Predicting Soil Phosphorus Fertilizer Rate Using Hierarchical Segmented Regression Models

Xiufu Shuai,[1] Russell S. Yost,[2] and T. Jot Smyth[3]

Abstract: Predicting soil phosphorus (P) needs and P fertilizer requirements is important for plant nutrition and reducing environmental risk. The P requirement (PR) can be calculated from three components: the current status of soil P (P_0), soil P buffer coefficient (PBC), and the soil P critical level (PCL). The PBC and PCL can be predicted from soil clay content using linear-plateau models. The PR, PBC, and PCL form a hierarchical model because PR depends on PBC and PCL, which, in turn, depend on soil clay content. The objective of this study is to estimate the parameters in this hierarchical model to ensure reasonable performance and behavior of PR in a large range of soil clay contents. Results showed that the linear-plateau model described the change of PBC with soil clay content in the range of 39 to 760 g kg^{-1}. This model also described the change of PCL with soil clay content in the range of 80 to 760 g kg^{-1} for six crops, including cotton, cowpea, maize, peanut, soybean, and wheat. The obtained PR showed irregular behavior of PR within soil clay content range from 288 to 357 g kg^{-1} when PBC and PCL were independently predicted from soil clay. When the join points in the linear-plateau models of PBC and PCL were set to be equal, the irregular change of PR with soil clay content disappeared. The hierarchically modeled system predicts a decrease in PR with increasing current status of soil P and a curve-plateau trend with soil clay content.

Key words: Phosphorus model, phosphorus fertilization, PDSS, NuMaSS.

(*Soil Sci* 2011;176: 303–306)

Predicting fertilizer phosphorus (P) need has long been a challenge and continues to be so for diverse reasons. In weathered soils of the tropics, P is frequently deficient and estimates of P fertilizer need are important to ensure sustainable productivity, which is a basis for food security. Phosphorus is also increasingly important for its role in biological nitrogen fixation through increased legume productivity. In both situations, accurate estimates of P needs and requirements are important (Yost et al., 1992).

A P fertilizer model was proposed (Yost et al., 1992) based on critical levels and buffer coefficients estimated from field studies on soybean for soils ranging in soil clay percentage from 120 to 680 g kg^{-1} (Lins, 1987; Cox and Lins, 1984). The model, the phosphorus decision support system (PDSS), included two parameters: (i) P buffer coefficient (PBC) to describe the linear relationship between the fertilizer added to soils and the increase in soil-extractable P, such as M1P, M3P, and Bray-1 P (Cox, 1994); and (ii) the P critical level (PCL) to describe soil nutrient concentrations at a given percentage of the maximum yield as well as economic considerations (Cox and Lins, 1984). The PBC was known to change with soil clay content (Cox, 1994). The PCL also varied with soil clay content (Cox, 1992).

Although the relationships between PCL and clay, as well as PBC and clay, have been described for maize and soybean (Cox and Lins, 1984; Lins et al., 1985); the range in soil clay contents needs to be expanded. Phosphorus buffer coefficients were determined for the coarse-textured low-clay soils (90 g kg^{-1}) of west Africa reported by Doumbia et al. (1992). Later, the range was extended to include soils of approximately 39 g kg^{-1} clay (M. Doumbia, 2000, personal communication). Although termed "fertilizer factors," PBC have been reported by numerous other researchers, including Sharpley et al. (1984), McCollum (1991), Johnston et al. (1991), and Yost et al. (1992).

For each crop, the PCL under several soil properties was reported (Cox, 1992; Cox and Lins, 1984; Kamprath, 1978; Lins and Cox, 1989; Lins et al., 1985). However, there is no general model to describe the change of PCL with soil properties when data from different crops are pooled together.

The objectives of this study were (i) to model the change of PBC with soil properties (soil clay content); (ii) to model the change of PCL with soil properties for six crops (maize, cotton, cowpea, peanut, soybean, and wheat); and (iii) to suggest how a P fertilization model might be modified to include six crops and a range in soil properties.

