Conference on Applied Statistics in Agriculture 1990 - 2nd Annual Conference Proceedings

LINEAR-PLATEAU REGRESSION ANALYSIS AND ITS APPLICATION TO SELENITE ADSORPTION RATE

Bahman Shafii

Kevin C. Harper

Steven L. McGeehan

LINEAR-PLATEAU REGRESSION ANALYSIS AND ITS APPLICATION TO SELENITE ADSORPTION RATE

Bahman Shafii, Statistical Programs, College of Agriculture
Kevin C. Harper, Computer Services, and
Steven L. McGeehan, Division of Soil Sciences, College of Agriculture
University of Idaho
Moscow, ID 83843

ABSTRACT

Simple computational methods are presented which facilitate fitting regression models to response data exhibiting a plateau effect. The iterative statistical program (called PLATFOR) is written in FORTRAN (a SAS version is also available), and produces all relevant regression statistics, plots, and information on goodness of fit. The presented procedures are empirically valuable, since linear-plateau models have many useful applications in agriculture, especially in soil fertility and soil chemistry experiments. The technique was employed in an experiment designed to determine the effect of soil volcanic ash content on selenite adsorption. Ion chromatographic methods were used to investigate selenite adsorption in three acidic North Idaho surface soils of varying ash content. Equilibration periods of 0.5 to 12.0 hours were evaluated and the time required to reach maximum adsorption was determined using the PLATFOR program. It was concluded that both the rate and magnitude of selenite adsorption were influenced by soil volcanic ash content.

Keywords: linear-plateau models, regression analysis, selenite adsorption.

I. INTRODUCTION

In certain situations in agricultural research, especially those of soil fertility and soil chemistry, the response function exhibits a plateau effect. In such situations, it is often appropriate to approximate the underlying function with two intersecting linear lines (Figure 1) accounting for the sloping as well as the plateau segments of the response. Here, the emphasis is to approximate the real regression function, E[y(x)], which is certainly nonlinear (and asymptotic), with a sequence of low order polynomials or segmented models. Statistical estimation of the optimal plateau, the point at which increases in the input under investigation (e.g. fertilizer) would have no significant effect on the magnitude of the response (e.g. yield), may also be desirable and empirically useful.

Linear-plateau regression analysis and its use in evaluating response to fertilizer nutrients was first introduced by Anderson and Nelson (1975). Agricultural researchers have since employed the techniques in describing other soil/crop situations (e.g. Mascagni and Cox, 1985; Potter *et al.*, 1985; Mahler and McDole, 1987; Singh, 1989). Procedures for fitting segmented curves and estimation of join points were introduced by Hudson (1966). However, computational methods outlining the iterative fitting program along with the associated statistics and the corresponding graphics have not been readily available. The main objective of this paper is to present simple computational techniques which facilitate fitting regression products to response data exhibiting a plateau effect. Empirical applications are demonstrated with reference to the selenite adsorption rate studies.

Importance of Selenite Adsorption Studies

The role of selenium (Se) as an essential trace element in human and animal nutrition as well as its toxic effects at slightly higher concentrations has been well established (Adriano, 1986). Selenium is present in detectable but highly variable amounts in soils. High soil selenium levels can lead to toxic responses including "alkali disease" and "blind staggers" (Adriano, 1986). Selenium leaching from agricultural soils has caused a high incidence of deformity and mortality of waterfowl in the San Joaquin Valley of California (Deverel and Millard, 1986). On the other hand, low soil selenium levels can lead to nutritional disorders such as "white muscle disease" (Adriano, 1986).

The chemical form of selenium determines its transport and bioavailability in soils. Inorganic Se occurs as metal selenides, elemental Se, selenite (SeO_3^{2-}), or as selenate (SeO_4^{2-}). Selenate and selenite are the predominant forms in well aerated soils (Neal *et al.*, 1987). Selenate is weakly adsorbed by soils while selenite forms a strong adsorption complex (Ahlrichs and Hossner, 1987; Neal *et al.*, 1987). Hence, the solubility of Se in soils depends on the predominant ionic form of Se and the interaction of the ionic species with adsorbing surfaces.

