ELSEVIER

Ecological Modelling 157 (2002) 209–225

ECOLOGICAL MODELLING

www.elsevier.com/locate/ecolmodel

Comparing five modelling techniques for predicting forest characteristics

Gretchen G. Moisen *, Tracey S. Frescino

USDA Forest Service, Rocky Mountain Research Station, 507 25th Street, Ogden, UT 84401, USA

Abstract

Broad-scale maps of forest characteristics are needed throughout the United States for a wide variety of forest land management applications. Inexpensive maps can be produced by modelling forest class and structure variables collected in nationwide forest inventories as functions of satellite-based information. But little work has been directed at comparing modelling techniques to determine which tools are best suited to mapping tasks given multiple objectives and logistical constraints. Consequently, five modelling techniques were compared for mapping forest characteristics in the Interior Western United States. The modelling techniques included linear models (LMs), generalized additive models (GAMs), classification and regression trees (CARTs), multivariate adaptive regression splines (MARS), and artificial neural networks (ANNs). Models were built for two discrete and four continuous forest response variables using a variety of satellite-based predictor variables within each of five ecologically different regions. All techniques proved themselves workable in an automated environment. When their potential mapping ability was explored through simulations, tremendous advantages were seen in use of MARS and ANN for prediction over LMs, GAMs, and CART. However, much smaller differences were seen when using real data. In some instances, a simple linear approach worked virtually as well as the more complex models, while small gains were seen using more complex models in other instances. In real data runs, MARS and GAMS performed (marginally) best for prediction of forest characteristics.
© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Predictive mapping; Forest inventory; Classification tree; Regression tree; Mulivariate adaptive regression spline; MARS; Artificial neural network

1. Introduction

Forest inventory data, like those collected by the Forest Inventory and Analysis (FIA) program in the United States, have historically been used to produce estimates of forest population totals over large geographic areas. Recent emphasis has been placed on expanding the traditional uses of this data by merging it with satellite-based information to produce regional maps of forest characteristics for use in a variety of forest land management applications. These applications include broad-scale activities like mapping wildlife habitat, assessing resource loss to fire, identifying lands suitable for timber harvest, and locating areas at high risk for insect and disease outbreaks.

* Corresponding author. Tel.: +1-801-625-5384; fax: +1-801-625-5723

E-mail addresses: gmoisen@fs.fed.us (G.G. Moisen), tfrescino@fs.fed.us (T.S. Frescino).

PII: S0304-3800(02)00197-7

There are numerous sources of ancillary data, and a tremendous amount of effort has been directed at acquiring finer resolution data from a wide variety of newly developed air- and space-borne platforms. There are also numerous ways in which forest class and structure variables from forest inventories may be modeled as functions of remotely sensed and other ancillary variables. Yet, little work has been directed at comparing modern statistical techniques to determine which tools are best suited to mapping tasks given multiple objectives and logistical constraints.

In this paper, five modelling techniques were compared for mapping forest characteristics in the Interior Western United States using forest inventory field data and ancillary satellite-based information. The research involved five statistical modelling techniques for predicting two discrete and four continuous forest inventory variables. The modelling techniques included: generalized additive models (GAM), classification and regression trees (CART), multivariate adaptive regression splines (MARS), and artificial neural networks (ANN). In addition, a simple linear model (LM) was used as a benchmark against which to judge the other models. The two discrete inventory variables included a forest/non-forest classification, as well as a binary classification within forested areas. The four continuous response variables were tree biomass per acre, average tree age, quadratic mean tree diameter, and percent tree crown cover. The analyses were conducted within five ecologically different regions (two each in Montana and Utah, and one in Arizona). The predictor variables included elevation, aspect, slope, geographic coordinates, unclassified spectral data from the Advanced Very High Resolution Radiometer (AVHRR) sensor, and a national vegetation cover map derived from Landsat Thematic Maper (TM) imagery. Predictive performance (map accuracy) of all discrete and continuous variables were compared across modelling techniques, ecoregions, and predictor variable sets using independent test data. All models were evaluated for suitability in a production environment.

2. Materials and methods

2.1. Data description

2.1.1. Study regions and sample design

Portions of five ecologically different regions defined by Bailey et al. (1994) were selected for analyses and are illustrated in Fig. 1. The ecoregions range from the coniferous forests of northwestern Montana, to the semi-desert conditions of the mountains of central Arizona. MT1 and MT2 refer to two ecoregions in Montana, UT1 and UT2 are two within Utah, and AZ1 is in Arizona. Dates of forest inventory, sample grid intensity, and field plot layout differ by ecoregion as well as by land owner and vegetation type, as summarized in Table 1. Standardized per-acre responses were retrieved under each layout.

2.1.2. Response variables

At each FIA field location, extensive stand- and tree-level measurements were collected. Individual tree measurements were compiled and combined with stand-level variables to produce location-level summaries. The two discrete response variables, FORTYP.2 and FORTYP.3, were created by collapsing a detailed forest type into forest/non-forest (FORTYP.2) and timberland/woodland (or spuce-fir/other in Montana) within forested areas (FORTYP.3), respectively. Data files for modelling the discrete FORTYP.2 include all data from forest and non-forest locations while data for modelling FORTYP.3 include only forested field locations. This is analogous to applying a forested ‘mask’ over a data set to focus modelling on within-forest conditions. The four continuous response variables were tree biomass per acre (BIOTOT), average tree age (STAGE), quadratic mean tree diameter (QMDALL), and percent crown cover (CRCOV).

2.1.3. Predictor variables

Predictor variables were extracted from four sources: (1) elevation, aspect, and slope from 1000-m digital elevation models; (2) spectral and positional data from a biweekly AVHRR composite; (3) vegetation cover type from the National Land Cover Data (NLCD); and (4) geographic coordi-

