Communications in Soil Science and Plant Analysis, 43:2535–2543, 2012

ISSN: 0010-3624 print / 1532-2416 online
DOI: 10.1080/00103624.2012.711875

Taylor & Francis
Taylor & Francis Group

Interlaboratory and Intralaboratory Testing of Soil Sulfate Analysis in Mollisols of the Pampas

DANIELA RUSSI, FLAVIO H. GUTIERREZ BOEM, PABLO PRYSTUPA, AND GERARDO RUBIO

Soil Fertility and Fertilizers, College of Agronomy, University of Buenos Aires Institute of Agricultural and Environmental Biosciences, National Scientific and Technical Research Council, Buenos Aires, Argentina

Sulfur (S) deficiencies in grain and forage crops have been detected in many agricultural regions of the world, but soil tests are not commonly used as the basis for S fertilizer recommendation programs. Errors of measurements of soil sulfate were determined to assess whether the variation among and within soil-testing laboratories could be a factor that prevent the adoption of soil testing to assess soil sulfate availability. Subsamples of 10 selected soils (Mollisols) from the Pampas (Argentina) were sent in two batches to five soil-testing laboratories. Laboratories were unaware of the existence of subsamples and performed routine sulfate analysis as if these soils came from 60 different fields. Soil sulfate ranged from 3.3 to 20.6 mg kg^{-1}. One laboratory reported sulfate values greater than the other ones, having a mean bias of 4.1 mg kg^{-1} S sulfate (SO_4). The other four laboratories reported similar sulfate values when soils had low sulfate availability (less than 10 mg S kg^{-1}), even when they used different extractants. Considering only these four laboratories, average interlaboratory coefficients of variations ranged from 6 to 24% for the 10 soils. Within-laboratory mean coefficients of variation (CVs) ranged from 12 to 22%. However, mean absolute errors of all laboratories were less than 1.2 mg kg^{-1} S-SO_4. Two laboratories reported different sulfate values for the two batches of shipment (an average difference of 4.7 and 3.8 mg kg^{-1} of S-SO_4). Laboratories using different extractants obtained similar results, suggesting that using the same extractant is not a prerequisite to standardize laboratory results in these soils. Differences between laboratories in our study were smaller than in other interlaboratory comparisons for soil sulfate. These differences could be easily detected and corrected if laboratories participate in an interlaboratory control system. The observed low mean absolute errors suggested that, in general, all laboratories achieve acceptable precision when evaluating within the same batch of determinations. Differences between batches of shipment (within laboratory error) stressed the importance of using reference material for internal quality control.

Keywords Soil fertility, sulfur, testing methodologies

Introduction

Sulfur (S) deficiencies in grain and forage crops have been detected in many agricultural regions of the world (Bansal, Sharma, and Singh 1979; Gutierrez Boem, Prystupa, and Ferraris 2007; Haneklaus, Bloem, and Schnug 2008; Islam and Ponnamperuma 1982).

Received 7 January 2011; accepted 14 March 2012.
Address correspondence to Flavio H. Gutierrez Boem, Av. San Martín 4453, C1417DSE Ciudad de Buenos Aires, Buenos Aires, Argentina. E-mail: gutierre@agro.uba.ar

However, a soil test is not commonly used as the basis for S fertilizer recommendation programs. The lack of consistency between soil test and crop response that has been observed in several regions may have prevented the adoption of a soil test for S recommendations (Mascagni, Harrison, and Padgett 2008; Rehm and Clapp 2008; Scherer 2001; Schnug and Haneklaus 1998; Vaughn, Jones, and Center 1987). Among the possible causes of this lack of consistency, variability in soil-test results from within and among laboratories is a major concern. Soil-testing laboratories use different solutions for the extraction of soil sulfate and different techniques to determine the concentration of sulfate in the extract (Crosland, Zhao, and McGrath 2001). In the Pampas (the main agricultural region of Argentina), all soil-testing laboratories determine sulfate in soil extracts by turbidimetry, while the extractant varies among laboratories. The most common extractants are calcium phosphate, ammonium acetate, and potassium phosphate. Several studies showed that the critical level of soil sulfate for crops and pastures varied with the extracting solution (Bansal, Montiramani, and Pal 1983; Blair et al. 1991; Islam and Ponnamperuma 1982). Different extracting solutions evaluate different forms of available sulfate such as sulfate in solution, adsorbed sulfate, and part of the organic S that mineralizes during crop season (Anderson et al. 1992; Blair et al. 1991; Watkinson and Kear 1996). Sulfate determination by turbidimetry is simple but has several problems that reduce its precision and accuracy such as interference from organic matter, color of the extract, and stability of the suspension (Ajwa and Tabatabai 1993; Anderson et al. 1992).

