from the line transect at the zero, 25, 50, 75, and 100 m points at each plot. The soil samples are analyzed at labs for soil properties related to soil erodibility factor, productivity, and botanical composition.

Land use is recorded for each plot. Surface disturbance, ground cover, and canopy cover are estimated by the point intercept method as described by Diersing et al. (1992). Along the 100m line transect along the center of each plot, surface disturbance, ground and canopy cover data are collected at 1m intervals (that is, 0.5m, 1.5m, 2.5m, …, 99.5 m). The categories for disturbance include: no disturbance; road; trail (semi-permanent traffic route receiving no maintenance); pass (random vehicle track that does not follow an established traffic pattern); and other disturbance. Ground categories are bare ground (no cover), rock, litter, and basal cover. Canopy cover is recorded by species at 0.1m height intervals up to 2m

and at 0.5m intervals up to 8m in height. For each transect, the cover percentage for a particular vegetation type can be obtained by dividing the total number of the covered points by the total points measured (× 100%). With this plot configuration, it is possible to map the covered points within and between transects across the entire area. Moreover, percent cover could be determined for different plot sizes by sub-sampling within each transect. In addition, species composition, density, and height distribution of woody and succulent vegetation are investigated for each plot. The standard area is 100 m by 6 m. However, the width can be reduced for high density species.

Three different types of monitoring are performed at LCTA field plots: initial inventory, short-tern monitoring, and long-term monitoring. Above is the procedure of the initial inventory that provides detailed information of land use and site conditions. Subsequent short-tern monitoring is conducted annually to detect changes of land use, disturbance, ground cover, canopy cover, and other natural resources at short time-scales. Long-term monitoring is carried out every 3 to 5 years using the same detailed procedure as the initial inventory. The short-term monitoring procedure yields much the same information as those in long-term monitoring, but lesser detail, particularly with regard to species composition.

Case study area – Fort Hood

This study took place at Fort Hood, Texas (Figure 2.1). This 87,890 ha installation is located in Central Texas in Bell and Coryell Counties approximately 160 miles southwest of Dallas, TX. This region has long, hot summers and short mild winters. Average temperatures range from a low of about 8 ^{o}C in January to a high of 29 ^{o}C in July. Average annual precipitation is 81 cm. The month of peak precipitation is May with a secondary peak in September. There are 230-280 frost-free days per year. Elevation at Fort Hood ranges from 180 to 375 m above sea level with 90 percent of Fort Hood below 260 meters. Most slopes are in the 2 to 5 percent range though slopes in excess of 45 percent occur as bluffs along the flood plain and as the sides of slopes of the mesa-hills. Soil cover is generally shallow to moderately deep and clayey and underlain by limestone bedrock. Fort Hood consists of four distinct regions that have different military training activities, general vegetation types and topography.

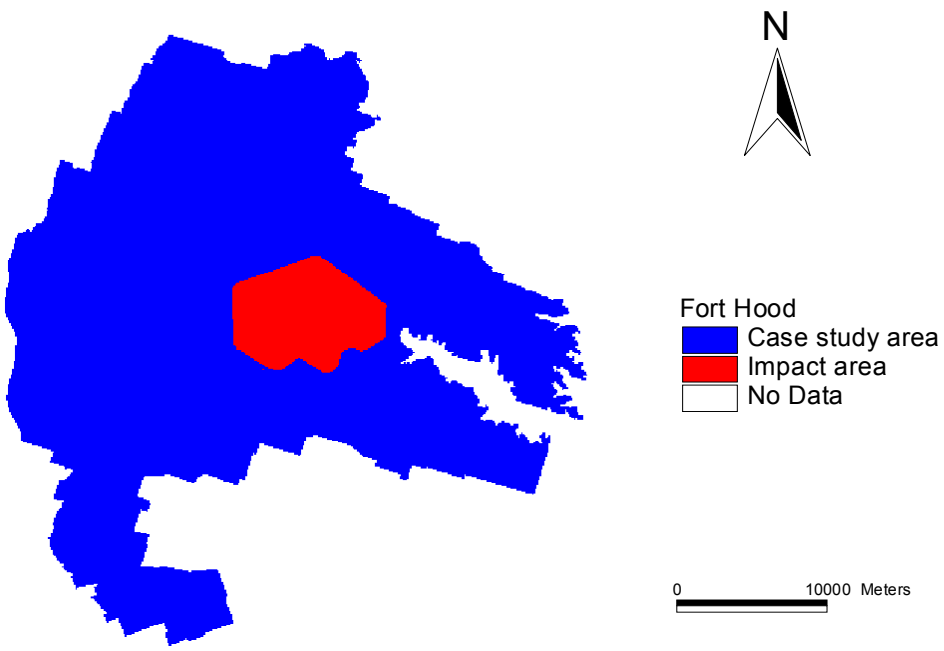

Figure 2.1. Case study area – Fort Hood, Texas.

Fort Hood lies in the Cross Timbers and Prairies vegetation area. The area is normally composed of oak woodlands with grass undergrowth. Traditionally the predominant woody vegetation consisted of ashe juniper (*Juniperus ashei*), live oak (*Quercus fusiformis*) and Texas oak (*Quercus texana*). Under climax conditions the predominant grasses consisted of little bluestem (*Schizachyrium scoparium*) and Indian grass (*Sorghastrum nutans*). East Fort Hood is dominated by oak-juniper woodlands, on high mesa-like hills with geologic cuts and slopes up to 45%. West and South Fort Hood are savannah type and dominated by mid-grasses, little bluestem (*Schizachyrium scoparium*) tall dropseed (*Sporobolus asper*) and Texas wintergrass (*Stipa leucotricha*) with scattered motts of live oak (*Quercus fusiformis*) on rolling topography and oak-juniper on hills and steep slopes along the major drainages. Central Fort Hood has a mixture of the savannah type on rolling topography and oak-juniper woodlands on mesa tops and along steep slopes of drainages.

The primary mission of Fort Hood is the training, housing and support of the III Corps and its two divisions (1st Calvary Division and 2nd Armored Division). Support is also provided to other assigned and tenant organizations such as the U.S. Army Reserve, the National Guard, the Reserve Officer Training Corps, and reservists from other services. Central Fort Hood

contains a 22,700 ha live-fire and artillery impact area and an additional 8,700 acre multi-purpose maneuver live-fire range. The range areas serve as familiarization and qualification firing ranges for all individual weapons, crew-served weapons, and the major weapons systems of active units assigned or attached to the III Corps and Fort Hood. Maneuver areas comprise 52,400 ha not including the multi-purpose live-fire area. Maneuver areas are used for armored and mechanized infantry forces in the conduct of task force and battalion-level operations, and for company and platoon level dismounted training, along with engineer, amphibious, combat support and combat services support training. West Fort Hood is used primarily for tracked and wheeled maneuver exercises at the Battalion level while South Fort Hood is used primarily for tracked and wheeled maneuver exercises at the smaller Platoon level. East Fort Hood is used primarily for small unit exercises, bivouac and foot soldier training because the terrain and dominant oak-juniper woodlands prevent large cross country exercises.

Case study data sets

LCTA database

At the Fort Hood case study area, a total of 219 field plots were established of which 163 were permanent field plots and the other 56 were special use plots. Special use plots were used for special issues that could not be addressed by core plots. These special issues included determining the success of land rehabilitation efforts, documenting the effects of burning, assessing natural recovery of degraded lands, etc. Special use plots were also used as control plots if they were placed in areas with little or no impact from military activities.

In the spring and summer of 1989, permanent field plots were established in a stratified random fashion using on LCTA methods based on an automated method of randomly selected plot locations using satellite imagery, soil surveys, and a computerized geographic information system (Warren et al., 1990). The number of plots allocated to each stratum was proportional to the percent of the land area occupied by the stratum. Each plot was 100 m by 6 m (600 m^2). The plots were measured in the initial inventory in 1989 for topographical information, land use, soil properties, disturbance, ground cover, canopy cover, botanical composition, etc., and annually re-measured through 1997. The inventory for long-term monitoring was carried out in 1992 and 1997. Because of missing plot markers, fire or other reasons, the number of the re-measured field plots generally decreased from 1989 to 1997 (Table 2.3).

Table 2.3. Number of the field plots in Fort Hood

Year	1989	1990	1991	1992	1993	1994	1995	1996	1997
Number of plots	215	214	220	220	200	166	178	0	0

The field plots were measured and re-measured using LCTA methods described above. A LCTA database for Fort Hood was established (Sprouse and Anderson, 1995) based on SQL commands. The database contains all the information measured and derived from the field plots and can be divided into nine distinct components including plot information, land use, vegetation, wildlife, climate, soil, supplementary information, summary, and validation tables. The input factors (soil erodibility, slope steepness, slope length, vegetation cover and management factor) related to soil erosion were calculated for all plots and included in the summary data.

Because not all the field plots were located using GPS when they were established in 1989, the coordinates of the field plots were re-measured using GPS in 1999. It was found that the root mean square error between the original and re-measured coordinates of the plots was 124.55m for the East direction, and 238.69 for the North direction. Because of the big differences in coordinates, the case study area was projected on the Universal Transverse Mercator (UTM) based on the coordinates re-measured by GPS.

Because the information from the original soil samples collected in 1989 was not enough to calculate plot soil erodibility factor values related to soil erosion, moreover, soil samples were re-collected from the field plots in 1999 (Wang et al., 2001c). The soil samples were analyzed in a soil lab for soil organic matter, sand and silt percentage, and classes of soil structure and permeability. The values of soil erodibility factor for the field plots were calculated using Eq. 2.6.

Rainfall data

No rainfall observation stations are located within the study area. Thus, it was necessary to use data from rainfall observation stations surrounding the study area to evaluate spatial variability in R factor estimates and their associated uncertainty. A total of 247 rainfall stations, located in Texas and surrounding states (Arkansas, Colorado, Kansas, Louisiana, New Mexico and Oklahoma) were used (Wang et al., 2001g). The data set of the maximum 26-year rainfall records came from the NCDC (National Climatic Data Center) Hourly and 15-minute Precipitation Database (provided by Steven Hollinger at the Illinois State Water Survey, Atmospheric Environmental Section). The value of rainfall-runoff erosivity factor (R)

was calculated for each rainfall station by the method developed by a research team headed by Hollinger. That is, Eq. 2.4 was used to calculate the energy contained in the volume of rain, and in an N year period, Eq. 2.5 was employed to calculate the annual R factor. In addition, the values of seasonal and half-month average rainfall-runoff erosivity (R) factors were computed using this data set. Based on traditional isoerodent map, annual R factor for Fort Hood is a constant 270 (Renard et al., 1997).

High-density soil sample data

In order to validate different mapping methods and to assess spatial uncertainty of soil erodibility in the National Cooperative Soil Survey (NCSS), a high-density soil sampling scheme was designed. A specific study area within Fort Hood was selected based on constraints imposed by Army training, and our desire to collect information from Fort Hood consisting of both Coryell and Bell counties. Thus, the center point of the sampling area was randomly selected from a larger area that would meet those requirements. Soil samples were collected in late summer of 1998, under the assumption that data collected during that time of the year would provide an approximate annual average based on the expected seasonal variability of the K factor (highest values in spring and lowest values in mid-fall and winter, Renard and Ferreira 1993).

We collected 576 soil samples on a grid whose points were located approximately 10m apart from each other. We obtained the real-time differentially corrected GPS location of some reference points, and completed the grid measuring distances with a tape. The end result was an approximate grid (as shown in Figure 1, Parysow et al., 2001a). The soil samples were obtained with a double-cylinder hammer-driven core soil sampler, which takes a solid cylinder of soil 76mm high by 76mm diameter, as described in Blake and Hartge (1986). Samples that fell on roads, edge of roads, and other highly disturbed areas were discarded, resulting in 524 usable samples for this study. Soil samples were stored in cardboard containers and transported to the soil laboratory at the University of Illinois at Urbana-Champaign, where they were analyzed to obtain all the necessary information to estimate K employing Eq. 2.6.

Ground control points and Digital Elevation Model (DEM)

A total of 24 road intersections were selected, measured for coordinates and elevation and used to assess accuracy of relevant topographical maps in position and elevation. For each of the intersections, two to four points controlling the intersection locations were measured for elevations and coordinates using a Trimble Pro XRS global position system (GPS). A total of 79 points across the whole area were obtained. The minimum and maximum elevation from the points was 183m and 333m with average of 262m and variance of 1403.

A 7-minute digital elevation model (DEM) at spatial and vertical resolution of 30m and 1m respectively for this area was acquired from the U.S. Geological Survey (USGS) (Figure 1 of Gertner et al., 2001d; or Figure 1 of Wang et al., 2001d). This DEM was classified into Level-2. The minimum and maximum elevation was 136m and 377m with average of 249.3m and variance of 1665.5. The root mean square error in elevation was 5.13 m.

Landsat TM images

For the case study area, multi-temporal Landsat TM images for the years 1989, 1990, 1991, 1992, 1993, 1994, 1995, and 1996 were obtained. The spatial resolution for all the images was 30m by 30m. These images consisted of band 1: 0.45-0.53 μm, band 2: 0.52-0.60 μm, band 3: 0.63-0.69 μm, band 4: 0.76-0.90 μm, band 5: 1.55-1.75 μm, and band 7: 2.08-2.35 μm and were geo-referenced to the UTM projection. The method used is as follows: 1) a set of digital orthophoto quads were acquired for AUG 1997 that were geo-referenced to UTM, WGS84; 2) these 113 DOQQ images were re-sampled to approximately 4 m resolution and mosaiced together to cover the case study area; 3) the first Landsat TM image was rectified to the map resulting from step 2; and 4) the remaining TM images were rectified to this first TM image.

METHODOLOGY

The important objective of this project is to develop a theoretical and methodological framework for optimizing sampling design, data collection, spatial modeling, mapping, uncertainty analysis, and management in terms of precision (errors) and/or expense as an integral part of the continuous monitoring-simulation process. By reviewing existing methods in these areas and assessing their advantages and disadvantages, we developed and presented a general methodology and its details for this purpose.

Existing methods and limitations

Traditional methods for sampling design, classification and mapping, accuracy assessment, and uncertainty analysis include the approaches used to determine plot size and shape, sampling pattern, and sample size, to perform image-aided spatial modeling, to calculate accuracy of spatial modeling, and to model uncertainty (i.e. variance) propagation from inputs to results. These methods are based on classical statistics theories and assume that sample data of a variable are spatially independent. However, sample data trend to be spatially correlated (i.e. samples from locations that are closer together tend to be more similar than samples from locations that are farther apart). The simplification of independence by traditional methods will lead to uncertainty far from the truth and limitations in application. The uncertainty and limitations vary depending on different methods and their applications. In recent years new methods have been applied to natural resources and ecosystems. Most of them were developed based on a theory of regionalized variables and geostatistics, and have shown good promise.

Sampling design

Sampling design is a cost-efficient procedure for collecting ground data about a variable to be estimated including determining plot size, plot shape, sample size, and sample patterns. The choice of plot shape depends on the variables to be investigated and can be readily determined from the published scientific literature. Generally, systematic sampling provides a better representation of a variable's spatial variability and is better used to collect data for mapping than stratified, random, and clustered sampling. Because the LCTA data have been made available for this project, and the data were obtained by a stratified random sampling

allows us to use different plot size and sample size in studies, the discussion for sampling design will thus be limited to determining plot size and sample size.

When designing an inventory program using traditional field sampling, it is usually desired to maximize the amount of information per unit cost. If there were a fixed budget for inventory, the objective would be to minimize the sampling variance. If there were a specified desired precision level for the sample estimate, the aim would be to minimize the cost of the inventory program. Based on either objective, plot size is related to both sampling variance and cost.

The traditional methods for determining appropriate plot size are optimization techniques that provide the optimal plot size given a budget (Smith, 1938; Freese, 1961; Zeide, 1980; Gambill et al., 1985; Reich and Arvanitis, 1992). These methods are based on the relationship between plot size and the coefficient of variation of a variable to be investigated. In a tropical forest inventory, for example, as the plot size increase, the number of tree species increases rapidly at the beginning, then slow and gradually becomes stable, and the plot size at which the number of tree species stabilizes can be considered to be appropriate. When the plot is very small, more generally, coefficient of variation of a variable decreases rapidly as the plot size increases, the decrease of coefficient becomes slow and eventually stable.

Estimation of population mean requires pre-calculation of sample size before sampling. Based on classical statistics theory, the sample size (n) for typical simple random sampling can be calculated:

$$n = \frac{t_\alpha^2 CV^2}{E^2} \tag{3.1}$$

where t_α is the value of student's t-statistics at a significant level of α, CV the coefficient of variation for the variable to be estimated, and E the maximum relative error. The corresponding equations for other sampling patterns can be derived. When auxiliary data sets such as remotely sensed images are used to help the estimation, the sample size can be reduced by a factor of $(1 - r^2)$ where r is the coefficient of correlation between the observed and estimated values using the auxiliary data sets. On the other hand, the sample size corresponds with coefficient of variation and thus with plot size based on the relationship of plot size with coefficient of variation. Furthermore, introducing costs such as travel and measurement time needed into Eq. 3.1 makes it possible to determine optimal plot size and sample size based on cost using traditional statistical theory.

However, these methods assume that sample data are independent and do not deal with spatial dependence of a variable and cross spatial variability between variables. The

similarity of data and interaction among variables should allow for a reduction of sample plots or uncertainty. Conversely, neglecting the spatial dependencies will require more sample plots or more cost. Moreover, the objective of traditional sampling design focuses more on unbiased estimation of population averages and less on local estimation. Therefore, the sample data obtained by traditional methods may not be suitable for generating spatial models (e.g. maps).

The theory of regionalized variables in geostatistics has been applied to sampling design (McBratney, et al., 1981; McBratney and Webster, 1981 * 1983; and Olea, 1984). Generally, the information representation obtained by systematic sampling is better than that by random sampling because variables are spatially dependent. The theory of regionalized variables enables the spatial dependence of a variable to be estimated from data under reasonable assumptions and then to be used to estimate means with minimum variance. The estimation variance depends only on the degree of spatial dependence. Given a known spatial dependence - semivariogram, the sampling variance of any regular scheme can be forecast before it is put into effect. If the desired precision is specified, the size of sample (in fact, sampling distance) required to achieve it can be determined.

Most of the applications focus on minimization of the estimation variance to find the minimum number of samples needed to attain a specific maximum level of error. For example, McBratney described a method of optimal sampling based on kriging and proposed two assumptions for the method. First, the maximum standard error of kriged estimates is a reasonable measure of the goodness of a sampling scheme. And second, the spatial dependence is expressed quantitatively in terms of the semivariogram. Arvanitis and Reich (1991) studied the effect of spatial pattern of trees on the accuracy and precision of sample estimates as well as taking the spatial factor into account.

Additionally, Englund and Heravi (1993) presented a practical application for sampling design optimization by conditional simulation, and generated detailed spatial model for case-specific optimization of sampling design. The entire process of the sampling estimation and decision is simulated by a Monte-Carlo approach. The optimization is realized through economic functions or on decision constraints, such as, unit sample cost, number of samples, total sampling cost, remediation cost and non-remediation cost, rather than minimization of estimation variance.

Scale and resolution

In addition to sampling design, another aspect that has to be clarified for spatial modeling and mapping is scale and resolution. In ecological modeling and management, scale is considered to be an attribute that affects spatial features, patterns, and processes of ecological variables

and resources in both space and time (Wu and Qi, 2000). The scale related issues include determining appropriate spatial and temporal scales or resolutions used to conduct the studies, interpolating or extrapolating results from one scale to another, including scaling up (from fine resolution to coarser - data aggregation) and vice versa (called scaling down), and modeling the change of spatial information change due to scaling.

Because of the scale dependency, choosing optimal spatial and temporal resolution is critical to capture spatial and temporal patterns, features, and processes of ecological and resource systems. The widely used methods are variance-based, texture analysis, fractal, and semivariogram. The variance-based methods include geographical variance (Moellering and Tobler, 1972) and local variance (Woodcock and Strahler, 1987). The geographical variance method works well for hierarchical structures such as landscape ecology (Wu et al., 2000). However, the hierarchical structure and assumption of data aggregation limit its application because the values of digital maps and images at a coarser resolution are usually not simple aggregation of the values at a finer resolution and pixels at different resolutions may be not nested. A local variance method is based on the relationship between spatial resolution and spatial dependence. The local variance is defined as the average value of the variances within a 3 by 3 moving window passing through the entire image. The local variance varies over spatial resolution and its maximum value is an indication of the appropriate resolution to capture spatial variability of the objects. Its disadvantage is that simple average of pixel values at a finer resolution may lead to quick disappearance of significant features at a coarser resolution.

Texture analysis is widely used in image processing, classification, and mapping, and varies depending on different measure indices such as variance, standard deviation (Holopainen and Wang, 1998), and Haralick textures (Haralick et al., 1973), etc. Similar to local variance, the spatial variability of image data in terms of textures varies with spatial resolution. The resolution with maximum variability can be considered to be optimal. A relative new alternative is the fractal method for determining optimal spatial resolution. Mandelbrot (1983) presented the fractal geometry and a key concept – statistical self-similar property that any portion of an object is similar in shape to the whole of the object at reduced scale. The similarity or dissimilarity can be measured by fractal dimensions of real world such as curves and surfaces as indices of roughness or complexity (Wang et al., 1997). The fractal dimension of an image decreases as the resolution becomes coarser. The scale at which the highest fractal dimension occurs may be the spatial resolution at which most of the interesting processes operate (Goodchild and Mark, 1987; Lam and Quattrochi, 1992). The method is very promising (Cao and Lam, 1997; Xia and Clarke, 1997), however, so far its development has not directly led to techniques that can be used to infer results across scales.

The semivariogram in geostatistics measures spatial variability of a variable, that is, the change of average dissimilarity between data over a lag distance h separating the data given a direction. When the lag distance is equal to a pixel size, the value of the semivariogram function is the semivariance at a lag of one pixel. The relationship between the pixel size and the semivariance at a lag of one pixel is similar to that between the spatial resolution and local variance mentioned above. The maximum semivariance is an indication of the appropriate spatial resolution to capture the desired spatial variability of the variable (Atkinson and Danson, 1988). Compared to the methods above, the semivariogram based method is more promising because it is based on capturing and modeling the spatial variability of a variable, it is the basis of all goestatistical methods used for spatial modeling and mapping, and it is expected that the corresponding methods for inferring results across scales can be derived.

Inferring the underlying spatial processes and results across scales is another difficult task in understanding ecological and resource systems and obtaining accurate and useful information for management decision-making. The existing methods for this purpose include moving average window, filtering, nearest neighbor, area-weighting average, expected-weighting average, explicit integration, spatial data aggregation (Moellering and Tobler, 1972; Jarvis, 1995; King, 1991; Wang et al. 1997; Wu, 1999). Some of them are related to the methods for determining optimal resolution. For example, using a moving average window local variance method results in digital values and variances of pixels at a coarser resolution from a finer, and the pixel variances decrease very quickly as the resolution increases (i.e. heterogeneity rapidly disappears). The nearest neighbor method can improve this, but may lead to misunderstanding of spatial patterns and processes because dominant values may be missed when going from a finer to coarser resolution. Other methods attempt to overcome the shortcomings, however, being very much subject to knowledge scientists have had in the areas. Furthermore, inferring uncertainties (variances of estimates) across scales in addition to obtaining estimates is problematic.

Modeling the change of spatial information due to scaling is a scale-related issue noted recently by scientists (De Cola, 1997; Vieux, 1995). It is important because scaling will result in changes of spatial patterns and processes, and modelers and managers need to know whether incorrect methods or different scales cause the changes. At the same time, the changes also mean uncertainties and managers need information on the uncertainties. De Cola (1997) suggested a measure by calculating global variance change across scales. Vieux, (1995) used the theory of entropy (Shannon and Weaver, 1964) to measure loss of spatial information content. The loss of entropy from finer to a coarser resolution can be represented as the difference of entropy between two scales. However, these are global measures and cannot be used to explain local changes of spatial information, for example, anisotropy of spatial variability in different directions. Another problem is how to link them with the

methods used to determine appropriate scales and infer results across scales. Therefore, there is a strong need to develop a systematic methodology for these purposes.

Mapping, accuracy and uncertainty assessment

In natural resource, ecological and environmental management, managers need accurate information in order to make the correct decisions. Accurately mapping the natural resources and ecosystems is very important. This is true especially when multiple variables are spatially correlated with each other and needs to be mapped jointly by aid of remotely sensed data. Separately mapping each of the variables and then overlapping them will lead to significant errors and loss of the correlation between the variables. However, jointly and accurately mapping multiple variables is usually very difficult mainly because of interactions among the variables and imperfection of existing methods.

The widely used methods for mapping are supervised and unsupervised classification or stratification, and methods that integrate both of the previous methods (Campbell, 1996; Holopainen and Wang, 1998; Lillesand and Kiefer, 2000; Wang, 1996; Wang et al., 1998). These methods result in homogeneous polygons or strata of pixels and, therefore, smoothing of estimates and the disappearance of spatial heterogeneity. This shortcoming can be improved by a regression method (Peng, 1987) and a k-nearest neighbors method (Tomppo, 1996). However, the regression can lead to illogical or extreme estimates, while it is not clear whether k-nearest neighbors can lead to unbiased population estimates. Moreover, a common assumption behind these methods is that sample data are not spatially correlated. This assumption makes it possible to provide unbiased estimates for populations. However, it is problematic in that reliable local estimates that reproduce the spatial variability of variables and interactions among them cannot be obtained. As detailed precision management planning becomes more common, the need for reliable local estimates will become essential.

In order to improve local estimates, Wang (1996) introduced a knowledge-based approach into remote sensing based estimation system of forest resources. Recent developments include spectral mixing analysis, uses of hyper-spectral remote sensing and fine resolution images (Campbell, 1996), and data fusion from different sensors (Wang, et al., 1998). However, real breakthroughs in methodology and accuracy have not been realized.

Using the methods described above, the uncertainty of resulting maps for unknown locations is not provided. Traditionally, accuracy is typically assessed by calculating correlation or root mean square error between estimated and observed values of a continuous variable, or an error matrix for a categorical variable. These traditional measures are for the global accuracy of a map. However, map accuracy often varies spatially depending on the complexity of landscape, soil properties, topographical features, density of sample data, and the accuracy of

remotely sensed data used (Congalton, 1988; Steele et al., 1998). The traditional methods lacks in the capability to measure spatial uncertainty. Moreover, errors from sampling, measuring, image processing, and models can propagate to the product maps. This error propagation is not accounted for by the traditional methods.

Another group of approaches used for spatial modeling and mapping are geo-statistical methods consisting of interpolation and simulation techniques (Chiles and Delfiner, 1999; Goovaerts, 1997; Journel and Huijbregts, 1978). These methods are based on the spatial variability theory, that is, spatial dissimilarity of the ground characteristics that varies depending on the separation vector of data or separation distance given a direction. They provide prediction maps of variables with their variance maps as uncertainty measure of estimates at any locations. These methods have been widely used in geology and recently expanded to applications in natural resource and environmental sciences. For example, Rogowski and Wolf (1994) investigated the variability in soil map unit delineation using kriging. Barata et al. (1996), Hunner et al. (2000), Wallerman (2000), and Xu et al. (1992) used cokriging and co-located cokriging methods to map forest variables with remotely sensed images and other auxiliary data, and a significant improvement was found. Mowrer (1997) used a Monte Carlo technique of sequential Gaussian simulation and studied propagation of uncertainty through spatial estimation processes for old-growth subalpine forests.

Various kriging and cokriging approaches are generalized least squares regression algorithms that interpolate variable values at unknown locations given a data set. Kriging estimates are best in terms of local minimum error variances in local areas. However, kriging estimates are smoothed, which leads to overestimation in the areas with small values and underestimation in the areas with large values. At the same time, the smoothing differs from place to place. The spatial variability of the estimated variable is higher in the areas with dense samples than in sparsely sampled areas. More importantly, kriging variances depend only on the data configuration and not on the actual observed data, and thus do not adequately reflect uncertainty. Indicator kriging methods have improved capabilities in this regard and provide a local uncertainty analysis by calculating conditional variances and probability maps of values larger than a given threshold (Goovaerts 1997). In this way, the conditional variance depends on not only data configuration but also, the data values.

In general when spatial simulation techniques are used, conditional distributions based on the collected data set are developed first, and then from these distributions the values of the stochastic variable at unknown locations are drawn at random. Once values at all the unknown locations are simulated, a realization of the stochastic variable is developed. After many realizations, the set of alternative realizations provides a visual and quantitative measure (actually a model) of spatial uncertainty (Deutsch and Journel 1998, Goovaerts

1997). The expected estimates and various uncertainty measures such as conditional variances and probability maps can be derived from these realizations. There are several spatial simulation approaches with the most widely used method being sequential Gaussian simulation. Sequential Gaussian simulation, however, requires the assumption of normality and may create underestimates or overestimates when there are extremely large or small values. As an alternative to Gaussian simulation, sequential indicator simulation can be used for the purpose of spatial uncertainty analysis by reproducing indicator covariance models. This method is especially useful when extreme values are very important to natural resource, ecological and environmental management.

When multiple variables are spatially correlated with each other, Gómez-Hernández and Journel (1992), Almeida (1993), Almeida and Journel (1994) presented a joint sequential simulation for mapping. In addition to prediction and variance maps, this method outputs co-variance maps indicating interactions among the variables and thus reproduces spatial cross variability between any two variables. Furthermore, remotely sensed data can be considered to be models of ground characteristic variables. The spatial variability of each variable and spatial cross variability between two variables are coded in the auxiliary data. The auto semivariogram and cross-semivariograms used in the method can capture the spatial dissimilarity and correlation between the ground characteristics and auxiliary data. Using the auxiliary data in the joint sequential simulation leads to a co-simulation, which can improve spatial modeling of variables and their correlation. This is very promising approach for spatial modeling and mapping of complex and multiple ecosystems.

An error budget is a comprehensive catalog of the different error sources in both surveys and models. In an error budget, the relative variance contributions of all uncertainty sources are calculated and main sources of the uncertainties are identified. This method is similar to an ANOVA table listing the contribution of each uncertainty source.

There are several methods for assessing the sources of uncertainty in models. They include Monte Carlo methods (Heuvelink, 1998), Fourier Amplitude Sensitivity Test (FAST) (Cukier et al., 1973), Taylor series (Gertner et al., 1995), Polynomial regression (Gertner et al., 1996), Sobol's method (Sobol, 1993), etc. All these methods have their advantages and disadvantages. For example, the Monte Carlo method and Sobol is computationally intensive when the number of input parameters increases, although they can be used to deal with interactions among the input parameters. The FAST method is computationally efficient, but assumes that all the input parameters are independent. The Taylor series expansion based methods can handle interactions among input parameters but, require the model functions can be continuously differentiable. The most important disadvantage is that all the methods were originally developed for an error budget of mean estimates for a population and cannot be directly applied to spatial uncertainty analysis.

As more attention is paid to detailed precision management planning, spatial uncertainty analysis becomes increasingly necessary. Additionally, there is a need to spatially assess major error sources because the relative uncertainty contributions vary over space (i.e. an error source is important to the output at a location, but may be not at another). Therefore, the error budget has to be done on a pixel-by-pixel basis to account for spatial variation of uncertainties. When multiple variables are highly correlated with each other, furthermore, considering interactions among the variables in mapping may result in an increase of accuracy. At the same time, there is abundant evidence to support that use of spatial information from neighboring locations can improve estimation at an unknown location. However, there are no existing methods available to assess the effect of the interactions and spatial information from neighbors on mapping.

When prediction is made using a Geographical Information System (GIS), the spatial error budget can become very complicated and difficult. Veregin (1992) proposed a hierarchy for modeling error in GIS operations. The hierarchy consists of five classes: error source identification, error detection and measurement, error propagation modeling, strategies for error management, and strategies for error reduction. The error sources are divided into several phases: data acquisition, data processing, data conversion, and data analysis and modeling. Within each phase, errors are further partitioned. For example, data analysis and modeling errors are divided into quantitative modeling and classification. Moreover, the errors can be due to incorrect position and/or measurements of variables. If remotely sensed data are used for mapping, various errors related to climate, sensor systems, image pre-processing, image rectification etc., will be included (Lunetta et al., 1991).

The errors in GIS propagate and accumulate to the outputs through operations such as data conversion, scaling up, data layer overlapping, and so on. Clarke (1985) examined the error involved in the conversion of polygonal data to a pixel-based format and found that the error was related to the complexity of the surface and the characteristics of the polygons. When the data are aggregated from finer resolution to coarser resolution, the errors are propagated. For example, the error propagation by scaling up in land surface process models was studied by Friedl (1997). Veregin (1992) summarized the methods used for modeling the error propagation and accumulation by data layer overlay. The methods are different from positional error to thematic error, from numerical data to categorical data, and also due to different operations such as “AND” and “OR”.

The errors in GIS operations are not always easy to identify and often very difficult to model their propagation. A general procedure for handling errors in GIS has been proposed by Openshaw (1992) based on Monte Carlo simulation (recommended method). As we mentioned above, however, this method is computationally very expensive and may be not practical especially if the spatial error budget is for a large grid (a large number of the

product of rows and columns). A very promising method may be polynomial regression (Gertner et al., 1996). This method can handle various source errors including interactions and effect of spatial information, but improvements to are needed. In a word, new methods need to be developed or existing methods have to be improved so that these methods can have the capacity to jointly map multiple variables, analyze spatial uncertainty and identify and quantify various resource errors.

Methodological framework

We developed a general GIS-based methodology to make spatial and temporal predictions, analyze uncertainty, and build error budgets (Figure 3.1). The methodology has been applied to a spatial and temporal version of models. The methodological framework (Gertner et al., 2001c; Wang et al., 2001a) integrated a map generation procedure – spatial modeling and simulation (in the right of Figure 3.1) and spatial uncertainty analysis procedure for resulting maps (in the left of Figure 3.1). The objective of spatial modeling and simulation is to create accurate maps with unbiased and reliable estimates for populations, sub-areas, and any specific location, and to provide spatial uncertainty measures of the estimates, including variance, covariance, and probability maps for each of input variables and interactions among them, in addition to the global accuracy measures. The aim of spatial uncertainty analysis is to identify various error and uncertainty sources and to derive relative uncertainty contribution maps for these error sources.