MATERIALS AND METHODS

Materials

Data sets of PBC and PCL were those described in Cox (1992, 1994), Cox and Lins (1984), and Lins et al. (1985), supplemented with those from Smyth and Cravo (1991) and Doumbia (2000, personal communication). The soil orders were Oxisol, Ultisol, Alfisol, and Entisol (Yost et al., 1992), all of which were noncalcareous. The soil-extractable P was the Mehlich-3 method in units of mg P kg^{-1} (Lins and Cox, 1989; Mehlich, 1984). Soil clay was determined by dispersion with 0.1 *M* NaOH (Medina and Grohman, 1986). The determination of PCL was selected as the soil P concentration at maximum net income or 95% of the maximum yield (Lins et al., 1985). The determination of PBC was based on the increase in M3P per unit applied P at field experiments after a certain time such as 1 year (Cox, 1994).

Mathematical Modeling of PBC, PCL, and P Requirement

The PBC and PCL were described by empirical linear-plateau functions of soil clay content (x, g kg^{-1}) as follows:

$$PBC = \begin{cases} a_0 + a_1 x & 0 \le x \le \beta_1 \\ a_0 + a_1\beta_1 & \beta_1 \le x \le 1000 \end{cases} \quad (1)$$

and

$$PCL = \begin{cases} b_0 + b_1 x & 0 \le x \le \beta_2 \\ b_0 + b_1\beta_2 & \beta_2 \le x \le 1000 \end{cases} \quad (2)$$

[1]Water Resources Research Center, University of Hawaii at Manoa, Honolulu, HI 96822. Dr. Russell S. Yost is corresponding author. E-mail: rsyost@hawaii.edu
[2]3190 Maile Way, Room 102, Department of Tropical Plant and Soil Sciences, University of Hawaii at Manoa, Honolulu, HI.
[3]Soil Science Department, North Carolina State University, Raleigh, NC.
Received November 18, 2010.
Accepted for publication March 10, 2011.

ISSN: 0038-075X
DOI: 10.1097/SS.0b013e318219db24

where a_0, a_1, b_0, b_1 are coefficients, and β_1 and β_2 are join points of the linear-plateau models.

The prediction of P requirement (PR) was based on a linear relationship between external fertilizer P added and the increase in soil-extractable P:

$$PR = \frac{PCL - P_0}{PBC} \tag{3}$$

where P_0 is the initial extractable P in soil. The models 1, 2, and 3 formed a hierarchical segmented regression model, and the model of PR was assembled by inserting Eq.(1) and Eq.(2) into Eq.(3). Simple simulation of the behavior of PR showed that there was an irregularity between the join points β_1 and β_2, thus suggesting that PR cannot be simply decomposed into two independent functions of PBC and PCL.

Algorithm to Estimate Parameters in Models of PBC and PCL With or Without Constraint of the Same Join Points

The algorithm described in Shuai et al. (2003) was used to estimate the parameters in models 1 and 2 with or without the constraint $\beta_1 = \beta_2$. When the constraint $\beta_1 = \beta_2$ was not imposed, PBC and PCL were fitted separately. When the constraint $\beta_1 = \beta_2$ was imposed, data sets of PBC and PCL were combined as:

$$\mathbf{X}_{\mathrm{m}} = \begin{bmatrix} \mathrm{X}_{\mathrm{m,PBC}} \\ \mathrm{X}_{\mathrm{m,PCL}} \end{bmatrix}, \quad \mathbf{Y}_{\mathrm{m}} = \begin{bmatrix} w_1\, \mathrm{Y}_{\mathrm{m,PBC}} \\ w_2\, \mathrm{Y}_{\mathrm{m,PCL}} \end{bmatrix} \tag{4}$$