Accuracy and precision are the parameters used to test the quality of measurements obtained with different laboratory techniques of soil analysis (Ajwa and Tabatabai 1993). Accuracy is the closeness of the results to a true or expected value and is affected by random and systematic errors. In laboratory comparisons, the true value is defined as the mean plus a confidence interval obtained from laboratories without discrepancies. Precision refers to the reproducibility and repeatability of the results obtained by each laboratory and is affected by random errors. Precision is evaluated through the variance or the variation coefficients obtained after analyzing the same sample several times. The precision measured in quick succession, in which a single operator uses the same reagents, soil batches, and materials (precision within run) is called *repeatability*. If the analyses are made by different operators or sample batches, the precision between-run replicates can be measured, which is the *reproducibility* (Miller and Miller 2005).

The aim of this study was to assess whether the variation among and within laboratories could be a factor that prevents the use of soil testing to assess soil sulfate availability. Our approach was to determine the errors of measurements of soil sulfate by the major laboratories of the Pampas region. Because soil-testing laboratories use different extractants, methods of determination of sulfate in the extract are not precise, and determination of sulfate is not yet included in any interlaboratory control system, our working hypothesis was that soil sulfate results would vary widely among laboratories.

Materials and Methods

Thirty soils were sampled within the Rolling Pampa of Argentina. Soil samples were collected from the top 20 cm of the soil profile, air dried, sieved to 2 mm, and homogenized using an open bin riffle splitter (Schumacher et al. 1990). Ten soils were selected that had ranges of sulfate levels that included the critical levels for most extractants reported in the literature (Dick, Kost, and Chen 2008). All soils were noncalcareous, loess-derived Argiudolls or Hapludolls (Soil Survey Staff 1999). Thirty subsamples from each soil were prepared with a riffle splitter. A first batch of three subsamples of each soil was sent to

the five most important soil-testing laboratories of the Pampas region. A second batch of three subsamples of each soil was sent 6 months later to the same laboratories. During this 6-month period, air-dried soil samples were stored at room temperature. Laboratories involved in this study were not informed, and soil subsamples were numbered at random as if they came from 60 different fields. Therefore, each laboratory analyzed the 60 samples received (two batches of 30) as samples from different soils. All laboratories performed the measurement of S sulfate (SO_4^{2-}) by turbidimetry, using the following extractants: ammonium acetate with acetic acid (laboratories 1 and 2), monocalcium phosphate (laboratories 3 and 5), and monopotassium phosphate (laboratory 4).

Differences among laboratories in the sulfate value reported for soils were evaluated by regression analysis, comparing the results from each laboratory to each other. A line representing the relationship between the amounts of sulfates measured by two different laboratories in the 10 soils was fitted using the standardized major-axis method. This method has been identified as more appropriate than linear regression when the objective is to describe the relationship between two variables measured with error or to test whether two measurement methods are in agreement (Warton et al. 2006; Webster 1997). The permutation of variables Y and X does not affect the line fitted by this method (i.e., fitting Y vs. X or X vs. Y yields the same line). Tests of whether slopes were different from 1 and intercepts were different from 0 were performed. The software SMATR was use for fitting the lines and testing the significance of their parameters (Warton et al. 2006).

Precision of sulfate determination within each laboratory was evaluated, calculating the variance, standard deviation, coefficient of variation [CV (%)], and mean absolute error (MAE) of sulfate values reported for each soil. The MAE was calculated as the average of the absolute values of differences between the measured value in each subsample and the mean of the three subsamples considered (Willmot and Matsuura 2005). Variances were compared using an F test.

The batch effect on the accuracy of the measurements was evaluated using generalized linear models adjusting a gamma function with identity link (Lindsey 1997). Additionally, a t test was performed to evaluate the significance of the batch effect within each laboratory.