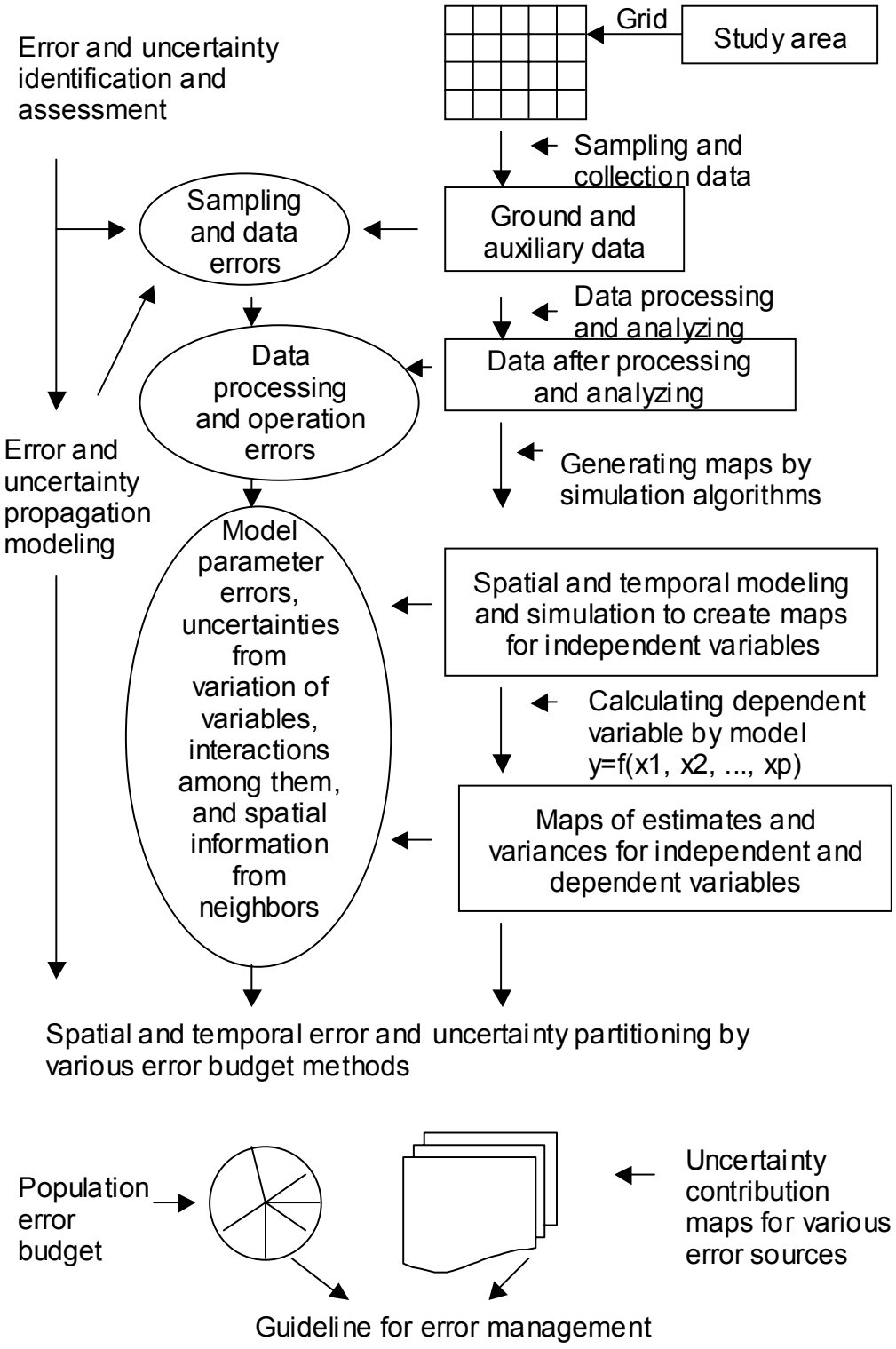

Fig. 3.1. A general methodology for spatial modeling and simulation, and uncertainty analysis.

In general, the steps taken for the spatial modeling and simulation are as follows:

- Generating a grid of the study area;
- Sampling and collecting data;
- Data processing and analysis;
- Generating maps by simulation algorithm;
- Calculating prediction and variance maps of dependent variable by model or function $y = f(x_1, x_2, ..., x_p)$.

A grid for the study area should be created and used for sampling and collecting ground and auxiliary data. The auxiliary data are co-located for the grid and include digital elevation models, soil type maps, and various remotely sensed images. Appropriate plot size should be determined for collecting data and mapping (Wang et al., 2001e). Data processing and analyzing include ground data grouping, transformation, statistical analysis, auxiliary data rectification, conversion, transformation, and scaling. The scaling means determining appropriate spatial and temporal resolution (pixel or cell size) for mapping and inferring results cross scales (Gertner et al., 2001d; Wang et al., 2001d). Selecting appropriate resolution should be integrated with determining optimal plot size (Wang et al., 2001e).

The methodology for map generation is based on simulation algorithms and spatial variability theory of variables in geostatistics. The simulation methods include sequential Gaussian simulation (Gertner et al., 2000; Wang et al., 2001f), sequential indicator simulation (Wang et al., 2001h; Wang et al., 2000b), and joint sequential simulation (Gertner et al., 2001a and 2001c; Wang et al., 2001b). These methods can be used for one or more than one variable. The auxiliary data such as remotely sensed images or other digital maps such as digital elevation models can be introduced into the simulation algorithms, which lead to co-simulation. When extreme values are not important, Gaussian simulation is a good choice. If the attention is paid to extreme values, indicator simulation should be taken into account. When multiple variables that are spatially correlated with each other are jointly mapped, joint sequential simulation or co-simulation with co-located auxiliary data. These methods can provide unbiased estimates of populations and reliable estimates of any sub-areas, and also reproduce the inherent spatial variability of the variables, and provide their spatial statistics in term of uncertainty. The prediction maps of the variables are employed to derive prediction and variance map of the dependent variable by relevant model or function.

The spatial uncertainty analysis procedure in the left of Figure 3.1 consists of error and uncertainty identification and assessment, modeling error and uncertainty propagation, error and uncertainty budget, and suggesting guidelines for error management. Various source errors and uncertainties in the GIS-based prediction system are assessed and shown in the middle of Figure 3.1 and their detailed classification is presented in Figure 3.2. There are many spatial and temporal errors in the subcomponents of models such as equations related

to soil erosion listed in Chapter 2. To obtain the input subcomponents, many different steps are taken and there are obviously many factors that cause uncertainties in the prediction of erosion both in small areas and large areas. These errors arise mainly from data, material, operations, modeling, and the inherent fuzziness of the real world. The errors and uncertainties are divided into three groups: sampling and data errors, data process and operation errors, and modeling and simulation errors. Within each of these groups, the error sources are further divided into sub-groups. The error sources, propagation, and accumulation are depicted in Figure 3.2. This figure is a very broad and general representation of some of the main errors that occur in the prediction of a natural resource and ecological system.

The error budget and partitioning into various sources of errors are generated (in the bottom of Figure 3.1). Error budgets can be used to assess the quality of the overall simulation system. An error budget can be considered as a catalog of the different error sources that allows the partitioning of the prediction variance and according to their origins. In table form, Table 3.1 displays how the error budget partitions error of a population prediction by sources based on Figure 3.2. As a specialized form of sensitivity analysis, an error budget shows the effects of individual errors and groups of errors on the quality of a multi-component model's predictions. The goal in developing the error budget is to account for all major sources of errors that can be expected in a system. By doing this, the sources of errors can be examined and partitioned in different ways. Additionally, an error budget can be generated for different time steps and spatial scales. The error budgets have been generated for both large and small areas.

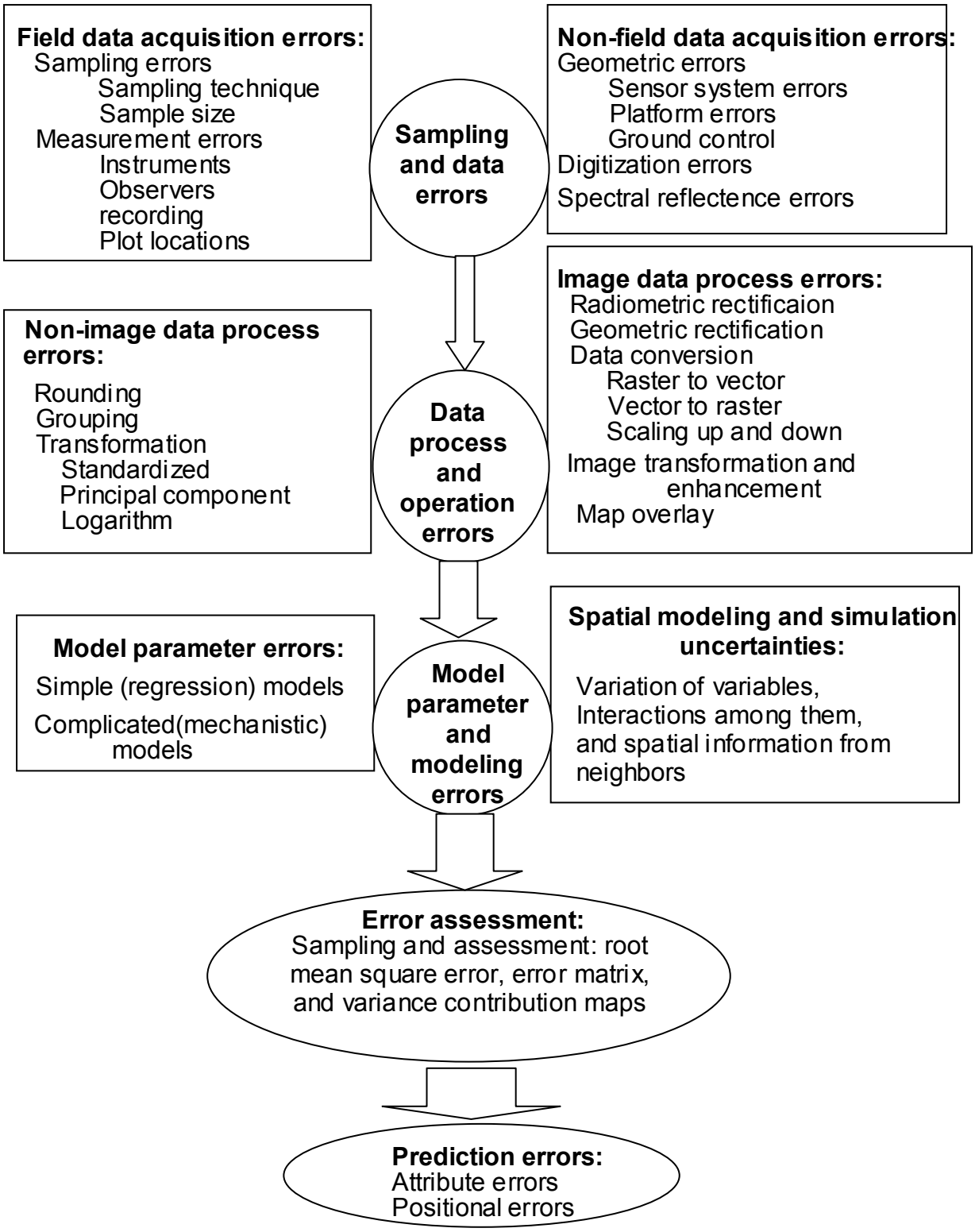

Fig. 3.2. Error sources and propagation of spatial modeling and simulation system.

When a spatial uncertainty budget is done, results will be relative variance contribution maps. As an example for spatial uncertainty budget, Figure 3.3 presents total variance maps of predicted ground cover, canopy cover, vegetation height, and vegetation cover and management factor C related to soil erosion, and relative variance contributions of the input variables to uncertainty of predicted C factor values for pixel at a transect line. The relative variance contribution varies over space and main uncertainty source differs from place to place.

Different approaches have been developed to generate the error budgets: deterministic and stochastic approaches. The approaches used depend on the structure of subcomponent models and the characteristics of errors. The deterministic approaches are based on analytical statistical estimators (expected mean square error models) and Taylor series approximations based on subcomponent models that are mathematically differentiable (Fang et al., 2001b; Parysow et al., 2001b). In terms of the stochastic approaches, they are Monte Carlo techniques based on simple random and Latin Hypercube sampling; and on Fourier analyses techniques (Fourier amplitude sensitivity test (FAST)) (Fang et al., 2001a; Gertner et al., 2001d; Wang et al., 2000a). Moreover, we have developed regression modeling for variance partitioning (Gertner et al., 2001a and 2001c). In addition, we are developing approaches that are a hybrid of both approaches based on surrogate models. These surrogate models are the simplification of the overall system that are computationally efficient and can be easily assessed in terms of their statistical properties. These will be the basis for our composite error variances and the partitioning of the error variances.

We will apply the GIS-based methodology to the case study – prediction and uncertainty analysis of soil loss using RUSLE. The flow of data and operations for this application is depicted in Figure 3.4. The study area – Fort hood is first sampled and ground data are collected for the primary variables related to soil erosion. The primary variables include soil properties, topographical features, vegetation cover variables, and rainfall. In addition, auxiliary data such as digital elevation model and remotely sensed data are acquired. A number of simulation algorithms with and without the auxiliary data are carried out to generate maps for each primary variable. The prediction maps of the primary variables together with empirical equations listed in Chapter 2 are then used to calculate the input factors including rainfall-runoff erosivity factor R, soil erodibility factor K, topographical factor LS, vegetation cover and management factor C, and support practice factor P. Finally, soil erosion is derived using Eqs. 2.1 and 2.2. The expected maps and their variance maps of the input factors and soil erosion status are obtained.

Using the prediction and variance maps above, a spatial uncertainty budget is first carried out for prediction of each input factor from its primary variables. The overall spatial uncertainty

budget is then made from the input factor to prediction of soil erosion. Finally, we will suggest guidelines of error management for prediction of soil erosion.

Table 3.1. A partition of final prediction variances and errors based on Figure 1.

Error sources	**Prediction variances %**	**Prediction errors %**
Data errors		
Sampling error		
Measurement error		
Geometric error		
Digitized error		
……		
Sub-total		
Data process errors		
Rounding		
Transformation		
Geometric rectification		
Image overlapping		
……		
Sub-total		
Experimental design error		
Sub-total		
Model parameter errors		
Component 1		
……		
Component n		
Sub-total		
Modeling and simulation uncertainties		
Variation of variables		
Interactions		
Neighboring information		
Sun-total		
Prediction value error		
Spatial error		
Human error		
Total		

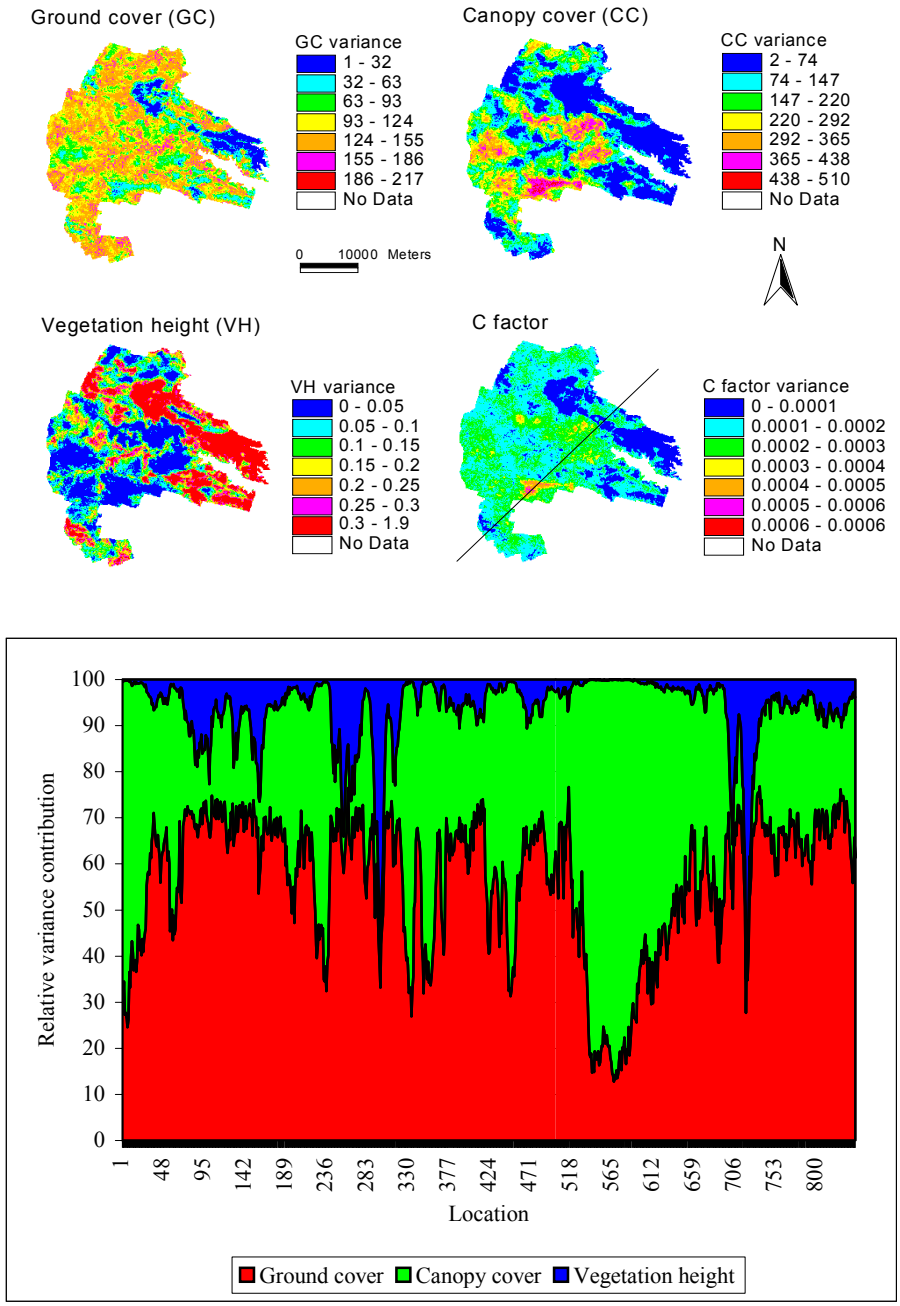

Fig. 3.3. Variance maps of predicted ground cover, canopy cover, vegetation height, and vegetation cover and management factor C related to soil erosion of Fort Hood, and relative variance contributions of the input variables to uncertainty of predicted C factor values for pixels at a transect line marked at the C factor variance.

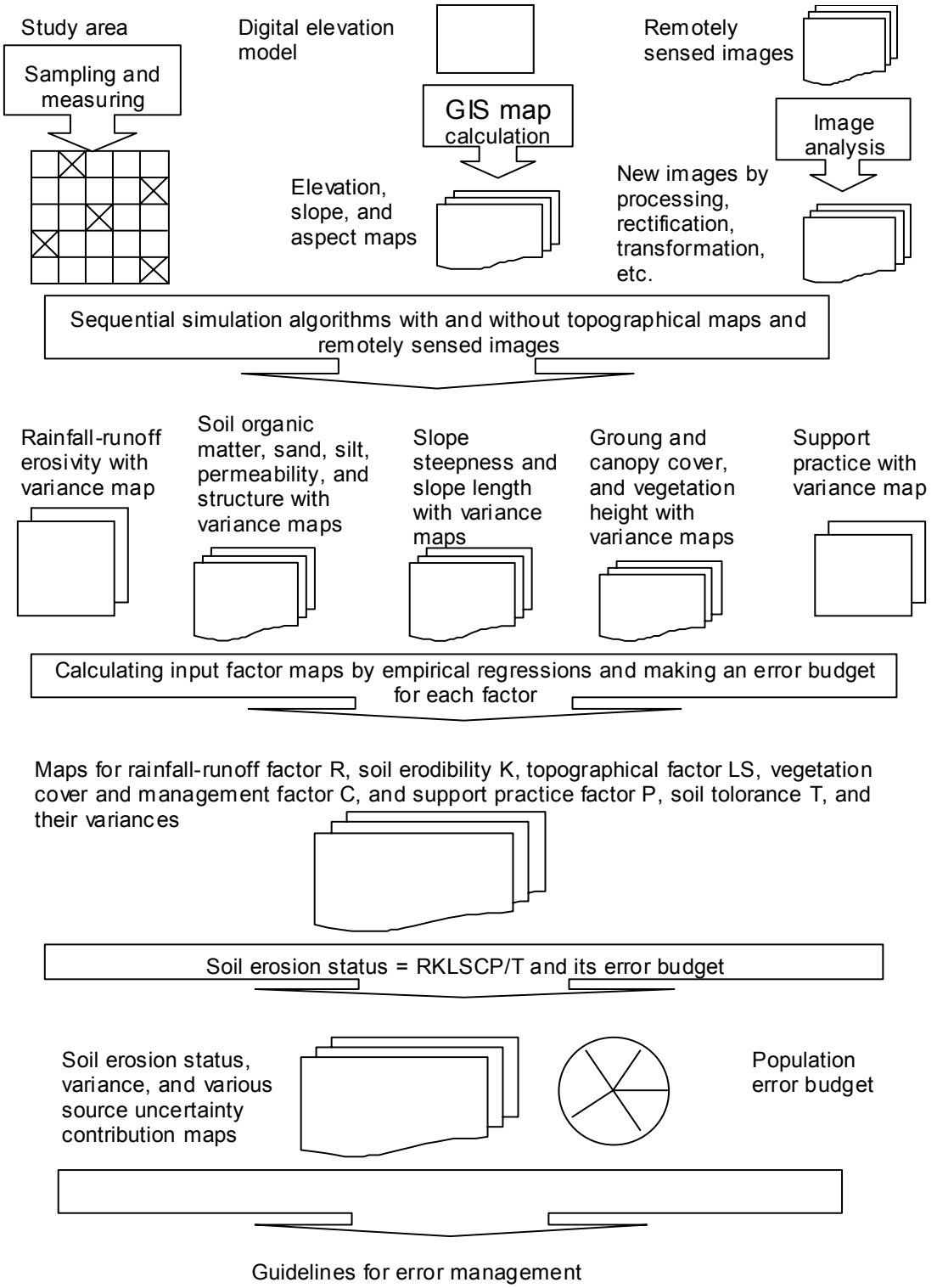

Fig. 3.4. Flow of spatial modeling, simulation, and uncertainty analysis for the case study – prediction of soil erosion using the Revised Universal Soil Loss Equation.

Spatial variability and cross variability

The GIS-based methodology mentioned above was developed based on spatial variability and cross variability of variables. Generally, a sample datum of a variable is similar with another sample datum separated by a distance h within a distance range given a direction, and the similarity becomes weaker and finally disappears as the separation distance h increases. That is, sample data separated by a distance h are only slightly dissimilar when they are close to each other, and the dissimilarity becomes stronger as the separation distance h increases, and finally the data get independent out of a certain distance range. The dissimilarity of data varies over space is called spatial variability of a variable.

Furthermore, the value of a variable at one location is related to the value of another variable a vector h apart. If both variables are positively related, an increase (decrease) in value of a variable from one location to another tends to be associated with an increase (decrease) in value of another variable. Conversely, a negative spatial correlation between two variables means that the increase (decrease) of a variable tends to be associated with the decrease (increase) of another variable. This is called spatial cross variability between two variables.

The spatial variability of a variable and cross variability between two variables can be modeled as realizations of random functions and by sampling. A study area can be divided into N pixels of a grid and P variables are estimated. In this area, a sample is drawn and the sample data set $\{ z_p(u_\alpha)$, $u_\alpha = 1, 2, \ldots, n, p = 1, 2, \ldots, P\}$ is obtained for P variables, and n is the number of sample data. The data of a variable p at location u_α is $z_p(u_\alpha)$. The expectation and variance for the variable p are m_p and σ_p^2, respectively. The cross covariance measuring the spatial cross variability between two variables is computed as:

$$C_{pp'}(h) = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} z_p(u_\alpha) \cdot z_{p'}(u_\alpha + h) - m_{p_{-h}} \cdot m_{p'_{+h}} \qquad (3.1)$$

with

$$m_{p_{-h}} = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} z_p(u_\alpha)$$

$$m_{p'_{+h}} = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} z_{p'}(u_\alpha + h)$$

where N(h) is the number of pairs of data locations a vector h apart, h is called lag given a direction, $m_{p_{-h}}$ and $m_{p'_{+h}}$ are the means of the tail values of variable p and head values of variable p' respectively. When p = p', Eq. 3.1 means covariance between data values of the same variable separated by a vector h, measuring spatial variability of the variable. On the other hand, cross semi-variograms, $\gamma_{pp'}(h)$ measures spatial cross variability between two variables and can be derived:

$$\gamma_{pp'}(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [z_p(u_\alpha) - z_p(u_\alpha + h)][z_{p'}(u_\alpha) - z_{p'}(u_\alpha + h)] \quad (3.2)$$

When p = p', Eq. 3.2 indicates semivariogram measuring spatial variability of a variable. When auxiliary data $x_q(u)$ (q = 1, 2, …, Q) for Q auxiliary variables are available at each location to be estimated, the spatial correlation between an estimated variable and an auxiliary variable can be obtained by Eqs. 3.1 and 3.2.

Eqs. 3.1 and 3.2 cannot be used to measure spatial variability of a categorical variable such as land use and cover. Various indicator methods have been developed so that probabilities of categories can be derived from sample data and used to obtain estimates at unknown locations. In the other word, the pattern of spatial variability for a continuous variable may differ depending on whether the variable values are small, medium, or large, and should be modeled separately. Thus, indicator approaches are also needed. The continuous variable z has to be subdivided into K+1 discrete intervals and K threshold values z_k are defined (k = 1,2,…,K). These threshold values are referred to as cutoff values. The indicator coding of the measurement data is then carried out as follows:

$$i(\mathrm{u}_\alpha; z_k) = \begin{cases} For \quad continuous \quad \mathrm{var}\, iables: \\ 1 \quad if \quad z(u_\alpha) \le z_k \quad k = 1,...,K \\ For \quad categorical \quad \mathrm{var}\, iables: \\ 1 \quad if \quad z(u_\alpha) = z_k \quad k = 1,...,K \\ 0 \quad otherwise \end{cases} \quad (3.3)$$

The spatial variability of the variable is estimated for each cutoff value using the indicator data and indicator semi-variograms. The indicator semi-variograms imply spatial similarity of indicator variables depending on the separation vector of data, that is:

$$\gamma_I(h; z_k) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [i(u_\alpha; z_k) - i(u_\alpha + h; z_k)]^2 \quad (3.4)$$

where $i(u_\alpha; z_k)$ and $i(u_\alpha + h; z_k)$ are the indicator data of the variable at spatial locations α and α + h, respectively.

As the separation distance of data given a direction increase, generally, semivariograms increase rapidly at the beginning, then slowly, and eventually become stable. Semivariogram or covariance inference provides a set of experimental values for a finite number of lags and directions. The spatial modeling and mapping by geostatistical methods such as simulation require semivariogram or covariance values at any separation distance h. Thus, continuous functions need to be fitted to the experimental values. In geostatistical methods, on the other hand, the semivariogram or covariance function will be used to derive weights λ_α of sample data given a neighborhood. In order to obtain non-negative variance of an estimate $Z(u)^*$ at any location *u*:

$$Var\{Z(u)^*\} = Var\{\sum_{\alpha=1}^{n} \lambda_\alpha z(u_\alpha)\} = \sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} \lambda_\alpha \lambda_\beta C(u_\alpha - u_\beta) \geq 0 \qquad (3.5)$$

the covariance function C(h) must be positive definite. Eq. 3.5 can be also repressed with semivariogram by following relationship:

$$\gamma(h) = C(0) - C(h) \qquad (3.6)$$

Thus, semivariogram models must be conditionally negative definite, the condition being that the sum of the weights λ_α is zero. Therefore, the experimental semivariograms are usually fitted using only linear combinations of permissible models. The models include spherical, exponential, Gaussian and power models with nugget effects:

$$\hat{\gamma}_{sph}(h) = \begin{cases} c_0 + c_1[1.5(\frac{h}{a_0}) - 0.5(\frac{h}{a_0})^3] & 0 \leq h < a_0 \\ c_0 + c_1 & h \geq a_0 \end{cases} \qquad (3.7)$$

$$\hat{\gamma}_{\exp}(h) = \begin{cases} c_0 + c_1[1 - \exp(-\frac{3h}{a_0})] & 0 \leq h < a_0 \\ c_0 + c_1 & h \geq a_0 \end{cases} \qquad (3.8)$$

$$\hat{\gamma}_{gau}(h) = \begin{cases} c_0 + c_1[1 - \exp(-\frac{3h^2}{a_0^2})] & 0 \leq h < a_0 \\ c_0 + c_1 & h \geq a_0 \end{cases} \qquad (3.9)$$

$$\hat{\gamma}_{pow}(h) = c_0 + c_1 h^{\omega} \qquad 0 < \omega < 2 \tag{3.10}$$

where c_0 and c_1 are the nugget variance and structure variance, respectively, and $c = c_0 + c_1$ is the sill variance. a_0 is the actual range parameter for the spherical model and the effective range parameter for the exponential and Gaussian models. The effective range is defined as the distance at which $\hat{\gamma}(a_0) = 0.95 \cdot c$. ω is a power of this power model. When $c_0 = 0$, the equations above represent pure spherical, exponential, Gaussian and power model.

The nugget variance c_0 of a semivariogram can be inferred by the intercept of the fitted model and arises from measurement error and micro-scale variance (Atkinson, 1997; Goovaerts, 1997). When the experimental semivariograms are calculated using raster data, the nugget variance implies a noise term, that is, measurement error variance and within-cell variability (Wang et al., 2001d). For spherical, exponential and Gaussian models, the semivariogram values increase as the lag h increases and gradually reach to the maximum, that is, sill variance as h reaches to the range parameter a_0 (Figure 3.5). This implies that out of the range parameter, the spatial similarity disappears. For power model, the semivariogram continuously increases and does not reach a sill value.

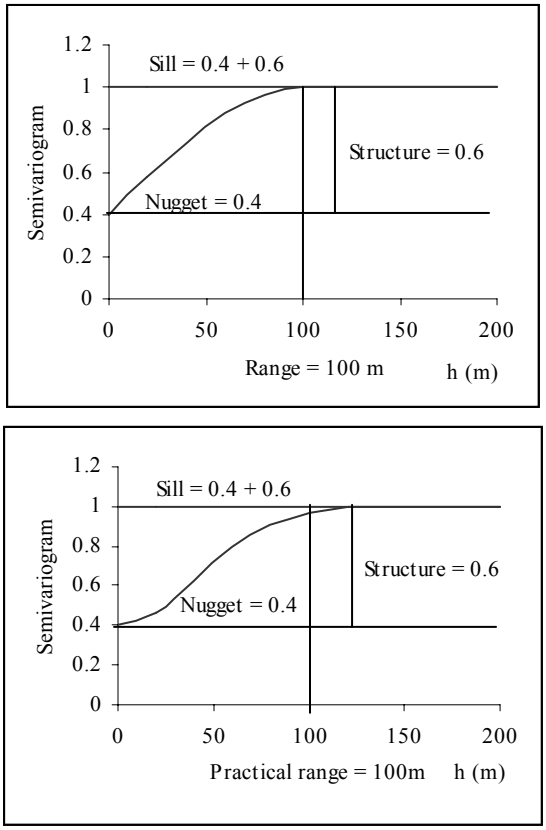

Fig. 3.5. Examples of Spherical (left) and Gaussian (right) models with their parameters.

In addition, different directions should be taken into account to determine whether the spatial variability is isotropic or anisotropic. Anisotropy means that semivariograms have different range or sill parameters in different directions. A method to detect the anisotropy is to calculate a semivariogram map centered at the origin of the semi-variogram and to derive a contour map of semivariogram values. The elliptical contour lines indicate anisotropy, while concentric contour imply isotropy. This method requires a data set of dense samples. Another alternative is to calculate experimental semivariograms in different directions and visually interpret the similarity. Semivariograms in different directions should be developed separately if anisotropy exists.

Sampling design

Sampling designing deals mainly with determining appropriate plots size and sample size. The average semivariance value at a lag of one pixel has been used to determine appropriate plot size and spatial resolution (Atkinson and Danson, 1988; Atkinson and Curran, 1997). In fact, its application is limited because of requiring a high dense sample. In this project

research, we improved this method by modeling the within plot spatial variability and regional spatial variability (Wang et al., 2001e). A plot size at which the within plot (micro) spatial variability and regional (macro) spatial variability of a variable is accurately captured simultaneously should be determined. The plot size should be an appropriate measurement unit for data collection and mapping. A semivariogram $\gamma_v(h)$ on plot size v can be derived from the punctual semivariogram by Journel and Huijbrets (1978):

$$\gamma_V(\mathrm{h}) = \overline{\gamma}(v, v_{\mathrm{h}}) - \overline{\gamma}(v, v) \tag{3.11}$$

where the first term at the right of the equality is the average punctual semivariance between two plots separated by a distance of h, that is, regional spatial variability; the second term is the average punctual semivariance within a plot, that is, within plot spatial variability. In practice, both semivariograms on the right of the equality in Eq. 3.11 are unknown. By sampling, these semivariograms can be obtained using experimental semivariogram Eq. 3.2. If spatial variability converges, the range parameter of spherical, exponential and Gaussian model provides the range of spatial dependence of the variable. Within the range, observations can be considered spatially dependent, and beyond the range, observations can be considered essentially independent.

The semivariogram models can be developed to describe the spatial variability within and between plots (Wang et al., 2001e). Within plot semi-variograms describes the within plot spatial variability over plot size, i.e., the length of transect line for LCTA plots. When using the spherical, exponential and Gaussian models, the within plot semivariance increases as plot size increases. The range parameter at which within plot semivariance reaches its maximum can be considered to be the maximum measure of appropriate plot size because the information beyond the range is independent (Wang et al., 2001e). This would correspond to maximizing the second term after the equality in Eq. 3.11.

Semivariograms can also be developed over the whole area by changing plot size. For each plot size, a regional experimental semivariogram is calculated and fitted using the permissible models mentioned previously. When the plot size increases, the modeled regional semivariograms vary in shape and parameters. For a specific variable, the structure variance increases and nugget variance decreases, and both gradually stabilize as the plot size arises. In remote sensing, this process implies enhancing structured variance and reducing noise - measurement error and micro variability, and this results in an improvement of correlation between field and remote sensing data. The plot size is considered appropriate when the ratio of the nugget variance to structure variance becomes stable (Wang et al, 2001e). This would correspond to stabilizing the first term after the equality in Eq. 3.11, that is, stabilize the estimate of regional variability. If there is a high correlation between field and image data, the appropriate plot size obtained using the field data will be consistent with the appropriate

spatial resolution using the images. This method is available for application of field data and remotely sensed data. When image data are employed, plot size means pixel or cell size, that is, spatial resolution. Thus, this method can be used to simultaneously determine plot size for ground data collection and spatial resolution for mapping.

Compared to traditional methods, the sampling design based on the theory of regionalized variables in geostatistics significantly reduced the number of samples with the same accuracy requirement because of considering spatial dependence of data of a variable (McBratney, et al., 1981; McBratney and Webster, 1981 * 1983). We have done the further improvement by introducing plot size and cost of data collection into the sampling design (Xiao et al., 2001). Kriging in geostatistics estimates localized unknown locations based on spatial variability of a variable and the estimates are unbiased with the sum of weights equal to one and minimizing local error variance. From Eq. 3.5, the estimation variance depends only on the separation distance $(u_\alpha - u_\beta)$ of data, and not data themselves. If the semivariogram is known, the kriging variances for any sampling schemes, that is, sampling distances, can be determined before sampling. Given a maximum error, the sampling distance can be determined and the sample size can be calculated with the interest area.

Moreover, a regional estimate obtained theoretically by kriging over the whole region is equal to the average of local estimates made for small neighborhoods (Journel and Huijbregts 1978). But the corresponding global estimation variance cannot be calculated simply by summing variances of local estimates because the neighboring locations are not independent. By an approximation, when S is a square with the observation point *u* at its center and side equal sampling interval, the variance of estimating its average value σ_s^2 equals to 2 times the average semivariance between the central point *u* and all other points in the square and minus the within square variance:

$$\sigma_s^2 = 2\bar{\gamma}(u,S) - \bar{\gamma}(S,S) \qquad (3.12)$$

If the area consists of n squares, the regional estimation variance σ_R^2 can be calculated (McBratney and Webster, 1983):

$$\sigma_R^{\ 2} = \frac{1}{n}\sigma_s^2 \qquad (3.13)$$

If semivariogram is known, the equations above can be solved for a range of sizes of square. The estimation variance is plotted against the sample size n and given a particular error, a sample size n can be determined. However, semivariogram function is usually estimated using experimental semivariogram that varies depending on plot size, as described above. If

the relationship between plot size and each parameter of the empirical semivariogram function obtained is established, on the other hand, plot size and sample size can be determined simultaneously (Xiao et al., 2001). Additionally, cost can be introduced into the analysis in terms of time for traveling between plots and measuring plots, and optimal plot size and sample size can be found.

