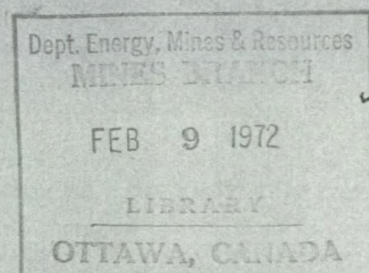




DEPARTMENT OF
ENERGY, MINES AND RESOURCES
MINES BRANCH
OTTAWA

THE DETERMINATION OF CALIBRATION FUNCTIONS FOR
THE OPTICAL EMISSION SPECTROCHEMICAL ANALYSIS OF STEEL
BY MEANS OF MULTIPLE REGRESSION CALCULATIONS



VON PAUL HÖLLER, KURT SCHNICK, KARL-HEINZ GALDA AND
CHRISTIAN THOMA COMMUNICATION FROM THE STRANSKI-INSTITUT
FÜR METALLURGIE DER HÜTTENWERK OBERHAUSEN AG

(TRANSLATED FROM GERMAN BY A. H. GILLIESON, AUGUST 21, 1971)

MINERAL SCIENCES DIVISION

AUGUST 1971

© Crown Copyrights reserved

Available by mail from Information Canada, Ottawa
and at the following Information Canada bookshops

HALIFAX

1735 Barrington Street

MONTREAL

Æterna-Vie Building, 1182 St. Catherine St. West

OTTAWA

171 Slater Street

TORONTO

221 Yonge Street

WINNIPEG

Mall Center Bldg., 499 Portage Avenue

VANCOUVER

657 Granville Street

or through your bookseller

Price *50* cents Catalogue No. *M38-3/274*

Price subject to change without notice

Information Canada

Ottawa, 1971

Mines Branch Information Circular IC 274

THE DETERMINATION OF CALIBRATION FUNCTIONS FOR THE OPTICAL
EMISSION SPECTROCHEMICAL ANALYSIS OF STEEL BY MEANS OF
MULTIPLE REGRESSION CALCULATIONS

by

Von Paul Höller, Kurt Schnick, Karl-Heinz Galda and Christian Thoma
(Translated from German by A. H. Gillieson, August 21, 1971)

FOREWORD

(by Translator)

Apart from its general interest for the application of regression calculations to spectrochemical calibration procedures, the details described are of particular value for emission spectrometric analysis of ferrous alloys by the direct-reading vacuum spectrometer, and it was considered that the information would be of interest in the production control of the Canadian steel industry.

SUMMARY

Characteristic features of regression calculation in its application to calibration in spectral analysis. Relevant fundamentals for the development of approximations and delimitation of approximations by means of regression analyses. Elimination of non-significant factors of influence by means of the t test. Stepwise regression calculation with widened approximations for each factor in order to optimize the reduced approximations with the residual scatter as criterion. Discussion of the results in the uniform calibration of a broad steelmaking program. Application of this method to process control.

*Research Scientist, Spectroscopy Group, Mineral Sciences Division,
Mines Branch, Department of Energy, Mines and Resources, Ottawa,
Canada.

THE DETERMINATION OF CALIBRATION FUNCTIONS FOR THE OPTICAL
EMISSION SPECTROCHEMICAL ANALYSIS OF STEEL BY MEANS OF
MULTIPLE REGRESSION CALCULATIONS

by

Von Paul Höller, Kurt Schnick, Karl-Heinz Galda and Christian Thoma
Communication from the Stranski-Institut für Metallurgie der Hüttenwerk
Oberhausen AG

Archiv für das Eisenhüttenwesen, 41, 811, (1970).

Translated from German by A.H. Gillieson, August 21, 1971.

To optimise the desired values as functions of those measured,
the classical aid for a collection of measured values is regression analysis.
The general expression for every regression calculation is:

$$\sum_{i=1}^N \left[Y_i - f(x_i, y_i, z_i) \right]^2 = \text{minimum} \quad (1)$$

Where Y are the desired values and f the function sought for the measured
values x, y, z, For f, expressions must be developed, e.g.,

$$f = a_0 + a_1x + a_2y + a_3z + \dots \quad (2)$$

Following the method of the analysis of variations, the coefficients
are optimised so that the sum of the squares of the deviations are minimized
(cf. equation 1). For the goodness (of fit) of each expression there are
statistical measures such as the precision -- from the resultant of the scatter
of the relevant residual scatters -- the residual scatter itself, and the
so-called t values which are measures of the significance of the particular
interaction quantity under consideration.

In conjunction with electronic computers, regression analysis gives an opportunity to experiment mathematically with various expressions. For further details of this method one should refer to the literature. It should, however, be mentioned that particularly comprehensive expressions are by no means the best. The residual scatter can be unfavourably influenced by insignificant interaction values, according to the t-test. In the choice of expressions it is necessary to consider relationships either based on laws of nature and measurement procedure or derived from experience in order to avoid a misdirected search for suitable expressions, based solely on statistical mathematical measures.

In an almost ideal way, (the help of) regression analysis is suitable for the acknowledged problem of calibration in optical-emission and X-ray fluorescence spectrochemical analysis. If, up till now, it has not or only rarely been used, this is chiefly because a vast amount of commercial material is required and because, in many cases, calibration curves are obtained which, at least in industrial laboratories, can be used only in an on-line computer. To this must be added that, only in the last few years, have there been to hand suitable computer programmes which allow of mathematical experimenting with various expressions and iterative execution of the regression analysis (with stepwise introduction of a variable) in reasonable time and at a reasonable cost.

Besides the already mentioned optimum derivation of the calibration functions, regression analysis offers the advantage that one can set up calibration functions for greater ranges of contents and a greater multiplicity of combinations of contents, i. e., types of samples. The prerequisite for this is that one introduces an adequate number of variables and appropriate expressions for each variable. For estimation of the number of measured values necessary for each variable, the following holds:

regression coefficient a , associated with the variable x , is

$$\Delta a_1 = \frac{t \cdot \sqrt{\sum_{i=1}^N [Y_i - f(x_i, y_i, z_i \dots)]^2}}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{N - P - 1}}$$

$$= t \cdot \sigma_R \sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \quad (3)$$

where σ_R = residual scatter.

If the extent of variation of the interacting value x , $\sum (x_i - \bar{x})^2$, is small, i. e., if one has chosen a small content range for the calibration, then one needs a large number N of measurements to obtain a small Δa_1 . The smaller the product of the t value and the residual scatter is, the smaller for a given Δa_1 can be the content range, $\sum (x_i - \bar{x})^2$, and N . The same considerations apply to the case where x is a transformed interaction value.

For numerical estimation, one introduces

$$Q = \frac{\sqrt{\sum (Y_i - y_i)^2}}{\sqrt{\sum (x_i - \bar{x})^2}}$$

and begins with Q values less than unity; then, for example:

with	$P = 5,$	$Q = 0.1$
and with	$N = 31,$	$\Delta a_1 = 0.2/\sqrt{25}$
		$= 0.04$
	$N = 55$	$\Delta a_1 = 0.2/\sqrt{44}$
		$= 0.03$
	$N = 106$	$\Delta a_1 = 0.2/\sqrt{100}$
		$= 0.02.$

The greatest effect is clearly not to achieve an increase in the number of measurements but to enlarge the extent of the variation of the

contents of those elements which have real influence. If one doubles the extent of the variation, one achieves the same effect as through a fourfold greater number of tests.

In what follows, the calibration by means of regression analysis, for the case of steel analysis with an optical-emission vacuum spectrograph, will be dealt with.