Results and Discussion

Accuracy of Measurements

Comparisons of sulfate values reported by each laboratory for the 10 soils are shown in Figure 1. Table 1 shows the parameters of fitted functions and its significance. Mean sulfate values reported by laboratories 2, 3, and 5 for each soil were not different, as fitted functions to compare their results were not different from the 1:1 line.

Sulfate values reported by laboratory 1 were greater than those of other laboratories in 39 out of 40 pairs of soil, having a mean bias of 4.1 mg kg^{-1} S-$SO_4^=$ (Figure 1). When comparing laboratory 1 with laboratories 4 and 5, the intercept was different from 0 ($P < 0.05$) and with laboratory 3 the intercept tended to be different from 0 ($P < 0.10$). The slope was not significantly different from 1 (except in comparison with the laboratory 4). These results suggested that laboratory 1 presented a systematic error in the sulfate determination, as the bias of laboratory 1 was independent of the measured value. This type of error is relatively easy to detect and correct in an interlaboratory control system because it does not depend on the sulfate concentration of soil.

When laboratory 4 was compared with its counterparts, slopes significantly different from 1, with a mean value of 1.45, were found (Figure 1). The intercept ranged between

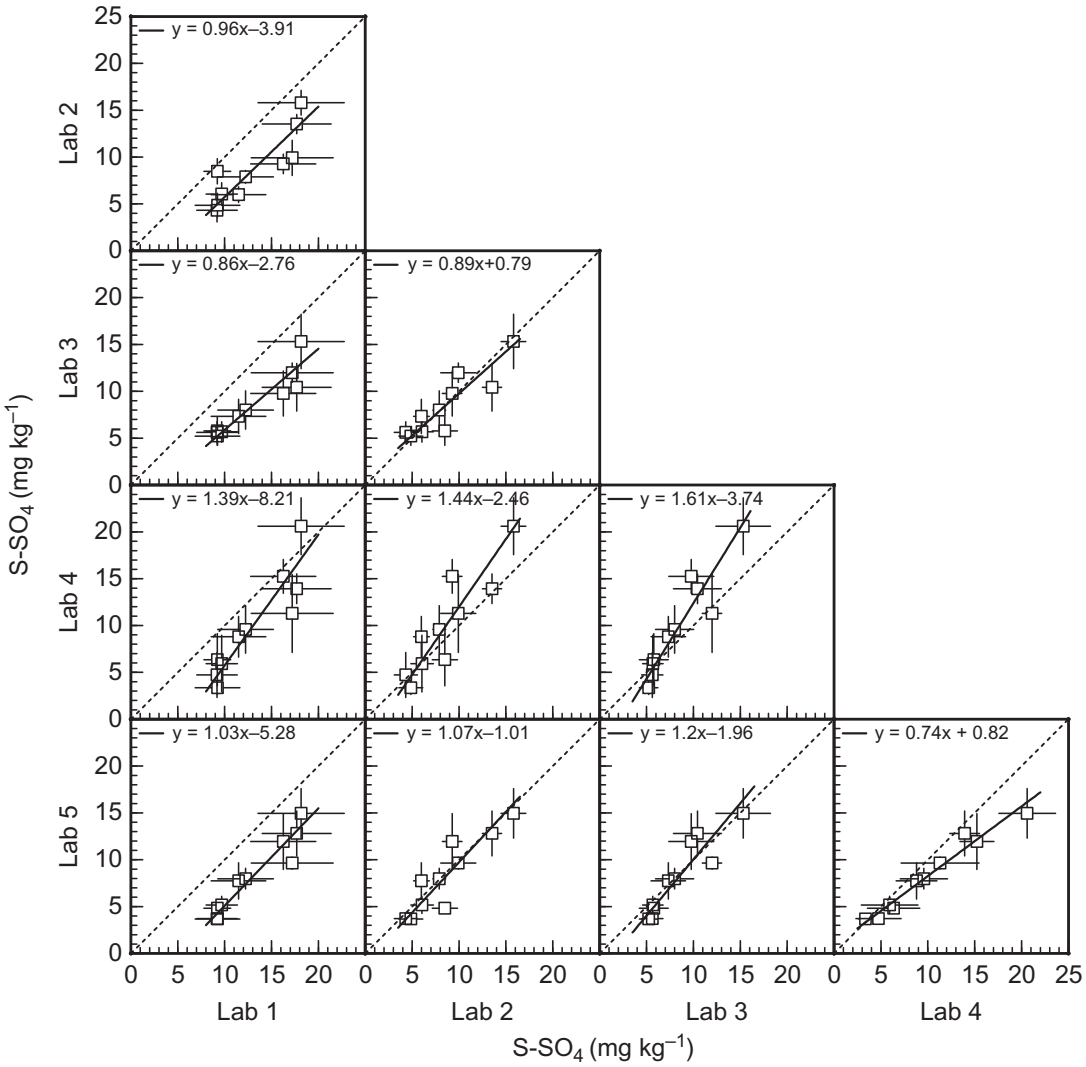

Figure 1. Comparisons of sulfate values reported by each laboratory for the 10 soils analyzed. Each point is the mean (from six subsamples) of each soil; bars indicate standard deviation of the mean. The dotted line is the function y = x, and the solid line the fitted line by SMA for the two laboratories.

Table 1
Comparison of sulfate measurements among laboratories: parameters (intercept and slope) of the fitted functions and *p* values of the tests

Laboratories (Y vs. X) (Y = a + b X)	Intercept (a)	*p* value H_0: a = 0	Slope (b)	*p* value H_0: b = 1
2 vs. 1	−3.919	0.123	0.9653	0.847
3 vs. 1	−2.767	0.095	0.8656	0.281
3 vs. 2	0.792	0.556	0.8967	0.512
4 vs. 1	−8.210	0.017	1.396	0.047
4 vs. 2	−2.468	0.260	1.447	0.038
4 vs. 3	−3.747	0.073	1.613	0.004
5 vs. 1	−5.286	0.008	1.039	0.728
5 vs. 2	−1.014	0.518	1.076	0.637
5 vs. 3	−1.965	0.221	1.200	0.215
5 vs. 4	0.823	0.185	0.7441	0.002

