Improving the Control of the Cement Making Process by Applying Multivariate Classical Least Squares Calibration

Jordi-Roger Riba, Joaquim Font, Daniel Rodriguez, Rita Puig

Escola Universitària d'Enginyeria Tècnica Industrial d'Igualada (EUETII). Universitat Politècnica de Catalunya, España.
Plaça del Rei, 15. 08700 Igualada (Barcelona), SPAIN

Mejoras en el Control del Proceso de Fabricación de Cemento Mediante Calibración Basada en el Método de Mínimos Cuadrados Clásicos

Millores en el Control del Procés de Fabricació de Ciment Mitjançant Calibració Basada en el Mètode de Mínims Quadrats Clàssics

RESUMEN

Las técnicas basadas en calibración multivariable permiten el control de procesos químicos a partir de la estimación de propiedades de las muestras analizadas, de modo no destructivo, rápido y económico, a diferencia de las técnicas tradicionales que generalmente son destructivas, lentas o caras. Este trabajo utiliza los mínimos cuadrados clásicos (CLS) para mejorar el control de la fabricación de cemento de aluminato de calcio. Se ha comprobado la eficacia de la técnica CLS mediante la predicción de muestras diferentes de las utilizadas durante la fase de ajuste del modelo de calibración. El tratamiento multivariable de los datos proporcionados por dos métodos instrumentales, fluorescencia de rayos X por dispersión de longitudes de onda (WDXRF) y colorimetría de reflexión (RC) permite una determinación rápida y precisa del ratio Fe_2O_3/FeO en muestras de cemento de aluminato de calcio. El sistema propuesto en este trabajo permite una mejora sustancial en el control del horno, permitiendo al mismo tiempo una mejora en la calidad del cemento producido. Los dos métodos, WDXRF i RC se aplican de forma rutinaria en cementeras a modo de controles continuos del proceso de fabricación del cemento. Por lo tanto, el sistema propuesto no requiere de equipamiento adicional.

Palabras clave: mínimos cuadrados clásicos, cemento de aluminato de calcio, colorimetría de reflexión, fluorescencia de rayos X por dispersión de longitudes de onda.

SUMMARY

Multivariate calibration techniques allow the control of chemical processes by estimating sample properties in a non-destructive, fast and consequently cheap way, which would otherwise require destructive, time-consuming or costly testing. This paper deals with the classical least squares (CLS) which is applied for improving the control of the calcium aluminate cement making process. The CLS has been successful since its performance has been verified through the prediction of samples different from those used to calibrate the model. The multivariate treatment of the data provided by two instrumental methods, wavelength dispersive X-ray fluorescence (WDXRF) and reflection colorimetry (RC) allows for rapid and accurate determination of the Fe_2O_3/FeO ratio in CAC. Rapid determination of the Fe_2O_3 content provides a better control of the kiln, producing higher quality calcium aluminate cement. As both WDXRF and RC methods are currently routinely applied in cement factories as continuous controls of the cement making process, no extra equipment is necessary.

Key words: classical least squares, calcium aluminate cement, reflection colorimetry, wave lenght dispersive X-ray fluorescence.

RESUM

Les techniques basades en el calibratge multivariant permeten el control de processos químics a partir de l'estimació de propietats de les mostres analitzades, de manera no destructiva, ràpida i econòmica en contrast amb les tècniques tradicionals que normalment són destructives, lentes o cares. Aquest treball utilitza els mínims quadrats clàssics (CLS) per millorar el control de la fabricació de ciment d'aluminat de calci. S'ha comprovat l'efectivitat de la tècnica CLS mitjançant la predicció de mostres diferents de les utilitzades durant l'ajust del model de calibratge. El tractament multivariant de les dades proporcionades per dos mètodes instrumentals, fluorescència de raigs X per dispersió de longituds d'ona (WDXRF) i colorimetria de reflexió (RC) permet una determinació ràpida i precisa del ràtio Fe_2O_3/FeO en mostres de ciment d'aluminat de calci. El sistema proposat en aquest treball permet un millor control del forn, permetent alhora una millora de la qualitat del ciment produït. Ambdós mètodes, WDXRF i RC s'apliquen de forma rutinària en fàbriques de ciment a mode de controls continus del procés de fabricació del ciment. Per tant, el sistema proposat no requereix d'equipament addicional.

