US Army Corps of Engineers
Construction Engineering Research Laboratories

USACERL Technical Report 98/83
May 1998

Modeling Training Site Vegetation Coverage Probability with a Random Optimization Procedure: An Artificial Neural Network Approach

by
Biing T. Guan, George Z. Gertner, and Alan B. Anderson

The objective of this project was to examine the feasibility of applying feed-forward neural networks to estimate training site vegetation coverage probability based on past disturbance pattern and vegetation coverage history. The rationale behind this project was the excellent approximation and generalization ability of feed-forward neural networks. Data used to train the networks were collected from Fort Sill, Oklahoma, using the U.S. Army's Land Condition Trend Analysis (LCTA) standard data collection methodology. Two types of vegetation covers were modeled in this project: ground cover and canopy cover. For both types of vegetation cover, the input vector of a transect point consisted of several variables; namely, the past disturbance, past vegetative cover, plant community type, and vegetation life form. The output from the model was the estimated conditional probability of a transect point having vegetation cover. Results from this project suggest that artificial neural networks are a suitable tool for predicting training site vegetation coverage probability.

Approved for public release; distribution is unlimited.

Foreword

This study was conducted for the Office of the Directorate of Environmental Programs, Assistant Chief of Staff (Installation Management) [ACS(IM)], under Project 4A162720A896 “Environmental Quality Technology”; Work Unit EN-TL6, “Integrated Natural and Cultural Resources Data Analysis.” The technical monitor was Dr. Vic Diersing, DAIM-ED-N.

The work was performed by the Natural Resources Assessment and Management Division (LL-N) of the Land Management Laboratory, U.S. Army Construction Engineering Research Laboratories (USACERL). The USACERL Principal Investigator was Alan B. Anderson. Dr. William D. Severinghaus is Operations Chief, CECER-LL. The USACERL technical editor was Gloria J. Wienke, Technical Information Team.

Biing T. Guan and George Z. Gertner are members of the Department of Forestry, University of Illinois at Urbana-Champaign and performed work on this report under contract with USACERL. The authors would like to acknowledge the National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign for providing computer time on the Connection Machine CM-2 computer.

COL James A. Walter is the Commander of USACERL, and Dr. Michael J. O’Connor is Director.

Contents

Foreword **2**

1 Introduction **5**

Background 5

Objectives 6

Approach 6

Mode of Technology Transfer 7

2 Data **8**

3 Training Algorithm and Network Structure **10**

4 Performance Comparison **13**

5 Conclusions and Recommendations **14**

References **16**

This publication was reproduced on recycled paper.

1 Introduction

Background

The United States Army is responsible for managing over 12 million acres of land. The Army's land management objective is to maintain realistic military training and testing environments while promoting land stewardship. To accomplish this objective, the U.S. Army Land Condition Trend Analysis (LCTA) program was developed at the U.S. Army Construction Engineering Research Laboratories (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 (Diersing, Shaw, and Tazik 1992) and is the Army's standard for land inventory and monitoring (Technical Note [TN] 420-74-3). Over 50 military installations and training areas in the United States and Germany have begun or plan to implement the LCTA program. LCTA data sets currently exist for more than 40 installations and contain from 1 to 10 years of monitoring data. Lands inventoried as part of the LCTA program include Army Materiel Command (AMC), Forces Command (FORSCOM), Training and Doctrine Command (TRADOC), and National Guard Bureau installations. More than 75% of the Army's land base is represented by LCTA data (Shaw and Kowalski 1996).

A central objective of the LCTA program is assessing how various site characteristics, both biotic and abiotic, respond to varying levels of disturbance. This project addresses part of this objective by estimating vegetation cover probabilities based on past disturbance history and site characteristics. Estimating vegetation cover is of primary interest since cover affects soil erosion and is the principal erosion factor that can be influenced by land managers. In fact, many of the carrying capacity models developed for the Army are erosion-based models (Diersing et al 1988, Shaw and Diersing 1989, U.S. Army Concepts Analysis Agency 1996, Anderson et al 1996, Warren and Bagley 1992). For land managers, cover and erosion levels are important in the decisionmaking process for scheduling military training on an installation or deciding when to begin reclamation procedures at a given site. As an example, if after prolonged distur-

This publication was reproduced on recycled paper.

bance a site has a higher probability of having adequate vegetation cover, then one may conclude that such a site can be used with relatively higher intensity. Also, if a site has a higher likelihood of recovering quickly by itself when compared to other sites, its recovery may not require reclamation efforts as intensive as other sites.