−8.21 (L4 vs. L1) and −1.10 (L4 vs. L5). Therefore, laboratory 4 tended to overestimate sulfate values, with increasing differences as soil sulfate concentration increased. As this error was associated with the soil sulfate concentration, is more difficult to detect in an interlaboratory control system, as it should include samples with a range of sulfate content. However, for soils with low sulfate availability (values less than 10 mg kg^{-1}), their measurements were similar to those obtained by laboratories 2, 3, and 5, even when laboratory 4 used potassium phosphate, a different extractant than the others laboratories. Laboratory 2 measurements coincide with those of laboratory 3 and laboratory 5, even when laboratory 2 uses a different extractant (ammonium acetate used by laboratory 2 and monocalcium phosphate used by laboratories 3 and 5). These results suggest that in these soils, employed extractants would not affect measured sulfate concentration. Accordingly, Crosland, Zhao, and McGrath (2001) argued that differences in reported values by different laboratories were the result of the extraction procedure (soil/extractant ratio, extract treatment, etc.) rather than a direct effect of the extractant.

Differences among laboratories in our study were smaller than in other interlaboratory comparisons for soil sulfate. In an interlaboratory comparison with two soils and 10 laboratories, results ranged from 2.5 to 8.7 mg kg^{-1} in one soil and from 3.0 to 15.2 mg kg^{-1} in the other one (Crosland, Zhao, and McGrath 2001). In our study, the soil with the greatest difference ranged from 9.6 to 17.2 mg kg^{-1}. When results from laboratory 1 were excluded from the analysis, interlaboratory CVs ranged from 6 to 24%, similar to those reported for common soil test for phosphorus (from 10 to 22%, Kleinman et al. 2001).

Precision

Comparison of variances, CV, and mean absolute error (MAE) revealed significant differences in repeatability among laboratories (Figure 2). Laboratory 2 was the most precise and laboratory 3 was least able to repeat the same value for the same soil sample. The other laboratories showed intermediate performance. Variances ranged from 0.2 to 3.6, values within the range of variance observed by Pandey and Girish (2007) for the standard turbidimetric technique (0.7 to 10.7) and a modified turbidimetric method proposed by these authors (0.3 to 4.7).