Mots clau: mínims quadrats clàssics, ciment d'aluminat de calci, colorimetria de reflexió, fluorescència de raigs X per dispersió de longituds d'ona

INTRODUCTION

Cements and binders are substances that can set and harden independently, allowing other materials to be binded together. The most important applications of cement are the production of concrete and mortar. The most commonly used cement in the world is Portland cement because it is a basic component of concrete, mortar, and stucco. This paper however, deals with calcium aluminate cements (CACs), which consist predominantly of calcium aluminates. CACs are hydraulic cements made predominantly from limestone and bauxite and are the major cements second to Portland cement in terms of production levels and use. CACs are resistant to high-temperature, which is why they are used in refractory concretes. They are mainly applied as a binder in monolithic refractories, in particular refractory castables. Other uses of CACs comprise industrial flooring, expansive grouts, chemical resistant concretes and mortars, sewer products, tile adhesives, domestic fireplaces and protective coatings among others. CACs are not used for general structural purposes. They are applied in niche uses justified by their distinctive properties including rapid strength development even at low temperatures, high temperature performance and resistance to a wide range of chemically aggressive conditions [1].

In contrast to Portland cement, in CACs some part of the iron present in the raw materials is reduced to Fe^{2+}. Atmospheric conditions of the kiln affect the proportion of ferrous oxide to ferric oxide of the cement. The color of the cement is highly influenced by the total amount of iron as well as by the ratio of ferrous to ferric iron. When most of the iron is in the oxidized form, the cement is brown in color. When iron is mainly in the reduced form, it is dark grey. Thus, a fast measure of the cement color can provide valuable information for controlling the atmospheric conditions of the kiln where the cement is produced.

The objective of this work is to improve the control of the cement making process by developing a fast and automatic multivariate method to calculate the amount of ferrous iron of the calcium aluminate cement. Rapid determination of the ratio Fe^{+3}/Fe^{+2} provides useful information on the operating conditions of the kiln and improves the control of its atmospheric conditions. Therefore, the quality of the final product can be improved.

It is now well known that there are several X-ray based non-destructive techniques available which provide important information about the crystallographic structure, the physical properties and the chemical composition of different kinds of materials [2,3,4]. In this work the data provided by the wavelength dispersive X-ray fluorescence (WDXRF) analysis is applied, which determines the quantitative elemental composition of the cement samples. This technique does not provide the ratio between Fe_2O_3 and FeO. Furthermore given that the cement's appearances are particularly important for many applications, color measurement is carried out by applying a reflection colorimetry (RC) analysis.

The determination of the Fe_2O_3 and FeO contents proposed in this paper, performed by applying a fast multivariate calibration method, will not involve the measurement of any new data in the task of the quality control. The proposed method will also avoid the need for the tedious volumetric analyses currently used to determine the content of ferrous iron of the cement, as explained in Experimental Results.

The aim of this work is to develop a fast multivariate method for improving the control of the cement making process as well as the final quality of the product. For this purpose, the classical least squares (CLS) multivariate calibration method is applied in order to verify if it is possible to make a fast and accurate prediction of the Fe_2O_3 content of a CAC sample. Thus, the method proposed here takes as input the data supplied by WDXRF and RC methods and outputs the predicted value of the Fe_2O_3 content of the sample analyzed.

Three multivariate calibration models have been studied. In the first one only the information provided by the WDXRF analysis method is considered, in the second only the data obtained from the colorimetric method is taken into account and in the third data provided by both WDXRF and RC methods is used to calibrate the multivariate regression method. Results show an obvious preference for the calibration model including the data provided by both two instrumental methods.

MULTIVARIATE ANALYSIS

Classical least squares (CLS)
In this section, the classical least squares (CLS) multivariate regression method is studied and their theoretical foundations are detailed. This mathematical method has been widely applied in chemometrics and in many other scientific disciplines [5,6,7,8]. CLS is an extension of the well-known univariate linear regression (LR).

When dealing with univariate data, perhaps the most applied linear regression method is the univariate linear regression (LR) which is based on the classical least-squares. Therefore, the former method can be extended to the multivariate case, resulting in the CLS regression, also called multiple linear regression. This method attempts to model the relationship between a set of independent variables x_1, x_2, ..., x_m and a response variable y by adjusting a linear equation to the data measured. Thus, it is assumed that the measured independent variables are related to the dependent variable y by means of a linear relationship which can be written as:

$$y = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + ... + b_m \cdot x_m + e \quad (1)$$

