DEVELOPMENT OF CHEMOMETRICS METHOD BASED ON INFRARED SPECTROSCOPY FOR THE DETERMINATION OF CEMENT COMPOSITION AND PROCESS OPTIMIZATION

A Thesis Submitted to

The Graduate School of Engineering and Sciences of

İzmir Institute of Technology

İn Partial Fulfillment of the Requirement for the Degree of

MASTER OF SCIENCE

in Chemistry

by

Dilek TEPELİ

December, 2021

İZMİR

ACKNOWLEDGEMENTS

I owe a debt of gratitude to many people without whom I would not have completed this master's thesis.

I would like to express my profound gratefulness to my advisor Prof. Dr. Durmuş ÖZDEMİR. He has endeared Chemometrics to me through his enthusiasm, motivation, and excellent attempts to describe clearly and simply. His effort was invaluable in the study and writing of this thesis.

Also, I would like to thank Selahattin ŞENTÜRK for helping me get through the difficult times and for all his support and friendship.

I would like to thank BATIÇİM Quality Contol Manager Mehmet Gökhan GÜMÜŞ, providing me with all necessary samples and the XRF analyses of the samples to carry out the study.

Finally, I want to thank my parents Nural TEPELİ and Metin TEPELİ, also my friends Nehir NALINCI, Tuğçe KAVURANPALA, Remziye YILDIZ, Barış YILDIRIM, Alper İNAN for being with me through the difficult times and for their encourage me.

ABSTRACT

DEVELOPMENT OF CHEMOMETRICS METHOD BASED ON INFRARED SPECTROSCOPY FOR THE DETERMINATION OF CEMENT COMPOSITION AND PROCESS OPTIMIZATION

Calcium, silicon, aluminum, iron oxide-containing raw materials are used in controlled portions to manufacture cement. (How cement Is Made, n.d.) This mixture is first converted to clinker, obtained by heating the mixture to 1500^{o}C; some additives are added and ground to obtain cement. Depending on the purpose, various types are produced, and therefore, the determination of cement composition is an essential task for the quality consideration and the sustainability of the production processes. The quantitative analysis of cement is performed with X-ray fluorescence spectroscopy. However, XRF generally requires tedious and lengthy analysis times. In this study, quantitative determination of the raw materials, intermediate products, and types of cement by using Fourier transform infrared spectroscopy coupled with chemometrics multivariate calibration method is aimed, which could be an alternative for the current XRF technique. Samples were collected from a local cement factory that has been in the sector for several years. Reference analyses of the samples were performed at the quality control laboratory of the same factory. The same samples were analyzed by the FTIR-ATR spectrometer. The resulting FTIR spectra combined with XRF reference composition data were used to construct calibration models using the partial least squares method (PLS). Based on the obtained results, the proposed method could generate quite successful results for the quantitative determination of all types of products used to produce cement. The regression coefficients (R^2) of the PLS models vary from 0.95 to 0.99. The standard errors of cross-validations were found as from 0.21 to 1.42 (w/w%).

ÖZET

ÇİMENTO KOMPOZİSYONUN BELİRLENMESİ VE ÜRETİM SÜREÇLERİNİN OPTİMİZASYONU İÇİN KIZILÖTESİ SPEKTROSKOPİSİNE DAYALI KEMOMETRİK METOT GELİŞTİRİLMESİ

Çimento ilk olarak kalsiyum, silisyum, alüminyum ve demir oksitleri içeren hammaddelerin belirli oranlarda karıştırılarak 1500^{o}C gibi oldukça yüksek sıcaklıkta pişirilmesi ile oluşan klinker'in, çeşitli katkı maddeleriyle birlikte öğütülmesi sonucunda elde edilir. Günümüzde kullanım amaçlarına bağlı olarak oldukça fazla çeşit ve kompozisyonda çimento üretiminin gerçekleştirildiği göz önünde bulundurulduğunda söz konusu ürünlerin bileşimin belirlenmesi ürün kalitesi ve sürdürülebilirlik açısından büyük bir önem arz etmektedir. Hali hazırda çimento kompozisyonunun kantitatif analizine yönelik olarak kullanılan yöntemler genel olarak klasik ıslak kimyaya dayalı gravimetrik ve titrimetrik yöntemler yanında X-ışını floresans (XRF) spektroskopisi tekniği ile yapılmaktadır. Ancak, bu yöntemler analizi yapılacak örneğe bağlı olarak zaman alıcı ve zahmetli süreçler gerektirebilmektedir. Bu çalışmada, çimento üretiminde hammaddeler, ara ürünler ve nihai ürünlere yönelik olarak çimento kompozisyonunun kantitatif belirlenmesi için mevcut X-ışını floresans (XRF) spektroskopisi tekniğine alternatif olarak Fourier dönüşümlü kızılötesi spektrometresi (FTIR) kullanılarak alınan spektral verilere kemometrik çok değişkenli kalibrasyon metotlarından biri uygulanarak yeni bir analitik metot geliştirilmesi hedeflenmiştir. Çalışma kapsamında uzun yıllardır çimento sektöründe üretim yapan bir özel sektör kuruluşundan elde edilen örnekler ile bu örneklerin aynı kuruluşta bulunan XRF sisteminden elde edilen referans analizleri kullanılarak sonrasında laboratuvarımızda bulunan FTIR sistemi ile aynı örneklerin FTIR-ATR spektrumları kaydedilerek ve elde edilecek spektral veriler ile XRF sisteminden gelen referans kompozisyon bilgisi kullanılarak kısmi en küçük kareler (PLS) yöntemi ile çok değişkenli kalibrasyon modelleri oluşturulmuştur. Ülkemizde gerçekleştirilen çimento üretiminin neredeyse tüm bölgelere yayılmış olarak oldukça yüksek üretim kapasitelerine ulaştığı göz önüne alındığından geliştirilen modellerin oldukça önemli bir yaygın etkiye sahip olacağı görülmektedir.

TABLE OF CONTENTS

LIST OF FIGURES…….....xii

LIST OF TABLES……...…….....vii

CHAPTER 1 INTRODUCTION 1
1.1. Literature Review..... 2
1.2. Aim of the study..... 7

CHAPTER 2 INSTRUMENTATION..... 9
2.1. Infrared spectroscopy 9
2.2. Attenuated Total Reflectance (ATR)..... 12
2.3. X-ray Fluorescense Spectroscopy..... 14

CHAPTER 3 MULTIVARIATE CALIBRATION METHODS..... 18
3.1. Overview 18
3.2. Classical Least Squares (CLS) 18
3.3. Inverse Least Squares (ILS)..... 19
3.4. Partial Least Squares (PLS) 21

CHAPTER 4 EXPERIMENTATION AND METHODS..... 26
4.1. Samples 26
4.2. FTIR-ATR Measurements 26
4.3. X-ray Fluoresence Spectrometer Measurements 27
4.3.1. Pressed Pellet Technique 28
4.3.2. Glass Bead Preparation..... 29
4.4. Loss of Ignıtıon Analysis 29
4.5. Data Analysis..... 30

CHAPTER 5 RESULTS AND DISCUSSION 31
5.1. The First Scenario. 31
5.1.1. The First Scenario Without pre-processing 37
5.1.1.1. The PLS results of the CaO component..... 37
5.1.1.2. The PLS results of the SiO_2 component 40

5.1.1.3. The PLS results of the Al_2O_3 component 42
5.1.1.4. The PLS results of the Fe_2O_3 component 44
5.1.1.5. The PLS results of the MgO component 45
5.1.1.6. The PLS results of the SO_3 component 47
5.1.1.7. The PLS results of the Na_2O component 49
5.1.1.8. The PLS results of the K_2O component 50
5.1.1.9. The PLS results of L.O.I 52
5.1.2. The First Scenario with Pre-Processing 54
5.1.2.1. The PLS result of CaO component 57
5.1.2.2. The PLS result of SiO_2 component 59
5.1.2.3. The PLS result of Al_2O_3 component 60
5.1.2.4. The PLS result of Fe_2O_3 component 62
5.1.2.5. The PLS result of MgO component 63
5.1.2.6. The PLS result of SO_3 component 65
5.1.2.7. The PLS result of Na_2O component 66
5.1.2.8. The PLS result of K_2O component 68
5.1.2.8. The PLS result of L.O.I component 69
5.2. The Second Scenario 72
5.2.1. The PLS results of CaO 76
5.2.2. The PLS results of SiO_2 77
5.2.3. The PLS results of Al_2O_3 78
5.2.4. The PLS results of Fe2O3 79
5.2.5. The PLS results of MgO 80
5.2.6. The PLS results of SO_3 81
5.2.7. The PLS results of Na_2O 82
5.2.8. The PLS results of K_2O 83
5.2.9. The PLS results of Loss of ignition 84

CHAPTER 6 CONCLUSION 90

REFERENCES 91

LIST OF FIGURES

Figure **Page**

Figure 2.1. Schematic illustration of vibrational and rotational transitions potential energies. 10
Figure 2.2. Types of molecular vibrations 11
Figure 2.3. Schematic illustration of Fourier transform infrared spectrometer 12
Figure 2.4. Schematic illustration of the ATR principle 14
Figure 2.5. Regions of the electromagnetic spectrum 14
Figure 2.6. Schematic illustration of fluorescent radiation 15
Figure 2.7. An illustration of the generation of X-ray fluorescence radiation 16
Figure 2.8. A schematic setup of an energy-dispersive X-ray fluorescence spectrometer with direct excitation 16
Figure 2.9. Schematic setup of a wavelength-dispersive X-ray fluorescence spectrometer. 17
Figure 3.1. An example of a number of components vs. PRESS plot. 25
Figure 4.1. Illustration of all the steps in producing a pressed pellet with a binder 28
Figure 4.2. Illustration of the steps in producing glass beads 29
Figure 5.1. Mid-IR spectra of all samples 36
Figure 5.2. Mid-IR spectra of each type of sample included in the data set 36
Figure 5.3. Number of PCs vs. PRESS plot to select the optimal number of LVs of CaO. 37
Figure 5.4. a) Reference CaO concentrations vs. Predicted CaO concentrations. 37
Figure 5.5. Reference CaO concentrations vs. Corresponding PLS prediction residuals of CaO 38
Figure 5.6. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of CaO 39
Figure 5.7. a) Reference CaO concentrations (w/w%) vs. PLS predicted CaO 39
Figure 5.8. Reference concentrations vs. corresponding PLS prediction residuals of CaO. 40
Figure 5.9. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of SiO2. 40
Figure 5.10. a) Reference SiO_2 concentrations (w/w%) vs. PLS predicted SiO_2 concentrations (w/w%). b)Statistical parameters of SiO_2. 41

Figure **Page**

Figure 5.11. Reference concentrations vs. corresponding PLS prediction residuals of SiO_2 42
Figure 5.12. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of Al_2O_3 42
Figure 5.13. a) Reference Al_2O_3 concentrations vs. Predicted Al_2O_3 concentrations by PLS. b) Statistical parameters of Al_2O_3 43
Figure 5.14. Reference concentrations vs. corresponding PLS prediction residuals of Al_2O_3 43
Figure 5.15. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of Fe_2O_3 44
Figure 5.16. a) Reference Fe2O3 concentrations vs. Predicted Fe_2O_3 concentrations by PLS. b) Statistical parameters of Fe_2O_3 44
Figure 5.17. Reference concentrations vs. corresponding PLS prediction residuals of Fe_2O_3 45
Figure 5.18. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of MgO 46
Figure 5.19. a) Reference MgO concentrations vs. Predicted MgO concentrations by PLS. b) Statistical parameters of MgO. 46
Figure 5.20. Reference concentrations vs. corresponding PLS prediction residuals of MgO (w/w%) 47
Figure 5.21. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of SO_3 47
Figure 5.22. a) Reference SO_3 concentrations vs. Predicted SO_3 concentrations by PLS. b) Statistical parameters of SO_3. 48
Figure 5.23. Reference concentrations vs. corresponding PLS prediction residuals of SO_3 48
Figure 5.24. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of Na_2O 49
Figure 5.25. a) Reference Na_2Oconcentrations vs. Predicted Na_2O concentrations by PLS. b) Statistical parameters of Na_2O 49
Figure 5.26. Reference concentrations vs. corresponding PLS prediction residuals of Na_2O 50

Figure **Page**

Figure 5.27. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of K_2O 50
Figure 5.28. a) Reference K_2Oconcentrations vs. Predicted K_2O concentrations by PLS. b) Statistical parameters of K_2O. 51
Figure 5.29. Reference concentrations vs. corresponding PLS prediction residuals of K_2O 51
Figure 5.30. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of L.O.I 52
Figure 5.31. a) Reference K_2Oconcentrations vs. Predicted L.O.I concentrations by PLS b) Statistical parameters of L.O.I 53
Figure 5.32. Reference concentrations vs. corresponding PLS prediction residuals of L.O.I 53
Figure 5.33. left: Raw FT-IR spectra of the samples, right : the EMSC spectra of the samples 57
Figure 5.34. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of CaO. 57
Figure 5.35. a) Reference CaO concentrations vs. Predicted CaO concentrations by PLS. b) Statistical parameters of CaO 58
Figure 5.36. Reference concentrations vs. corresponding PLS prediction residuals of CaO. 58
Figure 5.37. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of SiO_2 59
Figure 5.38. a) Reference SiO_2 concentrations vs. Predicted SiO_2 concentrations by PLS. b) Statistical parameters of SiO_2. 59
Figure 5.39. Reference concentrations vs. corresponding PLS prediction residuals of SiO_2 60
Figure 5.40. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of Al_2O_3 60
Figure 5.41. a) Reference Al_2O_3 concentrations vs. Predicted Al_2O_3 concentrations by PLS. b) Statistical parameters of Al_2O_3 61
Figure 5.42. Reference concentrations vs. corresponding PLS prediction residuals of Al_2O_3 61

Figure **Page**

Figure 5.43. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of Fe_2O_3 62
Figure 5.44 . a) Reference Fe_2O_3 concentrations vs. Predicted Fe_2O_3 concentrations by PLS. b)Statistical parameters of Fe_2O_3 62
Figure 5.45. Reference concentrations vs. corresponding PLS prediction residuals of Fe_2O_3 63
Figure 5.46. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of MgO 63
Figure 5.47. a) Reference MgO concentrations vs. Predicted MgO concentrations by PLS. b) Statistical parameters of MgO 64
Figure 5.48. Reference concentrations vs. corresponding PLS prediction residuals of MgO 64
Figure 5.49. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of SO_3 65
Figure 5.50. a) Reference SO_3 concentrations vs. Predicted SO_3 concentrations by PLS. b) Statistical parameters of SO_3 65
Figure 5.51. Reference concentrations vs. corresponding PLS prediction residuals of SO_3 66
Figure 5.52. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of Na_2O 66
Figure 5.53. a) Reference Na_2O concentrations vs. Predicted Na_2O concentrations by PLS. b) Statistical parameters of Na_2O 67
Figure 5.54. Reference concentrations vs. corresponding PLS prediction residuals of Na_2O 67
Figure 5.55. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of K_2O 68
Figure 5.56. a) Reference K_2O concentrations vs. Predicted K_2O concentrations by PLS. b) Statistical parameters of K_2O 68
Figure 5.57. Reference concentrations vs. corresponding PLS prediction of K_2O 69
Figure 5.58. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of L.O.I 69
Figure 5.59. a) Reference L.O.I concentrations vs. Predicted L.O.I concentrations by PLS. b) Statistical parameters of L.O.I 70

Figure **Page**

Figure 5.60. Reference concentrations vs. corresponding PLS prediction residuals of L.O.I 70
Figure 5.61. Raw FT-NIR spectra of a total of 56 CEM-I samples..... 74
Figure 5.62. The view of spectra of CEM-I samples after EMSC pre-processed method applied..... 75
Figure 5.63. a) Number of PCs vs. PRESS plot for selecting the optimal number of PCs. b) Reference CaO concentrations vs. Predicted CaO concentrations c) Reference CaO concentrations vs. corresponding PLS prediction residuals d) Statistical parameters of CaO..... 76
Figure 5.64. a) Number of PCs vs. PRESS plot for selecting the optimal number of PCs. b) Reference SiO_2 concentrations vs. Predicted SiO_2 concentrations c) Reference SiO_2 concentrations vs. corresponding PLS prediction residuals d) Statistical parameters of SiO_2..... 77
Figure 5.65. a) Number of PCs vs. PRESS plot for selecting the optimal number of PCs. b) Reference Al_2O_3 concentrations vs. Predicted Al_2O_3 concentrations c) Reference Al_2O_3 concentrations vs. corresponding PLS prediction residuals d) Statistical parameters of Al_2O_3..... 78
Figure 5.66. a) Number of PCs vs. PRESS plot for selecting the optimal number of PCs. b) Reference Fe_2O_3 concentrations vs. Predicted Al_2O_3 concentrations c) Reference Fe_2O_3 concentrations vs. corresponding PLS prediction residuals d) Statistical parameters of Fe_2O_3..... 79
Figure 5.67. a) Number of PCs vs. PRESS plot for selecting the optimal number of PCs. b) Reference MgO concentrations vs. Predicted MgO concentrations c) Reference MgO concentrations vs. corresponding PLS prediction residuals d) Statistical parameter of MgO 80
Figure 5.68. a) Number of PCs vs. PRESS plot for selecting the optimal number of PCs. b) Reference SO_3 concentrations vs. Predicted SO_3 concentrations c) Reference SO_3 concentrations vs. corresponding PLS prediction residuals d) Statistical parameters of SO_3..... 81

Figure **Page**

Figure 5.69. a) Number of PCs vs. PRESS plot for selecting the optimal number of PCs. b) Reference Na_2O concentrations vs. Predicted Na_2O concentrations c) Reference Na_2O concentrations vs. corresponding PLS prediction residuals d) Statistical parameters of Na_2O 82

Figure 5.70. a) Number of PCs vs. PRESS plot for selecting the optimal number of PCs. b) Reference K_2O concentrations vs. Predicted K_2O concentrations c) Reference K_2O concentrations vs. corresponding PLS prediction residuals d) Statistical parameters of K_2O 83

Figure 5.71. a) Number of PCs vs. PRESS plot for selecting the optimal number of PCs. b) Reference L.O.I concentrations vs. Predicted L.O.I concentrations c) Reference L.O.I concentrations vs. corresponding PLS prediction residuals d) Statistical parameters of L.O.I. 84

LIST OF TABLES

Table **Page**

Table 2.1. Infrared spectral regions 9
Table 2.2. Number of vibrations for nonlinear and linear molecules 11
Table 2.3. Properties of some common ATR Crystals 13
Table 5.1. The calibration data set of the first scenario 32
Table 5.2. Validation data set of first scenario. 34
Table 5.3. Calibration and Validation data set of first scenario for L.O.I. 35
Table 5.4. All PLS results of without pre-processing and with EMSC pre-processing . 70
Table 5.5. t-Test: Paired Two Sample for Means 72
Table 5.6. Calibration data set for second scenario samples. 73
Table 5.7. Validation data set for second scenario samples 74

CHAPTER 1

INTRODUCTION

Cement, derived from the Latin word "caementum" which means a piece of hewn-stone, is an adhesive substance produced by grinding raw meals of cement such as limestone and clay then, heating these materials to high-temperature degrees to form a rock-like substance and is called clinker (Kapkaç, 2013). Clinker is the intermediate product of cement. Lastly, the clinker is combined with many additives then ground into the fine powder (Türkiye Çimento Sanayicileri Birliği, n.d.-a). In civilization's history, binding materials such as limestone and gypsum have been known and used since ancient times. However, the first ferroconcrete buildings built using cement began to construct in England 1850s.Today, the first cement is known as the "Portland cement" was created by an English engineer Joseph Aspdin in 1824, and he has patented it. He called his product Portland Cement because it produced a substance similar to stones from the mines near Portland Island in the UK when it hardened (Teknik Bilgiler - Batıçim, n.d.).The manufacturing of cement on an industrial scale has been started in the 1848. It continues to be a fundamental input of the construction industry as the essential building material worldwide. In Turkey, the first manufacturing of cement was started with just a production capacity of 20.000 tons/year in 1910. The production capacity of cement came up to over 130 million tons, according to the reports published by the Turkish cement Manufacturer's Association (TÇBM) and other related institutions in recent years (Türkiye Çimento Sanayicileri Birliği, n.d.-b; Türkiye Çimento Sektörü 100 Milyon Ton Üretim Kapasitesine Ulaştı, n.d.). The standardization studies for cement production on an industrial scale were carried out by T.C 51, a European Standardization Committee subcommittee. Also, many types of cement are included in EN 197-1 standard. Later, the European Standard, which was directly accepted as the Turkish Standard, replaced Turkish cement's general purposes. cements named as "CEM" are produced in 5 main types with covering 27 sub-types (CEM I, CEM II, CEM III, CEM IV and CEM V) in TS EN 197-1. In addition to these, there are various types of cement produced for the particular use.

When cement is discussed in detail, it is a primary concrete element that is the most used globally. Each year, an enormous amount of cement is manufactured and used to construct buildings, roads, and many local purposes. cement is made by mixing calcium, silicon, aluminum, and iron oxide-containing raw meals in a carefully controlled combination. The manufacturing of cement is comprised of four main stages. First, crushing and grinding the raw meals such as limestone and clay, and second, blending the raw meals in controlled proportions. Third, clinker production by burning the mixed raw meal to higher temperature degrees at 1450 °C. The final stage is cement making by grinding the burned product with about 5% of calcium sulfate (gypsum).

On the other hand, The additives added to clinker affect the properties of cement and quality at high-temperature degrees (Arslanhan, 2016). Also, the production of the clinker process results in four main phases these are tricalcium silicate ($3CaO.SiO_2$) noted C3S, dicalcium silicate ($2CaO.SiO_2$) noted C2S, tricalcium aluminate ($3CaO.Al_2O_3$) noted C3A, and tetracalciumaluminoferrite ($4CaO.Al_2O_3Fe_2O_3$) noted C4AF. These four components are also known as the leading cement phases (Ferrari et al., 2012). The vital compounds C_3S, C_2S, C_3A, and C_4AF in Portland cement are effective for cement's physical features. The silicates, C_3S and C_2S, are responsible for the strength of hydrated cement paste. cement low in C_3A has a sulfate-resistant property because gypsum reduces the hydration rate of C_3A. Also, C_4AF is present in cement in small amounts and can influence the melting temperature of raw materials in the kiln but does not provide so much cement paste strength (Neville, 1995; TAYLOR, 1964). In general, the chemical composition of Portland cement clinker consists of calcium oxide (CaO; 62-68%), silicon dioxide (SiO_2; 19-25%) and the other components aluminum oxide(Al_2O_3; 2-9%) also iron oxide (Fe_2O_3; 1-5%) (De Schepper et al., 2013). However, these percentages can be slightly changeable for specific applications.

1.1. Literature Review

In the past years, classical wet chemical methods such as gravimetric and titrimetric analyses were applied for the quantitative determination of cement composition in terms of six essential oxides (CaO, SiO_2, Al_2O_3, Fe_2O_3, MgO, SO_3) However, these methods generally require tedious and lengthy analysis times depending

on the sample types to be analyzed. Then, various spectroscopic analysis methods have been developed to determine the chemical composition of cement to overcome the drawbacks of the method of classical wet chemical. In the field of this application, methods for determining cement composition are consisting of methods such as atomic absorption spectroscopy (AAS), inductively coupled plasma emission spectrometry (ICP), and X-ray fluorescence spectrometry (XRF) which give an elemental analysis of bulk composition in equivalent oxides.

Choi et al. (1994) were developed a new fusion method for the composition of cement raw mix by AAS and titrimetry as an alternative to X-ray fluorescence spectrometry. In the AAS method, the sample must be dissolved, and fusion is used to convert insoluble residue to an acid-soluble form. Many fusion agents are used, such as Na_2CO_3, $Li_2B_4O_7$, and $LiBO_2$ to achieve this aim, but all the common fusion agents have their disadvantages. A developed new fusion system was decided as the best in terms of efficiency and working times required. A new fusion system was developed by fusing the sample with a mixture of lithium tetraborate and sodium carbonate (1:3) at 925 ºC for 10 min. Also, hydrochloric acid was used to dissolve the fusion cake. The concentration of analyte solution was determined with both a classical method (volumetric titration) and an instrumental method (AAS). The analyses of interested oxides such as SiO_2, Al_2O_3, Fe_2O_3 and CaO completed at the end of one hour. The accuracy and precision were found to as comparable with X-ray Fluorescence spectrometry. As a result, a new fusion system was proposed as a backup method because of having time-saving, cost-effective instrumentation than ICP and XRF (Choi et al., 1994).