where $\mathbf{X}_{\mathrm{m,PBC}}$ and $\mathbf{X}_{\mathrm{m,PCL}}$ were the measured data sets of clay contents for PBC and PCL, respectively; $\mathbf{Y}_{\mathrm{m,PBC}}$ and $\mathbf{Y}_{\mathrm{m,PCL}}$ were the measured data sets of PBC and PCL, respectively, and w_1 and w_2 were relative weights of the two data sets. The weights were defined as:

$$w_1 = \sqrt{\frac{n_1}{SSE_1}}, \quad w_2 = \sqrt{\frac{n_2}{SSE_2}} \tag{5}$$

where n_1 and n_2 were the numbers of observations of PBC and PCL, and SSE_1 and SSE_2 were the sums of residual squares of models 1 and 2 when they were fitted without the constraint of equal join points in PBC and PCL. With the constraint of equal join points applied, the combined model was:

$$\mathrm{f}(x, \mathbf{P}) = \begin{bmatrix} w_1 \mathbf{PBC}(x, \mathbf{P}) \\ w_2 \mathbf{PCL}(x, \mathbf{P}) \end{bmatrix} \tag{6}$$

where $\mathbf{P}$ is the vector of parameters in models 1 and 2. The following cost function in a strict least squares sense is minimized by the Levenberg-Marquardt method (Seber and Wild, 1989).

$$K(\mathrm{P}) = [\mathrm{Y_m} - f(\mathrm{X_m}, \mathrm{P})]^T [\mathrm{Y_m} - f(\mathrm{X_m}, \mathrm{P})] \tag{7}$$

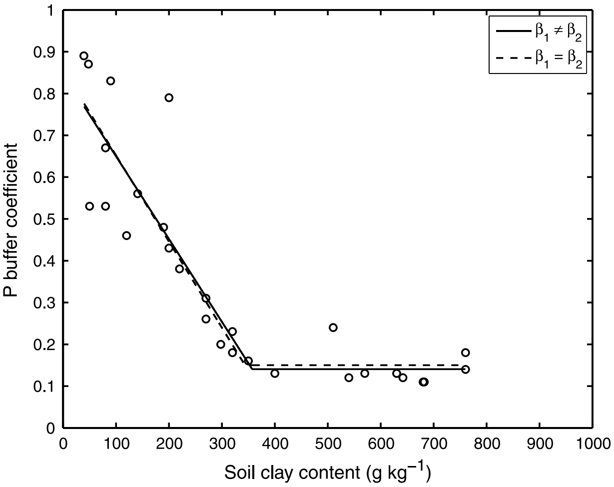

FIG. 1. Relationship between PBC and soil clay content (g kg^{-1}). Symbol circle represents measured data, solid line is linear-plateau model without the constraint of equal joint points in the models of PBC and PCL $PBC = \begin{cases} -0.00198x + 0.848 & 0 \le x \le 357.34 \\ 0.140 & 357.34 \le x \le 1000 \end{cases}$, and dashed line is linear-plateau model with constraint of equal join points in the models of PBC and PCL $PBC = \begin{cases} -0.00206x + 0.858 & 0 \le x \le 343.60 \\ 0.150 & 343.60 \le x \le 1000 \end{cases}$.