Assuming x is a $m + 1$ components row vector, its transposed representation is given by $x' = (1, x_1, x_2, ..., x_m)$. Additionally, b is a $m+1$ components column vector given by $b' = (b_0, b_1, b_2, ..., b_m)$ and e is the residual or signal noise. Thus, equation (1) can be expressed in matrix form as:

$$y = x'_{(1,m+1)} \cdot b_{(m+1,1)} + e \quad (2)$$

Equation (2) describes a multilinear dependency when dealing with a unique sample. Supposing that a total of n samples have been measured, for the i-th sample, the relationship between the independent variables $x_{i1}, ..., x_{im}$ and the dependent variable y_i can be computed as:

$$y_i = x'_{i(1,m+1)} \cdot b_{(m+1,1)} + e_i \quad (3)$$

being $i = 1,2,...,n$. Equation (3) can be extended to the total set of n samples and can be expressed in matrix form as:

$$y_{(n,1)} = X_{(n,m+1)} \cdot b_{(m+1,1)} + e_{(n,1)} \quad (4)$$

where matrix X can be written as,

$$X_{(n,m+1)} = \begin{vmatrix} 1 & x_{11} & ... & x_{1m} \\ 1 & x_{21} & ... & x_{2m} \\ 1 & ... & ... & ... \\ 1 & x_{n1} & ... & x_{nm} \end{vmatrix}$$

The CLS method determines the values of vector b by applying the least-squares method which is widely used to solve overdetermined systems, that is, systems of equations in which there are more equations than unknowns. The least squares method finds the fit in which the sum of squared residuals is as small as possible; the squared differences between measured data and the predictions of the model are minimized. Thus, the least-squares solution for coefficients vector b is as follows [7,9]:

$$b = (X' \cdot X)^{-1} \cdot X' \cdot y \quad (5)$$

When predicting the concentration of a set of n' new samples different than those used to calibrate the model, equation (6) must be applied,

$$\hat{y}_{(n',1)} = X_{pre(n',m+1)} \cdot b_{(m+1,1)} \quad (6)$$

However, CLS does present some drawbacks, especially when dealing with multicollinear data; when a linear dependence among the independent variables exists. Thus, it has been reported that when the calibration matrix $X_{(n,m+1)}$ is multicollinear, the least squares method leads to poor results [10].

Measure of the goodness of fit

Once the experimental data have been fitted by means of a statistical model it is highly recommended to establish a criterion to measure the goodness of fit [11]. This criterion should measure how well the statistical model fits the set of observations. The F-test is a method commonly used to measure the statistical significance of the regression models [12]. With this approach, the regression model is statistically significant to a fixed confidence level -usually set to 95%- if the following relationship is fulfilled,

$$F_{calc} > F_{tab}(\alpha=0.05; \nu_1 = p; \nu_2 = n-p-1) \quad (7)$$

where,

$$F_{calc} = \frac{MSR}{MSE} = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 / p}{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 / (n - p - 1)} \quad (8)$$

Being $F_{tab}(\alpha; \nu_1, \nu_2)$ the tabulated value of the Fisher-Snedecor distribution with ν_1 and ν_2 degrees of freedom for a fixed significance level $1-\alpha$, MSR the sum of squares due to regression divided by the corresponding degrees of freedom p, MSE the sum of squares due to residuals divided by the corresponding degrees of freedom $n-p-1$, n the number of samples used to build the model, p the number of independent variables ($p = 3$, 5 or 8 in the case under study), y_i the experimental response for the ith sample of the set of samples used to build the model, $\hat{y}_i$ the response of the ith sample predicted by the model and $\bar{y}$ the experimental average response for the samples used to build for the model.

If a data set does not fulfill equation (7) the regression model obtained is not statistically significant to the selected confidence level.

Additionally, the well known R^2 statistical -determination coefficient- that measures the closeness between a predicted regression line and the observed data is used to measure the goodness of fit.

EXPERIMENTAL RESULTS

In this work a total of 39 samples of CAC have been acquired. A subset of 20 samples –calibration set- was used for calibrating the CLS models analyzed, whereas the remaining 19 samples –prediction set- were used to validate these models. The calibration set was used to adjust the CLS model while the prediction samples were used for validation purposes, thus predicting the concentration of a set of samples which are different from those used in the calibration phase. The samples were supplied by the cement factory Cementos Molins SA (Sant Vicenç dels Horts, Spain).

In this work two kinds of samples of CACs have been analyzed. On one hand hard and nodular clinker samples having a granulometry of 2-10 cm, obtained from the kiln were studied, on the other, cement samples obtained by grinding the clinker into a fine powder of less than 200 μm were used.

Characterization of CAC samples was carried out by applying the reflection colorimetry (RC) and wavelength dispersive X-ray fluorescence (WDXRF) methods. While RC determines the color of the samples, giving useful information about the ratio of ferrous to ferric iron, WDXRF provides information about the chemical composition of the cement samples by detecting compounds such as CaO, Al_2O_3, SiO_2, TiO_2, and total iron oxides which are present in the samples.