Marjanovic et al. (2000) are performed an analysis of cement for major (Ca, Si, Mg, Al, Fe), minor and trace elements (S,K,Ti,Na,P,Mn and Sr) was determined by introducing it as a slurry into an inductively coupled plasma optical emission spectrometer (ICP-OES) for the feasibility of analyzing cement. In the experimental part, reference materials as slurries or simple aqueous standards were used for calibration. The mixture of hydrochloric acid (1% v/v), glycerol (0.01% m/v) and agitation in an ultrasonic bath (10 min) was utilized to ensure well dispersed and stable slurries. Both axial and radial plasma viewing was used for the measurements were made on simultaneous ICP-OES instrument. As a result, the values obtained for selected elements were compared to those measured on an X-ray fluorescence (XRF) instrument, and it was seen that ICP-OES gave an effective alternative to XRF for the analysis of cement. According to all

these results, the method could be easily utilized on a routine basis for the analysis of cement. In addition, this method also has great advantageous features such as cost-effectiveness and simplicity of sample preparation (Marjanovic et al., 2000)

In another study, Khelifi et al. (2017) was developed a reliable analytical program to adjust raw materials of cement proportions and control process conditions by XRF. The design of the analytical program includes three steps those are: choice of well-defined standards, the determination of the composition of standard solution and plotting the calibration curve. Calibration standards were prepared to give maximum and minimum values for each element to be analyzed. The standards are consisting of clinker samples, and the chemical analyses of clinker samples were performed with the European standard EN-196-2. The CaO, Al_2O_3, Fe_2O_3, and MgO concentrations were determined by complexometric titration, and sulfate content was determined by nephelometric turbidity. Silica content was determined by the gravimetric method, and Na and K ions were determined by flame spectrophotometry. Also, in this work, an X-ray fluorescence spectrometer was used to perform analysis of the clinker samples by preparing fused bead. The relative intensities obtained XRF were plotted against corresponding concentrations of the calibration solutions. The regression coefficients of the models were found as $R^2 > 0.97$ for CaO, SiO_2, Al_2O_3, Fe_2O_3, and were found $R^2 < 0.97$ for MgO, Cl, and Na_2O. (Khelifi et al., 2017). The mentioned methods above require sample preparation involves many steps that must be prepared carefully using many chemicals. Furthermore, the procedures are time-consuming.

Infrared spectroscopy (IR) has been used greater drastically in this region for mineral constituent detection, structural studies, and hydration rate determination (Dunstetter et al., 2006; Gastaldi et al., 2010). However, only a few studies have used infrared spectroscopy to carry out quantitative analysis of cement. Zaini et al. used Shortwave infrared (SWIR) spectroscopy with reflectance measurements in the 2100-2400 nm wavelength region that exhibits several diagnostic absorption features of mineral chemistries. The chemical composition and mineral abundance on the rock surfaces of Portland cement-grade lime stones was examined using shortwave infrared spectroscopy to determine the chemical composition and mineral abundance on the rock surfaces, which are used as one of the cement raw materials. Spectral parameters, consisting of the wavelength position and intensity of absorption characteristic derived from the SWIR continuum removed spectra suited for the second-order polynomial curves, are helpful in

reading and estimating carbonate (CO_3) and Al-OH absorption functions associated with CaO and Al_2O_3 contents and the compositions inside the rock samples, respectively. The authors used portable X-ray fluorescence measurements of CaO and Al_2O_3 in the rock samples to correlate the wavelength location and depth of CO_3 and Al-OH absorption features associated with carbonate and phyllosilicate minerals (Al_2O_3). The wavelength positions of the CO_3 and Al-OH absorption features vary with the chemical compositions of the samples. For each type of carbonate rock sample, separate models were developed. The regression coefficients were obtained for dark gray limestone samples as 0.774 for CaO and 0.842 for Al_2O_3. Also, for the light gray limestones, the regression coefficients were obtained as 0.787, 0.723 for CaO and Al_2O_3, respectively. The model obtained from the dolomitic limestone samples gave lower regression coefficients. As a result, the authors decided that in terms of CaO, MgO, Al_2O_3, and SiO_2 concentrations and compositions, dark gray and light gray limestone samples are more appropriate for manufacturing Portland cement clinker (Portland cement-grade limestone) than dolomitic limestone samples. (Zaini et al., 2016)

In another study, Fourier transforms infrared spectroscopy (FTIR) is applied for quantification determination of alkali concentrations in cement samples by Nasrazadani and Springfield. Alkali determination is crucial because alkali in cement is responsible for the alkali-silica reaction that shows itself inside the shape of early cracking in concrete structures, bridge decks, and concrete roads. X-ray fluorescence spectroscopy (XRF) is typically used for cement alkali quantification, but a more straightforward and faster analytical manner based on Fourier transform infrared spectroscopy (FTIR) has been improved for this motive in the study. The Na_2O linear regression coefficient, measured with FTIR and XRF, was taken into account. FTIR approach was performed efficiently to quantify alkali concentration in cement-based on absorbance band area ratio of 750 to a 923 cm^{-1} that belongs to tricalcium aluminate (C_3A). The linear regression coefficient between FTIR absorption ratio at 750/923 cm^{-1} and equivalent alkali Na2Oe (Na_2O= % Na_2O + 0.658x%K_2O) measured by XRF was found as 0.97. Analyzes were performed with FTIR faster than other techniques as XRF, XRD. However, in the methodology and material part of the study, sample preparation includes many stages, such as dilution with KBr (2 mg cement and 98 mg KBr), milling, and pellet making. The authors underlined the importance of meticulous control over pellet production to ensure reproducibility spectrally in results (Nasrazadani & Springfield, 2014).

In another study, Hughes et al. conducted qualitative and quantitative analyses using diffuse reflectance mid-infrared Fourier transform spectroscopy (DRIFTS). For the qualitative analysis, DRIFT spectra of synthetic C_3S, C_2A, C_3A, C_4AF, and pure gypsum, bassanite, anhydrite, and calcite are included in the study. For quantitative analysis, a large number of cement samples were included in the study. PLS regression calibration models were constructed to estimate the concentration of nine minerals and five minor oxides in samples. The elemental reference analysis of the samples was performed by inductively coupled plasma. Moreover, free lime, ignition loss, and acid-insoluble residue components were also examined for this study. Estimation of mineral composition was done based on the API Bogue calculation. For the construction of PLS models, the number of latent variables was used from 13 to 39 for the prediction concentration of many components. The linear regression coefficients (R^2) for the calibration set were found from 0.765 to 0.977. For the independent validation set, linear regression coefficients were found from 0.632 to 0.959. The accuracy of the independent validation set was found to change from 0.01 to 3.10 % (w/w). Although the proposed FTIR technique is recommended as a fast method, it is stated that the diffuse reflectance measurement technique included the sample dilution with potassium bromide at 90% of each sample, and the measurement is made after a pellet preparation stage. Therefore, the authors highlighted that special care must be taken in the sample preparation step for quantification analyses to reach high-quality spectra (Hughes et al., 1995).

On the other hand, Rebouças et al. used Near-infrared emission spectroscopy (NIRES) to analyze cement samples. For the study, NIRES-AOTF (Acousto-Optical Tunable Filter) spectrometer was assembled to develop the method. A total of 77 samples of cement and 13 samples of mixture design containing cement phases were included in the study. These cement samples are diversified with different cement types such as type II, type III, type IV, and type V. The estimation of the concentration of six major oxides was investigated using PLS regression calibration models. The reference analyses of the studied samples were performed with the X-ray fluorescence spectrometer (XRF). The regression coefficients (R^2) were obtained from 0.63 to 0.95 for the independent validation set. The worst performance of prediction the PLS model was found belongs to the Fe_2O_3 component. Also, Root means square error prediction (RMSEP) results were obtained from 0.18 to 1.67% (w/w). The authors stressed that there is no need to sample preparation, but temperature stabilization of the samples is required. Additionally, the

authors mentioned that there is no commercial instrument presently available (Rebouças et al., 2018).

1.2. Aim of the study

The present study aims to develop a new analytical method for determining the composition of cement raw materials, intermediate products, and finished products using Fourier Transform Infrared Spectroscopy combined with chemometrics multivariate calibration methods could be an alternative for the current XRF technique. With the present work, a new analytical method based on chemometrics calibration techniques provides a much more practical, fast, and cost-effective method to determine the composition of cement raw materials, intermediates, and final products are still carried out with mostly X-ray fluorescence spectroscopy. XRF or other techniques used in chemical composition determination of cement, raw materials, intermediate products require additional sample preparation, and analysis requires a long time. Compared to FTIR, X-ray fluorescence spectrometer exhibits relatively high costs and health concerns due to the radioactive source. Figure 1. shows a graphical illustration of the modeling process bringing together the FTIR and XRF analysis results, predicts a successful model for each parameter.

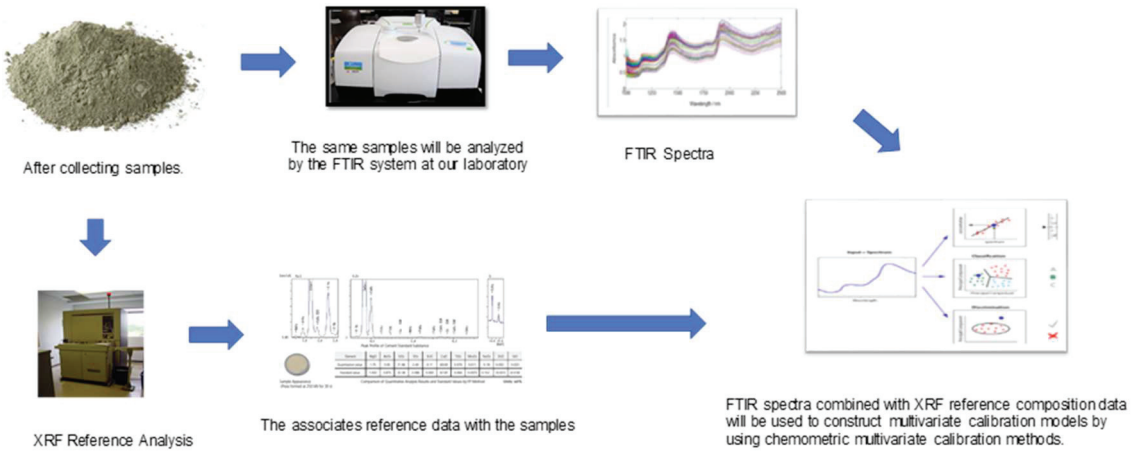


Figure 1.1. Graphical illustration of FTIR based chemometrics modeling process.

As can be seen from Figure 1.1, the first step begins with the collection of cement samples, raw materials, intermediate products from the BatıÇim cement factory. Then, analyses with FTIR-ATR of these samples are performed in our laboratory. Results of reference analysis achieved by XRF are associated with the obtained spectra of the samples. Then with the spectra of the samples and reference analysis are combined with the chemometrics multivariate calibration technique. Finally, concentration estimation is performed for each parameter.

CHAPTER 2

INSTRUMENTATION

2.1. Infrared spectroscopy

Infrared spectroscopy is one of the analytical techniques that any samples in any state can be studied to get information about the compound's structure with the infrared spectroscopy. Various types of samples such as liquids, pastes, powders, and gases can be inspected with a reasonable choice of sampling techniques. Infrared spectrometers have been on the market since the 1940s. At that time, the instruments bore on prisms to represent dispersive elements. After, diffraction gratings had been included in dispersive machines. The sizeable advances in infrared spectroscopy have emerged due to the introduction of Fourier transform infrared spectrometers. These instruments use an interferometer and a well-established mathematical process to improve the quality of infrared spectra and minimize the time required to occur analysis. With the feature, quantitative analysis of complex samples can be achieved both absorption and emission spectra by contrast with dispersive instruments used for qualitative analysis and structural determination on absorption spectra. Additionally, with continuous advancements to computers, infrared spectroscopy has made even more considerable progress. (Suart, 2004)

The infrared region occupies the electromagnetic spectrum between the visible and microwave region and includes radiations with wavenumbers from 12800 to 10 cm^{-1}. The IR region comprises three parts: the near-, mid- and far-infrared radiation is given in Table 2.1.

Table 2.1. Infrared spectral regions

Region	Wavelength (λ), μm	Wavenumbers (v), cm^{-1}
Near	0.75 to 2.5	12800 to 4000
Middle	2.5 to 50	4000 to 200
Far	50 to 1000	200 to 10
Most Used	2.5 to15	4000 to 670

Infrared spectroscopy investigates the interaction between matter and infrared radiation by giving IR radiation through a sample. Some of the infrared radiation is absorbed by the molecules as they undergo vibrational and rotational transitions causes changing of net dipole moment in molecules (IR active). Some of the infrared radiation is not absorbed by the sample when the molecule does not cause a net change in the dipole moment (IR inactive). Excitation of vibrational, rotational, and bending modes results from the absorption of infrared radiation by the molecules while the molecule stays in its electronic ground state. The potential energy diagram of vibrational and rotational transitions is given in Figure 2.1. Rotational and vibrational transitions often occur together, and vibrational transitions occur at higher energies than rotational transitions, but rotational transitions have slight energy differences.

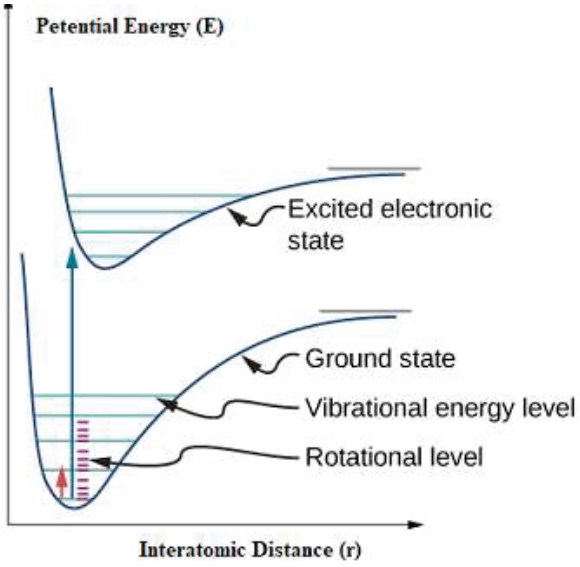


Figure 2.1. Schematic illustration of vibrational and rotational transitions potential energies.

(Source:chem.libretexts.org)

Stretching and bending are the types of molecular vibrations. Stretching vibrations consist of two types are; symmetrical and asymmetrical vibrations, while bending vibrations consist of four types are; Scissoring, rocking, wagging, twisting. The difference between the two types of vibrations is; bending motions change the bond angle of the molecule while stretching motions change bond length. A schematic illustration of the molecular vibrations is given in Figure 2.2

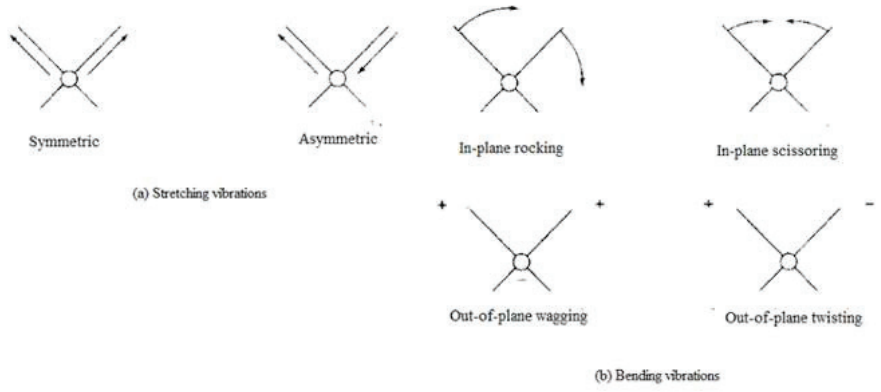


Figure 2.2. Types of molecular vibrations

(Source : Skoog et al 1998)

A molecule along with n atoms has a total of 3n ranges of freedom, similar to cartesian coordinates of every atom in the molecule. For nonlinear and linear molecules, the net number of vibrations is given in Table 2.2. Three degrees of freedom are rotational in a nonlinear molecule, three are translational, and the remaining correspond to fundamental vibrations; in a linear molecule, two degrees are rotational, and three are translational.

Table 2.2. Number of vibrations for nonlinear and linear molecules

Molecule	Degrees of freedom
nonlinear	3n-6
linear	3n-5

A classical dispersive infrared spectrometer consists of a radiation source, monochromator, and detector. While using dispersive instruments, some limitations can be observed, such as analysis needs long scan times, detector response is slow, the source is weak, and broadened peaks are obtained that affect the spectra' quality. Instead of visualizing each component's frequency sequentially, like in a dispersive IR spectrometer, all frequencies are examined simultaneously in FTIR with the developments of interferometer. The most used interferometer is a Michelson interferometer. It comprises three active components: a moving mirror, a fixed mirror, and a beam splitter. Both mirrors are vertical. In an FT system, three fundamental components exist. These are; radiation source, interferometer, and detector. In contrast to the dispersive instrument, the radiation source is more often water-cooled to provide better power and stability, and

a monochromator is replaced with the interferometer. Instrumentation of Fourier transform infrared spectrometer is shown in Figure 2.3.

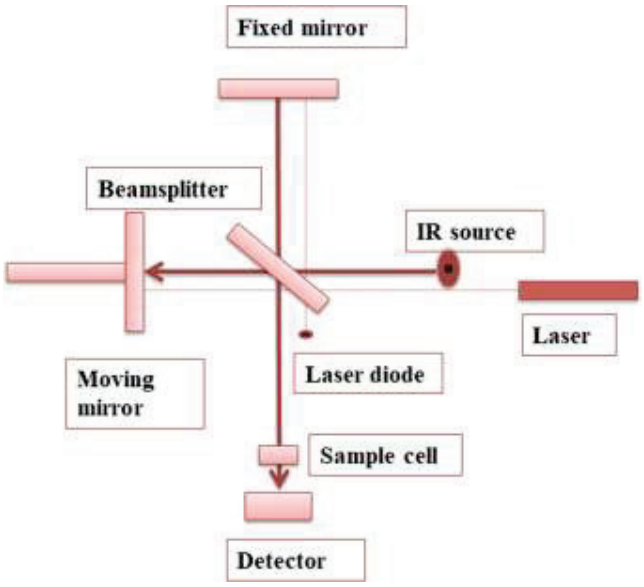


Figure 2.3. Schematic illustration of Fourier transform infrared spectrometer

(Source: (Titus et al., 2019)

The principle of FT system is radiation from the broadband IR source is collimated and delivered into the interferometer and impinges on the beam splitter. Half of the infrared beam is transmitted to the fixed mirror at the beam splitter, and the moving mirror reflects the remaining half. When the beams are divided and reflected by the two mirrors, they are recombined in the beam splitter. An interference pattern is generated due to Infrared changes in the moving mirror's corresponding position to the fixed mirror. The ensuing beam then passes through the sample and is finally directed to the detector (Suart, 2004).

2.2. Attenuated Total Reflectance (ATR)

Attenuated total reflectance (ATR) is an accessory used as a sampling technique to obtain high-quality data of various samples' infrared spectra. Samples that can not be easily examined through the usual transmission method can be analyzed with the Attenuated total reflectance (ATR) method. The ATR method is most suitable for studying thick or highly absorbing solid and liquid materials, such as films, coatings,

powders, threads, and polymer samples (Bağcıoğlu, 2011). Additional or sample preparation is not necessary with the ATR method, making it one of the versatile sampling techniques. Various types of crystals used in the ATR accessories change according to the samples to be analyzed. Characteristic ATR crystal materials and their features are listed in Table 2.3. Properties of some common ATR Crystals.

Table 2.3. Properties of some common ATR Crystals

(Source: Table courtesy of ThermoNicolet, Madison, WI.)

Crystal	Spectral Range (cm^{-1})	Refractive index	Depth	Uses
Germanium	5,500-675	4	0.66	Good for most samples; strongly absorbing samples such as dark polymers
Silicon	8,900-1500	3.4	0.85	Resistant to basic solutions
AMTIR[b]	11,000–725	2.5	1.77	Very resistant to acidic solutions
ZnSe	15,000–650	2.4	2.01	General use
Diamond	30,000–200	2.4	2.01	Good for most samples, extremely caustic or hard samples

The theory behind the ATR working principle is that some light is reflected in the sample medium when the first incident beam (IR beam) passes from a medium of crystal with a high refractive index to a medium of the sample that has a low refractive index. Almost the whole of the light waves is reflected at a distinguished angle of incidence. This phenomenon is described as a total internal reflection. In this case, some light energy gets out from the crystal and travels a short distance (0.1-5 μm) beyond the surface like waves. This invisible wave is known as an evanescent wave. The strength of reflected light is decreased at this point by the sample. This phenomenon is known as Attenuated total reflectance. This evanescent wave is a standing wave that occurs at the inflection point of the propagating light. After interaction with the sample, it carries the chemical information along with the detector. Some infrared radiation is absorbed by penetrating away the crystal, and the absorption is transformed into the infrared spectrum of the sample. A schematic illustration of the ATR principle is given Figure 2.4. The crystal's location must be positioned parallel to the ATR accessory's top surface. Actual physical contact must exist between the sample and high refractive index crystal such as zinc selenide, germanium, silicon, or diamond.

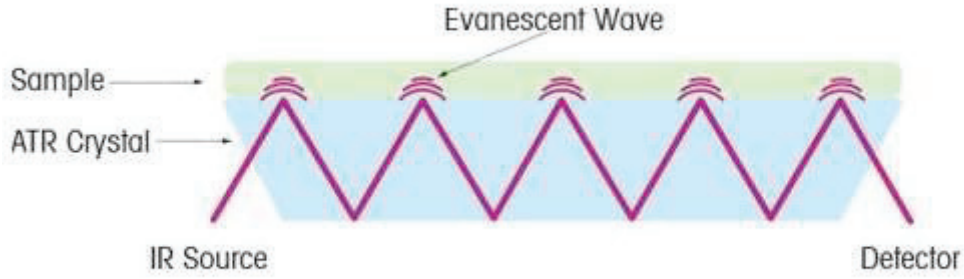


Figure 2.4. Schematic illustration of the ATR principle.

(Source: Attenuated Total Reflectance (ATR) | For FTIR Applications. (n.d.))

2.3. X-ray Fluorescence Spectroscopy

X-ray fluorescence (XRF) is used to analyze samples' elemental form at room temperature. XRF has many advantages compared to other elemental analysis techniques, such as the non-destructive method, and samples to be analyzed can also be in various forms, such as powders, crystals, and liquids. Consequently, XRF is carried out to a massive range of samples, including chemical, biological, environmental. (Skoog et al., 2017)

The X-rays are situated on the high energy side of the ultraviolet of the electromagnetic spectrum. The X-ray wavelengths are from $1x10^{-11}$ to $1.0\ x10^{-8}$ m (0.01 to 10 nm), which corresponds to the 0.25 keV to 125 keV area. The X-ray location on the electromagnetic spectrum is given in Figure 2.5 in terms of frequency (Hz) and wavelength (m).

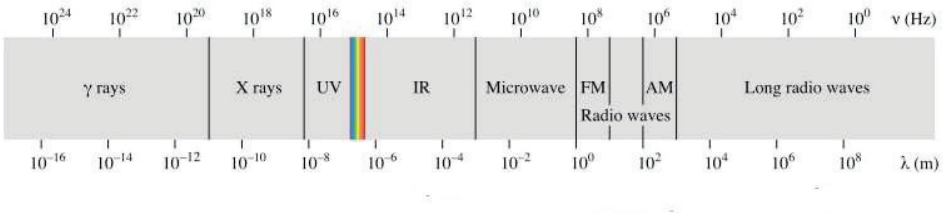


Figure 2.5. Regions of the electromagnetic spectrum

In the XRF, a sample is bombarded by an X-ray beam that emits secondary X-rays at lower energy, known as fluorescent radiation. It differs from conventional

fluorescence by not using visible light with poorly energetic and light-active molecules. A schematic illustration of fluorescent radiation is showed in Figure 2.6.

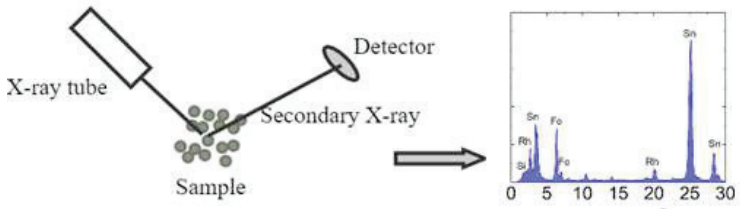


Figure 2.6. Schematic illustration of fluorescent radiation

Primary X-rays are produced by the X-ray source, generally an X-ray tube. When the sample is treated with an X-ray beam, generation of secondary X-ray is occurred by movement of atoms in the sample. If a photon of enough energy hits an atom, the energy is absorbed and excites one of the shell's outer electrons. As the electron relaxes, a secondary photon is emitted, and the secondary photons usually have low energy. This phenomenon is known as fluorescence. Low-energy photons are used in fluorescence microscopy; however, X-ray photons have sufficient energy to complete tight electrons' expulsion from an inner shell. When the excessive-energy primary x-ray collides with an atom, an electron at a high energy shell will fall into vacancy to re-establish its stability. The movement of electrons between two levels causes the release of excess energy, and it is emitted in the form of secondary X-rays specific for each element. This phenomenon is known to be X-ray fluorescence. It is used to identify the elemental composition of the sample and the amount found in the sample (X-Ray Fluorescence (XRF) | Protocol, n.d.). Figure 2.7, provides an illustration of the generation of X-ray fluorescence radiation.

