Sensitivity Analysis of a Model of Vegetative Growth of Soya Bean Plants

June Taeg Lim[1], Harvey J. Gold[1*], Gail G. Wilkerson[2] and C. David Raper, Jr.[3]

[1] Biomathematics Program, Department of Statistics, Box 8203, North Carolina State University, Raleigh, North Carolina 27695, U.S.A.,

[2] Department of Crop Science, Box 7620, North Carolina State University, Raleigh, North Carolina 27695, U.S.A., and

[3] Department of Soil Science, Box 7619, North Carolina State University, Raleigh, North Carolina 27697, U.S.A.

August, 1988

Institute of Statistics Mimeo Series 1921
Biomathematics Series 33
Department of Statistics
North Carolina State University
Raleigh, NC 27695

* For correspondence and reprint requests.

ABSTRACT

Sensitivity analysis of a dynamic model developed to predict growth of the soya bean (*Glycine max* (L.) Merrill) plant was approached by two methods: calculating relative sensitivity coefficients with 10 percent perturbation of parameter values, and a response surface method. Sensitivity analysis revealed that the model is most sensitive to changes in the coefficient of root growth rate and in the net flow coefficient of carbohydrate from the soluble carbohydrate pool to the storage starch pool. Of the six parameters with large relative sensitivity coefficients, three are associated with end-product inhibition. The results indicate that in order to build physiologically realistic models of growth of the soya bean plant, an understanding of the mechanism of end product inhibition may need to be improved.

Key words: Soya bean, sensitivity analysis, response surface method,
Monte Carlo method.

INTRODUCTION

In this paper, we evaluate the sensitivity of the vegetative plant growth model reported by Lim, Wilkerson, Raper, and Gold (1988), to uncertainties in parameter values. The analysis provides increased understanding of the mechanisms that underlie system behavior, a test of the accuracy of current knowledge of these mechanisms, and determination of the priorities that might be justified in improving parameter estimates.

Methods for sensitivity analysis are generally divided into two groups (Huson, 1984). One is the calculation of sensitivity coefficients by solving a system of differential equations (Tomovic and Karplus, 1963; Burns, 1975), and the other is calculation through response surface approximation of model outputs simulated by random inputs of parameter values (Burns, 1975; O'Neill, Gardner and Mankin, 1980). The first method is impractical when the model is complex and large (Steinhorst, Hunt and Haydock, 1978). The value of the second method is dependent upon the goodness of fit of the response surface. Since the plant growth model being examined is complex and has a large number of parameters, relative sensitivity coefficients are approximated numerically to get general ideas about system sensitivity with respect to parameters. A few parameters with large values of relative sensitivity coefficients are considered more carefully by the response surface method (Box, 1954).

The model of Lim et al. (1988) includes seven state variables: shoot dry weight (W_s), root dry weight (W_r), soluble carbon concentration in the shoot (C_s), soluble carbon concentration in the roots (C_r), storage carbohydrate concentration in the shoot (S), nonstructural nitrogen concentration in the shoot (N_s), and nonstructural nitrogen concentration in the roots (N_r). The growth rate of W_s (or W_r) is assumed to be proportional to the product of relative accumulation rate of nitrogen and C_s (or C_r). The rate of change of C_s is dependent upon the carbohydrate influx through photosynthesis, efflux through respiration, utilization for shoot growth, and translocation to the storage carbohydrate pool and to the carbon pool in the roots. The rate of change of C_r depends upon the amount of carbohydrate translocated from the shoot and the amount of carbohydrate utilized for root growth and nutrient uptake. The relative accumulation rate of nitrogen is dependent on C_r and nitrate concentration in nutrient solution. Photosynthetic rate in the leaves is a function of nitrogen content in the leaves. The model was used to describe growth of the soya bean plant for 25 days beyond unfolding of the third trifoliate. Lim et al. (1988) should be consulted for details of the model.

The model uses 31 parameters. Two of the state variables, dry weight of shoot and root and total plant nitrogen content (which is a function of the state variables), have been chosen to represent system behavior. Of the 31 parameters, the 20 parameters that seem most important and have direct physical interpretation have been chosen for this study and are listed in Table 1.

