Reprinted from *Agronomy Journal*
Vol. 86, No. 1

Higher-Order Analysis of Nutrient Accumulation Data

E. John Sadler* and Douglas L. Karlen

ABSTRACT

Biomass and nutrient accumulation data have often been obtained to determine rates of nutrient uptake. Traditionally calculated as the difference in accumulation divided by elapsed time, rate values thus obtained are slopes of linear interpolations between points on the accumulation curve. That implies an assumption of constant uptake rate during the observation period. Our objective was to illustrate a higher-order interpolant that is not subject to such assumptions. With it, one obtains smooth curves consistent with the assumption that daily uptake rates are somewhat related. The abrupt changes in rates determined with linear interpolation are consistent with daily rates that are unrelated. Analyses of historical and recent data showed that additional information may be obtained from higher-order analysis methods. Cubic interpolation methods were applied to the accumulation curve to obtain continuous, smooth nutrient uptake curves. The programs used are described, and two sample data sets of corn (*Zea mays* L.) growth and N accumulation illustrate the strengths, weaknesses, and inherent assumptions of this analytical technique. In general, this technique can be used if the objective is to analyze intraseasonal variation in growth or uptake rates determined from sparse data.

E.J. Sadler, USDA-ARS, Coastal Plains Soil & Water Conservation Res. Ctr. (CPSWCRC), P.O. Box 3039, Florence, SC 29502; D.L. Karlen, USDA-ARS, National Soil Tilth Laboratory (NSTL), 2150 Pammel Dr., Ames, IA, 50011. Joint contribution of the USDA-ARS CPSWCRC, in cooperation with the South Carolina Agric. Exp. Stn., Clemson, SC 29634 and the NSTL. Received 3 Mar. 1993. *Corresponding author.

Published in Agron. J. 86:26–31 (1994).

NUTRIENT REQUIREMENTS for crops are often inferred from amounts of nutrients accumulated by crops. Typically, such data sets consist of time series of concentrations and of corresponding aboveground dry matter amounts. The product of these two series results in a time series of aerial nutrient accumulation. Traditionally, average rates of accumulation between sampling dates have been calculated as the difference in accumulated nutrient divided by elapsed time (e.g., Henderson and Kamprath, 1970). The resulting series of estimates of accumulation rate is strictly applicable only to the midpoint of the sampling interval. Estimates of accumulation rate at any other time require further interpolation. Numerically, this procedure is equivalent to evaluating a linear interpolation of the accumulated curve at the midpoint of each interval. Evaluating the original interpolant at other points results in biases caused by the time shift from the midpoints, since the derived rate is constant during each interval.

General methods of interpolation and smoothing data date from the 1800s. For various reasons, researchers often smooth through variation in time series data rather than analyzing it. Typically, a least-square fit of a polynomial curve has been used for this purpose (e.g., Clawson et al., 1986). When one must detect seasonal trends underlying variation, smoothing through noisy data may be appropriate. However, if variation is real (and quantitatively important), then smoothing through it does two undesirable things. First, it hides the var-

iation itself, which reduces the chance of important information being acted upon; second, estimates are biased toward the running average value.

If one assumes that nutrient accumulation rates from day to day are related because of similarities in the crop over short periods, then one can estimate intraseasonal variation in both accumulation and accumulation rate using higher-order polynomial interpolation methods. Many techniques to analyze time series exist, each with its own limitations and strengths. Some, such as the fast Fourier transform (Kimball, 1974), require regularly spaced data, which is not generally the case with nutrient sampling. Others, such as convolution methods (Gorry, 1990), are designed to provide smoothed values or derivatives only at the time of sampling, which may not be useful for nutrient accumulation studies. Eliminating these techniques on the basis of data requirements leaves two classes of techniques that suit the general purpose of analyzing nutrient accumulation data, for which a differentiable expression is required over the entire range of data.

The first class of methods provides some smoothing of the data. This is desirable for noisy data, especially when the data sets are large. The least-square fitting of a single equation (linear regression, higher-order polynomial regression, exponential regression, etc.) belongs to this class. These are especially useful for simple patterns. Some patterns are described well by theoretical expressions (e.g., the Michaelis–Menton equation), which are also in this class. In general, variation in nutrient accumulation patterns occurs for reasons not yet fully understood, so that a general-purpose fitting method is required. Several published methods fit this description, including quadratic splines (DuChateau et al., 1972), cubic splines (Kimball, 1976), and sliding polynomials (Thomas et al., 1977).

