A comparison of neural network and process-based models for vegetation distribution under global climate change

Eric Chown
chown@cs.orst.edu

Thomas G. Dietterich
tgd@cs.orst.edu

Department of Computer Science
Oregon State University
Corvallis, OR 97331-3202

November 14, 1997

Abstract

Ecological models based on physical, chemical, and biological processes are difficult to develop and calibrate. Neural network methods provide an attractive alternative where the model is constructed automatically. We compared the MAPSS global vegetation model to a collection of neural network models trained on a data set covering current climate and vegetation distribution in the coterminous United States. Completely automated methods were employed to calibrate MAPSS and to train the neural networks. The models were then compared on two tasks: prediction of current vegetation distribution given current climate and prediction of future vegetation distribution given future $2{\times}CO_2$ climates produced by the OSU, GFDL, and UKMO global climate models. For predicting under current climate, the neural network models are more accurate than MAPSS. Unfortunately, the predictive accuracy of the models under future climate models cannot be measured without knowing future vegetation distribution. However, an upper bound on the accuracy can be computed using a statistic called the *classification scatter*. On future climate scenarios, the neural network models exhibit much higher classification scatter than MAPSS, which means that the highest accuracy that they could achieve is substantially less than the highest potential accuracy of MAPSS. This suggests—but does not prove—that process-based models such as MAPSS will make better predictions than neural network models for future climate scenarios.

1 Introduction

Policy-makers require accurate ecological models to help them make many difficult decisions. Consequently, the development of methodologies for constructing accurate models is a central problem in applied ecological research. Two methodologies for constructing such models are process models and empirical (e.g., neural network) models.

Process models attempt to model the physical, chemical, and biological processes underlying ecological phenemena. Examples of such models include the global circulation models (Michell & Warrilow, 1987; Wetherald & Manabe, 1988; Schlesinger & Zhao, 1989) and global vegetation models (Neilson, 1995). The construction of these models is difficult and time-consuming. The modeler must have expertise in many scientific fields. He or she must also make design tradeoffs between the fidelity of the model, the computational cost of computing model predictions, and the data that is available for calibrating and validating the model. In addition, the implementation and debugging of an efficient, process model is a major programming task. Most process-based models include an empirical component in the form of parameter values that must be calibrated from available data. This calibration is often performed manually, by making numerous model runs and adjusting the parameters to maximize the accuracy of the model on the calibration data. In some cases, algorithmic search procedures can automate the calibration process (Gupta & Sorooshian, 1985; Duan, Sorooshian, & Gupta, 1992; Klepper & Hendrix, 1994).

Neural network models, in contrast, do not make any attempt to model the physical, chemical, or biological causes of ecological phenomena. Instead, they learn a direct mapping from input features to output predictions. The mapping is constructed so that it fits a (preferably large) collection of *training examples* of the desired input-output relationship. There are many popular packages for training neural networks, so no additional programming is required aside from the computer work necessary to prepare the training data. Because of the ease of constructing neural network models, they have been growing in popularity in the ecosystem modeling community (Tan & Smeins, 1996; Hsu, Gupta, & Sorooshian, 1995).

Despite the huge difference in structure between process models and neural network models, the training or calibration problem is conceptually the same. Given a collection of data points, the parameters of the model must be adjusted to maximize the fit of the model to the data. Consequently, throughout this paper, we will use the work "training" to mean either training a neural network model or calibrating a process model.

In this paper, we compare the accuracy of a process-based model and a neural network model for predicting the continent-scale distribution of terrestrial ecosystems after a doubling of atmospheric carbon dioxide (CO_2). Our goal is to determine which modeling methodology provides more accurate results on this important task. We will show that if the task is to predict the distribution of terrestrial ecosystems under *current* climate, then the neural network methods are slightly more accurate than the process-based model. However, if the goal is to extrapolate beyond current climate to predict the consequences of global warming, then the process-based model is likely to be much more accurate.

This article begins with a description of the modeling task and the two modeling methodologies. We then report the results of two experiments. The first experiment involves predicting the distribution of plant ecosystems under current climate. The second experiment requires predicting the distribution of plant ecosystems under three alternative double-CO_2 climate predictions. The paper concludes with a discussion of the results.