When the precision was evaluated by means of the coefficients of variation, the ranking of laboratories changed for laboratory 1 because of its positive bias for sulfate determination. Mean sulfate value of laboratory 1 (13 mg kg^{-1}) was greater than the overall mean of all laboratories (9.8 mg kg^{-1}). The mean CV of the five laboratories was 16% and ranged from 12.3% to 22.0%. These values were within the range of CVs reported by Ajwa and Tabatabai (1993) for turbidimetric measurements of sulfate in soils (from 10.9% to 28.5%), but greater than those reported for a common soil test for phosphorus (from 2.8% to 10.6%, Wolf and Baker 1985). However, MAEs of all laboratories were less than 1.2 mg kg^{-1} $S\text{-}SO_4$, indicating that, in general, all laboratories had demonstrated acceptable precision within the same batch of determinations.

Variation between Batches of Shipment

The analysis of the effect of two temporally displaced sample sets on laboratory measurements showed a significant batch × laboratory interaction ($P < 0.01$). Differences between batches were significant only for laboratories 1 and 4 ($P < 0.01$)(Figure 3a). The differences between batches, average of 10 soils, were 4.7 and 3.8 mg kg^{-1} of $S\text{-}SO_4$ for laboratories 1 and 4, respectively. These results showed that the systematic error of laboratory 1 was mainly due to its determinations in the second batch of shipment. This variation

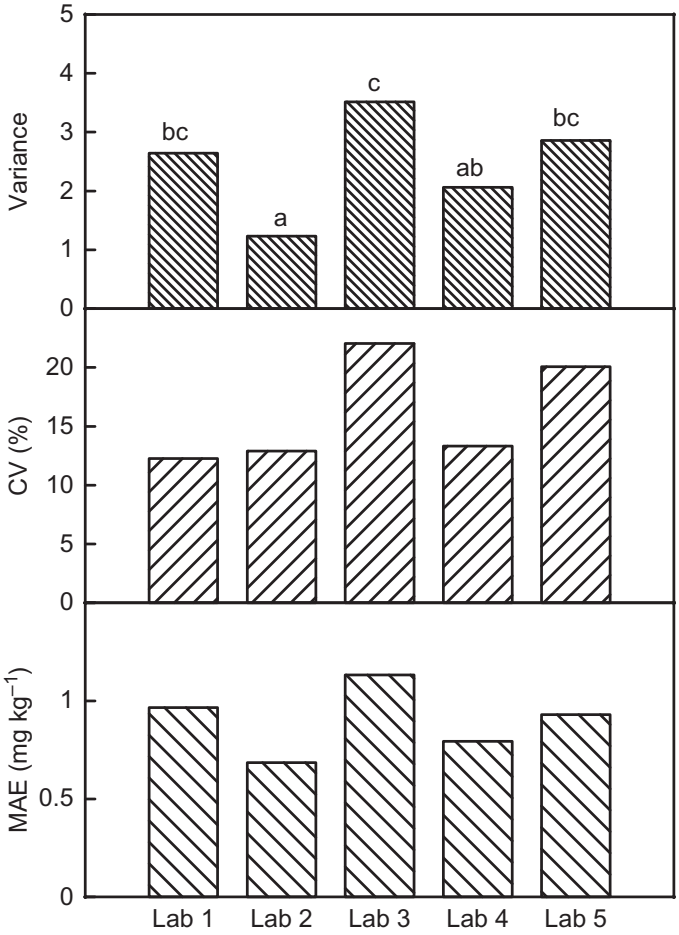

Figure 2. Variances, coefficients of variation (CV), and mean absolute errors (MAE) for each laboratory. Different letters denote significant differences between variances (F test, $\alpha = 0.05$).

in values obtained between two sample sets (i.e., low reproducibility) could be prevented by using reference material for internal quality control (Miller and Miller 2005). In this case, it also reveals the importance of retesting the same samples spaced out in time for interlaboratory control systems.

Figures 3b and 3c show the measurement variability of the two batches of shipment for each laboratory. Laboratories 4 and 5 had a different precision depending on batch of shipment, while the precision of the other laboratories was not affected by batch of shipment. Laboratories 4 and 5 significantly reduced the variance of the second sample set when compared to the first one (Figure 3b).