RELATIVE SENSITIVITY ANALYSIS

If y is an index of system behavior, and p is a model parameter, then the sensitivity coefficient is defined as the partial derivative (Frank, 1978), $\partial y/\partial p$. When a number of indices of system behavior and a number of parameters are being considered, comparision is aided by using the relative sensitivity. If y_i is the i-th measure of system behavior, and p_j is the j-th parameter, then the relative sensitivity coefficient of y_i with respect to p_j, R_{ij}, is $(\partial y_i/y_i)/(\partial p_j/p_j)$.

In the current study, we let N_t denote total nitrogen content in the plant, and define a vector $y \in R^3$ such that $y^T=(W_s, W_r, N_t)$. We define a vector parameter, $p^T = (\psi, S_o, \ldots, m)$. Since the true value of p is unknown, the estimated value of p in Lim et al. (1988) is taken as the nominal value, and the corresponding trajectory of y is taken as the nominal trajectory. Let y_o, and p_o denote the nominal trajectory and nominal parameter vector, respectively.

The system of differential equations developed in Lim et al. (1988) can be expressed as,

$$dx/dt = \mathbf{f}(x,p), \text{ where } x \in R^7 \text{ is the state vector.} \quad (1)$$

Then, y can be expressed by a vector function, g, in three dimensions as,

$$\mathbf{y} = \mathbf{g}(\mathbf{x}). \quad (2)$$

From eqns (1), and (2), y can be expressed as,

$$\mathbf{y} = \mathbf{h}(\mathbf{x},\mathbf{p}), \text{ where } \mathbf{h}^T = (h_1, h_2, h_3). \quad (3)$$

That is, the behavior of the system, y, is expressed as a function of the parameter values, p. With this notation, $\mathbf{y}_0 = \mathbf{h}(\mathbf{x},\mathbf{p}_0)$.

Letting R_{ij} be the relative sensitivity coefficient of the i-th element of y with respect to the j-th element of p,

$$R_{ij} = (\partial y_i/y_i)/(\partial p_j/p_j)|_{p_o},\ i=1,2,3,\ j=1,2, \ldots, 20, \tag{4}$$

where y_i is the i-th element of vector $\boldsymbol{y}$, and p_j is the j-th element of vector $\boldsymbol{p}$. In this report, R_{ij} is approximated at any time t by central difference, that is,

$$R_{ij} = (h_i(\mathbf{x},\mathbf{p}_o+\Delta p_j\mathbf{e}_j) - h_i(\mathbf{x},\mathbf{p}_o-\Delta p_j\mathbf{e}_j))p_{oj}/(2\ \Delta p_j y_{oj}), \tag{5}$$

where $\boldsymbol{e}_j$ is the j-th standard unit vector in R^{20}. Let

$$R_{\cdot j} = \sum_{i=1}^{3} | R_{ij} | . \tag{6}$$

Then, $R_{\cdot j}$ represents the sensitivity of the whole model with respect to the j-th parameter. With 10 percent perturbation of parameter value, the resulting values for R_{ij} and $R_{\cdot j}$ at day 25 are listed in Table 2. R_{ij} is an approximation of the relative sensitivity coefficient under the assumptions that: 1) there is no interaction between parameters; and 2) that the response surface of the trajectory generated by perturbation of parameter values is a monotonic function. Therefore, it can give only broad ideas about system sensitivity with respect to any given parameter.

RESPONSE SURFACE METHOD

In order to determine the sensitivities more precisely, a quadratic response surface method was used to approximate the surface of variables ***y*** at the 25-th day. A Monte Carlo method was employed, with three hundred random inputs of parameter values. Let y_i be the value of the i-th variable of y on the 25-th day. A quadratic response surface model, f_i,

$$y_i = f_i(p) = b_{oi} + \mathbf{b}_i^T\mathbf{p} + \mathbf{p}^T Q_i \mathbf{p} + e_i,\ i=1,2,3, \tag{7}$$

is used to approximate the surface, where b_{oi} is a scalar constant, $\boldsymbol{b}_i$ is a constant vector in parameter space, Q_i is a constant matrix, and e_i represents white noise.