The color measurement is a fast test, carried out with a reflection spectrophotometer model ColorFlex 45/0 from HunterLab, with 45°/0° measurement geometry. The D65 illuminant was used, and the samples color coordinates were provided in the CIE L*C*h color scale.

Chemical composition was determined by means of WDXRF using a spectrometer from Bruker, model S4 Pioneer. A drawback of the WDXRF method is that it does not allow distinguishing between oxidation states. Thus, determination of Fe_2O_3 content was carried out by means of a wet volumetric method. Cement and grinded samples of the clinker were dissolved with hydrochloric acid, and the amount of ferric iron was titrated with titanium trichloride under an inert atmosphere of CO_2 in order to prevent undesirable air oxidations [13]. Then, the content of ferrous iron of the cement is derived from the difference between the total iron determined by means of WDXRF analysis and the volumetrically determined ferric iron. Application of the wet classic method presents some drawbacks. It is time-consuming, very difficult to automate and is unable to provide on-line data, which is a requisite for applying an automatic control of the kiln atmospheric conditions. Fortunately, RC accepts the same 34 mm-diameter samples employed in the WDXRF analyses.

The variables measured of the calibration set samples and the ones of the prediction set are detailed in Tables Ia and Ib. In these tables variables x_1, x_2, ..., x_8 are respectively, CaO, SiO_2, Al_2O_3, total Fe_2O_3 and TiO_2 composition and L,

Samples	Measured Wet method (%)	WDXRF (%)					RC		
	$y = Fe_2O_3$	CaO	SiO_2	Al_2O_3	Fe_2O_3total	TiO_2	L*	C*	h*
1c	8.05	31.80	4.86	47.64	13.24	2.14	47.33	6.27	97.47
2c	8.50	38.05	4.01	40.67	14.13	1.90	45.57	4.44	93.59
3c	9.38	37.34	3.45	41.68	15.83	1.95	43.92	4.28	92.75
4c	9.50	37.18	3.69	41.62	15.71	1.94	44.04	4.64	90.09
5c	9.70	37.28	3.64	41.49	15.77	1.94	44.30	4.75	88.64
6c	9.90	36.81	3.68	41.85	15.82	1.95	44.37	4.71	93.51
7c	9.94	37.73	3.45	41.56	15.35	1.93	44.67	4.97	89.64
8c	9.95	37.08	3.55	41.71	16.02	1.95	44.29	4.73	92.18
9c	10.00	37.40	3.48	41.67	15.84	1.93	44.31	4.83	92.07
10c	10.00	36.51	3.09	42.84	15.29	1.99	45.40	5.91	89.55
11c	10.45	37.33	3.82	41.05	15.52	1.93	45.38	5.62	91.72
12c	10.62	37.54	3.53	41.42	15.56	1.93	45.12	5.57	88.89
13c	10.70	37.45	3.28	41.60	15.97	1.94	44.80	5.30	88.64
14c	10.94	37.66	3.45	41.23	15.53	1.93	45.45	5.90	88.55
15c	11.02	37.31	3.86	40.87	15.43	1.93	45.72	5.97	91.16
16c	11.05	37.94	3.54	41.09	15.07	1.93	45.97	6.08	87.68
17c	11.05	38.06	3.41	40.82	15.23	1.92	45.85	6.42	88.21
18c	11.20	37.76	3.35	41.37	15.49	1.93	45.60	6.33	88.42
19c	11.36	37.63	3.40	41.33	15.64	1.92	45.42	6.06	88.92
20c	12.70	37.83	3.69	40.65	14.79	1.91	47.70	8.29	86.89