(TRIAL) EXPRESSIONS

In optical emission spectrochemical analysis, calibration has little basis in physics: it is derived empirically from measurement of samples of known content. Nevertheless from the physics and chemistry of volatilization and excitation, from the measurement of the spectra, and from practice, there result a number of guide lines for the expressions.

1. Practice demonstrates that the calibration curves are not only continuous but also continuously differentiable more than once. Therefore, polynomials are suitable expressions.

2. In as much as line overlap on the secondary (exit) slit is unavoidable, the interfering element (effect) is additive and, to a first approximation, can be applied linearly.

3. The self-absorption leads to a continuous bending of the calibration curve and thus can reasonably be taken into account by a polynomial expression.

4. From several investigations, it is recognised that, with spark excitation of the spectra in an argon atmosphere, there originate a number of interelement effects which are favoured by the highly dispersed heterogeneous non-metallic phases -- e. g., sulphide, oxide, carbide etc. -- as cathode of the discharge. In the consideration of these interelement effects, reciprocal-effect terms which are products of the measured values of both the elements belonging to the compound suggest themselves. In comparison to other techniques, e. g., XRF, interelement effects are very small. Therefore, linear expressions or simple reciprocal terms are good approximations for the elements concerned.

These guide-lines lead to two expressions:

$$Y_k = \sum_{j=0}^2 j a_k^j \cdot (x_k)^j + \sum_{\substack{n=1 \\ n \neq k}}^{16} a_k^n \cdot x_n \quad (4)$$

For example: Carbon (k = 1)

$$Y_1 = a_1^0 + a_1^1 \cdot x_1 + a_1^2 \cdot x_1^2 + a_1^2 \cdot x_2 + a_1^3 \cdot x_3 + \dots + a_1^{16} \cdot x_{16}$$

$$Y_k = \sum_{j=0}^3 j a_k^j \cdot (x_k)^j + \sum_{\substack{n=1 \\ n \neq k}}^{16} a_k^n \cdot x_n + \sum_{n \neq k} b_k^n \cdot x_k \cdot x_n \quad (5)$$

For example: carbon (k = 1)

$$Y_1 = a_1^0 + a_1^1 \cdot x_1 + a_1^2 \cdot x_1^2 + a_1^3 \cdot x_1^3 + a_1^4 \cdot x_4 + a_1^6 \cdot x_6 + b_1^4 \cdot x_1 \cdot x_4 + b_1^6 \cdot x_1 \cdot x_6$$

The number of terms in the sum is suitable for the example treated later. In the expression for the first definition, equation (4), the measured values for the element in question are considered as a second-degree polynomial and all interelement effects as purely linear (no reciprocal effects). With this expression, the regression analyses are carried out with the goal of determining the significant elements producing interelement effects. This is not achieved by optimization of the residual scatter. For this is applied, the second extended expression (5) in which the measured values of the element in question are considered as a third-degree polynomial, with a simple interelement-effect in the second term and with a plausible reciprocal effect in the third term.

SAMPLES

The whole steel-programme of the Oberhausen foundry as well as types of imported trade samples were taken into account in the choice of (test) samples. These are: unalloyed steels up to 1% C, ship building steels, general building steels complying with DIN 17100, automatic (machining) steels (S, Pb, Te-alloying), weather-resistant steels, alloyed construction-steels for high-strength, heat resistant and low-temperature toughplate, preliminary test samples from the SM-furnace, 500 g cone samples taken from melts, poured into copper molds, and "killed" as necessary with aluminum, and samples from the various finished products with different additives and proportions, as well as samples from finished products remelted with addition of aluminum in the ore furnace. The samples selected for the investigation covered evenly the range of composition given in Table 1. In column 5 there are distinguished those elements considered as "target" and influencing elements (Z) or those considered only as influencing elements (be).

MEASUREMENT CONDITIONS

Two emission vacuum spectrometers (ARL - Quantovac 15000 and 17,500) were available for the determinations. The following excitation conditions were chosen throughout:

15 μ F, 360 μ H, 10 Ohm, 800 V, 4 litres Ar/min/over
critically damped spark discharge

Burn time $\sim 5 \times 10^{-4}$ sec;
 Spark repetition rate 50 per sec.;
 Anode: Silver electrode (6 mm diam);
 Cathode: sample;
 Cathode-Anode Separation: 6 mm;
 Analysis lines: refer to Tables 1 and 2;
 Prespark time: 20 sec;
 Integration time: 10 sec.

TABLE 1

Range of Element Content

Element	Range of Content (%)	"Target" or Influencing Element (Z, be)
C	0.01 - 2.0	Z
Si	0.01 - 3.5	Z
Mn	0.01 - 2.0	Z
P	0.001 - 0.4	Z
S	0.001 - 0.4	Z
Al	0.001 - 0.2	be
As	0.001 - 0.15	be
Co	0.01 - 0.3	be
Cr	0.01 - 3.5	be
Cu	0.01 - 1.5	Z
Mo	0.01 - 1.2	be
Ni	0.01 - 11.0	be
Nb	0.01 - 0.3	be
Sn	0.001 - 0.15	be
Ti	0.01 - 0.3	be
V	0.01 - 1.0	be

The line intensities were measured photo-electrically at a constant time of integration. By switching in an impedance-changer which is standardised electrically at attenuation 1:1 (no individual sensitivity - or zero-adjustment for the individual channels), the charge is digitalized and recorded. At intervals of 25 to 30 sparkings, standard samples were measured.

TABLE 2

Example of the Calibration Equation with Stepwise Determination of the
Optimum Coefficients

Element	Influence Factor No.	Normalized Coefficient β_i	Coefficient α_i	Inner Correlation Coefficient	t-values	Correlation Coefficient B	Constant Term α_0	Residual Scatter σ_R
C ²	02	0.49583	392.23	0.9983	5.796			
C	01	0.38179	135.32	0.9923	9.520			
C.Si	07	0.04423	155.55	0.9048	3.879			
S	05	0.02588	30.50	0.8043	3.254			
C ³	03	1.13232	248.90	0.9947	2.716			
C.S	06	-0.01497	-44.27	0.9469	0.980			
Si	04	-0.00591	- 3.05	0.8936	0.548	0.9973	-6.94	1.475
C ₂	01	0.38350	135.93	0.9925	9.610			
C ²	02	0.50493	399.43	0.9982	6.029			
C ³	03	1.22545	235.98	0.9944	2.670			
S	05	0.02732	32.20	0.7802	3.647			
C.S	06	0.01816	-53.70	0.9379	1.288			
C.Si	07	0.03842	135.11	0.3006	9.151	0.99773	-7.17	1.472
C ₂	01	0.38076	134.96	0.9922	9.583			
C ²	02	0.48735	385.53	0.9981	5.887			
C ³	03	0.13305	250.28	0.9943	2.849			
S	05	0.01933	22.78	0.3029	4.589			
C.Si	07	0.03850	135.41	0.0304	9.156	0.99771	-6.16	1.474

(Continued on next page.) -

TABLE 2 (Cont'd)

Element	Influence Factor	Normalized Coefficient	Coefficient	Inner Correlation Coefficient	t-values	Correlation Coefficient	Constant Term.	Residual Scatter
	No.	β_i	α_i			B	α_o	R
C ₂	01	0.27437	97.25	0.9379	19.068			
C ²	02	0.71998	569.55	0.9338	51.647			
S	05	0.01660	19.56	0.2648	3.971			
C.Si	07	0.03860	135.74	0.3004	9.008	0.99761	-4.35	1.502
C ₂	01	0.32974	116.88	0.9240	21.256			
C ²	02	0.67715	535.67	0.9240	42.649	0.99656	-3.65	1.792
C ²	02	0.99414	786.43	0.000001	126.410	0.98831	6.80	3.297