Scale and resolution

Scale and resolution affects spatial features, patterns, and processes of ecological variables and resources in both space and time. Before conducting studies, we have to determine appropriate spatial and temporal scales or resolutions to be used. When multiple variables are mapped and overlapped and if the appropriate scales differ, interpolating or extrapolating results cross scales, that is, scaling up or down, is needed. Furthermore, the change of spatial information due to scaling has to be modeled and its effect on management decisions being made based on the changed characteristics of ecosystems and natural resources has to be studied.

The scale-related issues are complicated and a lot of studies are needed. In this project, we have had a good start by developing the methods that can be used to determine appropriate spatial resolution for mapping and to model loss of spatial information due to scaling (Gertner et al., 2001d; Wang et al., 2001e and 2001c). We have also suggested the possibility to develop a systematical methodology to account for the effect of scale and resolution in ecological modeling and resource management. Explicitly modeling the spatial variability of variables and processes is critical to systematical methodology. These spatial variability models will provide a basis to derive the methods that can be used to detect optimal spatial resolution, to infer spatial information cross scales, to measure change of the information due to scaling, and further to analyze the effect of scaling on management decisions.

We have developed a method that can be used to determine appropriate spatial resolution for mapping multiple vegetation types (Wang et al., 2001e). This method is the same as that used to determine appropriate plot size. An appropriate plot size means a measure or support unit used to collect ground data so that spatial variability of an interest variable can be captured. This implies that if the support size is employed as spatial resolution to map the variable, its spatial statistics can be well reproduced. Additionally, we have suggested a method to model change of spatial information due to scaling, including information loss from a finer resolution to a coarser and information increase by interpolation from a coarser resolution to a finer (Wang et al., 2001c). The method consists of deriving and fitting the semivariograms of the interest variable at different scales, then calculating changes of spatial information by differentiation and integration of the semivariogram models. This method can not only lead to

the change of spatial information but also detect differences of the changes at different directions because of anisotropy in spatial variability of the variable.

Spatial modeling and simulation

The shortcomings in smoothing of estimates and kriging variances limits the applications of kriging methods in spatial modeling and mapping for natural resources and ecosystems. Especially, kriging variances cannot be employed for spatial uncertainty budgets (Gertner et al., 2000; Wang 2000a). The methodology we developed for spatial modeling and mapping is based on various simulation algorithms (Gertner et al., 2001a and 2001c; Wang et al., 2000b, 2001a, 2001b, 2001f and 2001h). However, simple and ordinary kriging, indicator kriging, and co-located cokriging will be used to determine conditional cumulative density function (CDF) in various simulation algorithms. Before we present simulation algorithms, the kriging methods are introduced.

Kriging

Simple and ordinary estimators

Given n data $\{z(\mathbf{u}_\alpha), \alpha = 1, 2,\ldots,n\}$ of a continuous variable z, sampled and measured over a study area, the value of the variable at any un-sampled location *u* can be estimated. The basic kriging estimator is:

$$Z^*(\mathbf{u}) \cdot m(\mathbf{u}) > \sum_{b>1}^{n(\mathbf{u})} m_b(\mathbf{u})[Z(\mathbf{u}_b) \cdot m(\mathbf{u}_b)] \qquad (3.14)$$

where $Z^*(u)$ is a kriging estimate at a unknown location *u*, $\lambda_\alpha(\mathbf{u})$ the weight assigned to datum $z(\mathbf{u}_\alpha)$, $m(\mathbf{u})$ and $m(\mathbf{u}_\alpha)$ are the expected values of the variables $Z(\mathbf{u})$ and $Z(\mathbf{u}_\alpha)$. Given a neighborhood centered on *u* being estimated, the number of data involved and weights derived in the estimation differ from one location to another. Based on this equation, various kriging methods can be derived (Goovaerts, 1997).

When the mean $m(\mathbf{u})$ is considered to be known and constant throughout the study area, simple kriging (SK) is obtained. When the mean $m(\mathbf{u})$ varies depending on the local neighborhood, and is filtered from the linear estimator by forcing the kriging weights to sum to 1, ordinary kriging (OK) is derived. The simple and ordinary kriging estimators respectively become:

$$Z^{*}_{sk}(u) > \sum_{b>1}^{n(u)} m_b^{sk}(u)Z(u_b) + [1 - \sum_{b>1}^{n(u)} m_b^{sk}(u)]m \tag{3.15}$$

$$Z^{*}_{ok}(u) > \sum_{b>1}^{n(u)} m_b^{ok}(u)Z(u_b) \quad \text{with} \quad \sum_{b>1}^{n(u)} m_b^{ok}(u) > 1 \tag{3.16}$$

To derive the weights, a linear equation system is created. The system for simple kriging and its minimum error variances are:

$$\sum_{\beta=1}^{n(u)} \lambda_\beta^{sk}(u)C(u_\alpha - u_\beta) = C(u_\alpha - u) \qquad \alpha = 1,...,n(u) \tag{3.17}$$

$$\sigma_{sk}^2(u) = C(0) - \sum_{\alpha=1}^{n(u)} \lambda_\alpha^{sk}(u)C(u_\alpha - u) \tag{3.18}$$

The kriging estimators are exact interpolators in that they honor data values at their locations. For the other notations and kriging estimators, readers should refer to Cressie (1991) and Goovaerts (1997).

If P variables are jointly estimated conditioning to the sample data of the P primary variables and the data of Q auxiliary variables available at each location to be estimated, a hierarchy of the primary variables can be defined according to their importance and the estimation starts from the most important variable. A simple co-located cokriging estimator can be selected with its estimate $Z_p^{sck}(u)$ for the pth variable at a location u (Almeida, 1993):

$$Z_p^{sck}(u) = \sum_{\alpha=1}^{n(u)} \lambda_\alpha^p z_p(u_\alpha) + \sum_{q=1}^{Q} \nu_q^p x_q(u) + \sum_{i=1}^{p-1} \tau_i^p Z_i^{sck}(u) \tag{3.19}$$

where n(u) is the number of the sample data for the primary variables given a neighborhood. $Z_i^{sck}(u)$ (i = 1, …, p-1) is the previously estimated value for the primary variable i. λ_α^p, ν_q^p and τ_i^p are weights of the data of the primary variable p, auxiliary variable q and previously estimated variable i. The weights for the variable p are the solutions of a linear equation system consisting of n + Q + p-1 equations containing the auto and cross co-variances. Instead of directly modeling, the cross co-variances are derived by a Markov model:

$$C_{Z_p,X_q}(h) \approx \frac{C_{Z_p,X_q}(0)}{C_{Z_p,Z_p}(0)} C_{Z_p,Z_p}(h) \tag{3.20}$$

Indicator kriging

Indicator approaches do not assume any particular shape or analytical expression for conditional distributions. As a first step in using the indicator approach, indicator coding of original data is carried out. The probability function F(**u**; z|(n)) is then modeled through a series of K threshold values z_k:

$$F(\mathrm{u}; z_k \mid (n)) = \begin{cases} \Pr ob\{z(u) = z_k \mid (n)\} & for \quad categorical \quad \mathrm{var}\, iable \\ \Pr ob\{z(u) \le z_k \mid (n)\} & for \quad continuous \quad \mathrm{var}\, iable \\ k = 1, \ldots, K \end{cases} \tag{3.21}$$

where |(n) means the condition of n sample data. The K conditional CDF values are interpolated within each class (z_k, z_{k+1}] and extrapolated beyond the two extreme threshold values z_1 and z_k for a continuous variable. The indicator approach is based on the interpretation of the conditional probability Eq. 3.21 as the conditional expectation of an indicator random variable I(**u**;z_k) given the information (n): F(**u**;z_k|(n)) = E{I(**u**;z_k)|(n)} with Eq. 3.3 for indicator coding. The conditional CDF value F(**u**;z_k|(n)) can be obtained by kriging the unknown indicator i(**u**;z_k) using indicator transforms of the neighboring information. Different kriging methods lead to the respective indicator krigings. For example, simple indicator kriging is given as follows:

$$\left[\mathrm{F}(\mathbf{u}; \mathrm{z_k} \mid (\mathrm{n}))\right]^*_{\mathrm{sik}} > \left[\mathrm{I}(\mathbf{u}; \mathrm{z_k})\right]^*_{\mathrm{sk}} > \sum_{b>1}^{\mathrm{n}(\mathbf{u})} m_b^{\mathrm{sk}}(\mathbf{u}; \mathrm{z_k})\, \mathrm{I}(\mathrm{u}_b; \mathrm{z_k}) \;,\; m_{\mathrm{m}}^{\mathrm{sk}}(\mathbf{u}; \mathrm{z_k})\, \mathrm{F}(\mathrm{z_k}) \tag{3.22}$$

where $\mathrm{E}\{\mathrm{I}(\mathbf{u}; \mathrm{z_k})\} = \mathrm{F}(\mathrm{z_k})$ and $m_{\mathrm{m}}^{\mathrm{sk}}(\mathbf{u}; \mathrm{z_k}) > 1 \,.\, \sum_{b>1}^{\mathrm{n}(\mathbf{u})} m_b^{\mathrm{sk}}(\mathbf{u}; \mathrm{z_k})$

When the data of an auxiliary variable such as image data are available at all locations to be estimated, a co-located indicator cokriging estimator can be used to introduce image information into the estimation process of statistical parameters of conditional CDF in simulation algorithms. The co-located indicator cokriging estimator is:

$$[I(u; z_k)]^*_{oICK} = \sum_{\alpha=1}^{n(u)} \lambda_\alpha^{ock}(u; z_k) i(u_\alpha; z_k) + \lambda_X^{ock}(u; z_k) x(u; z_k) \tag{3.23}$$

where $[I(u;z_k)]^*_{oICK}$ is a co-located indicator cokriging estimate of a primary variable, $i(u_\alpha;z_k)$ the indicator value of the primary variable, $x(u;z_k)$ the datum of the auxiliary variable at the location u to be estimated. $\lambda^{ock}_\alpha(u;z_k)$ and $\lambda^{ock}_X(u;z_k)$ are weights for the primary and auxiliary variable.

The linear equation system for the solutions of the weights includes n(u)+2 equations. The equations depend on not only the co-variance functions ($C_I(h;z_k)$ and $C_x(h;z_k)$) of the primary and auxiliary variables at a separation distance h, but also the cross co-variance function between the two variables, that is, $C_{IX}(h;z_k)$. The co-variance function of the primary variable is derived by the modeled semi-variogram. The co-variance of the auxiliary variable and the cross co-variance can be approximated by the co-variance of the primary variable based on a Markov model:

$$\begin{cases} C_X(h;z_k) = B^2(z_k)C_I(h;z_k) \\ C_{IX}(h;z_k) = B(z_k)C_I(h;z_k) \end{cases}$$
$$\begin{cases} B(z_k) = m^1(z_k) - m^0(z_k) \\ m^1(z_k) = E[X(u;z_k) \mid i(u;z_k) = 1] \\ m^0(z_k) = E[X(u;z_k) \mid i(u;z_k) = 0] \end{cases} \quad (3.24)$$

Each coefficient $B(z_k)$ is determined by the difference between the two conditional expectations. The difference to derive the coefficients $B(z_k)$ for a categorical and continuous variable is that the condition $i(u;z_k)=1$ for indicator coding of a categorical variable is $z(u_\alpha) = z_k$ and the corresponding condition for a continuous variable is $z(u_\alpha) \le z_k$. For the details of the linear equation system, readers can refer to Goovaerts (1997).

Simulation

Simulation algorithms provide not only estimates but also estimation variances and co-variance at any locations. The estimation variances and co-variances vary space depending on sample data themselves in addition to data configuration (sample density and distance of an estimated location from sample data). These methods can thus be integrated with uncertainty budget methods for spatial modeling, mapping, and uncertainty analysis (Gertner et al., 2001c; Wang et al., 2000a and 2001a). Several simulation algorithms have been developed and used in the research project. An important alternative is joint sequential co-simulation with auxiliary data such as remotely sensed images and digital elevation models (Gertner et al., 2001c; Wang et al., 2001b). This method can be used for jointly mapping one or multiple variables with more than one auxiliary variable. The time required to run the co-simulations mainly depends on the number of variables to be estimated and the number of co-simulation

runs. Sequential Gaussian simulation is the basis of joint co-simulation algorithm and can be used for the simplest case of one variable with and without auxiliary data (Gertner et al., 2000; Wang et al., 2001f, 2001i).

The Gaussian simulation algorithms assume normal distribution of variables to be estimated. When multiple variables are simulated, multiGaussian model is assumed for the multivariate distribution, which also implies univariate normality. When the data of the variables are not normally distributed, a normal score transform (Goovaerts, 1997) should be performed so that the transformed data have means of zero with unit variances. These methods usually lead to underestimation in the areas with large values and overestimation in the areas with small values. In the cases at which extreme values are important, another alternative is needed, that is, sequential indicator simulation that can improve estimation of extreme values and at the same time does not require normal distribution of variables (Wang et al., 2000a, 2000b, 2001a). Mapping a categorical variable also needs this method (Wang et al., 2001h). Because a semivariogram for each class of categorical variable or each of several cutoff values of a continuous variable has to be developed, the simulation thus becomes complicated and uncertainty from modeling semivariograms will be propagated into predictions. When multiple variables that are spatially correlated with each other are considered, using this method is very difficult. Therefore, choosing correct method for an application is very important.

Sequential Gaussian simulation

Suppose that a study area consists of N pixels in a grid and that $\{Z(u_j'),\ j = 1,\ 2,\ 3,\ ...,\ N\}$ is a set of random variables defined at N locations, u_j'. Conditional to sample data, L joint realizations (l = 1, 2, …, L) for these N random variables can be generated with the sequential Gaussian simulation. A realization implies that each of N pixels of the grid is provided with an estimate, that is, a prediction map is obtained. In each simulation, the N-point conditional cumulative dense function (CDF) is expressed as the product of N one-point conditional CDFs given the sample data values and estimates obtained previously (Goovaerts, 1997).

In a simulation (Figure 3.6), a random path to visit each pixel of the grid only once in the area is first defined. We suppose that an estimate of the *i*th pixel to be visited has a Gaussian conditional CDF that can be determined by a mean and variance. The mean and variance are estimated using a kriging estimator and the modeled semivariogram given normal score transformed values of n sample data and all simulated values at the locations previously visited. From the conditional distribution, a value is drawn and transformed back to the original distribution data, and that value is further added to the conditional data set. The

process is repeated until all N pixels have been visited and provided with estimates. Running L times, each time with a possible different path to visit the N pixels, will lead to L realizations, that is, L maps, from which an expected map and prediction variance map for the estimated variable can be derived.

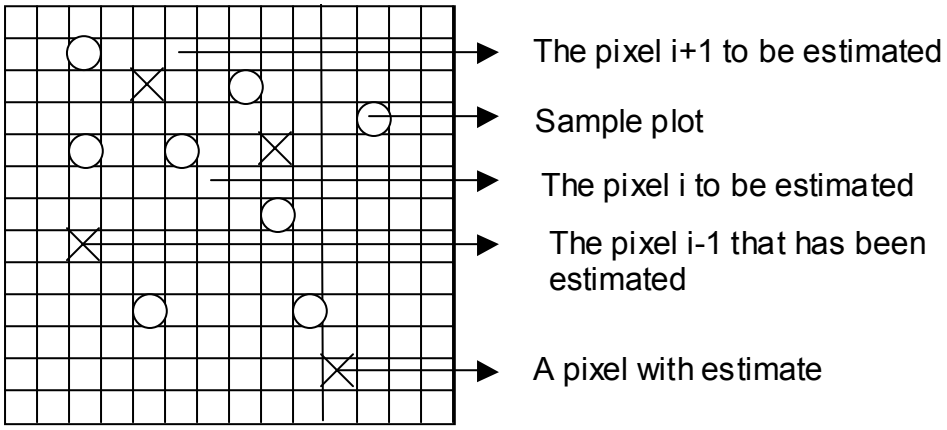

Fig. 3.6. One simulation run.

This method has been applied to generate prediction maps of rainfall-runoff erosivity factor (Wang et al., 2001f and 2001g), and soil erodibility factor (Gertner et al., 2000) for the case study of this project. Wang et al. (2001i) improved this simulation algorithm for mapping vegetation cover and management factor related to soil erosion by introducing Landsat TM images, which has become sequential Gaussian co-simulation. The co-simulation process is the same as above. However, the spatial cross variability between the variable and each auxiliary variable has to be modeled using Markov model described in kriging. In addition to sample data and previously simulated values, the co-simulation will also be conditional to the co-located auxiliary data. The co-located cokriging estimator is needed.

The conditional variances generated with the Gaussian simulation depend not on only data configuration but also data values, and in theory provide a more realistic assessment of uncertainty across space than the error variances obtained with kriging estimations (Gertner et al., 2000). As the number of L realizations increases, the variances decrease rapidly at the beginning, then slowly and gradually become stable. The number L, at which the estimation variances tend to become stable, can be chosen as the final number of realizations. For more details of mathematics on the sequential Gaussian simulation, the reader is referred to Chiles and Delfiner (1999) and Goovaerts (1997).

Sequential indicator simulation

The shortcoming of Gaussian simulation is that it may create under- and over-estimates when there are extremely large or small values. The advantages of sequential indicator simulation are that it does not require normal distribution of data and can handle different structures of the spatial variability. Moreover, indicator simulation is also needed for mapping categorical variables. With indicator simulation, the range of a continuous variable has to be discretized into several intervals and indicator transformation of original data must be done, which is called indicator coding. For a categorical variable, the indicator coding can be directly carried out according to categories. The indicator covariance or semivariogram models for these intervals are then developed and used for simulation. The sequential indicator simulation maintains the values of sample data at the sample locations and results in estimates of a variable at any non-sample locations of the study area using the sample data.

The sequential indicator simulation is similar to sequential Gaussian simulation (Goovaerts, 1997). The difference lies at that instead of deriving a mean value and variance of a normal distribution at each pixel to be estimated, K conditional CDF values $[F(u;z_k \mid (n))]^*$ (k = 1, …, K) are determined given the indicator transforms of original data and all previously simulated values using an indicator kriging. Because the probability estimates must lie in the interval [0,1] and their series has to be a non-decreasing function, the order relation deviations may be corrected to obtain a complete conditional CDF model using some interpolation or extrapolation algorithms. From the distribution function, a value is drawn and it becomes a conditional datum.

This method has been applied to map the topographical factor LS for prediction of soil erosion (Wang et al., 2000a, 2000b and 2001a). Wang et al. (2001h) further improved and used the method for mapping vegetation types at the case study of this project. At the case, the conditional CDF values determined are probabilities of occurrence of all categories at an estimated pixel. Landsat TM images are used to improve the simulation for classification, which becomes sequential indicator co-simulation. The spatial cross variability between the categorical variable and each auxiliary variable is modeled using Markov model described in indicator kriging. In addition to the sample data and previously simulated values, the conditional data include the co-located image data. Furthermore, an indicator co-located cokriging is needed to determine the conditional CDF values. A random number uniformly distributed in [0,1] is drawn and the estimated category at the location is derived based on the principle that if the random number is larger than the CDF value at the category k' –1 and less than or equal to the CDF value at the category k', the estimate is category $z_{k'}$.

As done in Gaussian simulation, independently repeating the indicator simulation or co-simulation L times with possibly different paths for each realization (run) lead to L maps.

The expected and variance maps can then be calculated. The uncertainties of the estimates at unknown locations can be expressed by conditional variances for continuous variables and by classification or misclassification probabilities for categorical variables (Wang et al., 2001h). The uncertainties depend on data configuration, data values used, and number of simulation runs (realizations), and if a continuously variable also on number of cutoff values and (Myers, 1997; Wang et al., 2001a).

When indicator coding of the sample data is done, the number of cutoff values, equal to the number of indicator semi-variograms used in simulation, will affect structure and information content of co-variance matrix. Generally, the more the indicator semi-variograms that are used, the more detail the information in spatial co-variance matrix and in theory more precise the estimated CDF will be. However, with more indicator semi-variograms, the computational time to perform the spatial simulation will increase and also more uncertainties might come from the semivariogram model parameters.

Joint sequential co-simulation

Suppose an implicit model of multiple variables:

$$Y = f(Z_1, Z_2, ..., Z_P) \tag{3.25}$$

Where Y is a dependent variable and Z_i is one of independent variables spatially correlated with each other. We will derive expected, variance and co-variance maps of all the variables using joint sequential co-simulation with auxiliary data.

Let us define a surface of the dependent variable for a study area and P sub-surfaces of the independent variables. The surface of the dependent variable can be derived using Eq. 3.25 from P sub-surfaces of the independent variables. The sub-surfaces are unknown. By sampling, however, we have obtained measurements of the variables. Based on the data set, co-located auxiliary data such as remotely sensed images, and the modeled semi-variograms, running a joint sequential co-simulation simultaneously generates all the sub-surfaces of P independent variables. Using the P sub-surfaces, the surface of the dependent variable is calculated. The co-simulation can be run L times, resulting in L sub-surfaces for each independent variable. Thus, L surfaces of the dependent variable can be obtained. Finally, an expected sub-surface for each independent variable and an expected surface for the dependent variable are derived as estimation of their truth surfaces.

The process above leads to L estimates for each variable at each location. Therefore, a matrix consisting of variances and co-variances of estimates at each location can be calculated as uncertainty measures. These variances and co-variances account for the uncertainties from

variation of the independent variables, their interactions, neighboring information, model parameters, and measurement errors and can be used to assess variance contributions of the components to variance of predicted dependent variable.

The joint sequential co-simulation process is similar to the co-simulation introduced above for estimation of a variable. In the joint sequential co-simulation, however, a hierarchy of the variables must be defined and a co-simulation starts from the most important one. For each co-simulation, a random path to visit each pixel of grid once needs to be set. At each pixel, an estimate of the first variable is first obtained by randomly drawing from a conditional CDF. The conditional CDF is determined by its mean and variance derived using a cokriging estimator in which auto and cross semi-variograms of the variables are included, given the sample data, previously simulated values, and co-located auxiliary data. The estimation is then performed for the second variable and the estimate of the first variable is also used as a conditional data. The co-simulation continues at this pixel until all the variables are estimated, and then moves to next pixel. The co-simulation is done when all the pixels are provided with estimates. The co-simulation process is repeated L times with possible different paths to visit the pixels of the grid, leading to L sub-surfaces for each variable. When auxiliary data are not used, the simulation is the same as above except for without auxiliary information.

The co-simulation algorithm is based on Bayes' axiom for conditional probability. That is, a joint P-variable CDF characterizing the P random events can be theoretically decomposed into a product of (P-1) univariate conditional CDFs and a marginal CDF. From the decomposition, the co-simulation can be developed to jointly simulate the P variables that are spatially correlated by drawing from the sequence of univariate conditional CDFs. Additionally, the cross semi-variograms between variables are generally approximated by a Markov model. For the details of the methods, readers can refer to Almeida (1993), Gómez-Hernández and Journel (1992), Goovaerts (1997), and Wang et al. (2001b)

Gertner et al. (2001a) and Parysow et al. (2001b) applied the joint sequential simulation without any auxiliary data for jointly mapping five soil properties and then deriving soil erodibility factor on soil erosion. Wang et al. (2001b) improved the joint sequential co-simulation with Landsat TM images and digital elevation models to derive expected and variance maps of soil erosion by jointly mapping rainfall-runoff factor R, soil erodibility factor K, topographical factor LS, vegetation cover and management factor C given one unit of support practice factor. Gertner et al. (2001c) integrated the joint sequential co-simulation and error budget for spatial prediction and uncertainty analysis of vegetation cover and management factor C by jointly mapping ground cover, canopy cover and vegetation height.

Accuracy assessment and uncertainty analysis

In terms of error and uncertainty, predicted maps should be assessed. Uncertainty of an estimate relates to probability at which the event will occur or the estimate falls within the confidential interval, and refers to a priori conditions. Thus, the estimate is only a rational guess as to the actual value. Error relates to a known outcome, a posteriori, and offers insights as to any potential biases with the sign of the discrepancy and magnitude. We use variances of estimates, probability for estimates falling confidential intervals, root mean square error and coefficient of correlation between estimated and observed values as measures of uncertainty and error analysis. Additionally, the measures for classifying a categorical variable include correct percentages and Kappa values, classification and misclassification probability. The potential sources of errors and uncertainties have been described in Figure 3.2. To assess predicted results, error and uncertainty analysis are carried out in several ways.

If observations of test samples are available, root mean square error and coefficient of correlation between the estimated and observed values of a continuous variable can be calculated. The root mean square error and correlation are mainly used to compare the results derived by different methods. For classification of a categorical variable, correct percentages and Kappa values are used to assess accuracy of classification and to compare results. The methods are applied to assess global accuracy of a study area. However, estimation or classification accuracy varies over space depending sample data (sampling and measure errors, sampling density), topographical features, landscape complexity, classification methods, and auxiliary data used (Steele et al., 1998; Wang et al., 2001h), and spatial accuracy assessment should be done. We have improved and developed following methods for spatial accuracy assessment and uncertainty budget based on classification and misclassification probability for categorical variables and estimation variance for continuous variables.

Spatial accuracy assessment

Variance and classification probability map by simulation

The various simulation algorithms described above generate L realizations of a variable, that is, L maps of estimates, thus provide not only an expected estimate at any unknown location, but also an estimation variance. The variances of estimates vary over space and directly indicate uncertainties of local estimates, and thus can be used to assess quality of a prediction map for a continuous variable. Similarly, classification probability maps of a categorical variable can be derived from L realizations (maps) (Wang et al., 2001h). If the simulation for classification is run 1000 times, and a pixel is classified into tree 600 times, grass 300 times,

and shrub 100 times, for example, the classification probability of this pixel is 0.6 for tree, 0.3 for grass and 0.1 for shrub. The spatial uncertainty information is documented. Thus, decision-makers can use the estimates with caution in terms of their uncertainties.

Misclassification probability map by simulation

Misclassification probabilities measure the probabilities that the predicted types are different from the true types. The misclassification probability varies over space depending sample data, topographical features, classification methods, auxiliary data used, etc. Based on the idea, we developed a method to do spatial assessment of classification, that is, by generating the misclassification probability map of a classification map by sequential indicator co-simulation with auxiliary data such as remotely sensed images (Wang et al., 2001h). This method suggests a significant improvement in accuracy assessment.

The classification of a categorical variable is first carried out using a sequential indicator co-simulation with remotely sensed data. Running this co-simulation L times results in L realizations, that is, L maps of classification. The expected classification map is derived based on prevailing category from L realizations at each location. Using a test data set, then, correctly classified probability can be calculated at the sample locations. If the simulation is run 1000 times and a sample plot that is dominated by tree is classified into tree 800 times, for example, the correctly classified probability is 0.8. On the other hand, the misclassification probability into other categories is 0.2.

The misclassification probabilities at the test sample locations can finally be interpolated to the unknown locations to generate a misclassification probability map. The interpolation is made using the sequential indicator co-simulation with the remote sensed data that have been used for the classification above. The misclassification probability is continuous and varies from 0 to 1. The range can be discretized into six intervals and five cutoff values, or ten intervals and nine cutoff values. The co-simulation for generating both classification map and misclassification probability map is similar.

User accuracy map by interpolation

Classification can be assessed using classification (producer) and application (user) accuracy, respectively. The accuracy measures can be estimated using bootstrap method at training data locations and then interpolated to unknown locations. We developed a method for the interpolation in which information from the satellite images is introduced into the interpolation process of user accuracy (Shinkareva et al., 2001). This method consists of three steps: classification, calculation of posterior probability, and derivation of user accuracy. This method also leads to an error partitioning by classes of a categorical variable.

In traditional classification, a discriminant function is derived by combining satellite image data and ground measurements. The discriminant analysis is then used for classifying pixels at unknown locations. The performance of the discriminant function can be evaluated by cross validation error matrix. It is computed by leaving out one observation from the training data set at a time, deriving a discriminant function based on the (*n*-1) remaining training points and classifying the left out observation. The procedure is repeated for all *n* observations and summarizing results of classification leads to a cross validation error matrix.

Entries n_{ij} in the error matrix corresponds to the number of plots of class *i* classified as class *j*. The diagonal of an error matrix shows the number of correctly classified plots. The entries in the error matrix can be divided by the corresponding column totals to compute sample conditional distribution of actual class membership given predicted membership, i.e. proportions $p(i|j)= (n_{ij}/n_{+j})$ for all *j* are computed. The diagonal of the resulting matrix is a measure of user's accuracy. Since the entries on the diagonal also represent correctly classified plots and have zero classification errors.

A posterior probability of a pixel belonging to a category i can be derived using the Bayes' theorem. The calculation of posterior probability for each pixel is done using the training data set, discriminant function, and a satellite image. This assumes that the prior probabilities are known. The posterior probabilities can be combined with the information of user accuracy from the error matrix to calculate a conditional probability that a pixel is of class *i* given that it has been classified as class *j*. The final user accuracy across classes is derived for each pixel.

Spatial uncertainty budget

The simulation algorithms above providing variance maps of estimates make it possible to do spatial uncertainty analysis with error budget methods. The traditional uncertainty analysis methods originally developed for an error budget of mean estimate of a population are widely used in uncertainty analysis for modeling of natural resources and ecosystems. However, they need to be improved so that an error budget can be done on the basis of pixel by pixel. The improved methods include Tayler series, response surface modeling, Fourier Amplitude Sensitivity Test (FAST), sequential sampling based method, and regression modeling. These methods have been applied to the case study of predicting soil erosion for spatial uncertainty budgets (Fang et al., 2001a, 2001b; Gertner et al., 2001a, 2001b, 2001c; Parysow et al., 2001; Wang et al., 2001a, 2000a).

Tayler series

The Taylor series methods are widely used in uncertainty analysis and recently have been expanded to spatial uncertainty analysis (Fang et al., 2001b; Heuvelink, 1998; Parysow et al., 2001). The methods do not need generating random numbers for computational experiments or simulation. As long as the partial derivatives, variance and covariance of the model input parameters are known, the uncertainty contribution of each input parameter as well as the uncertainty of the model can be computed (Dettinger and Wilson, 1981; Smith et al. 1992).

The first order Taylor Series method accounts for model uncertainty as the sum of individual contributions and co-contributions of the input parameters of the model:

$$\mathrm{var}(y) = \sum_{i=1}^{p} \mathrm{var}(z_i) \times \left(\frac{\partial y}{\partial z_i} \right)^2 + \sum_{i=1}^{p} \sum_{j=1}^{p} \mathrm{cov}(z_i, z_j) \times \frac{\partial y}{\partial z_i} \times \frac{\partial y}{\partial z_j} \qquad (3.26)$$

where y, z_i, and p are the response, the i[th] input parameter, and the total number of input parameters of the given model, respectively. Var(y),.var(z_i), and cov(z_i, z_j) are the variance of the model, variance of the i[th] input parameter, and covariance of the i[th] and j[th] input parameters of the model, respectively. $\partial y / \partial z_i$ is the partial derivative of the i[th] input parameter to the model. Individual contribution of an input parameter is the product of its variance and partial derivative and the co-contribution of a pair of input parameters to model uncertainty is the product of their covariance and partial derivatives (Gertner and Fang 2001). This method can handle interactions among the input parameters, however, assumes that the objective function is continuously differentiable.

Response surface modeling

The response surface modeling method is used to perform uncertainty analyses of complicated nonlinear models (Downing et al., 1985; Gertner et al. 2001; Iman and Helton, 1988). When nonlinear models are complicated, linear models can be used to represent them based on their responses surface relation. Then, the partial derivatives of the response surface models (linear models) can be easily obtained and the Taylor series method applied to investigate the uncertainty contribution of the model input parameters.

Assume the original nonlinear model is Eq. 3.25, that is, $y = f(z_1, \ldots, z_p)$, and y, ($z_1, \ldots, z_p$), and p are respectively the response, input parameters, and the total number of input parameters of the nonlinear model. By drawing a random sample of the input parameters and computing the model responses according to the random sample, a complete

data set can be obtained and thus used to fit a response surface model. The general form of the response surface model for our purpose is:

$$y = a_0 + \sum_{i=1}^{p} a_i z_i + \sum_{i=1}^{p} \sum_{j \neq i}^{p} b_{i,j} z_i z_j \qquad (3.27)$$

where a_i and $b_{i,j}$ are unknown coefficients that should be estimated using regression analysis with the obtained data. Based on this response surface model, the partial derivative of an input parameter is:

$$\frac{\partial y}{\partial z_i} = a_i + 2b_{i,i}\bar{z}_i + \sum_{j=2} b_{i,j}\bar{z}_j \quad , \quad i = 1, 2, \ldots, p \qquad (3.28)$$

where $\bar{z}_i$ is the mean value of the i^{th} input parameter of the original model. Applying the first order Taylor series method above, uncertainty contribution of the input parameters of the original nonlinear model can be obtained. Latin hyper cube sampling is used to generate random samples for this analysis method since it has been widely used in estimating the coefficients of response surface models.

Improved Fourier Amplitude Sensitivity Test (FAST)

FAST uses the behavior of the model variance to evaluate the variance contribution of the input parameters. It is a computationally efficient method that uses a small random sample to investigate the entire distribution of the input parameters. In FAST, Fourier coefficients are used to compute the proportional variance contribution (partial variance) of each input parameter. Cukier et al. (1973) and Collins and Avissar (1994) provided the details of method development and equations for sampling and computing Fourier coefficients and partial variance. Fang et al. (2001a) improved the sampling procedure for non-uniform distributions. The improved sampling procedure eliminated errors from the linear assumption and sequential sampling in the original sampling procedure. Wang et al. (2000a and 2001a) expanded the FAST to a spatial uncertainty analysis. Though the acronym FAST contains the word “sensitivity”, it does not estimate the sensitivity coefficients for the input parameters of the nonlinear model. This method assumes that all the input parameters are independent.

Sequential sampling based method

The sequential sampling based method investigates uncertainty propagation using the behavior of the model variance corresponding to the marginal distribution of input

parameters (Fang, 2000; Fang and Gertner, 2000; Jansen et al., 1999; Jansen et al., 1994; Sobol, 1993). In this method, a special random sequential sample needs to be generated. In the sequential sample, the first random vector (a vector containing p random numbers of the p input parameters of the model) is randomly generated. The second and thereafter random vectors are generated based on their immediately preceding random vectors. This is done by storing the random numbers of p-1 input parameters of the preceding random vector, generating a random number for a single input parameter given the p-1 input parameters, and combining the new random number with the stored p-1 input parameters to form a new random vector (Moriss, 1991; Sobol, 1993). In such a sequential random sample, the difference of a pair of immediate neighbors is just the random numbers of one input parameter. The difference of the model responses corresponding to one pair of random vectors is used to compute the variance caused by the change of the input parameter(s). Assume the original nonlinear model is:

$$y = f(Z) \tag{3.29}$$

where y , $Z = (z_1, \mathrm{L}\ , z_p)'$, and p are respectively the response, input parameter vector, and the total number of input parameters of the nonlinear model. With the notation:

$Z_i = (z_{1,i}, \mathrm{L}\ , z_{p,i})$ ' and $Z_i^{k+} = (z_{1,i}, \mathrm{L}\ , z_{k-1,i}, z_{k,i+1}, z_{k+1,i}, \mathrm{L}\ , z_{p,i})$ '

The variance of the model corresponding to the variation of the k^{th} input parameter is Eq. 3.30 that can be used to build error budgets.

$$\mathrm{var}(y)_{z_k} = \frac{1}{2N} \sum_{i=1}^{N} \left[f(Z_i^{k+}) - f(Z_i) \right]^2 \tag{3.30}$$

Regression modeling

Assume a multivariate model Eq. 3.25. The joint sequential co-simulation results in expected surfaces of the independent variables, their variance and co-variance maps (Wang et al., 2001b), and in addition to sample data, variation of variable themselves, and the interactions among them, the uncertainties of estimates are related to spatial information from neighbors used given a neighborhood. The uncertainties are propagated to the expected surface of the dependent variable. Spatial error budget has to be done so as to account for the effect of spatial information from neighbors. By improving a polynomial regression method proposed by Gertner et al. (1996), Gertner et al. (2001b and 2001c) presented a framework for this purpose.