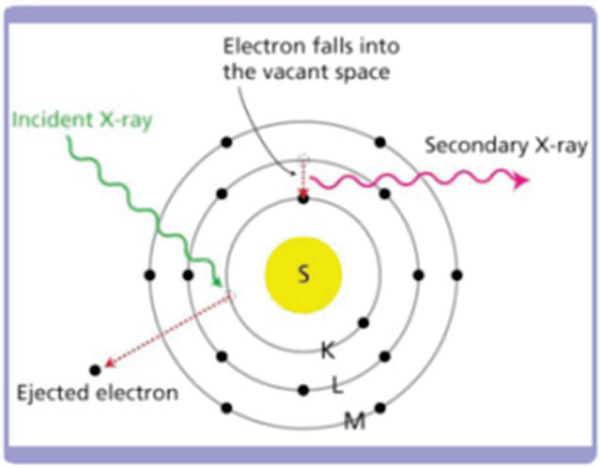


Figure 2.7. An illustration of the generation of X-ray fluorescence radiation (Source : www.projects.library.villanova.edu)

Instrumentation of X-ray spectrometry consists of a source, wavelength-energy selector, detector, collimator, and filter. In XRF spectrometry, optical system properties vary with each. For instance, the characterization of elements that emit X-rays is performed by measuring the emitted X-rays' energies or wavelengths. Therefore, the instrumentation is divided into energy dispersive (EDXRF) and wavelength dispersive (WDXRF) x-ray fluorescence analysis, depending on the detection method. In Figure 2.8 and Figure 2.9, basic schematic set-ups of an energy-dispersive and wavelength-dispersive X-ray fluorescence spectrometer are given, respectively.

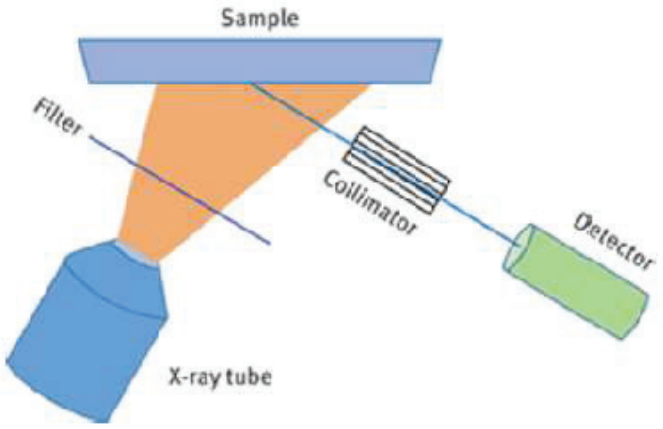


Figure 2.8. A schematic setup of an energy-dispersive X-ray fluorescence spectrometer with direct excitation

(Source: (Schramm, 2016))

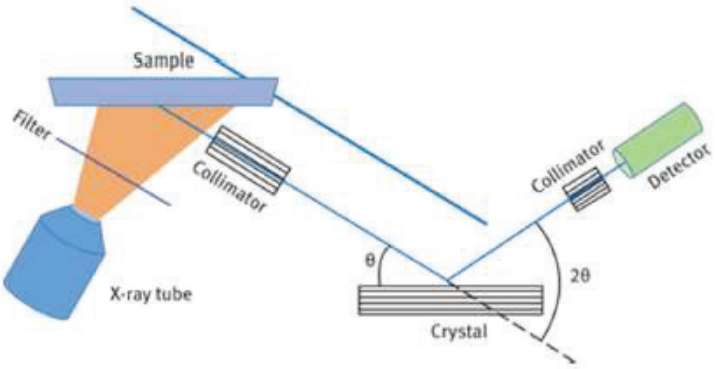


Figure 2.9. Schematic setup of a wavelength-dispersive X-ray fluorescence spectrometer.

(Source: (Schramm, 2016))

In the type of wavelength-dispersive analysis (WDXRF), a dispersive device is used to separate X-rays by deflecting them at different angles comparable to their wavelength. In energy-dispersive mode (EDXRF), each detected X-ray photons are records and measured by a detector, and a dispersive device is not used. With WDXRF systems, it may be viable to have many detector assemblies placed at constant angular places to examine some chosen elements repeatedly. WDXRF spectrometers usually provide more flexibility for the analyst and excellent sensitivities. Detector outputs are also more manageable to use directly. Disadvantages consist of the inability to quickly accumulate the whole x-ray spectrum for complete-element analyses, more expensive hardware prices, and a larger instrumental footprint than EDXRF systems.

CHAPTER 3

MULTIVARIATE CALIBRATION METHODS

3.1. Overview

Modern spectroscopic instruments can take many spectra of samples that have complex components over a short time. Univariate calibration methods are not suitable for those samples which have complex components. Multivariate calibration methods include applying mathematical methods to chemical analysis when each sample has more than one instrument signal. In spectroscopy, multivariate calibration is usually concerned about data comprised of instrument signals over multiple wavelengths for a multi-component sample. Multivariate calibration methods cope with some of the problems encountered in univariate calibration methods. In determining concentration from absorbance values, multivariate calibration can be seen an extension of Beer-Lambert's law, which may use the entire or partial spectra for calibration. It is not possible for the univariate calibration method. Other concerns are random interferences caused by interfering species and noises coming from instrument signals are decreased by averaging useful information on absorbance responses to predict more reliable and better calibration models (Brereton, 2000).This chapter will discuss the most used multivariate calibration techniques and their benefits, disadvantages, and implementation strategies.

3.2. Classical Least Squares (CLS)

Classical least squares (CLS) is one of the fundamental multivariate calibration techniques known as Beer's Law, where the absorbance value changes while depending on the concentration. The following Equation models CLS:

$$\mathbf{A}_{mxn} = \mathbf{C}_{mxl} \cdot \mathbf{K}_{lxn} + \mathbf{E}_{A_{mxn}} \tag{3.1}$$

In Equation 3.1, the *mxn* matrix of absorbance values is represented by **A**, *mxl* matrix of concentration is represented by **C**. Here, *m* is the number of samples in calibration set, *n* is the number of wavelengths, *l* is the number of components in a sample.

In this Equation, the unknown is the **K** matrix, which is the *lxn* matrix of absorptivity coefficients that depend upon absorbance values at *n* number of wavelengths to the concentration of *l* components of the *m* number of the calibration set. While using the least-squares method, Estimation of the **K** matrix is done by the described Equation 3.2.

$$\mathbf{K} = (\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{A} \tag{3.2}$$

After the calculation of the **K** matrix with Equation 3.2, the predicted component concentrations of an unknown sample from its spectrum can be found by Equation 3.3 as:

$$\mathbf{c} = (\mathbf{K}\mathbf{K}')^{-1}\mathbf{K}\mathbf{a} \tag{3.3}$$

a is the vector of absorbance values of the unknown sample obtained from its spectrum, and **c** is the vector of predicted concentration of components in the unknown sample. To obtain the best calibration model with the CLS method, all species' (components) concentrations with a significant effect on the spectrum in the calibration samples must be identified and included in the **C** matrix because the calculated coefficients are depended on all compositions of the samples. As all species' concentration is generally unknown, the response of instrument signal is used with these interfering species during modeling. This feature is the main disadvantage of the CLS method and can be eliminated by Inverse Least Squares (ILS) method.

3.3. Inverse Least Squares (ILS)

The necessity for the knowledge of the concentration of every interfering species, which is stated as the main drawback in the CLS method, can be solved by the Inverse Least Square(ILS) method, which is based on the multivariate extension of inverse Beer-Lambert's law and assumed that concentration is a function of absorbance.

$$\mathbf{C}_{mxl} = \mathbf{A}_{mxn} \cdot \mathbf{P}_{nxl} + \mathbf{E}_{c_{mxl}} \tag{3.4}$$

In this Equation 3.4 , **C** is the concentration matrix, and **A** is the absorbance matrix. **E** is the matrix of errors in concentrations not fitted by the model. Model error is assumed to derive from errors in the measurement of component concentration caused by the samples used for calibration, not the absorbance values. Modern spectroscopic

instruments provide excellent signal-to-noise(S/N) ratios with these instruments' development over the last decades. In Equation 3.4, **P** is the *nxl* matrix of regression coefficients associated with the absorbance values to the concentrations of the components in the calibration set. The solution of P is calculated as in Equation 3.5.

$$\mathbf{P} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' \times \mathbf{C} \tag{3.5}$$

The difference between ILS and CLS is, The ILS model has the advantage of modeling one component at a time, as given in Equation 3.6.

$$\mathbf{p} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' \times \mathbf{c} \tag{3.6}$$

p is an *nx1* vector of the regression coefficient for the component being modeled, and **c** is an *mx1* vector of concentration of the component modeled in the calibration set. After the **p** vector is reached, the prediction of the unknown sample's concentration is calculated by Equation 3.7.

$$c = \mathbf{a} \times \mathbf{p} \tag{3.7}$$

Although this method's strength is that without knowing the concentration of all interfering species, it can obtain a calibration model for a single component in a sample. Also, there are some problems with the ILS method. As shown in Equation 3.6, the absorbance matrix of **A**, which must be inverted while obtaining the p vector, has a longer dimension than the dimension of the concentration vector of **c**. Moreover, it contains the collinearity problem arising from the nature of the spectrum. Therefore, the number of columns equal to the number of wavelengths in the spectra of the matrix of **A** can not exceed the size of the vector of **c**, which is equal to the number of experiments or samples in the calibration set. Otherwise, makes ILS models overfitting can cause predicted concentrations are not reasonable. One solution to these problems is performing feature selection methods such as Genetic algorithms (GA) by leaving out the absorbances at wavenumber, which are unrelated to the concentrations. Another common solution based on principal component analysis (PCA) is using factor-based projection methods such as the partial least squares (PLS) method. (Fredrick Lindgren et al., 1994; Fredrik Lindgren et al., 1995). In the following section, brief information about these two methods is described.

3.4. Partial Least Squares (PLS)

In the late sixties, the pioneering studies in Partial Least Squares (PLS) were done by H.Wold in econometrics. PLS is becoming a significant regression method in many chemistry disciplines; analytical, physical, clinical chemistry, and industrial process control. For chemical applications, the groups of S. Wold and H. Martens led to the use of the PLS method in the seventies after a primary application by Kowalski et al. (Kowalski et al., 1982).Many articles (Otto & Wegscheider, 1985; Wold et al., 1983) show that PLS is a good choice to other regression methods as multiple linear regression and principal component regression because it is more robust. Also, there are many completed studies on PLS conducted by S. Wold et al. Besides, while other calibration methods assume that errors are in the samples' response or properties, PLS considers both cases .(Geladi & Kowalski, 1986)

Partial least square is a technique that creates new variables that are linear combinations of the original data by decomposing the data. These new variables are called principal components or factors. The way in which new variables are created can be visualized in a two-dimensional system. If the instrumental signals for a set of *m* samples at two wavelengths(*n*=2) are plotted against each other, the new axis is formed to represent the data's maximum variability. This new axis is called the first principal component or first eigenvector. Otherwise, the second eigenvector is perpendicular or orthogonal to the first eigenvector. The second eigenvector specifies the maximum amount of residuals not fit by the first one in the data set. If the instrumental signal matrix contains more than two wavelengths, the plotting space becomes multidimensional. Several eigenvectors can be found; each represents the maximum remaining variability and each orthogonal to the others.

The PLS model equation is described as:

$$\mathbf{A}_{mxn} = \mathbf{T}_{mxh} \cdot \mathbf{B}_{hxn} + \mathbf{E}_{A_{mxn}} \qquad \textbf{(3.8)}$$

A is an *mxn* matrix of spectral absorbance, **B** is a *hxn* matrix of loading vectors or loading spectra, and **T** is an *mxh* matrix of intensities or scores in the new coordinate system of the *h* loading vectors for the *m* sample' spectra. $\mathbf{E_A}$ is the *mxn* matrix residuals not fit by the model. The difference between the CLS and PLS methods is that **B'**s loading vectors

are not pure component spectra. They are linear combinations of original calibration spectra. Also, intensities in the new coordinate system are not dependent on concentrations as in CLS. However, the modeling is based on the relationship between scores in **T** and component concentrations. The algorithm determines the required number of basis vectors, *h*, (representing calibration spectra), during the calibration step. In the new coordinate system, spectral intensities are correlated with analyte concentrations using the ILS model. The concentration of the analyte is calculated by:

$$\mathbf{c} = \mathbf{Tv} + \mathbf{e_c} \tag{3.9}$$

In this equation 3.9, **c** is the *mx1* vector of component concentration, **v** is the *hx1* vector that relates spectral intensities to component concentrations, and $\mathbf{e_c}$ is the *mx1* vector of errors in reference values of analyzed components. The solution for **v** is the least-squares solution as in ILS:

$$\hat{\mathbf{v}}_\mathbf{h} = (\mathbf{T}'\mathbf{T})^{-1}\mathbf{T}'\mathbf{c} \tag{3.10}$$

Where $\hat{\boldsymbol{v}}_{\boldsymbol{h}}$ is the least-squares estimate of **v**. **T** and **B** matrices are calculated in a stepwise manner (one vector at a time) until the desired model has been obtained. The modified version of the NIPALS (nonlinear iterative partial least squares) algorithm is used to calculate loadings vectors with the concentration of components in PLS. There two different versions of PLS algorithms, which are PLS1 and PLS2. One component is examined at a time in PLS1, while all component is examined at a time in PLS2. That's why, PLS1 exhibits better predictive power than PLS2. Also, PLS1 is recommended for quantitative analysis, while PLS2 is recommended for qualitative analysis.

In the PLS1 algorithm, inverse classical least squares calibration is performed, assuming that only one component is known in the calibration samples. Then, the least-squares method estimates the first weighted loading vector, $\hat{\mathbf{w}}_\mathbf{h}$, with the substitution of 1 for h in following Equation 3.11.

$$\widehat{\mathbf{w}}_\mathbf{h} = \mathbf{A}'\mathbf{c}(\mathbf{c}'\mathbf{c})^{-1} \tag{3.11}$$

$\hat{w}_h$ is an *nx1* vector that represents the least square estimation (first-order approximation) of the pure component spectrum for *h*=1 of the interested component. The next step is forming the score (latent variable) vector $\hat{\mathbf{t}}_\mathbf{h}$ by using the weighted loading vector with an ILS prediction model as given as:

$$\hat{\mathbf{t}}_\mathbf{h} = \mathbf{A}\widehat{\mathbf{w}}_\mathbf{h} \tag{3.12}$$

Using a linear least square regression, the score vector representing the new PLS coordinate system's intensities can be associated with the concentrations. Equation 3.13 gives the estimate for $\hat{\mathbf{v}}_\mathbf{h}$ which is the scalar regression coefficient relating to the concentration of the component that is being analyzed.

$$\hat{\mathbf{v}}_\mathbf{h} = \hat{\mathbf{t}}_\mathbf{h}{}'\mathbf{c}\left(\hat{\mathbf{t}}_\mathbf{h}'\hat{\mathbf{t}}_\mathbf{h}\right)^{-\mathbf{1}} \tag{3.13}$$

Orthogonal $\hat{\mathbf{t}}_\mathbf{h}$ vectors are desirable to eliminate the collinearity problem caused by original inverse least square regression and can be acquired by forming a new model for A. Also, the calculated regression coefficient is used to obtain concentration residuals. Once more, the least square method is performed to find b_h, which is the PLS loading vector is given in Equation 3.14.

$$\hat{\mathbf{b}}_\mathbf{h} = \hat{\mathbf{t}}_\mathbf{h}{}'\mathbf{A}\left(\hat{\mathbf{t}}_\mathbf{h}'\hat{\mathbf{t}}_\mathbf{h}\right)^{-\mathbf{1}} \tag{3.14}$$

where $\hat{\mathbf{b}}_\mathbf{h}$ is the *nx1* PLS loading vector, it is possible to calculate the first PLS approximation to the concentrations related to calibration spectra by multiplying the score vector $\hat{\mathbf{t}}_\mathbf{h}$ with the transpose of the PLS loading vector. Spectral residuals, E_A, in the calibration spectra **A** are calculated by subtracting the PLS approximation from the measured calibration spectra A as given in Equation 3.15.

$$\mathbf{E}_\mathbf{A} = \mathbf{A} - \hat{\mathbf{t}}_\mathbf{h}\hat{\mathbf{b}}_\mathbf{h}' \tag{3.15}$$

Also, concentration residuals, $\mathbf{e}_\mathbf{c}$, is obtained by multiplying scalar regression coefficient $\hat{\mathbf{v}}_\mathbf{h}$, with the score vector $\hat{\mathbf{t}}_\mathbf{h}$, and substracting the obtained result from the original concentration vector as given in Equation 3.16.

$$\mathbf{e}_\mathbf{c} = \mathbf{c} - \hat{\mathbf{v}}_\mathbf{h}\hat{\mathbf{t}}_\mathbf{h} \tag{3.16}$$

PLS1 algorithm for obtaining concentration predictions involves calculating and using the vector of the final calibration coefficient, $\mathbf{b}_\mathbf{f}$. Once b*f* is calculated, it is possible to calculate the new sample's concentration using average analyte concentrations and its spectrum. Equation 3.17 shows the calculation of the final regression coefficient $\mathbf{b}_\mathbf{f}$ for the prediction in PLS1.

$$\mathbf{b}_\mathbf{f} = \widehat{\mathbf{W}}\left(\widehat{\mathbf{B}}\widehat{\mathbf{W}}'\right)^{-\mathbf{1}}\hat{\mathbf{v}} \tag{3.17}$$

Where $\widehat{\mathbf{W}}$ and $\widehat{\mathbf{B}}$ are comprised of vectors $\widehat{\mathbf{w}}_\mathbf{h}$ and $\hat{\mathbf{b}}_\mathbf{h}$, respectively, and $\hat{\mathbf{v}}$ is formed from individual regression coefficient, $\hat{\mathbf{v}}_\mathbf{h}$. The $\mathbf{b}_\mathrm{f}$ ensures an effective method to estimate concentrations from any unknown sample spectrum. The final estimation, Equation 3.18, is given below.

$$\hat{c} = \mathbf{a}'\mathbf{b}_\mathbf{f} + c_o \tag{3.18}$$

Here, c is the estimated unknown sample concentration, **a** is the spectrum of this sample, and c_o is the average concentrations of the calibration samples. (Haaland & Thomas, 1988)

One of the most critical concerns for the PLS model is selecting the optimum number of factors. The selection of the optimum number of factors allows us to model data with complexity without overfitting the concentration data. Cross-validation method leaving out one sample at a time is used to achieve this aim (Stone, 1974). For a set of m calibration spectra, the PLS1 calibration is performed on m-1 calibration spectra and with this calibration, prediction of the allocated sample's concentration is done. This process continues m times until each spectrum in the calibration set is used to test once. Then predicted concentration for every sample is compared with the initial concentration of the reference sample. Prediction error sum of squares or PRESS is a measure of how well the certain PLS model fits the concentration data. PRESS is determined in the same way each time a new factor is introduced to the PLS1 model. The equation of PRESS is given in below.

$$\mathbf{PRESS} = \sum_{i=1}^{m} (\hat{c} - c_i)^2 \tag{3.19}$$

The most appropriate PLS factor number should be selected according to the lowest PRESS value. Otherwise, we may face a model that gives excellent calibration results but cannot produce the same results at the prediction step. A comparison between the h^{th} and h+1$^{\text{th}}$ factors is usually made to those using the F-test method to avoid this situation. If the F-test method indicates no significant difference, the better model is usually the one obtained with a smaller number of factors. Figure 3.1 shows the PRESS values obtained as a result of cross-validation corresponding factor numbers are given.

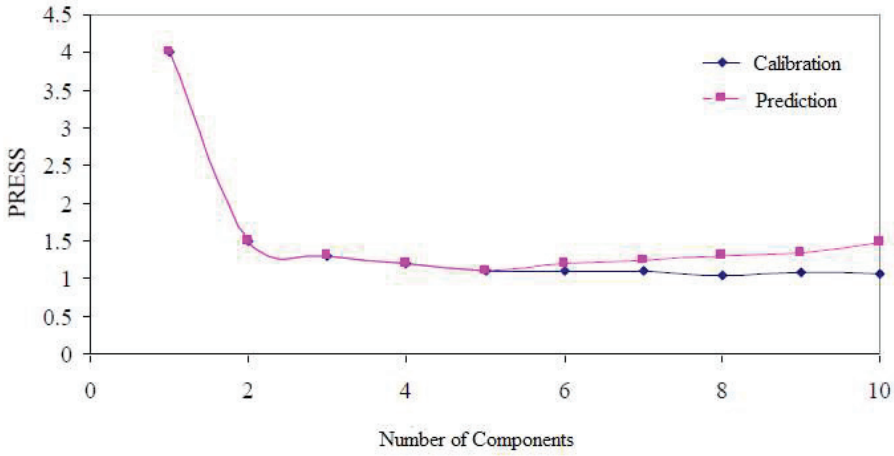


Figure 3.1. An example of a number of components vs. PRESS plot.

(Source: Özdemir Durmuş, 2020)

In Figure 3.1given above, PRESS values are given against the number of factors used to determine the number of PLS factors required to create the most suitable PLS model. As shown from the figure, the PRESS values obtained for the model show a decreasing trend, but after the fifth factor, the PRESS values in the prediction step increase. In this case, the most appropriate factor number should be 5

CHAPTER 4

EXPERIMENTATION AND METHODS

4.1. Samples

Samples for the study are obtained from the quality control laboratory in the BatıÇim factory at İzmir. Batı Anadolu cement Industry A.Ş. was established in 1966 with 100% of the Turkish capital. It continues its route inside the Batı Anadolu group, underneath its roof the Turkish cement industry's robust enterprises today.

4.2. FTIR-ATR Measurements

Fourier transform infrared spectrometer (PerkinElmer Spectrum 100 FT-IR) coupled with attenuated total reflectance (PIKE MIRacle) accessory as a sampling technique was employed to analyze the 113 samples that vary with raw materials, the intermediate product of cement (clinker), and cement samples. The diamond crystal was chosen because it generally is proper for hard samples. The spectrum analysis was performed between the wavenumber region 4000 cm^{-1} and 550 cm^{-1}. The optimum resolution was decided as an 8 cm^{-1} (data point interval 2 cm^{-1}), and 16 scans were done. For each sample, triplicate measurements were taken to limit spectral errors caused by sample particle size. An average spectrum of three measurements was calculated to use future in multivariate analysis. Before starting the sample analysis, the background measurement against air was done with the same parameters as the sample's measurements to decrease contamination caused by ATR crystal. For background, the empty, clean crystal was used, and after each measurement, the sample was removed from the sample holder part of the ART crystal by a vacuum cleaner. Then ethanol was used to clean the crystal surface and was allowed to dry.

4.3. X-ray Fluorescence Spectrometer Measurements

Elemental analysis of raw materials, raw meals, intermediate products, and cement is performed with an X-ray fluorescence spectrometer for most cement factories. XRF analyses of the study samples are performed at the BATIÇİM quality control laboratory with an X-ray Fluorescence spectrometer. Determination of elemental analysis of the products is a crucial procedure for the quality control stage and manufacture. Because the product's elemental composition constitutes their chemical composition that influences cement's performance, the nature of the raw materials used in the production of cement can significantly change the final cement's chemical composition. cement chemical analysis can thus be used to assess the cement's consistency and the applications to which it is ideally suited. The elemental ratios are detected in the sample by converting into weight fractions of each oxide present. The performance of XRF is strongly dependent on sample preparation. The measurements are achieved with XRF generally requires careful and controlled sample preparation, and analysis times take about 15 min per sample, changing with what type of sample is to be analyzed. The material types, such as raw materials and final products, can be analyzed at different time scales. Also, XRF presents high costs. Furthermore, the usage of radioactive source can be an issue for safety. Consequently, there may be room for improving the latest analytical techniques as an alternative technique for XRF that can carry out quicker, more secure, at a lower fee, and meet the requirements for quality control of cement samples.

There are two types of sample preparation techniques for XRF analysis. Pressed pellet techniques and glass bead preparation techniques are used for sample preparation. The pressed pellet technique is applied to the cement samples and clinker samples that are intermediate cement products, and glass bead preparation is applied to the raw materials.