EXECUTION OF THE COMPUTATION

The zero point and sensitivity deviations of the measuring equivalent are compensated in a preliminary programme, by normalization coefficients. For this, there are used precise previously given theoretical values for the upper and lower standard samples for each channel and the primary measured values occurring at definite intervals, for these samples. Theoretical and primary values are referred to an internal standard. If one defines with s_o and s_u the upper and lower theoretical values and with i_o and i_u the upper and lower primary values, then the normalization coefficients α and β are given by the following expression:

$$\begin{aligned} s_o &= \alpha \cdot i_o + \beta \\ s_u &= \alpha \cdot i_u + \beta \end{aligned} \quad (6)$$

The measured values referred to the internal standard, are normalized by these coefficients. In addition, the preliminary programme furnishes a test of the reproducibility. At small contents, absolute and, at higher contents, per cent maximum permissible deviations of the reproducibility are yielded for the individual measuring channels. In the further computation, there are permitted only means of the measured values which are reproducible within these previously given tolerances.

The expression in accordance with equation (4) is used to ascertain the significant element influences. Then, with respect to all terms of the summation, a_k^n is calculated. Finally the t values are determined by step-wise building up of the expression, i. e.; first for the expression

$$a_k^0 + a_k^1 \cdot x_k, \text{ then for } a_k^0 + a_k^1 \cdot x_k + a_k^2 \cdot x_k^2 \text{ and so on.}$$

In this way is ascertained which term of the summation has the smallest t value, i. e., is the least significant. Beginning with these terms, the regression equation is now built up stepwise. The coefficients are next calculated using the reduced equation and then, according to the previously described iteration, the t values. Again there results a term with the smallest t value which is left out of the next reduced expression and so on. In this stepwise building of the regression equation, at each step the correlation coefficient and residual scatter are determined along with the t values.

As an illustration, an example of an original printout of the computer is reproduced in Table 2. In column 3, there are shown the coefficients a_k^n ; in column 5, the t values for the elements entered in column 1; in column 2, the normalizing coefficients; and, in column 4, the inner correlation coefficients. The normalized coefficients are the coefficients (column 3) derived from the initial scatter. The inner correlation coefficient indicates the dependence of the influence factor under consideration (column 1) on the remaining (factors). The last columns of the printout contain the correlation coefficient, the absolute term of the expression, and the residual scatter.

The results obtained in this way are illustrated by three examples (Figures 1, 2 and 3). The logarithm of the t value, the correlation coefficient, and the residual scatter are plotted as ordinates. In the abscissa (upper edge of the figure) are given from left to right which elements in the appropriate approximation are no longer taken into account. The "target" elements are carbon (Figure 1), manganese (Figure 2), and phosphorus (Figure 3). It is recognized that both the correlation coefficient and the residual scatter are scarcely changed with omission of elements of little significance. A real decrease of the correlation coefficient and of the residual scatter appear only in the last steps of this successive regression analysis. The t value is a considerably more sensitive indicator of the influence of an element than is the correlation coefficient or the residual scatter.

A tabular summary of the examples reproduced in Figures 1 to 3 as well as similar computations for other "target" elements is contained in Table 3.

TABLE 3

Interelement Effects According to Linear Regression (Increasing Numbers Correspond to Decreasing Weight)

Target Element	A°	Influencing Element									
		C	Mn	P	Si	S	Al	Co	Ni	Mo	As
C	1930				2	1					
Si	2124	3						2		1	
Mn	2933					1		3	2		4
P	1783	1	4			3			2		
S	1807	1		3			2				
Cu	3274			1							

For each target element, the wavelengths of the spectral lines used, are arranged according to the significance of the influencing element quoted. The strong interelement effects -- distinguished in the table by "1" -- are without exception (already) known; in the case of copper-phosphorus one is dealing with a line interference. The weaker effects, 2, 3, and 4 are in (this) general form, not known up till now. They could well be included widely in the subdivision of calibration curves into "quality" groups.

In the computations of Table 1, all the samples are considered according to Section 3.

Finally the regression computations for optimization of the residual scatter must still be examined. For the final determination of the calibration functions, there were taken into account -- apart from some exceptions -- only those elements near the target element which, in the previously described regression analysis with the simplified expression (4), had shown themselves significant -- $t \geq 2$. The expression in accordance with equation (5) was used with linear interelement effect and simple reciprocal-effect terms. The individual expressions were judged only by

the residual scatter and not by correlation coefficient or by t values. It does not depend on which part of the initial scatter can be explained by the expression (correlation coefficient) but which residual scatter persists after use of the expression.

The execution of the computation is the same as with the first approximation, apart from (the fact) that a third-degree polynomial is considered for the measured values and a plausible reciprocal effect for the interelement effect. A number of examples are reproduced in Figures 4 to 9, and the buildup of a calibration function for the 6 basic elements in steel is given in Table 4. Figures 4 to 6 apply to the cone samples taken from the melt after relatively long cooling of the sample and correspondingly coarse non-metallic segregations. In the examples in Figures 7 to 9, one is in contrast, dealing with small samples which were remelted in the arc furnace and quickly cooled. Both from the absolute values of the residual scatter and from the significance of the reciprocal terms, it is clear that the metallurgical after-treatment has a favorable influence on the calibration in a way expected and described earlier. It is particularly noticeable in the great difference of the t value in the reciprocal-effect term for manganese-sulphur. This difference can undoubtedly be explained by the finer MnS precipitation during the faster cooling. This observation should have indicated that regression analysis is of great help in the quantitative evaluation of different measurement procedures or types of samples.

From Table 4 it can be inferred that the steel programme outlined in Section 3 can be homogeneously calibrated by an expression in accordance with equation (5), if 8 terms are allowed for in the calibration equation. The reciprocal effects play an important role therein. Polynomials up to the third degree in the target element are adequate. Such calibration equations are suitable for on-line computer equipment but not, on the other hand, for evaluation without automatic computation aids.

TABLE 3

Calibration Function Matrix with Consideration of Reciprocal Effects (Increasing Numbers Correspond to Decreasing Weight)

Target Element	Å	Expression x^1			Reciprocal Effect with						Linear Interelement Effect							
		i = 1	2	3	C	P	Si	S	Ni	Mo	C	Mn	S	Al	As	Co	Ni	Mo
C	1930	2	1	5			3	6					4					
Si	2881	1	3	7						4						2	6	5
Mn	2933	2	1				4	3							5	6		
P	1783	1	2		3			6			5							4
S	1807	1	4		2						3	5		6				
Cu	2135	2	1			3						4				5		

SUMMARY

The multiple regression analysis is the optimum mathematical aid for determining calibration functions for spectrochemical analysis. The application of this method is explained by an example from the optical emission spectrochemical analysis of steel with a vacuum spectrometer. Starting from physical, chemical guide-lines and from experience, formulae were set up and, in a first regression analysis, defined by simple expressions. For each influence factor thereby shown to be significant, the optimum functions and coefficients were determined with an expanded expression for each influence factor in a second approximation by similar stepwise regression calculation. In this, the residual scatter served as the criterion for optimization. For an extensive steel programme, a uniform calibration is possible with polynomials up to third-degree for the target elements as well as up to 4 reciprocal-effect - or linear interelement terms. Such calibration equations can be used in an on-line computer installation. The application of regression analysis to method testing is discussed in an example.