When many parameters are used as independent variables in the response surface method, the problem of multicollinearity arises; there may be at least one linear dependence among independent variables (Draper and Smith,1981; Chatterjee and Price, 1977). In this case, the regression coefficients estimated by the least squares method are very sensitive to small changes of values of independent variables, and it is meaningless to make statistical inferences based on the regression coefficients. To minimize this problem, a limited number of parameters are chosen for the response surface method, based on the magnitude of $R_{\cdot j}$. The six parameters with the largest values of $R_{\cdot j}$ are S_o, k_r, σ, ϕ_t, ϕ_d, and ϕ_n. Let ***p*** be redefined in R^6 such that,

$$\mathbf{p}^T = (\, p_1,\ p_2,\ p_3,\ p_4,\ p_5,\ p_6\,) = (\, S_o,\ k_r,\ \sigma,\ \phi_t,\ \phi_d,\ \phi_n\,).$$

The Monte Carlo method requires a joint probability density function for ***p***, which must be estimated from available information. For the variables being considered, reasonable ranges are listed in Table 3. Because most of the nominal values are located asymmetrically in the corresponding ranges, a beta distribution is used to describe the probability density for each parameter,

$$p_i \sim b(c_i, d_i), \; i=1,2, \ldots, 6, \tag{8}$$

where

$$b(c,d) = \begin{cases} \Gamma(c+d)/(\Gamma(c)\,\Gamma(d))\,(b_i - a_i)^{1-c-d}(p_i - a_i)^{c-1}(b_i - p_i)^{d-1}, & a_i \le p_i \le b_i, \\ 0 & , \text{otherwise,} \end{cases}$$

with $c > 0$, and $d > 0$. The distributions are assumed to be independent.

With this form for the probability density functions, the mode is taken to be the 'nominal value'. However, with only the modes and boundary values of $\boldsymbol{p}$, it is not possible to uniquely determine the probability density. The two parameters, c and d, are related to the mode by the following relation (Johnson and Kotz, 1970),

$$\text{mode}(p) = a + (b - a)(c - 1)/(c + d - 2). \tag{9}$$

Reasonable shapes for the distributions being considered are obtained using the value of $c=2$. The values for d are obtained from the above equation and are listed in Table 4.

In the Monte Carlo simulation, random numbers, $\boldsymbol{v}$, were generated, distributed according to the beta distribution, between 0 and 1, using the procedure of Yakowitz (1977). These values were then converted to values for parameters $\boldsymbol{p}$,

$$p_i = a_i + (b_i - a_i)\,v, \; i=1,2, \ldots, 6, \tag{10}$$

where a_i and b_i are the minimum and the maximum values of p_i listed in Table 3. The random vector $\boldsymbol{p}$ generated by the method described above is the random input to the model. The system of differential equations was solved by using the random $\boldsymbol{p}$ up to the 25-th day to give the value y. In this study, the fourth order Runge-Kutta method is used to solve the system equation. When an unrealistic solution is detected (negative values of to solve the system equation. When an unrealistic solution is detected (negative values of carbohydrate concentration in shoot or root), the parameter vector is discarded.

Since the magnitude of a regression coefficient in eqn (7) is dependent upon the units used for the independent variables, all the independent variables are standardized with respect to their means and standard deviations. Let m_i and s_i be the mean and standard deviation of three hundred randomly generated p_i. The standardized variable z_i is defined as

$$z_i = (p_i - m_i)/s_i \ , \ i=1,2,\ldots, 6. \tag{11}$$

All of the statistical analyses to be reported are based on the variable $z^T = (z_1, z_2, z_3, z_4, z_5, z_6)$. The response surface method (Box, 1954) is used to approximate responses of y with six independent variables z using the functional form

$$y_i = f_i(z) = b_{0i} + b_i^T z + z^T Q_i z + e_i \ , \ i = 1,2,3. \tag{12}$$

It should be noted that the y_i are highly correlated. It is therefore appropriate to represent them by a single canonical variable (Morrison, 1976). The first canonical variable, which will be labeled y_4, explains about 94 percent of total variance of y. The sensitivity coefficient of y_4 with respect to a given parameter may then be used to represent the system sensitivity. The response of y_4 to perturbation of parameter values is approximated by the functional form given in eqn (12). To estimate the values of parameters involved in eqn (12), the stepwise regression procedure in SAS was used, with significance level 0.05. In all cases, determination coefficients (R^2) of resulting equations are over 98 percent. Using these estimated values, sensitivity coefficients can be calculated at the standardized nominal values of parameters, z_n , where,

$$z_n^T = (-0.1377,\ -0.5085,\ 0.8135,\ -0.2211,\ -0.0593,\ -0.4541).$$

From these notations, sensitivity coefficient of the *i*-th element of *y* with respect to the *j*-th element of *p*, s_{ij} can be expressed as,

$$s_{ij} = \partial f_i(z_n)/\partial z_j\ ,\ i=1,2,3,4,\ \text{and}\ j=1,2,\ldots,6. \tag{13}$$

The calculated values of s_{ij} are listed in Table 5.