In this study, the samples obtained from the BatıÇim cement factory were analyzed in the quality control laboratory by X-ray fluorescence spectrometer with two different sample preparation techniques that change with sample types.

4.3.1. Pressed Pellet Technique

The method for forming pressed pellets for XRF analysis consists of five main stages: milling the sample to fine grain size, mixing it with a binding agent in a grinding vessel, pouring the mixture a pressing die, and pressing at constant a particular pressure. Illustration of all the steps in producing a pressed pellet are given in Figure 4.1. Making pellet technique was applied to our samples that are cement, clinker, raw meal. The specific procedure is given in below.

1. The sample was weighed to 20 g.
2. The grinding or tableting aid was added to the sample. The amount of grinding is to be added changes with the specificity of the sample
3. The sample and tableting aid were ground to fine particle size by pulverizing mill. Small particle size is a critical component in producing pellets that provide the best analytical results as it affects how the sample will compress and bind together when pressed.
4. After grinding, the mixture was poured into a mixing vessel, and triethanolamine was added to the mixture as a binding agent.
5. The mixing vessel was placed into pulverizing mill. The sample and bind were ground together into the mill.
6. The 5.0 g of the mixture was poured into a pressing die and pressed at a particular constant pressure.
7. The resulting pellet was then prepared for analysis.

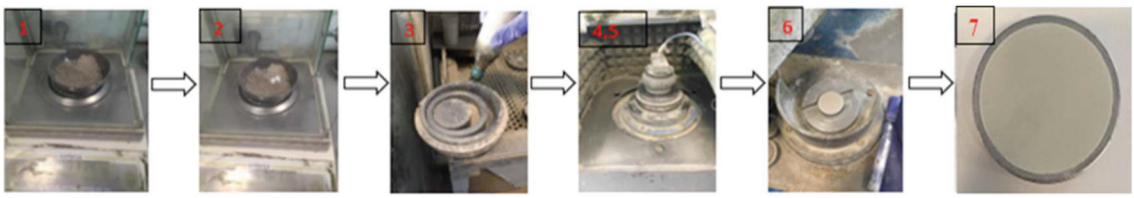


Figure 4.1. Illustration of all the steps in producing a pressed pellet with a binder

4.3.2. Glass Bead Preparation

Glass bead preparation for XRF analysis from the samples used for cement manufacturing has many advantages over pressed pellets. Forming samples into glass beads involves three main stages. First, a flux is used to mix with the sample. Second, melting the sample into glass beads at very high temperatures. Last, molten mixture can be formed into discs using a mold. Illustration of the steps in producing glass beads are given in Figure 4.2. There are challenges associated with the preparation of glass beads from cement samples. Identifying the right preparation strategy usually involves trial and error to find the best fusion parameters, ratios, and additives. The specific procedure used in the study is given in below.

1. The sample was to be analyzed, weighed, and tetraborate as a flux also weighed to specific amounts.
2. The sample and flux were mixed and transferred into a platinum crucible. Then, Calcium iodide was added to make an easy melting process.
3. The mixture was melted in a platinum crucible at a very high temperature, such as 900-1200 °C. While the mixture sample was melting, at the same time, the shaking process was constantly applying.
4. The molten mixture was formed into a disc using a mold.
5. The resulting glass bead was then prepared for analysis

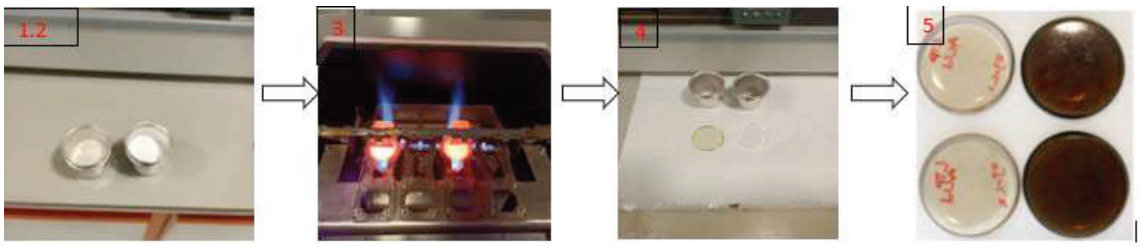


Figure 4.2. Illustration of the steps in producing glass beads

4.4. Loss of Ignition Analysis

The amount of weight lost by raising the temperature of raw material, cement, to a predetermined level is known as ignition loss. Loss of ignition analysis of the studied samples was made according to Turkish standards, TS EN 196-2.

Weight loss during heating can be caused by evaporation or volatilization of various components of the sample. Water vaporizes at 100-105 °C, organic matter is burned at approximately 550 °C, and most carbonates can be lost at 800-1000 °C. The procedure of the loss of ignition analysis is given in below

.

1. The sample was weighed 1.0000 ± 0.0005 g of on the analytical balance.
2. The sample was taken into a porcelain crucible, brought to a fixed weight, and had a fixed tare.
3. Put the porcelain crucible in the 950 °C muffle furnace and let stand 30 minutes.
4. After the porcelain crucible with sample has a fixed weight, the porcelain crucible was taken to the desiccator.
5. The porcelain crucible containing the sample was weighted After the crucible has cooled.

4.5. Data Analysis

The spectra obtained from the FT-ATR system were transferred to Microsoft Excel 365. The calibration data set and independent validation data set were created to predict the chemical composition of samples which are cement, clinker, raw meal, raw materials. The calibration data set was used to construct multivariate calibration models, and the independent validation set was used to train the predictive power and reliability of the created calibration models. The partial least squares (PLS) method was applied to the data to construct calibration models by chemometrics calibration toolbox belongs to OBA quantifier, OBA Chemometrics Inc Turkey (OBA Kemometri) in the environment of MATLAB R2018b. (Math Works Inc., MA). Different scenarios were tried to construct calibration models with the best predictive powers to predict cement's chemical composition as an alternative to XRF.

CHAPTER 5

RESULTS AND DISCUSSION

All samples obtained from the BatıÇim cement factory were analyzed by Fourier transform infrared spectrometer (FTIR) coupled with attenuated total reflectance accessory (ATR). The data acquired from the FTIR-ATR system were combined with multivariate calibration methods and used to determine major oxides that comprise the chemical composition of cement. The elemental reference analyses of samples were performed by X-ray fluorescence (XRF). For the thesis study, the different scenarios were developed to determine cement composition and process optimization.

The first scenario was that the studied samples were diversified using all kinds of samples used to manufacture cement, and the sample analysis was required rapid analysis times to determine the concentration of major oxides for the process optimization in the quality control stage. Types of samples, such as cement, raw materials, limestone, raw meals, and clinker samples, were included in the model. In conclusion, a calibration model was tried to construct using various types of samples that significantly affected the manufacturing stage of cement for the first scenario.

The data set was arranged into two groups as the calibration set and the independent validation set. Details of the data for each scenario were explained in the specific part of the scenarios. The partial least square (PLS) method was applied to the data for two scenarios.

5.1. The First Scenario.

A total of 113 samples was diversified as raw materials and types of cement. The 113 samples were consisted of 51 cement samples, 20 clinker samples, 11 of limestone, 9 of gypsum, 6 of raw meal, 5 of clay, 5 of iron ore, 3 of them trass, and 3 of the ash. The data set was divided into two groups; 89 samples were included in the calibration data set, and the remaining 24 samples were included in the independent validation set. (80% for calibration data set and 20% for validation data set). The data sets were arranged considering, the validation and the calibration set have samples with similar composition.

The validation data set was established to cover all types of samples. The calibration data set and validation data set were given with the percentages of major oxides comprise of cement composition in Table 5.1and Table 5.2, respectively.

Table 5.1. The calibration data set of the first scenario

Calibration Data Set										
No	**Type**	**Name**	**CaO (w/w%)**	**SiO_2 (w/w%)**	**Al_2O_3 (w/w%)**	**Fe_2O_3 (w/w%)**	**MgO (w/w%)**	**SO_3 (w/w%)**	**Na_2O (w/w%)**	**K_2O (w/w%)**
1	Iron ore	S42	0.61	50.53	1.15	42.42	0.05	0.72	0.02	0.09
2	Iron ore	S4	1.41	29.86	1.85	59.38	0.12	1.28	0.35	0.13
3	Iron ore	S33	4.36	32.82	2.51	48.77	0.21	0.93	0.41	0.21
4	Iron ore	S43	4.58	31.84	1.43	56.4	0.11	1.07	0.15	0.12
5	Clay	S9	5.33	54.22	15.64	5.88	2.09	0.14	2.41	2.52
6	Trass	S28	9.42	51.21	14.02	5.24	2.11	1.48	2.26	2.6
7	Clay	S70	5.6	52.05	16.2	6.13	2.31	0.16	1.11	3.2
8	Clay	S75	7.39	55.42	12.73	4.93	2.31	0.14	0.9	2.56
9	Clay	S73	9.22	49.11	14.46	5.83	2.53	0.35	1.11	2.71
10	Clay	S74	9.96	50.51	14	5.48	2.05	0.23	0.87	2.75
11	Ash	S37	29.86	37.44	17.66	4.3	2.16	5.09	0.39	1.21
12	Ash	S6	37.33	32.25	14.51	3.9	1.85	6.53	0.37	1.04
13	Gypsum	S19	33.22	1.21	0.29	0.1	2.93	37	0.12	0.06
14	Gypsum	S45	33.4	0.28	0.02	0.02	0.88	43.48	0.03	0.01
15	Gypsum	S76	34.02	0.82	0.13	0.11	0.27	42.45	0.05	0.02
16	Gypsum	S44	34.35	0.61	0.06	0.05	4.38	33.19	0.02	0.02
17	Gypsum	S49	34.97	0.76	0.1	0.12	0.06	41.99	0.05	0.02
18	Gypsum	S46	36.4	1.03	0.27	0.12	9.13	16.51	0.03	0.07
19	Gypsum	S14	36.5	0.74	0.13	0.12	3.54	33.56	0.04	0.03
20	Raw meal	S34	42.02	12.96	4.45	2.31	0.76	0.18	0.25	0.98
21	Raw meal	S36	42.05	12.98	4.42	2.28	0.75	0.18	0.6	0.98
22	Raw meal	S38	54.7	15.89	4.68	2.16	1.00	0.52	0.31	3.98
23	Raw meal	S39	55.27	16.28	4.74	2.39	0.99	0.5	0.28	3.38
24	Raw meal	S40	55.26	15.93	4.68	2.37	0.97	0.51	0.22	1.86
25	Limestone	S10	54.09	1.47	0.75	0.20	0.64	0.17	0.06	0.07
26	Limestone	S2	54.27	1.97	0.95	0.55	0.28	0.07	0.08	0.09
27	Limestone	S32	54.29	1.57	0.88	0.23	0.28	0.10	0.07	0.08
28	Limestone	S41	54.77	1.63	0.75	0.26	0.23	0.05	0.07	0.07
29	Limestone	S21	55.05	1.43	0.57	0.29	0.33	0.49	0.09	0.07
30	Limestone	S20	55.1	0.67	0.31	0.14	0.34	1.16	0.07	0.03
31	Limestone	S1	55.48	1.03	0.52	0.18	0.19	0.05	0.07	0.05
32	Limestone	S7	55.57	0.75	0.37	0.11	0.31	0.24	0.07	0.04
33	Limestone	S15	55.69	0.50	0.28	0.14	0.4	0.15	0.07	0.03
34	CEM IV	S71	48.38	28.82	7.84	3.74	1.45	2.8	0.69	1.18

(cont. on next page)

Table 5.1 (cont.).

35	CEM IV	S79	49.76	28.36	7.98	2.87	1.46	3	0.75	1.17
36	CEM II	S22	56.01	22.31	6.28	3.05	1.41	3.66	0.56	1.05
37	CEM II	S29	56.39	22.32	6.31	3.02	1.39	3.36	0.57	1.03
38	CEM-I	S80	57.4	22.39	7.35	2.66	1.23	3.42	0.39	0.77
39	CEM-II	S25	58.56	21	7.08	2.82	1.35	2.98	0.38	0.86
40	CEM-II	S24	58.67	20.8	6.98	2.95	0.36	3.05	0.38	0.85
41	CEM-I	S77	60.38	19.75	6.99	2.43	1.34	2.84	0.47	0.85
42	CEM-I	S99	61.96	21.23	5.76	2.1	1.9	3.07	0.21	0.9
43	CEM-I	S103	62	20.48	5.65	2.22	1.45	4.21	0.21	0.85
44	CEM-I	S111	62.33	20.82	5.76	2.32	1.47	3.11	0.22	0.88
45	CEM-I	S92	62.46	20.24	5.62	2.26	1.52	3.32	0.21	0.86
46	CEM-II	S26	62.58	21.11	7.28	2.63	1.42	3.11	0.46	0.9
47	CEM-I	S109	62.76	20.21	5.53	2.1	2.02	3.45	0.16	0.85
48	CEM-I	S110	62.76	20.52	5.66	2.18	1.58	3.3	0.18	0.86
49	CEM-I	S18	62.99	18.87	5.45	2.94	1.35	3.33	0.38	0.84
50	CEM-I	S90	63.02	19.99	5.54	2.12	1.59	3.57	0.18	0.83
51	CEM-I	S105	63.06	20.32	5.37	2.12	1.44	3.55	0.17	0.83
52	CEM-I	S12	63.07	18.92	5.44	2.91	1.36	3.29	0.39	0.85
53	CEM-I	S13	63.1	18.98	5.34	2.84	1.36	3.47	0.37	0.85
54	CEM-I	S97	63.25	20.36	5.41	2.17	1.67	3.21	0.17	0.85
55	CEM-I	S106	63.31	20.29	5.45	2.38	1.46	3.28	0.17	0.81
56	CEM-I	S84	63.34	20.24	5.4	2.12	1.67	3.24	0.16	0.82
57	CEM-I	S89	63.42	19.83	5.23	2.35	1.54	3.62	0.17	0.83
58	CEM-I	S93	63.45	19.81	5.47	2.91	1.49	3.17	0.18	0.9
59	CEM-I	S88	63.48	20.21	5.58	2.08	1.34	3.22	0.19	0.78
60	CEM-I	S104	63.48	20.21	5.58	2.08	1.34	3.22	0.19	0.78
61	CEM-I	S30	63.49	19.07	5.48	2.8	1.38	3.25	0.4	0.88
62	CEM-I	S102	63.65	20.2	5.37	2.29	1.28	2.8	0.18	0.81
63	CEM-I	S95	63.7	20.01	5.33	2.12	1.41	3.37	0.19	0.8
64	CEM-I	S100	63.75	20.29	5.26	2.18	1.3	3.28	0.17	0.75
65	CEM-I	S81	64.59	20.87	5.14	2.93	2.01	3.42	0.25	0.78
66	CEM-I	S113	62.27	18.94	4.89	2.86	2.48	3.39	0.49	0.79
67	CEM-I	S91	64.89	19.16	5.26	2.85	1.3	2.82	0.2	0.87
68	CEM-I	S94	64.96	21.63	5.93	2.26	1.49	2.65	0.21	0.86
69	CEM-I	S83	65.05	20.76	5.23	2.83	1.85	3.29	0.24	0.76
70	CEM-I	S85	63.38	19.82	5.14	2.66	1.43	3.52	0.19	0.83
71	CEM-I	S112	63.13	18.78	5.16	2.88	1.98	3.25	0.38	0.84
72	CEM-I	S96	63.16	20.44	5.53	2.14	1.23	3.01	0.21	0.86
73	CEM-I	S114	62.17	19.13	4.72	3.13	1.73	2.82	0.38	0.81
74	CEM-I	S82	66.16	20.32	4.83	3.05	1.8	2.9	0.26	0.67
75	Clinker	S60	65.75	21.49	5.81	3.8	1.04	0.35	0.34	0.92
76	Clinker	S50	65.82	21.47	5.84	3.64	1.06	0.39	0.35	0.91

(cont. on next page)

Table 5.1 (cont.).

77	Clinker	S56	65.89	21.43	5.76	3.65	1.01	0.46	0.34	0.94
78	Clinker	S57	65.91	21.64	5.65	3.49	1.03	0.47	0.34	0.94
79	Clinker	S16	65.92	21.21	5.74	3.62	1.05	0.41	0.34	1.01
80	Clinker	S52	65.96	21.44	5.76	3.57	1.01	0.39	0.34	0.93
81	Clinker	S54	66.03	21.43	5.75	3.59	1.04	0.4	0.33	0.92
82	Clinker	S58	66.09	21.55	5.56	3.43	1.02	0.52	0.34	0.96
83	Clinker	S67	66.16	21.49	5.64	3.46	1.01	0.44	0.34	0.95
84	Clinker	S66	66.2	21.46	5.7	3.41	1.01	0.43	0.34	0.94
85	Clinker	S53	66.22	21.32	5.71	3.49	1.09	0.4	0.33	0.92
86	Clinker	S55	66.24	21.21	5.72	3.5	1.09	0.51	0.31	0.9
87	Clinker	S68	66.24	21.56	5.61	3.37	1	0.44	0.34	0.93
88	Clinker	S63	66.27	21.21	5.73	3.49	1.08	0.6	0.32	0.94
89	Clinker	S59	66.38	21.15	5.72	3.52	1.04	0.4	0.33	0.96

Table 5.2. Validation data set of first scenario.

Validation Data Set										
No	**Type**	**Name**	**CaO (w/w%)**	**SiO_2 (w/w%)**	**Al_2O_3 (w/w%)**	**Fe_2O_3 (w/w%)**	**MgO (w/w%)**	**SO_3 (w/w%)**	**Na_2O (w/w%)**	**K_2O (w/w%)**
90	Iron ore	S3	4.36	32.82	2.51	48.77	0.21	0.93	0.41	0.21
91	Trass	S23	8.05	54.11	14.63	5.41	2.05	0.28	2.47	2.78
92	Clay	S72	7.27	53.74	13.83	5.34	2.24	0.18	0.97	2.75
93	Ash	S5	30.38	37.9	16.85	4.24	2.02	5.14	0.42	1.24
94	Gypsum	S48	34.21	0.3	0.03	0.02	3.53	34.92	0.02	0.02
95	Gypsum	S47	35.98	0.56	0.08	0.06	5.09	27.28	0.04	0.03
96	Raw meal	S35	43	12.58	4.19	2.16	0.7	0.14	0.25	0.89
97	Limestone	S8	54.87	1.34	0.5	0.97	0.2	0.06	0.07	0.06
98	Limestone	S69	55.11	1.56	0.51	0.21	0.23	0.02	0.07	0.05
99	CEM-II	S31	54.76	23.74	7.37	3.16	1.39	3.56	0.56	1.06
100	CEM-I	S78	56.14	22.48	6.34	3.17	1.45	3.2	0.58	1.04
101	CEM-II	S27	58.48	20.82	6.89	3.03	1.35	3.08	0.38	0.88
102	CEM-II	S17	60.4	19.97	5.63	3.07	1.32	3.45	0.44	0.9
103	CEM-I	S87	62.28	21.07	5.66	2.18	1.47	3.42	0.2	0.87
104	CEM-I	S108	62.5	20.32	5.69	2.28	1.46	3.55	0.2	0.84
105	CEM-I	S11	63.03	18.78	5.31	2.86	1.4	3.42	0.36	0.84
106	CEM-I	S101	63.28	20.18	5.54	2.16	1.36	2.79	0.19	0.84
107	CEM-I	S86	63.46	20.12	5.52	2.05	1.25	3.06	0.2	0.78
108	CEM-I	S98	63.53	20.41	5.27	2.15	1.38	3.22	0.17	0.83
109	Clinker	S61	65.87	21.41	5.76	3.79	1.05	0.34	0.36	0.9
110	Clinker	S51	65.92	21.38	5.76	3.63	1.06	0.38	0.41	0.92
111	Clinker	S65	66.18	21.33	5.68	3.65	1.02	0.39	0.33	0.92
112	Clinker	S64	66.25	21.28	5.73	3.49	1.08	0.41	0.31	0.93
113	Clinker	S62	66.29	21.3	5.7	3.49	1.04	0.37	0.33	0.97

The partial least squares (PLS) multivariate calibration method combined with FTIR-ATR spectroscopy has been used to determine the percent concentration of eight oxides: CaO, SiO_2, Al_2O_3, Fe_2O_3, MgO, SO_3, Na_2O, K_2O. Figure 5.1 illustrates the FTIR-ATR spectra of all 113 samples, which consist of raw materials, cement types, raw meals, and clinker between 4000-550 cm^{-1} wavenumber region. Each sample was analyzed in triplicate, and its mean spectrum was used to perform the PLS analysis.

Loss of ignition analysis of the samples was determined gravimetrically. Clinker samples were removed from the data because the loss of ignition analyses for the clinker samples do not perform in the quality control laboratory of the most cement factory. After all, clinker samples are produced by burning raw meals at high temperatures. They do not have humidity. Also, the loss of ignition result of some samples does not have. Therefore, these samples were also excluded from the data. Finally, loss of ignition results of 85 samples and their full range of spectra were mean-centered were used to construct the PLS model. The reference loss of ignition results was given in Table 5.3.The datasets were organized considering that the validation and calibration sets contain samples of similar composition. The validation dataset was designed to cover all types of samples.

Table 5.3. Calibration and Validation data set of first scenario for L.O.I

Calibration Data Set						Validation Data Set		
NO	Sample	Loss of Ignition (w/w%)	NO	Sample	Loss of Ignition (w/w%)	NO	Sample	Loss of Ignition (w/w%)
1	S37	1.03	35	S4	4.40	68	S5	1.03
2	S6	1.52	36	S25	4.45	69	S23	2.86
3	S93	2.60	37	S24	4.45	70	S84	3.00
4	S91	2.63	38	S77	4.45	71	S98	3.03
5	S87	2.83	39	S26	4.53	72	S88	3.11
6	S106	2.84	40	S114	4.56	73	S13	3.17
7	S99	2.86	41	S78	5.08	74	S113	3.42
8	S109	2.90	42	S29	5.10	75	S31	3.86
9	S97	2.90	43	S22	5.16	76	S27	4.57
10	S110	2.95	44	S75	5.48	77	S70	6.44
11	S89	3.00	45	S3	6.66	78	S72	6.77
12	S85	3.00	46	S33	6.69	79	S38	15.10
13	S100	3.01	47	S74	8.17	80	S49	21.93
14	S112	3.03	48	S73	8.99	81	S14	25.33
15	S111	3.07	49	S39	14.75	82	S34	34.57
16	S95	3.07	50	S40	17.36	83	S69	42.15
17	S43	3.11	51	S45	21.88	84	S20	42.72
18	S104	3.11	52	S76	22.30	85	S32	42.86
19	S105	3.12	53	S19	25.07			
20	S90	3.15	54	S48	26.95			
21	S108	3.15	55	S44	27.32			
22	S12	3.25	56	S47	30.88			
23	S42	3.32	57	S36	34.29			
24	S18	3.32	58	S35	34.94			
25	S102	3.41	59	S46	36.44			
26	S96	3.41	60	S2	41.91			
27	S28	3.48	61	S41	42.50			
28	S30	3.48	62	S15	42.50			
29	S86	3.55	63	S10	42.58			
30	S92	3.57	64	S21	42.63			
31	S9	3.64	65	S8	42.81			
32	S101	3.64	66	S1	43.19			
33	S80	3.88	67	S7	43.48			
34	S79	4.13						

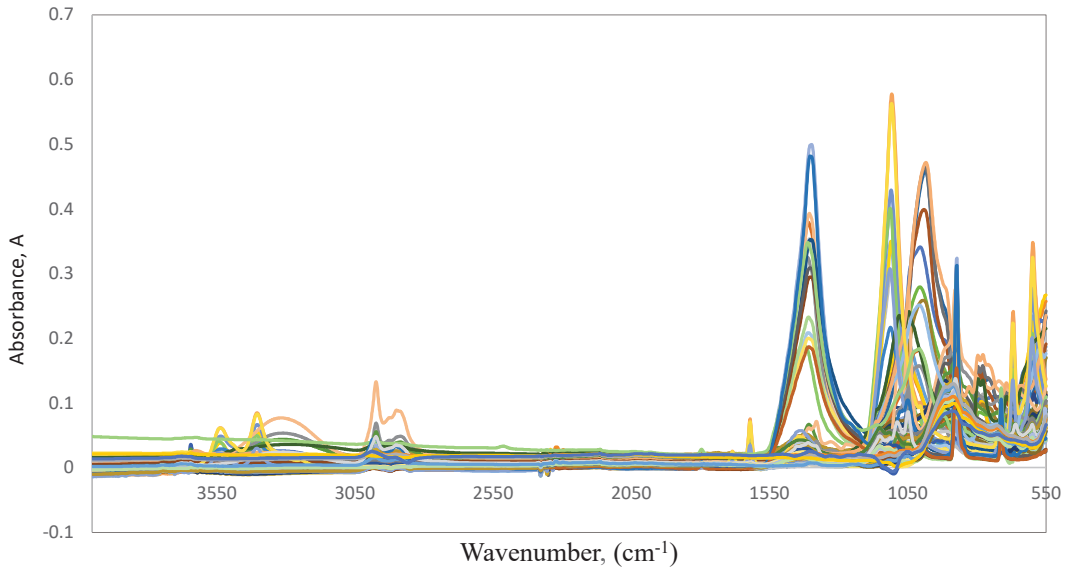


Figure 5.1. Mid-IR spectra of all samples

In Figure 5.2, the spectrum of one sample for each type was given in order to distinguish the types of samples included in the data set.