Samples	Measured Wet method (%)	WDXRF (%)					RC		
	$y = Fe_2O_3$	CaO	SiO_2	Al_2O_3	Fe_2O_3total	TiO_2	L*	C*	h*
1p	9.55	38.17	3.91	40.26	14.35	1.88	46.09	5.41	92.36
2p	10.50	37.56	3.54	41.27	15.72	1.93	44.85	5.18	89.18
3p	8.50	38.09	4.01	40.65	14.10	1.90	45.46	4.59	97.40
4p	9.80	36.85	3.69	41.78	15.81	1.96	44.39	4.57	91.74
5p	10.00	37.36	3.49	41.52	15.81	1.94	44.37	4.79	89.69
6p	10.90	37.40	3.28	41.72	16.03	1.94	44.72	5.43	89.38
7p	11.90	37.98	3.20	41.65	14.80	1.93	46.86	7.26	86.95
8p	10.50	37.56	3.54	41.27	15.72	1.93	44.85	5.18	89.18
9p	10.44	37.54	3.53	41.42	15.56	1.93	45.12	5.57	88.89
10p	9.70	37.73	3.45	41.56	15.35	1.93	44.67	4.97	89.64
11p	10.94	37.66	3.45	41.23	15.53	1.93	45.45	5.90	88.55
12p	10.89	37.94	3.54	41.09	15.07	1.93	45.97	6.08	87.68
13p	9.68	37.75	3.51	41.21	15.33	1.93	44.50	5.71	87.79
14p	10.00	37.36	3.49	41.52	15.81	1.94	44.37	4.79	89.69
15p	10.70	37.45	3.28	41.60	15.97	1.94	44.80	5.30	88.64
16p	11.05	38.06	3.41	40.82	15.23	1.92	45.85	6.42	88.21
17p	10.00	37.55	3.43	41.36	15.54	1.94	44.93	5.43	89.66
18p	9.38	37.34	3.45	41.68	15.83	1.95	43.92	4.28	92.75
19p	10.04	37.95	3.23	41.81	14.63	1.93	46.01	5.90	87.70

C, and h coordinates in the CIE color space. Additionally y is the reference Fe_2O_3 content of the CAC samples obtained from the wet volumetric classical analysis.

In order to measure the accuracy of multivariate calibration methods, different statistics based on the computation of the root mean square error (RMSE) are often calculated [14,15]. The RMSE is computed as,

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (9)$$

where y_i is the Fe_2O_3 content of the samples under study obtained from chemical analysis and $\hat{y}_i$ is the one predicted by the multivariate calibration method. The root-mean square error of calibration (RMSEC) and the root mean square error of prediction (RMSEP), derived from the RMSEC, are frequently computed. Note that the RMSEC is computed from the calibration set and the RMSEP is computed from the prediction set.

In the following subsections three CLS regression models are studied. The first one takes into consideration only the five variables provided by the WDXRF methodology. The

second makes the linear adjust by using the three color coordinates provided by the RC method, while the third model is calibrated from the whole set of eight variables. In order to evaluate the performance of the calibration model several statistics are calculated. These include the mean error of prediction, the RMSEC and RMSEP values, the determination coefficient R^2 and the results of the F-test. A good regression model should obtain low RMSEC and low RMSEP, together with a high determination coefficient R^2 [16]. Additionally, the F-test should report positive results, it is to say, values of F_{calc} as greater than F_{tab} as possible.

Results based in WDXRF analysis
Here, the five measured variables provided by the WDXRF method are applied to calibrate the CLS model.
Table II shows the measured Fe_2O_3 content y_i in the calibration set samples, the Fe_2O_3 estimated by the CLS model $\hat{y}_i$, the error of prediction of each sample $e_{i\%}$, the mean error, the RMSEC, the determination coefficient R^2 and the results of the F-test.

Table II. CLS results when dealing with the five measured variables provided by the WDXRF method. Calibration set estimation.

%Fe$_2$O$_3$ measured y_i	%Fe$_2$O$_3$ predicted $\hat{y}_i$	$e_{i\%}=100\left|\dfrac{y_i-\hat{y}_i}{y_i}\right|$
8.05	7.6513	4.9528
8.50	10.2798	20.9388
9.38	9.9231	5.7900
9.50	9.9373	4.6032
9.70	10.0347	3.4505
9.90	10.3649	4.6960
9.94	9.8388	1.0181
9.95	10.1146	1.6543
10.00	10.0843	0.8430
10.00	11.2706	12.7060
10.45	10.5504	0.9608
10.62	10.2279	3.6921
10.70	10.5381	1.5131
10.94	10.6029	3.0814
11.02	10.9270	0.8439
11.05	10.0571	8.9855
11.05	10.8612	1.7086
11.20	10.4269	6.9027
11.36	10.8234	4.7236
12.70	11.4957	9.4827
Mean-error (%)		**5.1273**
RMSEC		**0.6881**
R^2		**0.5441**
F_{cal} = 3.3421 < F_{tab} = 4.6001		

Results from Table II show that some of the samples have a high prediction inaccuracy. This results in an overall low determination coefficient and a high mean error of prediction. Furthermore, results of the F-test show that the regression model is not statistically significant to a confidence level of 95%.
Results from Table III confirm the poor accuracy achieved in the calibration model. Thus, as expected, when predicting samples different from those used to calibrate the model, all the statistics indicate even poorer results than in the calibration set. As a result, the variables measured from the WDXRF are insufficient to obtain an accurate calibration model.