2 Predicting Continental-Scale Vegetation Distribution

One of the primary impacts of the accumulation of greenhouse gasses will be on vegetation. Understanding this impact is critical for shaping environmental policy and assessing the consequences of global warming. Furthermore, the potential effects are complex. For example, if atmospheric carbon dioxide increases, plants can absorb this CO_2 more easily and with less water loss. This makes these plants more efficient water users, which enables the growth of more plants for a fixed amount of water. This could act as a negative feedback to the increase in CO_2. Conversely, if increased CO_2 causes an increase in temperature, this will increase the amount of water plants require, which may cause plants to die, decay, and liberate additional CO_2. This could act as a source of positive feedback.

To understand and predict the impact of global warming on the continent-wide distribution of plant ecosystems, several process-based ecosystem models are being developed. For a review and summary of these models, see the report of the VEMAP group (VEMAP-Members, 1995).

The specific task to be solved is the following. Given variables describing the climate at a given site (in terms of mean monthly precipitation, humidity, and high and low temperatures), the model should predict the class of vegetation that will grow at that site. There are 64 input variables (4 variables for each of 12 months, and other information such as elevation, and soil composition), and a single output variable, the vegetation class. We employ 74 possible vegetation classes such as "tall grass prarie", "conifer forest", "desert shrubland", and so on.

In this article, we compare two alternative approaches to constructing models for this task. We describe these methods in turn.

2.1 MAPSS

MAPSS (Mapped Atmosphere-Plant-Soil System; Neilson, 1995) is a process-based global vegetation model. It works by modeling water flows. Water enters a site in the form of precipitation, which may be either rain or snow. Depending on conditions such as temperature and humidity, some of this water may be added to the snowpack, evaporate, or turn into runoff. The rest of the water, along with melting snow, reaches the ground and is available to the vegetation or may be absorbed into the soil. Among other things, MAPSS models the rate at which water passes through the soil (via infiltration and percolation) and the rate at which vegetation can remove the water from the soil.

MAPSS groups all vegetation into three major types: grasses, trees, and shrubs. The key step inside MAPSS is to determine the amount of each of these three types of vegetation that is growing at a site. The amount of vegetation is captured by the *leaf area index* (LAI). MAPSS determines the LAI for grass, trees, and shrubs by modeling competition among these types of vegetation for water and sunlight. The competition is simulated through a series of iterations. The iterations begin by setting tree LAI to its maximum and then testing whether there is enough water available at the site to support all of the vegetation. If there is not, tree LAI is reduced and the process is repeated. If the tree LAI becomes low enough, then the canopy opens up, and sunlight can reach the forest floor to allow grass to grow. As a result, grass LAI becomes non-zero and the grasses compete with the trees for available soil water. This can cause tree LAI to drop even further. If it falls below a threshold, tree LAI is set to zero, shrub LAI is set to its maximum value, and the competition continues. In grasslands, shrub LAI becomes zero.

MAPSS performs this iterative search through the space of LAI values until it finds a configuration that balances the water use at the site. A site is balanced if the sum of water entering the site for a year equals the sum of water leaving the site (through transpiration and runoff). A site may temporarily store water, but it cannot accumulate water from one year to the next—MAPSS is an equilibrium model. Once LAI values have been computed, classification rules are applied to determine the vegetation class of the site. These classification rules consider features such as the LAI values, solar angle, minimum annual temperature, and leaf morphology in assigning the vegetation class.

As with most models, MAPSS contains a set of parameters that must be calibrated. The primary calibration parameters for the experiments performed in this article are five parameters that control how efficiently the various vegetation types transpire water under different conditions.