The rate of selenite adsorption decreases with equilibration time, often reaching a plateau. Equilibration periods for published selenite adsorption experiments are highly variable and range from 1 to 72 h (Rajan, 1979; Davis and Leckie, 1980; Singh *et al.*, 1981). Researchers have relied on a visual (qualitative) estimation of adsorption data to identify the plateau. Hence, statistical comparison of adsorption rate and maxima is important for evaluating soil selenite behavior.

II. METHODS

The iterative statistical program (PLATFOR) for performing the required regression analysis is written in Microsoft FORTRAN (version 3.0), and corresponding graphics are accomplished using Turbo PASCAL (version 5.0). Program codes are compiled into two exec files, one batch file, and five graphics drivers. The program requires minimal computer resources and runs easily on any IBM-PC (or compatible) machine. Copies of the program are available, upon request, from the senior author.

The first fitted equations considered by the program are of the form:

$$Y = \beta_0 + \beta_1 [\min(X,A)] \qquad (1)$$

where β_0 is the intercept, and β_1 is the slope of the linear line up to the point X = A. The value of A is selected by first setting A = max(X), i.e., the largest observed value of X. Successive iterations will set A equal to successively smaller and smaller observed values of X. At each iteration, the least square estimates of (1) along with the residual sum of squares, and the coefficient of determination (R^2) are calculated. The plateau model corresponding to the smallest residual mean square is then selected, and the "optimal plateau" is indicated.

A second equation,

$$Y = \beta_0 + \beta_1 [\min\ (X,A)] + \beta_2 [\max(X,A)-A] \qquad (2)$$

is also fitted to enable testing for the existence of significant slope (β_2) beyond the point X = A (Potter *et al.*, 1985). At this point the program provides all relevant regression statistics, including parameter estimates, standard error of estimates, t and corresponding p values (probability of obtaining a larger value of |t| under the hypothesis H_o:parameter = 0).

Graphics options include scatter plot, scatter plot with the fitted plateau regression line, and scatter plot with plateau and joined lines. These plots along with other statistics provided by the program are extremely useful in evaluating the goodness of fit as well as appropriateness of assuming a linear-plateau model to approximate a specified response data.

A SAS implementation of the PLATFOR program may alternatively be considered for the analysis. SAS codes required to perform the statistical computations and graphics are outlined in Table 1. This program was written using SAS (Statistical Analysis System) Version 5.18, and Procedures MATRIX (SAS Statistics), and GPlot (SAS Graphics). The program can be converted to IML codes via the MATIML Procedure.

III. APPLICATION

Linear plateau regression analysis was employed in an experiment designed to evaluate selenite adsorption in three north Idaho soils of varying volcanic ash content. Samples of the surface mineral horizon of each soil were air-dried and ground to pass a 2 mm sieve. Sodium fluoride pH and oxalate-extractable Al and Fe values were used as indicators of volcanic ash content (Table 2). The Huckleberry silt loam, which exhibited the highest sodium fluoride pH and highest extractable aluminum and iron values, was assumed to have the highest volcanic ash content. Rates of selenite adsorption were determined using soil suspensions (1:25 w/v) containing 2.0 mg/L selenite. Equilibration times ranged from 0.5 to 12 hr with three replications. Soluble selenite was determined using a Dionex 4000i Ion Chromatograph with Dionex AS4a separator column and 1.5 mM Na_2CO_3 + 1.0 mM $NaHCO_3$ eluent. Adsorption data were analyzed using the PLATFOR program.

Scatter plots along with the fitted plateau regression line and the corresponding prediction equation for each specified soil are given in Figure 2 (a, b, and c). It is evident that the linear-plateau model has provided a good fit for the adsorption data in all cases. All regression coefficients were statistically significant ($P<0.01$), and the tests for the β_2 parameter (Eqn. 2) indicated the lack of significant slope beyond the point of optimal plateau in each case.

Maximum adsorption was obtained in 4, 6, and 8 hr in the Huckleberry, Helmer, and Santa soils, respectively (Figure 2). Hence, the time required to reach the adsorption plateau decreased as the soil volcanic ash content increased. Furthermore, the greatest selenite adsorption occurred in the Huckleberry soil, which exhibited the greatest volcanic ash content (Figure 2). Thus, the results indicate that both the rate and magnitude of selenite adsorption depended on the soil volcanic ash content.