Objectives

Currently, there are two constraints to using LCTA data for estimating vegetation cover probabilities. First, only short time series data are frequently available for analysis. Second, except in some special cases, most installations do not have data that distinguish the exact cause of ground disturbance or distinguish the extent of each type of disturbance. Given these two constraints, it is not clear which procedures are most appropriate for estimating vegetation cover probabilities. The successful application of artificial neural networks (ANN) in pattern recognition and function approximation has prompted this test to determine whether feed-forward ANN can provide accurate probability estimates under these circumstances.

The specific objective of this project is to test and verify that feed-forward ANN are a valid approach when using LCTA data to predict vegetation cover probabilities. The question is: Can next years' vegetative coverage probability be adequately estimated by inputting historic vegetative cover and disturbance information along with next years' expected disturbance? If the proposed approach is indeed appropriate, careful modeling design will allow managers to predict the probability of future training site vegetation coverage based on past coverage.

Approach

A literature survey was conducted to identify artificial neural network analysis techniques applicable for processing LCTA data. Information from the survey was then used to process a selected installation's LCTA data. Logistic models were also developed to compare the performance of the neural network models with more traditional analysis techniques. Finally, results from the LCTA data processing were summarized and recommendations made.

Mode of Technology Transfer

Information from this study is intended to be incorporated into evolving Army land-based carrying capacity models such as the Army Training and Testing Area Carrying Capacity (ATTACC) model.

This publication was reproduced on recycled paper.

2 Data

Data used in this project was obtained from Fort Sill, Oklahoma, and was collected during the peak of the vegetation growing seasons in 1989, 1990, 1991, and 1992. Standard LCTA core plot data collection methodology was used (Tazik et al 1992). Core plots were allocated across an installation using a stratified random sampling design based on unique combinations of satellite imagery landcover (reflectance) categories and soil series (Warren et al 1990). Each unique landcover/soil combination is recognized as a separate category, with the number of plots assigned to each category proportional to the land area included in each. For example, a landcover/soil category covering 10 percent of the installation would receive approximately 10 percent of the plots. This procedure was intended to ensure that the data collected are representative of the installation as a whole.

Once a plot is located in the field, a 100-meter (m) long line transect is set on each plot. Along this transect, 100 points are sampled at 1-m intervals starting at the 0.5-m point. At each sample point, information is collected regarding the presence and type of surface disturbance, ground cover, and canopy cover. A point is considered disturbed if there is physical evidence of disruption of the soil surface or if the vegetation has been obviously crushed. Although LCTA recognizes five types of surface disturbance (NONE, PASS, TRAIL, ROAD and OTHER), only two are considered in this project: no disturbance (NONE) and disturbance due to a random vehicle pass (PASS). If vegetation is present at a sample point, the species is recorded and the point is considered to have a ground cover. If other soil-maintaining material (i.e., rock) is present, the soil-maintaining material is recorded and the point is also considered to have cover. If a sample point has any aerial vegetation cover above the point, the plant species and height are recorded and the point is considered to have canopy cover. Aerial vegetation cover is recorded at 0.1-m intervals from 0.1-m to 2.0 m, and at 0.5-m intervals from 2.0-m to 8.5 m. The top-most aerial recordings at each point were summed together to classify the transect's plant community. The plant community is a hierarchical classification scheme based on the transect's vegetation physiognomic structure, and categorizes a transect by overall life form (grass, forb, shrub, or tree), life form type (annual or perennial species), and general aerial cover density (sparse, open, dense, closed) (Anderson et al 1995).

It was assumed in this project that sampled points along a line transect are spatially independent; that is, whether a previous sampled point has vegetation cover has no effect in deciding whether the next point has cover. A total of 15,158 data points were included in the data set. Combinations of the data include (1) ground cover points with NONE or PASS disturbance, (2) canopy cover points with NONE or PASS disturbance, (3) no ground cover points with NONE or PASS disturbance, and (4) no canopy cover points with NONE or PASS disturbance.