One focus of our research on MAPSS has been to develop automatic methods for calibrating the model. This has proved to be a difficult task, because MAPSS is highly nonlinear. In this paper, calibration of MAPSS was performed by applying Powell's method (Press, Teukolsky, Vetterling, & Flannery, 1992). Powell's method is a "direction set" minimization method; in this case the quantity that is to be minimized is the error between the model outputs and the desired outputs. For the experiments reported here, the error measure is the least squares residual between the ideal LAI vector and the LAI vector obtained with the current parameter settings. Specifically, let Θ be the vector of five parameters to be calibrated. For a particular site s, define the error function

$$J(s,\Theta) \quad = \quad (\widehat{LAI}_{tree} - LAI_{tree})^2 + (\widehat{LAI}_{grass} - LAI_{grass})^2 + (\widehat{LAI}_{shrub} - LAI_{shrub})^2,$$

where $\widehat{LAI}$ is the LAI value predicted by the model using parameters Θ. We call this the *component-wise* error function, because it sums the error for each LAI component separately. For a set of sites S, we compute $J(s,\Theta)$ for each site s in S and then sum the errors to obtain an overall error measure $J(S,\Theta)$.

The calibration process can be viewed as a search in a 5-dimensional space for a value of Θ that minimizes $J(S,\Theta)$. Powell's method starts at an initial point Θ_0, which is chosen at random from pre-specified ranges of legal values for the parameters. The algorithm conducts a series of "line searches". Each line search evaluates new parameter vectors Θ that lie along a particular line to find one that minimizes $J(S,\Theta)$ *along that line.* It then chooses a new direction in which to search, and conducts a line search along that new direction. The initial search directions run along the coordinate axes, which means that initially, Powell's method varies only one parameter at a time. After each parameter has been varied separately, Powell's method attempts to choose new directions most likely to reduce the error measure. A complete description can be found in (Press et al., 1992).

2.2 The Neural Network Model

We employ a standard three-layer feed-forward neural network to learn a predictive vegetation model. This kind of neural network provides a very flexible space of non-linear functions. Each function has the following form. There are 64 input values (the same inputs as MAPSS), which we will denote by $x_i, i = 1,\ldots,64$. These are connected to 100 "hidden units". Each hidden unit j computes an "activation value" h_j according to the formula:

$$h_j = \sigma\left(\sum_{i=1}^{64} w_{ij} x_i\right),$$

where w_{ij} is the "weight" between input i and hidden unit j and $\sigma(\cdot)$ is the *sigmoid function*:

$$\sigma(u) = \frac{1}{1+e^{u}}.$$

The network also contains 74 "output units", one for each of the 74 vegetation classes. Each output unit k computes an activation value o_k according to the formula

$$o_k = \sigma\left(\sum_{j=1}^{100} w_{jk} h_j\right).$$

To train the network, the weights w_{ij} and w_{jk} (which we will refer to collectively as W) are adjusted to minimize the error between the output values o_k and the desired target values t_k. For a given site whose desired vegetation class is n, the target value t_n is set to one and all of the other target values $t_k (k \neq n)$ are set to zero. This is called the 1-of-n output encoding. The weights W are adjusted to minimize the squared error:

$$J(s, W) = \sum_{k=1}^{74} (o_k - t_k)^2.$$

We employed an implementation of the conjugate-gradient algorithm (Barnard & Cole, 1989; Press et al., 1992) to minimize $J(s, W)$. Given a setting for the weights W, the vegetation class for a new site is predicted as follows. The 64 variables describing that site are given to the network and the activations of the 74 output units are computed. The vegetation class corresponding to the output unit with the highest activation is the predicted class.

Whenever a highly flexible function approximator, such as a neural network, is applied to fit data, there is a danger of overfitting—that is, of memorizing the training data and failing to generalize well to new data points. Results of several studies, including our own work (Dietterich, Hild, & Bakiri, 1995) and the very thorough work of Caruana (1997) have shown that the best predictive accuracy for this form of neural network is obtained by using a large number of hidden units combined with the technique known as *early stopping* (Lang, Hinton, & Waibel, 1990).

Early stopping works as follows. Suppose we have a total of N data points available for training the neural network. We randomly set aside a subset of N_h of these points to form the *halting set*. The remaining $N - N_h$ points are given to the conjugate-gradient minimization algorithm. After each change to the weights, the predictive accuracy of the current network is measured on the halting set. If the error is smaller than any previous error, the weights are saved in a file. Training is conducted until a local minimum in the error measure J is attained. The saved weights are the weights that gave the best error on the halting set, so these are the weights that are employed to make future predictions.