Conclusions

The analysis of accuracy indicated that four of the five laboratories reported similar sulfate values when soils had low sulfate availability (less than 10 mg S kg^{-1}). One laboratory (laboratory 4) tended to determine sulfate values greater than the other laboratories when extractable soil sulfate concentration exceeded 10 mg S kg^{-1}, and another laboratory (laboratory 1) had a positive bias regardless of the measured value. An interesting finding of this work was that laboratories using different extractants obtained similar results, suggesting that using the same extractant is not a prerequisite to standardizing laboratory results in these soils. Differences between laboratories in our study were smaller than in other

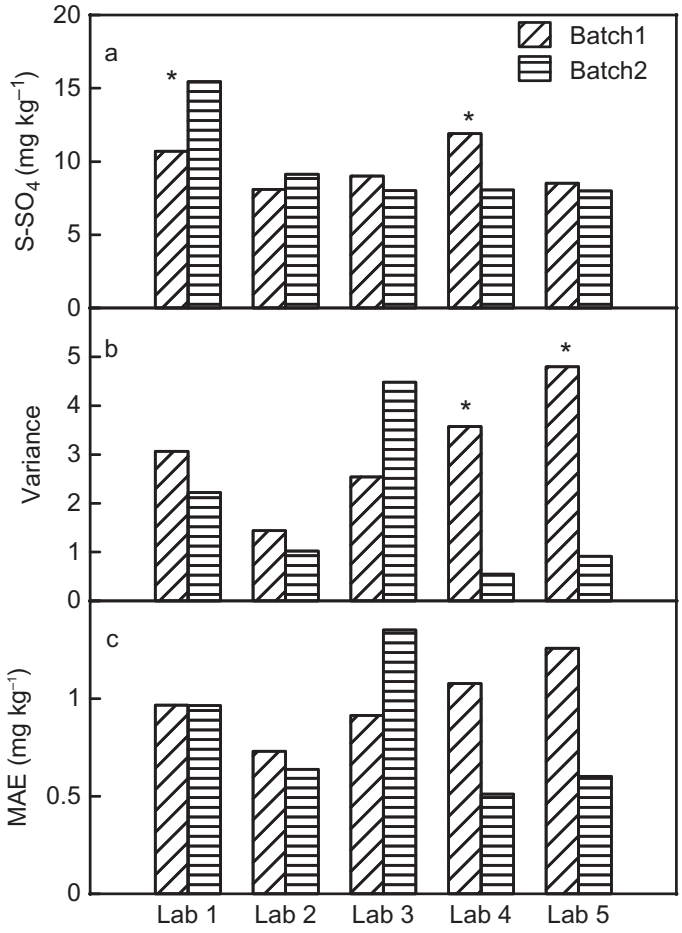

Figure 3. Effect of batch of analysis on (a) mean sulfate values (from three subsamples of each of the 10 soils), (b) variances of sulfate measurements, and (c) mean absolute errors. Asterisks indicates significant differences of sulfate values (t test) or variances (F test) between the two batches ($\alpha = 0.05$).

interlaboratory comparisons. These differences could be easily detected and corrected if laboratories were to participate in an interlaboratory control system.

The precision of sulfate determination was similar to those reported in previous studies. The observed low mean absolute errors suggested that a lack of precision in the analysis would not be a determining factor for not using this soil test as part of a fertilizer recommendation program. If in subsequent field experiments, crop response to S fertilization is found to be related to soil sulfate concentration, then the soil test does demonstrate both accuracy and precision and could be used as a routine soil-testing tool in soils similar to those included in this study.

Our results indicate that commercial laboratories measuring S should participate in interlaboratory assays. These interlaboratory assays should have to be conducted on at least two batches of soil samples covering a range of extractable sulfate concentrations. Thus, errors associated with the magnitude of the measured value could be detected.

References

Ajwa, H. A., and M. A. Tabatabai. 1993. Comparison of some methods for determination of sulphate in soils. *Communications in Soil Science and Plant Analysis* 24:1817–1832.

Anderson, G., R. Lefroy, N. Chinoim, and G. J. Blair. 1992. Soil sulphur testing. *Sulphur in Agriculture* 16:6–14.

Bansal, K. N., D. P. Montiramani, and A. R. Pal. 1983. Studies on sulphur in vertisols, I: Soil and plant test for diagnosing sulphur deficiency in soybean [*Glycine max* (L.) Merr.]. *Plant and Soil* 70:133–140.

Bansal, K. N., D. N. Sharma, and D. Singh. 1979. Evaluation of some soil test methods for measuring available sulphur in alluvial soils of Madhya Pradesh. *Journal of the Indian Society of Soil Science* 27:308–313.