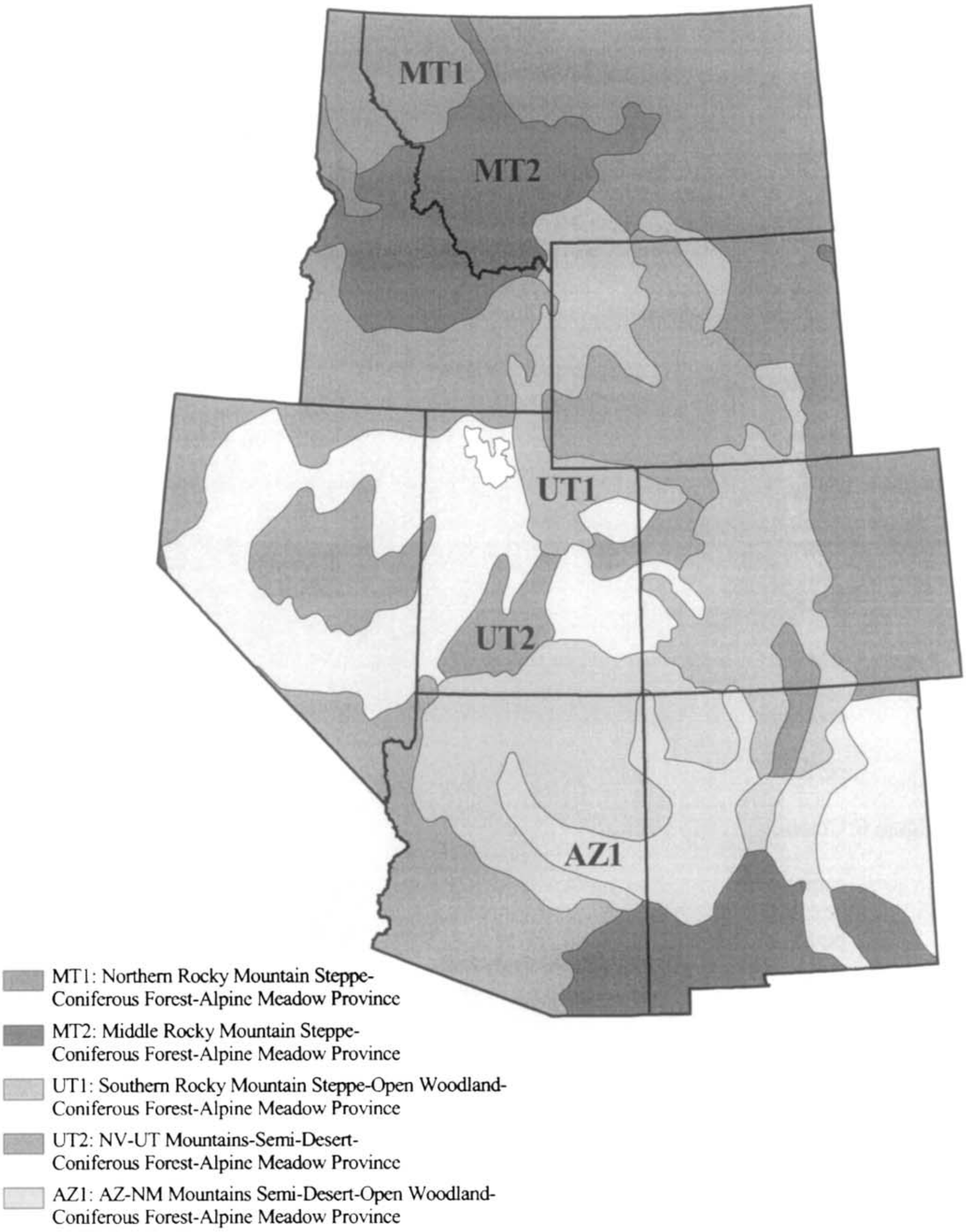
MT1
MT2
UT1
UT2
AZ1
▨ MT1: Northern Rocky Mountain Steppe-Coniferous Forest-Alpine Meadow Province
▨ MT2: Middle Rocky Mountain Steppe-Coniferous Forest-Alpine Meadow Province
▨ UT1: Southern Rocky Mountain Steppe-Open Woodland-Coniferous Forest-Alpine Meadow Province
▨ UT2: NV-UT Mountains-Semi-Desert-Coniferous Forest-Alpine Meadow Province
□ AZ1: AZ-NM Mountains Semi-Desert-Open Woodland-Coniferous Forest-Alpine Meadow Province

Fig. 1. Five ecologically different regions defined by Bailey et al. (1994) were selected for analyses. Ecoregions range from the coniferous forests of northwestern Montana, to the semi-desert conditions of the mountains of central Arizona. MT1 and MT2 refer to two ecoregions in Montana, UT1 and UT2 are two within Utah, and AZ1 is in Arizona.

nates in the Universal Transverse Mercator (UTM) projection. A list of predictor variables and their descriptions is provided in Table 2.

The circular aspect variable is transformed to a radiation index (TRASP) used by Roberts and Cooper (1989). This takes the form

Table 1
Description of six study ecoregions, sampling intensity, and number of plots

Label	Description	Size (ha)	Inventory dates	Grid intensity	# Plots forest	# Plots total
MT1	Northern Rocky Mountain Forest Steppe–Coniferous Forest–Alpine Meadow Province	4.43 M	NF: 1993–1996 Other: 1988–1989	All: 5 k	1393	1677
MT2	Middle Rocky Mountain Steppe–Coniferous Forest–Alpine Meadow Province	9.45 M	NF: 1996–1998 Other: 1988–1989	All: 5 k	1634	3727
UT1	Southern Rocky Mountain Steppe–Open Woodland–Coniferous Forest–Alpine Meadow Province	3.18 M	All: 1992–1996	NF: double 5 k Other: 5 k	531	968
UT2	NV/UT Mountains Semi-Desert–Coniferous Forest–Alpine Meadow Province	3.16 M	All: 1993–1996	NF: double 5 k Other: 5 k	829	1320
AZ1	AZ/NM Mountains Semi-Desert–Open Woodland–Coniferous Forest–Alpine Meadow Province	2.85 M	NF, res, Tmbr: 1996–1997 Other: 1983	NF, res, IR: 5 k Timber/other: double 10 k	664	1141

NF, National Forest; Other, lands outside NF; res, reserved lands; Tmbr, Timberland; IR, Indian reservations; Wdld, Woodland; F, Forested; Tot, Forested and Non-forested plots combined.

Table 2
Description of predictor variables

Type	Name	Description
Discrete	NLCD	0 = Non-forest 40 = Forest 50 = Shrubland with trees
Continuous	EASTING	UTM Easting—Zone 12
Continuous	NORTHING	UTM Northing—Zone 12
Continuous	ELEV.1K	Elevation (m) from 1 km DMA
Continuous	TRASP.1K	Radiation index derived by transforming aspect from 1 km DMA
Continuous	SLOPE.1K	Slope (%) from 1 km DMA
Continuous	AVH.1	Visible spectral band 1 from AVHRR composites
Continuous	AVH.2	Near-IR spectral band 2 from AVHRR composites
Continuous	AVH.3	IR spectral band 3 from AVHRR composites
Continuous	AVH.4	IR spectral band 4 from AVHRR composites
Continuous	AVH.5	IR spectral band 5 from AVHRR composites
Continuous	NDVI	NDVI from AVHRR composites

$$\text{TRASP} = \frac{1 - \cos((\pi/180)(\text{aspect} - 30))}{2}. \qquad (1)$$

This transformation assigns a value of zero to land oriented in a north-northeast direction, (typically the coolest and wettest orientation), and a value of one on the hotter, drier south-southwesterly slopes.

Daily observations from the AVHRR platform are compiled biweekly to produce spectral composites of the U.S. These composites result in a near cloud-free image depicting maximum vegetation greenness for the compositing period. One such composite dated (June 1986) was used in these analyses and contains six bands of ‘least cloud’ information including five spectral channels [one visible, one near infrared (NIR), and three infrared (IR)] as well as a Normalized Difference Vegetation Index (NDVI) that is computed NDVI = (NIR − IR)/(NIR + IR).

The NLCD (http://edcwww.cr.usgs.gov/programs/lccp) is a land cover data set produced through a cooperative effort involving the U.S. Environmental Protection Agency, U.S. Geological Survey, U.S. Forest Service, and National Oceanic and Atmospheric Administration. This Thematic Mapper (TM)-based national data set (released in 2000) provides 21 mapped cover-types at 30-m resolution. In this study, cover-types were collapsed to a simple forest, shrubland, and non-forest type.

2.1.4. *Test sets*

Within each ecoregion, both the total and forest-masked data files were randomly split into two data sets, 70% of the data for modelling and 30% for testing. The 30% test data set was chosen because this is the approximate proportion of plots collected on an intensified (not the standard 5 km) sampling grid and withholding this additional amount gives an indication of predictive abilities given customary sampling intensities in forest inventories.

2.2. *Modelling*

The following section describes each of the five modelling techniques along with model fitting details for this forest inventory application. DeVeaux et al. (1993), DeVeaux (1995) provide more general discussions comparing these techniques. All modelling and analyses were conducted in S-PLUS.

2.2.1. *NLCD benchmark models*

By far, the simplest mapping strategy that could be adopted in these analyses is to predict discrete variables by collapsing NLCD cover types, and predict continuous variables by assigning the mean of the continuous variable within each NLCD class. This approach is implemented by either using a function that collapses cover type classes for discrete variables, or using a simple linear model for continuous variables. This is the benchmark against which other models were judged.

2.2.2. *Generalized additive models*

Generalized additive models (Hastie and Tibshirani, 1986, 1990) have been described in detail by Guisan et al. (this issue). Illustrations of predictive modelling in forest inventory applications using generalized linear models and generalized additive models can be found in Moisen and Edwards (1999), Frescino et al. (2001), respectively.