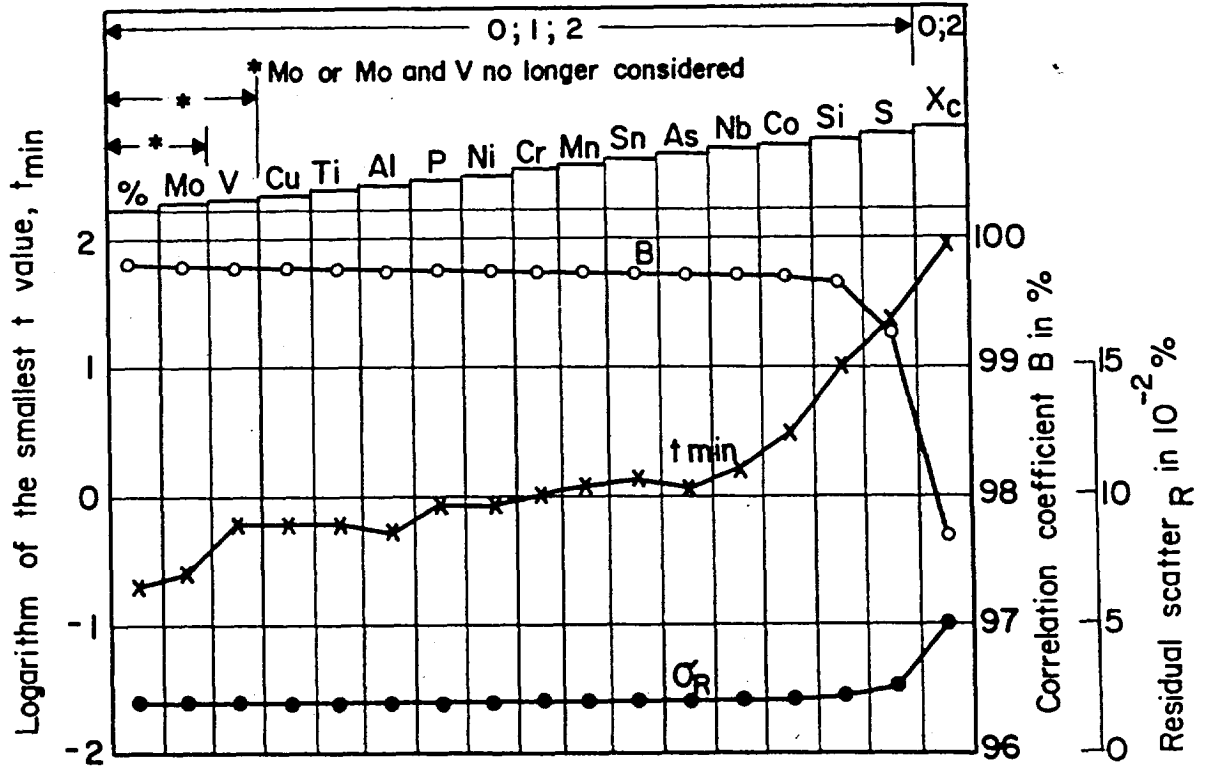


Figure 1. - Interelement effects in carbon determination.

First definition of the equations by stepwise regression.

Target value quadratic, third "partner" linear. The values in Figures 1 to 9 are, for clarity, joined together by a curve.

Abscissa: left: logarithm of the smallest t-value, t_{\min} ,
 right inner: correlation coefficient B in $\%$,
 right outer: residual scatter R in $10^{-2}\%$,
 *Mo or Mo and V no longer considered.

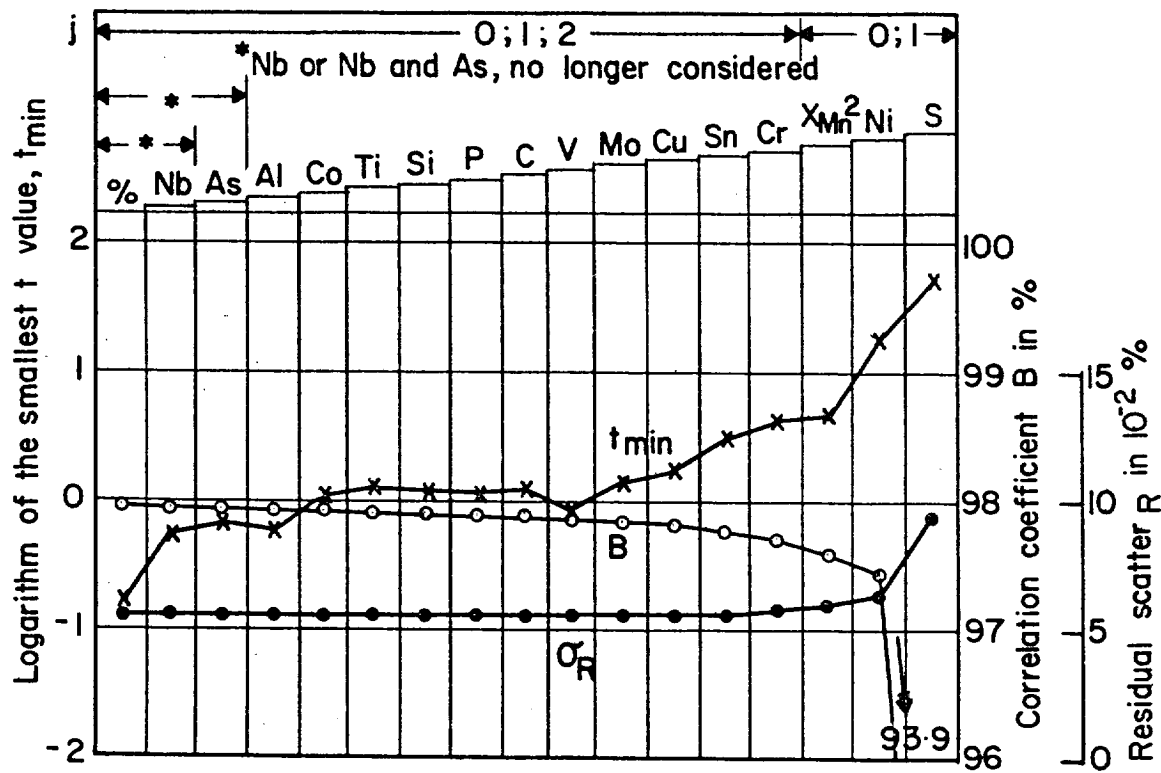


Figure 3. - Inter-element effects in phosphorus determination.

