Feedforward Neural Network Estimation of a Crop Yield Response Function

Wayne H. Joerding, Ying Li, and Douglas L. Young*

Abstract

Feedforward networks have powerful approximation capabilities without the "explosion of parameters" problem faced by Fourier and polynomial expansions. This paper first introduces feedforward networks and describes their approximation capabilities, then we address several practical issues faced by applications of feedforward networks. First, we demonstrate networks can provide a reasonable estimate of a Bermudagrass hay fertilizer response function with the relatively sparse data often available from experiments. Second, we demonstrate that the estimated network with a practical number of hidden units provides reasonable flexibility. Third, we show how one can constrain feedforward networks to satisfy *a priori* information without losing their flexible functional form characteristic.

Key Words: neural networks, biological process models, feedforward networks, production function

Introduction

Mitscherlich is credited as the first to suggest a nonlinear algebraic relationship between fertilizer and crop yield in 1909 (Mitscherlich, Heady and Dillon). A decade later, Spillman independently proposed and estimated a similar exponential fertilizer-yield response equation. These early single-input production functions helped provide a functional expression for the "law of diminishing marginal productivity." They also helped launch a revolution in agricultural production economics and farm management which wedded deductive microeconomic theory with empirical statistical analysis. Subsequent generations of production economists have made impressive advances in improving the flexibility of agricultural production functions and in statistical techniques for estimating these equations (Heady and Dillon, Beattie and Taylor, Chambers).

Explicit functional forms estimated statistically have several advantages for describing biological responses. They are usually simple to estimate using readily available software. The resulting equations are relatively easy to communicate to users. Desired theoretical properties often can be incorporated in the equations and tested statistically. Useful measures such as elasticities of production, elasticities of substitution, and marginal and average productivity equations are usually directly derivable.

The greatest disadvantage of explicit functional forms used for biological response has been their inflexibility. The functional forms themselves may impose specification error, thereby precluding the data from expressing itself. For example, the popular Cobb-Douglas functional form imposes zero production at zero input(s), unbounded output, constant unitary elasticity of substitution

*Authors are respectively: associate professor, Department of Economics, Washington State University, Pullman, WA; economist, FannieMae, Washington DC; and professor, Department of Agricultural Economics, Washington State University, Pullman, WA. This work has received partial support from NSF grant SES-9022773 and Torque Systems, Inc., and the Washington State University College of Agriculture Research Center.

J. Agr. and Applied Econ. 26 (1), July, 1994: 252-263

between inputs, and constant input elasticities. All these restrictions are unrealistic or at least unnecessary in describing crop response to commercial fertilizers, for example. To overcome these difficulties, production economists have made steady progress in identifying explicit functional forms which impose fewer theoretically unrealistic restrictions. For example, a simple quadratic response function eliminates all the inflexibilities cited above for the Cobb-Douglas form, but introduces others. Over the last two decades, economists have increasingly used "flexible" functional forms such as second-order Taylor series approximation or the even more flexible Fourier form to express general nonlinear relationships. Even more recently, feedforward neural networks have joined the collection of nonparametric regression estimators with desirable approximation characteristics, but these as yet have been rarely used to model production processes.

Recent advances in computer technology and in the modeling of biological systems, have resulted in rapid growth in the use of process simulation models to generate data on plant and animal responses to environmental and management inputs (Musser and Tew; Stockle; Dillon, Mjelde and McCarl). These process simulation models are often comprised of dozens of equations and subroutines which describe and link the fundamental chemical and physical subsystems which generate growth. The biochemical basis of these systems increases the likelihood that fundamental discontinuities in the input-output relationships will be revealed. Approaches which can efficiently describe input-output relationships with potential discontinuities and multiple outputs and inputs are needed to summarize the voluminous results of process simulation models.

Neural network computer models provide a potentially promising method for summarizing input-output results from biophysical simulators. Neural networks can also use data generated by traditional field experiments to describe biological input-output response functions. Agricultural scientists have successfully used neural networks to describe diverse relationships including that between weather and soil moisture, and sensory judgment and fruit coloration (Kunkel, Thai and Skewfelt). Some pioneering agricultural economics applications of neural networks include prediction of end-of-season corn yield from early season weather measurements, and prediction of planted acreage of corn and wheat from lagged prices and other variables (Uhrig, Engel, and Baker; Uhrig and Botkin).The objectives of this paper will be to describe the relatively new feedforward neural network approach for estimating flexible input-output relationships; to illustrate and evaluate the network approach in estimating a crop yield response function to fertilizer based on experimental data; to compare the network results to those of a traditional statistical approach; and finally, to assess the potential for networks in estimating biological response relationships using data from both experimental and process simulation sources. Specifically, a feedforward neural network and an OLS regression are used to describe the relationship of Alabama Bermudagrass hay to nitrogen and potash fertilizer. This two-input single product example facilitates graphical comparisons of the network and OLS results for a familiar agricultural response problem. Finally, the paper uses a new estimation algorithm for feedforward neural networks that avoids the local minima problem faced by traditional optimization algorithms and allows imposition of *a priori* information.