Both the binary forest/non-forest (FORTYP.2) and timberland/woodland (or spruce-fir/other) within forest (FORTYP.3) classifications were modeled using a binomial family. The selection of an appropriate link function and variance-to-mean relationship for the continuous variables, however, is more difficult. Encountering a large number of zeros (on non-forest lands) can confound the problem and dominate the mean/variance relationship. A non-forest mask was applied as described above in the data section and only continuous variables on forested plots were modeled, assuming the mask would be reapplied at time of mapping. The variances of continuous variables on forested plots (within bins defined by combinations of predictor variables) were plotted against the mean values of those bins, revealing no detectable patterns. Consequently, a simple gaussian family was specified for continuous the responses. Alternatively, an option can be implemented within the program to run a regression of variances on means, determining if the variance is proportional to 1, μ, μ^2, or μ^3, and then assigning a gaussian, poisson, gamma, or inverse gaussian family, respectively. For both continuous and discrete responses, predictor variables entered the model individually using a smoothing spline with a relatively large smoothing parameter to avoid fitting noise. Final models were selected by stepwise procedures.

2.2.3. Classification and regression trees

Classification and regression trees, also known as *recursive partitioning regression*, dates back to Morgan and Sonquist (1963) and has received more recent attention through Breiman et al. (1984), (the use of the acronym here is not to be confused with any proprietary software or trademarks). CARTs subdivide the space spanned by the predictor variables into regions $\{R_m\}$ for which the values of the response variable are approximately equal, and then estimate the response variable by a constant, a_m, in each of these regions. That is,

$$f(\boldsymbol{x}) = a_m, \quad \text{for } \boldsymbol{x} \in R_m. \tag{2}$$

The tree is called a *classification tree* if the response variable is qualitative, and a *regression tree* if the response variable is quantitative. The initial node on a tree is called the root. From the root, the model is fit using binary recursive partitioning. This means the data are successively broken into left and right branches with the splitting rules defined by the predictor variable values. For example, a first split might occur where $x_1 < c_1$, where c_1 is estimated. Then, $\hat{f}(\boldsymbol{x}) = a_1$, for $x_1 < c_1$, and $\hat{f}(\boldsymbol{x}) = a_2$, for $x_1 \geq c_1$. A second split might occur where $x_1 < c_1$ and $x_2 < c_2$, and so on. Splits are chosen that maximize the 'value' of a split, where this value may be computed in many different ways. For classification problems splits are chosen that most reduce the impurity of the distribution at the node, while in regression problems the value of a split is measured as the reduction in the residual sum of squares. Splitting continues down to the 'terminal' nodes where response values are all the same within a node or data are too sparse for additional splitting. At the terminal node, the predicted response is given that is the average or majority of the response values in that node for continuous or discrete variables, respectively. Pruning the tree to avoid overfitting the data can be accomplished a number of different ways, as described below.

In modelling the forest inventory data, an initial tree was fit using all the predictor variables. Tree pruning, analogous to variable selection in regression, is the methodology used to prevent overfitting the training data with too many splits. Although many methods of pruning are available, pruning through crossvalidation is most popular. The optimal size for these trees was identified via 10-fold cross validation. While this process was repeatable for classification, the 'optimal size' was very different under different crossvalidation runs for continuous variables. Consequently, 20 crossvalidatory splits were run and the 'majority rule' (i.e. optimal size getting the most votes) was used to determine pruning size for continuous variables.

2.2.4. Multivariate adaptive regression splines

MARS, developed by Friedman (1991) is a flexible nonparametric regression method that generalizes the piecewise constant functions of CART to continuous functions by fitting (multivariate) splines in the regions R_m, and matching up the values at the boundaries of the R_m. One form for writing the MARS model is

$$\hat{f}(\boldsymbol{x}) = a_0 + \sum_{K_m=1} f_i(x_i) + \sum_{K_m=2} f_{ij}(x_i,\ x_j) + \sum_{K_m=3} f_{ijk}(x_i,\ x_j,\ x_k) + \cdots, \qquad (3)$$

but the notation requires further explanation. Here, the first sum is over all basis functions that involve only one variable. Each function in this first sum can be expressed as

$$f_i(\boldsymbol{x}_i) = \sum_{\substack{K_m=1 \\ i\in V(m)}} a_m B_m(\underline{\boldsymbol{x}}_i). \qquad (4)$$

where $V(m)$ is the variable set associated with the mth basis function, $B_{m,}$ that survives backward selection strategies. The second sum is over all basis functions that involve two variables, where each bivariate function can be expressed as

$$f_i(\boldsymbol{x}_i,\ \boldsymbol{x}_j) = \sum_{\substack{K_m=2 \\ i,j\in V(m)}} a_m B_m(\underline{\boldsymbol{x}}_i,\ \underline{\boldsymbol{x}}_j). \qquad (5)$$

The third sum is over all basis functions that involve three variables, and so on.

Hastie and Tibshirani's (1996) *mars* function was loaded into S-plus and used to fit the MARS models. MARS automatically selects the amount of smoothing required for each predictor as well as the interaction order of the predictors. It is considered a projection method where variable selection is not a concern but the maximum level of interaction needs to be determined. Taking a conservative approach, only two-level interactions were specified.

2.2.5. Artificial neural networks

Neural networks have received considerable attention as a means to build accurate models for prediction, control, and optimization when the functional form of the underlying equations is unknown. This modelling technique has permeated literature in many fields including statistics (Ripley, 1994, 1996; Stern, 1996; Cheng and Titterington, 1994), remote sensing (Atkinson and Tatnall, 1997; Skidmore et al., 1997; Wang and Dong, 1997), and ecology (Lek et al., 1996; Lek and Guegan, 1999).

Although there are a variety of ways to construct these models, 'backpropagation networks' appear to be the most frequently used in practice. In this description we have used statistical terms with corresponding neural network terminology in parentheses. A backpropagation network with one hidden layer is a nonlinear statistical model of the form

$$f_l(\boldsymbol{x}) = \sigma\left(\sum_{k=1}^{Kl} w_{2kl}\sigma\left(\sum_{j=1}^{J} w_{1jk}\boldsymbol{x}_j + \theta_k\right) + \theta_l\right). \qquad (6)$$

The response (*output*) is a transformation of a weighted combination of the predictor (*input*) variables. The σ in the above equation is a bounded, monotonic, and differentiable function, with a logistic function the most common choice. That is,

$$\sigma(x) = \frac{\exp(x)}{1 + \exp(x)}. \qquad (7)$$

The numerous coefficients w (weights) and intercepts θ (bias terms) are estimated (undergo *training*, *learning*) through an optimization method similar to steepest descent (*backpropagation*). Because so many parameters can be estimated, there is danger in overfitting the model. By sacrificing an unlimited number of degrees of freedom, a modeller can eventually get a perfect fit. In that case one would be modelling noise as well as the underlying phenomenon, and prediction for unvisited sites could be severely compromised. The preferred method to avoid overfitting involves using a large enough network to avoid underfitting, then limiting the number of iterations of the fitting procedure through crossvalidation.

Nychka's FUNFITS S-Plus function library was obtained by ftp for fitting ANN's from http://www.cgd.ucar.edu/stats/Funfits/index.shtml. Although the computing time can be quite slow for full search options, the subjective choices about starting values, convergence criteria, and number of hidden units are done automatically.