The polynomial regression is integrated with the joint sequential co-simulation to make a spatial error budget for mapping multiple variables. By sampling the variance and co-variance maps of the dependent and its P independent variables, the variance and co-variances for sampled pixels are obtained. The cross co-variances to represent spatial information from neighbors are then calculated in terms of the variance and co-variance for the sampled pixels, and auto and cross semi-variograms. A polynomial regression model cab be further developed to establish the relationships of auto variances and cross co-variances from the variables, their interactions, and the components to account for the effect of neighboring information with the variances of estimates of the dependent variable. The initial regression model is a non-intercept model:

$$Var(Y) = \sum_{h=0}^{H} \sum_{i=1}^{P} \sum_{j \geq i}^{P} b_{ijh} Cov[Z_i(u), Z_j(u+h)] + e \qquad (3.31)$$

where Y, Z_i, P are the dependent variable, the i^{th} input component, and the total number of input components, respectively, H is the maximum distance of a center pixel from the neighbors. Var(Y) is variance of the dependent variable Y, and u a location to be estimated, pixel. The separation distance h varies from zero, meaning a center pixel itself to be estimated, to H pixels, meaning the neighbors having a distance of H pixels from the estimated center pixel. b_{ijh} is the coefficient of the regression model, and e is error term.

$Cov[Z_i(u), Z_j(u+h)]$ is the auto variance or cross co-variance of the independent variables Z_i and Z_j at a separation distance h of the estimated location from its neighbor. It is a traditional variance of a variable when i=j and h=0, implying uncertainty propagation from variation of the variable itself; it is a traditional co-variance of two variables when i≠j and h=0, implying interactions between two variables; it is a cross variance of a variable when i=j and h≠0, meaning effect of neighboring information of the variable itself; and it is a cross co-variance between two variables when i≠j and h≠0, meaning effect of neighboring information through interactions between the variables.

A stepwise regression is used to reduce the insignificant terms in the model. The model obtained by stepwise regression contains all the components that significantly contribute their variances and co-variances to the estimation variances of the dependent variable. The variance contribution from a component is the sum of the variance and co-variances related to the component. For example, the variance contribution from an independent variable at an estimated location is itself variance plus all the co-variances between it and other variables. At each location, the relative variance contribution for each component can be thus derived by calculating the total variance proportion of this component to the estimation variance of the dependent variable.

RESULTS AND DISCUSSION CASE STUDY AT FORT HOOD

Appropriate plot size and sample size

Based on the method developed by Wang et al., (2001e), we studied appropriate plot sizes for collection of field data of five vegetation cover types, including tree, shrub, grass, mixed land, bare land, and water. This method is based on field data and a geostatistical theory that spatial variability of a variable is divided into within support (plot) and regional spatial variability, represented by within support semi-variogram and regional semi-variogram. The range parameters of the within support semi-variograms implies the maximum range of appropriate plot sizes. The ratio of nugget variance to structure variance from regional semi-variograms at different plot sizes generally decreases from rapidly to slowly and gradually stabilizes as plot size increases. The plot size at which the ratio becomes stable can be considered appropriate.

The results show that the appropriate plot size varied depending on vegetation types (Wang et al., 2001e). It was about 60m for grass and shrub, 70m for forb and 80m for tree and half-shrub, and would not be less than 80m for woody. An integrated appropriate plot size for ground data collection was determined using overall vegetation cover and Landsat TM images. All six TM images led to an appropriate spatial resolution of 90m (Wang et al., 2001e). The result was reasonable partly because each of the images was an integrated model in spectral signals from the objects on the ground. On the other hand, the appropriate support size from the images should imply the appropriate measurement unit in the integrated spatial variability of the variables, thus might correspond with the maximum appropriate plot size. The spatial resolution should be applied for mapping multiple vegetation types. The comparison of the vegetation classification at different plot and image window sizes by cross validation proved the appropriate plot size and spatial resolution. This suggested that this method is practical to determine appropriate plot size for ground data collection and spatial resolution for mapping together.

This method of determining appropriate plot sizes for individual variables suggested a possible improvement in classification and interpretation of spectral mixtures due to cover types and extents. When using a fixed pixel size for classification, it was expected there

would be great variation in accuracy for different cover types. The possible improvement may be that the percentage cover map for each vegetation type is first derived using its appropriate plot size and spatial resolution by integrating remote sensing data and geo-statistical methods such as co-kriging. The cover maps are then overlapped and vegetation classification on the maps is carried out according to the defined classification rules. Further study for this idea is needed.

The appropriate plot sizes were also studied by a traditional method, that is, coefficients of variation. However, the plot sizes with the stable coefficients were much less than those by the geo-statistical method. The reason might be that the traditional method did not deal with spatial dependence.

In a sampling design, plot size deals only with within plot cost. The final plot size chosen and the cost within plots might be larger than those required for individual vegetation types, but this might be inevitable. This study was based on a given sample density of field plots. However, the sample density also affected the spatial variability and cost for collecting field data through spatial pattern of plots and travel time respectively. An optimal sampling design, cost and effectiveness analysis of the entire sampling strategy was done by Xiao et al. (2001).

The optimal sampling design developed by Xiao et al. took spatial variability of a variable into account and solved the estimation of appropriate plot size and optimal sample size by geo-statistical methods on the basis of efficient-cost. Meanwhile, the results were compared with those by traditional method without spatial correlation. It was found that the present plot size of 100m referring to LCTA data could be used to reveal the spatial variability but not cost-efficient. However, the combination of plot size and sample size may affect the precision of estimation and cost significantly. Therefore, cost introduced as a factor and then optimal sampling design was developed on the basis that estimated error and budget were considered simultaneously. The plot size and sample size with efficient cost by optimal method were thus found.

The traditional theories of survey sampling, which expects the independence of sampling unites with each other, might not work well for spatial sampling of continuous resources, since it does not take spatial dependence into account, thus leads to uncertainty. Semivariograms have indirect relations to both the sizes of the plot and sample. Based on semivariogram, kriging variances were derived and plotted against grid spacing. The grid spacing indicated the distance of sample plots. Thus, the sample size could be calculated.

Fixed plot size 100m, the kriging variance for overall vegetation cover was smaller than that from traditional approach. Therefore, kriging method was recommended as the basic method

of processing optimal solution. Because the different combination of plot size and sample size might have different precision and cost, the cost was taken into account and the appropriate integration of plot size and sample size was searched for in terms of accepted error and efficient cost.

The cost functions were derived and used in this case study. As an example, the cost analysis reported here dealt only with overall vegetation cover. The regression equations for the nugget, sill and range parameters of the semivariogram for overall vegetation versus plot sizes were developed to evaluate the cost by plot size changing from 10m to 100m with 1m interval.

According to the kriging variance and grid spacing (distance between plots), the corresponding sample sizes were calculated given the area of study region. It was found that the precision did not vary very much as plot size increased from 20m to 100 m, given a sample size in regional estimation. Plot size very slightly affected regional kriging standard error, but the sample size greatly did, given a cost.

In local estimation, the different plot sizes significantly affect local kriging standard error. For the same plot size, the kriging standard error curves had similar trend over grid spacing and sample size. With the plot size increasing, the local kriging stand error decreased from rapidly to slowly given a grid spacing and sample size. That indicates that the larger the plots, the higher the precision, while sample size did not improve precision very much. In terms of cost and precision, the plot size of 60 m was large enough to estimate local overall percent cover and it was about 20 m for regional percent cover.

Correspondingly, sample sizes for local and regional estimation was 40 and 200 for overall vegetation cover. However, the sample size by the traditional method differed significantly from that by kriging method. The optimal sample size by kriging was much smaller than that by the traditional method, which implied that the kriging was more cost-efficient.

So far the above methods for sampling design were applied to vegetation cover that affects vegetation cover and management factor C in this case study. Using these methods, the appropriate plot size and sample size for other input factors related to prediction of soil erosion, including soil erodibility factor K, topographical factor LS, and rainfall-runoff erosivity factor R, can be determined. It is expected that the appropriate plot size and sample size will differ from one input factor to another. However, more attention should be paid to the most sensitive factors – topographical factor LS and vegetation cover and management factor C to soil erosion. Because appropriate plot size corresponds with appropriate spatial resolution, furthermore, next section we will introduce the results of spatial resolution for mapping topographical factor LS.

Appropriate spatial resolution for mapping

There are two sets of empirical models involved in the USLE and RUSLE, respectively, which can be used to calculate the slope length factor L and steepness factor S with the field measurements of slope length λ in meters and slope angle β in degrees. There is a shortcoming of this method, that is, for converging and diverging terrain the empirical models does not differentiate net erosion and those areas experiencing net deposition. In order to improve this, a physically based topographical factor LS equation developed based on a digital elevation model (DEM) can be used to map the topographical factor LS (a product of L and S). However, the precision for predicting the LS factor is related to the DEM accuracy and spatial resolution, and the methods to derive topographical variables related to LS.

Wang et al. (2001d) investigated the use of DEM and appropriate DEM spatial resolution for mapping the LS factor, and modeled the loss of spatial variability due to data resampling. The DEM spatial resolution should be chosen considering simultaneously the required prediction precision and the detailed spatial information of the LS factor. In choosing a single DEM spatial resolution optimally for both requirements, a compromise may be needed, depending on the users' emphasis on one of the requirements or both. Global variance and semivariance at a lag of one cell can be used in combination to achieve the above purpose. In addition, modeling the experimental semivariograms and using them to estimate spatial variability loss due to data resampling can help users determine the appropriate DEM spatial resolution.

For the same spatial direction, the nugget variance and total variance of the modeled variograms generally decrease as cell size increases, while the range parameter generally increase. The more complex the topographic features, the larger the nugget variances and range parameters. In addition to the within-cell spatial variability, the nugget variances may be considered as an estimate of micro variability and noise caused by errors from elevation measurements, data resampling, models used, and calculation of the variables related to the LS factor. Developing a method to separate the noise from the within-cell spatial variability is important in order to determine an appropriate DEM spatial resolution.

In addition to entropy and global variance as a general measure of information loss due to scaling up (from a finer resolution to coarser), Wang et al. (2001d) developed a new method to directly measure the loss of spatial variability. This method is based on the modeled variograms and varies depending on the variogram model (e.g. spherical) chosen. Once a model is determined, the loss measure function of spatial variability can be easily derived and calculated by differentiation and integration. The results showed that the losses of spatial variability calculated by the new method are similar in three of the four directions, but different in one direction. This implies that the new method can reveal differences in spatial

variability and its loss due to data resampling in different directions, when anisotropy exists, while the existing methods including entropy and global variance cannot.

Most of existing DEMs have a spatial resolution of 30m by 30m. However, the resolution may not be insufficient for calculating the up-slope contributing areas when using the physically based topographical factor LS equation mentioned above to derive LS values. Gertner et al. (2001d) further investigated appropriate DEM spatial resolution by interpolating the DEM at 30m into 20m, 10m and 5m based on uncertainty analysis and error budget method to generate a topographical factor LS map. Because of IBM computers used, a small area of 10,020 m by 10,020 m was extracted from the DEM. In the small area, the slope, up-slope contributing area, LS factor and their variance maps were then calculated using physically based topographical factor LS equation. The accuracy and uncertainty of the maps at different spatial resolutions was assessed and compared in terms of root mean square error derived using field measurements, also based on spatial distribution and spatial variability of the predicted values, and significant difference test of the average values. The error propagation from slope, up-slope contributing area, and two model parameters to the prediction of LS factor was further modeled and relative variance contributions were generated using Fourier Amplitude Sensitivity Test (FAST). Through the procedure above, the effect of spatial resolutions was successfully illustrated in terms of prediction variance, main sources of uncertainty in predicting LS factor were identified, and selection of spatial resolutions was suggested. The results provided users and decision-makers with useful information in error management of soil loss estimation for options and applications of DEMs, and plans of agricultural and environmental management.

When LS factor map is derived using a DEM, the uncertainty depends, to a great extent, on the spatial resolution determining the accuracy of estimates of slope and up-slope contributing area related to LS factor in the equation of physically based topographical factor. In practice, the spatial resolution for most of DEMs available is equal to or coarser than 30m by 30m and this resolution are too coarser for spatial prediction of up-slope contributing area and further LS factor. The interpolation of elevation data into finer resolution is thus needed. But, the interpolation may lead to degradation of accuracy in elevation and thus in estimation of other variables. Therefore, it is important to select a good interpolation method.

According to studies by Mitášová and Mitáš (1993), we selected the regularized spline with tension and smoothing for the interpolation. The results in this study showed that the interpolation from the spatial resolution of 30m to 20m, 10m and 5m did not lead to significant difference of average values and variance of elevation compared to those from the original DEM. The spatial distribution of the interpolated elevation was similar to that of the original DEM, and spatial variability of the elevation overlapped each other at all four resolutions. These implied that the DEMs at finer spatial resolution by the interpolation

provided spatial information in more detail without degradation of accuracy compared to the original one.

On the other hand, the interpolation did result in different estimates of slope, up-slope contributing area, and LS factor. The effect of spatial resolution on the prediction uncertainty of the LS factor was assessed comparing the maps. According to the results in this study, statistically, the average values and variances of slope, up-slope contributing area and LS factor obtained at all four spatial resolutions were significantly different although the spatial distributions of the estimates were similar, that is, large or small estimates were consistently located at the maps of all the different resolutions. Moreover, the semi-variogram functions (measuring spatial variability) of these variables given a separation distance of data obviously differed while the function structures were similar.

The finer the spatial resolution, the larger the predicted slope values, their variance and semi-variogram given a separation distance of data. The reason may be that the shorter distance used to calculate the slopes due to smaller pixel size at the finer resolution lead to larger uncertainty. When the original DEM at 30m resolution was used, however, extremely large values of maximum value, estimated mean, estimation variance, and semi-variogram of up-slope contributing area was obtained. The interpolation from spatial resolution of 30m to 20m, 10m and 5m resulted in the rapid decrease of these estimates for up-slope contributing area, especially in the steep areas. This might be mainly because at the areas where there was high spatial variability of elevation, the finer spatial resolution made smaller the pixel size and boundary errors of watershed areas, and thus lower uncertainty in the prediction of the up-slope contributing area. These features above for up-slope contributing area could be applied to estimates of LS factor versus spatial resolution because of uncertainty propagation. That is, the finer the resolution, the smaller the mean estimate, its variance and semi-variogram of LS factor.

Using Fourier Amplitude Sensitivity Test (FAST), moreover, we modelled the uncertainty propagation from slope, up-slope contributing area, and two model parameters in the equation for calculation of LS factor to the prediction of LS factor. The results of variance partitioning suggested that given a spatial resolution, the uncertainty in predicting the topographical factor LS using a DEM mainly came from slope in the areas of gentle slopes and up-slope contributing area in steep areas. Two model parameters contributed little in terms of variance. Although the relative variance contributions of the four components were similar at different resolutions, the absolute variance contribution from slope slightly increased and that from up-slope contributing area extremely decreased as the spatial resolution varies from 30m to 20m, 10m and 5m. Thus, the total variance of predicted LS factor decreased rapidly with finer spatial resolution. As the predicted LS values rose,

additionally, the total variance and the partial variances from up-slope contributing area and slope increased.

The results above suggested that to derive a better slope map in terms of smaller estimation variance, the DEM at the spatial resolution of 30m can be used while in calculation of up-slope contributing area, the spatial resolution of the DEM used should be finer than 30m. Using the existing DEMs at spatial resolution of 30m might lead to extremely large estimation variance of up-slope contributing area and thus LS factor due to error propagation. For our particular case study, a DEM at the spatial resolution courser than 5 m could be considered problematic for the prediction of the LS factor.

The interpolation of a DEM to finer resolution for a large area, however, will result in the requirement of computers that are extremely fast and with large memory. One solution may be to divide the large area into several smaller areas, then the DEM for the smaller areas can be used to interpolate to finer resolution. However, it is necessary to further study the techniques for setting up the overlapping areas in order to avoid breaking a watershed into different small areas, and then for mosaicking the smaller areas together to get the whole area at a finer resolution.

Comparison of methods for mapping

Gertner et al. (2000) compared three geostatistical methods including ordinary kriging, sequential Gaussian and indicator simulation for spatial prediction and uncertainty analysis of soil erodibility factor K based on a data set from a very intensive soil survey (524 observations, 10 m by 10 m grid). Half the data was used for calibration, the other half used for validation.

Three spatial statistical methods produce similar prediction maps of soil erodibility K values and the spatial distribution of the predicted values is consistent with that of the model and test data sets, although there was slight overestimation when the K value is small and underestimation when the K value is large. Compared to these three spatial methods, the traditional point-in-polygon method results in smoothed spatial prediction and variance maps. At the same time, the use of published soil erodibility K values from soil surveys may lead to large over- and underestimation compared to the field sample K values.

According to the mean square error calculated from the test sample K values and their estimates, suggest that sequential Gaussian simulation is the best method for mapping the soil erodibility factor, then ordinary kriging, and finally sequential indicator simulation. The main reason may be that Gaussian simulation requires normal distribution of data sets and the

normal distribution of the model data set used has led to the most suitable use of Gaussian simulation. Theoretically, sequential indicator simulation is very flexible because the distribution of data set need not be predefined. However, unlike Gaussian simulation and ordinary kriging, indicator simulation needs several indicator semivariograms to be developed. The modeling of these indicator semivariograms can be complicated and can lead to additional errors and uncertainty. However, the variance estimates obtained using indicator simulation were consistent with the spatial variation of the data set, while those obtained by Gaussian simulation and ordinary kriging were overly smoothed. For ordinary kriging the reason may be that the error variances depend only on the data configuration. For the Gaussian simulation, the reason may due to two factors, only one semivariogram is used, and that the k value samples are geographically dense. With indicator simulation, using more than one semi-variogram results in modeling spatial variability close to the reality. This is true especially for the variables that are not normally distributed, such as topographical factor LS that has a reverse J shape distribution.

For the LS factor, Wang et al. (2000b) compared different geo-statistical methods including ordinary kriging, indicator kriging, and sequential indicator simulation. In previous studies related to mapping LS factor in the case study area, point-in-polygon and point-in-stratum methods were used. The traditional methods led to smoothing prediction values without uncertainty measures. Furthermore, the traditional methods usually result in underestimates of soil loss in sub-areas where soil loss is serious. The comparisons suggested that sequential indicator simulation was a better method for spatial prediction and uncertainty assessment of the topographic factors in the soil loss model RUSLE than ordinary and indicator kriging. The sequential indicator simulation provided not only reliable spatial conditional variance maps of the predicted values, but also probability maps for predicted values larger than a given threshold value. The simulation realization is conditional and the histogram of simulated values reproduces the declustered sample histogram. Moreover, spatial variability is modeled by reproducing the set of indicator covariance models for various cutoff values. The prediction is not smoothed and thus, spatial variability and uncertainty is modeled in more detail. Compared to the other geo-statistical methods used, the sequential indicator simulation can provide better results in the cases where the distance between sample points is relatively large, the sample data may not be normally distributed, and where extreme values are key factors for decision-making. Also, the simulation can easily integrate other variables into the conditional distribution used in sequential simulation.

We also presented a comparison of three methods for vegetation classification and accuracy assessment at Fort Hood (Wang et al., 2001h). The methods included a traditional image-aided classification with six original TM images, and two geo-statistical methods, that is, sequential indicator co-simulation with the ratio image of TM3/TM4 and sequential indicator simulation without TM images. Based on the percentages correct and Kappa values of the

classified test plots, and the spatial distributions of five vegetation categories in three classification maps, the sequential indicator co-simulation with the ratio image resulted in slightly better classification than the traditional method. The sequential indicator simulation without TM images was the worst. Moreover, the co-simulation made it possible to directly generate classification probability and misclassification probability maps. The classification assessment was thus improved by spatially investigating uncertainty in classification. The spatial assessment of classification can also provide users with detailed information on uncertainty when they use the product maps. The results suggested that the image-aided co-simulation method might be promising in vegetation classification and accuracy assessment.

Compared to the maps from the two methods with TM images, the classification map by the method without TM images had fewer pixels classified as tree and many more as grass. The two methods with TM images also led to higher percentages correct and Kappa values for the classified test plots, and more similar spatial distributions of the classified pixels to the sample plots used for developing the methods than the one without TM images. This indicates that the use of Landsat TM images significantly improved the classification. As expected, using the TM images made it possible to model the spatial trend of the vegetation cover categories in traditional classification through the discriminate function, and in the co-simulation through the spatial cross co-variance function, between the image data and the vegetation types. The trend models thus provided useful spatial information at non-sampled locations for the vegetation classification. Furthermore, using the Markov model might lead to a reasonable approximation for the cross co-variance between the image data and vegetation categories in the co-simulation. However, the approximation depends very much on the correlation between the primary and secondary variable.

Although the sequential indicator co-simulation with the TM3/TM4 ratio image produced only slightly better classification than the traditional method, the former created the classification probability maps for the five vegetation categories and the overall misclassification probability map. Because the co-simulation method generated many realizations (estimates) of the vegetation category variable at each location, the classification probability maps were direct measures of uncertainty of classification. As an uncertainty measure, especially, the misclassification probabilities were directly derived when the realizations were compared to the ground measurements. The misclassification probability maps obtained by interpolation showed the spatial uncertainty information of the vegetation classification at any location and the range of the classification errors, which is deemed a shortcoming for traditional classification assessment using error matrix.

The probability maps obtained in this study presented reasonable classification and misclassification probabilities over the study area and five vegetation categories. For example, grass was adjacently distributed at the southwest parts, and at which the co-

simulation method classified most of the pixels into grass at the probabilities higher than 0.8, and the misclassification probabilities were less than 0.2. The misclassification probabilities varied depending on the vegetation categories and might be related to other factors such as soil properties and geographical features. To generate the misclassification map, furthermore, the test plots and the sample plots used for the classification were together applied in this study because of a relatively small test sample. The purpose was to demonstrate the method. In fact, the observations to be employed for production of the misclassification maps should be from a test sample only. In the classification by the sequential indicator co-simulation, additionally, only one transformed image was used. Using more than one TM image might result in further improvement in classification. Therefore, further studies are needed to recommend this method in vegetation classification and accuracy assessment.

Using a sample data set and a scene of six Landsat TM images, Wang et al. (2001i) compared three traditional and three geostatistical methods for mapping vegetation cover and management factor C for the USLE in soil loss prediction. Three traditional methods were typically point-in-polygon or point-in-stratum, that is, vegetation classification with pixel value assignment using (i) average cross category; (ii) linear regression model cross category; and (iii) log linear regression models cross category. Three geostatistical methods were (i) co-located cokriging with a TM ratio image; and sequential Gaussian cosimulation (ii) with and (iii) without the TM ratio image. From all 215 sample plots, 31 plots were randomly selected and used as the test data set. The remaining 184 sample plot data were used for developing spatial interpolation models. For the traditional methods, the image data used were all six TM bands. For two geostatistical methods, the image data employed were a ratio image having the highest correlation with the C factor values.

The coefficient of correlation between estimates and observations varied from 0.4888 to 0.7317, and the root mean square error (RMSE) from 0.0159 to 0.0203. The sequential Gaussian cosimulation with a TM ratio image resulted in the highest correlation and the smallest RMSE, and reproduced the best and most detailed spatial variability of C factor. This method may thus be recommended for mapping the C factor. It is also expected that this method can be applied to image based mapping in other input factors with normal distribution for USLE or RUSLE and also other disciplines. The vegetation classification with linear regression was the worst.

Although it is easy to obtain remotely sensed data now, many investigators still map natural resources using geostatistical methods without any auxiliary data. Wang et al (2001i) showed that the simulation without TM images resulted in much worse prediction than the co-located cokriging and sequential Gaussian cosimulation with the ratio image 5. The simulation without TM images created even worse estimates than two traditional methods, vegetation classification with average and log linear regression. As expected, the TM images and cross

semivariogram between the image data and the C factor values provided useful spatial information at the non-sampled locations in terms of coding spatial variability of the C factor. In other words, geostatistical methods without any auxiliary data should be used with caution for mapping natural resources. Furthermore, using Markov models might lead to a reasonable approximation of the cross correlogram. However, the approximation depends very much on the correlation of the primary and secondary variables.

Compared to the three traditional methods, two geostatistical methods, i.e., the co-located cokriging and Gaussian cosimulation with the ratio image 5 reproduced better and more detailed spatial variability of the vegetation cover C factor. At the same time, both gave uncertainty measures, that is, error variances at the non-sample locations and areas. Theoretically, the co-located cokriging, as an interpolation method, aims at providing the best estimates at every location, and does not care about spatial variability. On the other hand, the Gaussian cosimulation tries to reproduce spatial variability and probably may not result in the best predictions. In this study, the co-simulation led to slightly better estimates than the co-located cokriging. The differences may probably be mainly due to the normal score transformation done and different Markov model used for the co-simulation. Although the co-simulation was about ten times more expensive than the co-located cokriging in terms of computing time, the former was very worthwhile in this study because spatial variability was very important in prediction of soil erosion and uncertainty analysis.

Additionally, a simulated value at a non-sample location was drawn from conditional cumulative density function derived conditional to the sample data, the previously simulated values and the image datum at this location. Thus, the Gaussian co-simulation with the ratio image 5 avoided illogical estimates such as negative and extremely large values, which deemed to be a shortcoming for two traditional methods with linear or log linear regression modeling.

In this study, the values of vegetation cover C factor from the sample data were assumed to be the observations. In fact, the values were calculated as a function of ground cover, aerial cover and minimum average height of vegetation. The spatial uncertainty and error propagation from these three variables and the function parameters to the C factor prediction was not analyzed. This was done and showed in other articles.

Wang et al. (2001b) developed a joint sequential co-simulation to jointly create prediction maps of soil erodibility factor K, topographical factor LS, and vegetation cover factor C. In the co-simulation, the factors were defined as primary variables, while the Landsat TM images and slope map were defined as secondary variables. The primary variable data were available only at the sample locations and the secondary variable data were available over the study area at a grid spacing of 90 m by 90 m. The prediction maps were also produced by a

traditional stratification with same TM images and slope map. The two methods were compared in terms of statistical parameters and spatial distribution of observed and estimated values. The rainfall-runoff erosivity factor R was not considered correlated with other factors and was thus simulated independently without any auxiliary data.

The joint sequential co-simulation led to all means of factors LS, C, and K falling into the confidence intervals at the probability of 95%. The coefficient of correlation between the estimated and observed values was 0.4589 for factor LS, 0.7093 for factor C, and 0.4053 for factor K. The spatial distribution of the estimates was consistent with that of the observed values. For example, the co-simulation created large estimates of factor K at the west areas and small estimates at the east areas as the observed values shown.

Compared to the co-simulation, the stratification only estimated the average of factor C into the confidence interval and resulted in lower correlation between the estimated and observed values. Furthermore, the stratification might underestimate the factors at the areas with large observed values and overestimate them at the areas with small observed. On the other hand, the stratification might spatially smooth the local estimates.

For both co-simulation and stratification methods, however, the correlation between the estimated and observed values was low for factors LS and K. This was mainly due to the low correlation between the factors and the slope map and TM images used. In addition to the methods used, on the other hand, the correlation between the estimated variables and auxiliary variables used is the basis on which accuracy of jointly mapping multiple variables can be improved. This is true especially important to improve local estimates and to reproduce the cross-spatial variability of the variables.

The co-simulation jointly created a set of estimation vectors for the factors, thus an expected vector and covariance matrix at each unknown location. The probability maps for the expected estimates being larger or smaller the given threshold values can also be derived. The variance and co-variance maps can further be used as the input information for spatial error budget of predicting soil loss. That is, the relative contributions of the factors and their interactions to the uncertainties of predicting soil loss can be determined. The uncertainty measures thus provide decision-makers with useful information to assess the risk of the decisions being made. This is deemed to be an advantage of simulation based methods compared to traditional stratification.

Mapping soil erosion and spatial uncertainty

Rainfall-runoff factor R

Wang et al. (2001f and 2001g) generated the rainfall-runoff erosivity factor R map and analyzed its spatial prediction uncertainty using a sequential Gaussian simulation without any auxiliary data. Within Fort Hood, there were no rainfall observation stations. It was thus necessary to use the data from the rainfall observation stations around this area. A total of 248 rainfall stations were used and they were located at Texas and those states around Texas. Out of the stations, 30 stations were sampled at random and used as a validation data set and the left ones were used to develop the models.

Because the rainfall stations at the expanded area were not systematically located, the data sets of the rainfall and runoff erosivity R factor were first de-clustered. Normal score transformation of the original data was done in order to make the transformed data normally distributed. The spatial variability of the transformed data was then modeled using semi-variograms for annual, seasonal, and half-month rainfall-runoff erosivity respectively. Experimental standardization semi-variograms were derived and fit using authorized models including spherical, Gaussian, exponential and power models. Most of the semi-variograms were best fit by Gaussian model.

The spatial and temporal prediction and uncertainty analysis of annual, seasonal and half-month R factors was further carried out using sequential Gaussian simulation for the large rainfall station area at two dimensions. The simulations were tested using validation data sets and prediction errors were calculated. The spatial and temporal variation of the predicted values was analyzed in terms of variance and error. The prediction and uncertainty maps for Fort Hood were extracted from those making up the large rainfall station area. The results were compared with these obtained using traditional isoerodent maps.

The sequential Gaussian simulation provided the spatially and temporally predicted values and their uncertainty measures in terms of prediction variances for rainfall-runoff erosivity R factor in prediction of soil loss at the unknown locations and areas. The spatial and temporal distributions of the predicted values were similar to the observed data from the rainfall stations. This method can thus be recommended as a monitoring and mapping strategy for spatial and temporal prediction and uncertainty analysis of rainfall-runoff erosivity R factor in prediction of soil loss.

The rainfall-runoff erosivity R factor is an important variable in the prediction of soil loss. However, it is usually difficult to derive the R factor in the areas where there are no rainfall stations. Traditionally, the most widely used method is to interpolate the R factor values from

the isoerodent maps where R factor values are assumed constant over time. Global climate change may result in the false assumption (Nearing, 2001). This method presented here suggested the possible improvement in deriving the R factors.

In fact, the results showed that the average estimates by the simulation for annual, seasonal, and half-month rainfall R factor fell into the confidential intervals, while the annual rainfall R factor estimate by the isoerodent map was out of its corresponding interval, had a serious and systematical negative bias. The annual rainfall R fcator obtained by the simulation for the area of Fort Hood without any rainfall stations varied from 350 to 376 falling into the R factor values of four rainfall stations around it, but much higher than the R factor of 270 based on the isoerodent map.

The results in this study also implied that the annual rainfall and runoff erosivity R factor had large spatial variability over space. Even within a relative small area such as Fort Hood with an area of 87,890 ha, the spatial variability may not be neglected. This suggests that it should be very careful to use a constant R factor over space for a specific area. Additionally, there was a high temporal variability of the R factor in the time series of seasons and half months. As expected, the summer had the largest R factor values, then autumn, spring and winter. The half-month rainfall R factor increased from January to June, then fluctuated and decreased slowly to October, and after that tended to a rapid decrease to December. This implies an importance of vegetation cover to reduce soil loss in summer by cutting down water runoff.

When an isoerodent map is used to estimate the rainfall R factor, moreover, its uncertainty is unknown. This simulation method gave estimates with their variances at any unknown locations. Where the rainfall stations used for model development were dense and the rainfall R factor was low, the small variances turned out, and otherwise large. Thus, the rainfall R factor estimates can be applied carefully by decision-makers based on assessment of their uncertainties.

Soil erodibility factor K

Soil erodibility may be defined as the inherent susceptibility of the soil to be lost due to erosion. The water erosion model RUSLE (Revised Universal Soil Loss Equation) is partly a function of soil erodibility, which in that model is also known as the K factor. The National Cooperative Soil Survey (NCSS) provides information about this factor by assigning soil series (minimum mapping unit) one value of K, which in turn represent classes of soil erodibility. Thus, information contained in those surveys assumes that K factor values remain unchanged both across whole soil series and over time, and are mostly free of estimation errors (except for grouping error, which arises from clumping values into classes). However, evidence provided by soil science literature suggests that those assumptions may

not hold. Although prediction of the K factor traditionally may not bring a large amount of uncertainty into the prediction of soil erosion compared to other factors, on the other hand, the uncertainty analysis to its prediction is necessary to overall uncertainty budget. Therefore, the studies we have completed for the K factor included assessing uncertainty of soil erodibility in the national cooperative soil survey (NCSS) in a small area using a high dense soil sample and the whole Fort Hood using the existing LCTA sample, respectively, and an uncertainty budget of a soil erodibility map for Fort Hood by jointly sequential simulation of five soil properties and regression modeling.

Parysow et al (2001a) evaluated variability and uncertainty in the K factor as reported in the NCSS soil surveys in a small area of 230m by 230m at the southwest and cross two counties within Fort Hood using a high dense soil sample. Within the small area, Parysow et al. collected 524 soil samples in late summer of 1998, following a square grid whose points were 10m apart of each other. After laboratory analysis, K values were obtained for each of those points and then compared with the K values published in the NCSS soil surveys.

Several important results were obtained in this study. First, assuming that one K value could be considered representative of each series, sample results do not support concurrence with the information provided by the NCSS. This fact is apparent by the highly significant differences between the sampled mean and NCSS K values for the three soil series analyzed. The direction of these differences for Coryell County do not suggest a specific pattern in relation to the information provided by NCSS since we found both a positive (Krum) and a negative (Brackett-Topsey) average difference. The difference for the only series analyzed in Bell County turned out to be positive, although we have to be cautious on this statement due to the small sample size that result was based on.

Secondly, the assumption that each series might be represented by only one K value does not seem to agree with the sample results either. In fact, it is apparent that there exists considerable variation within each of those series as shown by the estimated coefficients of variation. Whereas the only exception to that statement might be found in the Denton series, it is worth noting here that the estimate of variation for Denton is based only on eight samples collected in a small area. Based on the fact that small areas tend to be more homogeneous, we can expect this estimate of variation to be lower than the estimate that would have likely been obtained by sampling a larger area of that series. Additionally, the trend observed in the sampled K values supports the fact that soil characteristics tend to vary smoothly rather than presenting sharp changes that would coincide with soil series boundaries. Nevertheless, even using discrete mapping units we would expect sampled values to approximate the NCSS K in areas farther away from soil series boundaries (which should represent a purer form of the series), and become fuzzier by intermingling with values of the next series as distance to the boundary decreases. However, the trend suggested by the

data in this study seems to behave in an opposite fashion with respect to that view. As shown in the results section, this trend causes sampled values to depart from NCSS values as the distance from the soil series boundary increases. Results also suggest that this inverse behavior appears to apply within the Brackett-Topsey association series. In that association, portions located in lower areas (closer to the soil series boundary with Krum in this case) usually correspond to the Topsey series, which was assigned a K=0.32, whereas higher lands usually belong to the Brackett series, having a K = 0.17. As seen in this study, the trend would present the opposite pattern as that suggested by the description of the association.