IV. CONCLUSION

The PLATFOR program offers a simple computational technique which facilitates fitting regression models to empirical data exhibiting a plateau effect. The program is easy to run and provides relevant regression statistics and plots that can be useful in evaluating the suitability of linear-plateau models in approximating a specific response function. Further extensions of the program (multi-input situations, confidence intervals for the optimal plateau, comparison of plateau regression lines, other splining techniques) are being considered, and the authors intend to include the findings in future work.

ACKNOWLEDGEMENTS

Idaho Agricultural Experiment Station manuscript number 9001. The authors wish to thank Dr. Denny V. Naylor for his assistance in the selenite adsorption studies, and an anonymous referee for comments and suggestions made on a previous draft of this paper.

REFERENCES

Adriano, D.C. 1986. Trace Elements in the Terrestrial Environment. Springer-Verlag, New York.

Ahlrichs, J.S., and L.R. Hossner. 1987. Selenate and selenite mobility in overburden by saturated flow. J. Environ. Qual. 16:95-98.

Anderson, R.L., and L.A. Nelson. 1975. A family of models involving intersecting straight lines and concomitant experimental designs useful in evaluating response to fertilizer nutrients. Biometrics 31:303-318.

Davis, J.A., and J.O. Leckie. 1980. Surface ionization and complexation at the oxide/water interface. 3. Adsorption of anions. J. Colloid Interface Sci. 74:32-43.

Deverel, S.J., and S.P Millard. 1986. Distribution and mobility of selenium and other trace elements in shallow ground water of the western San Joaquin Valley, California. U.S. Geological Survey, Open-File Report 86-538.

Hudson, D. J. 1966. Fitting segmented curves whose joined points have to be estimated. JASA 64:1096-1129.

Mahler, R.L., and R.E. McDole. 1987. Effect of soil pH on crop yield in northern Idaho. Agron. J. 79:751-755.

Mascagni, H.J. Jr., and F.R. Cox. 1985. Effective rates of fertilization for correcting manganese deficiency in soybeans. Agron. J. 77:373-375.

Neal, R.H., G. Sposito, K.M. Holtzclaw, and S.J. Traina. 1987. Selenite adsorption on alluvial soils: I. Soil composition and pH effects. Soil Sci. Soc. Am. J. 51:1161-1165.

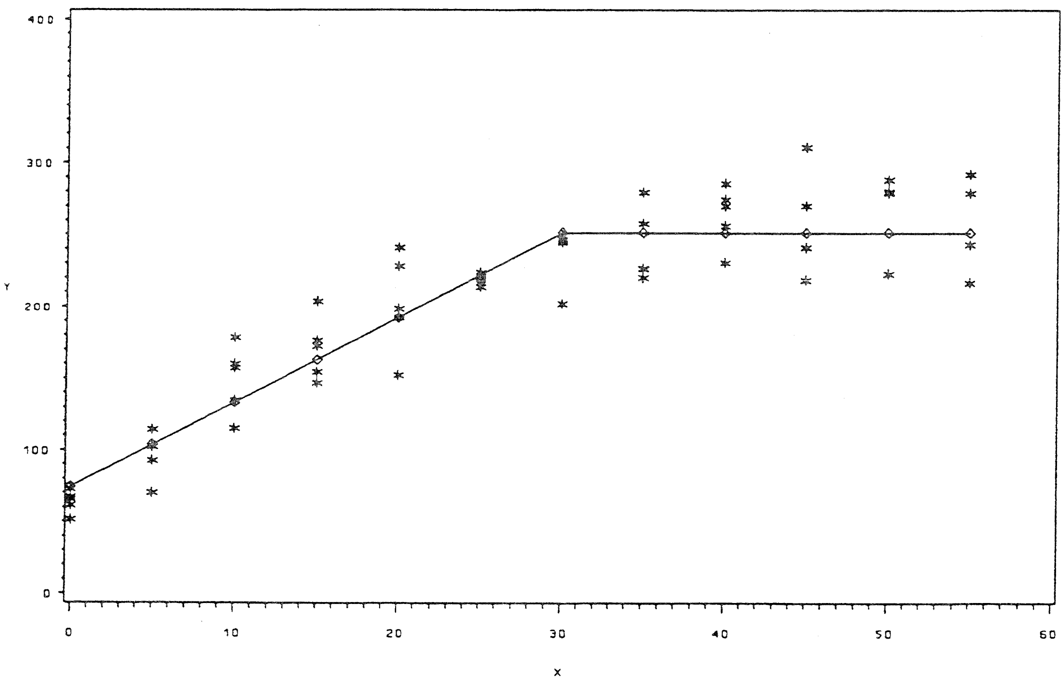
Potter, D.A., B.L. Bridges, and F.C. Gordon. 1985. Effect of N fertilization on earthworm and microarthropod population in Kentucky bluegrass turf. Agron. J. 77:367-372.