Preliminary data analysis identified the seven most relevant variables in determining vegetation cover probability for the year 1991. These seven variables include: disturbance history (NONE or PASS) in 1989 and 1990, vegetation cover (covered or not covered) in 1989 and 1990, disturbance in year 1991, the transect plot's plant community classification, and the vegetation cover's life form. Thus, each training pattern consisted of seven input variables and one target output variable. In this project, ground and canopy vegetation cover were modeled separately, with the training based on transect point data.

This publication was reproduced on recycled paper.

3 Training Algorithm and Network Structure

Due to the imbalanced representation of the training data, widely used gradient-based methods (i.e., back-propagation and its variants) failed to produce useful results. All the gradient-based training methods actually settled down at a local minimum that corresponds to classifying all noncovered points as covered points; that is, all the noncovered were filtered out as data noise. An adaptive and directional random optimization method (ADRO) was developed for this project as an alternative training method. The algorithm can be regarded as a hybrid between gradient-based and random search optimization methods. It has a self-adjusting variance term, a directional component, and can conduct backward searches. In this project, fixed structure, single hidden layer feed-forward networks with one or two hidden units were employed.

The ADRO algorithm adopted was first developed as a random optimization procedure (Matyas 1965). The algorithm was then modified by adding a backward search process and an adaptive variance component (Solis and Wets 1981). The algorithm was then introduced as a training algorithm for finding the global error minimum for feed-forward neural networks (Baba 1989). This algorithm not only has the capability to locate the global error minimum, but it is also fast.

Like simulated annealing, the ADRO procedure also uses a variance term to determine the size of weight changes (delta-weights). However, the ADRO procedure differs from simulated annealing in one significant way; the ADRO algorithm conducts searches not only in a forward manner (i.e., adding delta-weights to the current weight vector), it can also conduct backward searches (i.e., subtracting delta-weights from the current weight vector) if the forward search is unable to lower the training error. If both forward and backward searches are unable to improve the training error, the weight vector remains unchanged and a new set of delta-weights will be generated.

The second unique feature of the ADRO algorithm is its variance term. The variance of this algorithm is controlled by the results of the search process, and depending on the progress, it can go either up or down. In this implementation, the original rule for variance adjustment was adopted (Solis and Wets 1981). If

each of the five successive forward search steps (or three backward search steps) is able to lower the training error, the variance for the next step will be doubled from its current value. If, however, each of the five consecutive forward steps (or three backward search steps) fails to lower the training error, the current variance will be halved. The variance remains unchanged otherwise. This practice encourages larger steps to be taken when the search is going well, and forces smaller steps to be taken if the search is not going well. An added advantage of this self-adjusting process is that the variance term will approach zero rapidly if there is no improvement in training. This feature allows unnecessary training to be avoided, especially at the beginning of a training session.

The third unique feature of the ADRO algorithm is its directional component. This component is similar to a memory or momentum term. It allows the algorithm to search along the directions where it has been successful, and accelerates the search process. In this application, the rules for adjusting the directional component reported in Baba (1989) were adopted. Since the ADRO algorithm is still a random optimization procedure, it can escape local minima and locate the global minima. The backward search, the self-adjusting variance, and the directional component together make the ADRO a fast random search procedure.

In the original algorithm, the delta-weights are generated by either a Gaussian or a uniform distribution. For speed enhancement in this implementation, the new steps are generated by a Cauchy distribution similar to the fast simulated annealing (Szu and Hartley 1987). Since the variance of a Cauchy distribution is unbounded, occasionally large steps in the right directions will be taken, which will improve the training speed. Using some benchmark test data sets in an artificial neural network (e.g., multi-bits parity problem), we concluded that the modified algorithm is at least as fast as the original algorithm.

Because of the inherent parallelism of the algorithm and because the problems of interest typically involve a large number of observations, a parallel version of the algorithm was implemented on a Connection Machine CM-2 computer using CM FORTRAN under field-wise (or PARIS) mode. The implementation is a training set parallel implementation where each training point occupies a processor, and the weights are broadcast to each processor when needed (Singer 1990). Under this implementation, the majority of the computationally intensive tasks will be done on CM-2. Because inter-processor communication is kept to a minimum, this implementation is probably the fastest one for this algorithm on a

This publication was reproduced on recycled paper.

Connection Machine.