2.3 Training and Test Data Sets

To compare MAPSS and the neural network models, we used a data set developed by the VEMAP group for their inter-model comparison study (VEMAP-Members, 1995). This data set divides the United States into a grid of 3,261 sites, each one-half degree of latitude by one-half degree of longitude. At each site, the climate is described by the mean monthly precipitation, humidity, wind, and low and high temperatures (i.e., 5 quantities per month) computed over a 50-year period. Each site is also assigned to one of 74 vegetation classes such as "tall grass prarie", "conifer forest", and so on. These vegetation classes were based on the vegetation classes produced by an earlier hand-calibrated version of MAPSS that was run on a high-resolution climate data set where each cell was 10km square (Neilson, 1995). It might seem that this would give MAPSS an unfair advantage, because it appears to guarantee that the target vegetation classes can be produced by MAPSS. However, because the input climate data is based on a completely different data source that is noisier and provides the data at a much coarser spatial scale, MAPSS has no particular advantage and needs to be completely recalibrated to make useful predictions.

After obtaining this data set, we divided the 3,261 sites into two subsets, called the "black" and "white" subsets, by imposing a checkerboard pattern on the map. Each square of the checkerboard contained a 5x5 square of 25 sites. This arrangement ensures that each of the two subsets contains representative sites from all over the U.S. while reducing local spatial correlations between sites in one subset and sites in the other. These subsets were used for training and testing the neural network and MAPSS models. To assess the performance of a model, we compute the percentage of sites correctly classified by the model.

In addition to evaluating the models on current climate, we also collected three data sets describing future climate assuming a doubling of atmospheric CO_2. These climate data sets were computed by the Geophysical Fluid Dynamics Laboratory (GFDL), Oregon State University (OSU), and United Kingdom Meteoroological Office (UKMO) global climate models (Wetherald & Manabe, 1988; Schlesinger & Zhao, 1989; Michell & Warrilow, 1987). For these future climates, we do not have the corresponding vegetation distribution, so we cannot measure the percentage of sites correctly classified by the neural network and MAPSS models.

We can, however, obtain a bound on the best achievable accuracy by the following procedure. Suppose that we could repeat the following experiment N times:

1. Measure current climate and vegetation classification for the "black" sites.
2. Train one of our models (e.g., a neural network model) on these sites.
3. Apply the trained model to predict future vegetation classifications for one of the future climate data sets.

Consider one particular site x_{23} in the future climate data set. Let $c(x_{23}, j)$ be the number of times that site x_{23} was classified as vegetation class j by the trained model. We can define $p(x_{23}, j) = c(x_{23}, j)/N$ to be the proportion of the N trials in which x_{23} was predicted to have vegetation class j. Now suppose the correct vegetation class was class 10. Then $p(x_{23}, 10)$ estimates the probability that x_{23} is classified correctly. More generally, if $f(x)$ is the correct class for site x, then $\sum_x p(x, f(x))$ will be the average number of sites correctly classified. If there are m sites, then $\sum_x p(x, f(x))/m$ is the expected proportion of correct classifications of the model.

Unfortunately, we do not know the correct vegetation classes $f(x)$. But we can make the optimistic assumption that the correct class is the one that was predicted most often by the trained models: $f^*(x) = \max_j p(x, j)$. Then $\sum_x p(x, f^*(x))/m$ is the best possible accuracy that could be achieved by the model. We define the *classification scatter* to be $1 - \sum_x p(x, f^*(x))/m$. This quantity measures how much the model "scatters its guesses" when it is trained on different training sets. It is the lowest possible error rate that could be achieved by the model.

The key idea underlying this approach is that there are two fundamental sources of error in any prediction scheme. First, there is the choice of the space of models to consider. One could choose to consider neural network models or one could choose to consider possible parameterizations of MAPSS. Second, there is the choice of the data to use for training the models. The actual training data that we have available is presumably corrupted by various sources of noise. If we were to repeat the experiment, the new data set would also be corrupted by these noise sources. This would produce different values in the data set, which would in turn result in different trained models. The above scheme of repeating the data collection, training, and prediction would permit us to measure how much the choice of the space of models and the noise in the data cause our guesses to scatter.