Overview of Neural Networks

Wasserman provides an introduction to neural networks. For a more specialized treatment see McClelland and Rumelhart. As the name suggests, neural networks model certain aspects of neural activity by linking together many simple units, called neurons, into a complicated network of interconnections. This network can then perform calculations or store information as determined by the network connections.

This study uses a class of neural network models called feedforward networks, which organize a network into an input layer, one or more hidden layers, and an output layer, each layer containing one or more units. Each layer feeds forward to the next layer, no feedback of signals allowed.

Each layer of the network has an input vector denoted by **z** and an output vector denoted by **a**. Let, $\mathbf{z}_i = (z_{i1},\ldots,z_{ik_i})$ represent a $k_i \times 1$ vector of inputs to the ith layer and $\mathbf{a}_i = (a_{i1},\ldots,a_{ik_i})$ a $k_i \times 1$ vector of outputs for the ith layer. The outputs of one layer define the inputs to the next layer by a linear function $\mathbf{z}_k = W_{k-1}\,\mathbf{a}_{k-1}$, where W_ℓ denotes a $k_{\ell+1} \times k_\ell$ matrix of parameters or weights. Observe that all the outputs of the preceding layer potentially influence each of the inputs of the next layer. By convention, the output to layer zero (the input layer) equals the input of layer zero, i.e. $\mathbf{a}_0 = \mathbf{z}_0$. Each unit in a hidden layer takes its input and performs a nonlinear transformation, represented by $f(\cdot)$ and called an activation function, to produce the output of the unit. That is, $a_{\ell,k} = f(z_{\ell,k})$ for $k = 1,\ldots, k_\ell$. The final output of the network occurs at the output layer, say layer L. The output layer may or may not have an activation function, we choose no activation, so letting y represent the output, $y \equiv \mathbf{z}_L = W_{L-1}\mathbf{a}_{L-1}$.

A sequence of vector operations conveniently expresses this structure. Let $\mathbf{F}_{k_\ell}(\cdot)$ represent an array of activation functions, i.e. $\mathbf{F}_{k_\ell}(\mathbf{z}_\ell) \equiv [f(z_{\ell,1}),\ldots, f(z_{\ell,k_\ell})]^T$. Also, let $\mathbf{x} = (x_1,\ldots, x_{k_0})^T$ represent a vector of network inputs to a feedforward network. Then,

$$
\begin{aligned}
\mathbf{z}_0 &= \mathbf{x}\\
\mathbf{a}_0 &= \mathbf{z}_0\\
\mathbf{z}_1 &= W_0\mathbf{a}_0\\
\mathbf{a}_1 &= \mathbf{F}_{k_1}(\mathbf{z}_1)\\
&\;\;\vdots\\
\mathbf{z}_{L-1} &= W_{L-1}\mathbf{a}_{L-1}\\
\mathbf{a}_{L-1} &= \mathbf{F}_{k_{L-1}}(\mathbf{z}_{L-1})\\
y &= W_{L-1}\mathbf{a}_{L-1}
\end{aligned}
\tag{1}
$$

For a single hidden layer network with a scaler output (i.e. $L = 2$ and $k_2 = 1$) the above sequence reduces to a much simpler model. Let $\psi(\mathbf{x})$ denote a feedforward network. Then

$$\psi(\mathbf{x}) \equiv \mathrm{W}_1\mathbf{F}(W_0\mathbf{x}) \tag{2}$$

simplifies to

$$\psi(\mathbf{x}) = \sum_{i=1}^{k_1} w_{11i} f(\mathbf{w}_{0i}\mathbf{x}) = \sum_{i=1}^{k_1} w_{11i}\, f\Big(\sum_{j=1}^{k_0} w_{0ij}x_j\Big)\,, \tag{3}$$

where $\mathbf{w}_{0i}$ represents row i of matrix W_0 and $w_{\ell ij}$ the i,j element of matrix W_ℓ.

Our interest in feedforward networks arises from their powerful approximation characteristics (Carroll and Dickinson; Cybenko; Hornik, Stinchcombe and White (1989); Ito). Most useful for our purposes, Hornik, Stinchcombe and White (1990) have shown that for a suitable choice for $f(\cdot)$ a single hidden layer feedforward network can approximate any piecewise differentiable function *and* its derivatives on open bounded subsets of $\mathbb{R}^k$ to any desired degree of accuracy. This study uses the widely applied logistic function $f(v) = 1/(1 + exp(-v))$ which satisfies the required conditions on $f(\cdot)$.