All the networks used in this project were standard 3-layer feed forward networks. The input layer consisted of seven units; the output layer had only one unit, and the outputs from this unit were regarded as the conditional vegetation coverage probabilities for the current year estimated by the trained networks. For each training set, two networks were trained, one with one hidden unit and the other with two hidden units. The purpose was to investigate whether a more complex network would provide better probability estimation. Thus, there were 10 weights (including the weights for the bias unit) in each of the one-hidden-unit networks (referred to as ADRO-1 networks hereafter) and 19 weights for the two-hidden-unit networks (ADRO-2 hereafter). The error function for both networks was a squared-error function, and the activation function for all process units was a logistic function. The best initial variances were determined through trial and error.

The training procedure had three stopping rules:

1. The maximum training cycle was set at 10,000. Training stopped if this limit was exceeded.

2. Training stopped if the error improvement for 10 consecutive cycles was smaller than 1.0E-3.

3. Training stopped and the network was considered to have accomplished the approximation if the training produced an error smaller than the pre-defined error criterion.

4 Performance Comparison

Logistic models were developed to compare to the performances of the ANN models. The logistics model used the same input and output variables as the neural network models. The logistic regression procedure of the SAS statistical software (PROC LOGIST) was used to obtain the parameter estimates of the logistic models.

The main goodness-of-fit statistic used in this project is the χ^2 statistic (Snedecor and Cochran 1980). In order to use this statistic, validation data (i.e., data for predicting the coverage probability for 1992) were cross-classified according to the seven input variables. Then, for each combination, an average coverage ratio was estimated and treated as the long-term coverage probability for that combination. For χ^2 to be effective, each combination has to have at least five observations in it. Therefore, in this project, combinations with less than five observations were removed from the validation data sets. Thus, the validating data set for ground coverage has 86 combinations, and the validation data set for aerial coverage has 74 combinations.

The χ^2 goodness-of-fit statistic is defined as:

$$\chi^2 = \sum_{i=1}^{n} \frac{(E_i - O_i)^2}{E_i}$$

where O_i and E_i are the observed and estimated, respectively, number of covered transect points in combination i; and $E_i = O_i \times P_i$ where P_i is the estimated probability of vegetation coverage from the ANN or the logistic models for combination i.

This publication was reproduced on recycled paper.

5 Conclusions and Recommendations

For both types of vegetation covers, training stopped for ANN models after roughly 5,000 iterations once good initial variances were located. In all instances, network training stopped due to a lack of significant improvement in lowering the training error. It should be noted that for ANN models, the best initial variances were in the order of 10^{-5}, which corresponded to a set of weights with small values. This set of small initial weights produced outputs around 0.5 for each input pattern. In a certain respect, one can regard that initially the ANN models treated every pattern as having a coverage probability of 0.5 (i.e., a form of noninformative prior). Through repeated training, the ANN models gradually adjusted the weights to reflect the probability of coverage until the systems settled to some solutions. The residue errors as well as the validation results for the ANN models are given in Table 1.

Both logistic models were able to converge. Most of the asymptotic 95% confidence intervals for the parameter estimates for logistic models did not contain 0, which indicates that the logistic models were statistically valid. The residue errors as well as the validation results for the logistic model are given in Table 1.

As shown in Table 1, ANN models had a better fit to the training data than the corresponding logistic models for both types of vegetation covers. The improvement of ANN models over the corresponding logistic model was substantial. The ANN models were particularly effective in predicting the canopy cover. This result is likely due to the fact that most of the training data had canopy cover in 1991. As expected, ADRO-2 models fit the training data better than the ADRO-1 models, though the improvement was less significant than anticipated.

Table 1. Performances of various models for predicting vegetation coverage probability.

	Ground Vegetation Coverage			Canopy Vegetation Coverage		
	Logistic	**ADRO-1**	**ADRO-2**	**Logistic**	**ADRO-1**	**ADRO-2**
Residue error	1756.7	1474.5	1471.6	266.4	221.1	281.0
χ^2	739.7	57.3	54.9	89.5	6.3	2.5

The main reason that the ADRO algorithm worked with these data sets, while back-propagation failed, is attributed to the way network weights were adjusted. For the ADRO algorithm, the overall training error is lowered through random weight adjustments regardless of the errors of individual patterns. Thus, if a set of weight changes can lower the overall training error, the weight changes are adopted; otherwise the algorithm retains the current weights and generates another set of weight changes. In contrast, gradient-based methods reduce the overall training error through the minimization of each individual pattern error. For training sets with roughly balanced representation, this strategy works well. For imbalanced training sets, the main contribution of each weight change will be due to the corrections of the dominant pattern. The weight adjustments will continue until the dominant pattern is correctly classified for all cases. Consequently, the rare patterns will be filtered out as data noises. This result will occur regardless of what the initial weights are.