Rajan, S.S.S. 1979. Adsorption of selenite, phosphate and sulphate on hydrous alumina. J. Soil Sci. 30:709-718.

Singh, K.P. 1989. Statistical models for the analysis of dose-response data. Proc. Conference on Applied Statistics in Agriculture, April 30-May 2. G.A. Milliken and J.R. Schwenke (Eds.). Kansas State University, Manhattan, Kansas.

Singh, M., N. Singh, and P.S. Relan. 1981. Adsorption and desorption of selenate and selenite on different soils. Soil Sci. 132:134-141.

Figure 1. Linear Plateau Response.



New Prairie Press
https://newprairiepress.org/agstatconference/1990/proceedings/12

Table 1. SAS Implementation of the PLATFOR Program.

```
/*
     THIS PROGRAM CAN BE CONVERTED TO IML CODE BY USING PROC MATIML
     THIS PROGRAM IS SELF-CONTAINED ONCE THE CORRECT POINTER
     IS SET IN THE FILEDEF STATEMENT BELOW.
     GRAPHICS:  WITHOUT ADJUSTMENT THIS PROGRAM PRODUCES AN OVERLAID
                SCATTER PLOT.  TO HAVE CONNECTED LINES USE THE SAS
                GRAPH CODE WHICH IS IN COMMENTS AT THE BOTTOM OF THIS
                PROGRAM.
*/

CMS FI IN1 DISK TUBEWS DATA A;   /*  SET POINTER TO DATA FILE  */
 GOPTIONS  DEVICE = TEK4662;
DATA ALL;
INFILE IN1;
 INPUT X Y @@;                   /*  READ IN OBSERVATIONS   */
PROC SORT DATA=ALL OUT=SALL;     /* SORT BY X VARIABLE      */
BY X;
PROC MATRIX;
 FETCH A DATA=SALL;              /*  GET SORTED DATA        */
 NOBS = NROW(A);
 RSQRMAX = -999;
 X = J(NOBS,1,1) ]] A(,1);       /*  INITIALIZE X MATRIX    */
 Y = A(,2);
 XPY = J(3,1,0);
   SY =  J(1,NOBS,1) * Y;        /*  COMPUTE CONSTANTS      */
   SYY = Y' * Y;
   SSTO = SYY - SY * SY #/ NOBS;
   XPY(1,1) = SY;
   XHOLD = J(NOBS,2,0);
   converge = 1;
   MIDPOINT = NOBS#/2;
   MIDPOINT = ROUND(MIDPOINT,1);
DO WHILE (CONVERGE = 1);       /*  PERFORM BISECTION FITTING  */
   RIGHT = MIDPOINT + ((NOBS - MIDPOINT) #/ 2);
   LEFT = MIDPOINT - ((NOBS - MIDPOINT) #/ 2);
   RHLD = X(RIGHT,2);
   RXHOLD = (X(1:RIGHT,1:2) //
     (J((NOBS-RIGHT),1,1) ]] J((NOBS-RIGHT),1,RHLD)));
  RHX1 =  J(2,NOBS,1) * RXHOLD;
  RSX1 = RHX1(1,2);
  RHX1X1 = RXHOLD' * RXHOLD;
  RSX1X1 = RHX1X1(2,2);
  IF DET(RHX1X1) NE 0 THEN DO;
     RXPXI = INV(RXHOLD' * RXHOLD);     /*   O.L.S.       */
     RBETA = RXPXI * RXHOLD' * Y;
     RYHAT = RXHOLD * RBETA;
     RRESID = Y - RYHAT;
     RSSE = RRESID' * RRESID;
     RRSQR = 1 - RSSE #/ SSTO;
   LHLD = X(LEFT,2);
   LXHOLD = (X(1:LEFT,1:2) //
     (J((NOBS-LEFT),1,1) ]] J((NOBS-LEFT),1,LHLD)));
  LHX1 =  J(2,NOBS,1) * LXHOLD;
  LSX1 = LHX1(1,2);
  LHX1X1 = LXHOLD' * LXHOLD;
  LSX1X1 = LHX1X1(2,2);
  IF DET(LHX1X1) NE 0 THEN DO;
     LXPXI = INV(LXHOLD' * LXHOLD);     /*   O.L.S.       */
     LBETA = LXPXI * LXHOLD' * Y;
     LYHAT = LXHOLD * LBETA;
     LRESID = Y - LYHAT;
     LSSE = LRESID' * LRESID;
```

```
    LRSQR = 1 - LSSE #/ SSTO;
    OMIDPOINT = MIDPOINT;
    IF LRSQR > RRSQR THEN DO;
       MIDPOINT = LEFT;
       SIDE = 1;
       END;
       ELSE DO;
       MIDPOINT = RIGHT;
       SIDE = 2;
       END;
    IF MIDPOINT = OMIDPOINT THEN CONVERGE = 2;
    IF CONVERGE=2 THEN DO;         /* PREPARE OUTPUT MATRIX   */
       IF SIDE = 1 THEN DO;
         BETA = LBETA;
         SSE = LSSE;
         RSQR = LRSQR;
          END;
        ELSE DO;
          BETA = RBETA;
          SSE = RSSE;