Table 1: Median proportion of correct classifications for the two modeling methods based on 8 runs of each.

Method	% Correct Predictions
Neural Network	68.35
MAPSS	61.75

Of course we can't repeat the experiment of gathering multiple data sets and training on each of them. However, we can simulate this process through a technique known as "bootstrapping" (Efron & Tibshirani, 1993). The idea is to generate a series of "bootstrap replicate data sets" from our initial data set. If the given data set contains n data points, we generate a boostrap replicate data set by drawing n points *randomly with replacement* from the given data set. This simulates the process of drawing data from the true underlying distribution. The rest of the procedure is the same as described above. We train the model on each of the bootstrap replicate data sets and then compute the classification scatter on a separate test data set. In practice, this will tend to underestimate the classification scatter. But it will provide a useful bound on the best accuracy that these models could achieve in predicting future vegetation distribution.

3 Results

3.1 Predicting vegetation distribution under current climate

To compare neural network models to MAPSS on the task of predicting current climate, we performed the following experiment. MAPSS and the neural network were trained eight times on the sites in the "black" squares and then evaluated on the sites in the "white" squares.

For MAPSS, a subset of only 72 "black" sites was used for training. The sites were chosen randomly while ensuring that sites from each vegetation class were chosen in proportion to the frequency of occurrence of that class. The reason for choosing such a small set of sites is that the automatic calibration of MAPSS is very expensive, and previous experiments had shown that samples of approximately this size gave good fits to the training data while minimizing the computational expense.

For the neural networks, 200 "black" sites were chosen at random to form the halting set, and the remaining 1,448 sites formed the training set.

In each of the eight training runs of each model, the parameters of the model were randomly initialized.

Table 1 shows the median proportion of correct classifications of the "white" sites for each of the models. This difference is statistically significant (according to McNemar's test, $p < 0.001$), and it shows that the neural network model is giving greater accuracy than MAPSS for predicting current climate.

3.2 Predicting vegetation distribution under future climate scenarios

We took the 8 calibrations of MAPSS and the 8 neural networks trained above and applied them to predict future vegetation distribution based on the three global climate models mentioned above. The classification scatter was then measured and used to compute a bound on the maximum possible prediction accuracy.

Table 2: Estimated maximum possible predictive accuracy on three future climate scenarios

	Maximum possible % accuracy	
Climate scenario	Neural networks	MAPSS
GFDL	62	91
OSU	77	91
UKMO	70	91

Table 2 summarizes the results. Caution is needed when interpreting these numbers. First, it is important to remember that the numbers are based on using the bootstrap to approximate the real variability in the training data. Second, it is possible for a completely trivial model to obtain a score of 100%. Consider a model that predicts "Conifer forest" for all sites regardless of the training data used to calibrate the model. It would obtain a score of 100% in this experiment. However, we have very good reasons to believe that its true accuracy would be much less. The value of 91% for MAPSS in the table says only that the true accuracy of MAPSS will lie somewhere between 0 and 91%.

We see that in all cases, MAPSS has higher potential accuracy than the neural network models. This is because the neural network models are scattering their guesses over a wider range of classes. This reflects the fact that the neural network models are much less constrained than MAPSS—they know nothing about plant ecosystems, whereas MAPSS incorporates substantial knowledge of hydrology and plant physiology. Without the constraint of prior knowledge, the neural network models are guided only by the training data. They can be expected to be accurate if the test data is similar to the training data. But the future climates are so different from current climate that the neural network predictions are unconstrained in these future climates, so they "bounce around" wildly.

This spread of predictions is most striking for the GFDL climate. The maps in Figure 1 show two of the ten runs of the neural networks and of MAPSS. We can see that the MAPSS predictions are much more consistent than the neural network predictions. The MAPSS runs are generally in agreement that the western mountains will remain predominantly shrubland and that the eastern United States will largely become incapable of supporting forests and will shift over to grasslands. The results are less clear for the neural networks. The pattern of shrubs in the west is fairly consistent, and the two runs agree fairly well in the south central portion of the country. However, there are glaring differences in the north. Figure 2 plots those sites where the two runs agree.