The second class of methods fits the curve through each data point without smoothing. These methods are preferable when the uncertainty at each point is small compared with the differences among points, or when one can explain differences using collateral data. Technically, such methods are interpolants. Linear interpolation is most common, and quadratic and cubic interpolants (Erh, 1972) are simply extensions of this general technique to second- and third-order polynomials.

Choosing between these two classes of techniques requires knowledge of the objectives of the analysis and the capabilities of the methods. If the objective is finding trends in noisy data, one usually uses some smoothing technique. On the other hand, for analysis of variation in time series, one usually cannot use smoothing and therefore requires an interpolation technique. Mathematical characteristics of the candidate equations also differ. For instance, differentiating a linear interpolation results in a sequence of plateaus, or stair-steps, in the derivative of the interpolated function. Similarly, a quadratic interpolation, when differentiated, results in a derivative that looks like a saw-tooth pattern. At the cubic or higher order, the derivative is smooth and continuous in mathematical terms.

Such considerations resulted in our choice of a cubic spline interpolant to analyze nutrient and dry matter accumulation data to produce accumulation rates

Table 1. Sample data set for cubic spline interpolation. Data are aerial N accumulations for corn grown near Florence, SC, in 1980 (Karlen et al., 1987b). Plant population was 89 000 plants ha^{-1}; final yield was 14.0 Mg ha^{-1}.

Input data set†

```
7
  0          0
 43.0       26.8
 50.0       64.3
 57.0      126.2
 64.0      160.2
 82.0      191.0
113.0      237.5
```

Output data set for interpolated values‡

Days	Amount	Rate
0.0000000E+00	0.0000000E+00	−0.5960464E−07
0.1000000E+01	−0.4062169E−01	−0.7993070E−01
0.2000000E+01	−0.1572362E+00	−0.1519856E+00
	•	
	•	
	•	
0.1110000E+03	0.2371584E+03	0.3693156E+00
0.1120000E+03	0.2374385E+03	0.1893706E+00
0.1130000E+03	0.2375340E+03	0.0000000E+00

Coefficient file ($Y = c_0 + c_1X + c_2X^2 + c_3X^3$)§

X_{low}	X_{high}	c_0	c_1	c_2	c_3
6					
0.0000000E+00	43.00000	0.0000000E+00	−5.9604645E−08	−4.1934248E−02	1.3126194E−03
43.00000	50.00000	26.82600	3.674754	0.1273936	1.6263664E−02
50.00000	57.00000	64.37000	7.849023	0.4689306	−4.6649452E−02
57.00000	64.00000	126.2900	7.556582	−0.5107079	1.7888077E−02
64.00000	82.00000	160.2970	3.036219	−0.1350583	3.4092853E−03
82.00000	113.0000	191.0730	1.487945	4.9043078E−02	−1.5707999E−03

† The first line holds the number of pairs. Time, the first column, is in days after planting. Values for N, the second column, are in kg N ha^{-1}.
‡ Amount is in kg N ha^{-1}, and rate is in kg N ha^{-1} d^{-1}. Dots indicate omission of Days 3 to 110, to shorten the table.
§ Coefficient file for data above. The first line holds number of subranges.

Example of MathCad worksheet to compute cubic spline interpolant.

Define ha as a substitute for the built-in unit hectare: $ha := 1 \cdot hectare$

and d as a substitute for the built-in unit day: $d := 1 \cdot day$

Given the following data set (Table 1):

$$Days := \begin{bmatrix} 0 \\ 43 \\ 50 \\ 57 \\ 64 \\ 82 \\ 113 \end{bmatrix} \cdot d \qquad AccN := \begin{bmatrix} 0 \\ 26.8 \\ 64.3 \\ 126.2 \\ 160.2 \\ 191.0 \\ 237.5 \end{bmatrix} \cdot \frac{kg}{ha}$$

Define the index for these vectors.

$i := 0, 1 .. 6$

Perform the MathCad function to provide derivatives at each point.

$vs := lspline(Days, AccN)$

Make a time vector t on which to evaluate the interpolant.

$t := 0 \cdot d, 1 \cdot d .. 113 \cdot d$

Define the interpolant function.

$Acc(t) := interp(vs, Days, AccN, t)$

Evaluate the functions for some point, say 100 days.