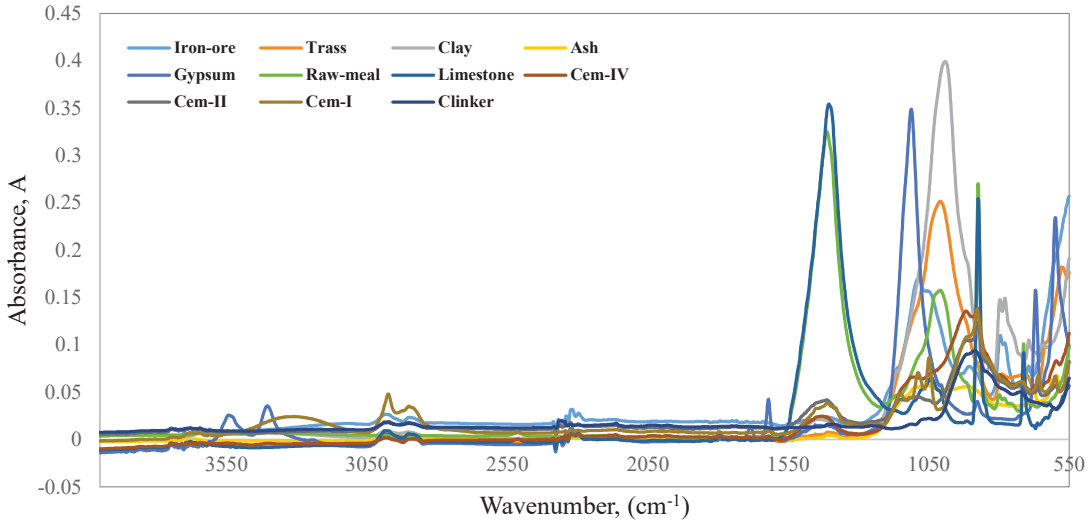


Figure 5.2. Mid-IR spectra of each type of sample included in the data set

Without any pre-processing technique and, also, with performing a suitable pre-processing technique, the full range of spectra was mean-centered and used to construct the PLS models. The Leave-one-out cross-validation pattern was used to determine the optimal number of latent variables (principal components) to prevent the overfitting problem.

5.1.1. The First Scenario Without pre-processing

5.1.1.1. The PLS results of the CaO component

The predicted residuals errors sum of squares (PRESS) values were determined using cross-validation with mean-centering for the first 30 LVs to find the optimal number of Latent variables, and the results are shown in Figure 5.3.

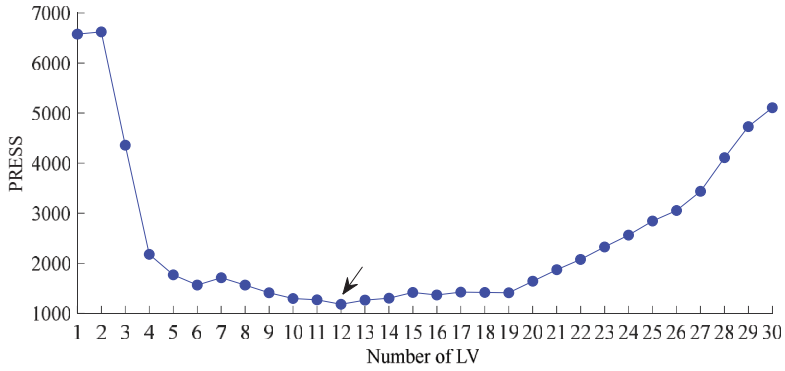
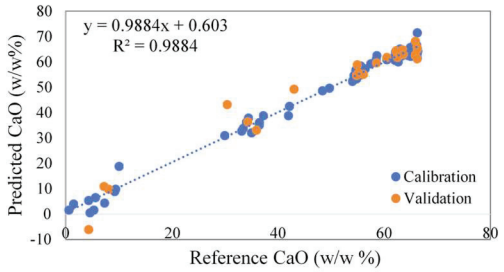


Figure 5.3. Number of PCs vs. PRESS plot to select the optimal number of LVs of CaO.

By considering Figure 5.3, 12 LVs were chosen after the mean centering of the data. Next, the model's performance was decided by plotting predicted CaO (w/w%) values against reference CaO (w/w%) values. Finally, Figure 5.4 represents the correlation graph of the calibration and validation sets.

a)



b)

CaO	
SECV (w/w%)	2.06
SEP (w/w%)	4.23
R^2	0.99
Max	66.38
Min	0.61
Range	65.77
Latent Variables	12

Figure 5.4. a) Reference CaO concentrations vs. Predicted CaO concentrations.
b) Statistical parameters of CaO.

As can be seen in Figure 5.4, the regression coefficient (R^2) of the model was found as 0.988, and the standard error of cross-validation (SECV) was found 2.065 (w/w%), and the standard error of prediction (SEP) was found as 4.234 (w/w%). Thanks to having

a high regression coefficient value, the Pls model has successfully predicted the concentration of CaO (w/w%) for the samples. Figure 5.5 shows the residuals graph, which can be used to determine the error range and potential residual trends.

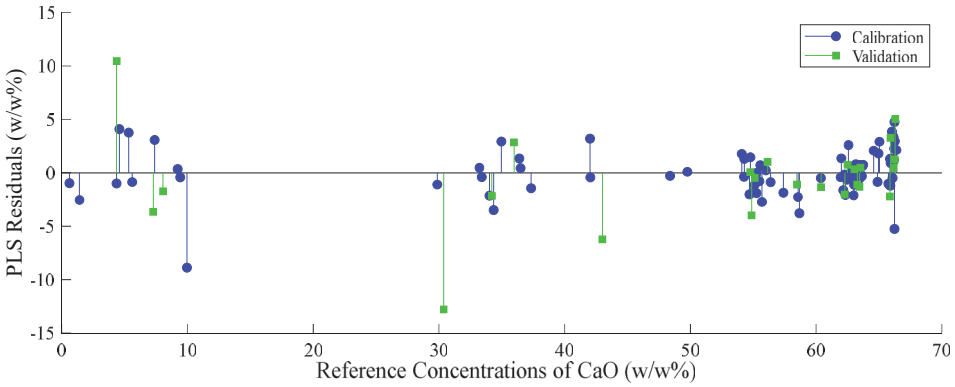


Figure 5.5. Reference CaO concentrations vs. Corresponding PLS prediction residuals of CaO

Although most residuals are in the ±5 (w/w%) range, there are exceptions, especially for some samples, as shown in Figure 4.7. When the calibration and validation data were examined, the s74 sample has the highest cross-validation absolute error in the calibration data set, and the S3-S5 samples have the highest absolute errors in the validation data set. The s74 sample has three times more cross-validation errors than other samples in the calibration dataset. The S3 and S5 samples have the close absolute errors and have the two highest errors compared to other samples in the validation set and were a vast difference between its absolute errors the other samples. First, S74 was removed from the data to improve the calibration model. Then S3 and S5 samples were also excluded from the validation dataset. After all outlier samples are removed, the PLS model was performed on the rest of the data by choosing 10 LVs after mean centering of the data and number of PCs vs. PRESS plot for selecting the optimal number of PCs was illustrated in Figure 5.6.

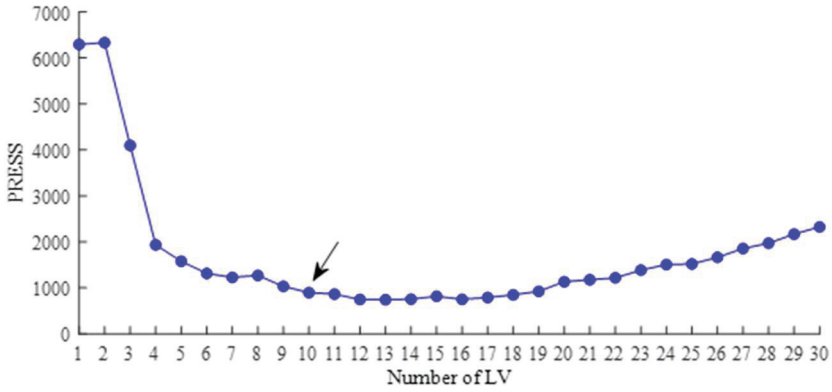
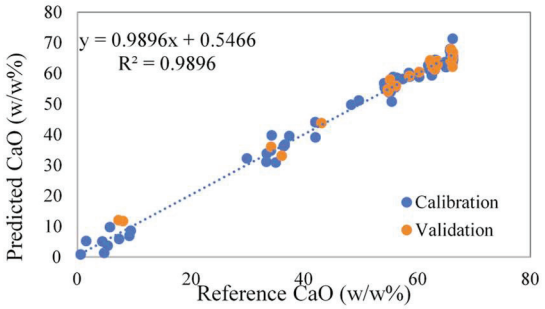


Figure 5.6. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of CaO

For the prediction of concentration of CaO (w/w%), the PLS model was constructed using the first ten principal components (PCs). To evaluate performance of model, a plot which is reference concentration of CaO (w/w%) values vs. predicted concentration of CaO (w/w%) values were obtained and was showed in Figure 5.7.

a)



b)

CaO	
SECV (w/w%)	1.91
SEP (w/w%)	2.09
R^2	0.99
Max	66.38
Min	0.61
Range	65.77
Latent Variables	10

Figure 5.7. a) Reference CaO concentrations (w/w%) vs. PLS predicted CaO concentrations (w/w%) b) Statistical parameters of CaO.

As seen in Figure 5.7, the model performance is considerably close for calibration and validation set predictions with showing no overfitting. Additionally, SECV and SEP values were found to be 1.912 (w/w %) and 2.086 (w/w %), respectively. The predicted power of value for calibration set predictions are calculated as 0.99, and the regression coefficient (R^2) value for the validation set is 0.989. For determining the error range and possible residual trends, the residuals plot is given in Figure 5.8.

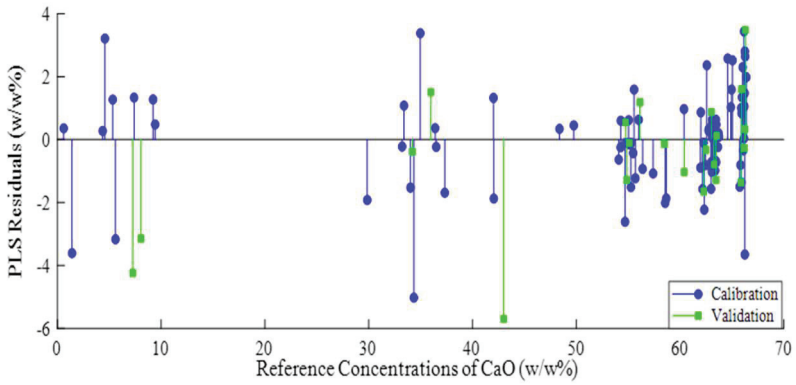


Figure 5.8. Reference concentrations vs. corresponding PLS prediction residuals of CaO.

With the elimination of the outlier samples from the data set, most of the residuals exist in the range of ±4 (w/w %). As compared to the first model, the final PLS model was produced using a lower number of principal components that ensure avoiding the model to noises and with lower SECV and SEP values, and with higher predictive power. The same strategy has been followed for outlier detection for other components.

5.1.1.2. The PLS results of the SiO_2 component

Prediction of SiO_2 component concentration of scenario one samples was obtained by applying the PLS method. The predicted residuals errors sum of squares (PRESS) values were determined using cross-validation with mean-centering for the first 30 LVs to decide the optimal number of Latent variables, and the results are presented in Figure 5.9.

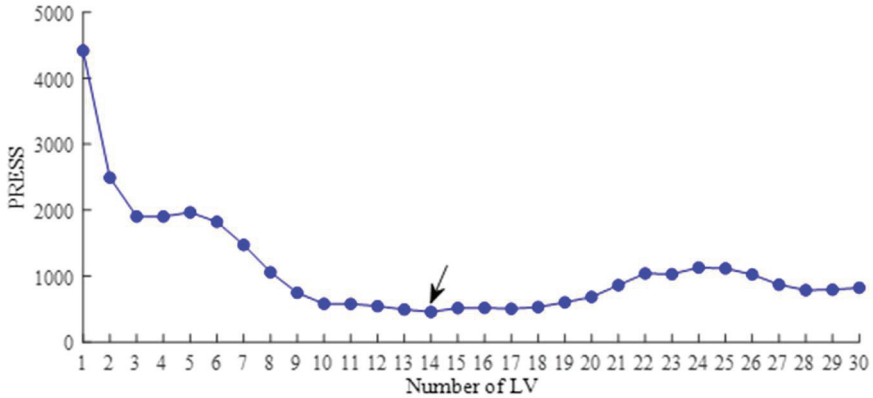
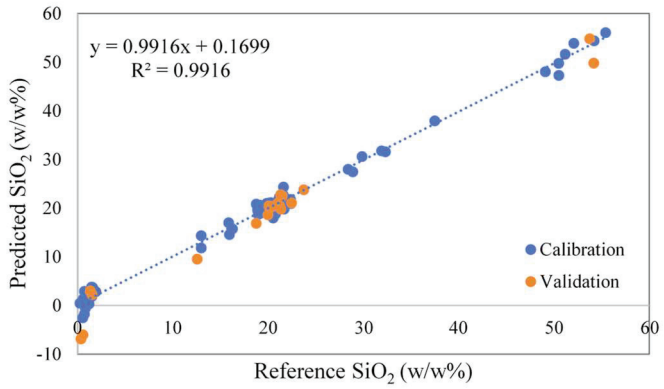


Figure 5.9. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of SiO_2.

By considering Figure 5.9, 14 LVs were chosen after mean centering of the data. The model's performance was decided by plotting a graph that predicted SiO_2 (w/w%) values vs. reference SiO_2 (w/w%) values. Figure 5.10 was then reached to represent the performance of the model.

a)



b)

SiO_2	
SECV (w/w%)	1.16
SEP (w/w%)	2.59
R^2	0.99
Max	55.42
Min	0.28
Range	55.14
Latent Variables	14

Figure 5.10. a) Reference SiO_2 concentrations (w/w%) vs. PLS predicted SiO_2 concentrations (w/w%). b)Statistical parameters of SiO_2.

The same strategy has been followed for outlier detection for the SiO_2 component. Samples that have names as S46, S33 was removed from the calibration data set and, S5, S3 were removed from the validation data set. Those samples were decided as outlier samples because they had high cross-validation absolute errors for calibration set samples and high absolute errors for validation set samples. After the outlier elimination, the regression coefficient of the final calibration model (R^2) was found as 0.99, and the predictive power of the validation data set (R^2) was found as 0.97. The performance of the model is remarkably close for both calibration and validation set predictions. Also, the Standard error cross-validation (SECV) and the standard error prediction (SEP) values were calculated as 1.161 and 2.594 (w/w%), respectively. For the detection of error range, the residual plot of reference concentrations vs. corresponding PLS predictions of SiO_2 the residual plot of reference concentrations vs. corresponding PLS predictions of SiO_2 was given in Figure 5.11.

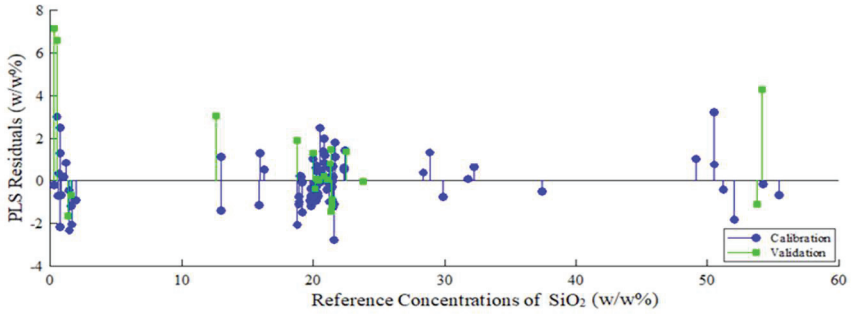


Figure 5.11. Reference concentrations vs. corresponding PLS prediction residuals of SiO_2

As can be seen in Figure 5.11 , most of the residuals are in the region of ±4 (w/w%), but the range of residuals for the validation set was expanded +7 for the positive side of the plot. Therefore, the SEP value was higher than the SECV value.

5.1.1.3. The PLS results of the Al_2O_3 component

For the concentration prediction of Al_2O_3 component, 18 principal components were used while applying PLS calibration method. The responsible figure that number of principal component vs. PRESS was showed in Figure 5.12.

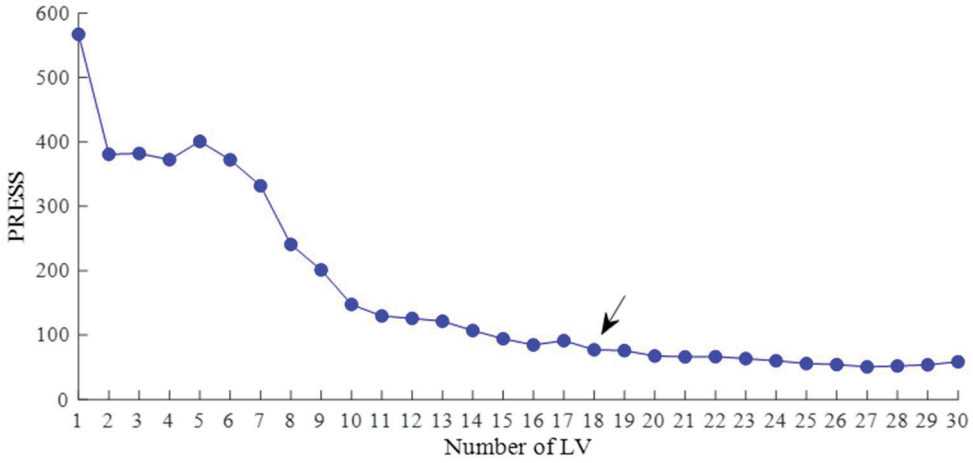
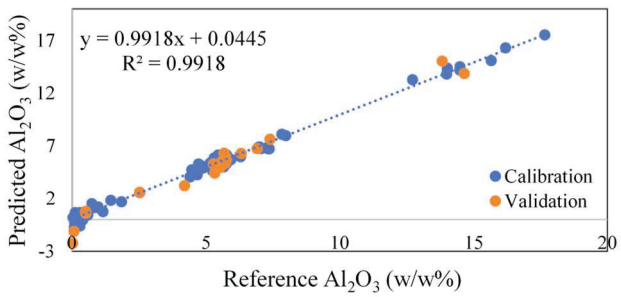


Figure 5.12. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of Al_2O_3

The reference Al_2O_3 concentrations (w/w%) against predicted Al_2O_3 concentrations (w/w%) gained by the PLS model were plotted to obtain the model's

performance for the Al_2O_3 concentration prediction. The corresponding figure was given in Figure 5.13.

a)



b)

Al_2O_3	
SECV (w/w%)	0.34
SEP (w/w%)	0.73
R^2	0.99
Max	17.66
Min	0.02
Range	17.64
Latent Variables	18

Figure 5.13. a) Reference Al_2O_3 concentrations vs. Predicted Al_2O_3 concentrations by PLS. b) Statistical parameters of Al_2O_3

In Figure 5.13, the regression coefficient, SECV, SEP values, and concentration range of the Al_2O_3 were given to indicate the performance of the PLS model. The regression coefficient values were found as 0.99, and 0.92 for calibration and validation sets, respectively. For the concentration prediction of Al_2O_3 components, high regression coefficients were obtained. The standard error of cross-validation was calculated as 0.339 (w/w%), and the standard error of prediction was calculated as 0.728 (w/w%). Before constructing the last model, outlier elimination was made by following the same strategy as modeled previous components. S46 and S5 samples are removed from the data as outliers. Those were gypsum and ash samples. For the determination of the range of errors and possible residual trends, the residuals graph is shown in Figure 5.14.

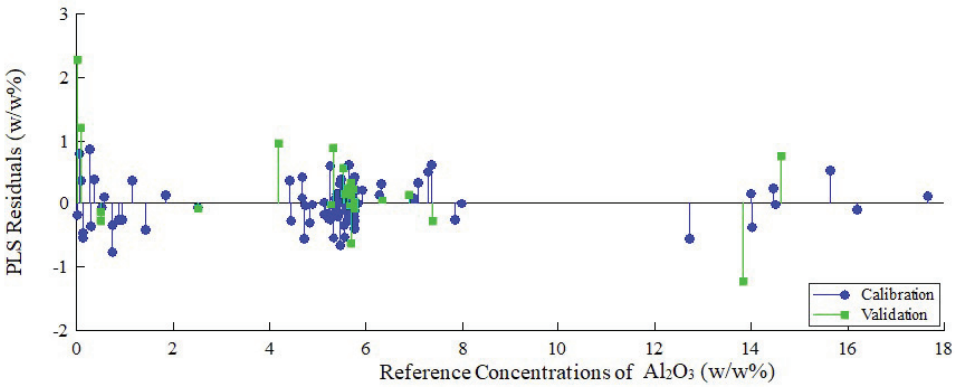


Figure 5.14. Reference concentrations vs. corresponding PLS prediction residuals of Al_2O_3

The residuals were found in ±1 region for the calibration set and were found in the -1.5,+2 region for the validation set.

5.1.1.4. The PLS results of the Fe_2O_3 component

Before starting construction of a calibration model, S4, S43, S33, S42, S3 samples were removed from the data set because those samples are iron ore samples and have significantly higher concentrations of Fe_2O_3 than the rest of the data. PRESS values were calculated using CV predictions with mean centering for the first 30 LVs to decide the optimal number of LVs, and the results are presented in Figure 5.15.

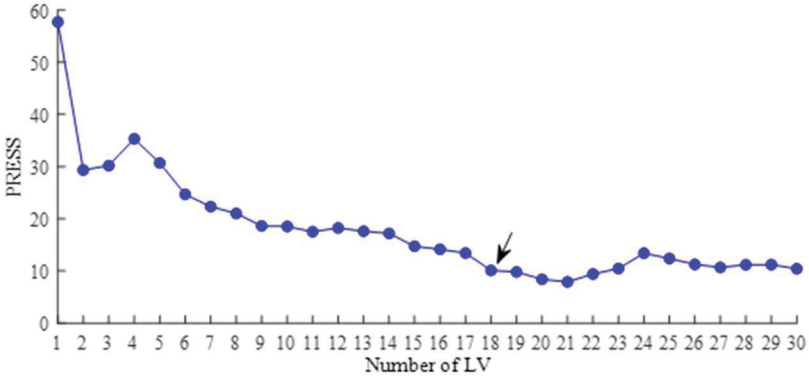
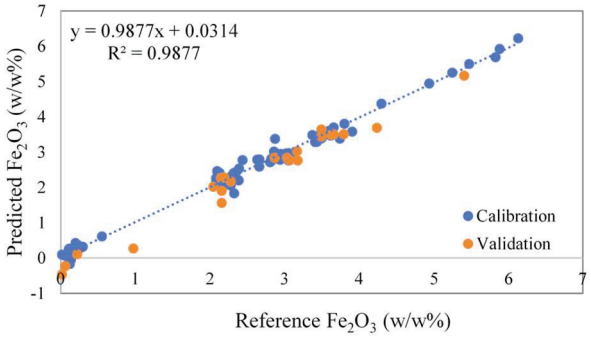


Figure 5.15. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of Fe_2O_3

Using the first eighteen principal components, the PLS model was constructed for the concentration prediction of the Fe_2O_3 component. The model performance was found by plotting the reference concentration of Fe_2O_3 (w/w%) against the predicted concentrations of Fe_2O_3 (w/w%). The results were shown in Figure 5.16.

a)



b)

Fe_2O_3	
SECV (w/w%)	0.16
SEP (w/w%)	0.31
R^2	0.99
Max	6.13
Min	0.02
Range	6.11
Latent Variables	18

Figure 5.16. a) Reference Fe_2O_3 concentrations vs. Predicted Fe_2O_3 concentrations by PLS. b) Statistical parameters of Fe_2O_3

The regression coefficients of the model for Fe_2O_3 concentration prediction were found as 0.988 that belongs to the calibration set, and 0.987 which belongs to the validation set. These results show that the model performance to predict Fe_2O_3 concentration is excellent because calibration and validation set regression coefficients are almost equal. In conclusion, The PLS model results are reliable to predict the concentration of Fe_2O_3. SECV and SEP values were calculated as 0.159 and 0.314 (w/w%), respectively. SECV value is higher than the SEP value that shows the model does not have an overfitting problem. Although the dynamic range of samples is narrow, the Fe_2O_3 model showed high performance with high regression coefficients. For the determination of error range along , the residuals plot was given in Figure 5.17.