4 Discussion

The performance of both models on current-climate predictions is somewhat disappointing. We believe that it reflects the coarse spatial and temporal resolution of the data. It may also reflect the fact that year-to-year variation in precipitation and temperature can be important causes of biome variation. Neither model can detect this, since the inputs only provide data averaged over a 50-year period. The errors are most frequent in the mountainous regions of the U.S.

The fact that neural networks perform better than MAPSS for current climate is somewhat surprising. This suggests that some of the approximations in the MAPSS process model (e.g., grouping all plants into trees, grasses, and shrubs and dividing the soil into only three layers) may hurt accuracy. On the other hand, the fact that MAPSS can attain performance similar to neural

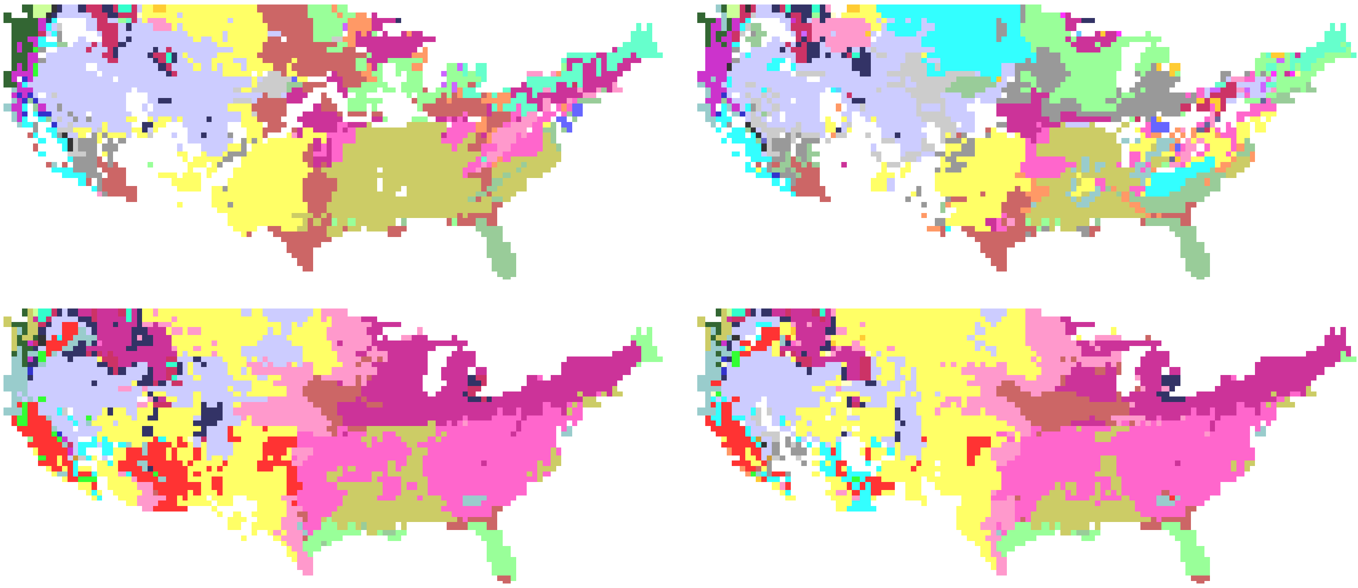

Figure 1: Vegetation predictions under GFDL climate for two neural network runs (top) and two MAPSS runs (bottom). Note the radical differences in the neural network predictions.

networks while using only one twentieth of the amount of training data suggests that the MAPSS model has captured many constraints that are important for making accurate predictions.

We cannot draw any strong conclusions about the accuracy of MAPSS and neural network models for future climate scenarios. However, there are two arguments in favor of believing MAPSS rather than the neural network predictions. First, MAPSS incorporates a substantial body of knowledge about plant physiology and soil hydrology. This knowledge is likely to be invariant to future climate changes. For example, the freezing point of water will certainly not change. The efficiency with which roots extract water from the soil will probably not change over the space of a few decades. This provides strong support for MAPSS and other process-based models.

Second, the neural network models are obviously scattering their predictions widely. This shows that the training data are not sufficient to constrain their predictions. The performance of both MAPSS and the neural networks will certainly drop from the maximum values shown in Table 2. But MAPSS will need to drop much further to end up below the level of performance of the neural networks.