          RSQR = RRSQR;
          END;
  RESULTS1 = X(MIDPOINT,2) ]] BETA(1,1) ]] BETA(2,1) ]] SSE ]] RSQR;

       RSQRMAX = RSQR;
       XMAX = X(MIDPOINT,2);
       POSITION = MIDPOINT;
       BMAX = BETA;
       END;
   BO = BETA(1,1);
   B1 = BETA(2,1);
   END;
END;
YHAT1 = X * BMAX;                /* COMPUTE YHAT VECTOR   */
DO I = POSITION TO NOBS;
   YHAT1(I,1) = XMAX * BMAX(2,1) + BMAX(1,1);
END;
PLOT = X(,2) ]] Y ]] YHAT1;        /* GET PLOT DATA SET   */
OUTPUT PLOT OUT=P1(RENAME = (COL1 = X COL2 = Y COL3 = YHAT1));
MAX = XMAX;
OUTPUT MAX OUT=XMAX(RENAME = (COL1 = X COL2 = Y COL3 = YHAT));
XHOLD1 = X(1:POSITION,1:2) // (J(NOBS-POSITION,1,1) ]]
J(NOBS-POSITION,1,XMAX));
COL1 = J(NOBS,1,1);               /* COMPUTE MODEL:                  */
COL2T = J(POSITION,1,0);          /* Y = BO + B1*MIN(X,X_MAX) +      */
DIFF = NOBS - POSITION;           /*      B2*MAX(0.0,X-X_MAX)        */
COL2B = X(POSITION+1 : NOBS ,2) - J(DIFF,1,XMAX);
XHOLD2 = COL1 ]] (COL2T //COL2B);
HX1 = J(2,NOBS,1) * XHOLD1;
SX1 = HX1(1,2);
HX2 = J(2,NOBS,1) * XHOLD2;
SX2 = HX2(1,2);
SX1X1 = XHOLD1(,2)' * XHOLD1(,2);
SX1X2 = XHOLD1(,2)' * XHOLD2(,2);
SX2X2 = XHOLD2(,2)' * XHOLD2(,2);
XPY(2,1) = XHOLD1(,2)' * Y;
XPY(3,1) = XHOLD2(,2)' * Y;
XBAR1 = SX1 #/ NOBS;
XBAR2 = SX2 #/ NOBS;
SSX1 = SX1X1 - SX1*SX1 #/ NOBS;
SSX2 = SX2X2- SX2*SX2 #/ NOBS;
SSX1X2 = SX1X2- SX1*SX2 #/ NOBS;
DN = SSX1 * SSX2 - SSX1X2 * SSX1X2;
NXPXI = J(3,3,0);
```

```
  NXPXI(1,1) = 1 #/ NOBS - (2*XBAR1*XBAR2*SSX1X2 - XBAR1*XBAR1*SSX2
               -XBAR2*XBAR2*SSX1) #/ DN;
  NXPXI(2,1) = (XBAR2*SSX1X2 - XBAR1*SSX2) #/ DN;
  NXPXI(3,1) = (XBAR1*SSX1X2 - XBAR2*SSX1) #/ DN;
  NXPXI(2,2) = SSX2 #/ DN;
  NXPXI(3,2) = -1*SSX1X2 #/ DN;
  NXPXI(3,3) = SSX1 #/ DN;
  NXPXI(1,2) = NXPXI(2,1);
  NXPXI(1,3) = NXPXI(3,1);
  NXPXI(2,3) = NXPXI(3,2);
  B = J(3,1,0);
  DO I = 1 TO 3;
    DO J = 1 TO 3;
      B(I,1) = B(I,1) + NXPXI(I,J) * XPY(J,1);
    END;
  END;
  SSREG = B' * XPY;
  MSE = (SYY - SSREG) #/ (NOBS-3);
  SEB = ( VECDIAG(NXPXI) * MSE) ## .5;
  TB = B #/ SEB; DF = J(3,1,NOBS-3);
  RESULTS2 = B ]] SEB ]] TB ]] DF;
  OUTPUT RESULTS1 OUT=R1(RENAME = (COL1 = X COL2 = B0 COL3 = B1
           COL4 = SSE COL5=RSQR));
  OUTPUT RESULTS2 OUT=R2(RENAME = (COL1 = B COL2 = SEB
                  COL3 = TB COL4 = DF));
DATA NEW1; SET R1; DROP ROW;
TITLE1 'FITS FOR PROGRESSIVELY DECREASING X_MAX FOR THE MODEL:';
TITLE2 'Y = B0 + B1 * MIN(X,X_MAX)';
LABEL X = 'X_MAX'
      SSE = 'SS_ERROR'
      RSQR = 'R_SQUARE';
PROC PRINT LABEL NOOBS;
RUN;
DATA XXX; DROP ROW;
  SET XMAX;
  CALL SYMPUT('XXMAX',LEFT(X));
 RUN;
TITLE 'FIT FOR THE MODEL Y = B0 + B1*MIN(X,X_MAX) + B2*MAX(0.,X-X_MAX)';
TITLE2 "WHEN X_MAX = &XXMAX (OPTIMAL PLATEAU)";
DATA NEW2; SET R2; DROP ROW TPROB ATB DF;
LABEL  B = 'BETA_HAT'
       SEB = 'STANDARD ERROR'
       TB = 'T VALUE'
       PVALUE = 'P-VALUE';
ATB = ABS(TB);
TPROB = PROBT(ATB,DF,0);
PVALUE = 2*(1 - TPROB);
RUN;
PROC PRINT LABEL NOOBS;
PROC PLOT DATA=P1;
  PLOT Y*X = '*' YHAT1*X = 'P' / OVERLAY;
PROC PRINT DATA=P1;
TITLE1 '              ';
TITLE2 '              ';
TITLE 'FIT FOR THE MODEL Y = B0 + B1*MIN(X,X_MAX) + B2*MAX(0.,X-X_MAX)';
TITLE2 "WHEN X_MAX = &XXMAX (OPTIMAL PLATEAU)";
RUN;
/*
SYMBOL1 V = STAR;
SYMBOL2 I = J;
PROC GPLOT DATA=P1;
  PLOT Y*X  YHAT1 * X / OVERLAY;
```