RESULTS AND DISCUSSION

According to the magnitude of relative sensitivity coefficients approximated by 10 percent perturbation of parameter values (Table 2), the model is most sensitive to six parameters. These parameters are listed in Table 6 after ordering based on the magnitude of the relative sensitivity coefficients. The parameters, ϕ_d, ϕ_n, ϕ_t, S_o, and σ are the parameters related to carbohydrate concentration in the plant, while k_r is related to root growth. These results are reasonable because plant growth is dependent upon soluble carbohydrate concentration in the plant, and also upon nitrogen uptake rate which in turns depends on the root growth. Note that the system is also moderately sensitive to changes in values of k_s, and δ, and is relatively insensitive to changes in other parameter values.

The order of sensitivities computed on the basis of the response surface (Table 5) is slightly different. This discrepancy may come from interactions among parameters and from the nonlinear response of the system to changes of parameter values. This emphasizes that it may be incorrect to order sensitivity coefficients according to the magnitudes of correlation coefficients between state variables and parameters on the basis of a relatively wide range of perturbations of the parameter value; the correlation coefficient

Carney, 1981).

Both analyses are in agreement that ϕ_d, and k_r are the parameters to which the system is most sensitive. In fact, ϕ_d is the most important parameter for controlling carbon input from the environment, and k_r is the parameter for controlling nitrogen input. It is especially important to note that of the six parameters with a large relative sensitivity coefficient, three (ϕ_d, ϕ_n, S_o) are associated with end-product inhibition. This may indicate that in order to build physiologically realistic models of soybean plant growth, the understanding of mechanisms of end product inhibition must be improved.

ACKNOWLEDGEMENT

Part of this study was supported by National Aeronautics and Space Administration grant NCC #2-101. Paper no. 11509 of the journal series of the North Carolina Agric. Res. Serv., Raleigh, NC 27695-7601, U.S.A.

LITERATURE CITED

Box, G. E. P., 1954. The exploration and exploitation of response surface: some general considerations and examples. *Biometrics* 10, 16 - 60.

Burns, J. R., 1975. Error analysis of nonlinear simulations: application to world dynamics. *IEEE Transactions on System, Man, and Cybernetics* (5), 331 - 40.

Chatterjee, S., and Price, B., 1977. *Regression Analysis by Example.* John Wiley & Sons, New York.

Draper, N. R., and Smith, H., 1981. *Applied Regression Analysis.* John Wiley & Sons, New York.

Frank, P. M., 1978. *Introduction to System Sensitivity Theory.* Academic Press, New York.

Gardner, R. H., O'Neill, R. V., Mankin, J. B., and Carney, J. H., 1981. A comparision of sensitivity analysis and error analysis based on a stream ecosystem model. *Ecological Modelling* 12, 173 - 90.

Huson, L. W., 1984. Definition and properties of a coefficient of sensitivity for mathematical models. Ibid. 21, 149 - 59.

Johnson, N. L., and Kotz, S., 1970. *Continuous Univariate Distribution - 2.* Houghton Mifflin Company. Boston.

Lim, J. T., Wilkerson, G. G., Raper, C. D. Jr., and Gold, H. J., 1988. A dynamic growth model of vegetative soya bean plants: Model structure and behaviour under varying root temperature and nitrogen concentration. *Journal of Experimental Botany* 39, (accepted for publ.).

Morrison, D. F., 1976. *Multivariate Statistical Methods.* McGraw-Hill, New York.

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

Steinhorst, R. K., Hunt, H. W., and Haydock, K. P., 1978. Sensitivity analysis of the ELM model, pp. 231 - 56, In *Grassland simulation model*, ed. G. S. Innis. Springer Verlag, New York, NY.

Tomovic, R., and W. J. Karplus, 1963. *Sensitivity Analysis of Dynamic Systems.* McGraw-Hill, New York.