Table III. CLS results when dealing with the five measured variables provided by the WDXRF method. Prediction set estimation.

%Fe$_2$O$_3$ measured y_i	%Fe$_2$O$_3$ predicted $\hat{y}_i$	$e_{i\%}=100\left|\dfrac{y_i-\hat{y}_i}{y_i}\right|$
9.55	11.4375	19.7644
10.50	10.2838	2.0590
8.50	10.2200	20.2353
9.80	10.1142	3.2061
10.00	10.2912	2.9120
10.90	10.3858	4.7174
11.90	10.4063	12.5521
10.50	10.2838	2.0590
10.44	10.2279	2.0316
9.70	9.8388	1.4309
10.94	10.6029	3.0814
10.89	10.0571	7.6483
9.68	10.2957	6.3605
10.00	10.2912	2.9120
10.70	10.5381	1.5131
11.05	10.8612	1.7086
10.00	10.5245	5.2450
9.38	9.9231	5.7900
10.04	10.2050	1.6434
Mean-error (%)		**5.6248**
RMSEP		**0.7708**
R^2		**0.0286**

Results based in RC analysis

In this subsection, results from the calibration model obtained from the three measured variables obtained by means of the RC method are reported. Results obtained when dealing with the 20 samples of the calibration set are reported in Table IV.

Table IV. CLS results when dealing with the three measured variables provided by the RC method. Calibration set estimation.

%Fe$_2$O$_3$ measured y_i	%Fe$_2$O$_3$ predicted $\hat{y}_i$	$e_{i\%}=100\left|\dfrac{y_i-\hat{y}_i}{y_i}\right|$
8.05	8.6943	8.0037
8.50	8.1414	4.2188
9.38	9.3493	0.3273
9.50	10.0870	6.1789
9.70	10.2183	5.3433
9.90	9.4701	4.3424
9.94	10.0887	1.4960
9.95	9.7351	2.1598
10.00	9.8682	1.3180
10.00	10.7844	7.8440
10.45	10.1255	3.1053
10.62	10.6363	0.1535
10.70	10.5604	1.3047
10.94	10.8623	0.7102
11.02	10.3993	5.6325
11.05	10.8047	2.2199
11.05	11.2882	2.1557
11.20	11.3385	1.2366
11.36	11.0529	2.7033
12.70	12.5052	1.5339
Mean-error (%)		**3.0094**
RMSEC		**0.3760**
R^2		**0.8638**
F_{cal} = 33.8370 > F_{tab} = 4.4940		

Results from Table IV show that five of the calibration samples have prediction inaccuracy greater than 5 %. How-

ever, the overall mean error of calibration set samples is around 3%. Additionally, results of the F-test show that the regression model is statistically significant when dealing with a confidence level of 95%. Thus, these results are clearly better than those obtained from the WDXRF data. Similar results are obtained for the prediction set samples, as displayed in Table V. As a result, the model calibrated from the RC measured data improves the predictive capability of the one calibrated, through the data obtained with the WDXRF method.

Table V. CLS results when dealing with the three variables provided by the RC method. Prediction set estimation.

%Fe$_2$O$_3$ measured y_i	%Fe$_2$O$_3$ predicted $\hat{y}_i$	$e_{i\%} = 100\left\|\dfrac{y_i - \hat{y}_i}{y_i}\right\|$
9.55	9.1940	3.7277
10.50	10.2883	2.0162
8.50	7.9311	6.6929
9.80	9.4979	3.0827
10.00	10.0787	0.7870
10.90	10.7020	1.8165
11.90	11.7799	1.0092
10.50	10.2883	2.0162
10.44	10.6363	1.8803
9.70	10.0887	4.0072
10.94	10.8623	0.7102
10.89	10.8047	0.7833
9.68	11.4623	18.4122
10.00	10.0787	0.7870
10.70	10.5604	1.3047
11.05	11.2882	2.1557
10.00	10.4981	4.9810
9.38	9.3493	0.3273
10.04	10.5281	4.8616
Mean-error (%)		**3.2294**
RMSEP		**0.4942**
R^2		**0.6788**

Results based on both WDXRF and RC analysis

In this section, the CLS model is calibrated by taking into account the data provided by both WDXRF and RC methods. In total, eight variables are used for characterizing each sample. Table VI shows the results obtained when dealing with the calibration set.

Table VI. CLS results when dealing with the whole set of measured variables provided by both WDXRF and RC methods. Calibration set estimation.