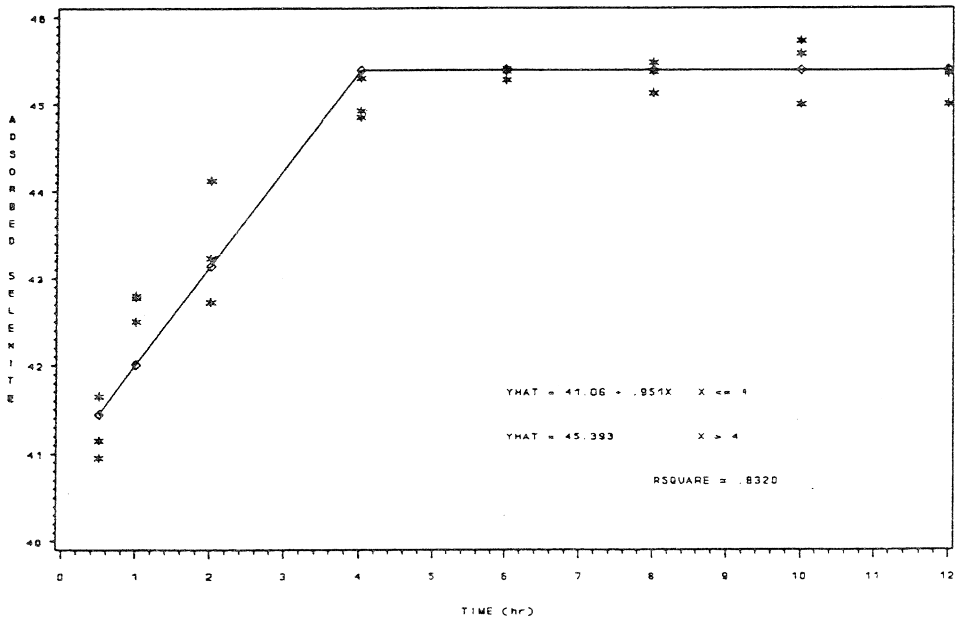
Table 2. Selected properties of the Huckleberry, Helmer, and Santa silt loam soils.