Single hidden layer feedforward networks belong to a class of semi-nonparametric series estimators that have approximation abilities similar to that described for feedforward networks. Andrews describes series estimators by $y = \sum_{i=1}^{k1} \theta_i g_i(\mathbf{x})$, where $g_i:\mathbb{R}^k \to \mathbb{R}$ belong to some pre-specified family of functions. The family of functions $\{g_i(\cdot)\}$ traditionally consists of polynomial spline or trigonometric functions. However, there exists many other families of functions that enable series estimators to have good approximation characteristics. The work cited in the previous paragraph simply adds a family of sigmoid functions to the list of function families with good approximation abilities by setting $g_i(\mathbf{x}) = f(\mathbf{w}_{0i}\ \mathbf{x})$.

The well-known Fourier Flexible Form (FFF), (Gallant, 1982) based on the trigonometric family, has the same powerful ability as feedforward networks to approximate piecewise continuous functions and their derivatives. Interestingly, one can express the Fourier form as a special case of the feedforward network with sine and cosine activation functions, but such a network does not satisfy the conditions required by Hornik, *et al* (1990), because sine and cosine functions do not satisfy the ℓ-finite condition. Thus, the approximation ability of feedforward networks depends on a fundamentally different mechanism from the mechanism supporting the same ability in the Fourier form.

There exists an important difference between feedforward networks and other traditional series estimators since each of the $g_i(\cdot)$ functions of a network, i.e. the activation functions, contain estimable parameters, i.e. $g_i(\mathbf{x}) \equiv f(\mathbf{w}_{0i}\ \mathbf{x})$, while polynomial and trigonometric families use predetermined $g_i(\mathbf{x})$ functions, e.g. $g_i(\mathbf{x}) = sin(i\mathbf{x})$. This enables networks to avoid the "explosion of parameters" problem associated with polynomial and trigonometric families while obtaining a better rate of approximation than series estimators based on polynomial and trigonometric families. Specifically, letting d represent the dimension of an input vector and n represent the number of terms in the series expansion, (i.e., for the networks described above $n = k_1$ and $d = k_0$), the number of parameters in a polynomial series approximation or the FFF increase at order n^d while the parameters of a feedforward network increase at order nd. Despite the more slowly growing number of parameters, Barron shows that feedforward networks can achieve an integrated squared approximation error of order $O(1/n)$ while other series estimators can only achieve an approximation error of order $O((1/n)^{2/d})$, for functions satisfying the same smoothness property. Thus, only for problems with a scaler input can polynomial and Fourier methods outperform networks, and networks dominate for problems with more than two inputs.

Networks face some disadvantages compared to traditional series estimators. First, the statistical properties of networks remain undeveloped. The work of Gallant (1981) and of Andrews does not apply directly because the $g_i(\cdot)$ functions of networks contain estimable parameters. Second, feedforward network models are nonlinear in the parameters, so commonly used estimation algorithms may not find the globally optimal set of connection weights. This problem can be solved with considerable computational resources, thus we expect this handicap to decline over time.

Some criticize networks for being black boxes with which the researcher cannot determine the mechanism by which independent variables affect the dependent variable. This criticism has been overstated since networks can address many of the same questions for which other flexible functional forms have been used. For example, one can easily obtain estimates of elasticities and marginal products from a network in the same manner that these values are computed from other forms. Moreover, by developing methods to impose *a priori* constraints on networks, as shown in the example below, one can develop tests of economic hypotheses.

The neural network literature often refers to estimation as training, but by either name uses a least squares criterion to find optimal values of $W = (W_0, W_1)$ in a parameter space $\mathbf{W}$. The most common algorithm uses a steepest descent method, called back-propagation, to minimize the SSE (sum-of-squared errors). Unfortunately, back-propagation, or other gradient methods such as those available in SAS, can easily stick at local minima to the SSE surface and show sensitivity to starting values. The alternative we adopt uses a random search algorithm, called Simulated Annealing, to search $\mathbf{W}$ for a globally minimizing set of parameters. Unfortunately, this approach requires considerable computation and has no general purpose implementations available, thus we have written our own algorithm.

The actual algorithm uses a hybrid linear least squares and simulated annealing method, (Joerding and Li; Li, Joerding, and Genz). Briefly, this hybrid algorithm takes a random step in the parameter space for W_0, computes outputs for the hidden layer units, then uses a QR decomposition and back-substitution to compute the value of W_1 that minimizes the SSE. (See Golub and Van Loan for a discussion of the QR decomposition. We use the QR decomposition instead of the faster Gaussian elimination method because of its superior numerical stability.) The algorithm proceeds by comparing the resulting new SSE to the old SSE generated by the previous W_0. If the old SSE exceeds the new SSE, then update the W_0 weights to the new W_0, otherwise, keep the old W_0 and make a new guess. If the new SSE exceeds the old SSE, then keep the "inferior" new W_0 with a probability determined by an exponential distribution. The probability of accepting a "bad" step declines to zero as the algorithm proceeds. With unlimited iterations this approach always finds the global minimum to the SSE function (Aarts and Korst). In practice, we limit the number of iterations meaning that the algorithm may stop at a local minimum inferior to the global minimum. However, simulated annealing has worked well in a variety of non-convex optimization problems (Aarts and Korst).