We would like to emphasize that NCSS has made a significant contribution to our understanding of soil composition by comprehensively surveying soils across the nation. Even though the specific information provided in those surveys about soil erodibility do not seem to agree with the results of this study, values originally proposed by the NCSS surveys are not necessarily erroneous. In fact, the sampling phase of this study was carried out in 1998, whereas the Coryell County survey was conducted in 1985 and the Bell County survey in 1977. Therefore, besides any possible problems or limitations in K factor estimation by the original soil surveys such as soil series misclassification, misrepresentation of assigned erodibility factors, or lack of accounting for intraseries variation, other factors such as compaction and/or erosion of whole soil layers over time may have changed soil properties since the time the original surveys were conducted.

Based on the results found in this study, it would appear that employing the K values reported by NCSS for making soil erosion predictions with RUSLE would cause considerable uncertainty in those predictions. However, in light of the evidence showing that K values tend to vary considerably and in a smooth fashion, the application of geostatistical methods may prove to be a valuable modeling tool, and thus contribute to reducing uncertainty in erosion predictions. Finally, it is worth noting that changes in soil properties over time may prove to be a considerable force affecting soil erodibility in lands exposed to disturbance. This scenario would in turn call for the implementation of a monitoring strategy of soil properties for periodically updating information on this critical factor to sustainable land management.

Wang et al. (2001c) analyzed the uncertainty of the published K values for the whole Fort Hood. The methods used for assessing the uncertainty included statistically comparing the published and sampled soil erodibility K values in terms of their differences, analyzing error properties of the published K values, and performing spatial prediction and uncertainty analysis of the K values with the sample data using sequential Gaussian simulation. Soil samples were collected in summer of 1999 from 186 LCTA plots over the area and measured at a laboratory for soil properties including: %silt, %sand, %clay, %organic matter, and classes for structure and permeability. The soil erodibility factor K values of these soil

samples were calculated using K factor equation. The results showed that unbiased estimation of soil erodibility K values using the published information was possible only at a few sample locations and for a few soil types. Biased estimation, especially underestimation, was observed at most of the sample locations, in most of the study area, and for most of the 25 soil types.

For the whole area, using the published K values led to the underestimation of soil erodibility. Thus, the published soil erodibility K values should be used with caution. This underestimation can be explained by the change of soil properties over space and time (see p.133, Hudson, 1995). The published soil erodibility K values were determined twenty years ago using average values within the same soil series. In fact, the soil erodibility K values within a soil series varied within a certain range, and using an average value might thus result in uncertainty.

On the other hand, the change of soil properties over time was caused by many factors such as plants, climate, human activities, and so on. Because off-road vehicular impact activities took place in recent years in this area, we looked into the correlation between the cumulative disturbance (Demarias et al., 1999) caused by the off-road vehicular impact activities and sampled soil erodibility K values, and their differences with the published K values. The correlations were found to be weak. That is, the soil erodibility increased due to many factors or their integrated effect, but not solely from these activities.

Determining the soil erodibility factor (K) directly from soil loss data collected from repeat measurement plots measured over the long term (over 20 years) is the most reliable method for assessing soil erodibility (Wischmeier and Mannering, 1969, SWCS, 1995, Renard et al., 1997). This method, however, is very expensive and can take a long time to obtain results, which can be impractical for many situations (Renard et al., 1997). The second alternative is using the published soil erodibility K values by USDA-Natural Resources Conversation Service (USDA-NRCS). Its advantages include low cost and ease of acquisition of soil erodibility K values. However, the assumption that soil erodibility K values are constant over time and the use of an average K value for each soil type (class) introduces uncertainty into the estimate of soil erodibility. In addition, using the published K values introduces spatial discreteness in the soil erodibility values.

Another alternative to determine soil erodibility K values is the application of geostatistical methods such as sequential Gaussian simulation with soil erodibility K values from soil samples. The sequential Gaussian simulation produced not only a spatial prediction map of soil erodibility K values, but also, uncertainty measures, prediction variance images and probability maps for a specific feature such as soil loss larger than a given value. The spatial distribution of soil erodibility K values predicted using this method is very similar to that of

the soil samples. The variance images and probability maps of the predicted values measured the uncertainty caused not only by the variation of the soil erodibility values based on the soil samples, but also by the spatial orientation of the sample plots. Thus, the procedure used in this study for spatial prediction and uncertainty assessment of soil erodibility can be recommended as a potential monitoring strategy to periodically update soil erodibility K value maps. When this method is applied to all factors in the RUSLE, the uncertainties obtained can provide decision-makers with useful information to reduce the risks in soil and land management

Using the soil sample of the whole Fort Hood mentioned above, Parysow et al (2001b) evaluated the use of joint sequential simulation for mapping soil erodibility, as well as to partition the individual and joint variance contribution of soil properties used to predict soil erodibility. Our study area for the simulation consisted of 5,776 square cells (76 rows by 76 columns), the side length of each cell being 200 meters. We carried out both independent and joint sequential simulation to generate spatially-explicit predictions and variance of all soil properties as well as covariance between pairs of soil properties for each cell. We also obtained estimates of soil erodibility (K factor) and its variance for each cell as a function of the soil property predictions generated across all simulation runs.

The results showed that incorporating spatial cross-correlation information through joint sequential simulation reduced the average predicted variance of the K factor to less than half the variance produced assuming independence between soil properties. Although the range of predicted K values between independent and joint sequential simulation were similar, results from the latter presented significantly less variability and a clearer spatial pattern than those from the former. Therefore, our results agree with the theoretical postulates that favor including cross-correlation information as a more precise alternative to estimating spatially-explicit variables. Furthermore, the variances of predicted K values by independent simulation appear randomly distributed in space, whereas those produced by joint sequential simulation vary depending on the K values of the samples and the distances of locations to be estimated from those samples.

The net result of both inherent variance/covariance and error propagation sensitivity through the K factor equation resulted in individual and pairs of input soil properties having a markedly different contribution to K factor variance. Individually, Structure contributed the least (6.53%), whereas very fine sand plus silt contributed the most (46.19%) to the K factor variance. Thus, improving accuracy of data measurement and semivariogram modeling accuracy as well as increasing sample size of very fine sand plus silt may cause a significant reduction in the uncertainty of the predicted K values. It is worth noting that in this application all soil properties are estimated from the same soil samples and, therefore, increasing the sample size of one would also increase the sample size of the others.

Furthermore, the variance contributions from the interactions between soil properties may be either positive or negative, depending on the spatial correlation between those properties. Jointly, sand/very fine sand plus silt caused the largest reduction (-19.19%), whereas permeability/structure contributed the most (9.32%) to K factor variance. This implies that accurately modeling the cross semivariograms of two pairs of the variables may lead to significant reduction of uncertainty for spatial prediction of soil erodibility. Our finding showed that the variance percent contribution of soil properties varied across space. However, very fine sand plus silt contributed the most uncertainty to the variance of predicted K values across the whole area. This was probably because the study area was small, resulting in a fairly homogeneous distribution of soil properties.

Taylor series expansion provided a very close approximation to the K variance obtained from joint sequential simulation. The resulting mean difference between estimated variances by Taylor series expansion and the variances from the simulation was very close to zero, suggesting that positive and negative differences virtually canceled out across the study area. More specifically, since the actual mean difference was very slightly positive, we can infer that the approximation resulted in slightly lower variances than those from the joint sequential simulation. Likewise, the minimum and maximum differences (as well as the frequency distribution of differences) support a minor tendency toward accumulation of positive differences. Although inclusion of higher-order terms in the Taylor series expansion might provide an even better approximation, the potential gain would be too small to justify its implementation.

For the whole Fort Hood area, Gertner et al. (2001a and 2001b) did the uncertainty budget for prediction of soil erodibility by integrating a joint sequential simulation with uncertainty analysis procedure – regression modeling. The data set used was the same as mentioned above. The cross-spatial variability between the variables was introduced into the joint simulation, which should be basis on which spatial uncertainty analysis was performed in the error budget. The joint simulation well reproduced the joint spatial statistics of the variables. Figure 4.1 shows the predicted maps of these soil properties. The spatial distribution of predicted K factor values is more similar to that of predicted soil sand and very fine sand than those of other soil properties. This joint simulation also led to prediction variance maps of all the variables and covariance between them. As an example, the variance maps of predicted K factor, sand, structure, and the covariance maps between them are given in Figure 4.2. The spatial distribution of the variances of predicted K factor values is more dependent on that of the variances and covariances of predicted sand values than on the corresponding distribution from predicted structure values. On the other hand, more contribution to variances of predicted K factor values may come from soil sand variable.

The results of the joint sequential simulation for several soil properties in this case study showed that both the variances of a spatially explicit model and elements of covariance matrix of the model components might have symmetric and approximate normal distribution. Assumption of normality can thus be acceptable for the distribution of variation of spatial simulation. The approximate normal distribution makes it possible to analyze the relationship between variation of models and that of model components using Ordinary Least Square.

In initial model of stepwise regression to construct uncertainty budget models, it is practical to include auto/cross covariance terms from the pixels of the first 3 neighbor groups. Since spatial correlation depends on distance, the closer neighbors have higher priority to be introduced into the initial regression model. Including more neighbor groups would reduce information from the immediate neighbor groups because of spatial correlation. Both final regression models and spatial uncertainty partitioning showed that the first three neighbor groups are sufficient for an initial regression model.

The final regression model obtained can explain the uncertainty propagation of the spatially explicit model from its components. The model coefficients express the sensitivity of the corresponding components. Since there is no intercept in the regression model, the uncertainty propagated from the variation of an independent variable to the model is the product of its variance or covariance or cross covariance and the corresponding coefficient.

The integration of the joint sequential simulation with the uncertainty analysis procedure in this study has made it possible to take the spatial correlation of multiple variables and effect of neighborhood into account in modeling uncertainty propagation. Most of uncertainty of a pixel comes from the variation of the model components at the concerned (host) pixel. The interaction and spatial correlation between the model variables may contribute positive or negative covariance to the total uncertainty of the model. Discarding the interaction and spatial correlation between the variables might result in large bias in prediction variance of dependent variable. On the other hand, the neighbors of a host pixel usually contribute negative uncertainty through cross correlation, indicating a reduction in total uncertainty of the host pixel, although the uncertainty contribution from neighbor pixels occasionally is positive. This implies that neglecting the cross-spatial correlation in spatial simulation may lead to overestimation in uncertainty contribution of model components for most pixels of a study area. The uncertainty contribution of neighbor pixels could be totally different even in the case in which they have the same distance to a host pixel. The largest and smallest uncertainty contributors vary depending on locations.

Topographical factor LS

Based on previous studies, soil erosion is most sensitive to the combined topographical factor LS (as a product of slope length L and slope steepness S), thus, the factor LS is very important for uncertainty analysis of soil erosion system. We completed four studies for spatial prediction and uncertainty analysis of the LS factor.

By comparing different geo-statistical methods, Wang et al. (2000b) suggested that sequential indicator simulation was a better method for spatial prediction and uncertainty assessment of the topographic factors than ordinary and indicator kriging. The sequential indicator simulation provided not only reliable spatial conditional variance maps of the predicted values, but also probability maps for predicted values larger than a given threshold value. The variance and probability maps can be used to assess the quality of modeling and simulation systems. The variance maps can be used to further develop error budgets over space and time into various error sources. Probability maps for predicted values larger than a given threshold value such as soil loss tolerance can help decision makers in management of ecological and environmental resources in these cases where some extreme values are important. The uncertainty measures and loss functions can be combined and used for estimating loss due to mistakes in decision-making.

Wang et al. (2000a and 2001a) accomplished spatial prediction and uncertainty budget from slope length, slope steepness, their model parameters and measurement errors to the combined topographical factor LS by integrating the sequential indicator simulation above and a variance partitioning method - Fourier Amplitude Sensitivity Test (FAST). This method produced not only similar spatial distribution of estimates to that of the observed data but also spatial variance contribution maps. The variance contribution varied spatially and depending on different components, and thus provided spatial information of uncertainty for system modellers and decision-makers for the purpose of error management. Using the spatial information, modellers can improve predicted maps (local estimates) of soil loss by paying attention to reduction of local errors from main factors (main sources of uncertainty) in sampling, measuring and simulation, and further by obtaining reliable spatial variability of the factors. On the other hand, decision-makers can use the maps with caution for local plans in agricultural and rangeland management.

The sequential indicator simulation successfully generated spatial prediction maps of the variables. The simulation held the data values at the sampling locations where the error variances were zero. However, reducing the prediction uncertainty at the unknown locations depended greatly on the simulation techniques including determining number of simulation runs, number of indicator semivariograms, semivariogram parameters (nugget, sill and range), and data search radius used. As the numbers of the runs and the indicator

semivariograms increased, the prediction variance for slope steepness and length decreased. The use of more than 500 runs and seven indicator semivariograms led to stable prediction and uncertainty maps.

Based on the spatial variance partitioning, the slope steepness contributed the largest uncertainty to prediction of LS factor, followed by slope length because slope steepness was much higher correlated with LS than slope length. The contribution due to the model parameters was relatively small. Reducing uncertainty of slope steepness is thus critical to increase the precision in spatial prediction of the topographical factor LS and soil loss. This implies that obtaining accurate spatial variability of slope and slope length for a specific area by sampling and measuring enough field plots may be more important to improvement in the spatial prediction than calibrating the model parameters. However, how many plots that are consider enough for this purpose may depend on the landscape complex. Moreover, these results do not mean that the method of uncertainty analysis used in this study can replace the calibration of the model parameters. But, the uncertainty information obtained by this method does suggest a direction for future error reduction. This is important especially when the cost for collecting calibration data is high.

The sensitivity of the LS factor to the components was also analyzed using the field data set. The measurement errors of slope steepness and length were evaluated. For example, when measurement errors of slope steepness and length were assumed to be 10% of their means, the percentage of the variance contribution from slope steepness, length, both measurement errors, and the total of the model parameters were 78.8%, 15.9%, 0.2%, 2.2%, and 2.9%, respectively. The variance of the LS factor was still mainly due to the uncertainty in slope steepness. The uncertainty analysis for measurement errors was only done for the population using the field data set and was not involved in the spatial predictions. Additionally, the input components using FAST were assumed to be independent. Using the FAST for a system where the correlation between input components exists should be done with caution.

Mapping the LS factor above for Fort Hood was based on the set of models in the RUSLE. In the models the upper contribution area is not taken into account. Based on the physically based LS equation, Wang et al. (2001d) investigated the use of DEM and appropriate DEM spatial resolution for mapping the LS factor, and modeled the loss of spatial variability due to data resampling. The predicted LS map and its variance map derived using the physically based topographical factor LS equation and DEMs are spatially consistent and correlated with the topographical features. That is, in the hilly areas the predicted LS values and variances are high, and in flat areas LS values and variances are low. The lake areas are filled with LS values of zero. The improved correlation of the predicted LS values with the topography is obvious compared to the corresponding maps by a spatial simulation based on the empirical models and sample data.

Gertner et al. (2001d) further investigated spatial resolution of DEMs to generate a topographical factor LS map using a physically based topographical factor LS and carried out an uncertainty budget. The error propagation from slope, up-slope contributing area, and two model parameters to the prediction of LS factor was modeled and relative variance contributions were generated using Fourier Amplitude Sensitivity Test (FAST). The results of variance partitioning suggested that given a spatial resolution, the uncertainty in predicting the topographical factor LS using a DEM mainly came from slope in the areas of gentle slopes and up-slope contributing area in steep areas. Two model parameters contributed little in terms of variance. For our particular case study, a DEM at the spatial resolution courser than 5 m could be considered problematic for the prediction of the LS factor.

Vegetation cover and management factor C

The vegetation cover and management factor C together with topographical factor LS is very important variable for monitoring soil erosion. In the USLE, the C factor varies in both space and time, depending on ground cover, canopy cover, and minimum rain drip vegetation height. In the RUSLE, the calculation of the C factor is more complicated, and existing LCTA database does not provide enough information to derive the C factor based on the RUSLE. The C factor relevant studies of this project focus on its prediction based on the USLE.

We developed a method to determine appropriate plot size and spatial resolution for mapping multiple vegetation types using remote sensing data for a large area, and applied this method to Fort Hood area (Wang et al., 2001e). This study suggested that the existing LCTA plot size of 100m transect line was appropriate for collecting vegetation cover types. If the measurements of vegetation cover types at 100m transect lines are used as estimates of 100m by 100m pixels, the spatial resolution of 100m by 100m is probably an optimal choice.

We have done the optimal sampling design for investigating vegetation cover at Fort Hood area (Xiao et al., 2001). In the design, both plot size and sample size were considered in terms of cost and variance estimated for regional and local situations to obtain spatial information of overall vegetation type. We found that the sample size of 200 plots for plot size 100m could be recommended since it achieved high precision. When the cost was introduced into the design, sample sizes for local and regional estimation was 40 and 200 for overall vegetation cover. However, the sample size by the traditional method differed significantly from that by kriging method. The optimal sample size by kriging was much smaller than that by the traditional method, which implied that the kriging was more cost-efficient. The sample size of 200 plots may be enough for regional estimation of overall vegetation cover in percent. The sample sizes for predicting cover percentages of individual vegetation types and for classification of land cover types should be more than that.

Accurately mapping vegetation types and spatially assessing classification accuracy is difficult because of the high cost of collecting field data at a high density, spectral mixtures, low correlation between remote sensing and field data, and limitations of traditional methods. In Wang et al. (2001h) study, we developed an image-aided sequential indicator co-simulation method that models the spatial variability of an estimated variable based on the spatial cross variability between this variable and an auxiliary variable such as a Landsat TM image. The co-simulation is a geostatistical method that provides a number of realizations (estimates of probability) at each location given field data and previously simulated values within a neighborhood, and an image datum at the location estimated. An expected vegetation type and classification probability is derived as the final estimate. A spatially explicit misclassification probability map was also obtained.

The results showed that the classification and misclassification probability varied over space and depended on the field and auxiliary data sets used, land cover types, landscape complexity, and topographical features. At the southwest, center, and north parts of Fort Hood, the probabilities at which the pixels were classified into tree and mixed vegetation category were very low (less than 0.2), while the classification probabilities into grass were very high (larger than 0.8). At the east and northeast parts, on the other hand, there were high probabilities (higher than 0.5) for classifying the pixels into tree, and very low probabilities at which the pixels were classified into grass. The lowest misclassification probability happened at the southwest, center, and north parts. At the east, northeast, and south parts, the prevailing misclassification probability varied from 0.2 to 0.4, and some pixels might be incorrectly classified at the probabilities from 0.4 to 0.6. This method can be used to spatially assess classification accuracy and realizes significant improvement in accuracy assessment compared to traditional methods such as percentage correct and Kappa value.

Based on a case study at Fort Hood, it was found that image-aided co-simulation improved the classification compared to a simulation without TM data and a traditional image-aided classification. However, the classification accuracy of six land cover types at Fort Hood, including tree, shrub, grass, mixed vegetation land, bare land, and water, was still low, about 58%. Many reasons including insufficient sample size of existing 200 LCTA plots, improper classification definitions, the poor quality of satellite images used, difficulty of separating spectral mixture pixels, etc., might cause the result.

Shinkareva et al. (2001) presented another new method to spatially assess classification accuracy. This method was developed by combining a classification error matrix by cross validation and posterior probabilities calculated using auxiliary data such as satellite images used for classification. This method was applied to Fort Hood where six land cover categories were incurred. The results showed that the misclassification probability of unknown locations classified as six land cover classes varied over space and depends not only on the

density of sample plots used for classification, but also spatial distribution of the land cover classes. Furthermore, the misclassification probability by the proposed method were the lowest for water, then grass, tree, bare land, mixed land, and shrub, and they corresponded with the spatial distribution of the land cover classes. For example, the misclassification probability at the south and southwest area, where most was classified as grass were small.

The vegetation cover and management factor C represents the effect of cropping and management practices on erosion rates in agriculture, and the effect of ground, tree and grass canopy covers on reduction of soil loss in non-agriculture situation. We studied mapping of the C factor using the LCTA data set of ground and canopy cover and Landsat TM images at For Hood (Wang et al., 2001i). The measurements of ground cover, canopy cover, and minimum rain drip vegetation height from the LCTA field plots were used to calculate the plot C factor values based on the C factor empirical equations mentioned in Chapter 2. The plot values were then employed for mapping the C factor. After comparing six methods, we found out that the sequential Gaussian co-simulation with a TM ratio image resulted in the highest correlation (0.7317) and the smallest root mean square error (0.0159) between the estimated and observed C factor values, and reproduced the best and most detailed spatial variability of the C factor.

In the USLE, the C factor depends on ground cover, canopy cover, and minimum rain drip vegetation height. The variables are spatially correlated with each other. Theoretically, considering interactions between variables can improve correlation between resulting maps and using spatial information from neighbors can increase map accuracy. However, the difficulties lie mainly at how to model the interactions and uncertainty propagation from the variables, their interactions, and spatial information from neighbors. Gertner et al. (2001c) integrated a joint sequential co-simulation with Landsat TM images for mapping and a polynomial regression for spatial uncertainty analysis. This method was applied to the Fort Hood case study in which ground cover, canopy cover, and vegetation height were jointly mapped to derive the vegetation cover and management factor C, and variance contributions from variation of three variables, their interactions, and neighboring information to the uncertainty of the predicted factor were spatially assessed.

In addition to unbiased maps, this method well reproduced the spatial variability of the vegetation variables and spatial correlation between them, and successfully quantified effect of variation from the variables, their interactions and spatial information from neighbors on prediction of the vegetation cover factor. The spatial variability and spatial correlation – spatial interactions were modeled in terms of auto and cross semi- variograms respectively. The role of Landsat TM images is to provide a control surface to reproduce spatial variability of the estimated variables, and also to establish a bridge surface for modeling the interactions between the estimated surfaces through the cross semi-variogram models. Therefore,

acquiring remotely sensed data of high quality and high correlation with interest variables is critical to derive accurate maps of multiple variables and their correlation.

The variance of predicted vegetation cover factor was mainly contributed by variation of ground cover and canopy cover, and the contribution from vegetation height was very small. This suggests drawing a representative sample and accurately measuring both ground cover and canopy cover is very important to derive the map of vegetation cover factor for prediction of soil loss. The variance contributions from the interactions between ground and canopy cover, and between canopy cover and vegetation height, were significant. The effect of spatial information from neighbors on the uncertainty of the predicted vegetation cover factor decreased as increased the separation distance of the neighbors from the estimated location. The total variance contribution of the spatial information from the neighbors was –17.8%, suggesting use of spatial information from neighbors can significantly increase accuracy of maps.

The joint sequential Gaussian co-simulation means that ground cover, canopy cover, and vegetation height were first mapped jointly, and from these prediction maps, the C factor map was then calculated. Using this method, we obtained a coefficient (0.7056) of correlation between the estimated and observed C factor values for the test data set (Gertner et al., 2001c). This accuracy was slightly lower than that (0.7317) by directly mapping the C factor using the sequential Gaussian co-simulation (Wang et al., 2001i). The reason is that by the joint sequential Gaussian co-simulation, the uncertainties from ground cover, canopy cover, and vegetation height were propagated to the prediction of the C factor. However, the joint mapping method provides the possibility to do uncertainty budget and to understand spatial correlation between the variables and spatial information from neighbors.

Disturbance

We completed spatial modeling and spatial uncertainty analysis of ground surface and vegetation cover disturbance due to training activities (Fang et al., 2001b). The model used to predict the spatial and temporal distribution of disturbance probability/intensity in this research area is modified from the model of Guertin et al (1998). One modification added a new term, the number of battalions training at the facility in a given year, to represent the change of activity intensity over time. The other modification reinterpreted the disturbance observations as a continuous variable ranging from 0 to 1 indicating the proportion of subplots disturbed within a plot. The original model of Guertin et al (1998), considered the disturbance observations as a binary (presence/absence) variable from each subplot. This modified model avoids the questionable assumption of subplot observation independence within plots and has the form (Fang et al., 2001b):

$$y = \frac{e^{(b_0 + \sum_{i=1}^{8} b_i \cdot x_i)}}{1 + e^{(b_0 + \sum_{i=1}^{8} b_i \cdot x_i)}} \quad (4.1)$$

where b_i's and x_i's are parameters and independent variables of the model, respectively. The variables in this model included number of battalions, the shortest distance to road, slope, region codes, and vegetation types. Distance to roads of the plots was calculated from the coordinates of the plots and roads. The number of battalions training at Ft. Hood in a given year was taken from facility records. The region code of the plots was copied from a facility map. Disturbance predictions are extrapolated across the facility using the maps of the independent variables at the spatial resolution of 50m by 50m.

Uncertainty from these sources fell into four general categories – modeling, mapping, decision error, and measurement errors. The uncertainty of the model parameter estimates was referred to as modeling error and measured as the variance of parameter estimates. Mapping error referred to the error in the distance, slope, and vegetation classification maps used to spatially extrapolate disturbance across the training facility. Decision error was uncertainty contributed from inaccurate management decisions or projections. Measurement error was the uncertainty contributed by the dependent variable - disturbance due to sampling, measuring and data processing. It was estimated using an unbiased estimator of variance and data from a 1998 validation study in which two observers independently assessed disturbance at 20 plots. A Taylor series expansion method was applied to partition the uncertainty of predicted disturbance into the uncertainty sources.

Spatial and temporal variation in training activity induced disturbance at Ft. Hood from 1989 to 1996 was presented in Figure 4.3. The disturbance over time first decreased by year from 1989 to 1991, then increased from 1991 to 1996. The disturbance increased with the number of battalions training at the facility; decreased with distance to roads and

Slope (Fang et al., 2001b). The disturbance was the highest in the west region, lower in the east and southern regions and lowest in the central region. The disturbance was the highest in grass, followed by shrub and lowest in tree. High disturbance intensity/probability (>0.6) occurred mainly in grassy areas of the west and east training areas. Low disturbance intensity/probability (<0.3) occurred in the central and south training areas and in the roadless portions of the east training area.

The total uncertainty ranged from 0.00 to 0.195 variance units of the disturbance prediction. The uncertainty contribution from mapping error was the largest source of prediction uncertainty. It was broken out into the uncertainty contributions of the distance to road map,

slope map and the vegetation map. Since the vegetation map uncertainty was the dominant source of mapping contributed uncertainty, and mapping uncertainty was the dominant source of total prediction uncertainty, the spatial distribution of predicted disturbance uncertainty was largely determined by the vegetation map as well as the predicted disturbance map. The central, northeast and southwest parts of the study area had little predicted disturbance and therefore had relatively low uncertainty (<0.04) associated with those predictions. The west region and the parts of the east region with roads had more predicted disturbance and therefore, had relatively higher prediction uncertainty for those areas falling within the vegetation map categories (tree and shrub) that produced the greatest amount of uncertainty.

The spatial distribution of prediction uncertainty was heterogeneous and corresponded to the spatial distribution of components of the prediction model. The majority of the prediction uncertainty was caused by high classification error rates for vegetation types shrub and tree in the vegetation map. When error rate of vegetation classification was low, as in vegetation type grass, the total amount of uncertainty was greatly reduced. Under such conditions, vegetation misclassification contributed only a minor amount of uncertainty to the model prediction and modeling error became the dominant source of prediction uncertainty. Decision and measurement error of disturbance contributed only a small amount to prediction uncertainty.

Based on the behavior of the model components in uncertainty propagation, reducing the error rate of vegetation classification is probably the most efficient way to increase the precision of disturbance prediction. Using an updated high quality vegetation map should reduce a large proportion of the variance at the pixels whose vegetation type is tree or shrub.

Soil erosion

Figure 4.4 show the location of ground plots and spatial distribution of data values for topographical factor LS, soil erodibility factor K, and vegetation cover factor C in 1989, 1992 and 1995. The data sets were obtained by calculation with the empirical regression equations of the input factors from the LCTA plot measurements of slope steepness and slope length for LS, five soil properties mentioned above for K, and three vegetation cover variables (ground cover, canopy cover, and minimum rain drop vegetation height) for C. The LS factor and K factor is assumed constant at the same locations during the period from 1989 to 1995, and two data sets were obtained in 1989.Within Fort Hood, there were no rainfall stations. Three data sets for the C factor were available for 1989, 1992 and 1995. Larger LS values are located at the east parts of Fort Hood, larger K and C values at the west parts. The C values decrease from 1989 to 1992 and then increase at the west and north parts to 1995. The statistical parameters of the data sets are listed in Table 4.1.

Fig. 4.4. Sample locations and spatial distribution of data for topographical factor LS, soil erodibility factor K, vegetation cover factor C in 1989, 1992 and 1995.

The coefficients of correlation between the input factors are shown in Table 4.2. There is a significant but not strong correlation between these factors, except for the C factor in 1992. The C factor values at different years are highly correlated with each other. Furthermore, we studied the correlation of the input factors with Landsat TM data and their various ratio images, elevation and slope data from a digital elevation model. We found that LS factor was highly correlated with slope. The K factor and C factor in 1989 have the highest correlation with spectral data of 89's Landsat TM7. The C factor in 1992 was most correlated with 92's Landsat TM2. There was the highest correlation of the C factor 1995 with a 95's ratio image - TM7/TM4.

Figure 4.5 presents four auto semivariograms and one cross semivariograms of the input factors. Gaussian model was used to fit the experimental semivariogram of R factor and spherical model for other factors. The cross semivariogram between LS and C was approximated by Markove model. According to the correlation above, the joint sequential co-simulation of LS, K factor and C factor for 1989 was accomplished with aid of slope and 89's TM7 (Wang et al., 2001b). The results of LS and K were used for 1992 and 1995 prediction. The 92's C factor was co-simulated with 92's TM2 and the 95's C factor with the ratio image of 95's TM 7/TM4. The rainfall-runoff erosivity factor R was simulated using a data set of 218 rainfall stations covering a large area of six states around Texas without any auxiliary data, then from result map, the R factor map of Fort Hood was extracted.

The predicted maps of factors LS, C, K, R, and soil loss in 1989 are demonstrated in Figure 4.6 (also see Wang et al., 2001b). Average values of all the predicted maps fall into the confident intervals at a probability of 95% (Table 4.1). The spatial distribution of the predicted values is similar to that of corresponding data set in Figure 4.4. For example, larger LS factor values were predicted at the east parts, larger C factor and K factor values were obtained at the west parts. The predicted R factor values slightly increase from the west to the east, and they are much higher than the value 270 of R factor obtained from a published isoerodent map for Fort Hood. The calculated values of soil loss are higher at the west and north parts. Figure 4.7 shows the variance maps of the predicted values. Generally, at the areas with larger prediction values the estimation variances are higher, and vice versa. Presented in Figure 4.8 are the co-variance maps of the input factors with soil loss and between the factors. All the input factors have positive co-variances to soil loss. However, most of the co-variances between factors LS and C are negative. This is because a steeper area implying larger LS factor, but less training activities and disturbance of vegetation resulting in higher vegetation cover and less C factor.

Fig. 4.7. Variance maps of predicted values for topographical factor LS, vegetation cover factor C, soil erodibility factor K, rainfall and runoff factor R, and soil loss in 1989 using joint sequential co-simulation with slope map from a DEM and 89's TM7.

Figure 4.9 show the expected, variance and probability maps of predicted soil erosion status in 1989. The predicted values were derived by predicted soil loss values divided by soil tolerance values at the same locations. The erosion status is regarded as a measure of land conditions. According to training land carrying capacity standards, the erosion status is grouped into four classes: less than 1.0, from 1 to 1.5, from 1.5 to 2.0, and equal and larger than 2.0. High erosion status values (e.g. greater than 2) reflect a poorer land condition, whereas lower erosion status (e.g., less than 1) implies a better land condition. From Figure 4.9, the east and northeast parts of Fort Hood have better land conditions than the west parts. That is, at the east and northeast parts the probability at which erosion status of less than 1.0 may take place is higher than 0.5, while the probability at which erosion status of greater than 2.0 may occur is less than 0.5. At the west parts, on the other hand, erosion status of less than 1.0 may take place at the probability less than 0.5, while erosion status of greater than 2.0 may occur at a probability more than 0.5.

Figure 4.10 presents the change of predicted values for vegetation cover and management factor C and soil loss during the period from 1989 to 1992 and 1995. From 1989 to 1992, the predicted maps of the C factor and soil loss decreased at the east and central parts, and increased at the south parts. From 1992 to 1995, the predicted values significantly increased at the west and north parts because the west and north parts of Fort Hood are flat and more training activities have taken place, resulting in more disturbance to vegetation cover.

A spatial uncertainty budget for prediction of soil loss was done at a small area (5010 by 5010 m^2) and high spatial resolution of 5 m. The location of the small area is shown in Figure 4.11. The C and K factors were predicted using the above sequential joint co-simulation with Landsat TM images, and R factor using a sequential Gaussian simulation without auxiliary data. The LS factor was derived using a Digital Elevation Model (in Figure 4.11) at 5 m resolution and a physically based LS equation. When the expected maps were generated, we assumed the input factors were independent. The expected soil loss map was then calculated as a product of R, K, LS, and C factors by overlapping the maps. The variance of soil loss was derived and partitioned into the input factors by Taylor series expansion described in previous reports.

Figure 4.12 presents the predicted maps of the input factors and soil loss at the spatial resolution of 5 m for the small area. There, the hilly areas go from the northwest to southeast. Along the boundaries of the hilly areas, large LS values were predicted. At the flat areas, small LS values were obtained. Large C and K factor values took place mainly at the

southwest areas and small values at the southeast. The R factor was evenly predicted. The spatial distribution of predicted soil loss is similar to that of predicted LS values and thus reflects the topographical features.

In Figure 4.13, the variances of predicted soil loss were high along the boundaries of hilly areas and low at the southeast areas. The largest relative variance contribution came from LS factor, then C factor, K factor, and R factor. The average contribution is 89% for LS factor, 8% for C factor, 2% for K factor, and 1% for R factor. Along the hilly area boundaries, large slopes and up-slope contributing areas determined the amount of soil erosion, but high vegetation cover might significantly reduce soil loss. At the flat areas, slope was very close to zero, thus LS also close to zero, and very little or no soil erosion happened. The overall error budget above was carried out supposing that the spatial correlation between the input factors was not significant. Assuming a measurement error for LS of 20%, C of 25%, K of 10% and R of 10%; the overall error budget is shown in Table 4.3. The error budget assuming no measurement error in these factor is displayed in Table 4.4.

Table 4.1. Statistical parameters of sample data and predicted maps for topographical factor LS, soil erodibility factor K, and vegetation cover factor in 1989, 1992 and 1995. (Min, Max, Stdev, Lower, and upper are minimum and maximum value, standard deviation, lower and upper limit of confidential interval at probability of 95%).

	Min	Max	Average	Stdev	Lower	Upper
		LS factor				
Sample (211 plots)	0.0762	15.8393	0.7014	1.3173	0.5232	0.8796
Predicted Map	0.1133	9.1463	0.7237	0.5529		
Variance Map	0.0021	219.56	1.879			
		K factor				
Sample (211 plots)	0.095	0.447	0.27093	0.06555	0.2621	0.2798
Predicted Map	0.13678	0.53868	0.26852	0.05451		
Variance Map	0.0013	0.2651	0.00393			
		C factor 1989				
Sample (211 plots)	0.009	0.17091	0.05112	0.02416	0.0478	0.0544
Predicted Map	0.01017	0.2552	0.05119	0.02434		
Variance Map	0	0.04195	0.00032			
		C factor 1992				
Sample (208 plots)	0.009	0.20773	0.03684	0.02895	0.03289	0.0408
Predicted Map	0.009	0.2077	0.0408	0.03105		
Variance Map	0	0.0031	0.0006			
		C factor 1995				
Sample (171 plots)	0.009	0.3937	0.0570	0.0643	0.04734	0.0667
Predicted Map	0.009	0.3937	0.0641	0.06162		
Variance Map	0	0.0076	0.0014			

Table 4.2. Coefficients of correlation between the input factors

	K	C89	C92	C95
LS	-0.12044	-0.21051	-0.06918	-0.15444
K		0.225473	0.090653	0.168034
C89			0.406999	0.616338
C92				0.327694

Table 4.3. Overall error budget assuming a measurement error for LS of 20%, C of 25%, K of 10% and R of 10%.