2.3. Evaluation criteria

Because the utility of maps for different management applications cannot be captured in a single map accuracy number, several global measures were used to assess the predictive performance of the models. Let $\boldsymbol{x}$ be an $r \times r$ contingency table or error matrix set out in rows and columns that express the number of sample plots (of which there are n) predicted to belong to one of r classes relative to the true ground class (on the diagonal). The percent of correctly classified (PCC) plots is calculated

$$\mathrm{PCC} = \left(\frac{1}{n}\sum_{i=1}^{r} x_{ii}\right) \times 100\%. \tag{8}$$

Note that PCC can be deceptively high when frequencies of zeros and ones in binary data are very different. For example, if a model predicts only zeros for a data set with 10% ones and 90% zeros, the PCC is 90%. The Kappa statistic (Cohen, 1960) measures the proportion of correctly classified units after the probability of chance agreement has been removed. It has been used extensively in map accuracy work (Congalton, 1991), and is calculated

$$\mathrm{Kappa} = (\theta_1 - \theta_2)/(1 - \theta_2), \tag{9}$$

where

$$\theta_1 = \sum_{i=1}^{r} x_{ii}/n$$

and

$$\theta_2 = \sum_{i=1}^{r} x_{i\cdot}x_{\cdot i}/n^2.$$

Predictive performance of models of the continuous variables were evaluated through independent estimates from test sets of global root mean squared error (RMSE),

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}, \tag{10}$$

and proportion of plots within some user-specified range (PWI),

$$\mathrm{PWI} = \frac{1}{n}\sum_{i=1}^{n} I\{|\hat{y}_i - y_i| < R\}, \tag{11}$$

(e.g. proportion of plots predicted to within 50 cubic feet of the true volume). The correlation coefficient (ρ) between observed and predicted values

$$\rho = \frac{\sum y_i\hat{y}_i - \sum y_i \sum \hat{y}_i/n}{\sqrt{\sum y_i^2 - \left(\sum y_i\right)^2/n}\sqrt{\sum \hat{y}_i^2 - \left(\sum \hat{y}_i\right)^2/n}} \tag{12}$$

was also calculated for each model.

In addition to the evaluation criteria above, the amount of time it took to run each model was recorded and considered in discussions about suitability of each of the models for a production environment.

2.4. Mapping

Predictions were produced for each response variable within each and imported into ArcView for display and analysis. The scale of the resulting maps is a function of the intensity at which predictor variables (as ArcInfo grids) are resampled. Here, a coarse 1 km grid was used for mapping to keep size and prediction times in check. Finer resolution maps may also be produced.

When mapping over large geographic areas, one is guaranteed to run into values of predictor variables outside the range seen in the modelling data set and extrapolation is unavoidable. In addition, high dimensional models with interaction confound the extrapolation problem and it is likely that nonlinear and nonparametric models, such as those used here, may produce unrealistic estimates. To prevent these few extreme values from completely overpowering evaluation criteria and map color schemes, model predictions were restricted from going below zero or above the maximum value seen in the model data set.

3. Results

3.1. Test simulations

Before running data from all the ecoregions through the modelling system, a simple test was conducted to illustrate the known advantages and disadvantages of the modelling techniques. Following DeVeaux et al. (1993), 1000 each of ten uniformly distributed predictor variables $X1$–$X10$ were generated. Next, a response Y was specified as a function of only $X1$–$X5$,

$$Y = 2\sin(\text{ð}{*}X1{*}X2) + 0.4(X3 - 0.5)^2 + 0.2(X4) + 0.1(X5), \qquad (13)$$

with no error term. A simple linear model along with a GAM, CART, MARS and ANN were used to fit the relationship between Y and the $X1$–$X10$. Results are shown in Table 3. These illustrate the effectiveness of MARS and ANNs in deciphering complex relationships. CART models identified the contributing predictor variable ($X1$–$X5$), but had an RMSE that was 10% higher than a linear model, and 10 times the RMSE of ANNs. LM also had a high RMSE because of its inability to detect the non-linearity or interaction between terms. GAM residuals were considerably better, but the model's stepwise procedures incorrectly identified $X8$ and $X10$ as contributing predictor variables in addition to the correct ones. Both MARS and ANN did exceptionally well, and MARS correctly identified the contributing variables and order of interaction. Again, ANNs and MARS performed best overall but MARS had a much faster computing time.

Next, simulations were run to illustrate the effect of random noise on the performance of each modelling technique. Following from the example above, the response was generated with increasing error. As expected, differences between performance measures diminished rapidly with increasing noise in the system.

3.2. Discrete variables

The PCC and Kappa values obtained using independent test sets for each modelling technique, response variable and ecoregion are illustrated in Figs. 2 and 3, respectively. These graphics allow for quick visualization of a very large number of total model fits. Each individual dotplot shows the value of the evaluation criteria by modelling technique (y axis) and response variable (columns) within ecoregion (rows). Modelling techniques were ordered from best to worst (descending down Y axes in each plot) according to the mean value of each performance measure across all variables and ecoregions.

The PCC and Kappa results suggest little difference between modelling techniques for identification of forest/non-forest but illustrate substantial gains over the NLCD approach when

Table 3
Simulation results where first, ten uniformly distributed predictor variables $X1$–$X10$ were generated (1000 each)

Model	Selected variables[a]	RMSE	PWI (25%)	RHO	Time
CART	$X4$, $X1$, $X3$, $X5$, $X2$	0.030	76	0.843	202
LM	All	0.027	83	0.873	1
GAM	s($X1$), s($X2$), s($X3$), $X4$, $X5$, s($X9$), s($X10$)	0.014	95	0.966	201
MARS	$X1{*}X2$, $X3$, $X4$, $X5$	0.004	100	0.997	43
ANN	All	0.001	100	1.000	336

Next, a response Y was specified as a function of only $X1$–$X5$, $Y = 2\sin(\text{ð}{*}X1{*}X2) + 0.4(X3-0.5)^2 + 0.2(X4) + 0.1(X5)$, with no error term. A simple linear model along with a GAM, CART, MARS and ANN were used to fit the relationship between Y and the $X1$–$X10$

[a] Here, s indicates the variable came into the model in a nonlinear fashion using a nonparametric smoother, and * indicates interaction between variable.