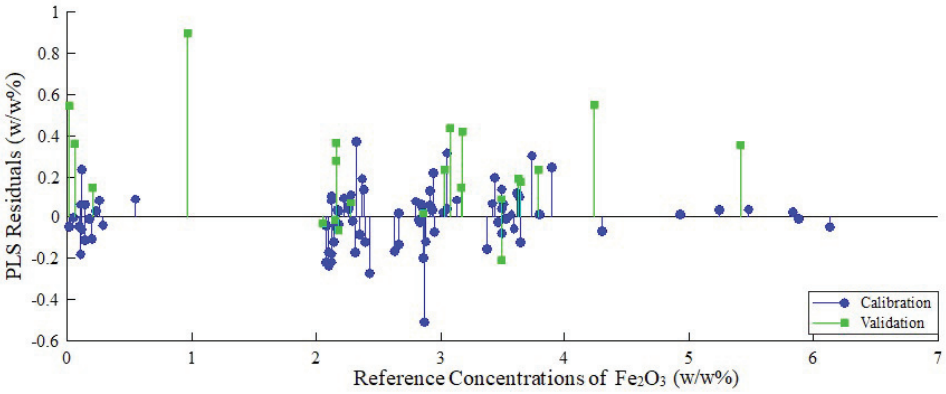


Figure 5.17. Reference concentrations vs. corresponding PLS prediction residuals of Fe_2O_3

Figure 5.17. shows the residual trends, and most of the residuals are in the region between ±0.6 (w/w %).

5.1.1.5. The PLS results of the MgO component

The MgO model was constructed using the first seventeen principal components that were decided as the optimal number of principal components. The number of latent variables vs. PRESS plot was given in Figure 5.18.

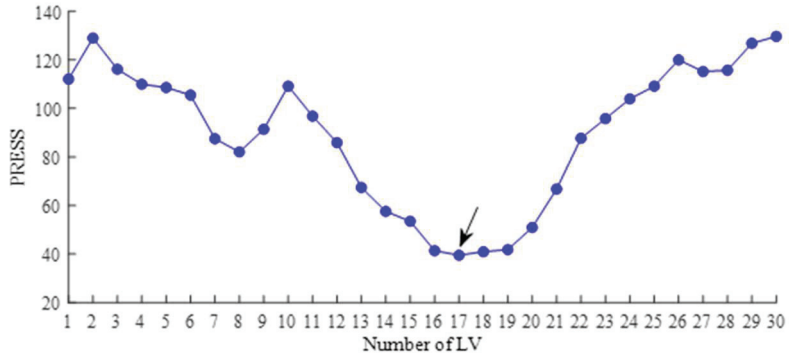
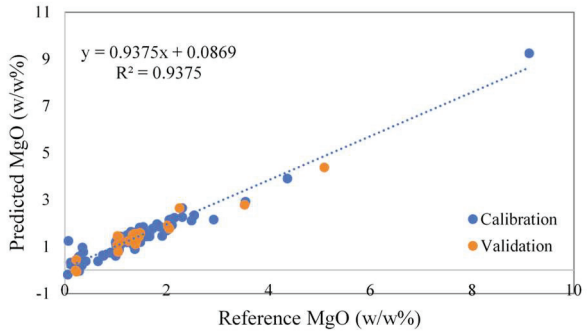


Figure 5.18. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of MgO

The reference MgO concentrations (w/w%) against predicted MgO concentrations (w/w%) obtained by the PLS model were plotted to obtain the model's performance for the MgO concentration prediction. The corresponding figure was given in Figure 5.19.

a)



b)

MgO	
SECV (w/w%)	0.28
SEP(w/w%)	0.31
R^2	0.94
Max	9.13
Min	0.05
Range	9.08
Latent Variables	17

Figure 5.19. a) Reference MgO concentrations vs. Predicted MgO concentrations by PLS. b) Statistical parameters of MgO.

The regression coefficients of model were found as 0.937 and 0.922 for calibration, validation sets, respectively. SECV and SEP values were calculated as 0.279, 0.310 (w/w%), respectively. The results are close each other, the performance of the model is succeeded to predict concentration of MgO for each sample. Before the constructing final model, the outlier samples were evaluated for MgO component according to their results of absolute errors. Only S35 sample was defined as outlier and removed the data set. For the determination of the range of errors and possible residual trends of constructed final PLS models, the residuals graph is shown in Figure 5.20.

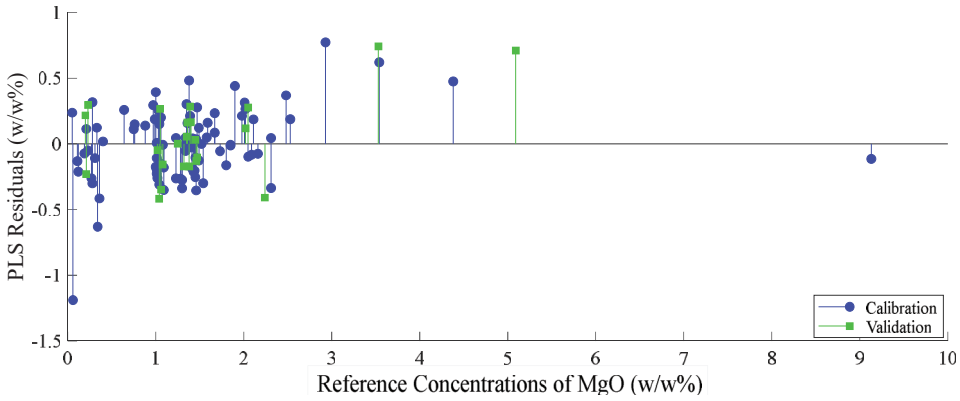


Figure 5.20. Reference concentrations vs. corresponding PLS prediction residuals of MgO (w/w%)

Figure 5.20, shows the residual trends, and most of the residuals are in the region between ±0.5 (w/w %).

5.1.1.6. The PLS results of the SO_3 component

For the determination of SO_3 concentrations in studied samples, the PLS model was constructed using the first sixteen principal components given in Figure 5.21 obtained to find the optimal number of principal components.

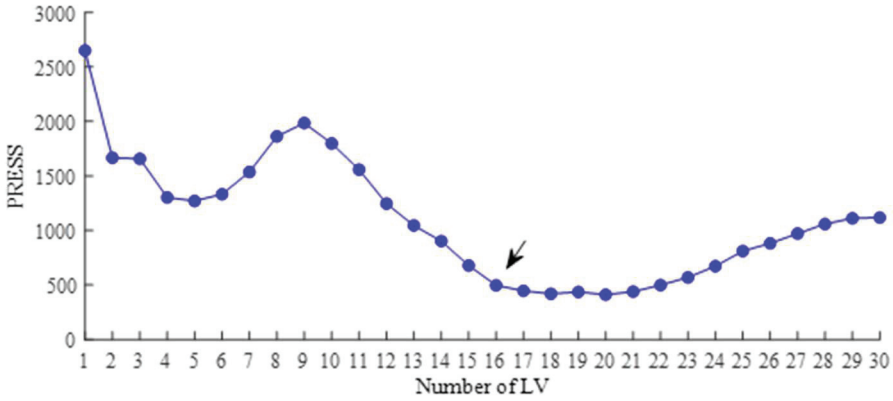
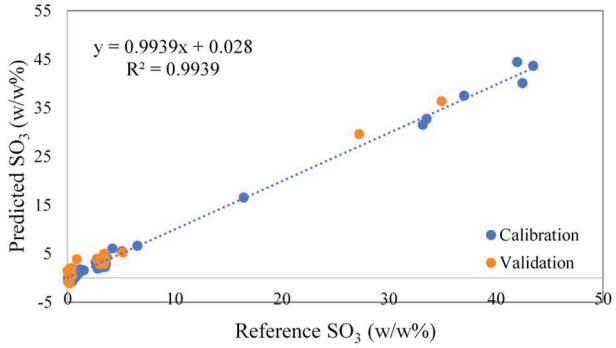


Figure 5.21. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of SO_3

The model performance was found by plotting the reference concentration of SO_3 (w/w%) against the predicted concentrations of SO_3 (w/w%). The results were shown in Figure 5.22.

a)



b)

SO_3	
SECV (w/w%)	0.75
SEP (w/w%)	1.20
R^2	0.99
Max	43.48
Min	0.05
Range	43.43
Latent Variables	16

Figure 5.22. a) Reference SO_3 concentrations vs. Predicted SO_3 concentrations by PLS. b) Statistical parameters of SO_3.

The regression coefficients of the model were found as 0.994 and 0.988 for calibration and validation sets, respectively. SECV and SEP values were calculated as 0.74 and 1.205 (w/w%), respectively. The model performance is close enough for the calibration and validation set predictions, with the calibration performance only slightly higher, indicating that no overfitting can be observed. For the determination of the error range and possible residual trends, the residuals plot was given in Figure 5.23.

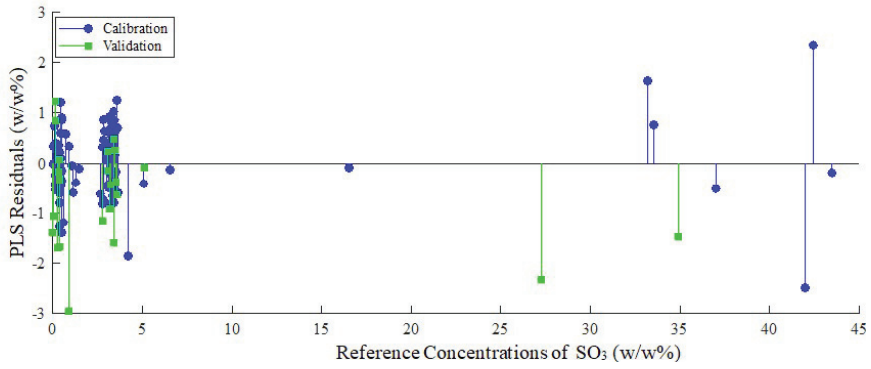


Figure 5.23. Reference concentrations vs. corresponding PLS prediction residuals of SO_3

As shown in Figure 5.23, most residuals are in the range between ±1 (w/w %). However, there are exceptions for some samples which lie on more extensive error ranges. The samples were checked and realized that all samples with high error values are gypsum samples that comprise $CaSO_4.0.5\ H_2O$. The constructed model can trouble in predicting concentrations of SO_3 of gypsum samples.

5.1.1.7. The PLS results of the Na_2O component

The PLS model of Na_2O was formed using the first ten principal components after mean centering of the data. PRESS values were calculated using cross-validation predictions for 30 LVs. Results were presented in Figure 5.24.

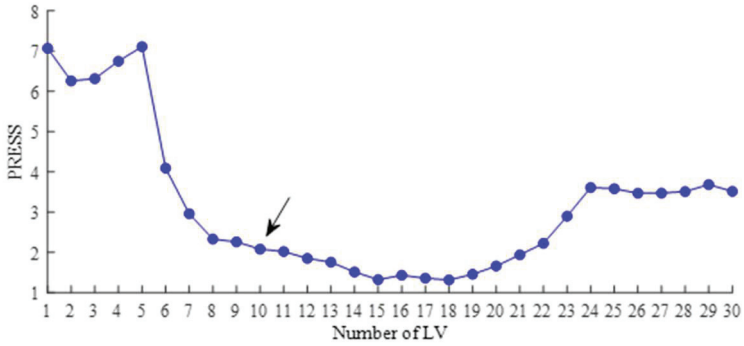
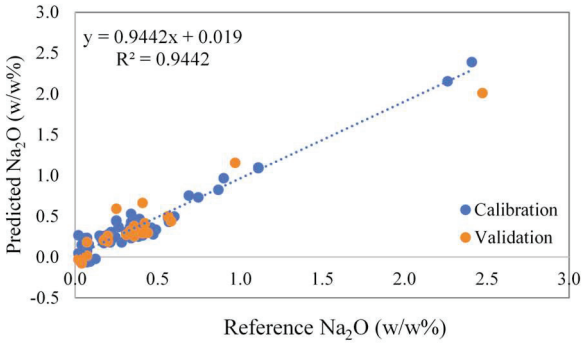


Figure 5.24. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of Na_2O

The model performance evaluation was done by plotting the reference concentrations of Na2O (w/w%) against predicted concentrations of Na2O (w/w%) obtained by the constructed PLS model. The corresponding figure is illustrated in Figure 5.25.

a)



b)

Na_2O	
SECV (w/w%)	0.09
SEP (w/w%)	0.15
R^2	0.94
Max	2.41
Min	0.02
Range	2.39
Latent Variables	10

Figure 5.25 a) Reference Na_2Oconcentrations vs. Predicted Na_2O concentrations by PLS. b) Statistical parameters of Na_2O

The regression coefficients of the model were found as 0.944, 0.90 for calibration and validation set, respectively. SECV and SEP values were calculated as 0.089 and 0.154 (w/w%), respectively. Any sample was not removed from the data as an outlier. The Na_2O is a minor oxide that includes cement composition; therefore, the range of the Na_2O concentration is narrow. Although the concentration range of Na_2O was narrow, the

constructed PLS model successfully predicted the concentration Na_2O of the sample thanks to having high regression coefficients. For the determination of the error range and possible residual trends, the residuals plot was given in Figure 5.26.

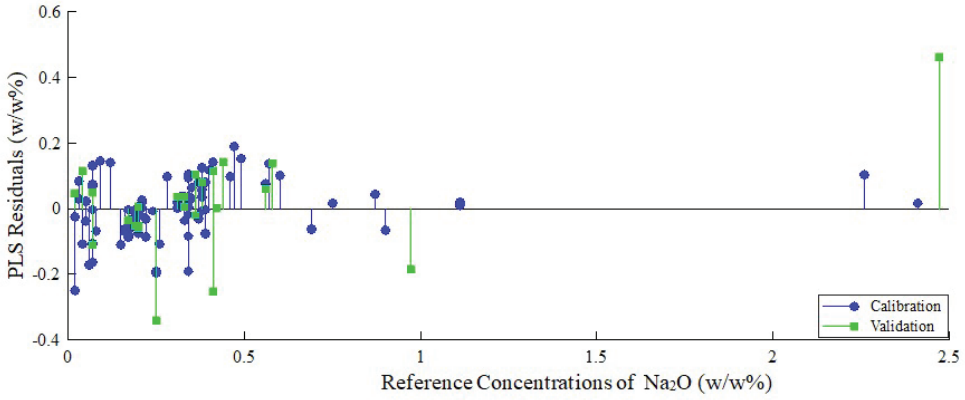


Figure 5.26. Reference concentrations vs. corresponding PLS prediction residuals of Na_2O.

Figure 5.26. shows the residual trends, and most of the residuals are in the region between ±0.2 (w/w %). The reference concentration range is quite narrow, making it more difficult to predict, and more samples can improve the ability to predict Na_2O concentration.

5.1.1.8. The PLS results of the K_2O component

For deciding optimal LVs, PRESS values were calculated using cross-validation predictions with mean-centering for the first 30 LVs, and the results are given in Figure 5.27. Using the first fourteen principal components, the PLS model was constructed to predict concentration K_2O in samples.

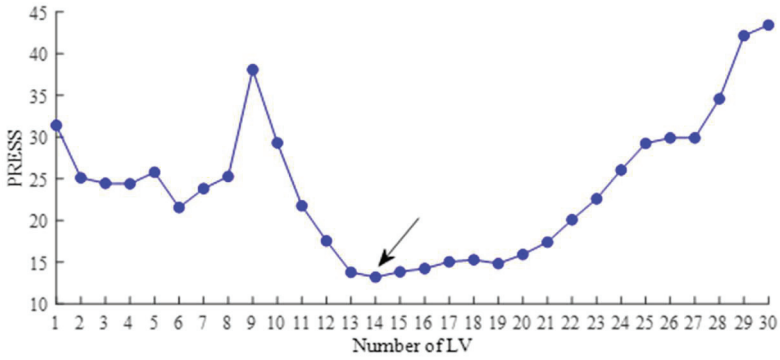
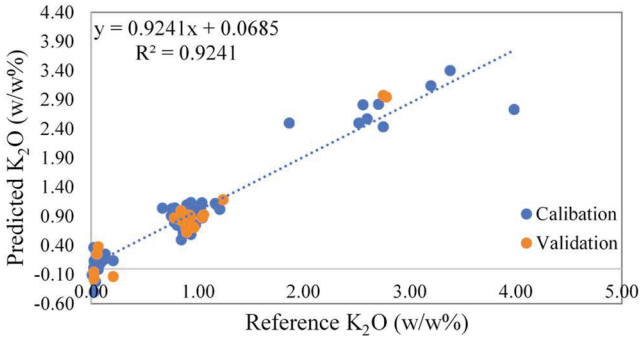


Figure 5.27. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of K_2O

The model performance evaluation was done by plotting the reference concentrations of K2O (w/w%) against predicted concentrations of K2O (w/w%) obtained by the constructed PLS model. The corresponding figure is illustrated in Figure 5.28.

a)



b)

K_2O	
SECV (w/w%)	0.21
SEP (w/w%)	0.17
R^2	0.92
Max	3.98
Min	0.01
Range	3.97
Latent Variables	14

Figure 5.28. a) Reference K_2Oconcentrations vs. Predicted K_2O concentrations by PLS. b) Statistical parameters of K_2O.

The K2O concentration range of samples is narrow because K2O is also minor oxide that includes in the cement composition. The performance of the model was found 0.924 for the calibration set and 0.9525 for the validation set. The regression coefficient of the validation model is higher than the calibration model. The SECV and SEP values were calculated as 0.21 and 0.169 (w/w%), respectively. Therefore, SEC value of the validation model is higher than the SEP value for the calibration set. This result shows a slight overfitting problem in constructed PLS model for the estimation of K2O concentration. For the determination of the error range and possible residual trends, the residuals plot was given in Figure 5.29.

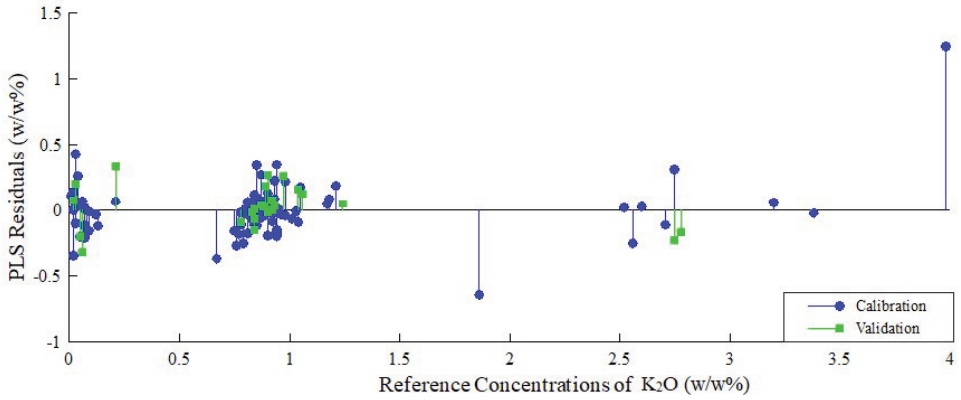


Figure 5.29. Reference concentrations vs. corresponding PLS prediction residuals of K_2O

As shown in Figure 5.29, most residuals are in the region between ±0.5 (w/w %). One sample was out of the region between ±0.5 (w/w%). As this sample is examined, the S38 was found to as one of the raw-meal samples with the highest concentration of K2O. By adding the raw-meal samples with the concentrations close as the s38 sample in the data set, the predicting ability of the model can be improved.

5.1.1.9. The PLS results of L.O.I

The PLS model was constructed to estimate the amount of loss of ignition of 85 samples using the optimal number of principal components. PRESS values were calculated by using cross-validation predictions with mean centering of 30 LVs. Figure 5.30 represents the results of calculated PRESS values vs. the Number of Latent variables.

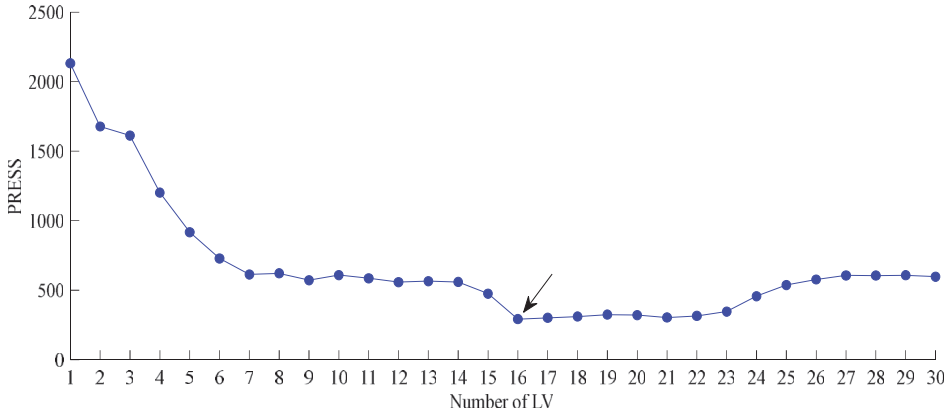
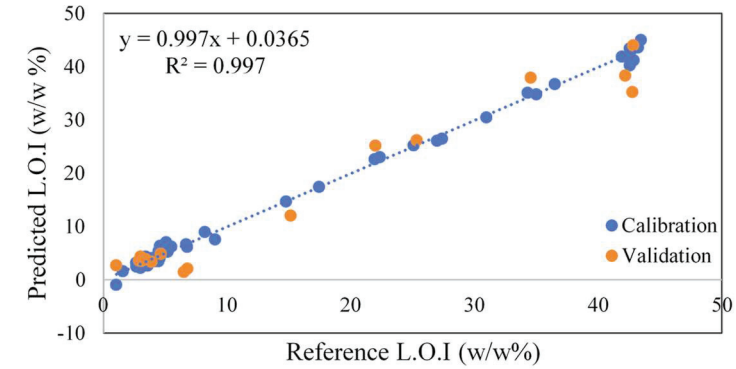


Figure 5.30. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of L.O.I

Using the first sixteen principal components, the Pls model was constructed to estimate the loss of ignition concentration of samples. Also, the constructed PLS model performance was evaluated by plotting the reference concentration of loss of ignition (w/w%) against the predicted concentration of loss of ignition. The PLS model results and the plot were given in Figure 5.31.

a)



b)

L.O.I	
SECV (w/w%)	0.79
SEP (w/w%)	3.03
R^2	1.00
Max	43.48
Min	1.03
Range	42.45
Latent variables	16

Figure 5.31. a) Reference K_2Oconcentrations vs. Predicted L.O.I concentrations by PLS b) Statistical parameters of L.O.I

The regression coefficients of the model were found as 0.99, 0.96 for calibration and validation, respectively. The regression coefficients of the two sets were quite close to each other. The performance of the calibration model is too good, and the performance of the formed validation set to test the calibration model was also quite good. The constructed PLS model was successful in estimating the concentration of loss of ignition of studied samples. The SECV and SEP values are calculated 0.79, 3.034 (w/w%), respectively. The residual plot for both calibration and validation set was given in Figure 5.32.

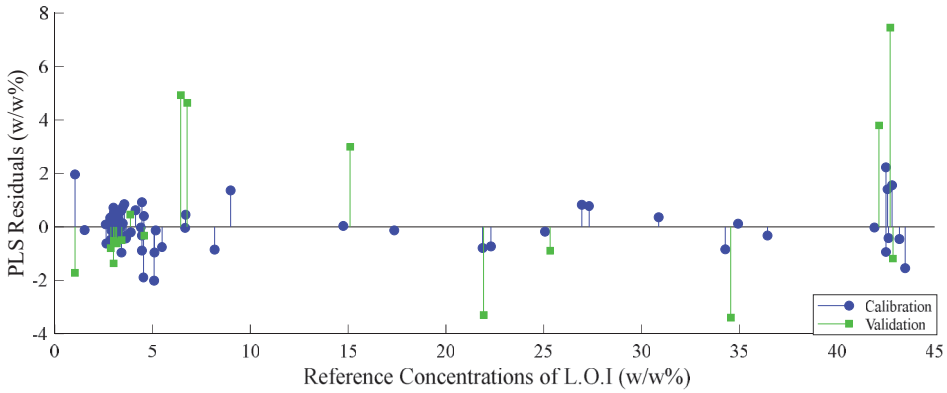


Figure 5.32. Reference concentrations vs. corresponding PLS prediction residuals of L.O.I

As shown in Figure 5.32, the residual plot of the validation set was distributed in a wide range between ±4 (w/w %), and the calibration set residuals are in the lower range between ±2 (w/w %), which shows that the SEP value of the validation set much higher than calibration set.