Source	Variance Contribution (%)
Direct Contribution	
LS	67.4
C	14.2
K	3.2
R	0
Due to Measurement Error	
K	7.9
LS	5.5
C	1.1
R	.7

Table 4.4. Overall error budget assuming no measurement error.

Source	Variance Contribution (%)
Direct Contribution	
LS	72.9
C	15.2
K	11.2
R	.7

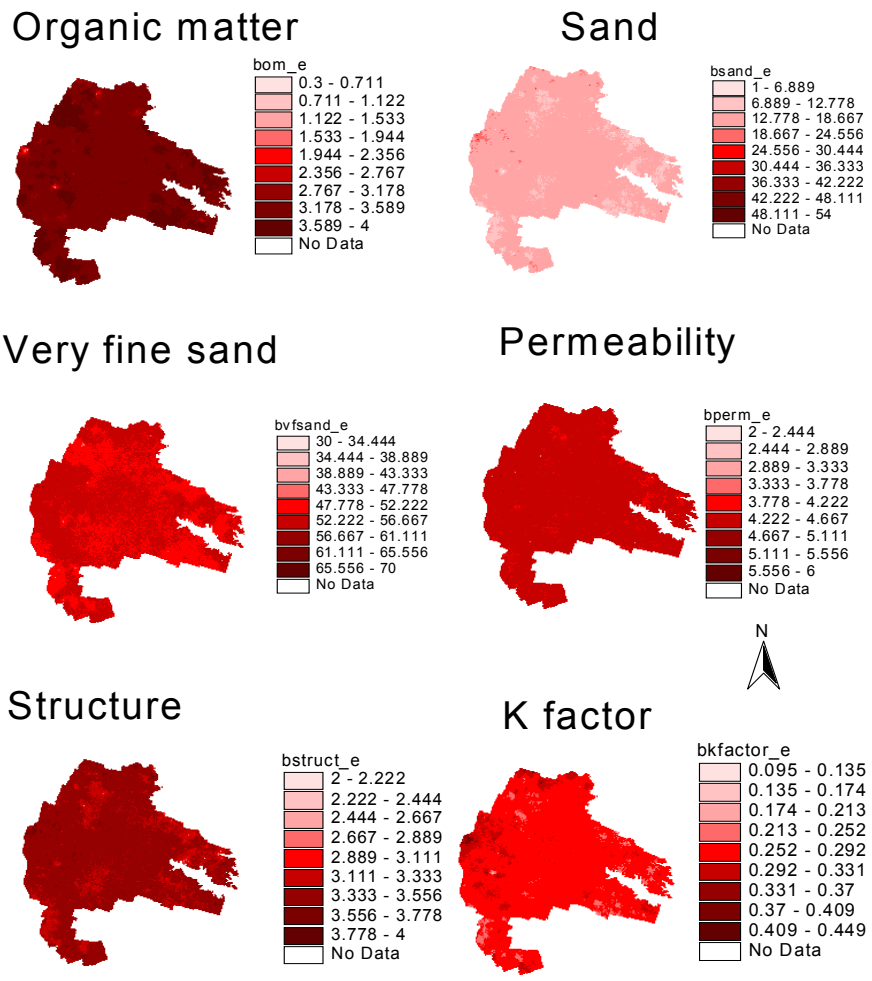

Figure 4.1. Predicted maps of soil organic matter, sand, very fine sand, permeability, structure, and soil erodibility K factor using multiple variable joint simulation.

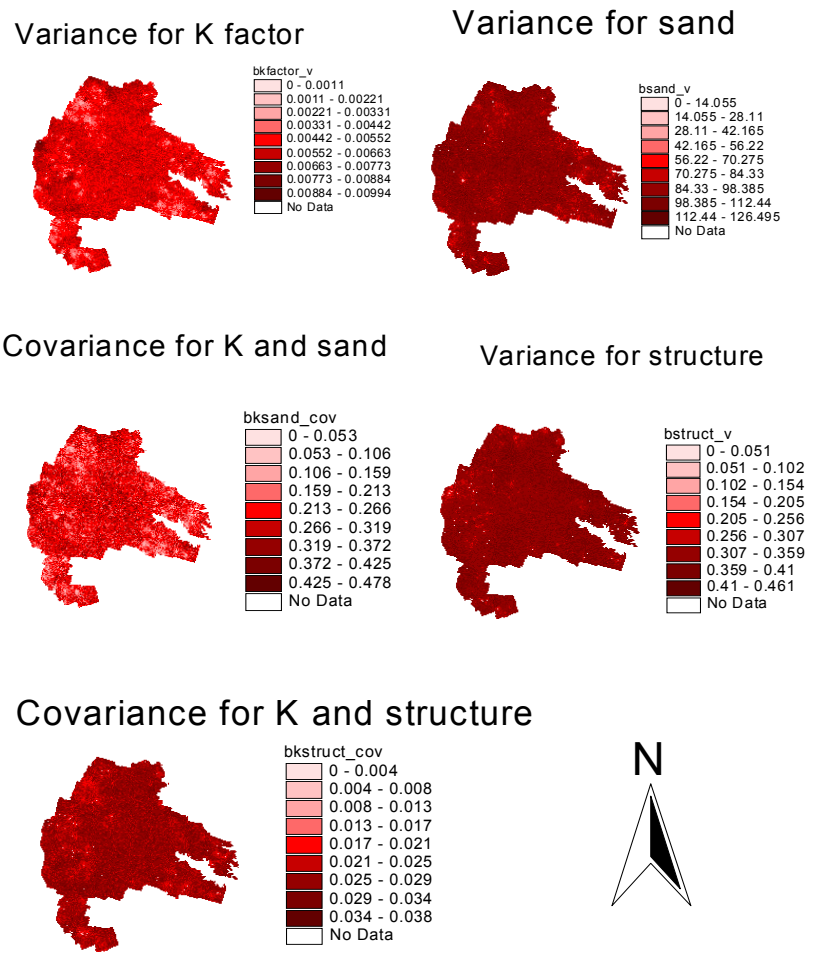

Figure 4.2. Predicted variance and covariance maps of K factor, soil sand, structure, and between them using multiple variable joint simulation.

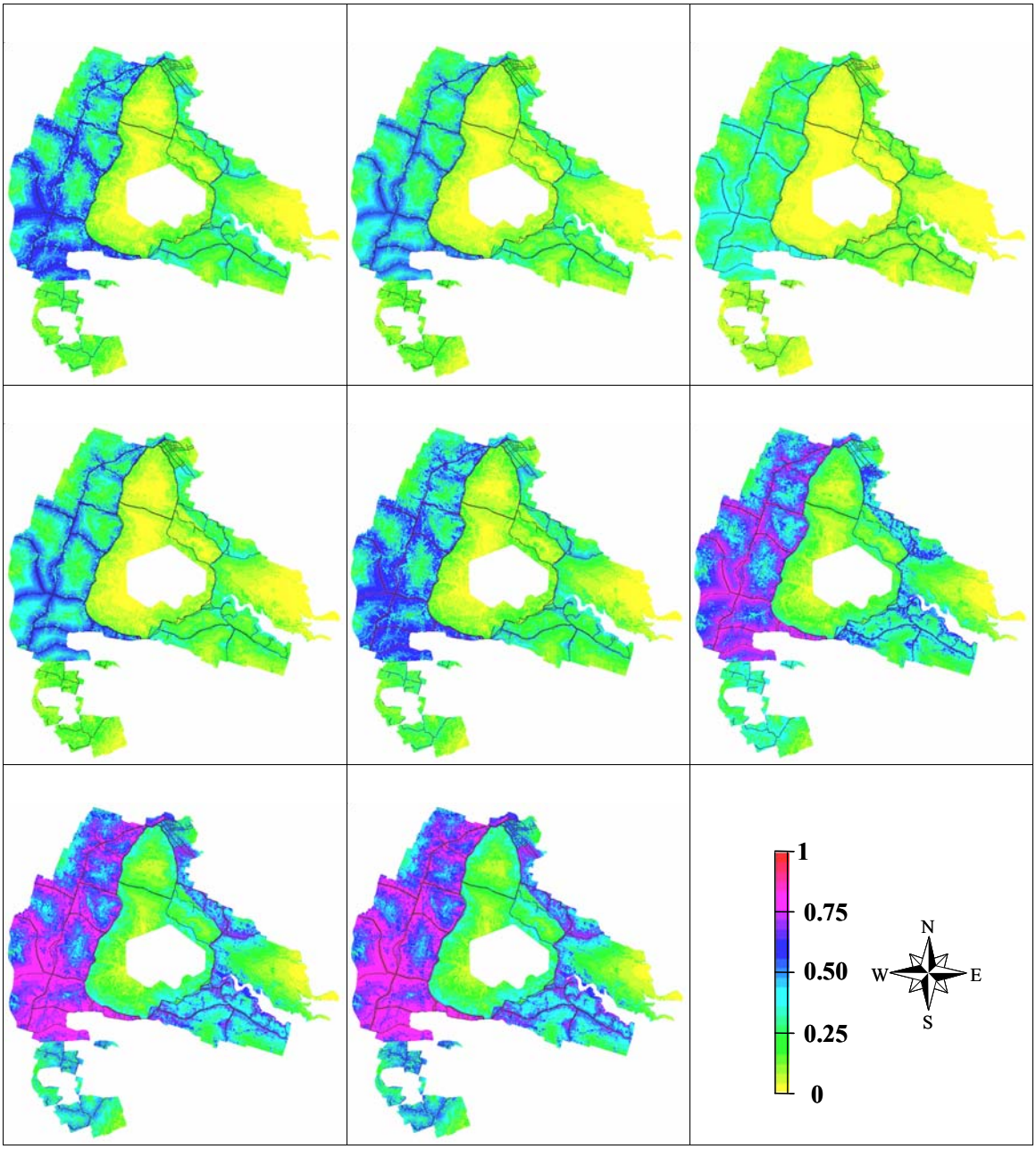

Figure 4.3. Spatial and temporal variation in training activity induced disturbance at Fort Hood from 1989 to 1996.

.

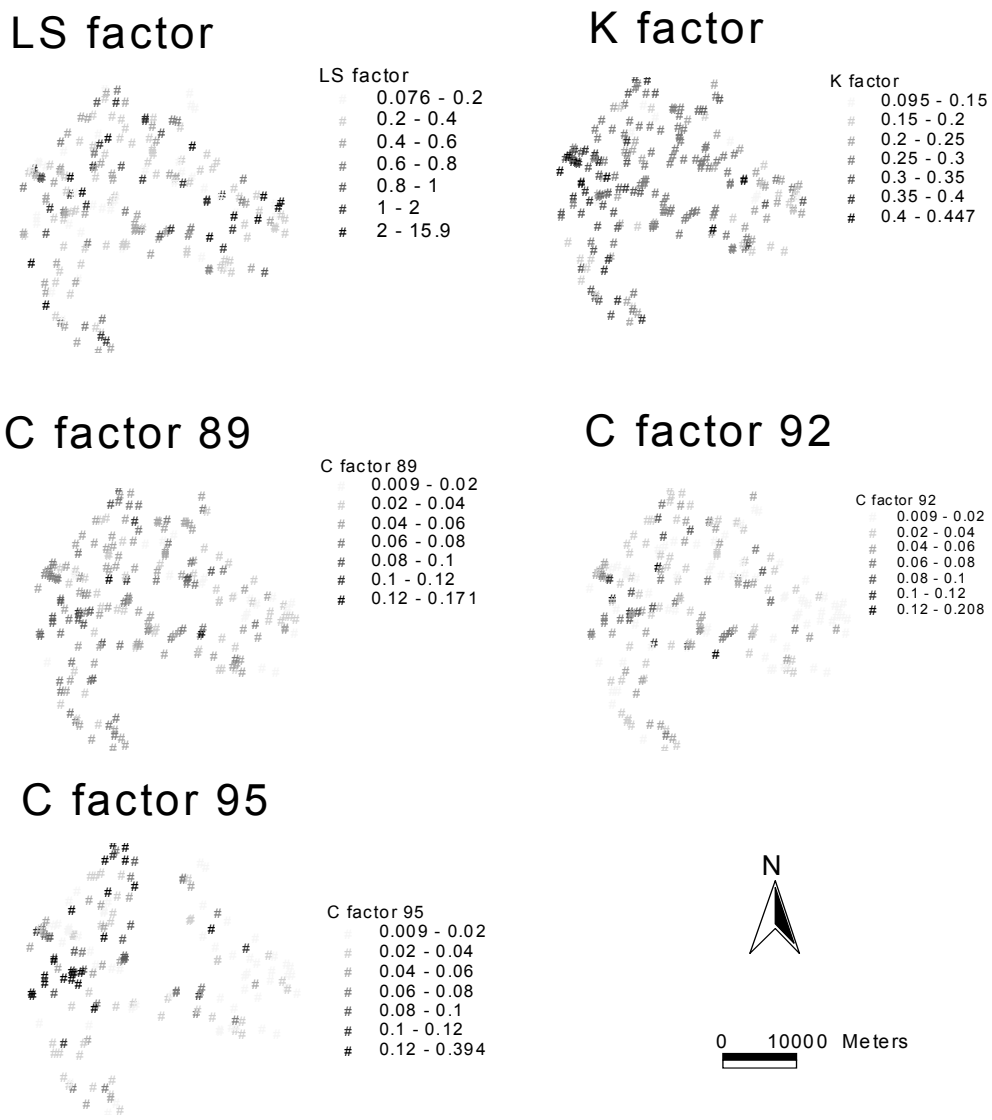

Fig. 4.4. Sample locations and spatial distribution of data for topographical factor LS, soil erodibility factor K, vegetation cover factor C in 1989, 1992 and 1995.

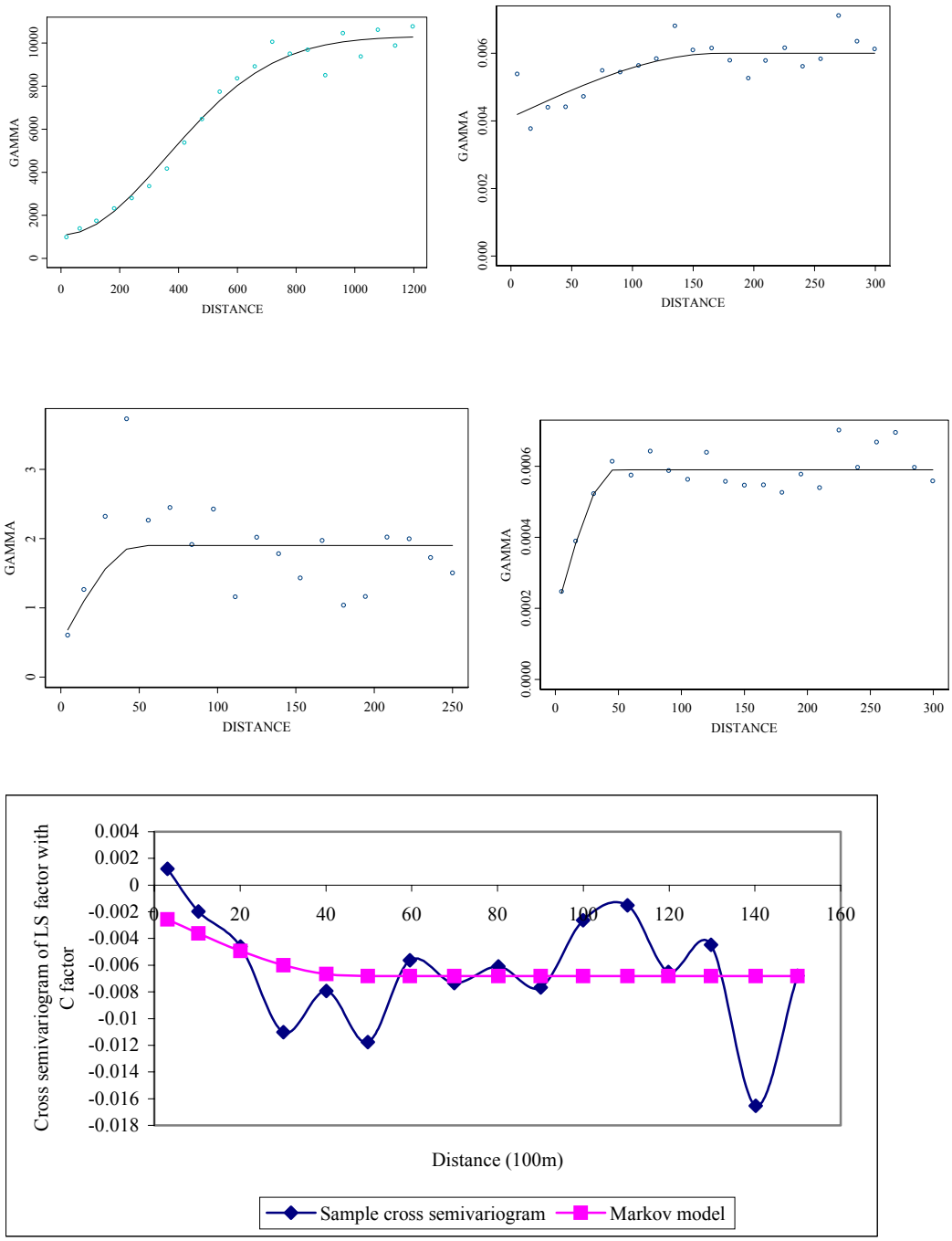

Fig. 4.5. Experimental and modeled semivariograms of rainfall and runoff factor R (upper left), soil erodibility factor K (upper right), topographical factor LS (middle left), vegetation cover factor C (middle right), and cross semivariogram of LS with C factor (below).

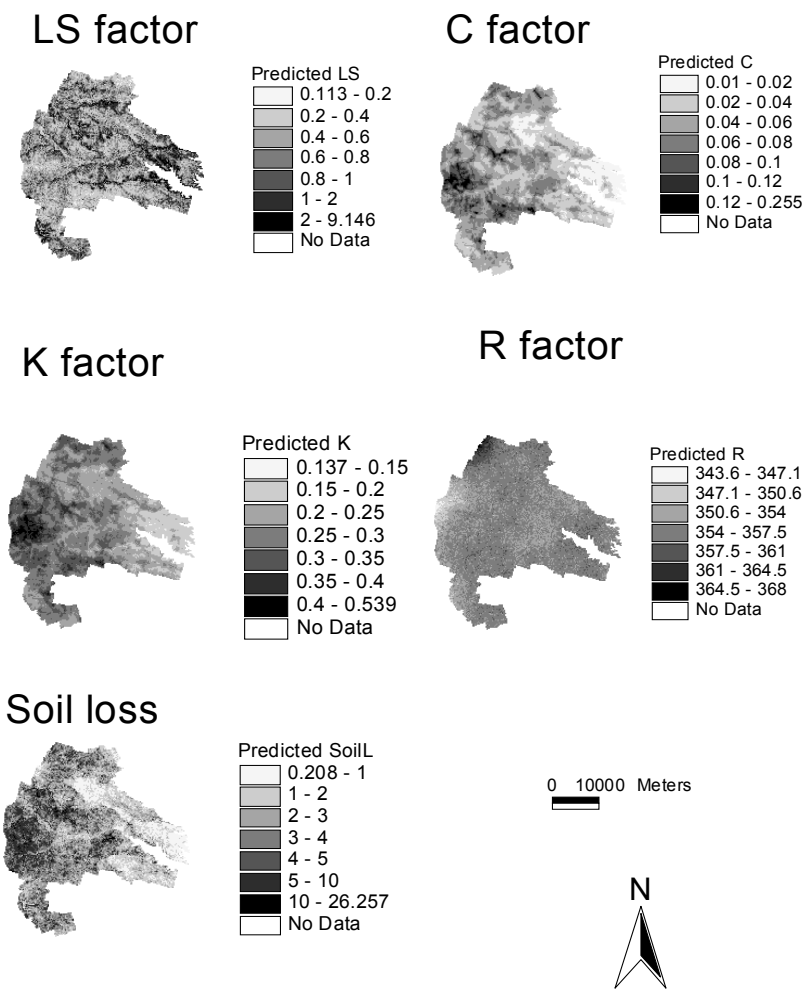

Fig. 4.6. Predicted maps of topographical factor LS, vegetation cover factor C, soil erodibility factor K, rainfall-runoff factor R, and soil loss in 1989 using joint sequential co-simulation with slope map from a DEM and 89's TM7.

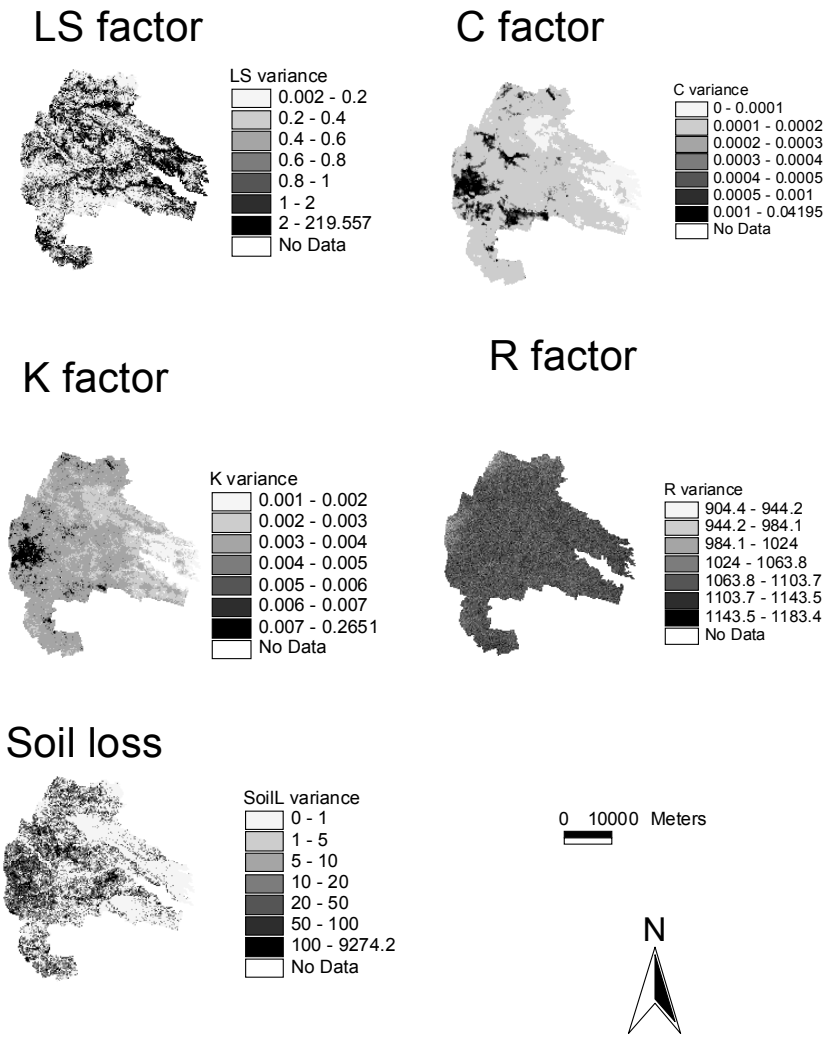

Fig. 4.7. Variance maps of predicted values for topographical factor LS, vegetation cover factor C, soil erodibility factor K, rainfall and runoff factor R, and soil loss in 1989 using joint sequential co-simulation with slope map from a DEM and 89's TM7.

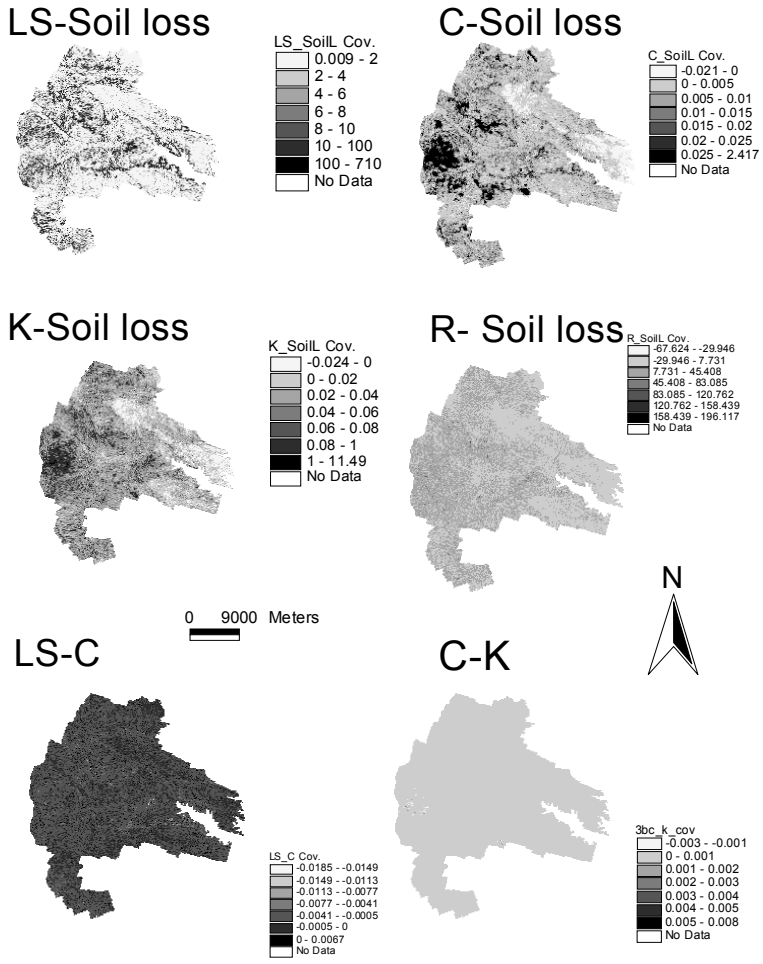

Fig. 4.8. Co-variance maps of predicted values for topographical factor LS, vegetation cover factor C, soil erodibility factor K, rainfall-runoff factor R, and soil loss in 1989 using joint sequential co-simulation with slope map from a DEM and 89's TM7. (LS-soil loss implies the co-variance between predicted LS factor and soil loss, and LS-C is the co-variance between predicted LS and C factor, and so on).

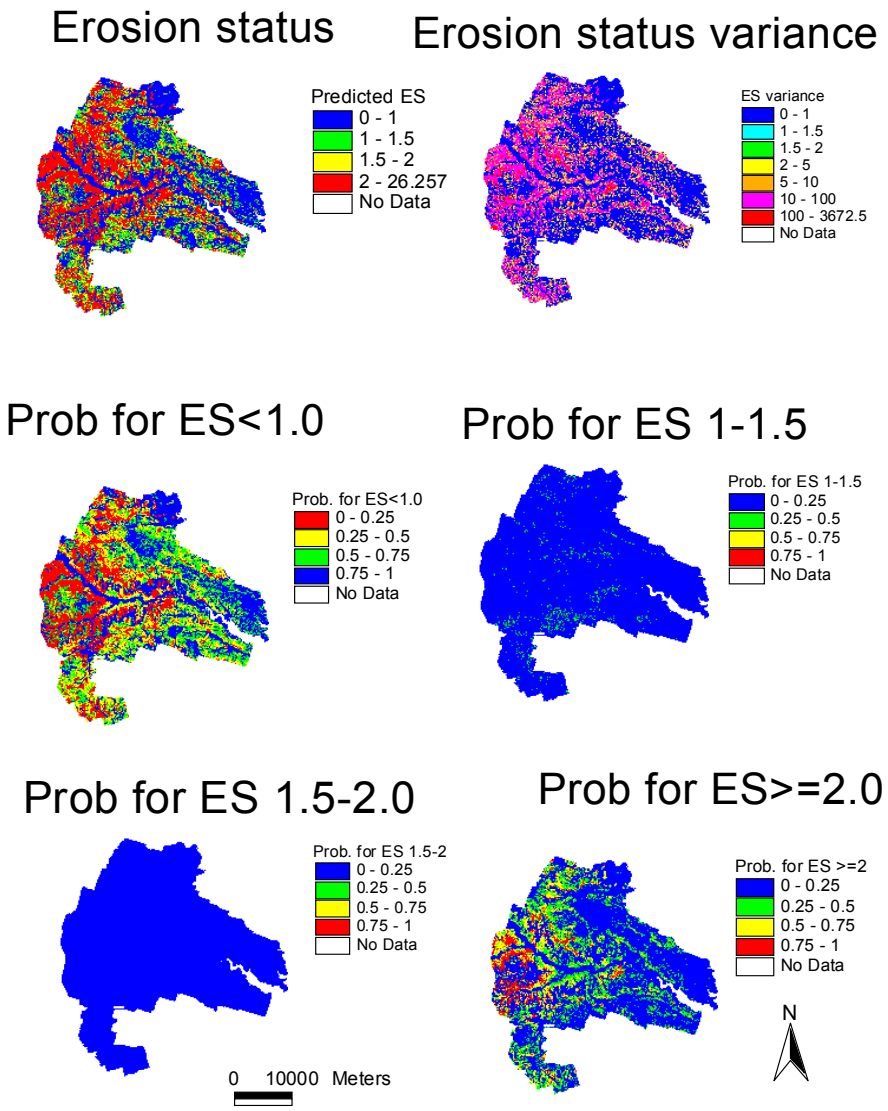

Fig. 4.9. Expected, variance and probability (Prob) maps of predicted erosion status in 1989 using joint sequential co-simulation with slope map and 89's TM7. The predicted erosion status values were derived by dividing the predicted soil loss with soil tolerance values at the same location. The erosion status is divided into four classes: less than 1.0, from 1 to 1.5, from 1.5 to 2.0, and equal and larger than 2.0.

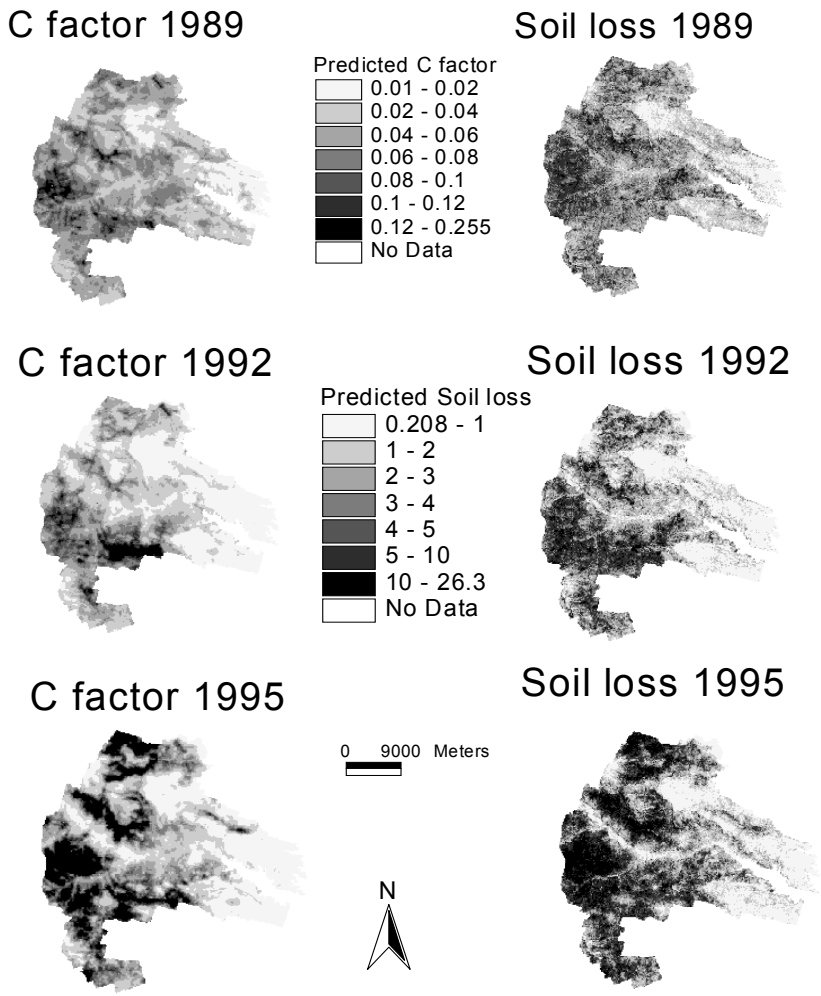

Figure 4.10. Change of predicted values for vegetation cover factor and soil loss during the period from 1989 to 1992 and 1995 using joint sequential co-simulation with slope map from a DEM and corresponding Landsat TM images.

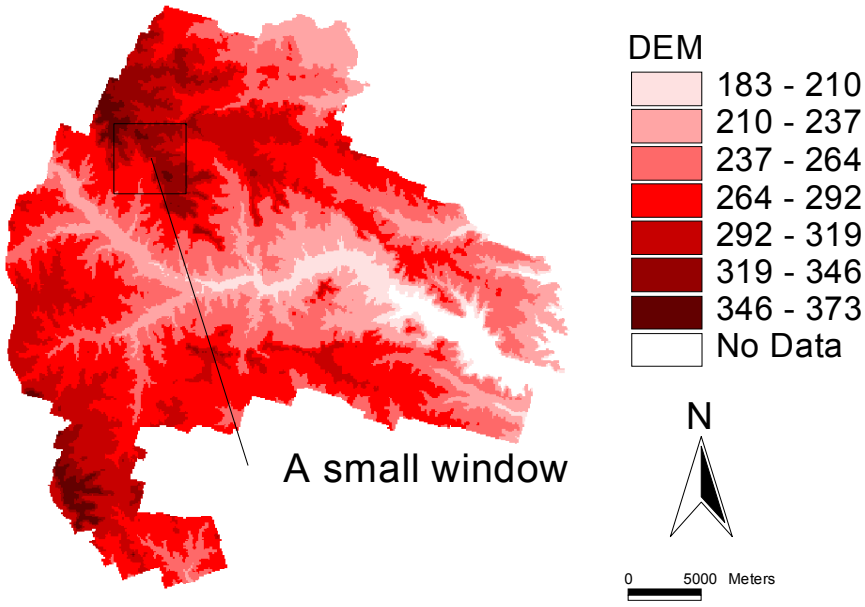

Figure 4.11. Digital Elevation Model (DEM) for Fort Hood and a small window area indicated for uncertainty partitioning at spatial resolution of 5m.