Data

The data utilized in this study for estimating a traditional statistical response function and neural network relationship are drawn from a Bermudagrass hay fertilizer response study (Engibous and Young). The data were collected in a five year study over 1972-76 at the Brewton experimental field at Auburn University in Alabama. Annual harvest of Coastal Bermudagrass hay is measured in pounds of oven-dried forage. Soil tests revealed that soil phosphorus was high and precluded the need for supplemental phosphate fertilization. Plots received combinations of nitrogen (N) and potash (K_2O) fertilizer varying between 0 and 600 pounds per acre of nitrogen, and 0 and 500 pounds per acre of potash.The training data contain twenty configurations of nitrogen and potash, distributed as shown in figure 1 over five years for a total of 100 observations. The response data were available only as averages over equal numbers of spatial replications for each treatment within years. Since the inputs were identical within each averaged group and the number of replications for each input combination were equal, averaging the responses involves no efficiency loss (Theil).

Results

We estimated a network with five hidden units and no constraints, since we wanted the network to approximate a general yield relationship. This five unit model produced a sum of squared errors equal to 255,580,000 and an R^2 equal to .89. Figure 2 displays the estimated network production yield surface and contour map. Biological processes such as these can exhibit negative marginal returns to fertilizer input, which in our case shows up most clearly with nitrogen, especially with little or no potash application. We can see this more clearly in figure 3 which shows sections of the production yield surface and marginal product surface at the 400 pounds/acre of nitrogen and 250 pounds/acre of potash input. Notice that the network estimates negative yields at 600 pounds of nitrogen and zero of potash which is still within the range of the data. Of course, no actual biological process can generate negative pounds of Bermudagrass, which shows how a network, or any unconstrained functional form, can violate *a priori* information.

For comparison we estimated a quadratic approximation to the biological response function by ordinary least squares regression. The Student t-ratio is presented in parenthesis:

$$\begin{aligned} Yield = \underset{(2.9)}{1820.3} &- \underset{(-8.7)}{0.079*N^2} - \underset{(-3.9)}{0.055*(K_2O)^2} \\ &+ \underset{(3.4)}{0.061*N*K_2O} + \underset{(12.4)}{54.79*N} + \underset{(3.6)}{17.33*K_2O}, \\ R^2 &= .866, \\ SSE &= 291{,}183{,}662 \qquad (4) \end{aligned}$$

The quadratic approximation is plotted in figure 4. Notice that the quadratic representation does not violate the negativity restriction at $N = 600$ and $K_2O = 0$, as with the network representation. However, the symmetry of a quadratic representation means that for some input combinations the quadratic representation will predict negative output. Evaluating equation (4), we determine that the quadratic representation will predict negative output for N greater than 733.27 when no potash is applied.

Typically, one doesn't impose a positivity constraint on a quadratic representation because of the obvious bias introduced. For example, only a convex quadratic function (open side up bowl) has positive output over the entire positive orthant, hardly a realistic representation for a biological response function. A network representation doesn't face this problem because of its flexibility. That is, one can constrain the network to predict only positive values without biasing the estimated output surface over the range of the data set as long as the true response functions satisfies piecewise continuity. Indeed, including the *a priori* information that biological response functions must produce positive values should improve the estimation efficiency.

We re-estimated the network, using techniques described in Joerding, Li, Hu, and Meador; and Joerding and Meador, imposing a positivity constraint at the point $N = 600$ and $K_2O = 0$, which lies at the boundary of the data domain, with results displayed in figure 5. The basic shape of the relationship remains unchanged except for

Figure 1. Experimental Combinations of Potash (K_2O) and Nitrogen (N) in Pounds/Acre Applied in Bermudagrass Hay Experiment over Five Years