%Fe$_2$O$_3$ measured y_i	%Fe$_2$O$_3$ predicted $\hat{y}_i$	$e_{i\%} = 100\left\|\dfrac{y_i - \hat{y}_i}{y_i}\right\|$
8.05	8.0550	0.0621
8.50	8.5572	0.6729
9.38	9.3467	0.3550
9.50	9.4680	0.3368
9.70	9.8061	1.0938
9.90	9.7372	1.6444
9.94	9.8185	1.2223
9.95	10.0567	1.0724
10.00	10.0819	0.8190
10.00	10.0198	0.1980
10.45	10.6342	1.7627
10.62	10.6408	0.1959
10.70	10.7434	0.4056
10.94	10.9456	0.0512
11.02	10.8378	1.6534
11.05	10.9088	1.2778
11.05	11.1139	0.5783
11.20	11.2896	0.8000
11.36	11.2114	1.3081
12.70	12.7375	0.2953
Mean-error (%)		**0.7903**
RMSEC		**0.1004**
R^2		**0.9903**
F_{cal} = 140.41 >> F_{tab} = 4.8443		

Results from Table VI show clearly that this is the best calibration model among the three studied. Therefore, the information provided by both WDXRF and RC instrumental methods is necessary in order to calibrate an accurate model.

The CLS algorithm outputs the values of the coefficient vector $b_{(m+1,1)}$, which are computed as detailed in equation (5), resulting in

$$y = -104.0705 + 0.8038 \cdot CaO + 0.7247 \cdot SiO_2 + 0.5769 \cdot Al_2O_3 + 1.6571 \cdot Fe_2O_3 - 3.8724 \cdot TiO_2 + 0.8297 \cdot L + 0.4577 \cdot C - 0.0024 \cdot h \qquad (10)$$

Thus, this simple equation gives a very fast and accurate prediction of the Fe$_2$O$_3$ content of an unknown cement sample.

Table VII shows the results for the prediction set obtained by applying equation (10).

Table VII. CLS results when dealing with the whole set of measured variables provided by both WDXRF and RC methods. Prediction set estimation.

%Fe$_2$O$_3$ measured y_i	%Fe$_2$O$_3$ predicted $\hat{y}_i$	$e_{i\%} = 100\left\|\dfrac{y_i - \hat{y}_i}{y_i}\right\|$
9.55	9.6651	1.2052
10.50	10.4395	0.5762
8.50	8.4963	0.0435
9.80	9.6377	1.6561
10.00	9.9192	0.8080
10.90	10.8632	0.3376
11.90	11.8505	0.4160
10.50	10.4395	0.5762
10.44	10.6408	1.9234
9.70	9.8185	1.2216
10.94	10.9456	0.0512
10.89	10.9088	0.1726
9.68	9.8451	1.7056
10.00	9.9192	0.8080
10.70	10.7434	0.4056
11.05	11.1139	0.5783
10.00	10.2463	2.4630
9.38	9.3467	0.3550
10.04	10.3291	2.8795
Mean-error (%)		**0.9570**
RMSEP		**0.1251**
R^2		**0.9739**

The Fe$_2$O$_3$ content predicted by the CLS model in front of the results provided by the wet volumetric method is plotted in Fig. 1. Both the calibration and the prediction data sets are superimposed in the same plot. A strong correlation between the analytical results and those provided by the CLS model is shown, indicating the accuracy and feasibility of the methodology applied in this work.
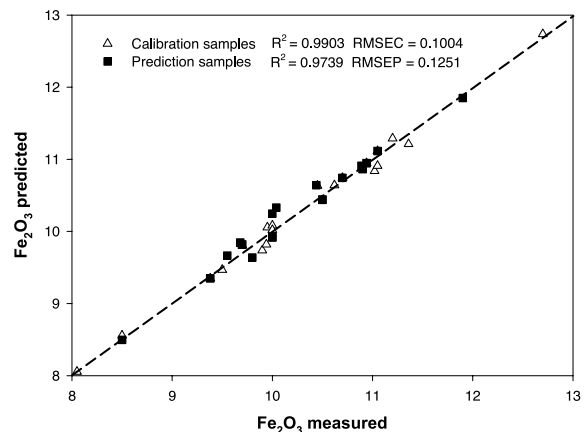

Fig. 1. Generalized correlation for calibration and prediction sets, respectively, when applying MLR

CONCLUSION AND DISCUSSION

In this work the classical least squares method has been successfully applied to predict the Fe_2O_3 content of calcium aluminate cement samples. The data provided by both WDXRF and RC methods has been used to calibrate three CLS models. The first model only considers the five variables provided by the WDXRF methodology. The second model makes the linear adjust by using the three color coordinates provided by the RC method, whereas the third model is calibrated from the whole set of eight variables. Results clearly show that the model including the eight measured variables is the best calibration model among the three studied. Therefore, the information provided by both WDXRF and RC instrumental methods is necessary in order to calibrate an accurate model.