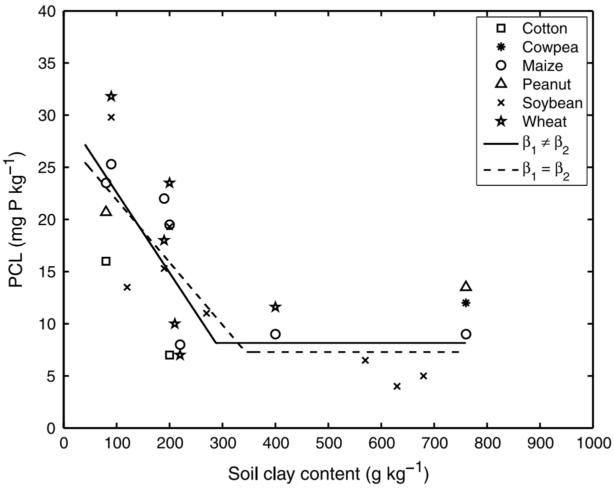

FIG. 2. Relationship between PCL and soil clay content (g kg^{-1}). Symbols represent measured data, and solid line was fitted linear-plateau model without the constraint of equal join points in the models of PBC and PCL $PCL = \begin{cases} -0.0769x + 30.277 & 0 \le x \le 287.63 \\ 8.158 & 287.63 \le x \le 1000 \end{cases}$, and dashed line was fitted linear-plateau model with constraint of equal join points in the models of PBC and PCL $PCL = \begin{cases} -0.0599x + 27.870 & 0 \le x \le 343.60 \\ 7.288 & 343.60 \le x \le 1000 \end{cases}$.

RESULTS

Behavior of PR When PBC and PCL Were Fitted Without the Constraint of Equal Join Points in PBC and PCL

The fitted curves of PBC and PCL without the constraint of equal join points are shown in Figs. 1 and 2, respectively. The sums of residual squares, SSE_1 and SSE_2, were 0.2881 and 627.8 for PBC and PCL, respectively. The plateau of model 1 is speculated to be caused by the effects of clay aggregation on PBC in highly weathered soils, with clay percentages greater than 357 g kg^{-1} (Linquist et al., 1997; Wang et al., 2000). We speculate that this effect of aggregation of the high-clay soils may also be resulting in a plateau of model 2. The join point for PBC occurred at a higher clay content than that of PCL.

The PR was predicted by inserting models 1 and 2 into 3 as:

$$PR = \begin{cases} \dfrac{-0.0769x + 30.277 - P_0}{-0.00198x + 0.848} & 0 \le x \le 287.63 \\[2ex] \dfrac{8.158 - P_0}{-0.00198x + 0.848} & 287.63 \le x \le 357.34 \\[2ex] \dfrac{8.158 - P_0}{0.140} & 357.34 \le x \le 1000 \end{cases} \tag{8}$$

The predicted PR is shown in Fig. 3. When soil clay content was a constant, PR decreased with increasing P_0 because of the increase of soil P supply. This result is consistent with the models of P fertilization rates for maize and soybean (Cox and Lins, 1984; Lins et al., 1985). When P_0 was a constant, PR fluctuated drastically within the interval of the join points of PBC and PCL, that is, soil clay contents (287.63, 357.85), which was different from an expected monotonic change with soil clay content in the P fertilization rate models for maize and soybean (Cox and Lins, 1984; Lins et al., 1985).

Behavior of PR When the Two Join Points in PBC and PCL Were Restricted to Be Equal

The weights were calculated from Equation 5, and they were 9.858 and 0.204 for PBC and PCL, respectively. The fitted curves of PBC and PCL are shown in Figs. 1 and 2, respectively. The join point of soil clay content was 343.6 g kg^{-1} under the constraint of an equal join point.

The PR was then assembled by inserting models 1 and 2 into 3 as:

$$PR = \begin{cases} \dfrac{27.870-0.05991x-P_0}{0.858-0.00206x} & 0 \le x \le 343.60 \\ \dfrac{7.288-P_0}{0.150} & 343.60 \le x \le 1000 \end{cases} \tag{9}$$