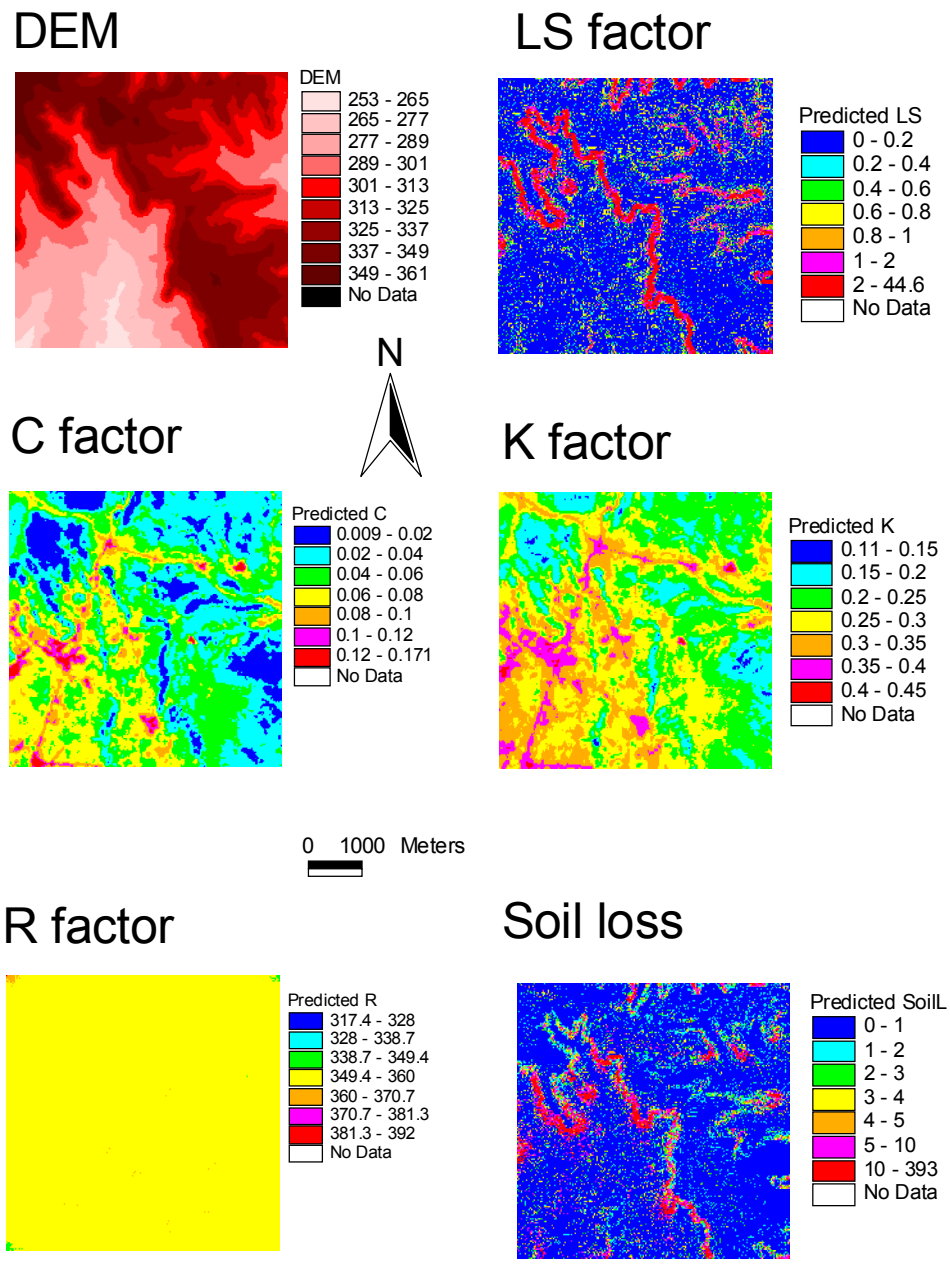

Fig. 4.12. DEM at spatial resolution of 5m for a small window area indicated in Figure 19, predicted maps of input factors and soil loss. LS factor was derived using the DEM and a physically based LS equation, C and K factors using sequential co-simulation with a ratio image (TM3*TM7)/TM4 and TM7 respectively, and R factor using a sequential Gaussian simulation without auxiliary data.

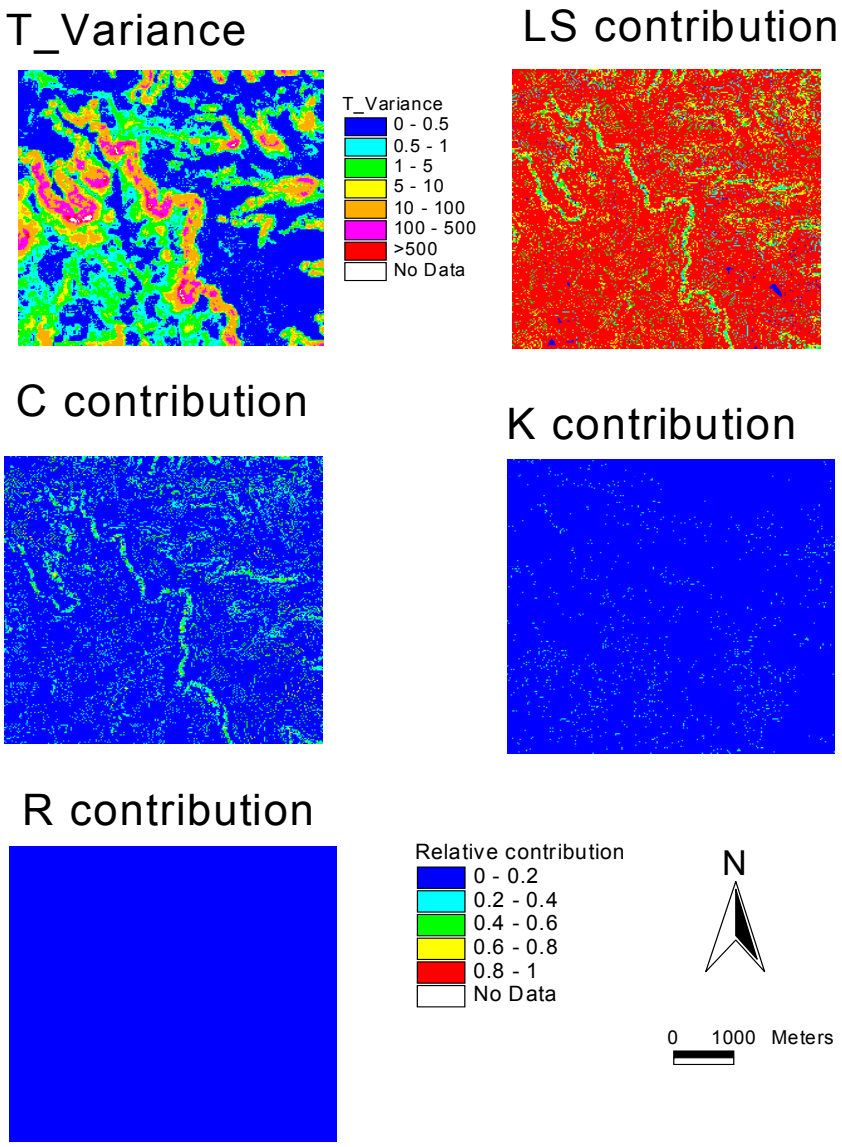

Fig. 4.13. Total variance of predicted soil loss and relative contribution maps of input factors at spatial resolution of 5m for a small window area indicated in Figure 4.11.

PRESENTATIONS, MEETINGS, TECHNICAL PAPERS, SOFTWARE, AND WEB SITE IN SUPPORT OF TECHNOLOGY TRANSFER PLAN

The following summarize our steps in support of the SERDP Error and Uncertainty Project Technology Transfer Plan (Gertner, G.Z. SERDP Project CS1096 Transition Plan. Submitted August 2001 and approved October 2001. University of Illinois Department of Natural Resources White Paper). A copy of the transfer plan is enclosed.

Presentations in support of the SERDP error and uncertainty project technology transfer plan.

Gertner, G., P. Parysow, A. Anderson, J. Westervelt, and D. Tazik. 1998. Error and Uncertainty for Ecological Modeling and Simulation: Case Study of Two Modeling Systems at Fort Hood. Partners in Environmental Technology, Technical Symposium and Workshop. Sponsored by the Strategic Environmental Research and Development Program (SERDP) and the Environmental Security Technology Certification Program (ESTCP). Arlington, VA 1-3 December 1998.

Gertner, G., G. Wang, A. Anderson, and P. Parysow. 1999. Error and Uncertainty for Ecological Modeling and Simulation: Error Identification and Estimation. Partners in Environmental Technology, Technical Symposium and Workshop. Sponsored by Strategic Environmental Research and Development Program. November 30 to December 2, 1999.

Gertner, G., A.B. Anderson, and B. MacAllister. 2000. Error Budgets For Predicted Disturbance Due To Training Activities At Fort Hood. 2000 6th Annual SERDP/ESTCP Symposium "Environmental Challenges for the Next Decade", Alexandria, VA.

Gertner, G., A.B. Anderson, and B. MacAllister. 2001 Effect and Uncertainty of DEM Spatial Resolutions on Predicting Topographical Factor for Soil Loss Estimation. 2001 7th Annual SERDP/ESTCP Symposium. November 2001. Washington, DC.

Technical presentations that are part of the technology transfer process to communicate project results to others in the R&D community.

Wente, S., S. Fang, G. Gertner, and A. Anderson. 2000. Error Budgets For Predicted Disturbances Due to Training Activities. 2000 ASA Annual Meetings, Minneapolis, MN, Nov 5-9.

Wang, G., S. Fang, G.Z. Gertner & A.B. Anderson. 2000. Uncertainty propagation and partitioning in spatial prediction of topographical factor for RUSLE. Proceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, July 12-14, 2000, at Amsterdam, the Netherlands. p.717-722.

Fang, S., G.Z. Gertner, G. Wang, and A.B. Anderson. 2001. An Uncertainty Analysis Procedure for Analyzing Joint Multilevel Spatial Simulations of a Model. 13th Annual Kansas State University Conference on Applied Statistics in Agriculture. April 29- May 1, 2001, Manhattan KS.

Wang, G., G.Z. Gertner, S. Wente, and A.B. Anderson. 2001. Vegetation classification and accuracy assessment using image-aided sequential indicator co-simulation. Conference proceedings (CD) of American Society of Photogrammetry and Remote Sensing (ASPRS) 2001 - Gateway to the New Millennium, April 23-27, America's Center St. Louis, Missouri, USA.

Gertner, G., D. Jones, S. Wente, and A. Anderson. 2001. Appropriate Spatial Resolution For Vegetation Cover Mapping Based On LCTA At Fort Hood, Texas. 2001 ITAM Workshop, Nashville TN.

Wang, G., G.Z. Gertner, V., Singh, and P., Parysow. 2000c. Temporal and spatial prediction and uncertainty of rainfall-runoff erosivity for revised universal soil loss equation. Modeling Complex Systems Conference, July 31 – August 2, 2000, in Montreal, Canada.

Gertner, G., G. Wang, D. Jones, Shinkareva, P. Parysow, A.B. Anderson, and B. MacAllister. Spatial And Temporal Prediction And Uncertainty Analysis of RUSLE R Factor. 2001 ITAM Workshop, Nashville TN.

Wente, S., D. Jones, G. Gertner, and A.B. Anderson. 2000. Error Budgets For Predicted Disturbance Due To Training Activities at Fort Hood. 2000 ITAM 9th Annual Workshop, 23-28 August 2000, Richmond, VA.

Wente, S., D. Jones, G. Gertner, and A.B. Anderson. 2000. Uncertainty Assessment For Ecological Modeling and Simulation. 2000 ITAM 9th Annual Workshop, 23-28 August 2000, Richmond, VA.

Technical presentations that are part of the technology transfer process to communicate project results to the ITAM user community.

Anderson, A. Improved Units of Measure for Training and Testing Area Carrying Capacity (SERDP CS01102). Range Commanders Council (RCC), Environmental Group 12th Meeting, 19-21 October 1999, Yuma Proving Ground, Yuma Arizona.

Wente, S. S. Fang, G. Gertner, and A. Anderson. 2000. Uncertainty Assessment for Ecological Modeling and Simulation: Error Budgets for an Erosion Model at Fort Hood. 9th Annual Integrated Training Area Management (ITAM) Workshop, 22-24 Aug 2000, Richmond Va.

Wente, S. S. Fang, G. Gertner, and A. Anderson. 2000. Error Budgets For Predicted Disturbances Due to Training Activities at Fort Hood. 9th Annual Integrated Training Area Management (ITAM) Workshop, 22-24 Aug 2000, Richmond Va.

Gertner, G., D. Jones, S. Wente, and A. Anderson. 2001. Appropriate Spatial Resolution For Vegetation Cover Mapping Based On LCTA At Fort Hood, Texas. 2001 ITAM Workshop, Nashville TN.

Gertner, G., G. Wang, D. Jones, Shinkareva, P. Parysow, A.B. Anderson, and B. MacAllister. Spatial And Temporal Prediction And Uncertainty Analysis of RUSLE R Factor. 2001 ITAM Workshop, Nashville TN.

Programmatic presentations are part of the technology transfer process to coordinate integration of project products with organizations that manage the technology transfer processes. At each of these meetings project status and product development of error and uncertainty tools was discussed.

Carrying Capacity Research and Development. Annual Conservation Technology Team (CNTT) Meeting, 11-12 October 1999, Aberdeen Proving Ground, MD.

Land Capability/Characterization R&D Initiatives. Army Training and Testing Area Carrying Capacity (ATTACC) Executive Management Committee (EMC) Annual Meeting. 16 December 1999, Arlington Va.

Carrying Capacity Research and Development, Integrated Training Area Management (ITAM) Program Management Review (PMR), St. Cloud, MN. August 1999.

Carrying Capacity, LMS Workshop, 16-17 Nov 1999. Vicksburg, MS.

Anderson, A. LMS Carrying Capacity Related Projects. Fort Hood Military Field Application In-Progress Review, 4-5 April 2000, Killeen, Texas.

Carrying Capacity Research and Development. Annual Conservation Technology Team (CNTT) Meeting, 9-10 May 2000, Champaign, IL.

Land Capability/Characterization R&D Initiatives. Integrated Training Area Management (ITAM) Program Management Review (PMR) Annual Meeting. 29 February through 1 March 2000, Fort Eustis, Va.

Carrying Capacity Research and Development briefing to the ITAM Installation Steering Committee Chairman (IISC), Norfolk, VA. March 2000.

Carrying Capacity Research and Development briefing to the Army Training and Support Center (ATSC), Norfolk VA. November 2000.

Carrying Capacity Research and Development briefing to the Army Environmental Center (AEC), Aberdeen Proving Ground, MD. November 2000.

Carrying Capacity Research and Development, Integrated Training Area Management (ITAM) Executive Management Committee (EMC), Aberdeen Proving Ground, MD. December 2001.

Carrying Capacity Research and Development, Integrated Training Area Management (ITAM) Program Management Review (PMR), Norfolk VA. March 2001.

Carrying Capacity Research and Development briefing to the Army Training and Support Center (ATSC), Norfolk VA. November 2001.

Technical manuscripts (reviewed) in support of our transition plan.

Wang, G., G.Z Gertner, X. Xiao, Steven Wente and A.B.Anderson 2001. Appropriate plot size and spatial resolution for mapping multiple vegetation types. Photogrammetric Engineering and Remote Sensing, 67(5):575-584.

Parysow, P., G.Z. Gertner and J. Westervelt 2001. Efficient approximation for building error budgets for large and computationally-intensive process models. Ecological Modeling. 135:111-125.

Wang, G., G.Z. Gertner, X. Liu, and A. Anderson 2001. Uncertainty assessment of soil erodibility factor for Revised Universal Soil Loss Equation. CATENA 46: 1-14.

Fang, S., G.Z. Gertner and D. Price. 2001. Uncertainty analyses of a process model when vague parameters are estimated with Entropy and Bayesian Methods. Journal of Forest Research. J. For. Res. 6:13-19.

Wang, G., G.Z. Gertner, P. Parysow and A. Anderson. 2001. Spatial prediction and uncertainty assessment of topographic factor for RUSLE using DEM. ISPRS Journal of Photogrammetry and Remote Sensing, 56 (1) 65-80.

Parysow, P., G. Wang, G., G.Z. Gertner and A. Anderson. 2001. Assessing uncertainty of erodibility factor in the National Cooperative Soil Survey: A case study at Fort Hood, Texas. Journal of Soil and Water Conservation, 56 (3) 206-210.

McIsaac, G., M. David, G.Z. Gertner and D. Goolsby 2001. Net anthropogenic N input to the Mississippi River Basin and nitrate flux to the Gulf of Mexico. Nature (Brief Communication) 414: 166-167. (**Uncertainty analysis done with SERDP software**)

Gertner, G., G. Wang, S. Fang, and A. Anderson 2001. Error budget assessment of the effect of DEM spatial resolution in predicting topographical factor for soil loss estimation. Soil and Water Conservation (accepted).

Wang, G., G.Z. Gertner, V. Singh, S. Shinkareva, P. Parysow and A. Anderson 2001. Spatial and temporal prediction and uncertainty for complex systems – a case study in rainfall and runoff erosivity for soil loss. Ecological Modeling (Special Issue on Modeling Complex Ecological Systems) (accepted).

Wang, G., S. Wente, G. Z. Gertner, and A. Anderson 2001. Improvement in mapping vegetation cover factor for universal soil loss equation by geo-statistical methods with Landsat TM images. International Journal of Remote Sensing (accepted).

Wang, G., S. Fang, S. Shinkareva, G.Z. Gertner, and A. Anderson 2001. Uncertainty propagation and error budgets in spatial prediction of topographical factor for Revised Universal Soil Loss Equation (RUSLE). Transactions of the American Society of Agricultural Engineers (Accepted).

Parysow, P. and D. Tazik 2001. Assessing the effect of estimation error on population viability analysis: an example using the black-capped vireo. Submitted to Conservation Biology.

Gertner, G., G. Wang , S. Fang, and A. Anderson 2001. Mapping and uncertainty of predictions based on multiple primary variables from joint co-simulation with TM image. Remote Sensing of Environment. (In review)

Parysow, P., G.Wang, G.Z. Gertner and A. Anderson 2001. Spatial uncertainty analysis for mapping soil erodibility based on joint sequential simulation. CATENA (In review).

Gertner, G., S. Fang, G. Wang and A. Anderson 2001. Partitioning spatial model uncertainty when inputs are from joint simulations of correlated multiple attributes. International Journal of Geographic Information Systems. (In review)

Wang, G., G.Z Gertner, S. Fang, and A.B.Anderson 2001. Mapping multiple variables for predicting soil loss by joint sequential co-simulation with tm images and slope map. Photogrammetric Engineering and Remote Sensing. (In review)

Fang, S., G.Z. Gertner, S. Shinkareva, and G. Wang. 2001. An Improved Sampling Procedure for Non-uniform Distributions in Fourier Amplitude Sensitivity Test (FAST). Computational Statistics (In review).

Fang, S., S. Wente, G.Z. Gertner, G. Wang, and A.B. Anderson. 2001. Uncertainty analysis of predicted disturbance from off-road vehicular traffic in complex landscapes. Environmental Management (In review).

Mendoza, G., A. Anderson, and G.Z. Gertner 2001. Uncertainty analysis of predicted disturbance from off-road vehicular traffic in complex landscapes. Environmental Management (In review).

Mendoza, G., A. Anderson, and G.Z. Gertner 2001. Allocating training areas in military installations: An integreated multicriteria analysis and GIS approach. Journal of Environmental Planning and Management (In review).

McIsaac, G., M. David, G.Z. Gertner and D. Goolsby 2001. Relating N inputs to the Mississippi River Basin and nitrate flux in the Lower Mississippi River: A comparison of approaches. Journal of Environmental Quality. (In review).

Wang, G., G.Z. Gertner, P. Parysow and A. Anderson 2000. Spatial prediction and uncertainty analysis of topographic factors for the Revised Universal Soil Loss Equation (Rusle). Journal of Soil and Water Conservation. Third Quarter 2000, p.373-382.

Gertner, G.Z.; S. Fang and J.P. Skovsgaard 1999. A Bayesian approach for estimating the parameters of a forest process model based on long-term growth data. Ecological Modelling 119:249-265.

Additional technical manuscripts in support of our transition plan.

Pablo, P. and D. Tazik 2001. Assessing the effect of estimation error of population viability Analysis: An example using the black-capped vireo. USACERL Technical Report. ERDC/EL MP-01-1.

Gertner, G. 2001. Comparison of computationally intensive spatial statistical methods for generating inputs for spatially explicit error budgets. In: Proceedings of Conference on Forest Biometry, Modeling and Information Sciences. Greenwich, UK. Sponsored by University of Greenwich School of Computing and Mathematical Sciences; and the International Union of Forestry Research Organization. (In press).

Fang, S. and G. Gertner 2001. Analysis of parameters of two growth models estimated using bayesian methods and nonlinear regression. In: Proceedings of Conference on Forest Biometry, Modeling and Information Sciences. Greenwich, UK. Sponsored by University of Greenwich School of Computing and Mathematical Sciences; and the International Union of Forestry Research Organization. (In press).

Cao, X. and G. Gertner 2001. Error Budgets for a Spatially Explicit Biodiversity Monitoring/Modeling System. In: Berichte der Schriftenreihe Freiburger Forstliche Forschung. XXI IUFRRO World Congress 2000. 7-12 August 2000, Kuala Lumpur Asia. (Ed. Barbara Koch). In press.

Wang, G., G. Gertner, S. Wente and A. Anderson 2001. Vegetation classification and accuracy assessment using image-aided sequential indicator co-simulation. American Society for Photogrammetry and Remote Sensing Annual Conference Proceedings. St. Louis, MO. April 23-27, 2001. 12p.

Gertner, G.Z., G. Wang, P. Parysow, A. Anderson 2000. Application and comparison of three spatial statistical methods for mapping and analyzing soil erodibility. In: Proceeding of entitled, Agricultural, Biological, and Environmental Statistics Conference. Manhattan, Kansas p.204 to 216.

Wang, G., G. Gertner, V. Singh, S. Shinkareva, P. Parysow and A. Anderson 2001. Spatial and temporal prediction and uncertainty analysis of rainfall and runoff erosivity for revised universal soil loss equation. USACERL Technical Report ERDC/CERL TR-01-39.

Fang, S. and G. Gertner 2000. Uncertainty analysis of a pipe model based on correlated distributions. In: Proceedings entitled, Agricultural, Biological, and Environmental Statistics Conference. Manhattan, Kansas. p.66 to 79.

Fang, S. and G. Gertner 2000. Uncertainty estimation of the self-thinning process by maximum-entrophy principle. In: Proceeding of Integrated Tools for Natural Resources Inventories in the 21 Century. IUFRO Conference. Editors: Mark Hansen and Thomas Burk. August 16-20, 1998.

Wang, G., S. Fang, G.Z. Gertner and A. Anderson 2000. Uncertainty propagation and partitioning in spatial prediction of topographical factor for RUSLE. IN: Proceedings of Fourth International Conference on Spatial Uncertainty. Amsterdam, Holland. p. 717-722.

Cao. Xiangchi, G. Gertner, B. MacAllister and A. Anderson. 2000. Errors in environmental assessments: A error-budget model for plant populations. USACERL Technical Report ERDC/CERL TR-00-12.

Cao, X., G. Z. Gertner, and A. Anderson 2000. Stochastic Models of Plant Diversity: Application to White Sands Missile Range. USACERL Technical Report. ERDC/CERL TR-00-5.

Parysow, P., G.Z. Gertner and J. Westervelt 1998. Efficient approximation for building error budgets for large and computationally-intensive process models. IN: Proceedings entitled, Modeling Complex Systems Conference. New Orleans, U.S.A.

Uncertainty analysis software in support of our transition plan.

We developed three versions of an uncertainty analysis software mainly focusing on the variance-based methods widely used for uncertainty assessment in studies of natural resources, ecological and environmental systems, chemistry and nuclear reactions. In these versions, the same methods have been included. Three versions are referred to as LEVEL 1, LEVEL 2 and LEVEL 3. Each of the three levels is briefly described below.

LEVEL 1 UNCERTAINTY SOFTWARE: ATTACC COMMUNITY

The first version of the uncertainty software generates error budget for the fixed ecological component models of ATTACC. Environmental component of ATTACC is a spatially explicit version of Revised Universal Soil Loss Equation (RUSLE). The software is very easy to use, and is an integral part of the ATTACC software toolkit. The inputs for the uncertainty software are maps of means, predictions and variances and resulting outputs are the regional and local uncertainty maps. The software is ArcView 3.2 (Hutchinson and Daniel, 1997) compatible. USACERL programmers have actively been involved in the development of software. The software has been verified by undergraduate and graduate students in the Department of Natural Resources and Environmental Sciences at the University of Illinois. They were widely used in class course assignments. The software is internally documented with a number of real examples.

This software will be distributed by Alan Anderson this January (2002) to the military community (Integrated Training Area Management and Configuration Management Working Group). In Appendix 1 is the letter that will accompany the software.

LEVEL 2 UNCERTAINTY SOFTWARE: MILITARY RESEARCH AND DEVELOPMENT COMMUNITY

The second version of uncertainty software has been developed for the military research and development community. A series of programs have been written to work in an integral fashion with the commercially available software package S-Plus for Windows (MathSoft, Incorporated). S-Plus is a widely used both in and out the military and is known for its statistics and graphics. The uncertainty software can be used to analyze typical models used in land management modeling and decision support. The documentation for this level will be incorporated into the software. A website being developed at the University of Illinois will allow easy downloading of the software with corresponding documentation. Public access to the website will be available soon. The software has been verified over the last two years with analytical approaches.

LEVEL 3 UNCERTAINTY SOFTWARE: UNIVERSITY RESEARCH COMMUNITY

The third version of uncertainty software has been developed for the academic research community. A series of programs have been written to work in an integral fashion with the Geostatistical Software Library (Deutsch,C. and A. Journel, 1997, Geostatistical Software Library and User's Guide. Applied Geostatistics Series. Oxford University Press.). The Geostatistical Software is public domain. Our uncertainty software was written in FORTRAN. Documentation has been written describing the methodology and to how used the software. Worked examples are included. The source code is documented in detail so it can be easily adapted for the users with particular applications. A website developed at the University of Illinois allows easy downloading of the software with documentation. Through out the project, the programs have been utilized extensively in developing the error budgets for the ATTACC case study at Fort Hood, Texas. The software has been verified by undergraduate and graduate students in the Department of Natural Resources and Environmental Sciences at the University of Illinois. These programs can be directly adapted to future enhancements of ATTACC. The User's Guide of Level 3 is organized by analysis method. The first chapter provides a very brief introduction to uncertainty analysis and other general information about uncertainty analysis software. Subsequent chapters each describe a method and provide listings of corresponding FORTRAN programs, a description of how to use the method and programs, required input information, and examples to demonstrate the application of the method and programs. The last chapter describes the software utility programs and their usage.

The software was developed in the FOTRAN language on Microsoft Developer Studio (Fortran PowerStation 4.0, 1993-1994) software. Its source files make use of procedures from IMSL (MSIMSL) for random number generation, probability computation, and regression analysis. The executable files of the FORTRAN programs already include these procedures. For spatial studies, data files for predicted and variance maps of variables have to be generated using a Geostatistical Software Library and re-formatted to the general format of the ASCII input/output data files for ArcInfo® or ArcView GIS© (Hutchinson and Daniel, 1997). Programs for performing these transformations are included in the software as utility programs.

Website used to disseminate the SERDP develop uncertainty software in support of our transition plan.

We have developed a website for the project 'Error and Uncertainty Analysis for Ecological Modeling and Simulation'. It will be fully functional at end of January, 2002. This website

briefly describes this project, methodology, research team, and achievements including publications and software. Level 2 and Level 3 software can be download from the website. The website address is:

http://uncertainty.nres.uiuc.edu

CONCLUSION

Methodology and software.

In this project, we developed a GIS based methodology for spatial modeling, mapping, and uncertainty analysis of natural resources, ecological and environmental systems. The methodology deals with the methods that can be used for optimal sampling design, determining appropriate spatial resolution, spatial modeling and mapping, and spatial uncertainty budgets. The methodology assumes that sample data of a variable are spatially similar to each other within a range of separation distance of data given a direction and sample data of a variable may also be spatially correlated with sample data of another variable. The development of the methodology is based on measuring and modeling spatial variability of variables, and spatial cross variability between the variables. The methodology is characterized:

The methodology was developed on a GIS platform so that prediction and uncertainty analysis could be done on the basis of pixel by pixel. The methodology thus provides users and managers with detailed spatial information for management plans and error reduction.

To understand and obtain spatial variability of a variable is the basis for accurately mapping the variable and making the spatial uncertainty budget. Thus, simultaneously capturing within plot spatial variability and regional spatial variability of a variable is the key to determine plot size. This can be realized by developing the within plot semi-variogram and regional semi-variogram at different plot sizes. Further, introducing measurement cost into the relationship between the semi-variogram models and plot sizes makes a linkage of plot size and sample size. This method is applicable for ground data and auxiliary data such as satellite images. Using the method, optimal plot size and sample size can be successfully determined so that ground data are collected at cost-efficient and spatial variability of variables is captured.

Before spatial modeling, appropriate spatial resolution should be chosen so that desired information of spatial variability and accuracy requirements are met. Optimal plot size for collecting ground data is thus consistent with appropriate spatial resolution for spatial modeling. Both appropriate plot size and spatial resolution should be determined together. This method was tested at Fort Hood for vegetation classification and prediction of

topographical factor LS. Moreover, we developed a method to model loss of spatial information due to scaling (from one resolution to another), successfully applied it to the LS factor when it is calculated using different digital elevation models, and derived the loss of spatial information and detected the anisotropy of this factor.

The simulation algorithms we developed for mapping include sequential Gaussian simulation, sequential indicator simulation, and joint sequential simulation. When auxiliary data such as satellite images and digital elevation models are used, the methods become co-simulation. The simulation and co-simulation algorithms provide expected and unbiased estimates for areas and sub-areas, reliable estimates for any unknown locations, their estimation variances for continuous variables, and their classification and misclassification probabilities for categorical variables. The uncertainty measures provide users and managers with spatial uncertainty information, help them use the maps with caution and make detailed management plans, and also make it possible to do spatial variance partitioning. Integrating simulation algorithms and error budget methods can thus realize the spatial error and uncertainty analysis for ecological modeling and simulation.

In the co-simulation algorithms, use of auxiliary data can significantly improve estimation of variables especially reproduction of spatial statistics including spatial distribution and spatial variability of estimates, and spatial cross variability between variables. The auxiliary data provide control surfaces of the spatial variability and bridge the interactions among the variables for spatial cross variability. When variables are correlated with each other, joint simulation or co-simulation can reduce uncertainties of estimates compared to individual simulations or co-simulations. Using spatial information from neighboring locations can also reduce variances of estimates. Compared to traditional methods such as supervised and unsupervised classification and stratification, and regression modeling, the co-simulations can generate more accurate maps in addition to uncertainty measures. The uncertainty measures such as variance maps, classification and misclassification maps make it possible to do spatial accuracy assessment, while traditionally global accuracy assessment is only done. The sequential Gaussian co-simulation should be selected for mapping of variables that have normal distributions, while the sequential indicator co-simulation should be applied to categorical variables and the variables that are not normally distributed and the extreme values are important for management plans. The joint sequential co-simulation should be used for mapping multiple variables that are spatially correlated with each other.

We improved and developed several error budget methods so that that can be used to do spatial uncertainty analysis. That is, an error budget can be done on the basis of pixel by pixel. The improved methods include Tayler series, response surface modeling, Fourier Amplitude Sensitivity Test (FAST), sequential sampling based method, and regression

modeling. These methods have been applied to the case study for prediction of soil erosion and led to reasonable results of uncertainty budgets.

The FAST method is computationally efficient, but requires all the input parameters are independent. The Taylor series expansion based methods can handle interactions among input parameters but assume the model functions can be continuously differentiable. The response surface modeling method is improvement of Taylor Series methods and can be used to perform uncertainty analyses of complicated nonlinear models. When nonlinear models are complicated, linear models can be used to represent them based on their responses surface relation. Then, the partial derivatives of the response surface models (linear models) can be easily obtained and the Taylor series method applied to investigate the uncertainty contribution of the model input parameters. The sequential sampling based method investigates uncertainty propagation using the behavior of the model variance corresponding to the marginal distribution of input parameters. The regression modeling is integrated with the joint sequential co-simulation to make a spatial error budget for mapping multiple spatially correlated variables, and the integration can partition the total variance of a dependent variable into the variation of its independent variables, interactions among them, and the effect of spatial information from neighboring locations.

The methodology above has been computationally programmed into the uncertainty analysis software. The package refers to three level versions: LEVEL 1, LEVEL 2 and LEVEL 3. At level 1, the software can be used by ATTACC community to generate error budgets for the fixed ecological component models of ATTACC. Once the required maps are input, the error budgets are created. At level 2, the software can be applied by military research and development community to analyze typical models used in land management modeling and decision support. At level 3, the software can be used by university research communities. Assuming any spatial models, users have to generate prediction and variance maps of variables before doing error budgets. These three levels of this software have been tested by the research team at Fort Hood area for prediction of soil erosion. It is expected that the methodology above and its software can be applied to other areas of natural resources, ecological and environmental systems.

Case study.

We applied the methodology and software to the case study at Fort Hood, where soil erosion is predicted by the Universal Soil Loss Equation (USLE) or the Revised USLE (RUSLE). Soil loss (A) is a function of six input factors including rainfall-runoff erosivity (R), soil erodibility (K), slope length (L), slope steepness (S), vegetation cover and management (C), and support practice (P), the case study was first done for spatial prediction and uncertainty

analysis of each input factor from its primary variables, and then for spatial prediction and uncertainty of soil erosion from the input factors. At the same time, various methods for mapping and uncertainty analysis were compared.

Using joint sequential co-simulation with Landsat TM images and digital elevation model (DEM), we generated prediction and variance maps of the input factors and soil erosion at Fort Hood. These maps are unbiased at the probability of 95%. The spatial distribution of the predicted values is similar to that of corresponding data set. Large LS factor values and small C and K factor values were predicted at the east parts, while small LS factor values and large C and K factor values were obtained at the west parts. The predicted R factor values slightly increase from the west to the east. The calculated values of soil loss thus are higher at the west and north parts. Therefore, the east and northeast parts of Fort Hood have better land conditions than the west parts. At the east and northeast parts the probability at which erosion status of less than 1.0 may take place is higher than 0.5, while the probability at which erosion status of greater than 2.0 may occur is less than 0.5. At the west parts, erosion status of less than 1.0 may take place at the probability less than 0.5, while erosion status of greater than 2.0 may occur at a probability more than 0.5. Because of change of the C factor over time due to disturbance, the soil erosion at Fort Hood decreases from 1989 to 1991, then increases from 1991 to 1996.

Generally, at the areas with larger prediction values the estimation variances are higher, and vice versa. All the input factors have positive co-variances to soil loss. However, most of the co-variances between factors LS and C are negative. This is because a steeper area implying larger LS factor, but less training activities and disturbance of vegetation resulting in higher vegetation cover and less C factor. Generally, relative variance contributions of the input factors to the uncertainty of predicted soil loss vary spatially, depending on locations. The largest relative variance contribution to the uncertainty of predicted soil loss comes from LS factor, then C factor, K factor, and R factor. That is, main uncertainty source is the LS and C factor. Along the hilly area boundaries, large slopes and up-slope contributing areas determine the amount of soil erosion, but high ground and vegetation cover may significantly reduce soil loss. At the flat areas, slope is very close to zero, thus LS also close to zero, and very little or no soil erosion happens.

We compared the results of predicted LS factor based on the empirical equations of LS using the sample data and based on a physically based LS calculation equation using digital elevation model (DEM). The results showed the use of DEM led to more reasonable and consistent prediction map of soil erosion with the topographical features at Fort Hood. That is, soil erosion may be high along the hilly area boundaries and low at the flat areas. This feature is not so clear when the sample data were used to generate the map of soil loss. In other words, a high dense sample may be needed at the case. We studied appropriate spatial

resolution of DEM and found that at Fort Hood a sufficient spatial resolution (pixel size) of DEM should be less than 5m by 5m. We also detected the anisotropy of spatial variability of the LS factor derived using a DEM. We completed an uncertainty budget of the LS factor using Fourier Amplitude Sensitivity Test. Given a spatial resolution, the uncertainty in predicting the topographical factor LS using a DEM mainly come from slope in the areas of gentle slopes and up-slope contributing area in steep areas. The model parameters contributed little in terms of variance.

The LS factor has reverse J shape distribution. When the sample data of slope steepness and slope length are used for its prediction, the sequential indicator simulation should be selected. The number of indicators (cutoff values for indicator coding of original data) should be equal or larger than seven, and the number of simulation runs should not be less than 500. The slope steepness contributes the largest part of uncertainty for the LS factor, then slope length, and the model parameters and measurement errors contribute a little.