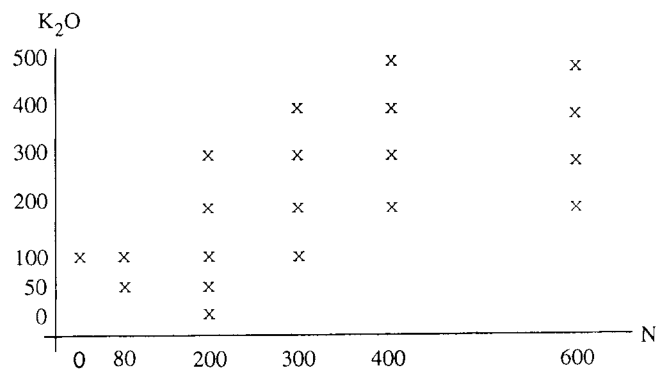

Figure 2. Estimated Unconstrained Network Bermudagrass Yield and Isoquant Surfaces (Yield, Nitrogen (N), and Potash (K_2O) in Pounds/Acre)

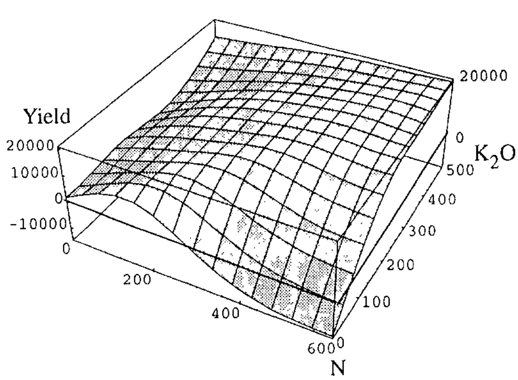

Figure 2a

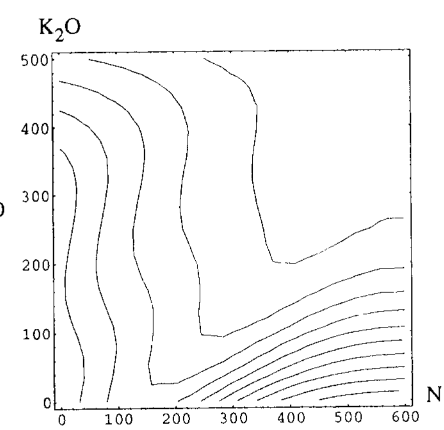

Figure 2b

Figure 2

Figure 3. Estimated Unconstrained Network Yield and Marginal Product Curves (Yield, Marginal Product (*MP*), Nitrogen (*N*), and Potash (K_2O) in Pounds/Acre)

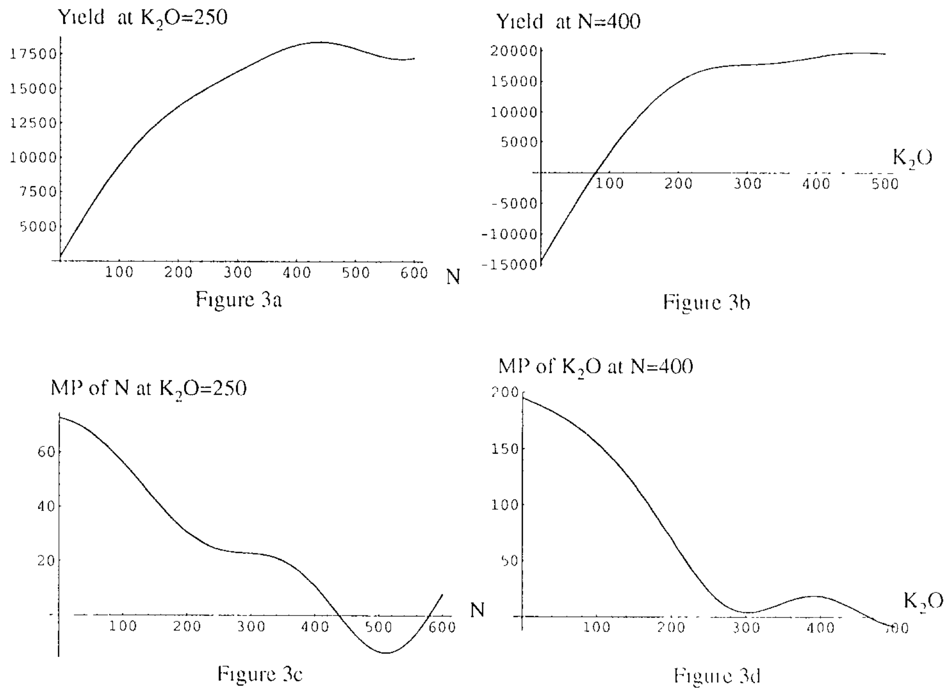

Figure 3.