The revised predicted PR is shown in Fig. 4, and the interpretation of the curves is as follows: (i) When soil clay content was a constant, PR decreased with increasing P_0, which is consistent with the models of P fertilization rates for maize and soybean (Cox and Lins, 1984; Lins et al., 1985). (ii) When soil clay content was less than 343.6 g kg^{-1}, the curves of PR depended on P_0, that is, monotonic increase with increasing soil clay content if P_0 was less than 2.9 mg P kg^{-1}, or monotonic decrease if P_0 was greater than 2.9 mg P kg^{-1}, or constant if P_0 was equal to 2.9 mg P kg^{-1}. This effect of P_0 on PR was similar to that in regression models for maize and soybean (Cox and Lins, 1984; Lins et al., 1985). In the models for maize (Cox and Lins, 1984), the PR increased or decreased as a function of the square of soil clay content if P_0 was less or greater than 4.93 mg P kg^{-1} or kept constant if P_0 was equal to 4.93 mg P kg^{-1}. In the models for soybean (Lins et al., 1985), the PR increased or decreased as a function of the square of soil clay content if P_0 was less or greater than 3.70 mg P kg^{-1} or kept constant if P_0 was equal to 3.70 mg P kg^{-1}. (iii) When soil clay content was greater than 343.6 g kg^{-1}, PR was independent of soil clay content. This result was different from the models of P fertilization rates for maize and soybean (Cox and Lins, 1984; Lins et al., 1985), in which PR was a function of soil clay content squared in its whole range and unreasonably high PR was observed when soil clay content was high. The latter prediction occurred when P_0 and clay contents were beyond their measured values, which may happen as a result of extrapolating regression predictions beyond the measured range.

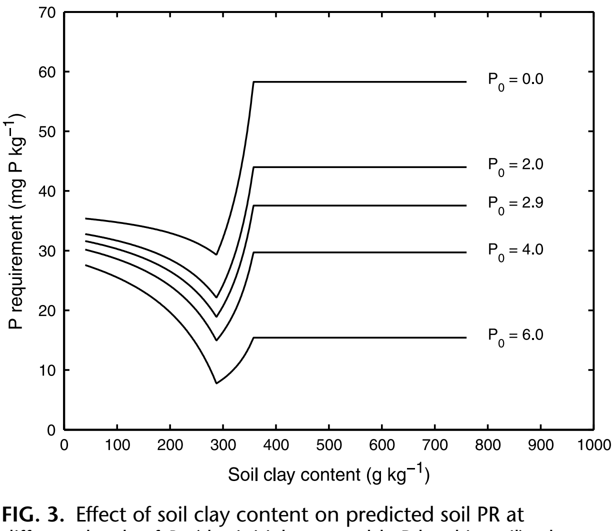

FIG. 3. Effect of soil clay content on predicted soil PR at different levels of P_0 (the initial extractable P level in soil) when PBC and PCL were estimated without the constraint of equal join points in PBC and PCL.

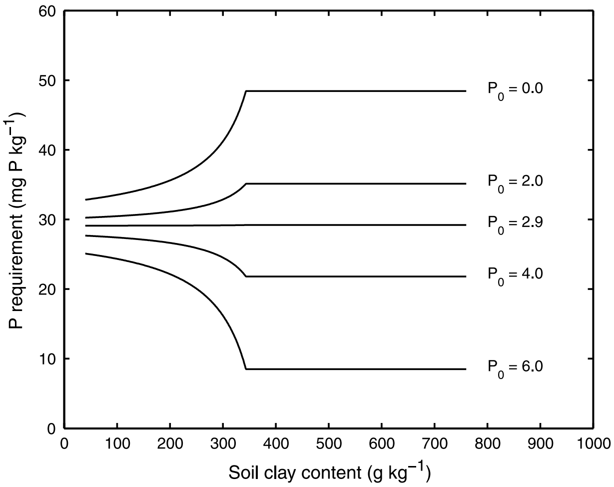

FIG. 4. Effect of soil clay content on predicted soil PR at different levels of P_0 (the initial extractable P level in soil) when PBC and PCL were fitted jointly with equal join points of the linear-plateau models.