$Acc(100 \cdot d) = 216.227 \cdot \frac{kg}{ha}$

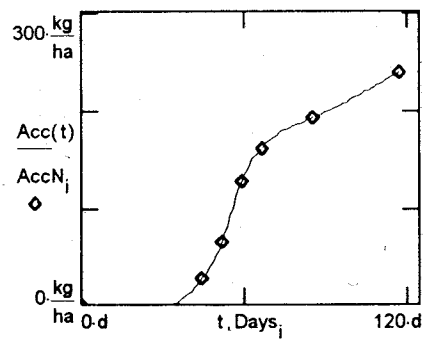
And its derivative.

$Rate(t) := \frac{d}{dt} Acc(t)$

$Rate(100 \cdot d) = 1.578 \cdot \frac{kg}{(ha \cdot d)}$

Plot the accumulation function...

...and the rate function.



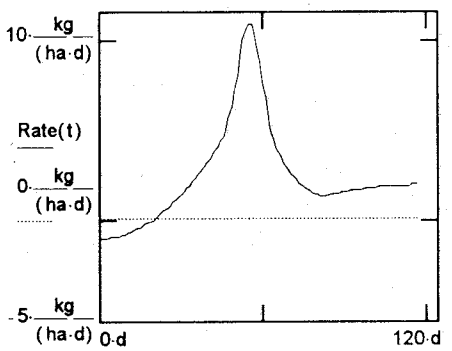


Fig. 1. Sample MathCad worksheet to compute the cubic spline interpolant for the data in Table 1.

at any point in time. This was done originally to compare recent and historical nutrient accumulation for corn (Karlen and Sadler, 1986), and since has been used to analyze differences in accumulation rate for high- and low-population corn (Karlen et al., 1987b), for partitioning of nutrients among plant parts for corn (Karlen et al., 1987a, 1988), for describing nutrient accumulation in soybean [*Glycine max* (L.) Merr.] (Karlen and Sadler, 1989), and for describing nutrient accumulations for soft red winter wheat (*Triticum aestivum* L.) (Karlen and Sadler, 1990).

Continued requests for information about this technique, continued confusion about its application, and the development of an easier way to perform it all led to our decision to document the technique as applied to nutrient accumulation rates. The utility of the method hinges on ease of use, the assumptions used, and interpretation of the results. Our objective was to document the rationale and procedures for the analysis of nutrient accumulation data using the cubic spline interpolation, with an illustration of the procedure, including data form and results.

MATERIALS AND METHODS

The cubic spline interpolation algorithm (Burden et al., 1981) was programmed in FORTRAN-77 (ANSI X3.9-1978). It has

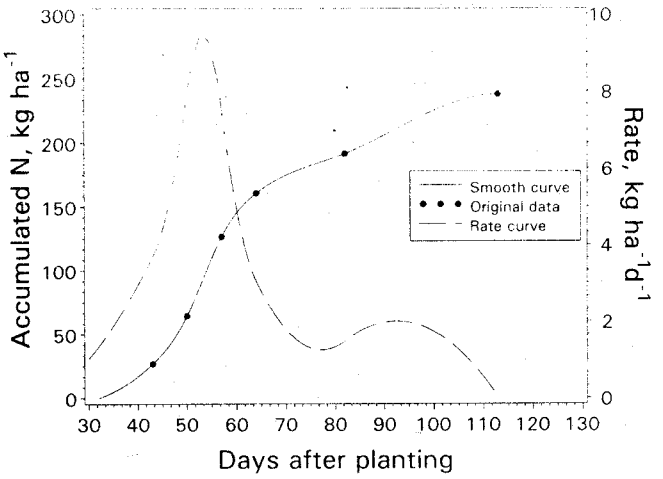


Fig. 2. Sample results from the input data set shown in Table 1, computed using the FORTRAN implementation. Original data are aerial N accumulations for corn grown near Florence, SC, in 1980 (Karlen et al., 1987b).

been run on both a desktop personal computer and a minicomputer and required <5 s CPU time on either. Similar algorithms are available from computer software companies (e.g., Turbo Pascal's Numeric Toolbox, from Borland International,[1] Scotts Valley, CA) and from numerical methods texts (e.g., Press et al., 1989, which has Pascal, FORTRAN, and C versions). Microcomputer-based mathematical processors can be used as well (e.g., MathCad, from Mathsoft, Inc., Cambridge, MA). Those inclined to understand the numerical details of the solution of the cubic spline are referred to Burden et al. (1981) or Press et al. (1989). The FORTRAN program is ≈100 lines of FORTRAN code, and totals <4K in size for the source code. Users may request a listing from the senior author.