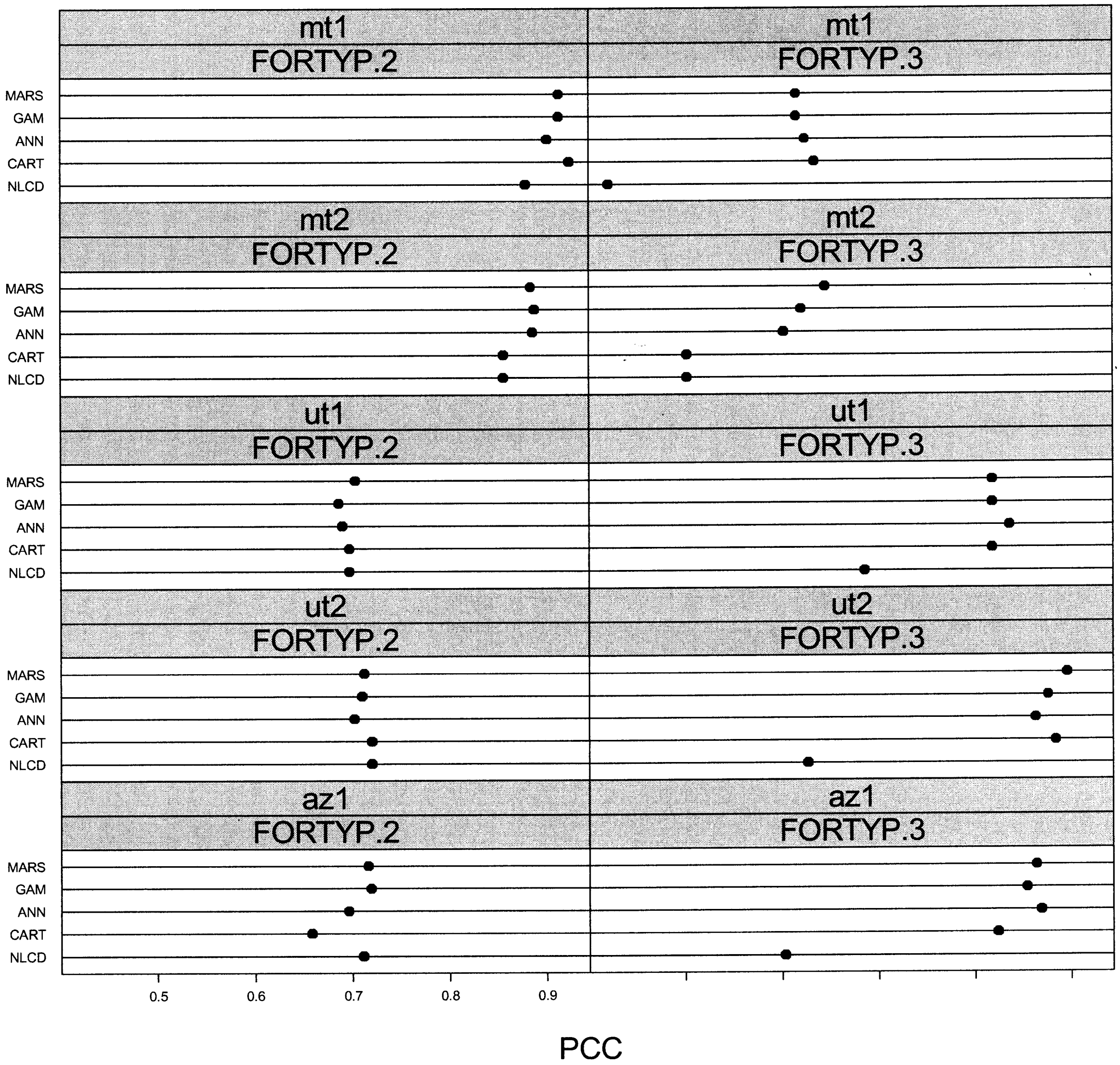

Fig. 2. PCC by modelling technique (y axis) and response variable (columns) within ecoregion (rows), ordered from best to worst according to the mean value of each performance measure across all variables and ecoregions.

separated into three classes (FORTYP.3). These gains are made regardless of the nonlinear or nonparametric model and reflect the inability of the NLCD vegetation type maps to identify woodland areas in Utah and Arizona, or to identify spruce/fir forest in MT2. The similarity in Kappa values in MT1 reflect the fact that the majority of forests in the ecoregion are, in fact spruce/fir, and one is highly likely to get a correct classification simply by chance. The top two modeling techniques (based on mean values for individual performance measures) were MARS and GAMs for both PCC and Kappa (by a very slim margin). The NLCD and MARS models were fastest, computationally.

An example of a 1 km resolution map of predicted forest/non-forest in UT2 is given in Fig. 3. Files of UTM coordinates and predicted

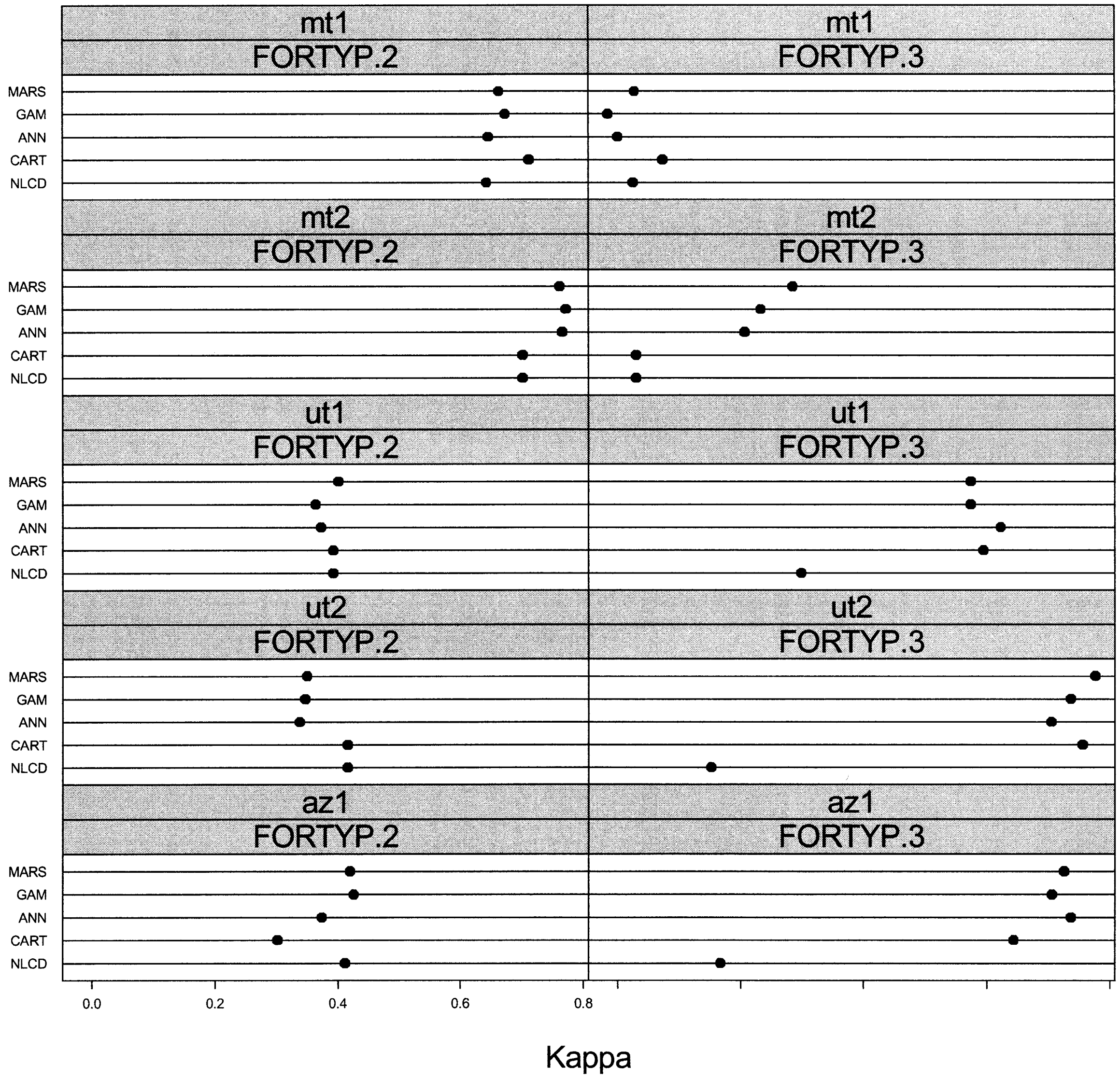

Fig. 3. Kappa by modelling technique (*y* axis) and response variable (columns) within ecoregion (rows), ordered from best to worst according to the mean value of each performance measure across all variables and ecoregions.

values can be brought into pre-made ArcView layouts, easing the chore of generating map displays (Fig. 4).

3.3. Continuous variables

The RMSE, RHO, and PWI obtained using independent test sets for each modeling technique, response variable and ecoregion are illustrated in Figs. 5–7, respectively. Results suggest that all five models often perform competitively for RMSE and PWI, but occasional erratic behavior by ANN, MARS, and CART can be anticipated. As with the discrete variables, GAMs and MARS performed marginally best based on mean values of the performance measures. Only small gains are realized through alternative modelling techniques. The lowest values of RHO (~ 0.1) were seen in