Soil	Taxonomic Classification	pH*	NaF pH	oxalate-extractable Al	Fe
Huckleberry	Typic vitrandept	6.6	9.9	19.5	7.6
Helmer	Andepic Paleboralf	6.5	9.3	10.2	5.9
Santa	Ochreptic Fragixeralf	6.0	7.8	1.7	4.8

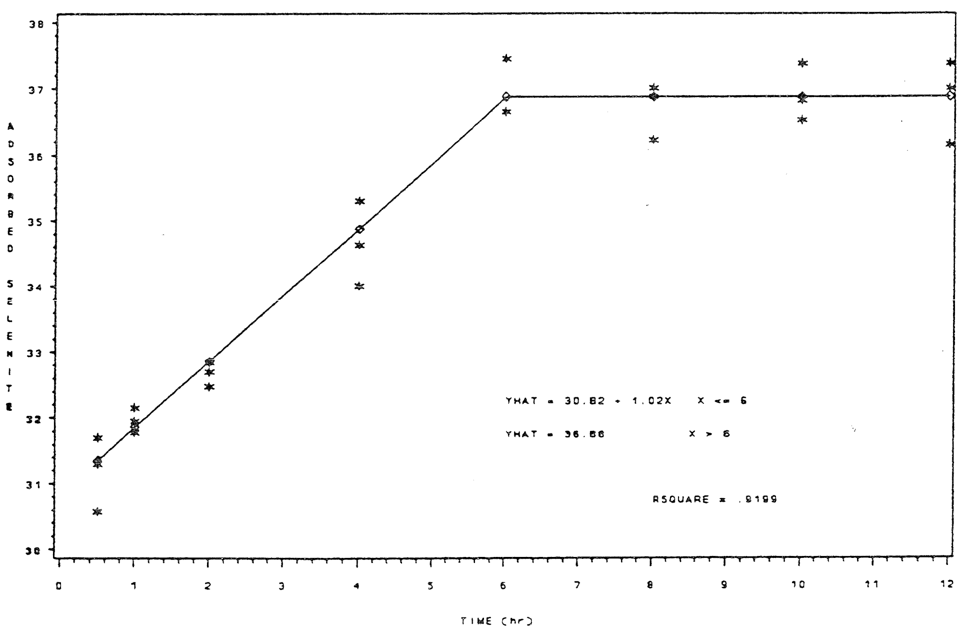
* saturated water-paste

Figure 2. Adsorption Response for (a) Huckleberry (b) Helmer (c) Santa Soils and (d) Overlay of Mean Responses.