Blair, G., N. Chinoim, R. Lefroy, G. Anderson, and G. J. Crocker. 1991. A soil sulfur test for pastures and crops. *Australian Journal of Soil Research* 29:619–626.

Crosland, A. R., F. J. Zhao, and S. McGrath. 2001. Interlaboratory comparison of sulphur and nitrogen analysis in plants and soils. *Communications in Soil Science and Plant Analysis* 32:685–695.

Dick, W. A., D. Kost, and L. Chen. 2008. Availability of sulfur to crops from soil and other sources. In *Sulfur: A missing link between soils, crops, and nutrition*, ed. J. Jez, 59–82. Madison, Wisc.: ASA, CSSA, and SSSA.

Gutiérrez Boem, F. H., P. Prystupa, and G. Ferraris. 2007. Seed number and yield determination in sulfur-deficient soybean crops. *Journal of Plant Nutrition* 30:93–104.

Haneklaus, S., E. Bloem, and E. Schnug. 2008. History of sulfur deficiency in crops. In *Sulfur: A missing link between soils, crops, and nutrition*, ed. J. Jez, 45–58. Madison, Wisc.: ASA, CSSA, and SSSA.

Islam, M. M., and F. M. Ponnamperuma. 1982. Soil and plant tests for available sulfur in wetland rice soils. *Plant and Soil* 68:97–113.

Kleinman, P. J. A., A. N Sharpley, K. Gartley, W. M. Jarrell, S. Kuo, R. G. Menon, R. Myers, K. R. Reddy, and E. O. Skogley. 2001. Interlaboratory comparison of soil phosphorus extracted by various soil test methods. *Communications in Soil Science and Plant Analysis* 32: 2325–2345.

Lindsey, J. K. 1997. *Applying generalized linear models*. New York: Springer.

Mascagni, H. J., S. A. Harrison, and G. B. Padgett. 2008. Influence of sulfur fertility on wheat yield performance on alluvial and upland soils. *Communications in Soil Science and Plant Analysis* 39:2133–2145.

Miller, J. N., and J. C. Miller. 2005. *Statistics and chemometrics for analytical chemistry*. Harlow, UK: Pearson Prentice Hall.

Pandey, R. N., and B. H. Girish. 2007. An improved turbidimetric method for the estimation of sulphur in soil extracts. *Journal of the Indian Society of Soil Science* 55:73–79.

Rehm, G. W., and J. G. Clapp. 2008. Sulfur in a fertilizer program for corn. In *Sulfur: A missing link between soils, crops, and nutrition*, ed. J. Jez, 143–152. Madison, Wisc.: ASA, CSSA, and SSSA.

Scherer, H. W. 2001. Sulphur in crop production. *European Journal of Agronomy* 14:81–111.

Schnug, E., and S. Haneklaus. 1998. Diagnosis of sulphur nutrition. In *Sulphur in agroecosystems*, ed. E. Schnug, 1–38. Dordrecht, the Netherlands: Kluwer Academic.

Schumacher, B. A., K. C. Shines, J. V. Burton, and M. L. Papp. 1990. Comparison of three methods for soil homogenization. *Soil Science Society of American Journal* 54:1187–1190.

Soil Survey Staff. 1999. *Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys* (USDA Agricultural Handbook 436). Washington, D.C.: U.S. Government Printing Office.

Vaughn, C. E., M. B. Jones, and D. M. Center. 1987. Sulfur test on northern California subclover-annual pasture surface soils. *Soil Science* 143:184–191.

Warton, D. I., I. J. Wright, D. S. Falster, and M. Westoby. 2006. Bivariate line-fitting methods for allometry. *Biological Reviews* 81:259–291.

Watkinson, J. H., and M. J. Kear. 1996. Sulfate and mineralizable organic sulfur in pastoral soils in New Zealand, II: A soil test for mineralizable organic sulfur. *Australian Journal of Soil Research* 34:405–412.

Webster, R. 1997. Regression and functional relations. *European Journal of Soil Science* 48:557–566.

Willmot, C. J., and K. Matsuura. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research* 30:79–82.

Wolf, A. M., and D. E. Baker. 1985. Comparison of soil test phosphorus by Olsen, Bray P1, Mehlich I, and Mehlich III methods. *Communications in Soil Science and Plant Analysis* 16:467–484.