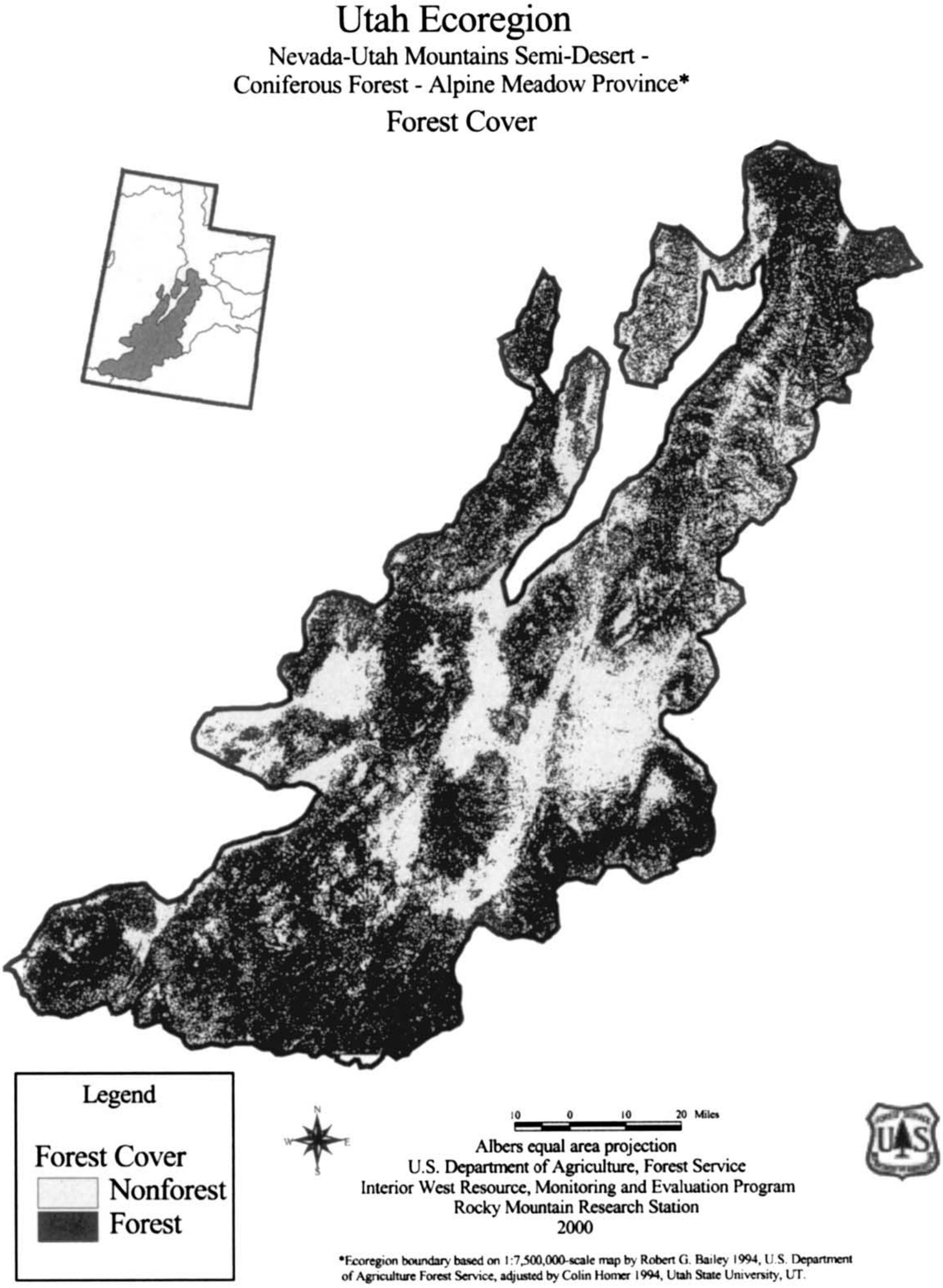
Utah Ecoregion
Nevada-Utah Mountains Semi-Desert -
Coniferous Forest - Alpine Meadow Province*
Forest Cover
Legend
Forest Cover
Nonforest
Forest
N
W
E
S
10
0
10
20 Miles
Albers equal area projection
U.S. Department of Agriculture, Forest Service
Interior West Resource, Monitoring and Evaluation Program
Rocky Mountain Research Station
2000
*Ecoregion boundary based on 1:7,500,000-scale map by Robert G. Bailey 1994, U.S. Department of Agriculture Forest Service, adjusted by Colin Homer 1994, Utah State University, UT.

Fig. 4 Example of a 1-km resolution map of predicted forest/non-forest in UT2 using a MARS model.

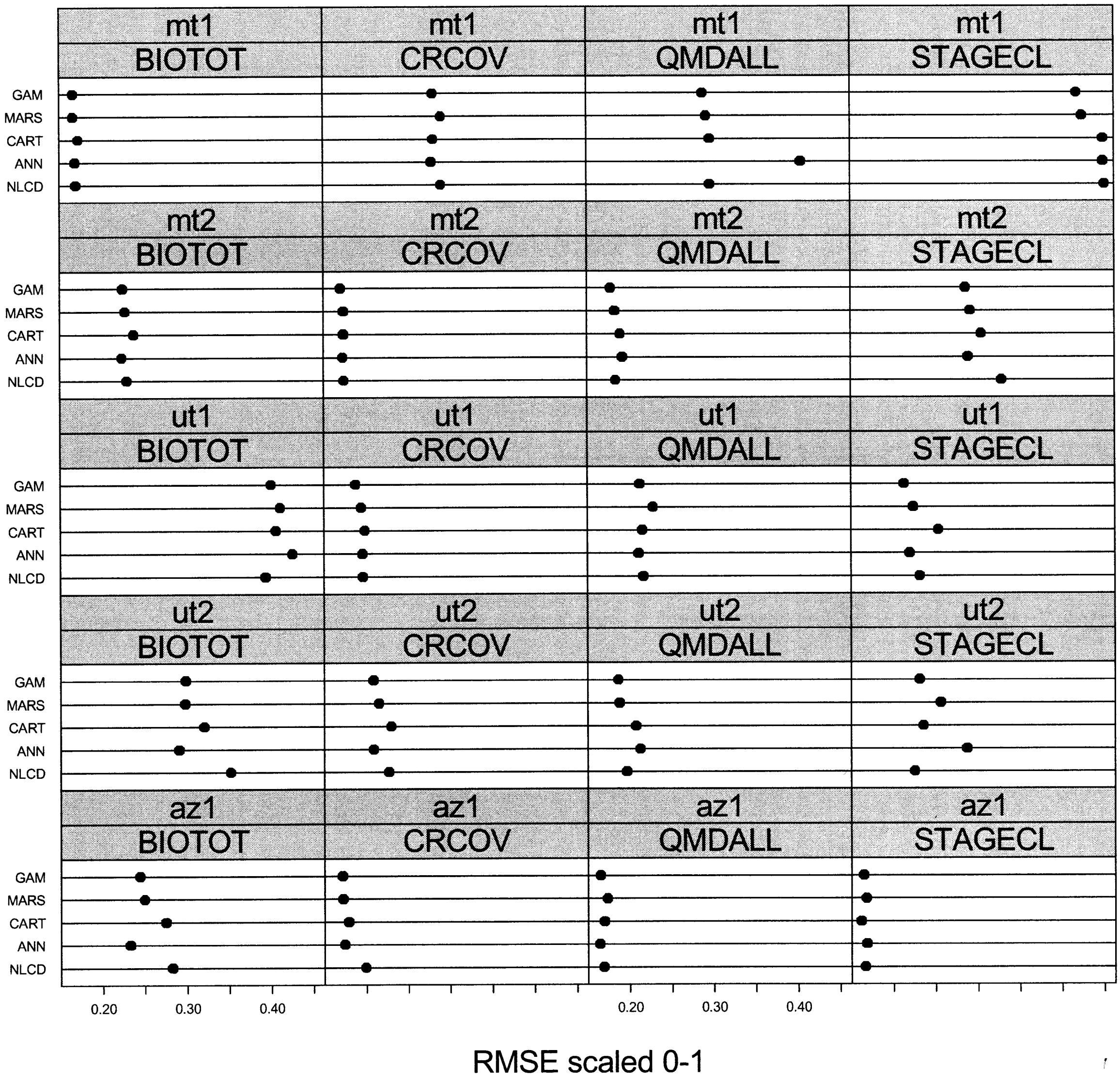

Fig. 5. RMSE by modelling technique (y axis) and response variable (columns) within ecoregion (rows), ordered from best to worst according to the mean value of each performance measure across all variables and ecoregions.

UT1 in the BIOTOT models while higher values (~ 0.6) were seen in MT2's STAGECL, UT2's BIOTOT, and AZ1's BIOTOT and CRCOV. It is important to note the overall low PWI values in Fig. 7, illustrating how difficult it is to accurately predict continuous response variables. As expected, simple NLCD and MARS models ran much faster than the others.