These arguments suggest that faced with the choice of believing MAPSS or believing a neural network model, we should choose MAPSS.

5 Concluding Remarks

Neural networks provide an attractive approach to modeling in ecological applications, because of the accuracy they can achieve without the need for large amounts of prior knowledge. However, as with any modeling methodology, neural networks have both strengths and weaknesses. If the only goal of the MAPSS project was to model vegetation under climate conditions very similar to those that exist currently, then a neural network solution would be appealing, because it would not require knowledge engineering and its accuracy would be comparable to (or even better than) systems that are knowledge intensive.

In any knowledge-intensive system, there is a risk that the knowledge built into the model embodies approximations that cause errors. However, these approximations are typically not too far off. Hence, when extrapolating to conditions very different from those where the model was trained, the builtin knowledge—even if it is approximate—gives us more confidence in the model than we would have in a neural network model that is not constrained by any prior knowledge.

This paper has presented evidence that the accuracy of neural network models in predicting future vegetation distribution is likely to be lower than the accuracy of the MAPSS process-based model. The paper has also introduced a new methodology for obtaining an upper bound on the best achievable accuracy. The methodology applies the bootstrap technique from statistics to measure the classification scatter of a model and thereby bound the potential accuracy of its predictions. This methodology should be applicable in a wide range of modeling applications.

Acknowledgements

The authors thank Ron Neilson and Jesse Chaney for their support and assistance in carrying out this research. The authors gratefully acknowledge the support of the NSF under grants IRI-9204129, ASC-9503976, and CDA-9216172.

References

Barnard, E., & Cole, R. A. (1989). A neural-net training program based on conjugate-gradient optimization. Tech. rep. CSE 89-014, Oregon Graduate Institute, Beaverton, OR.

Caruana, R. (1997). *Multitask Learning*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA.

Dietterich, T. G., Hild, H., & Bakiri, G. (1995). A comparison of ID3 and backpropagation for English text-to-speech mapping. *Machine Learning*, *18*, 51–80.

Duan, Q., Sorooshian, S., & Gupta, V. (1992). Effective and efficient global optimization of conceptual rainfall-runoff models. *Water Resources Research*, *28*(4), 1015–1031.

Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New York, NY.

Gupta, V. K., & Sorooshian, S. (1985). The automatic calibration of conceptual catchment models using derivative-based optimization algorithms. *Water Resources Research*, *21*(4), 473–485.

Hsu, K., Gupta, H. V., & Sorooshian, S. (1995). Artificial neural network modeling of the rainfall-runoff process. *Water Resources Research*, *31*, 2517–2530.

Klepper, O., & Hendrix, E. M. T. (1994). A method for robust calibration of ecological models under different types of uncertainty. *Ecological Modelling*, *74*, 161–182.

Lang, K. J., Hinton, G. E., & Waibel, A. (1990). A time-delay neural network architecture for isolated word recognition. *Neural Networks*, *3*(1), 23–43.

Michell, J. F. B., & Warrilow, D. A. (1987). Summer dryness in northern mid latitudes due to increased CO2. *Nature*, *330*, 238–240.

Neilson, R. P. (1995). A model for predicting continental-scale vegetation distribution and water balance. *Ecological Applications*, *5*, 362–385.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical Recipes in C* (2d edition). Cambridge:New York, NY.

Schlesinger, M. E., & Zhao, Z. C. (1989). Seasonal climatic change introduced by double CO2 as simulated by the OSU attomospheric GCM/mixed-layer ocean model. *Journal of Climatology*, *2*, 429–495.

Tan, S. S., & Smeins, F. E. (1996). Predicting grassland community changes with an artificial neural network model. *Ecological Modelling*, *84*, 91–97.

VEMAP-Members (1995). Vegetation/ecosystem modeling and analysis project: Comparing biogeography and biogeochemistry models in a continental-scale study of terrestrial ecosystem responses to climate change and co2 doubling. *Global Biogeochemical Cycles*, *9*.

Wetherald, R. T., & Manabe, S. (1988). Cloud feedback processes in a general circulation model. *Journal of Atmospheric Science*, *45*, 1397–1415.