Figure 4. Estimated Quadratic Bermudagrass Yield and Isoquant Surfaces (Yield, Nitrogen (*N*), and Potash (K_2O) in Pounds/Acre)

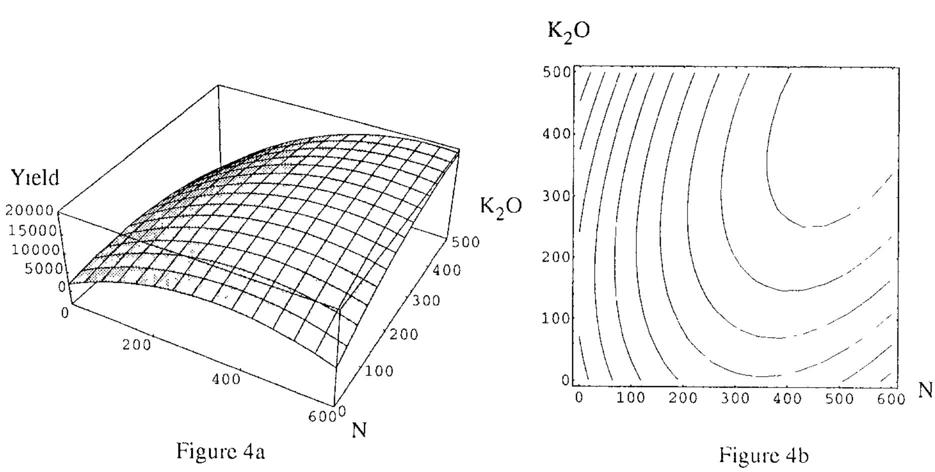

Figure 4.

Figure 5. Constrained Network Bermudagrass Yield and Isoquant Surfaces (Yield, Nitrogen (N), and Potash (K_2O) in Pounds/Acre)

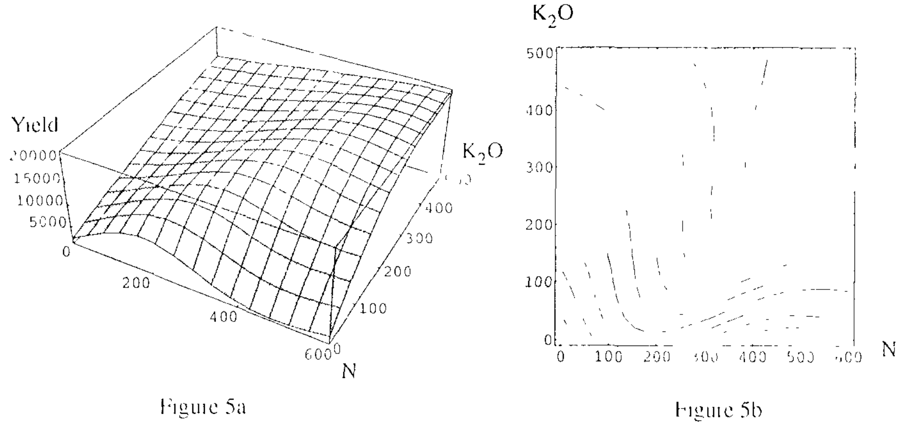

Figure 5.

smoothing the non-convexity in the isoquants over the top left portion of figure 5b. Of course the quadratic form used for the isoquant surface in figure 4b imposes symmetric oval isoquants.

Both the constrained and unconstrained network estimates show concave isoquants on the fringe of the sample domain. This concavity may result from extrapolating into regions outside the sample domain, but they may reflect the actual response function. Concave isoquants have been found in other biological processes; Brokken presents a livestock feeding example. They also raise the possibility of cost minimization at corner solutions. Corner solutions appear feasible at certain points (figure 5b). The agronomic feasibility of such fertilization patterns on Bermudagrass hay in the local soils merits further investigation. Single nutrient fertilization, as suggested by corner solutions on concave isoquants, is fairly common for some crops and regions. Of course, single nutrient fertilization could also be profit maximizing with convex isoquants having weak curvature.

Conclusions

Previous theoretical work, described earlier, has shown feedforward networks to have powerful approximation capabilities similar to the Fourier flexible form but without the "explosion of parameters" problem faced by Fourier and polynomial expansions. Nevertheless, several practical questions need answers before recommending feedforward networks for empirical analysis. Our study has sought to address some of these questions.

First, can networks provide reasonable results with the relatively sparse data available from experiments? Our data set had only 100 observations yet produced fairly reasonable results compared to a simple quadratic reference model. In some ways the two models barely differed. For example, holding potash at 250 pounds/acre, the network estimated maximum hay output at about 465 pounds/acre of nitrogen while the quadratic approximation estimated maximum output at about 445 pounds/acre of nitrogen.

Second, can networks with a practical number of hidden units provide the necessary flexibility to "let the data speak"? Our estimated network contained only five hidden units yet found possible evidence of non-convex isoquants in two regions of the input space. Non-convex isoquants are consistent with observed single-nutrient fertilization and suggests a possibly fruitful area for more experimentation with a broader range of nitrogen and potash combinations.