The main purpose of this project was to test whether a random optimization procedure can be used to model vegetation cover probabilities on military installations given manmade impacts and natural variability in vegetation cover. In general, ANN models had a better fit to the training data than the corresponding logistic models. However, since the data in this project have a short time sequence, results from the project should be interpreted cautiously. It should be noted that traditional statistical methods might be more suitable for this problem as further data accumulate. It will be of great interest to determine whether the ADRO algorithm will continue to perform better than the logistic model as long-term time sequence data becomes available. Ultimately though, data consistency through time is imperative for any type of time-dependent modeling. Even short to mid-term data is crucial given our lack of knowledge on the relationships between manmade impacts and the natural resources.

This publication was reproduced on recycled paper.

References

Anderson, A., L. Chenkins, L. Winters, R. Hunt, C. Couvillon, D. McFerren, S. Sekscienski, T. Shirnia, and P. Sydelko, “Army Training and Testing Area Carrying Capacity (ATTACC) and Evaluation of Land Value Study (ELVS),” pp 8-17, in: *Integrated Training Area Management Workshop*, 5th Annual ITAM/LRAM Workshop, LaCrosse, WI, August, 1996.

Anderson, A.B., W.L. Sprouse, D.G. Kowalski, and P.J. Guertin. 1995. *LCTA Users Interface, Users Manual Version 1.0*, ADP Report 95/24/ADA300797 (USACERL August 1995) 156 pp.

Baba, N., “A new approach for finding the global minimum of error function of neural networks,” *Neural Network*, vol 2 (1989) pp 367-373.

Diersing, V.E., R.B. Shaw, and D.J. Tazik, “U.S. Army Land Condition-Trend Analysis (LCTA) Program,” *Environmental Management*, vol 16 (1992) pp 405-414.

Diersing, V.E., R.B. Shaw, S.D. Warren, and E.W. Novak, “Users Guide for Estimating Allowable Use of Tracked Vehicles on Non-wooded Military Training Lands,” *Journal of Soil and Water Conservation*, vol 43 (1988) pp 191-195.

Matyas, J., “Random optimization,” *Automation and Remote Control*, vol. 26 (1965) pp 246-253.

Shaw, R.B., and V.E. Diersing, “Allowable Use Estimates for Tracked Vehicular Training on Pinon Canyon Maneuver Site, Colorado, US,” *Environmental Management*, vol 13 (1989) pp 773-782.

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).

Singer, A., *Implementations of artificial neural networks on the Connection Machine*, Thinking Machine Corporation, Technical Report RL90-2 (Thinking Machine Corporation, Cambridge, Massachusetts, 1990).

Snedecor, G.W., and W.G. Cochran, *Statistical Methods* (Iowa State University Press, Ames, IA, 1980).

Solis, F.J. and J.B. Wets, “Minimization by random search techniques,” *Mathematics of Operations Research*, vol. 6 (1981) pp 19-30.

Szu, H., and R. Hartley, “Fast simulated annealing,” *Physics Letter (Series A)*, vol 122 (1987) pp 157-162.

Tazik, D.J., S.D. Warren, V.E. Diersing, R.B. Shaw, R.J. Brozka, C.F. Bagley, and W.R. Whitworth, *U.S. Army Land Condition-Trend Analysis (LCTA) Plot Inventory Field Methods*, TR N-92/03/ADA247931 (USACERL, February 1992).

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).

U.S. Army Concepts Analysis Agency, *Evaluation of Land Value Study (ELVS),* Study Report CAA-SR-96-5 (U.S. Army Concepts Analysis Agency, 1996).

Warren, S.D., and C.F. Bagley, “SPOT Imagery and GIS in Support of Military Land Management,” *Geocarto International*, vol 1 (1992) pp 35-43.

Warren, S.D., M.O. Johnson, W.D. Goran, and V.E. Diersing, "An Automated, Objective Procedure for Selecting Representative Field Sample Sites," *Photogrammetric Engineering and Remote Sensing,* vol 56, no. 3 (1990) pp 333-335.

This publication was reproduced on recycled paper.