We investigated appropriate plot size and sample size. The results suggest the plot size of 100 transect line is appropriate for mapping of multiple land cover categories. The existing sample size of 200 LCTA plots might be sufficient for mapping overall vegetation cover and also for mapping vegetation cover and management factor C because the C factor is related to the overall vegetation cover in percent. But, the sample size might be insufficient for classification of land cover types at Fort Hood. We mapped land cover types at Fort Hood using various methods. The sequential indicator co-simulation with Landsat TM images led to the best results. Especially, the misclassification map provided spatial information of classification accuracy, and the accuracy varied spatially. However, the overall classification accuracy was still low and the reason might be because of insufficient sample size as mentioned above.

The vegetation cover and management factor C has a distribution close to normal. We directly simulated the C factor using various methods. The traditional methods produce much worse results than the sequential Gaussian co-simulation with Landsat TM images. We also created the prediction and variance map of the C factor using the joint sequential co-simulation with Landsat TM images by jointly mapping ground cover, canopy cover, and minimum rain drip vegetation height. The C factor values were higher at the west parts of Fort Hood due to disturbance of ground and canopy cover and lower at the east parts due to higher vegetation of wood land and less disturbance. The disturbance and the C factor at Fort Hood decreased from 1989 to 1991, and then increased from 1991 to 1995, especially at the west parts. The C factor was most sensitive to the ground cover, then canopy cover, and vegetation height. However, the main uncertainty source varied depending on locations. To estimation of disturbance, mapping especially vegetation mapping was the main uncertainty source, and the uncertainty of the model parameters were relative not important.

If one K value could be considered representative of each soil series, results from this project do not support agreement with the information provided by the NCSS. In fact, the sampled K means for all the soil series are significantly different from the K values proposed by the NCSS. The assumption that each soil series might be represented by only one K value does not seem to agree with the sample results either, based on the coefficients of variation estimated within those soil series. Considering that those surveys were conducted in 1977 and 1985, changes over time in soil structure may have contributed to the differences found. The published values tend to underestimate soil erodibility.

We integrated a joint sequential simulation and regression model for mapping the soil erodibility factor K from five soil properties and making the spatial uncertainty budget. The uncertainty of soil erodibility of a pixel was mainly propagated from its own soil properties. Overall, the largest uncertainty source was very fine sand and silt, and the smallest uncertainty source was structure. The largest and smallest uncertainty contributors are different soil properties at different locations. Considering the correlation between the soil properties led to reduction of uncertainty. The soil properties of neighbor pixels contributed negative uncertainty to soil erodibility.

The rainfall-runoff erosivity factor R map was created by extracting from the predicted map derived for a large area with about 250 rainfall stations. Although soil erosion is less sensitive to the R factor compared to other factors, the study showed that the R factor had large spatial variability over space, and even within a relative small area such as Fort Hood with an area of 87,890 ha, the spatial variability may not be neglected. This suggests that it should be very careful to use a constant R factor over space for a specific area. Additionally, there was a high temporal variability of the R factor in the time series of seasons and half months. The R factor value of Fort Hood from the published isoerodent map is 270, however, it is much lower than that obtained in this study. This difference may be because of global climate change. Additionally, we suggested a new R factor map might be needed and it might be created using a Gaussian simulation algorithm.

REFERENCES

Almeida, A.S., 1993, Joint simulation of multiple variables with a Markov-type coregionalization model, The Department of Applied Earth Sciences, Stanford University, Ph.D. Dissertation, 199 p.

Almeida, A.S., and A.G. Journel, 1994. Joint simulation of multiple variables with a Markov-Type coregionalization model. Mathematical geology, 26:565-588.

Anderson, A.B., R. Hunt, L. Chenkin, D. McFerren, & P. Sydelko, 1996. Evaluation of Land Value Study (ELVS)/ Army Training and Testing Area Carrying Capacity (ATTACC). In 5th Annual LRAM/ITAM Workshop Proceedings. August 27-29, 1996. LaCrosse, Wisconsin.

Anderson, A.B., W.L. Sprouse, D.G. Kowalski, and P.J. Guertin, 1995. LCTA Users Interface Program, Users Manual Version 1.0. USACERL ADP Report 95/24, August 1995, ADA 300797. 156 pp.

Anderson, A.B., W.L. Sprouse, D.G. Kowalski, and R.B. Brozka. 1995. Land Condition Trend Analysis Data Collection Software Users Manual: Version 1.0. USACERL ADP Report 95/13, July 1995, ADA 299981. 88pp.

Arvanitis, L.G., and R.M. Reich, 1991. Spatial patterns and precision of estimates, Conference on Optimal Designs of Forest Experiments and Surveys, p.107-117.

Atkinson, P.M., & F.M. Danson, 1988. Spatial resolution for remote sensing of forest plantations. Proceedings of IGARSS ’88 Symposium, 13 – 16 September, Edinburgh, Scotland, p. 221-223.

Atkinson, P.M., & P.J. Curran, 1997. Choosing an appropriate spatial resolution for remote sensing investigations: Photogrammetric Engineering & Remote Sensing, 63(12):1345-1351.

Atkinson, P.M., 1997. On estimating measurement error in remotely-sensed images with the variogram: Int. J. Remote Sensing, 18(14):3075-3084.

Barata, M.T., M.C. Nunes, A.J. Sousa, F.H. Muge, and M.T. Albuquerque, 1996. Geostatistical estimation of forest cover areas using remote sensing data. E.Y. Baafi and N.A. Schofield (eds.), Geostatistics Wollongong '96, 2:1244-1257.

Benkobi, L., M.J. Trlica, and J.L. Smith, 1994. Evaluation of a refined surface cover subfactor for use in RUSLE, Journal of Range Management, 47(1):74-78.

Biesemans, J., M.V. Meivenne, and D. Gabriels, 2000. Extending the RUSLE with Monte Carlo error propagation technique to predict long-term average off-site sediment accumulation, J. of Soil and Water Conservation, First Quarter:35-42.

Blake, G.R., and K.H. Hartge, 1986. Bulk density. *In*: Methods of Soil Analysis. Part 1, 2nd ed. American Society of Agronomy and Soil Science Society of America Book, Madison, WI. A. Klute (editor)

Campbell, J.B., 1996. Introduction to Remote Sensing. The Guilford Press.

Cao, C. & N.S. Lam, 1997. Understanding the scale and resolution effects in remote sensing and GIS. In D.A. Quattrochi and M.F. Goodchild (Eds.), Scale in remote sensing and GIS. p. 57-72.

Cao, X. and G. Gertner 2001. Error Budgets for a Spatially Explicit Biodiversity Monitoring/Modeling System. In: Berichte der Schriftenreihe Freiburger Forstliche Forschung. XXI IUFRRO World Congress 2000. 7-12 August 2000, Kuala Lumpur Asia. (Ed. Barbara Koch). In press.

Chiles, J.P. and P. Delfiner, 1999. Geostatistics: Modeling Spatial Uncertainty. Wiley, New York.

Clarke, K.C., 1985. A comparative analysis of polygon to raster interpolation methods: Photogrammetric Engineering and Remote Sensing, 51(5):575-582.

Collins, D.C., & R. Avissar, 1994. An evaluation with the Fourier Amplitude Test (FAST) of which land-surface parameters are of greatest importance in atmospheric modeling: Journal of Climate, 7:681-703.

Congalton, R.G., 1988. Using spatial autocorrelation analysis to explore the errors in maps generated from remotely sensed data. Photogrammetric Engineering and Remote Sensing, 54:587-592.

Cressie, N.A.C., 1991. Statistics for spatial data. John Wiley and Sons, Inc.

Cukier, R. I., C. M. Fortuin, K. E. Shuler, A. G. Petschek, and J. H. Schaibly, 1973. Study of the Sensitivity of coupled reaction systems to uncertainties in rate coefficients. I. Theory. Journal of Chemical Physics, 59: 3873-3878.

Dale, V.H., H.I. Jager, R.H. Gardner, and A.E. Rosen, 1988. Using sensitivity and uncertainty analyses to improve predictions of broad-scale forest development. Ecological Modelling 42:165-178.

De Cola, L., 1997. Multi-resolution covariation among Landsat and AVHRR vegetation indices. In: Quattrochi, D.A., Goodchild, M.F. (Eds.), Scale in remote sensing and GIS, Lewis Publishers, CRC Press, Inc., Boca Raton, Florida, p.73-92.

Demarais, S., D. Tazik, P. Guertin, & E. Jorgensen, 1999. Disturbance associated with military exercises. Ecosystems of the World 16. Ecosytems of Disturbed Gound. Lawrence Walker (editor). Elvevier, Amsterdam. p.385-396.

Dettinger, M. D. and J. L. Wilson, 1981. First order analysis of uncertainty in numerical models of grounderwater flow. Part I. Mathematical development. Water Resource Research 17: 149-161.

Deutsch, C.V., and A.G. Journel, 1998. Geostatistical software library and user's guide. Oxford University Press, Inc.

Diersing, V.E., R.B. Shaw, and D.J. Tazik, 1992. US Army Land Condition-Trend Analysis (LCTA) Program. Environmental Management , 16:405-414.

Downing, D.J., R.H. Gardner, and F.O. Hoffman, 1985. An Examination of response-surface methodologies for uncertainty analysis in assessment models. Technometrics. 27(2):151-163.

Englund E. J., and N. Heravi, 1994. Phased sampling for soil remediation, Environmental and Ecological Statistics, 1: 247-263.

Fang, S., 2000. Uncertainty analysis of biological nonlinear models based on Bayesian Estimation. Ph.D. Dissertation, University of Illinois at Urbana-Champaign, U.S.A.

Fang, S., G.Z. Gertner, S. Sinkaliva, and G. Wang. 2001a. An Improved Sampling Procedure for Non-uniform distributions in Fourier Amplitude Sensitivity Test (FAST). Computational Statistics and Data Analysis (in review).

Fang, S., S. Wente, G.Z. Gertner, G. Wang, & A.B. Anderson. 2001b. Uncertainty analysis of predicted disturbance from off-road vehicular traffic in complex landscapes. Environmental Management (in review).

Foster, G.R., L.D. Meyer, and C.A. Onstad, 1977. A runoff erosivity factor and variable slope length exponents for soil loss estimates. Trans. ASAE 20:683-687.

Freese, F., 1961. Relation of plot size to variability: an approximation. J. For., 56:679.

Friedl, M.A., 1997. Examining the effects of sensors resolution and sub-pixel heterogeneity on spectral vegetation indices: Implications for biophysical modeling. Scale in remote sensing and GISD (Eds. Quattrochi D., & M.F. Goodchild), CRC Press, Inc. p.113-140.

Gambill, C.W., H.V. Wiant, Jr., and D.O. Yandle, 1985. Optimum plot size and BAF. Forest Science, 31(3):587-594.

Gardner, R.H., and R.V. O'Neill, 1981. A comparison of sensitivity and error analysis based on a stream ecosystem model. Ecological Modelling 12: 173-190.

Gelb, A., J. Kasper Jr., R. Nash Jr., C. Price, and A. Sutherland Jr., 1974. Applied Optimal Estimation. The MIT Press, Cambridge, MA.

Gertner, G. 2001. Comparison of computationally intensive spatial statistical methods for generating inputs for spatially explicit error budgets. In: Proceedings of Conference on Forest Biometry, Modeling and Information Sciences. Greenwich, UK. Sponsored by University of Greenwich School of Computing and Mathematical Sciences; and the International Union of Forestry Research Organization. (In press).

Gertner, G., G. Wang, S. Fang, & A.B. Anderson. 2001c. Mapping and uncertainty of predictions based on multiple primary variables from joint co-simulation with TM image. Remote sensing of Environment (in review).

Gertner, G., S. Fang, G. Wang, & A.B. Anderson. 2001a. Partitioning spatial model uncertainty when inputs are from joint simulations of correlated multiple attributes. International Journal of Geographic Information Systems (in review).

Gertner, G., S., Fang, G., Wang , and A.B., Anderson. 2001b. An uncertainty analysis procedure for spatially joint simulation of multiple attributes. Thirteenth annual - Kansas State University Conference on applied statistics in agriculture, April 30 - May 2, 2001 (in press).

Gertner, G.Z., G. Wang, P. Parysow, & A.B. Anderson. 2000. Application and comparison of three spatial statistical methods for mapping and analyzing soil erodibility. Twelfth annual - Kansas State University Conference on applied statistics in agriculture, April 30 - May 2, 2000 p.66-79.

Gertner, G.Z., G. Wang, S. Fang, and Alan Anderson. 2001d. Error budget assessment of the effect of DEM spatial resolution in predicting topographical factor for soil loss estimation. Journal of Soil and Water Conservation (in press).

Gomez-Hernandez, J.J. and A.G. Journel, 1992. Joint sequential simulation of multiGaussian fields. In: Geostatistics Tróia 1992, Soars, A. (ed.), 1:85-94, Kluwer Academic Publishers, Kluwer, Dordrecht.

Goodchild, M.F., & D.M. Mark, 1987. The fractal nature of geographic phenomena. Ann. Assoc. Am. Geogr., 77, 265.

Goovaerts, P., 1997. Geostatistics for natural resources evaluation. Oxford University Press, Inc., 198 Madison Avenue, New York.

GRASS, 1993. Geographic Resources Analysis Support System (GRASS) version 4.1 user's reference manual. USA Corps of Engineers, Construction Engineering Research Laboratories, Champaign, Illinois, http://www.baylor.edu/grass/ (accessed May 6, 2001).

Hannes, S.J., L.A. Roberts, J.J. Warwick, and W.G. Cale, 1991. Testing the utility of the first order uncertainty analysis. Ecological Modelling, 58:1-23.

Haralick, R.M., K. Shanmugam, & I. Dinstein, 1973. Texture features for image classification. IEEE Transactions on Sustems, Man, and Cybernetics SMC-3: 610-621.

Heuvelink, G.B.M, 1998. Error propagation in environmental modelling. Taylor & Francis.

Holopainen, M. & G. Wang, 1998a. Calibration of digital aerial photographs for forest inventory and monitoring. International Journal of Remote Sensing. 19(4):677-696.

Holopainen, M. & G. Wang, 1998b. Accuracy of digitized aerial photographs for assessing forest habitats at plot level. Scandinavian Journal of Forest Research. 13:499-508.

Hudson, N., 1995. Soil Conservation. Iowa State University Press, Ames, Iowa, USA.

Hunner, G., Mowrer, H.T., and Reich, R.M., 2000. An accuracy comparison of six spatial interpolation methods for modeling forest stand structure on the Fraser Experimental Forest, Colorado. In Accuracy 2000, proceedings of the 4th international sysmposium on spatial accuracy assessment in natural resources and environmental sciences (G.B.M. Heuvelink, and M.J.P.M. Lemmens, eds.), Amsterdam, July 2000, Delft University Press, The Netherlands, p.305-312.

Hutchinson, S., and L. Daniel, 1997. Inside ArcView GIS. OnWord Press.

Iman, R. L. and J. C. Helton, 1988. An investigation of uncertainty and sensitivity analysis techniques for computer models. Risk Analysis 8(1):71-90.

Jansen, M.J.W., 1999. Analysis of variance designs for model output. Computer Physics Communications, 117:35-43.

Jansen, M.J.W., W.A.H. Rossing, and R.A. Daamen, 1994. Monte Carlo estimation of uncertainty contributions from several independent multivariate sources. In Proc. Congress Predictability and Nonlinear Modelling in Natural Sciences and Economics, 5-7 April 1993, The Netherlands.

Jarvis, P.G., 1995. Scaling process and problems. Plant, cell and environment, 18:1079-1089.

Jenny, H., 1941. Factors of soil formation. McGraw-Hill Book Co., New York.

Journel, A.G., and C.J. Huijbregts, 1978. Mining geostatistics. Academic Press, New York.

King, A., 1991. Translating models cross scales in the landscape. In M.G. Turner and R.H. Gardner (eds.), Quantitative methods in landscape ecology, Springer-Verlag, New York, p.479-517.

Kremer, J.N., 1983. Ecological implications of parameter uncertainty in stochastic simulation. Ecological Modelling, 18:187-207.

Lam, N., & D.A. Quattrochi, 1992. On the issues of scale, resolution, and fractal analysis in the mapping sciences, Prof. Geogr., 44, 88.

Lillesand, T.M., and R.W. Kiefer, 2000. Remote Sensing and Image Interpretation. John Wiley & Sons, Inc.

Lunetta, R., R. Congalton, L.K. Fenstermaker, J.R. Jensen, K.C. McGwire, & L.R. Tinney, 1991. Remote sensing and geographic information system data integration: Error sources and research issues. Photogrammetric Engineering & Remote Sensing, 57(6):677-687.

Mandelbrot, B.B., 1983. The fractal geometry of nature. W.H. Freeman and Company, New York.

McBratney, A.B., and R. Webster, 1981. The design of optimal sampling schemes for local estimation and mapping of regionalized variables—II program and examples, *Computers & Geosciences,* 7(4):335-365.

McBratney, A.B., and R. Webster, 1983. How many observations are needed for regional estimation of soil properties?, *Soil Science,* 135(3):177-183.

McBratney, A.B., R. Webster, and T.M. Burgess, 1981. The design of optimal sampling schemes for local estimation and mapping of regionalized variables—I theory and method, *Computers & Geosciences*, 7(4):331-334.

McCarthy, M.A., M.A. Burgman, and S. Ferson, 1995. Sensitivity analysis for models of population viability analysis. Biological Conservation, 73: 95-100.

McGregor, K.C., C.K. Mutchler, and A.J. Bowie, 1980. Annual R values in North Mississippi. Journal of Soil and Water Conservation, 35(2):81-84.

Mendoza, G., A. Anderson, and G.Z. Gertner 2001. GIS-based multicriteria land condition models for military training areas. Environmental Management (In review).

Mitášová, H., J. Hofierka, M. Zlocha, & L.R. Iverson, 1996. Modeling topographic potential for erosion and deposition using GIS. International Journal of Geographical Information Science, 10(5):629-641.

Moellering, H., and W. Tobler, 1972. Geographical variances. Geographical analysis, 4:34-64.

Moore, I.D., & G.J. Burch, 1986. Physical basis of the length-slope factor in the Universal Soil Loss Equation. Soil Science Society of America Journal, 50(5): 1294-1298.

Moore, I.D., and J.P. Wilson, 1992. Length-slope factors for Revised Universal Soil Loss Equation: simplified method of estimation. J. Soil and Water Cons., 47(5):423-428.

Morris, M. D., 1991. Factorial sampling plans for preliminary computational experiments. Technometrics, 33(2):161-174.

Mowrer, H.T., 1997. Propagating uncertainty through spatial estimation processes for old-growth subalpine forests using sequential Gaussian simulation in GIS. Ecological modeling, 98:73-86.

Myers, J.C., 1997. Geostatistical Error Management. Van Nostrand Reinhold, New York, NY.

Olea Ricardo A., 1984. Sampling Design Optimization for Spatial Functions: *Mathematical Geology,* 16(4):369-392.

O'Neill, R.V., and R.H. Gardner, 1979. Sources of uncertainty in ecological models. In B.P. Gardner, M.S Elzas, G.J. Klir and T.I. Oren (editors), Methodology in System's Modelling and Simulation. North Holland Publishing Co., Amsterdam, p.447-463.

O'Neill, R.V., R.H. Gardner, and J.B. Mankin, 1980. Analysis of parameter error in a nonlinear model. Ecological Modelling, 8: 297-311.

Openshaw, S., 1992. Learning to live with errors in spatial databases. Accuracy of spatial databases (Eds. Goodchild, M., & S. Gopal), Taylor & Francis Ltd. p.263-276.

Parysow, P., G. Wang, G.Z. Gertner, & A.B. Anderson. 2001. Assessing Uncertainty of Soil Erodibility Factor in the National Cooperative Soil Survey: A Case Study at Fort Hood, Texas. J. of Soil and Water Conservation, 56 (3) 206-210.

Parysow, P., G. Wang, G.Z. Gertner, & A.B. Anderson. 2001. Spatial uncertainty analysis for mapping soil erodibility based on joint sequential simulation (in research group review).

Peng, S., 1987. On the combination multi-temporal satellite and field data for forest inventories. Acta Forestalia Fennica, 200.

Reich, R.M., and L.G. Arvanitis, 1992. Sampling unit, spatial distribution of trees, and precision. North Journal of Applied Forest, 9(1):3-6.

Renard, K.G., and V.A. Ferreira, 1993. RUSLE model description and database sensitivity, J. of Environmental Quality, 22(3):458-466.

Renard, K.G., C.R. Foster, G.A. Weesies, D.K. McCool, and D.C. Yoder, 1997. Predicting soil erosion by water: A guide to conservation planning with the Revised Universal Soil Loss Equation (RUSLE). U.S. Department of Agriculture, Agriculture Handbook Number 703, U.S. Government Printing Office, SSOP Washington, DC.

Risse, L.M., M.A. Nearing, A.D. Nicks, and J.M. Laflen, 1993. Error assessment in the Universal Soil Loss Equation, Soil Science Society of America Journal, 57(3):825-833.

Rogowski, A.S., and J.K. Wolf, 1994. Incorporating variability into soil map unit delineations. J. Soil Sci. Soc. Am.. 58:163-174.

Rossing, W.A.H., R.A. Daamen, and M.J.W. Jansen, 1994a. Uncertainty analysis applied to supervised control of aphids and brown rust in winter wheat. Part 1. Quantification of uncertainty in cost-benefit calculations. Agricultural Systems 44: 419-448.

Rossing, W.A.H., R.A. Daamen, and M.J.W. Jansen, 1994b. Uncertainty analysis applied to supervised control of aphids and brown rust in winter wheat. Part 2. Relative importance of different components of uncertainty. Agricultural Systems 44: 449- 460.

Shannon, C.E., & W. Weaver, 1964. The mathematical theory of communication. University of Illinois Press, Urbana, IL.

Shaw, R.B., and D.G. Kowalski, U.S. Army Lands: A National Survey, The Center for Ecological Management of Military Lands CEMML TPS 96-1 (Colorado State University, Fort Collins, CO., 1996).

Shinkareva, S., G.Z Gertner, S. Fang, G. Wang 2001. Image-aided spatial accuracy assessment of land cover classification (in research group review).

Siegel, S.B., R.P. Hunt, C.L. Couvillon, A.B. Anderson, and P. Sydelko, 1996. Evaluation of Land Value Study. Proceedings of the 22nd Environmental Symposium & Exhibition. March 18-21, 1996., Orlando FL. p. 469-475.

Smith, A. E., P. B. Ryan, and J. S. Evans, 1992. The effect of neglecting correlation when propagating uncertainty and estimating the population distribution of risk. Risk Analysis, 12(4): 467-474.

Sobol, I. M., 1993. Sensitivity estimates for nonlinear mathematical models. Mathematical Modeling and Computational Experiments, 1(4):407-414.

Sprouse, W.L., and A.B. Anderson, 1995. Land Condition Trend Analysis (LCTA) Program Data Dictionary: Version 1.0. USACERL ADP Report EN-95-03, April 1995, ADA295608.

Steele, B.M., J.C. Winne, and R.L. Redmond, 1998. Estimation and mapping of misclassification probabilities for thematic land cover maps. Remote sensing of environment, 66:192-202.

Summers, J.K., H.T. Wilson, and J. Kou, 1993. A method for quantifying the prediction uncertainties associated with water quality models. Ecological Modelling, 65: 161-176.

SWCS., 1995. User Guide: Revised Universal Soil Loss Equation version 1.04. Soil and Water Conservation Society. p.145.

Tazik, D.J., S.D. Warren, V.E. Diersing, R.B. Shaw, R.J. Brozka, C.F. Bagley and W.R. Whitworth, 1992. U.S. Army Land Condition Trend Analyis (LCTA) plot inventory field methods. USACERL, Tech. Rep. N-92/03. Dept. of the Army, Construction Engineering Research Laboratories, Champaign IL.

Technical Note [TN] 420-74-3, Army Land Inventory and Monitoring Procedures on Military Installations (U.S. Army engineering and Housing Support Center[USAEHSC], Fort Belvoir, VA, 1990).

Tomppo, E., 1996. Multi-source national forest inventory of Finland. (Eds. Paivinen, R., J. Vanclay, & S. Miina) Caring for the Forest: New Thrusts in Forest Inventory, Proceedings of the Subject Group S4.02-00 "Forest Resource Inventory and Monitoring" and Subject Group S4.12-00 "Remote Sensing Technology" Volume I, IUFRO XX World Congress 6-12 August 1995, Tampere, Finland. EFI Proceedings No. 7:27-42.

Veregin, H., 1992. Error modeling for the map overlay operation. Accuracy of spatial databases (Eds. Goodchild, M., & S. Gopal), Taylor & Francis Ltd. p.3-18.

Vieux, B.E., 1995. DEM aggregation and smoothing effects on surface runoff modeling. In: Lyon, J.G., McCarthy, J. (Eds.), Wetland and environmental applications of GIS, Lewis Publishers, CRC Press, Inc., Boca Raton, Florida, p. 205-229.

Wallerman, J., 2000. Co-Kriging of forest stem volume using Landsat TM data and detected edges. Proceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, July 12-14, 2000, at Amsterdam, the Netherlands. p.709-716.

Wang, G., 1996. An expert system for forest resource inventory and monitoring in the frame of multi-source data. 173 pages. University of Helsinki, Department of Forest Resource Management, PUBLICATIONS 10 (Ph.D. dissertation), ISBN 951-45-7289-0.

Wang, G., G. Gertner, S. Fang, & A.B. Anderson. 2001b. Mapping Multiple Variables for Predicting Soil Loss by Joint Sequential Co-simulation with TM images and slope map. Photogrammetric Engineering & Remote Sensing (in review).

Wang, G., G., Gertner, V., Singh, S., Shinkareva, P. Parysow and A.B., Anderson. 2001g. Spatial and temporal prediction and uncertainty analysis of rainfall and runoff erosivity for revised universal soil loss equation. US Army Corps of Engineers, Engineer Research and Development Center, Construction Engineering Research Laboratories, ERDC/CERL TR-01-39.

Wang, G., G.Z. Gertner, P. Parysow, & A.B. Anderson. 2000b. Spatial prediction and uncertainty analysis of topographic factors for the Revised Universal Soil Loss Equation (RUSLE). J. of Soil and Water Conservation. Third Quarter, p.373-382.

Wang, G., G.Z. Gertner, P. Parysow, & A.B. Anderson. 2001d. Spatial prediction and uncertainty assessment of topographic factor for RUSLE using DEM. ISPRS J. of Photogrammetry and Remote Sensing. 56(1):65-80.

Wang, G., G.Z. Gertner, S. Wente, and A.B. Anderson. 2001h. Vegetation classification and accuracy assessment using image-aided sequential indicator co-simulation. Conference proceedings (CD) of American Society of Photogrammetry and Remote Sensing (ASPRS)

2001 - Gateway to the New Millennium, April 23-27, America's Center St. Louis, Missouri, USA.

Wang, G., G.Z. Gertner, V., Singh, and P., Parysow. 2000c. Temporal and spatial prediction and uncertainty of rainfall-runoff erosivity for revised universal soil loss equation. Modeling Complex Systems Conference, July 31 – August 2, 2000, in Montreal, Canada (abstract).

Wang, G., G.Z. Gertner, V., Singh, S., Shinkareva, P., Parysow, and A.B., Anderson. 2001f. Spatial and temporal prediction and uncertainty of soil loss using revised universal soil loss equation: A case study in rainfall and runoff erosivity for soil loss. Ecological Modeling (in press).

Wang, G., G.Z. Gertner, X. Liu, & A.B. Anderson. 2001c. Uncertainty assessment of soil erodibility factor for revised universal soil loss equation. CATENA. 46:1-14.

Wang, G., G.Z. Gertner, X. Xiao, S. Wente, and A.B. Anderson. 2001e. Appropriate plot size and spatial resolution for mapping multiple vegetation cover types. Photogrammetric Engineering & Remote Sensing. 67(5):575-584.

Wang, G., S. Fang, G.Z. Gertner & A.B. Anderson. 2000a. Uncertainty propagation and partitioning in spatial prediction of topographical factor for RUSLE. Proceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, July 12-14, 2000, at Amsterdam, the Netherlands. p.717-722.

Wang, G., S. Fang, S. Shinkareva, G.Z. Gertner, & A.B. Anderson. 2001a. Uncertainty propagation and error budgets in spatial prediction of topographical factor for Revised Universal Soil Loss Equation (RUSLE). Transactions of American Society of Agricultural Engineer (in press).

Wang, G., S. Poso, & M. Waite, 1997. SMI user's Guide for Forest Inventory and Monitoring. University of Helsinki, Department of Forest Resource Management, PUBLICATIONS 16. ISBN 951-45-7841-4. 336 p. Note: SMI is an abbreviation of Satelliittikuvat Metsien Inventorinnissa in Finnish - satellite based monitoring and inventory in English.

Wang, G., S. Poso, M. Waite, & M. Holopainen, 1998. Use of Digitized Aerial Photographs and Local Operation for Classification of Stand Development Classes. Silva Fennica 32(3):215-225.

Wang, G., S. Wente, G. Gertner, and A.B. Anderson. 2001i. Improvement in mapping vegetation cover factor for universal soil loss equation by geo-statistical methods with Landsat TM images. International Journal of Remote Sensing (in press).

Warren, S.D., and C.F. Bagley, 1992. SPOT imagery and GIS in support of military land management. GEOCARTO international, 7:35-43.

Warren, S.D., M.O. Johnson, W.D. Goran, and V.E. Diersing. 1990, An automated, objective procedure for selecting representative field sample sites. Photogrammetric Engineering and Remote Sensing, 56:333-335.

Wheeler, P.H., 1990. An Innovative County Soil Erosion Control Ordinance. Journal of Soil and Water Conservation, 45:374-378.

Wischmeier, W.H., 1959. A rainfall erosion index for a universal soil loss equation. Soil Science Society of America Proceedings, 23(3):246-249.

Wischmeier, W.H., & D.D. Smith, 1978. Predicting rainfall erosion loss: a guide to conservation planning. U.S. Department Agric. Handb. No. 537.

Wischmeier, W.H., & J.R. Mannering, 1969. Relation of soil properties to its erodibility. Soil Sci.Soc.Am.Proc. 33:131-137.

Wischmeier, W.H., and D.D. Smith, 1958. Rainfall energy and its relationship to soil loss. Trans. American Geophysical Union, 39: 285-291.

Woodcock, C.E., & A.H. Strahler, 1987. The factor of scale in remote sensing. Remote sensing of environment, 21(3):311-322.

Wu, J., & Y. Qi. 2000. Dealing with scale in landscape analysis: An overview. Geographical Information Sciences. 6(1):1-5.

Wu, J., 1999. Hierarchy and scaling: extrapolating information along a scaling ladder. Canadian Journal of Remote Sensing, 25(4):367-380.

Wu, J., D.E. Jelinski, M. Luck, and P.T. Tueller, 2000. Multiscale analysis of landscape heterogeneity: scale variance and pattern metrics. Geographical Information Sciences, 6(1):6-19.

Xia, Z.G., & K.C. Clarke, 1997. Approaches to scaling of geo-spatial data. In D.A. Quattrochi and M.F. Goodchild (Eds.), Scale in remote sensing and GIS. p. 309-360.

Xiao, X., G. Gertner, and G. Wang. 2001. Optimal Sampling Scheme for Mapping Vegetation Cover (in research group review).

Xu, W., T.T. Tran, R.M. Srivastava, and A.G. Journel, 1992. Integrating seismic data in reservoir modeling: the colocated cokriging alternative. The 67th Annual Technical Conference and Exhibition of the Society of Petroleum Engineers: 833-842. Washington, DC, October 4-7, 1992.

Zeide, B., 1980. Plot size optimization. Forest Science, 26(2):251-257.

APPENDIX

Appendix 1 (Draft of letter that will be sent with Level 1 Software to Integrated Training Area Management and Configuration Management Working Groups for integration into ATTACC)

CEERD-CN-N (70-1s)

MEMORANDUM FOR Commander, Army Environmental Center,
ATTN: Mr. George Teachman, SFIM-AEC-EQN,
Aberdeen Proving Ground, MD 21010-5401

SUBJECT: Submission of Uncertainty Analysis products to the Integrated Training Area Management (ITAM) Configuration Management Working Group (CMWG) for integration into the Army Training and Testing Area Carrying Capacity (ATTACC) Methodology.

1. Reference prior meetings between Mr. Larry Chenkin (ATSC), Mr. Gordon Weith (ATSC), Mr. George Teachman (USAEC), Mr. Tom Macia (ODCSOPS), and Mr. Alan Anderson (ERDC-CERL) concerning integration of Army research and development products into the ATTACC methodology.

2. Request ITAM CMWG evaluate Uncertainty Analysis products for integration into the Army Training and Testing Area Carrying Capacity (ATTACC) methodology.

3. Uncertainty Analysis products are submitted to the ITAM CMWG under guidance from Mr. Tom Macia (ODCSOPS) and the Conservation Technology Team (CNTT) and in accordance with the ITAM technology transfer process documented in the "ITAM Technology Configuration Management Process Standard Operating Procedure" dated November 2000. Mr. George Teachman (USAEC) is the Point of Contact to initiate the technology transfer process.

4. Uncertainty Analysis products were developed to address Army Conservation User Requirement #3 "Land Capability and Characterization", Exit Criteria FY00 #1 "Develop a protocol, tool(s) and/or factors for installation level use that reflects a probable range of results in the ATTACC methodology". Full documentation of the Army Conservation User Requirements can be found online (http://denix.cecer.army.mil/denix/DOD/Policy/Army/Aerta/tnstop.html).

5. Uncertainty Analysis products were developed by the Dr. George Gertner and his research team at the University of Illinois, Urbana Illinois. Uncertainty Analysis product research was funded the Strategic Environmental Research and Development Program (SERDP). SERDP is the Department of Defense's (DoD) corporate environmental research and development (R&D) program. Dr. Robert Holst is SERDP Program Manager for Conservation.

6. Uncertainty Analysis products were developed within current ITAM ATTACC development guidelines. Uncertainty Analysis products and algorithms were incorporated into software developed for ESRI ArcView GIS software using Avenue scripts and ESRI Spatial Analyst. Uncertainty Analysis products were developed with the version of the ATTACC methodology available at the time of the study.

7. The ERDC-CERL point of contact for this action is Mr. Alan Anderson 217/352-6511 ext 6390, alan.b.anderson@cecer.army.mil. Correspondence may be sent to: CEERD-CN-N/Alan Anderson, Engineering R&D Center, P.O. Box 9005, Champaign IL 61826-9005. The University of Illinois point of contact for this action is Dr. George Gertner 217/333-9346, gertner@uiuc.edu. Correspondence may be sent to: George Gertner, W503 Turner Hall, Department of Natural Resources and Environmental Sciences, University of Illinois, Urbana, Illinois 61801.

8. Uncertainty Analysis products are provided in this package. A CD with the ATTACC Uncertainty Software and a publication highlighting some of the uncertainty analyses that can be conducted with the software are enclosed. For more details, please refer to the following website: http://uncertainty.nres.uiuc.edu. In the website, papers related to the project are listed. Dr. George Gertner will provide reprints upon request. The enclosed publication is

Fang, S., S. Wente, G.Z. Gertner, G. Wang, and A.B. Anderson. 2001. Uncertainty analysis of predicted disturbance from off-road vehicular traffic in complex landscapes. Environmental Management (In review).

3 Encls

Mr. Alan B. Anderson
Principal Investigator
Ecological Processes Branch

CF:
Tom Macia (ITAM EMC Chair)
Larry Chenkins (ITAM CMWG)
Bob Decker (ITAM CMWG)
William Severinghaus (Co-Chair CNTT