The two forms of the cubic spline in Burden et al. (1981) differ with respect to the conditions at either end of the data series. The free, or natural, spline has no constraints; the curve is free to pass through the endpoints at any slope. The clamped spline, which we used, requires input values for the slopes at the two endpoints. For our purposes, assuming data were taken up to maturity suggested that zero rates are usually sufficient for both endpoints.

An example input data set from Karlen et al. (1987b), shown in Table 1, illustrates the input requirements and operation of the program. For input, the program prompts for the source filename and then for the initial and final derivatives. For output, the program creates two files. The first has the same filename as the input file and a file extension of GEN. Each record corresponds to 1 d, and has three columns: the time in days after planting, the interpolated value for accumulated nutrient for that day, and the rate of accumulation (slope) for that day. The second output file uses the same filename as the input, but with a file extension of COE, for coefficients. It has a line with the number of subranges (computed as the number of points minus one), which is the number of lines that follow. Each of these lines holds the low and high x-value for the subrange, plus the four coefficients for the cubic polynomial that applies to that range.

For most purposes, such as creating graphs of these data, the generated table of daily values is sufficient. However, the resulting compound polynomial can be evaluated for any value. The coefficient array shown in Table 1 has a number of rows equal to the number of subranges (original number of points minus 1), and four columns, which are the coefficients c_0, c_1, c_2, and c_3 of the cubic polynomial for each subrange. Users wishing to manually evaluate these polynomials must define an x-value that is the duration from the start of the proper subrange, and use it in the evaluation of the corresponding polynomial. Because of this convention, the first coefficient is always the same as the original y-value. From this point, evaluating the polynomial or its derivative is a simple matter. Integration between two points is similarly easy, with the precaution that, if the two points are in different subranges, the polynomial must be integrated separately for each subrange and then summed.

Interpolants similar to that programmed in FORTRAN are available in the computer-based mathematical scratchpad MathCad. A worksheet illustrating the cubic spline interpolant, with one variation from the FORTRAN version, is shown in Fig. 1. The data series from Table 1 is input, and an intermediate function is computed. The interpolant is then computed from the intermediate. The difference between the worksheet version and the FORTRAN version is that MathCad provides linear, quadratic, and cubic constraints at the endpoints, where the Burden et al. (1981) source provides only fixed derivatives, which we used, or unconstrained derivatives. In the scratchpad, the most constrained form appears to be the one with linear endpoints.

The effect of data frequency was illustrated with a data set that had sufficient length and frequency that subsamples could be taken from the base set and compared among themselves and to the base data set. Sayre (1948) published aerial dry matter accumulation for corn with 31 points, one taken every 3 d. From this set, two sets of 6-d samples were selected by allocating alternate points to one of two sets. Similarly, every third point was put in three 9-d data sets, and every fifth in five 15-d sets. The cubic spline was run on each of these data sets.

RESULTS AND DISCUSSION

The results for the example data set are shown in Fig. 2, which was developed using the FORTRAN program and the generated data shown in Table 1. They illustrate the value of the interpolation technique during the period in which data were taken about weekly. Results from any interpolation method depend on data frequency. The 31-d period with no data between the last two points leaves the curve's shape dependent on the assumption of zero rate at the endpoint. The small bulge in the curve between Days 85 and 100 could have been reduced by setting the final derivative to some value approximating the average rate during the final period. This was not done, however, because the assumption of zero growth at physiological maturity requires that the growth rate at that point has decreased to zero. A second aspect, also related to data spacing, is shown by the period between 0 and 30 d after planting, in which the curve is actually below zero, and therefore nonsensical, for this data set. Although the input data file started with zero at planting, the initial subrange resulted in poor information because of the wide spacing of the data. Normally, this interval is of less interest than the period of rapid growth and is disregarded. If it is of interest, however, data during this interval must be collected at an adequate frequency.

The results from the MathCad worksheet are shown in Fig. 1. There are few differences from the FORTRAN version, until the last subrange. The linear constraint on the final derivative produced a nearly constant rate for the final interval (Fig. 1). The assumption of zero uptake at physiological maturity forced a small rise before the final decrease to zero (Fig. 2). In absolute terms, the difference between the two curves is very small. However, interpretation of the assumptions used and their effect on the results is particularly salient. The nearly

[1] Mention of a trademark is for the reader's convenience and does not imply endorsement by the USDA.