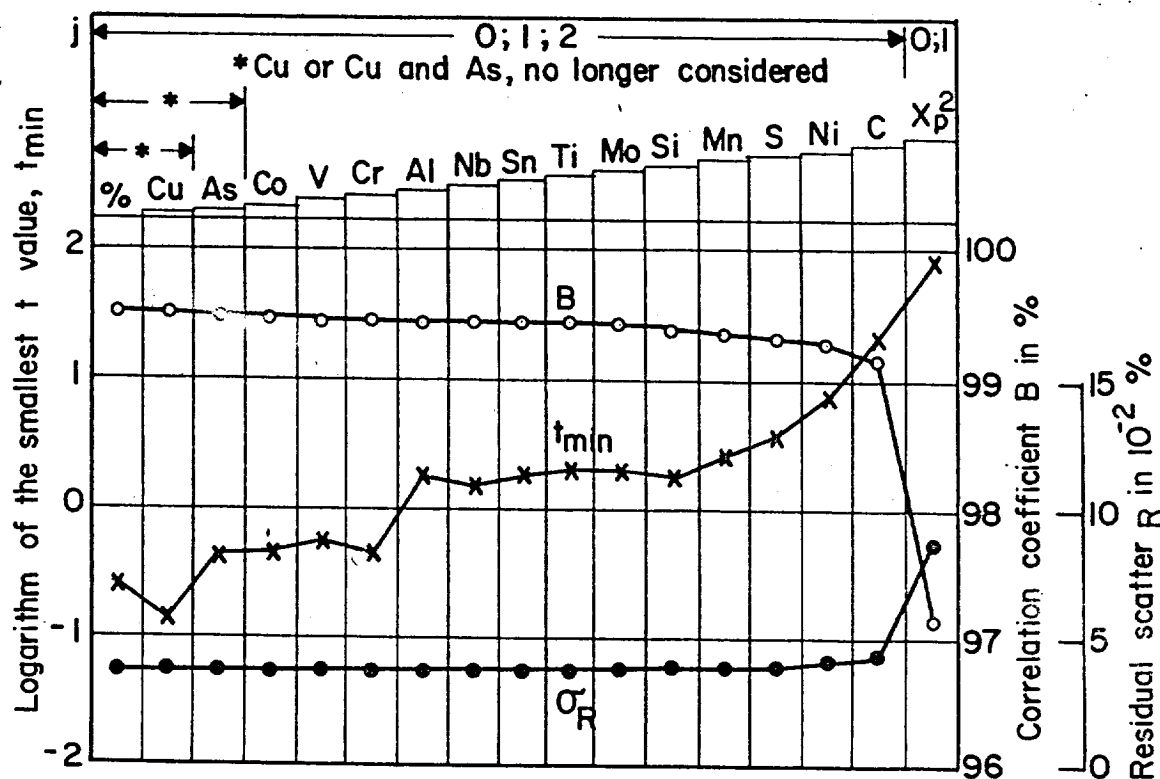


Figure 2. - Inter-element effect in manganese determination.

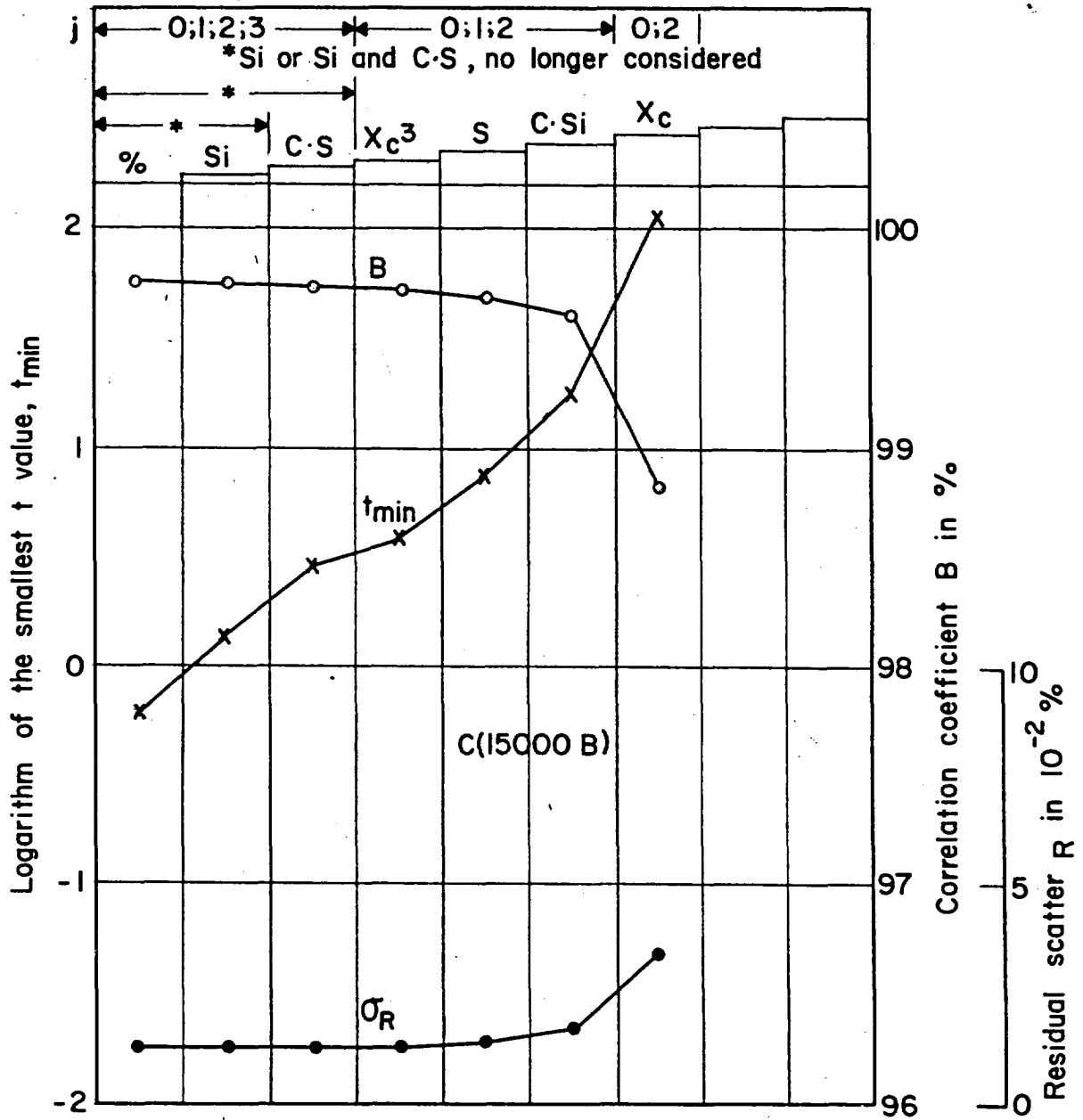


Figure 4. - Carbon content of cast samples "Polynome 3 Grades" - Reciprocal effect and linear interelement effect of two components with stepwise regression.