Yakowitz, S. J., 1977. *Computational Probability and Simulation.* Addison-Wesley Publishing Company, New York.

Table 1. List of parameters used in sensitivity analysis.

index	symbol	description
1	ψ	Structural concentration of nitrogen in the plant
2	S_o	Threshold storage CH_2O concentration
3	ω_d	End product inhibition coefficient
4	α	Maximum photosynthetic rate
5	K_p	Michaelis-Menten constant in photosynthetic rate
6	n	Power of the Hill equation
7	K_n	Michaelis-Menten constant of nitrogen uptake
8	K_c	Michaelis-Menten constant of CH_2O uptake
9	k_s	Coefficient of shoot growth rate
10	k_r	Coefficient of root growth rate
11	η	Leaf-weight leaf-area ratio
12	G_r	Growth respiration coefficient
13	σ	Structural carbohydrate concentration
14	ε	Energy cost for nitrogen uptake
15	ϕ_t	Net translocation coefficient of CH_2O from shoot to root
16	ϕ_d	Net flow coefficient of CH_2O from soluble CH_2O pool to storage CH_2O pool
17	ϕ_n	Net flow coefficient of CH_2O from storage CH_2O pool to soluble CH_2O pool
18	ζ	Net translocation coefficient of nitrogen from root to shoot
19	δ	Maximum nitrogen uptake rate at 24 °C root temperature
20	m	Maintenance respiration rate at 24 °C root temperature

Table 2. R_{ij}, and $R_{\cdot j}$ at 25-th day

j	i = 1	2	3	$R_{\cdot j}$
1	0.0184	0.0223	0.0146	0.0552
2	2.2600	2.7692	1.6189	6.6481
3	-0.5941	-0.7253	-0.4292	1.7486
4	0.9028	1.1077	0.6440	2.6545
5	-0.2106	-0.2606	-0.1493	0.6205
6	0.2157	0.2641	0.1530	0.6328
7	-0.0648	-0.0360	-0.0662	0.1670
8	-0.3250	-0.1903	-0.3308	0.8462
9	1.5796	-1.3494	-0.7903	3.7193
10	2.5567	5.5367	2.9829	11.0762
11	0.0192	0.0240	0.0140	0.0572
12	-0.6458	-0.9877	-0.5809	2.2143
13	-1.7207	-2.6526	-1.5473	5.9206
14	-0.6951	-1.2449	-0.7296	2.6696
15	1.9086	3.5768	2.1033	7.5887
16	-4.2883	-5.2040	-3.0084	12.5007
17	2.7354	3.3573	1.9649	8.0576
18	0.0178	0.0223	0.0127	0.0528
19	1.3521	0.7545	1.3852	3.4917
20	-0.4211	-0.6670	-0.3909	1.4791

Table 3. Estimated ranges of **p** and their nominal values

parameter	range	nominal value
p_1	0.11 - 0.17	0.145
p_2	14.0 - 27.0	18.0
p_3	0.80 - 0.95	0.92
p_4	45.0 - 75.0	65.0
p_5	6.0 - 16.0	8.0
p_6	30.0 - 45.0	33.0

Table 4. Parameter values of the beta distribution function of **p**.

variable	c	d
p_1	2	1.5
p_2	2	3.25
p_3	2	1.25
p_4	2	1.5
p_5	2	5.0
p_6	2	5.0

Table 5. The values of sensitivity coefficient (s_{ij}) by the response surface method.

	j					
i	1	2	3	4	5	6
1	4.7965	7.9136	-1.9940	5.1645	-13.4524	4.0196
2	0.9192	2.5140	-0.4058	1.4322	-2.5775	0.7672
3	0.1549	0.3833	-0.0689	0.2351	-0.4230	0.1320
4	3.3597	6.1263	-1.3558	3.8733	-9.3914	2.8598

Table 6. Rank ordering of the relative sensitivity coefficients with the six parameters to which the model is most sensitive.

Rank	W_s	W_r	$\dot{W}_t$	$R_{.j}$
1	ϕ_d	k_r	ϕ_d	ϕ_d
2	ϕ_n	ϕ_d	k_r	k_r
3	k_r	ϕ_t	ϕ_t	ϕ_n
4	S_o	ϕ_n	ϕ_n	ϕ_t
5	ϕ_t	S_o	S_o	S_o
6	σ	σ	σ	σ