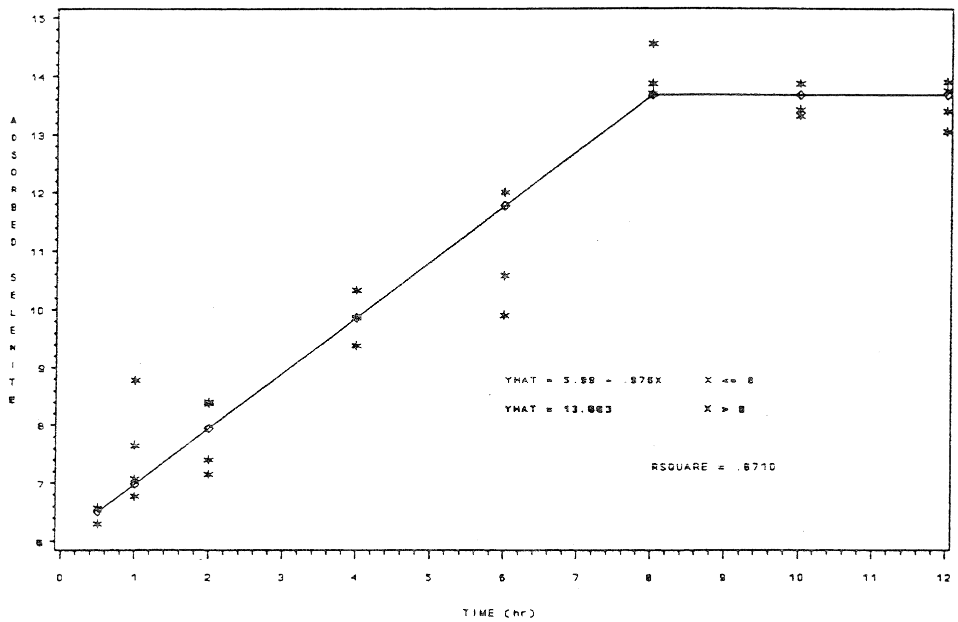
a.



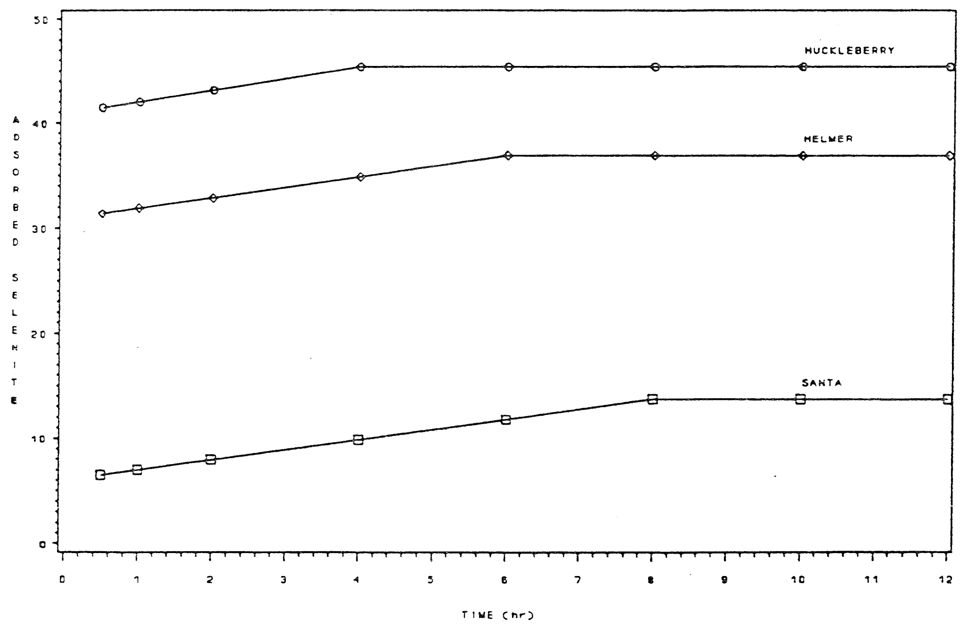
b.



c.


YHAT = 5.99 + .976X X <= 8
YHAT = 13.863 X > 8
RSQUARE = .6710
ADSORBED SELENITE
TIME (hr)

d.


HUCKLEBERRY
HELMER
SANTA
ADSORBED SELENITE
TIME (hr)