DISCUSSION

The empirical models of P fertilization rates for maize and soybean (Cox and Lins, 1984; Lins et al., 1985) were developed based on multiple-variable polynomial regressions, that is, the dependent variable was soil PR and the independent variables were soil clay content and soil-extractable P. Usually fitting multiple polynomial regressions, the experimental designs are carefully constructed, for example, response surface or central composite designs, which specify preplanned values of the independent variables to give an unbiased regression model (Cochran and Cox, 1957). These requirements become difficult and usually impossible when locating soils to match the specified preplanned combinations characteristic of response surface designs (Cochran and Cox, 1957). Consider, for example, the difficulty in finding combinations of large ranges of soil clay contents and soil-extractable P of certain soil orders (mainly Oxisol and Ultisol) in several continents of the world.

In the hierarchical modeling approach implicit in the model proposed by Yost et al. (1992), the components PBC and PCL, comprising the lower level of the hierarchical structure, were functions of a single frequently measured variable (soil clay content), although at the upper level of the hierarchical structure, the prediction of fertilizer requirement (PR) was a two-variable function (soil clay content and soil-extractable P). This decomposition of a two-variable function into two one-variable functions reduced the complexity of requirements of the experimental design, and as a result, there was no need to seek an exceedingly difficult or impossible response surface combination of soil clay contents and soil-extractable P. To illustrate further, if the levels of soil clay contents and soil-extractable P were both integers, for example, N, then the number of soils needed for the empirical modeling approach explodes as N^2 to represent the full combination of these two variables (soil clay contents and soil-extractable P), whereas the maximum number of soils needed for the hierarchical modeling approach was only $2N$ to represent two one-variable (soil clay content) functions. Clearly, with more variables, the benefits of the hierarchical structure increase.

Because the components in the hierarchical regression model are not biased, a result of the simplified experimental design, the accuracy of the hierarchical regression model is only determined by the structure of the model. The decomposition of a two-variable function into two one-variable functions in the hierarchical modeling of soil PR, as illustrated in this study, runs a risk of oversimplification, however. The decomposition of the two-variable function into two one-variable functions ignores the possible interaction between the two component variables PBC and PCL. This interaction, in our view, resulted in the irregular behavior shown in Fig. 3. The modification of the structure of a hierarchical model such as the restriction of equal join points in the segmented regression models of PBC and PCL was required in this study to obtain the behavior consistent with expectations and experience as shown in Fig. 4. Our expectation is that a more detailed understanding of the processes of the system and the interactions between the PBC and PCL will provide a further improved hierarchical model. Nonetheless, the current result is consistent with our expectations and experience.

The hierarchical model in this study illustrates a general approach in the development of mathematical models for other complex systems. The advantage of this approach is to reduce the complexity of a system based on introducing components and a hierarchical structure to describe the relationships of the components of less complexity.

CONCLUSIONS

A PR prediction model was proposed as a hierarchical model because of its dependence on three components: the current status of soil P, the soil PCL, and soil PBC. The latter two components, in turn, depended on soil clay content. Although the same linear-plateau model described the changes in soil PCL and soil PBC with soil clay content, the assembled PR behaved irregularly within a certain range of soil clay content. An additional constraint, such as the equal join points in the linear-plateau models for soil PBC and PCL, was needed to refine the assembled model, and then a reasonable performance of PR, consistent with subject matter expectations, was obtained. This additional constraint may be generally required for future hierarchical modeling because of the nature of uncertainty in the prediction of the components from noisy measurements and/or possible interaction between the components.

REFERENCES

Cochran, W. G., and G. M. Cox. 1957. Experimental Designs, 2nd Ed. Wiley, New York, NY.

Cox, F. R. 1992. Range in soil phosphorus critical levels with time. Soil Sci. Soc. Am. J. 56:1504–1509.

Cox, F. R. 1994. Predicting increases in extractable phosphorus from fertilizing of soils of varying clay content. Soil Sci. Soc. Am. J. 58:1249–1253.

Cox, F. R., and I. D. G. Lins. 1984. A phosphorus soil test interpretation for corn grown on acid soils varying in crystalline clay content. Commun. Soil Sci. Plant Anal. 15:1481–1491.