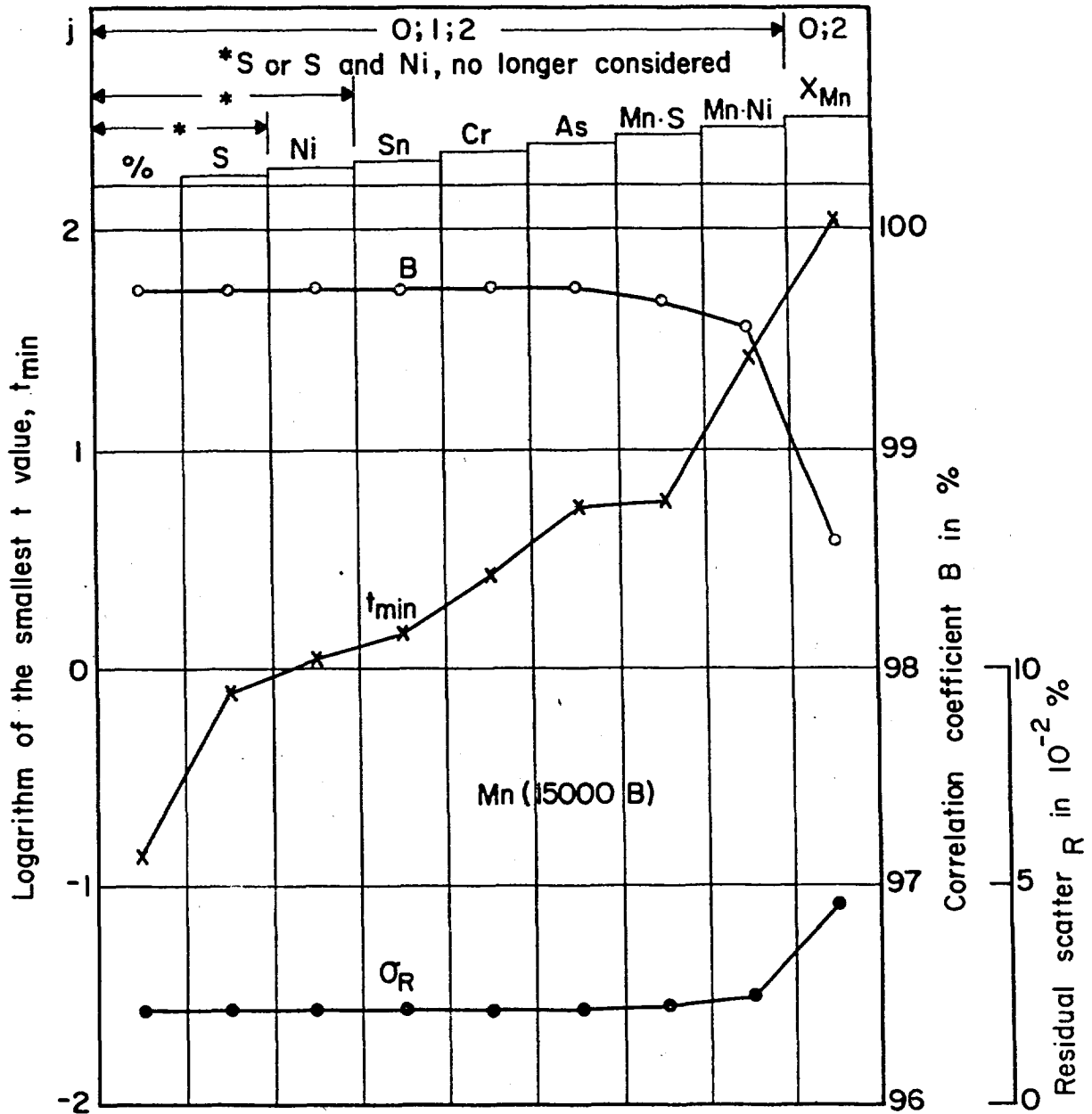


Figure 5. - Manganese content of cast samples "Polynome 2 Grades" - Reciprocal effect and linear interelement effect of two components with stepwise regression.

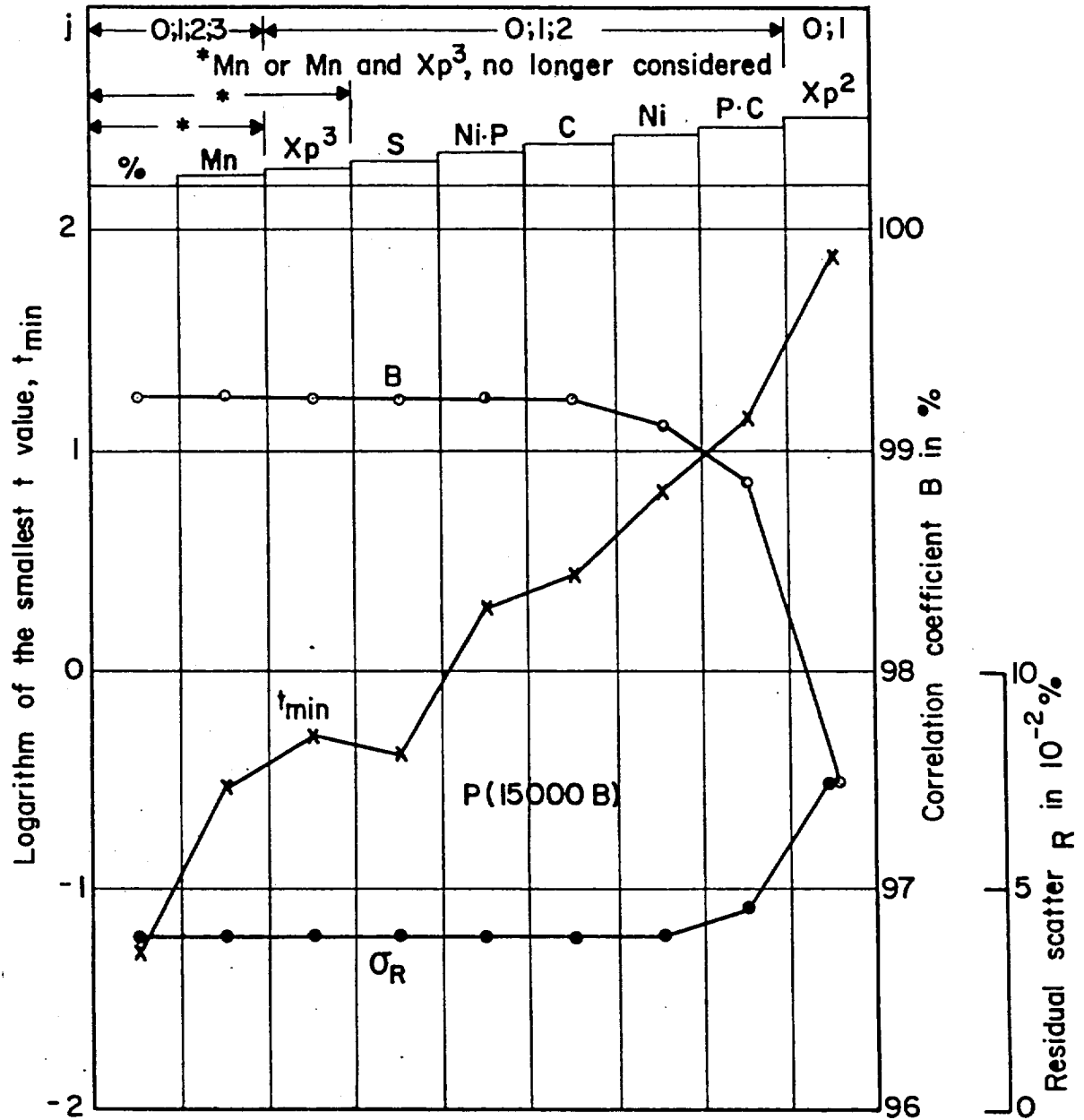


Figure 6. - Phosphorus content of cast samples "Polynome 3 Grades" - Reciprocal effect and linear interelement effect of two components with stepwise regression.