4. Discussion

All techniques tried here proved themselves workable in an automated environment, although ANNs were a bit more problematic. Computation run time is one area the modelling techniques differed substantially. Naturally, the simple NLCD model was extremely fast and straightfor-

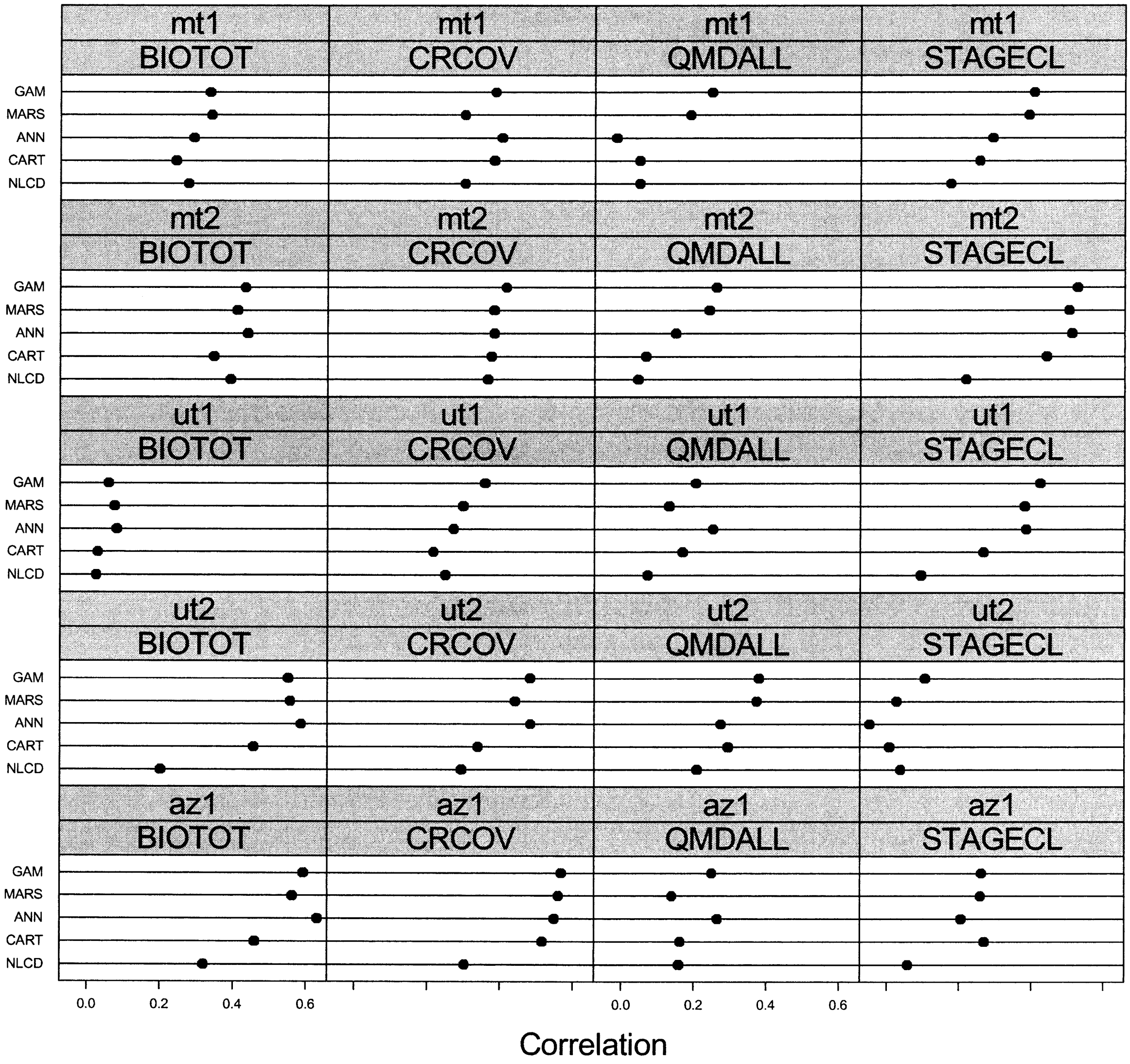

Fig. 6. Correlation by modelling technique (y axis) and response variable (columns) within ecoregion (rows), ordered from best to worst according to the mean value of each performance measure across all variables and ecoregions.

ward. GAMs and CARTs are normally quite fast but were considerably slower here because of the stepwise procedures for GAM and iterative runs searching for best tree size for CART. ANNs were the slowest in these applications and have the potential to be cripplingly slow for 'slow but safe' parameter optimization procedures in FUNFITS. Obviously, the simplest NLCD approach or another simple linear model is most readily incorporated into a production process. But of the more flexible techniques, MARS showed promise in a production environment because of its fast computing rate, little need for user 'steering', and tendency to produce reasonable models when

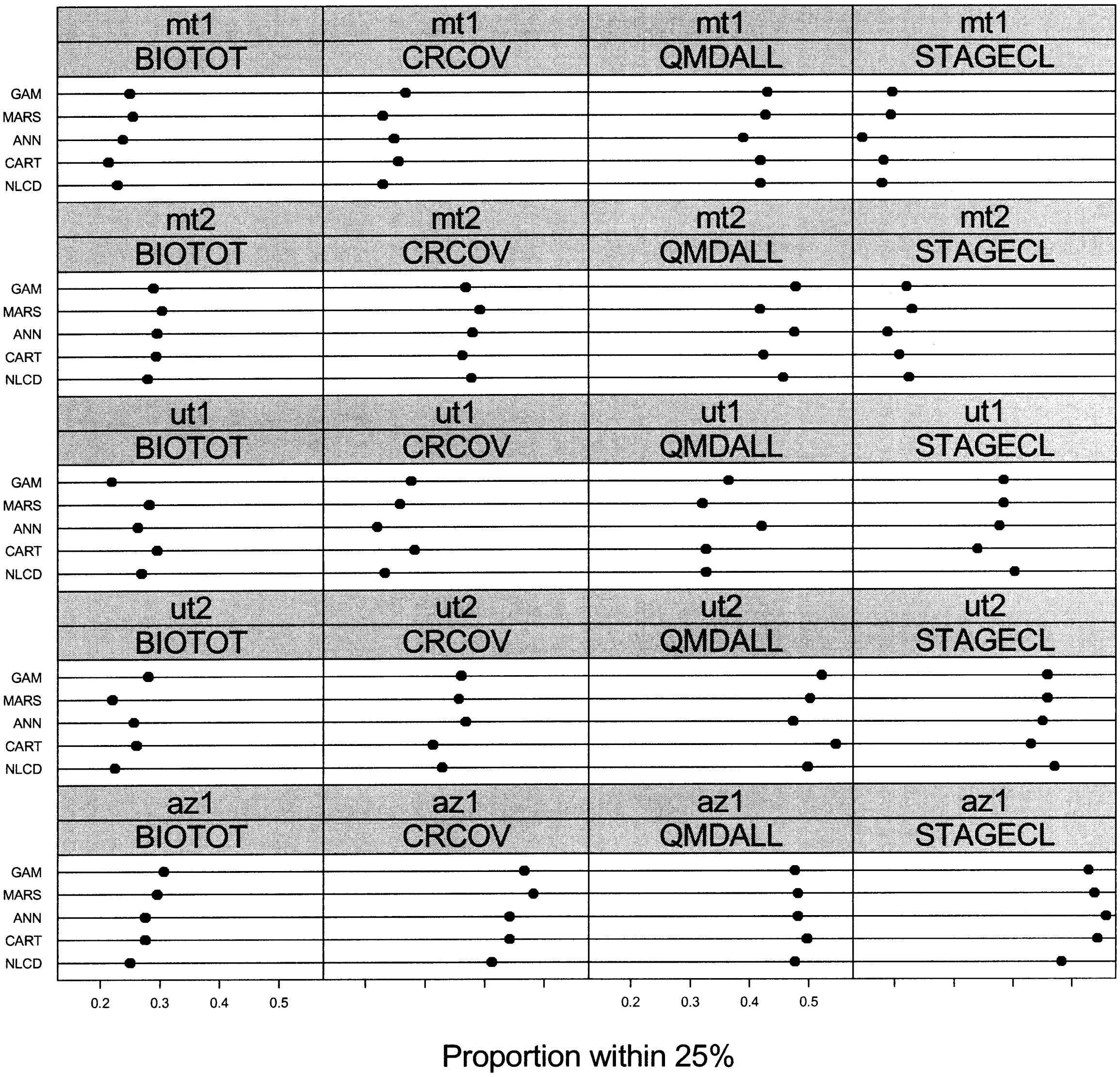

Fig. 7. Proportion of predictions with 25% of the truth by modelling technique (*y* axis) and response variable (columns) within ecoregion (rows), ordered from best to worst according to the mean value of each performance measure across all variables and ecoregions.

optimal parameters for an ANN were not found. Certainly, any of the models could be made production suitable, and a sensible strategy may be to keep all the tools in the toolkit, using several for each application.