Doumbia, M., M. Ouattara, L. R. Hossner, and A. B. Onken. 1992. Phosphorus status and sorption in sandy soils of the Sahel region of west Africa. TropSoils Bulletin No. 92-01, 1992. Proceedings of the TropSoils Phosphorus Decision Support Workshop, College Station, Texas, March 11–12, 1992. Department of Agronomy and Soil Science, University of Hawaii at Manoa, Honolulu, HI, pp. 45–58.

Johnston, M. A., N. Miles, and G. R. Thibaud. 1991. Quantities of phosphorus required to raise the soil test value. S. Afr. J. Plant Soil 8:17–21.

Kamprath, E. J. 1978. The role of soil chemistry in the diagnosis of nutrient disorders in tropical situations. *In*: Mineral Nutrition of Legumes in Tropical and Subtropical Soils. C. S. Andrew, and E. J. Kamprath (eds.). CSIRO, Victoria, Australia, pp. 313–327.

Linquist, B., P. W. Singleton, R. S. Yost, and K. G. Cassman. 1997. Aggregate size effects on the sorption and release of phosphorus in an Ultisol. Soil Sci. Soc. Am. J. 61:160–166.

Lins, I. D. G., and F. R. Cox. 1989. Effect of extractant and selected soil properties on predicting the correct phosphorus fertilization rate of soybean. Soil Sci. Soc. Am. J. 53:813–816.

Lins, I. D. G., F. R. Cox, and J. J. Nicholaides III. 1985. Optimizing phosphorus fertilization rates for soybean grown on Oxisols and associated Entisols. Soil Sci. Soc. Am. J. 49:1457–1460.

Lins, I. D. G. 1987. Improvement in Soil Test Interpretations for Phosphorus and Zinc. Dissertation, North Carolina State University, Department of Soil Science, Raleigh, NC, UMI No. 8717286.

McCollum, R. E. 1991. Buildup and decline in soil phosphorus: 30-year trends on a Typic Umbraquult. Agron. J. 83:77–85.

Medina, H. P., and F. Grohman. 1986. Disponilibilidade de água em algumas solos sob Cerrado [in Portuguese]. Bragantia 25:65–75.

Mehlich, A. 1984. Mehlich 3 soil test extractant: A modification of the Mehlich 2 extractant. Commun. Soil Sci. Plant Anal. 15:1409–1416.

Seber, G. A. F., and C. J. Wild. 1989. Nonlinear Regression. Wiley, New York, NY.

Sharpley, A. N., C. A. Jones, C. Gray, and C. V. Cole. 1984. A simplified soil and plant phosphorus model. II. Prediction of labile, organic, and sorbed phosphorus. Soil Sci. Soc. Am. J. 48:805–809.

Shuai, X., Z. Zhou, and R. S. Yost. 2003. Using segmented regression models to fit soil nutrient and soybean grain yield changes due to liming. J. Agric. Biol. Environ. Stat. 8:240–252.

Smyth, T. J., and M. S. Cravo. 1991. Lime and P requirements for peanut in a clayey Oxisol. *In*: TropSoils Technical Report, 1988–1989. T. M. McBride (ed.). Soil Management Collaborative Research Support Program, North Carolina State University, Raleigh, NC.

Wang, X., J. M. Jackman, R. S. Yost, and B. A. Linquist. 2000. Predicting soil phosphorus buffer coefficients using potential sorption site density and soil aggregation. Soil Sci. Soc. Am. J. 64:240–246.

Yost, R. S., A. B. Onken, F. Cox, and S. Reid. 1992. The diagnosis of phosphorus deficiency and predicting phosphorus requirement. *In*: TropSoils Bulletin 92-01, 1992. Proceedings of the Tropical Phosphorus Decision Support System Workshop, College Station, Texas, March 11–12, 1992. S. Balas (ed.). Department of Agronomy and Soil Science, College of Tropical Agriculture and Human Resources, University of Hawaii, Honolulu, HI.