5.1.2. The First Scenario with Pre-Processing

Data pretreatment is carried out by using several pre-processing techniques for each FTIR-ATR spectrum of first scenario samples are given before. Several pre-processing techniques, such as Multiplicative Scatter Correction (MSC), Baseline correction, Extended Multiplicative Scatter Correction (EMSC), were applied to spectra of samples before the PLS models of each oxide and loss of ignition are computed. The best PLS models have been obtained with the Extended Multiplicative Scatter Correction (EMSC) pre-processing technique. Before extended multiplicative scatter correction is applied, the region between 2450-2350 cm^{-1} was excluded from the spectra because of showing absorption caused by atmospheric carbon dioxide. The calibration and validation sets are the same as the without-preprocessing part of the thesis. Outlier detection was evaluated for every specific component according to the high error values of the sample for each component.

Extended multiplicative signal correction (EMSC) is a preprocessing tool for distinguishing physical light scattering effects from vibrational effects in samples ranging from powders to turbid solutions. In spectroscopy, the spectra are often influenced by various phenomena other than chemical components of interest. Depending on the analysis type, the phenomena are changing, for example, from measurement noise to systematic error to interfering effects of undesired chemical and physical variations that cause undesired scattering variations in the sample of interest. This uncontrolled physical variation in samples could be due to the particle size and shape, sample preparation, sample surface. These issues could be minimized or eliminated if light scattering could be mathematically modeled and corrected in a reliable preprocessing point. EMSC preprocessing produces a spectrum by isolating the chemical absorption of light from additive, multiplicative and wavelength-dependent effects of uncontrolled variations in light scattering (Afseth & Kohler, 2012; Martens et al., 2003).

The Lambert-Beer law, a convenient starting point for deriving the basic EMSC model, is used to make quantitative measurements in vibrational absorbance spectroscopy. According to the Lambert- Beer law, an absorbance spectrum is directly proportional to the effective optical path length. The absorbance A(v) of transparent samples containing a single light-absorbing chemical species is calculated as follows:

$$A(v) = k(v).c.b \quad (5.1)$$

where *k(v)* denotes the characteristic absorptivity of a particular component at a given wavenumber *v*, *b* denotes the optical path length, and *c* denotes the concentration of light-absorbing chemical species in the sample. Spectral variations caused by changes in optical path length are generally referred to as "multiplicative" variations. Many different component spectra $k_J(v)$ (where *j* denotes the different component from *j=1...J*) have strongly overlapping characteristics since many different types of samples are very complex and encounter many different species *k(v)*. The Lambert-Beer law can thus be written as a superposition of absorbance of several species for *J* absorbing species:

$$A(v) = (\sum_{j=1}^{j} c_j.k_j(v)).b \quad (5.2)$$

where $k_j(v)$ is the spectrum, and c_j is the concentration for component j. The length of the optical path b is assumed to be comparable for all components Sample's chemical-constituent spectra are often not recognized also difficult to obtain. However, the overall shapes of infrared spectra from samples are similar. The average spectrum of a sample set is an excellent approach for each spectrum. A measured absorbance spectrum A(v) may thus be expressed as the mean $\bar{X}(v)$ of all spectra in given data set plus deviations $\Delta k_j(v)$ from this mean :

$$k_j(v) = \bar{X}(v) + \Delta k_j(v) \quad (5.3)$$

Inserting Equation (5.3) into Equation (5.2) obtained :

$$A(v) = \left(\sum_{J=1}^{J} C_J.k_j\,(v)\right).b = \left(\sum_{j=1}^{j} c_j\bar{X}(v) + \sum_{j=1}^{j} c_j\,\Delta k_j(v)\right).b \quad (5.4)$$

where the total of the concentrations is 1 :

$$\sum_{j=1}^{j} c_j = 1 \quad (5.5)$$

Then inserting Equation (5.5) into Equation (5.4), then obtained:

$$A(v) = \overline{(X}(v) + \sum_{j=1}^{j} c_j.\Delta k_j(v)).b \quad (5.6)$$

The statistical model can be used to replace Equation 5.6 in all situations where the overall form of the measured spectrum $A(v)$ is similar to the average spectrum $\bar{X}$, anf the satistical model represented as :

$$A(v) = \bar{X}(v).b + e(v) \tag{5.7}$$

Where $e(v)$ is the residual. The multiplicative signal correction (MSC) model is an additive effect provided by the constant baseline that extends the Lambert–Beer-type model shown in Eq. (5.7).The MSC model can be written as :

$$A(v) = a + \bar{X}(v).b + e(v) \tag{5.8}$$

Parameters that are not known are estimated by weighted least squares regression, then, a and b are evaluated, the spectra are corrected for:

$$A_{corr}(v) = \frac{A(v)-a}{b} \tag{5.9}$$

In some instances, the spectra contain baseline variations that cannot be expressed by a straight line. A baseline with an arbitrary slope, a quadratic term, or terms of higher polynomial order can be added to the MSC model in Eq. (5.8) to produce simple extensions of MSC, known as an extended multiplicative signal correction and represented as :

$$A(v) = a + \bar{X}(v).b + d_1 v + d_2 v^2 + \cdots + d_n v^n + e(v) \tag{5.10}$$

The least square solution of the unknown parameters is given by :

$$A_{corr}(v) = \frac{A(v)-a-d_1 v-d_2 v^2-\cdots-d_n v^n}{b} \tag{5.11}$$

Figure 5.33, Raw FT-IR spectra of the samples are given, and after the EMSC pre-processing applied spectra of the samples are given.

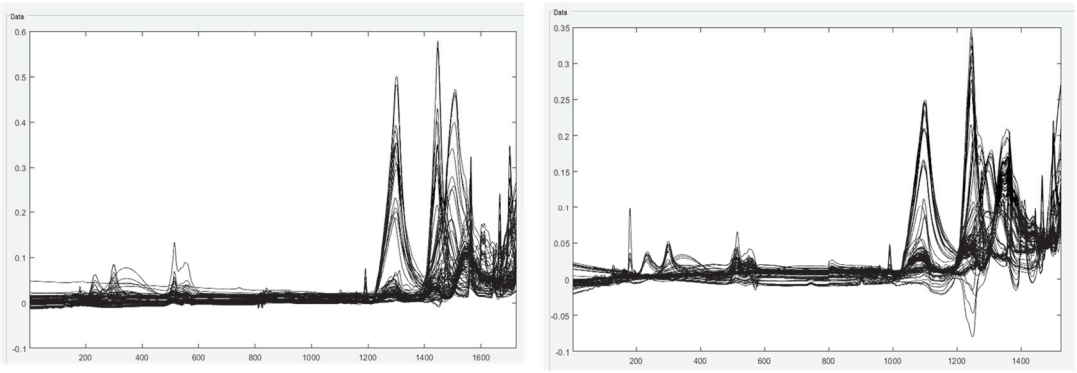


Figure 5.33. left: Raw FT-IR spectra of the samples, right : the EMSC spectra of the samples.

5.1.2.1. The PLS result of CaO component

To find the best-suited LV number, the predicted residual error sum of squares (PRESS) was calculated for the first 30 LVs, and the results are given in Figure 5.34.

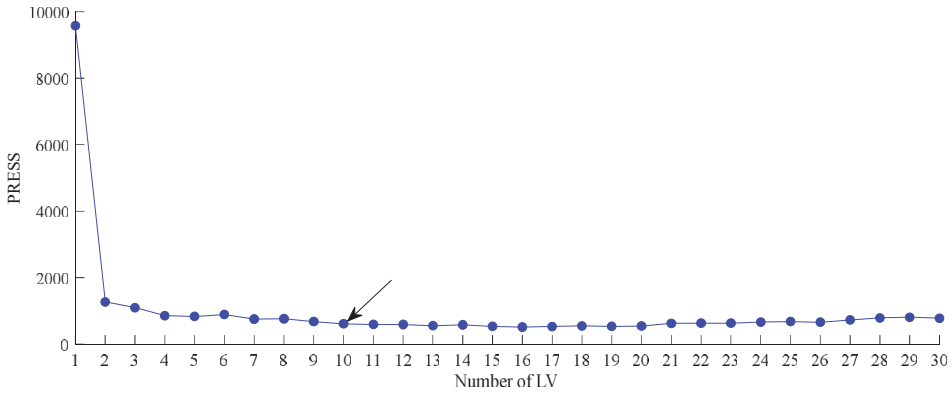
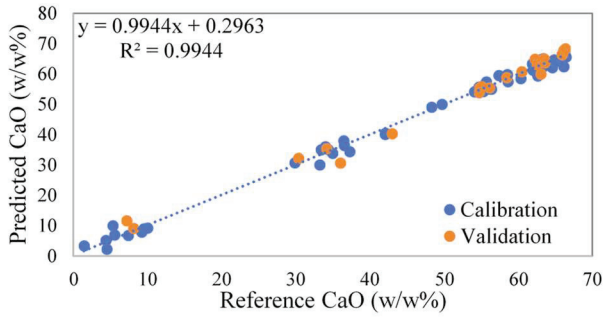


Figure 5.34. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of CaO.

Using the first ten principal components, the PLS model to predict the concentration of CaO (w/w%) was constructed. The model's performance was calculated by plotting the regression graph of two sets that involves reference concentration of CaO (w/w%) vs. predicted concentration of CaO (w/w%) by the PLS model. Furthermore, the related graphic was illustrated in Figure 5.35.

a)



b)

CaO	
SECV (w/w%)	1.39
SEP(w/w%)	2.07
R^2	0.99
Max	66.38
Min	1.41
Range	64.97
Latent Variables	10

Figure 5.35. a) Reference CaO concentrations vs. Predicted CaO concentrations by PLS. b) Statistical parameters of CaO

The regression coefficients of the model were found as 0.994, 0.987 for calibration and validation sets, respectively. The model performance of the calibration set, and the validation set is quite close each other which show that the PLS model predictive ability are too reliable. SECV and SEP values were calculated as 1.391, 2.075 (w/w%), respectively. The residual graph of the two set was given in Figure 5.36.

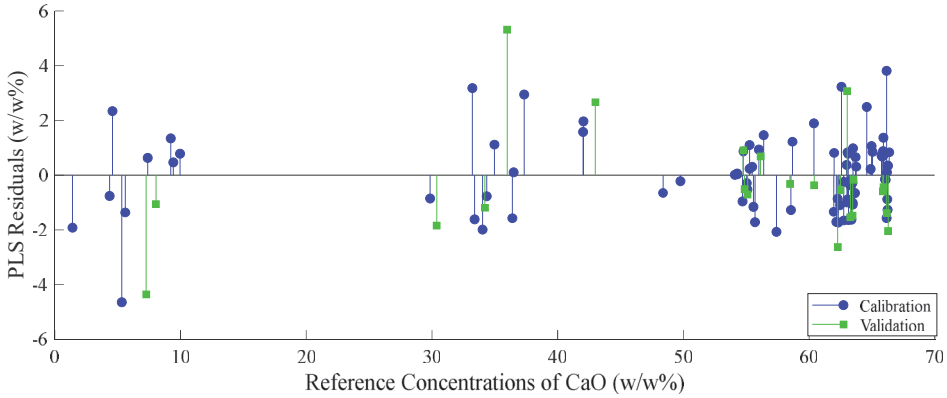


Figure 5.36. Reference concentrations vs. corresponding PLS prediction residuals of CaO.

As shown in Figure 5.36, most of the residuals are in the region between ±4 .

When The PLS model, which is created without the pre-processing technique, and with the EMSC applied PLS model were compared, the concentration of CaO (w/w%) in the studied samples was predicted more successfully for the EMSC applied PLS model. Lower SEC and SEP values and higher regression coefficients were obtained with the EMSC technique.

5.1.2.2. The PLS result of SiO_2 component

The predicted residual error sum of squares (PRESS) were calculated for 30LVS for detection optimal number of principal component that will use for the construction the PLS model. The PRESS values and number latent variables results are given Figure 5.37.

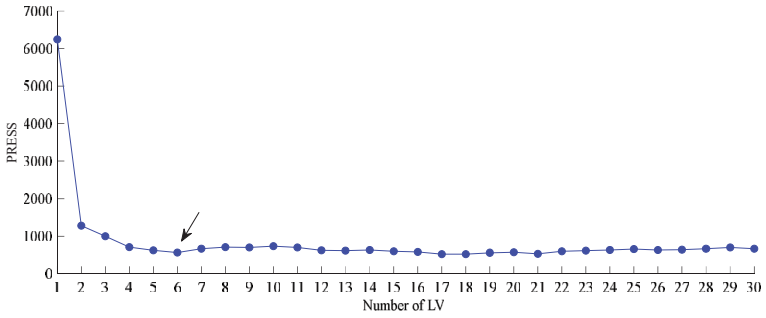
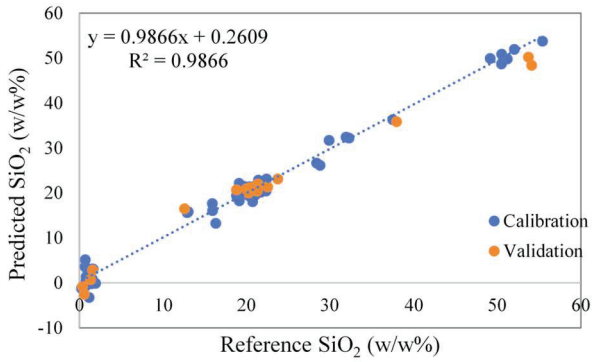


Figure 5.37. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of SiO_2.

The first six principal components were found as the optimal number principal components and with using these first six principal components the PLS model was obtained for prediction of SiO_2 concentration. The regression coefficient of the model was decided to detect the reliability of the model for the prediction stage. Therefore, correlation graph of the calibration and validation set was plotted by using reference concentrations of SiO_2 and the predicted concentrations of SiO_2 (w/w%). The graph was given in Figure 5.38.

a)



b)

SiO_2	
SECV (w/w%)	1.42
SEP (w/w)	1.97
R^2	0.99
Max	55.42
Min	0.28
Range	55.14
Latent Variables	6

Figure 5.38. a) Reference SiO_2 concentrations vs. Predicted SiO_2 concentrations by PLS. b) Statistical parameters of SiO_2.

The regression coefficients of the model were found as 0.986, 0.984 for calibration and validation sets, respectively. SECV and SEP values were calculated as 1.416 and 1.96 (w/w%), respectively. The PLS model's prediction of SiO_2 concentration was made successful thanks to having high and close regression coefficients of the two sets. The residual range of samples were given in Figure 5.39.

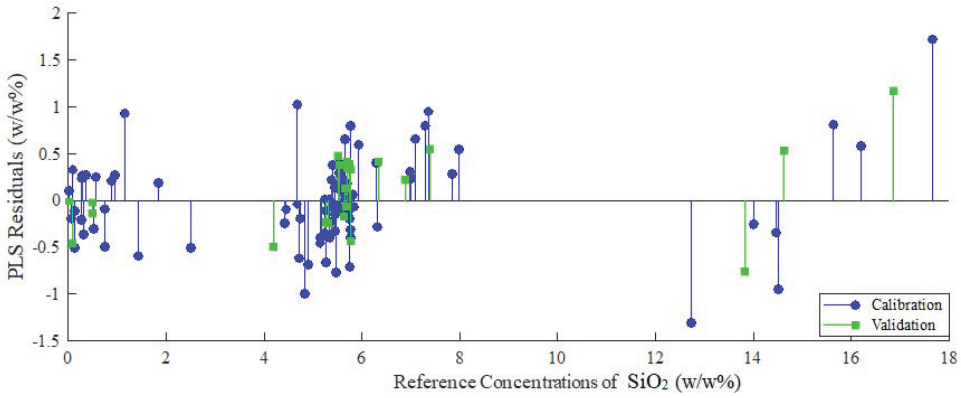


Figure 5.39. Reference concentrations vs. corresponding PLS prediction residuals of SiO2

As shown in Figure 5.39, most of the residuals are in the range of ±1.5. The calibration set residuals were found to be scattered in a little wide range than the validation set, which shows that the SECV value was found higher than the SEP value

5.1.2.3. The PLS result of Al_2O_3 componentt

To find the best-suited LV number, the predicted residual error sum of squares (PRESS) were calculated for the first 30 LVs, and the results are given in Figure 5.40.

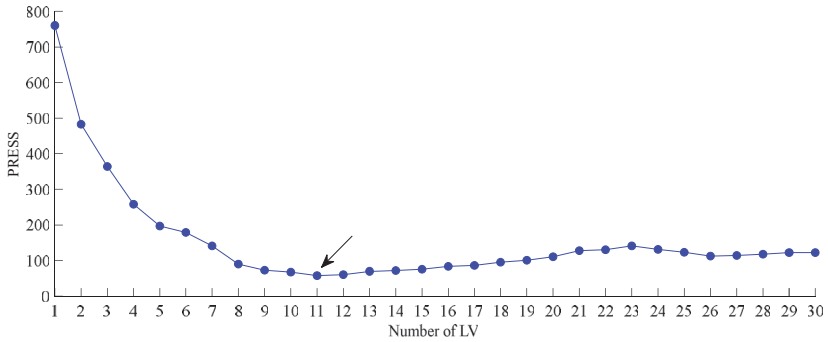
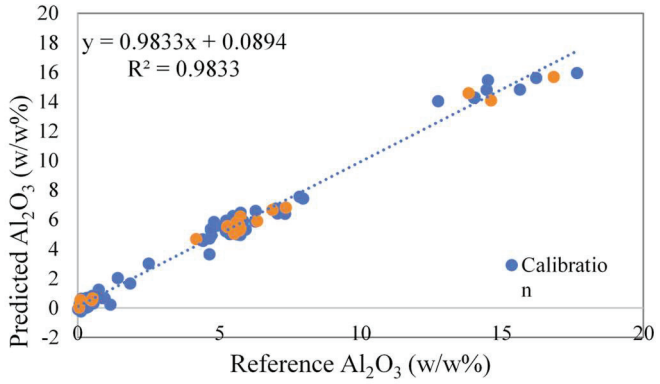


Figure 5.40. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of Al_2O_3

Using Figure 5.40, the PLS model was created with the first eleven principal components. Figure 5.41 is a regression graph obtained from the calibration and

validation's reference concentrations and predicted concentration of the Al_2O_3 (w/w%) to determine the performance of the PLS model.

a)



b)

Al_2O_3	
SECV(w/w%)	0.49
SEP(w/w%)	0.45
R^2	0.98
Max	17.66
Min	0.02
Range	17.64
Latent Variables	11

Figure 5.41. a) Reference Al_2O_3 concentrations vs. Predicted Al_2O_3 concentrations by PLS. b) Statistical parameters of Al_2O_3

The regression coefficients of the model were found as 0.98, 0.99 for calibration and validation set, respectively. SECV and SEP values were calculated as 0.486, 0.455, respectively. SEP value was higher than the SECV value, indicating a little overfitting problem in the PLS model, but not too much because the quantities of regression coefficients are close. Also, having high regression coefficients indicates that the model's predictability was quite satisfying to estimate Al_2O_3 (w/w%) concentration in the samples. The residual scatter graph of the two set was given in Figure 5.42.

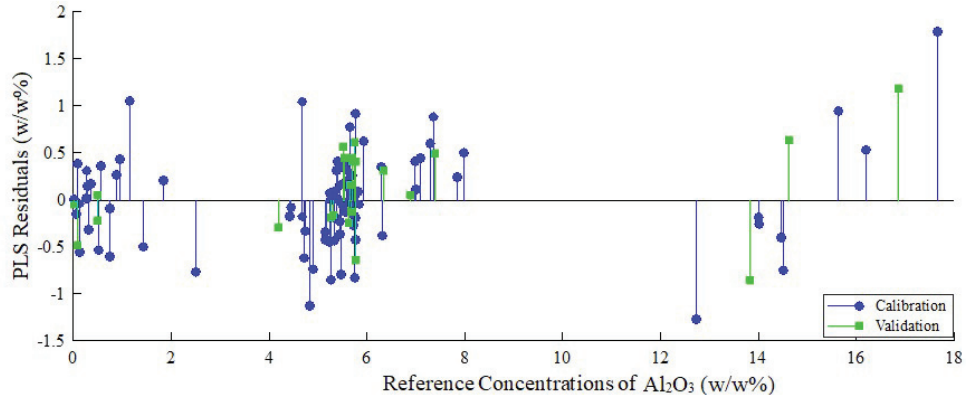


Figure 5.42. Reference concentrations vs. corresponding PLS prediction residuals of Al_2O_3

As can be seen in Figure 5.42, the residuals of the PLS model for the calibration set distributed in a little wider range than the residuals of the validation set. Therefore, the SECV value was found higher than the SEP value.

5.1.2.4. The PLS result of Fe_2O_3 component

For the determination of Fe_2O_3 concentration (w/w%), the PRESS values were calculated for the first 30 LVs, then the optimal number of the latent variables was decided. The results were shown in Figure 5.43.

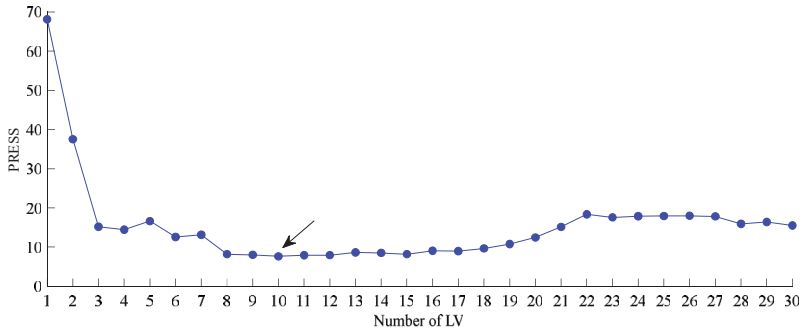
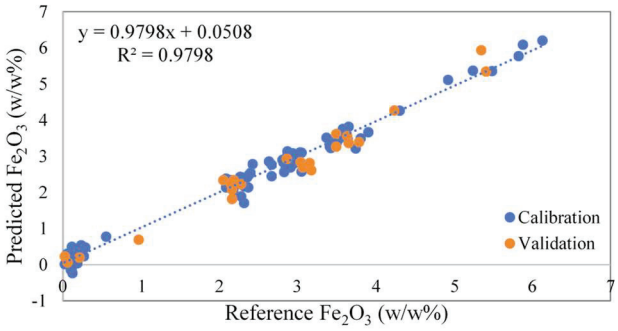


Figure 5.43. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of Fe_2O_3

Using the first ten latent variables, the PLS model was constructed, and Figure 5.44 was obtained to evaluate the model's performance.

a)



b)

Fe_2O_3	
SECV(w/w%)	0.21
SEP(w/w%)	0.27
R^2	0.98
Max	6.13
Min	0.02
Range	6.11
Latent Variables	10

Figure 5.44 . a) Reference Fe_2O_3 concentrations vs. Predicted Fe_2O_3 concentrations by PLS. b)Statistical parameters of Fe_2O_3

The regression coefficients of the model were found as 0.978, 0.968 for the calibration and validation set, respectively. SECV and SEP values were calculated as 0.205, 0.273 (w/w%), respectively. Although the Fe_2O_3 concentration range of the sample is narrow, the model's predictive ability was found high thanks to having high regression coefficients and low SEC and SEP values. Furthermore, the model was constructed with a relatively low number of the latent variable. The residual plot was given in Figure 5.45.

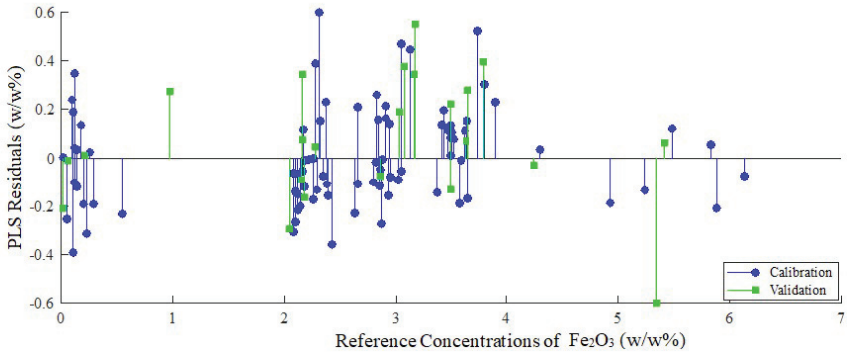


Figure 5.45. Reference concentrations vs. corresponding PLS prediction residuals of Fe_2O_3

As shown in Figure 5.45, most of the residuals are in the range of ±0.6, and the two sets are almost equally scattered.