When thinking about accuracy of maps produced through an automated modelling system, it is important to note that high scores for global performance measures do not necessarily mean that the maps will be better for management applications on the ground, and there is no substitute for application-specific testing. However, valuable lessons were still learned using global performance measures obtained both through simulations and diverse data sets.

This simple simulation described in the beginning of Section 3 illustrated that use of a flexible and powerful modelling technique can make a huge difference in predictive performance when one has a high signal to noise ratio (recall that our data did not). The test also shed some light on the character of each technique. It was surprising that CART performed worse than a simple linear model. It was also surprising that GAMs stepwise procedure was not able to exclude all the non-contributing variables. In addition, the ease with which both MARS and CART established the relationship of Y to the predictor variables was very informative.

The differences between modelling techniques using real data were far less distinct. In fact, for a number of variable/ecoregion combinations, only small differences were realized using any of the modelling techniques over a simple NLCD approach, particularly for distinguishing forest/non-forest, or in RMSE for continuous variables. Larger gains were realized, however, for further classification of forested areas (FORTYP.3) and in getting predictions that fell within a user-specified range. In addition, slightly higher correlations were realized for MARS and GAMs. This was seen in residual plots where more realistic predictions were obtained for extreme lows (in both MARS and GAMs) and extreme highs (for MARS).

When starting this analysis with the real data, we had anticipated seeing marked differences between modelling techniques. The small gains seen with these data sets were at first surprising, but understandable given the tremendous amount of noise in the data. Sources of noise are numerous and include: positional error in field plots, registration difficulties between plots and images, scale differences between data collected in the field and the imagery, differences in date, and definitional differences. Based on the results one might be inclined to stick with a simple linear model for mapping. Yet, the data are in a constant state of change. GPS coordinates with national standards are now being collected on all field plots, better resolution imagery with standardized registration procedures are becoming available, softcopy low altitude photography is under development, and better resolution topographic information is also available. Given all that, the true benefit of a new predictor variable might be overlooked if only linear models were in place. So, building more flexible modelling techniques like GAMs or MARS into a predictive mapping system up front is likely to yield large predictive gains in the future, even if differences between that and a much simpler approach are only small right now.

5. Conclusions

In comparing the different modelling techniques, all proved themselves workable in an automated environment, though the simple NLCD and MARS required the least amount of user guidance. When explored through a simple simulation, tremendous advantages were seen in use of MARS and ANN for prediction, but much smaller differences were seen when using real data because of noise or possible lack of nonlinear relationships between the response and predictor variables. While the simple NLCD model was the fastest and easiest of all to apply, MARS and GAMS performed marginally better than the others for prediction of forest characteristics. Although, little appreciable difference was seen between the models, as better predictor variables become available in the future, better predictions may be realized using more flexible statistical techniques.

Acknowledgements

We would like to thank many people at the Rocky Mountain FIA unit including Dwane Van Hooser and Renee O'Brien for their support and enthusiasm, and Ron Tymcio and Shirley Waters for their technical expertise. We are grateful to Richard Cutler for his numerous reviews of the work and useful suggestions, Tom Edwards for launching the FIA mapping efforts and keeping them rolling, and, of course, the ever present T.B. Murphy.

References

Atkinson, P.M., Tatnall, A.R.L., 1997. Neural networks in remote sensing. Int. Remote Sens. 18, 699–709.

Bailey, R.G., Avers, P.E., King, T., McNab, W.H. (eds.), 1994. Ecoregions and Subregions of the United States (map). Washington DC, U.S. Geological Survey. Scale 1:7,500,000, colored, accompanied by a supplementary table of map unit descriptions, prepared for the U.S. Department of Agriculture, Forest Service.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth and Brooks/Cole, Monterey, CA, 358 p.

Cheng, B., Titterington, D.M., 1994. Neural networks: a review from a statistical perspective. Stat. Sci. 9, 2–54.

Cohen, J., 1960. A coefficient of agreement of nominal scales. Educ. Psychol. Meas. 20, 37–46.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens. Environ. 37, 35–46.

DeVeaux, R.D., 1995. A guided tour of modern regression methods, Proceedings of the Section on Physical and Engineering Sciences, 1995 Fall Technical Conference, St. Louis, MO.

DeVeaux, R.D., Psichogios, D.C., Ungar, L.H., 1993. A comparison of two nonparametric estimation schemes: MARS and Neural Networks. Comput. Chem. Eng. 8, 819–837.

Frescino, T.S., Edwards, T.C., Jr., Moisen, G.G., 2001. Modelling spatially explicit forest structural attributes using generalized additive models. J. Veg. Sci. 12, 15–26.

Friedman, J.H., 1991. Multivariate adaptive regression splines. Ann. Stat. 19, 1–141.

Guisan, A., Edwards, T.C., Jr., Hastie, T. [This issue.] Generalized linear and generalized additive models in studies of species distributions: setting the scene.

Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models. Chapman and Hall, New York, 335 p.

Hastie, T., Tibshirani, R.J., 1986. Generalized additive models. Stat. Sci. 1, 297–318.

Hastie, T., Tibshirani, R.J., 1996. S Archive: mda, StatLib (http://lib.stat.cmu.edu/S/).

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., Aulagnier, S., 1996. Application of neural networks to modelling nonlinear relationships in ecology. Ecol. Model. 90, 39–52.

Lek, S., Guegan, J.F., 1999. Artificial neural networks as a tool in ecological modeling: and introduction. Ecol. Model. 120, 65–73.

Moisen, G.G., Edwards, T.C., Jr., 1999. Use of generalized linear models and digital data in a forest inventory of northern Utah. J. Agric. Biol. Environ. S 4, 372–390.

Morgan, J.N., Sonquist, J.A., 1963. Problems in the analysis of survey data and a proposal. J. Am. Stat. Assoc. 58, 415–434.

Ripley, B.D., 1994. Neural networks and related methods for classification. J. Roy. Stat. Soc. B 56, 409–456.

Ripley, B.D., 1996. Pattern Recognition and Neural Networks. Cambridge University Press, New York, p. 403.

Roberts, D.W., Cooper, S.V., 1989. Concepts and techniques of vegetation mapping. In: Land Classifications Based on Vegetation: Applications for Resource Management. USDA Forest Service General Technical Report INT-257, Ogden, UT, pp. 90–96.

Skidmore, A.K., Turner, B.J., Brinkhof, W., Knowls, W., 1997. Performance of a neural network: mapping forests using GIS and remotely sensed data. Photogramm. Eng. Rem. S. 63, 501–514.

Stern, H.S., 1996. Neural networks in applied statistics. Technometrics 38, 205–220.

Wang, Y., Dong, D., 1997. Retrieving forest stand parameters from SAR backscatter data using a neural network trained by a canopy backscatter model. Int. J. Remote Sens. 18, 981–990.