Third, can one constrain feedforward networks to satisfy a priori} information without losing the flexible functional form characteristic? Our unconstrained network violated positivity for hay yield, so we developed an algorithm to impose the positivity condition. The resulting network retained enough flexibility to capture the non-convex isoquants found in the unconstrained model without apparently biasing the results elsewhere.

We answered the above questions in the affirmative, but feedforward networks also have disadvantages. First, estimating feedforward networks in such a way as to avoid local minima requires considerable computational resources. Our unconstrained results required almost 24 hours for a solution on a special purpose parallel processing computer capable of up to 60 million floating point operations per second. (The comparable figure for an IBM-type 486 machine is probably around 2 million.) Advances in computer technology should reduce this problem in the future. Second, while off-the-shelf programs for estimating feedforward networks exist, these programs use back-propagation, a method that can stick at local minima to the sum of squared errors. Moreover, these back-propagation programs do not allow imposition of *a priori* information, like positivity. Thus, we found it necessary to write our own program that implements a global minimization routine and provides a way to impose *a priori* constraints. This programming effort has been non-trivial, resulting in over 8,000 lines of C code. Third, and most important, networks share with other flexible forms, like the Fourier flexible form, a difficulty in displaying results. A network doesn't have a particular parameter to measure an input's marginal product. Instead, we have needed to rely on figures and descriptive statistics. Users of these flexible forms need to develop better methods to communicate their results. Fourth, as with the other flexible functional forms, users face the difficult problem of deciding on the optimal number of hidden units. Too few and the network fails to extract all the information from the data, too many and the network overfits the data. As yet, no widely accepted or practical solution to this problem has been found.

We were reasonably successful in estimating a crop yield response relationship using feedforward networks and only 100 experimental observations. While this two-input single output example was useful for illustrating the relatively new neural network procedure, it is unlikely that networks will replace regression analysis for routine estimation of small dimensional production functions.There are good reasons, however, why networks might be useful in the future for estimating complex multiple input-multiple output relationships from experimental or process simulation results. There is growing interest in simulating multiple production and environmental outputs such as multiple crop yields, wildlife populations, soil erosion, and nitrogen leaching from multiple management inputs. Neural networks are well adapted for summarizing such complex processes. However, our ongoing work in this area indicates that the computer programming and processing requirements are considerably more formidable for multiple input-multiple output applications.

Networks are also well adapted for describing innate discontinuities in biological phenomena modeled in process simulations. For example, plants or animals may die or cease growing when certain inputs (or toxins) reach threshold levels. The ability of networks to approximate such relationships derives from their denseness in spaces of piecewise differentiable functions.

Finally, successful estimation of complex input-output relationships using neural networks may require very large quantities of data. The marginal cost of generating additional observations from a stochastic process simulation model is very low once the model is built and validated, providing an inexpensive source for data to estimate feedforward networks.

References

Aarts, E. and J. Korst. *Simulated Annealing and Boltzmann Machines*. New York: John Wiley and Sons, 1989.

Andrews, D. W. K. "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models." *Econometrica,* 59(1991):307-345.

Beattie, B. R., and C. R. Taylor. *The Economics of Production*. New York: John Wiley and Sons, 1985.

Barron, A. R. "Universal Approximation Bounds for Superpositions of a Sigmoidal Function." Manuscript, Department of Statistics, University of Illinois, IL 61820, 1991.

Brokken, R. F. "The Case of a Queer Isoquant: Increasing Marginal Rates of Substitution of Grain Roughage in Cattle Finishing." *West. J. Agr. Econ.,* 1(1977):221-24.

Carroll, B. W. and B. D. Dickinson. "Construction of Neural Net Using the Radon Transform." *'89IJCNN Proceedings,* I(1989):607-611.

Chambers, Robert G. *Applied Production Analysis: A Dual Approach*. New York: Cambridge U. Press, 1988.

Cybenko, G. "Approximation by Superpositions of a Sigmoidal Function." *Mathematical Control Signal Systems,* 2(1989):303-314.

Dillon, Carl R., James W. Mjelde, and Bruce A. McCarl. "Recommendations on the Implementation and Development of Biophysical Simulation Models in Agricultural Economics Research." Monograph. Dept. of Agr. Econ., Texas A & M U., College Station, 1989.

Engibous, J. C., and D. L. Young. "Economics of Fertilizer Use." Ch. 10 in *The Fertilizer Handbook*. The Fertilizer Inst., Washington, DC, 1982.

Gallant, A. R. "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form." *Journal of Econometrics,* 15(1981):211-245.

Gallant, A. R. "Unbiased Determination of Production Technologies." *Journal of Econometrics,* 20(1982):285-323.

Gallant, A.R. and G. H. Golub. "Imposing Curvature Restrictions on Flexible Functional Forms." *Journal of Econometrics,* 26(1984):295-321.

Golub, G. H. and C. F. Van Loan. *Matrix Computations*. Baltimore: Johns Hopkins University Press, 1989.