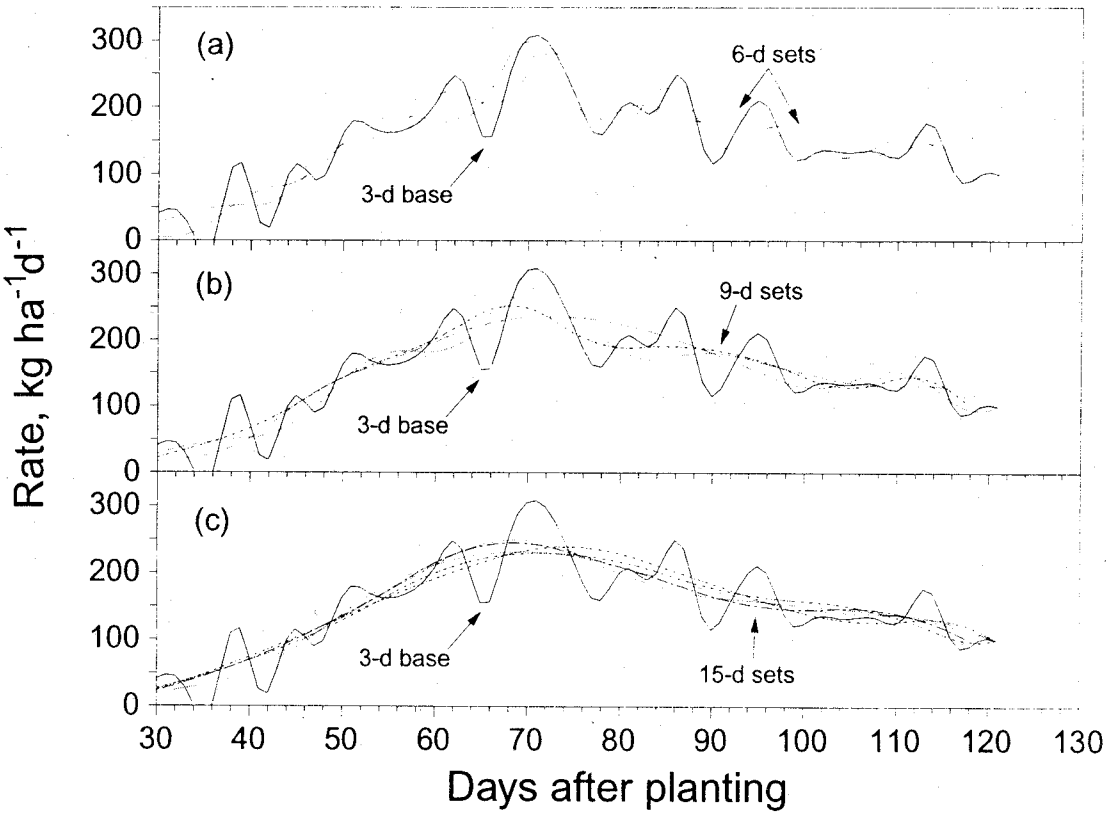


Fig. 3. Comparison of rates derived from the 3-d frequency base data set (Sayre, 1948) with those derived from (a) the two 6-d frequency data sets, (b) the three 9-d frequency data sets, and (c) the five 15-d frequency data sets.

constant rate of uptake for the last few weeks is visually satisfactory; nothing exists in the data itself to suggest otherwise. However, if it is more likely that the plant slowly decreased uptake to zero when it reached maturity, then the curve in Fig. 2, in spite of the small hump during the last interval, may be a closer approximation to the actual rate of uptake. Questions such as these must be considered when choosing the methods and assumptions, and also when choosing the boundary conditions.

Effects of data frequency are illustrated by the data of Sayre (1948). Results from the 6-d sets are compared with the base 3-d set in Fig. 3a. Both 6-d sets show less short-term variation than the 3-d base set, though they all show nearly the same seasonal maximum (within 10%) between Day 70 and 75. The 9-d sets show correspondingly less variation than the 6-d sets (Fig. 3b), and the results for all five of the 15-d sets show very nearly the same seasonal trends (Fig. 3c). Decreasing data frequency moderates the variation in the rate curve. The average rate between points must integrate to the correct total on the day of measurement, because the spline curve passes through each point in the accumulated amount data series. Spacing observations farther apart makes the rate curve approach the average rate for the now-longer period.