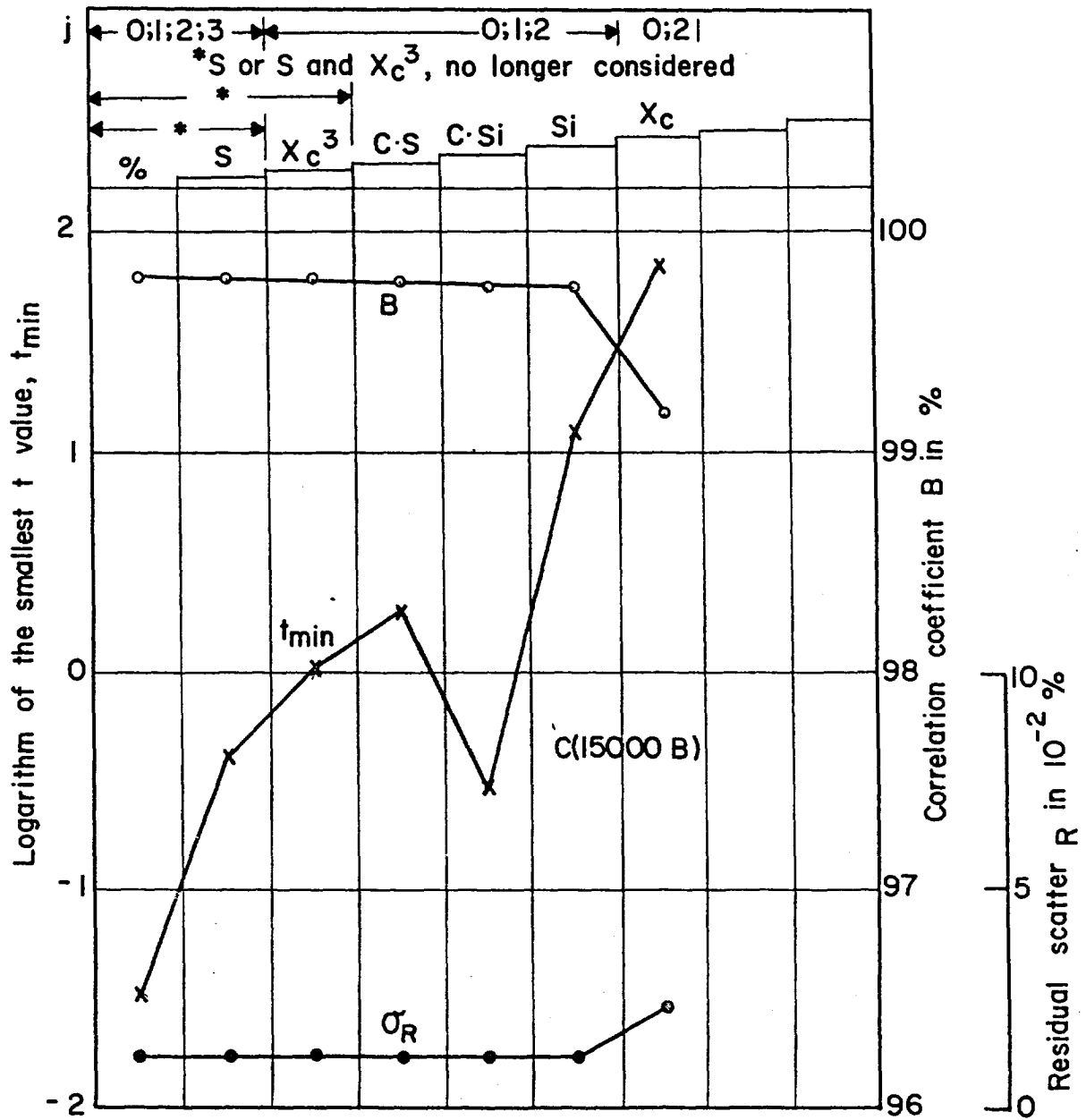


Figure 7. - Carbon content remelt samples "Polynome 3 Grades" - Reciprocal effect and linear interelement effect of two components with stepwise regression.

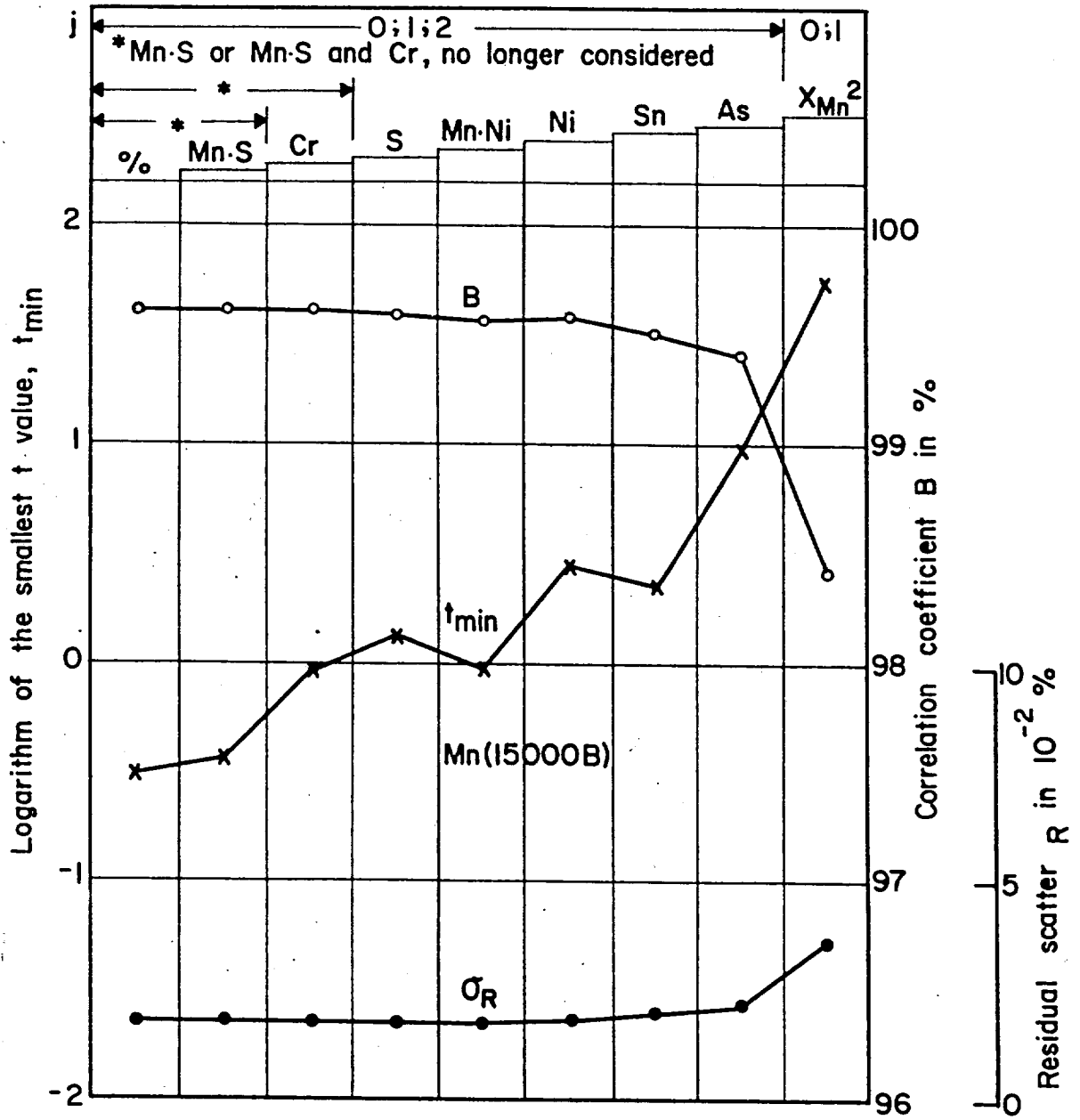


Figure 8. - Manganese content in remelt samples "Polynome 2 Grades" - Reciprocal effect and linear interelement effect of two components with stepwise regression.