Heady, Earl O., and John L. Dillon. *Agricultural Production Functions*. Ames: Iowa State U. Press, 1961.

Hornik, K., M. Stinchcombe, and H. White. "Multilayer Feedforward Networks are Universal Approximators." *Journal of Neural Networks,* 2(1989):359-366.

Hornik, K., M. Stinchcombe, and H. White. "Universal Approximation of an Unknown Mapping and Its Derivatives Using Multilayer Feedforward Networks." *Journal of Neural Networks,* 3(1990):551-560.

Ito, Y. "Approximation of Functions on a Compact Set by Finite Sums of a Sigmoid Functions Without Scaling." *Journal of Neural Networks,* 4(1991):817-826.

Ito, Y. "Approximation of Continuous Functions on $\mathbb{R}^d$ by Linear Combinations of Shifted Rotations of a Sigmoid Function With and Without Scaling." *Journal of Neural Networks,* 5(1992):105-116.

Joerding, W. H. and Y. Li. "Global Training with *A Priori* Constraints." Manuscript, Department of Economics, Washington State University, Pullman, WA, 99164, 1992.

Joerding, W. H. and Y. Li, S. Hu, and J. Meador. "Approximating Production Technologies with Feedforward Neural Networks." Forthcoming *Advances in Artificial Intelligence in Economics, Finance, and Management.* Greenwich: AI Press, 1992.

Joerding, W. H. and J. L. Meador. "Encoding *A Priori* Information in Feedforward Networks." *Journal of Neural Networks,* 4(1991):847-856.

Kunkel, K. E. "Operational Soil Moisture Estimation for the Midwestern United States." *J. Appl. Meteorology, Notes and Correspondence,* 29(11):1158-66, 1990.

Li, Y., W. H. Joerding and Genz, A. "A Hybrid LS/Simulated Annealing Training Algorithm for Feedforward Networks." Manuscript, Department of Economics, Washington State University, Pullman, WA, 99164, 1992.

McClelland, J. and D. Rumelhart. *Parallel Distributed Processing.* Cambridge: The MIT Press, 1988.

Mitscherlich, E. A. "Das Gesetz des Minimums und das Gesetz des Abnehmenden Bodenertrages." *Landw. Jahrb.,* 38(1909):537-52.

Musser, W., and B. Tew. "Use of Biophysical Simulation in Production Economics." *So. J. Agr. Econ.,* 16(1984):77-86.

Spillman, W. J. "Applications of the Law of Diminishing Returns to Some Fertilizer and Feed Data." *J. Farm Econ.,* 5(1923):36-52.

Stockle, Claudio O. "Simulation Models in Agriculture: From Cellular Level to Field Scale." Proceedings of the 1989 Summer Computer Simulation Conf., Austin, TX, July 24-27, 1989.

Thai, C. N. and R. L. Shewfelt. "Modeling Sensory Color Quality of Tomato and Peach: Neural Networks and Statistical Regression." *Transactions of the ASAE,* 34(1991):950-955.

Theil, H. *Principles of Econometrics.* New York: John Wiley & Sons, Inc., 1971.

Uhrig, J. W., and C. J. Botkin. "An Application of Neural Networks in Predicting Planted Acreage of Corn and Wheat in the United States." Paper presented at the Amer. Agr. Econ. Assoc. annual meeting, Vancouver, BC, Aug. 4-8, 1990.

Uhrig, J. W., B. A. Engel, and W. L. Baker. "An Application of Neural Networks: Predicting Corn Yields." Paper presented at the Fourth International Congress for Computer Technology in Agriculture, Paris-Versailles, France, June 1-3, 1992.

Wasserman, P.D. *Neural Computing: Theory and Practice.* New York: Van Nostrand Reinhold, 1989.

Endnote

[1]Estimated values for W_0 and W_1 equal

$$\hat{W}_0 = \begin{pmatrix} 7.845E\ {-03}, & -1.093E\ {-02}, & -1.538E\ {+00} \\ 7.262E\ {-03}, & -1.349E\ {-02}, & -4.562E\ {-01} \\ 7.084E\ {-03}, & -1.660E\ {-02}, & 9.312E\ {-01} \\ 4.577E\ {-03}, & -1.103E\ {-02}, & 1.735E\ {+00} \\ 1.068E\ {-05} & 3.284E\ {-03} & 1.312E\ {+00} \end{pmatrix} \tag{5}$$

$$\begin{aligned} \hat{W}_1 = (&-1.227E\ {+05}, \quad 1.4940E\ {+05}, \quad -3.997E\ {+04}, \\ &3.345E\ {+4}, \quad 3.022E\ {+05}, \quad -2.718E\ {+05}) \end{aligned} \tag{6}$$