Users of this technique must consider the significance of the variation from the long-term mean rate. If the variation is important, or the uncertainty in the observation is small compared with the variation, this technique should be valid. If variation about an observation is noise in a signal, the user might wish to use one or the other of the smoothing techniques mentioned earlier.

Data spacing in a series may be critical if data are irregularly spaced and rates are highly variable. Increased frequency of observation within a data set may lead to susceptibility to oscillation in the polynomial as it attempts to pass through each point, especially if errors accumulate in opposite directions. For instance, an underestimate followed closely by an overestimate will result in a rate much higher than if the opposite occurred.

SUMMARY AND CONCLUSIONS

The cubic spline interpolation can be a useful tool for a wide range of rate determinations. Because the curve passes through each datum, the method is best suited to data for which variation at a time is less than variation between times. The user must decide whether the variation is information or noise. The information gained is directly related to the data input and assumptions made; higher sampling frequencies should be used if short-term variation is of interest.

REFERENCES

Burden, R.L., J.D. Faires, and A.C. Reynolds. 1981. Numerical analysis. Prindle, Weber & Schmidt, Boston, MA.

Clawson, K.L., J.E. Specht, B.L. Blad, and A.F. Garay. 1986. Water use efficiency in soybean pubescence density isolines: A calculation procedure for estimating daily values. Agron. J. 78:483–487.

DuChateau, P.C., D.L. Nofziger, L.R. Ahuja, and D. Swartzendruber. 1972. Experimental curves and rates of change from piecewise parabolic fits. Agron. J. 64:538–542.

Erh, K.T. 1972. Application of the spline function to soil science. Soil Sci. 114:333–338.

Gorry, P. 1990. General least squares smoothing by the convolution (Savitzky–Golay) method. Anal. Chem. 62:570–573.

Henderson, J.B., and E.J. Kamprath. 1970. Nutrient and dry matter accumulation by soybeans. N.C. Agric. Exp. Stn. Tech. Bull. 197.

Karlen, D.L., R.L. Flannery, and E.J. Sadler. 1987a. Nutrient and dry matter accumulation rates for high yielding maize. J. Plant Nutr. 10:1409–1417.

Karlen, D.L., R.L. Flannery, and E.J. Sadler. 1988. Aerial accumulation and partitioning of nutrients by corn. Agron. J. 80:232–242.

Karlen, D.L., and E.J. Sadler. 1986. Seasonal N, P, and K uptake rates associated with high yield corn. *In* R.D. Munson (ed.) Physiology, biochemistry, and chemistry associated with maximum yield corn. Proc. Res. Roundtable for Maximum Yield Research (MYR) of the Future, St. Louis, MO. 11–12 Nov. 1985. Foundation for Agron. Res. and Potash & Phosphate Inst., Atlanta, GA.

Karlen, D.L., and E.J. Sadler. 1989. Aerial dry matter accumulation rates for soybean at different yield levels. *In* R.D. Munson (ed). The physiology, biochemistry, nutrition, and bioengineering of soybeans: Implications for future management. Proc. Res. Roundtable with Focus on the Future, St. Louis, MO. 14–15 Nov. 1989, Foundation for Agron. Res. and Potash & Phosphate Inst., Atlanta, GA.

Karlen, D.L., and E.J. Sadler. 1990. Nutrient accumulation rates for wheat in the southeastern Coastal Plains. Commun. Soil Sci. Plant Anal. 21:1329–1352.

Karlen, D.L., E.J. Sadler, and C.R. Camp. 1987b. Dry matter, nitrogen, phosphorus, and potassium accumulation rates by corn on Norfolk loamy sand. Agron. J. 79:649–656.

Kimball, B.A. 1974. Smoothing data with Fourier transformations. Agron. J. 66:259–262.

Kimball, B.A. 1976. Smoothing data with cubic splines. Agron. J. 68:126–129.

Press, W.H., B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. 1989. Numerical recipes (FORTRAN version). Cambridge Univ. Press, New York.

Sayre, J.D. 1948. Mineral accumulation in corn. Plant Physiol. 23:267–281.

Thomas, A.W., W.M. Snyder, and R.R. Bruce. 1977. Smoothing, interpolation, and gradients from limited data. Agron. J. 69:747–750.