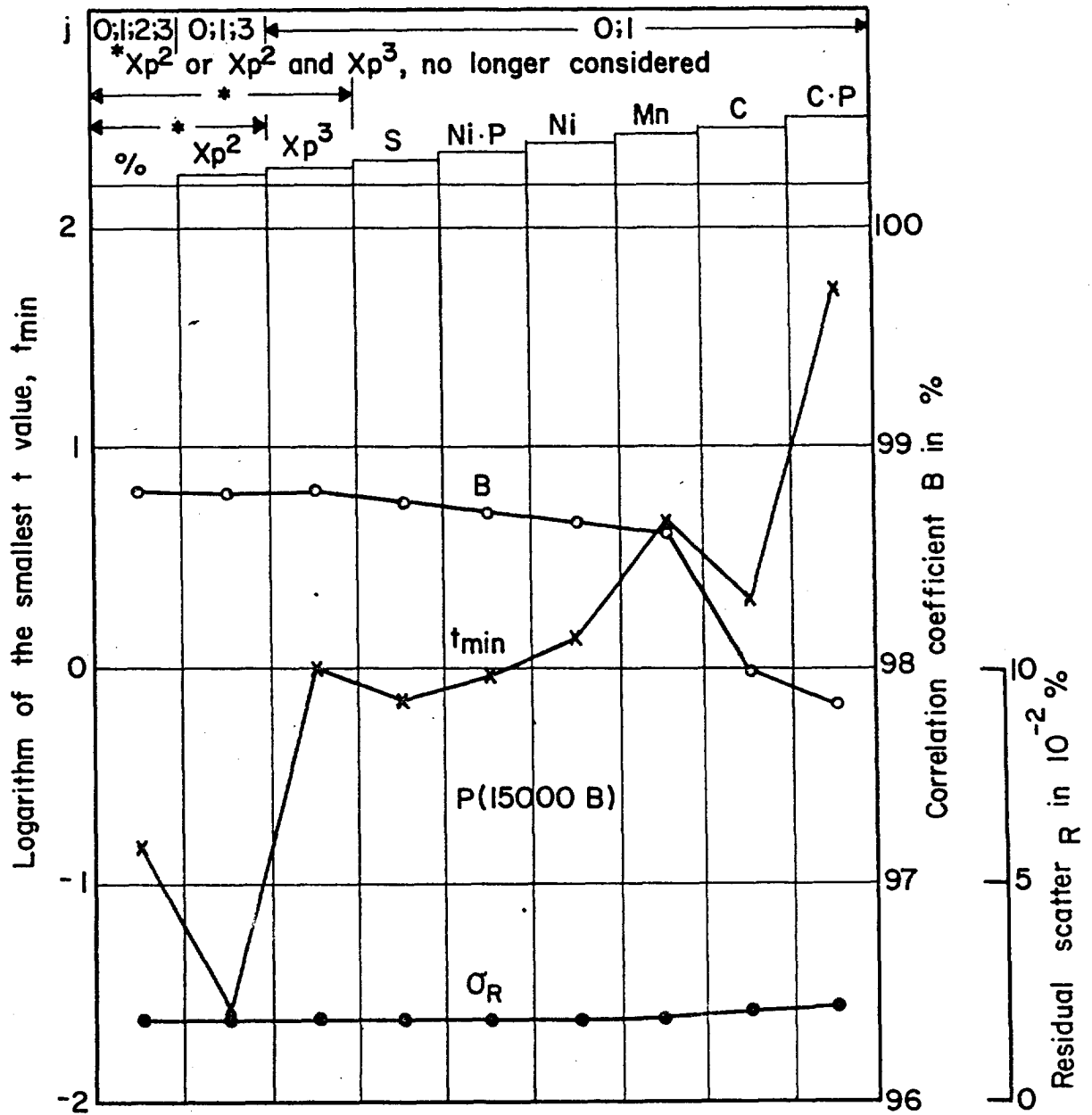


Figure 9. - Phosphorus content in remelt samples "Polynome 3 Grades" - Reciprocal effect and linear interelement effect of two components with stepwise regression.

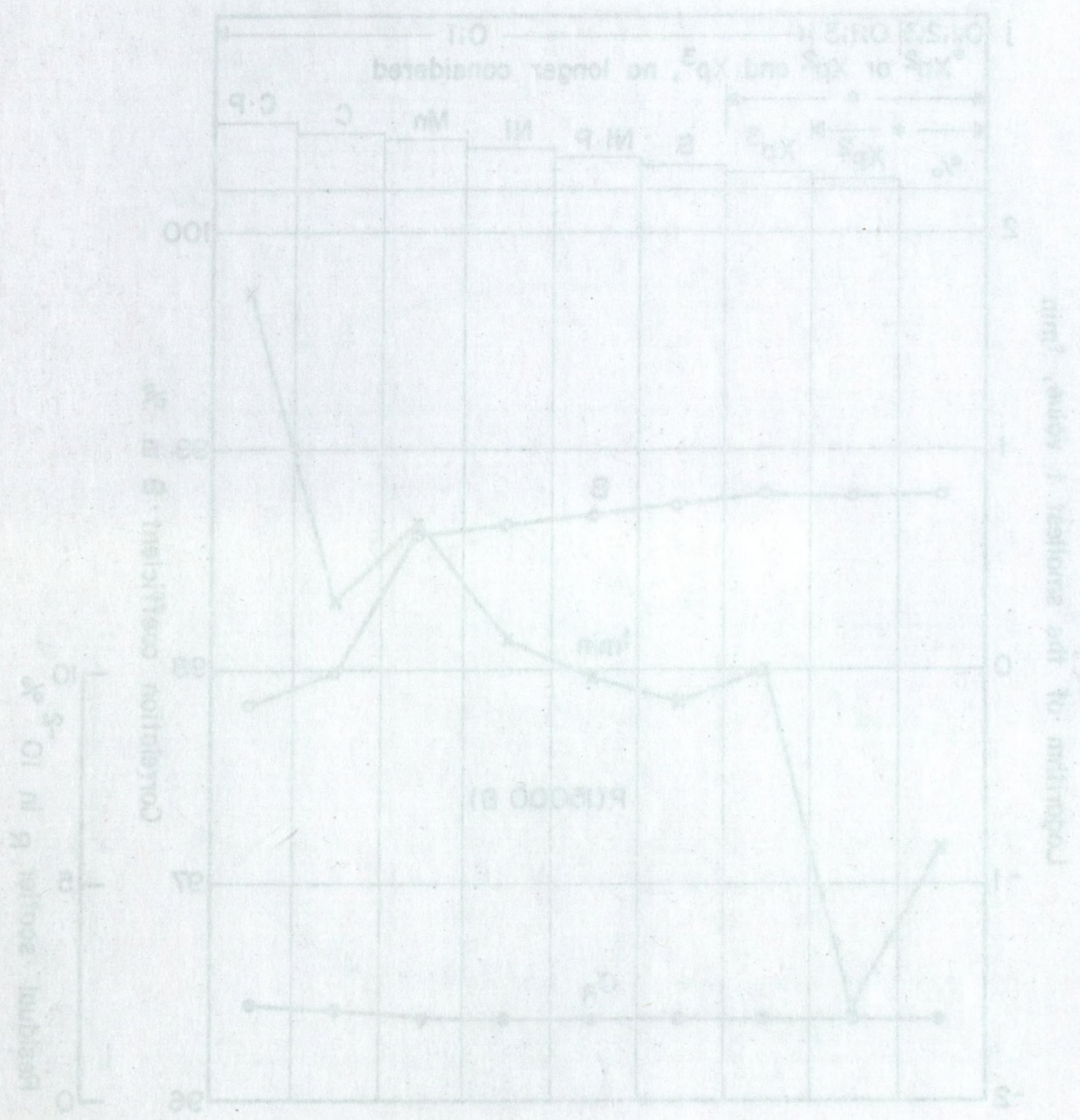


Figure 9 - Phosphorus content in remelt samples "Polymers 3 Grades" - Reciprocal effect and linear interrelationship effect of two components with stepwise regression.