The predictive performance of the CLS calibration models studied in this work has been checked by means of the prediction of samples different from those used to calibrate the model. The best model results in mean prediction errors close to 1 percent. Thus, the proposed methodology allows a fast and accurate determination of the Fe_2O_3 content (and consistently also the FeO content) in calcium aluminate cement. This, in turn, provides better control of the kiln where calcium aluminate cement is made, allowing improving the quality of the final product.

The authors of this work expect to go beyond this study through the application of more advanced multivariate methods to discover if even more accurate calibration models can be obtained.

BIBLIOGRAPHY

1. Hewlett, Peter C. (Editor), Lea's Chemistry of Cement and Concrete. Fourth Edition. Butterworth-Heinemann, Oxford, pp. 713-722, 1998, reprinted in 2001.
2. Diaz, DB; Camargo, G; Molano, M, et al., Evaluation of zinc oxide catalysts and their catalytic activity in the reaction for the production of transesterification biodiesel, Afinidad, **65**(536), pp. 299-306, 2008.
3. Diaz, RV; Dominguez, ER, Natural and anthropogenic component of the chemical elementary composition of the suspended particles (PM2.5) and his possible emission sources, at the metropolitan zone of the Mexico city, Afinidad **65**(534), pp. 120-128, 2008.
4. Garcia, AA; Moreno, PY; Garcia, JA, et al., Study of structural parameters in the carbonization and activation of a Colombian carbon Colombian through X ray diffraction, Afinidad, **65**(535), pp. 230-235, 2008.
5. Kowalski, B.R.; Seasholtz, M.B.; Recent Developments in Multivariate Calibration, Journal of Chemometrics, **5**, pp. 129-145, 1991.
6. Gaetano Ragno, Claudio Vetuschi, Antonella Risoli, Giuseppina Ioele, Application of a classical least-squares regression method to the assay of 1,4-dihydropyridine antihypertensives and their photoproducts, Talanta **59**, pp. 375-382, 2003.
7. Rivera, D.; Alama, M.K.; Yelton, W.G.; Staton, A.W.; Simonson, R.J.; Use of classical least squares/partial least squares (CLS/PLS) hybrid algorithm for calibration and calibration maintenance of surface acoustic wave (SAW) devices, Sensors and Actuators, **B 99,** pp. 480–490, 2004.
8. Lopes, Marta B.; Wolff, Jean-Claude; Bioucas-Dias, José M.; Figueiredo, Mário A.T.; Determination of the composition of counterfeit Heptodin tablets by near infrared chemical imaging and classical least squares estimation, Analytica Chimica Acta, **641**(1-2), pp. 46-518, May 2009.
9. Draper, N. R.; Smith, H.; Applied Regression Analysis, 3rd. ed., Wiley Series in Probability and Statistics, John Wiley & Sons Inc., New York, 1998.
10. Donachie, A.; Walmsley, A.D.; Haswell, S.J.; Application and comparisons of chemometric techniques for calibration modelling using electrochemical/ICP-MS data for trace elements in UHQ water and humic acid matrices, Analytica Chimica Acta, **378**, pp. 235-243, 1999.
11. Marengo, E.; Longoa, V.; Bobba, M.; Robotti, E, Zerbinati, O.; Di Martino, S., Butene concentration prediction in ethylene/propylene/1-butene terpolymers by FT-IR spectroscopy through multivariate statistical analysis and artificial neural networks, Talanta, **77**(3), pp. 1111-1119, 2009.
12. Montgomery, Douglas C.; Peck, Elisabeth A.; Vining, Geoffrey C.; Introduction to linear regression analysis 4th Edition New York John Wiley & Sons, cop. 2006
13. Monnier, A.; Use of Titanium Trichloride in Volumetric Analysis. Ann.Chim. anal., **21**, pp. 109-113, 1916.
14. Esbensen, Kim H.; Multivariate data analysis in practice: an introduction to multivariate data analysis and experimental design, Camo Process AS, Oslo 2002.
15. Donald A. Burns, Emil W. Ciurczak, Handbook of near-infrared analysis, CRC Press, Third Edition, 2007.
16. Lammertyn, J.; Peirs, A.;De Baerdemaeker, J; Nicolai, B.; Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment. Postharvest Boil and Technol., **18**, pp. 121–132, 2000.