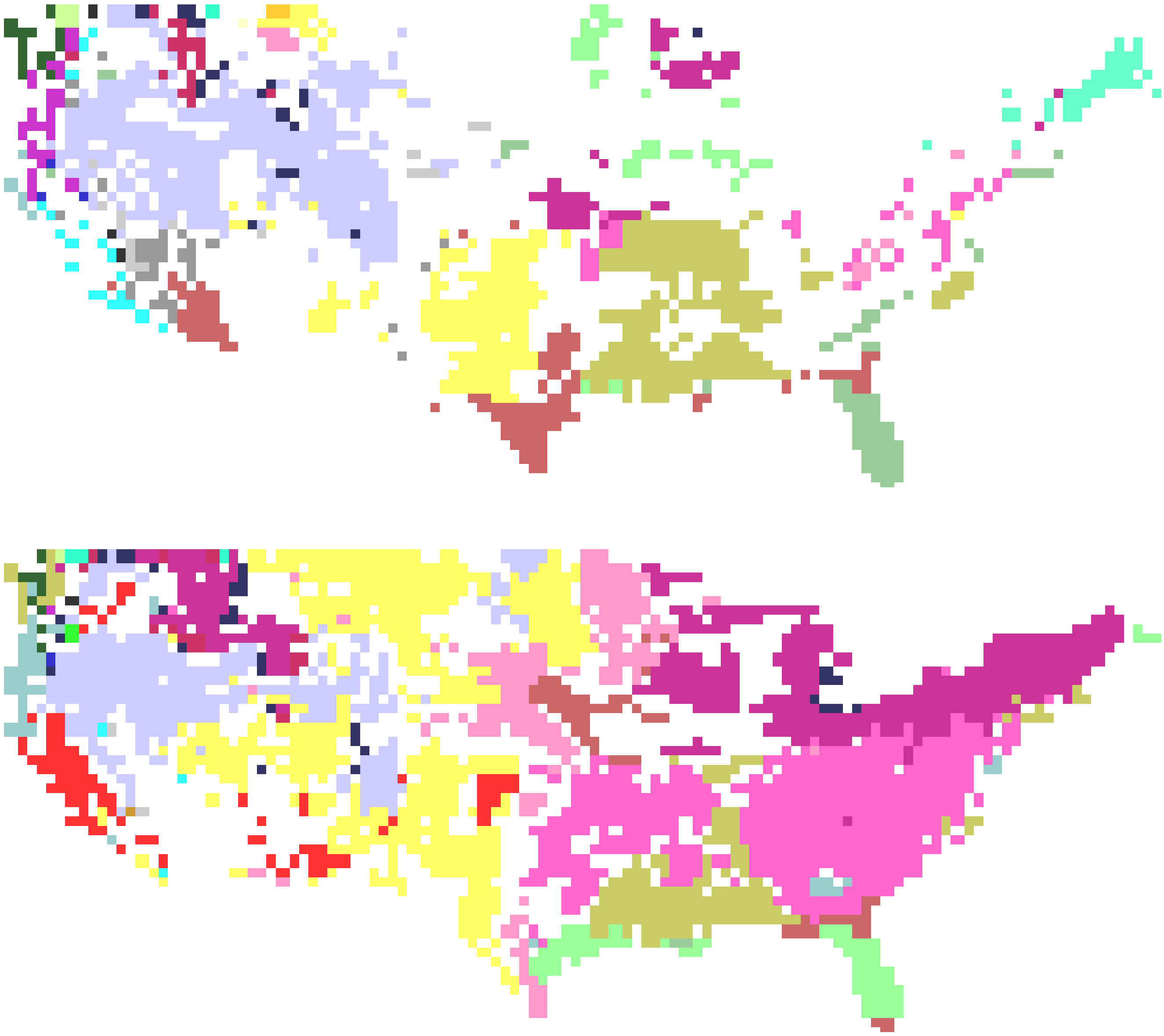

Figure 2: Sites that have the same GFDL prediction in both neural network runs (top) and both MAPSS runs (bottom). Note that the areas of agreement are much more sparse for the neural network runs.