5.1.2.5. The PLS result of MgO component

PRESS values were calculated by using cross-validation predictions for the first 30 LVs to find the optimal number of latent variables, and the results are shown in Figure 5.46.

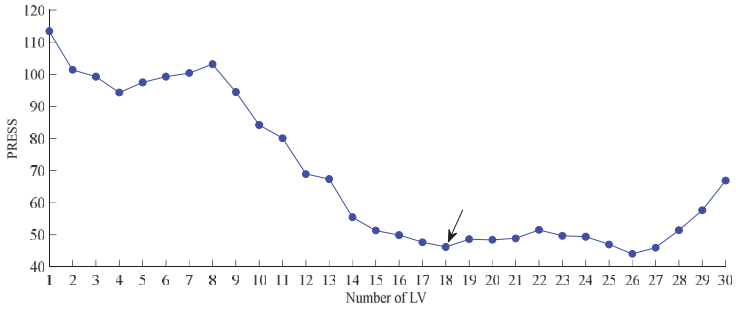
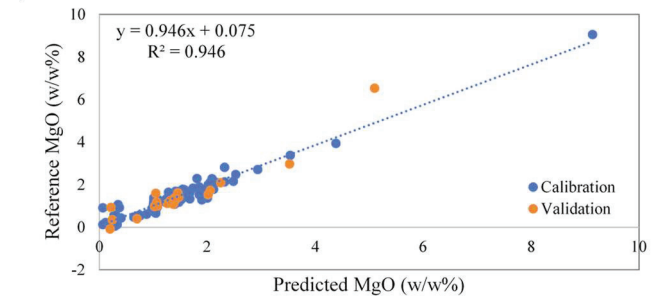


Figure 5.46. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of MgO

The PLS model was constructed with the first 18 LVs to estimate the MgO concentrations of samples (w/w%). The model performance was presented in Figure 5.47.

a)



b)

MgO	
SECV	0.26
SEP	0.42
R^2	0.95
Max	9.13
Min	0.05
Range	9.08
Latent Variables	18

Figure 5.47. a) Reference MgO concentrations vs. Predicted MgO concentrations by PLS. b) Statistical parameters of MgO

The regression coefficients of the model were found as 0.94, 0.89 for calibration and validation sets, respectively. The SECV and SEP values were calculated as 0.26 and 0.417, respectively. The regression coefficient of the validation set showed inferior performance for the prediction of MgO, which can be explained by the low variability of samples with high MgO content. Also, can be explained low content of the MgO component in the sample set. The residual graph of two set was given in Figure 5.48.

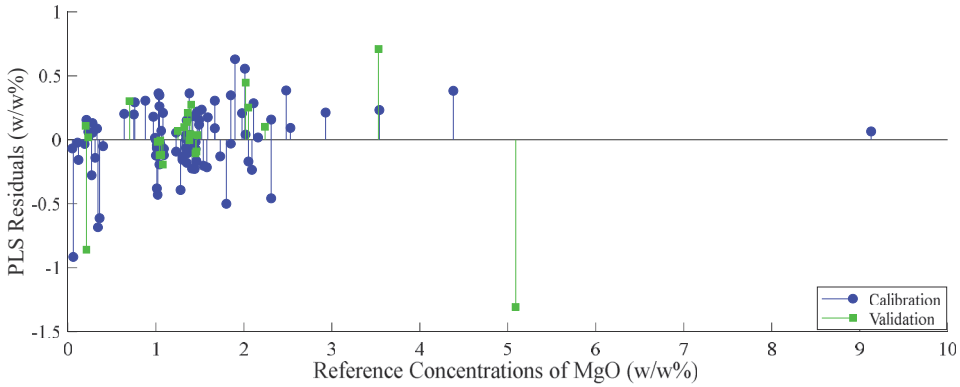


Figure 5.48. Reference concentrations vs. corresponding PLS prediction residuals of MgO

As can be seen in Figure 5.48, most residuals are in the range ±1, but there are a exception for a specific sample that is S47-gypsum sample. Among all gypsum samples, S47 has the highest MgO concentration. Although the PLS model is relatively successful, in the prediction stage, the estimation of MgO can be doubtful, especially for gypsum samples.

5.1.2.6. The PLS result of SO_3 component

PRESS values were calculated with using cross-validation prediction of the first 30 LVs to obtain the optimal principal component to create PLS model. The results a given in Figure 5.49.

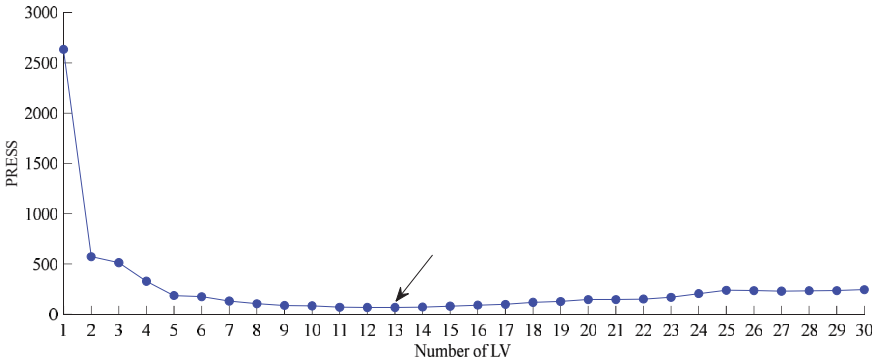
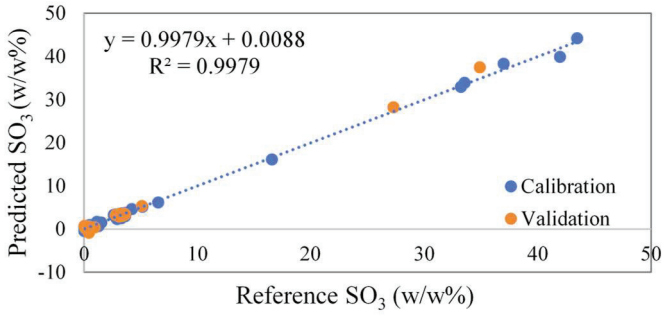


Figure 5.49. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of SO_3

The first 13 principal components were used to construct the PLS model for the estimation of SO_3 concentrations of the samples. The model performance evaluation was done by plotting the correlation graph of the calibration and validation sets. The correlation graph was given in Figure 5.50.

a) b)



SO_3	
SECV	0.400
SEP	0.689
R^2	0.998
Max	43.480
Min	0.050
Range	43.430
Latent Variables	13

Figure 5.50. a) Reference SO_3 concentrations vs. Predicted SO_3 concentrations by PLS. b) Statistical parameters of SO_3

The regression coefficients of the model were found as 0.9979, 0.9977 for calibration and validation sets, respectively. SECV and SEP values were calculated as 0.39, 0.68 (w/w%), respectively.The regression coefficients of the two set are very close that show the predicted abilility of the model for SO_3 concentration is quite high thanks

to having high regression coefficients. SECV values is lower than the SEP value that indicates that there is no overfitting problem in PLS model. The residual plot of the two set is given in Figure 5.51.

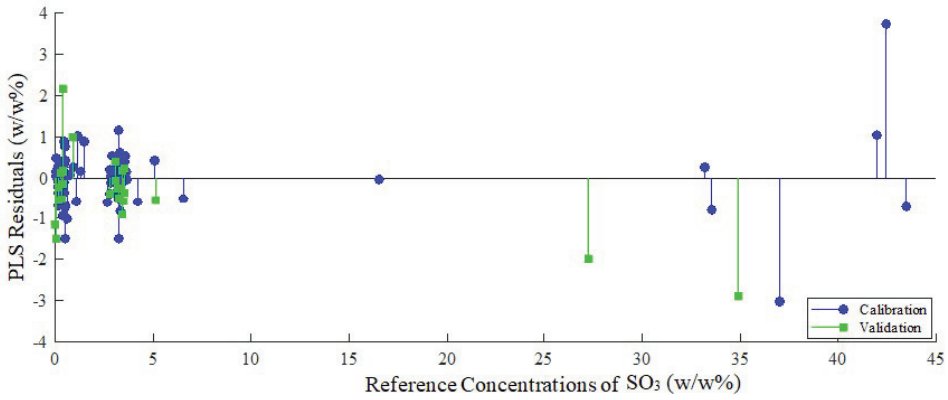


Figure 5.51. Reference concentrations vs. corresponding PLS prediction residuals of SO_3

As can be seen in Figure 5.51, most residuals are in the range of ±1. Some samples exceed the ±1 range. These samples are the gypsum samples with a much higher concentration of SO_3 than the rest of the data.

5.1.2.7. The PLS result of Na_2O component

PRESS values were calculated with using cross-validation prediction of the first 30 LVs to obtain the optimal principal component to create PLS model. The results are given in Figure 5.52.

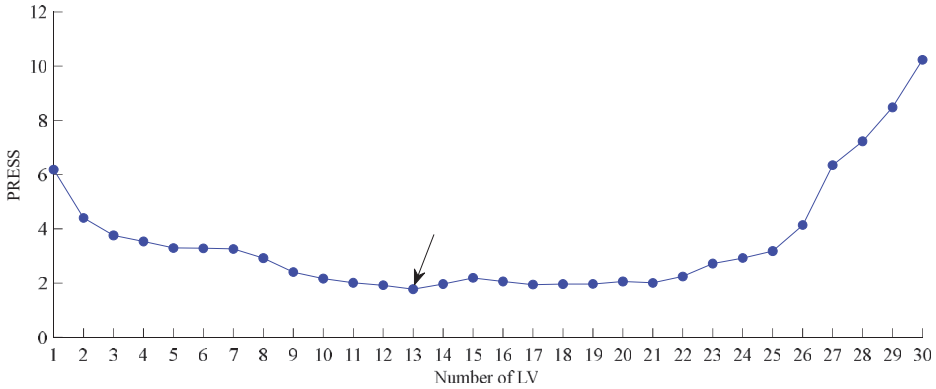
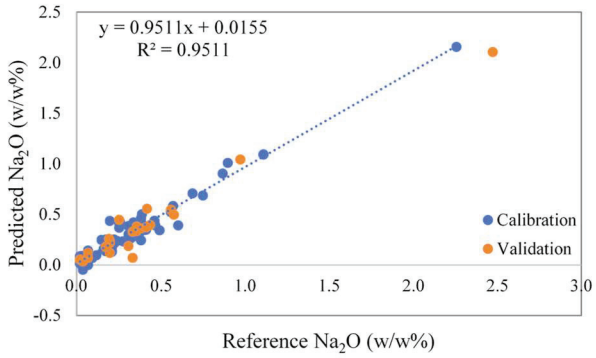


Figure 5.52. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of Na_2O

The first thirteen principal components were decided as the optimal number of latent variables and were used to construct PLS model to estimate the concentration of

Na_2O (w/w%). The performance of model were decided by plotting regression graph of the two set. The correlation graph is shown in Figure 5.53.

a)



b)

Na_2O	
SECV(w/w%)	0.07
SEP(w/w%)	0.12
R^2	0.95
Max	2.26
Min	0.02
Range	2.24
Latent Variables	13

Figure 5.53. a) Reference Na_2O concentrations vs. Predicted Na_2O concentrations by PLS. b) Statistical parameters of Na_2O

As shown in Figure 5.53, the regression coefficients of the model were found 0.95, 0.95 for calibration and validation set, respectively. The SECV value for calibration set and SEP value for validation set were calculated as 0.067, 0.118 (w/w%), respectively. If the Na_2O concentration range of the sample is considered, the range is found to be so narrow. Although the samples have low content of Na_2O component, the model has a high predicted ability because of obtained high regression coefficient for two sets.Figure 5.54, shows the residuals plot, which can be used to determine the error range and potential residual trend.

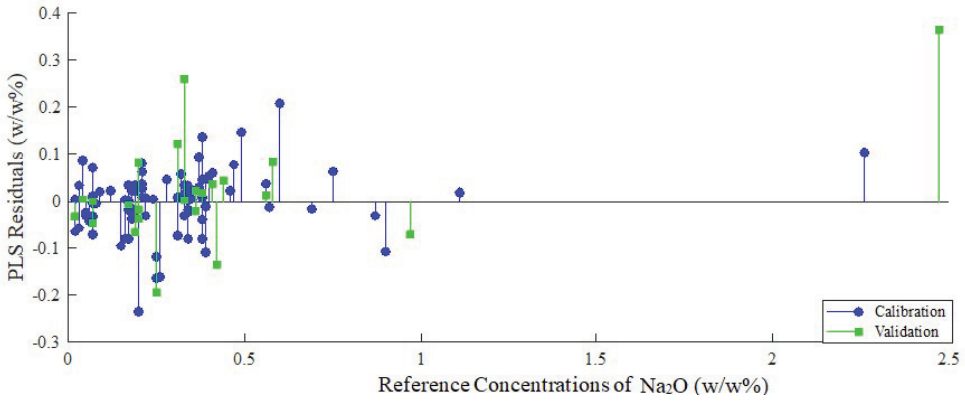


Figure 5.54. Reference concentrations vs. corresponding PLS prediction residuals of Na_2O

The most residuals are in the range ± 0.3. One sample in the validation set exceeded the 0.3 range. Also, that sample has the highest concentration of Na_2O. If more samples

added that near have that concentration, the model performance, especially for this type of sample, can be more robust.

5.1.2.8. The PLS result of K_2O component

For determination of optimal number of latent variables, PRESS values were calculated for the first thirty latent variables, and results are given in Figure 5.55..

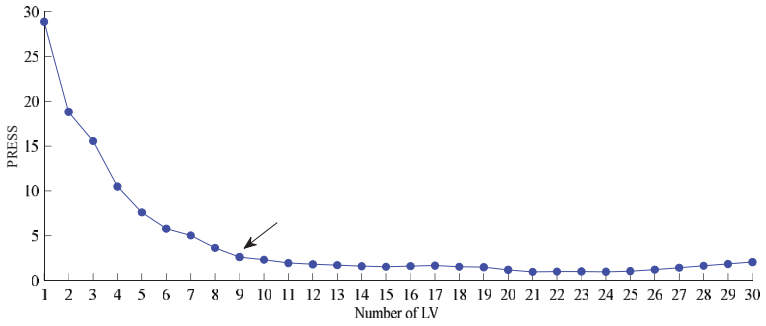
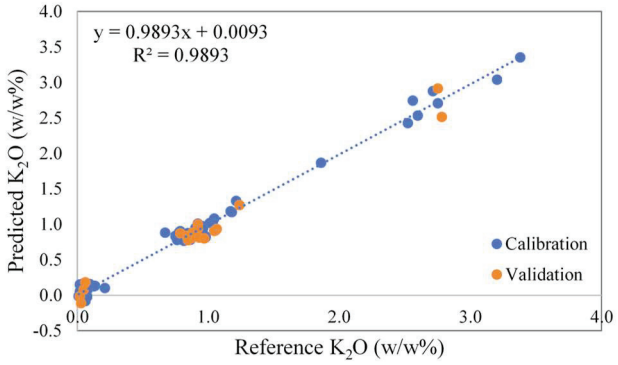


Figure 5.55. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of K_2O

The PLS model was constructed with the 9 latent variables. The evaluation of the model performance was shown in Figure 5.56.

a)



b)

K_2O	
SECV(w/w%)	0.07
SEP(w/w%)	0.10
R^2	0.99
Max	3.38
Min	0.01
Range	3.37
Latent Variables	9

Figure 5.56. a) Reference K_2O concentrations vs. Predicted K_2O concentrations by PLS. b) Statistical parameters of K_2O

The regression coefficients of the model were found as 0.98, 0.97 for calibration and validation sets, respectively. SECV and SEP values were calculated as 0.072 and 0.104 (w/w%), respectively. Although the concentration range so narrow, the model's predictive ability is found quite well shows that the constructed PLS model is robust.

Figure 5.57, shows the residuals plot, which can be used to determine the error range and potential residual trend.

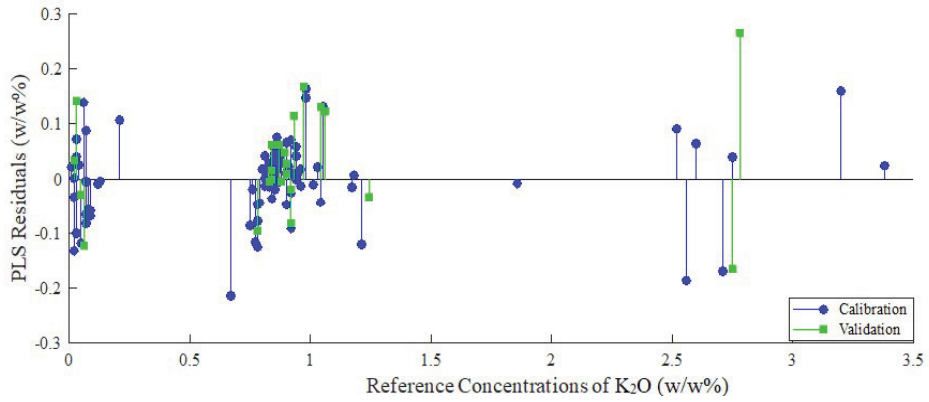


Figure 5.57. Reference concentrations vs. corresponding PLS prediction of K_2O

As shown in Figure 5.57, most residual are in the range of $\pm$ 0.3.

5.1.2.8. The PLS result of L.O.I component

For the loss of ignition estimation of samples given in Table 5.3, the optimal number of latent variables was determined by calculating PRESS values for the first thirty latent variables to construct a PLS model. The results were shown in Figure 5.58.

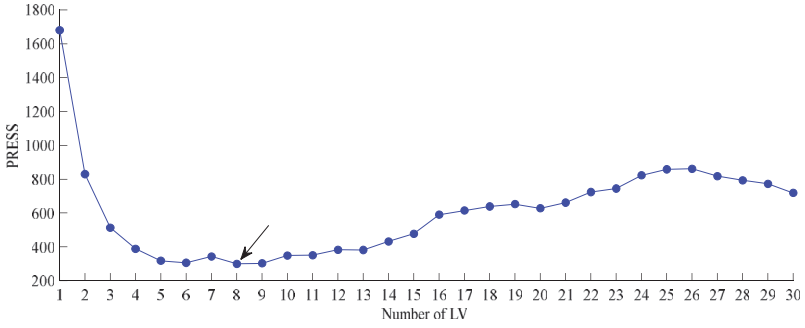
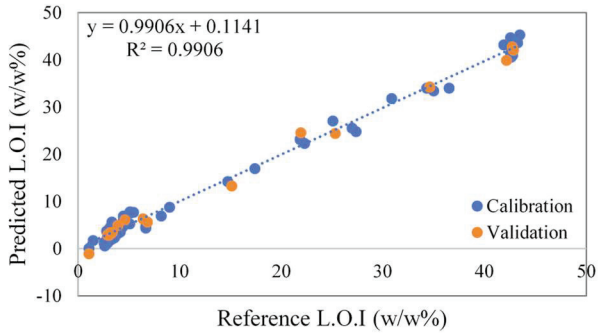


Figure 5.58. Number of PCs vs. PRESS plot for selecting the optimal number of PCs of L.O.I

By considering Figure 5.58, the eight principal components were used to construct a PLS model to predict the loss of ignition content of the studied samples. The model's performance was examined by plotting a correlation graph of the model that How

consistent the concentration values predicted by the PLS model are with the actual values, and the corresponding correlation graph and model parameters were given in Figure 5.59.

a)



b)

L.O.I	
SECV(w/w%)	1.40
SEP(w/w%)	1.24
R^2	0.99
Max	43.48
Min	1.03
Range	42.45
Latent variables	8

Figure 5.59. a) Reference L.O.I concentrations vs. Predicted L.O.I concentrations by PLS. b) Statistical parameters of L.O.I

The regression coefficients of the model were found as 0.99, 0.99 for calibration and validation set, respectively. The SECV and SEP values were calculated as 1.40, 1.24 (w/w%), respectively. The model's performance was found quite good, and also, there is a slight overfitting problem in PLS because the SEP value was found lower than the SECV value. The residual graph of the two sets was given in Figure 5.60.

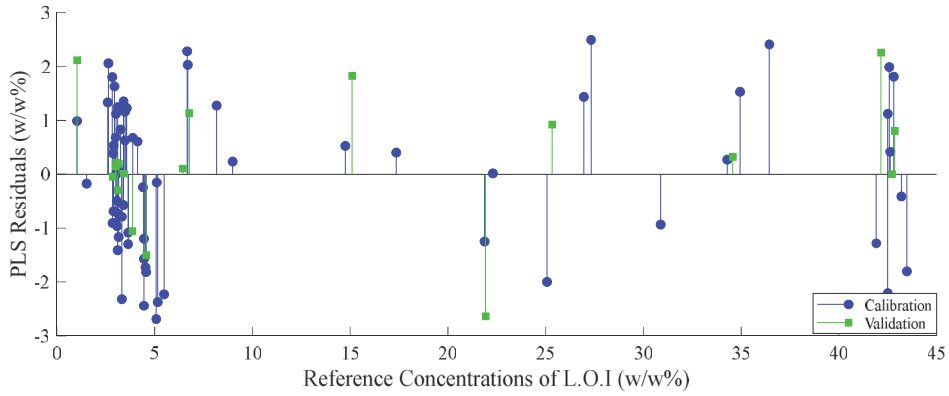


Figure 5.60. Reference concentrations vs. corresponding PLS prediction residuals of L.O.I

As shown in Figure 5.60, most residuals are in the range of ±3.

As a conclusion, all obtained results of PLS both without preprocessing and with EMSC preprocessing are given in Table 5.4.

Table 5.4. All PLS results of without pre-processing and with EMSC pre-processing

	Without Pre-Processing							
	Latent Variables	SECV (w/w%)	SEP (w/w%)	R^2(cal)	R^2(val)	Max	Min	Range
CaO	10	1.91	2.09	0.99	0.99	66.38	0.61	65.77
SiO_2	14	1.16	2.59	0.99	0.97	55.42	0.28	55.14
Al_2O_3	18	0.34	0.73	0.99	0.97	17.66	0.02	17.64
Fe_2O_3	18	0.16	0.31	0.99	0.97	6.13	0.02	6.11
MgO	17	0.28	0.31	0.94	0.92	9.13	0.05	9.08
SO_3	16	0.75	1.20	0.99	0.99	43.48	0.05	43.43
Na_2O	10	0.08	0.15	0.94	0.94	2.41	0.02	2.39
K_2O	14	0.21	0.17	0.92	0.95	3.98	0.01	3.97
L.O.I	16	0.79	3.03	0.99	0.96	43.48	1.03	42.05
	With EMSC pre-processing							
	Latent Variables	SECV (w/w%)	SEP (w/w%)	R^2(cal)	R^2(val)	Max	Min	Range
CaO	10	1.39	2.07	0.99	0.99	66.38	1.41	64.97
SiO2	6	1.42	1.97	0.99	0.98	55.42	0.28	55.14
Al_2O_3	11	0.49	0.45	0.98	0.99	17.66	0.02	17.64
Fe_2O_3	10	0.21	0.27	0.98	0.97	6.13	0.02	6.11
MgO	18	0.26	0.42	0.95	0.89	9.13	0.05	9.08
SO_3	13	0.40	0.69	0.99	0.99	43.48	0.05	43.43
Na_2O	13	0.07	0.12	0.95	0.95	2.26	0.02	2.24
K_2O	9	0.07	0.10	0.99	0.97	3.38	0.01	3.37
L.O.I	8	1.41	1.25	0.99	0.99	43.48	1.03	42.45

As shown in Table 5.4, the obtained PLS model after applied the EMSC pre-processing method of the spectra of the samples has more effective than the first PLS model with raw spectra of the samples. The PLS model with EMSC is more successful in predicting each component's concentration because the regression coefficients of the calibration and validation sets were found higher than PLS with the without pre-processing method. Also, SECV and SEP values were found much lower, and the PLS model was constructed for each component with a relatively low number of latent variables that show the model has less noise. Only the MgO component has better predicted with the first PLS model. Finally, the Paired t-test method was applied to SECV and SEP values obtained from both PLS models with raw FTIR spectra, and EMSC applied FTIR spectra. The results are shown in Table 5.5.

Table 5.5. t-Test: Paired Two Sample for Means

	SECV (w/w%)-Raw Data	SECV (w/w%)-EMSC	SEP (w/w%)-Raw Data	SEP (w/w%)-EMSC
Mean	0.63	0.63	1.18	0.82
Variance	0.36	0.35	1.26	0.59
t Stat	-0.03		1.84	
P(T<=t) one-tail	0.49		0.05	
t Critical one-tail	1.86		1.86	
P(T<=t) two-tail	0.98		0.10	
t Critical two-tail	2.31		2.31	

The purpose of applying Paired t-test is to compare SECV values obtained with the two different PLS calibration models and compare SEP values obtained from the validation sets of two different PLS models. As seen in Table 5.5, the t-stat is calculated as -0.03. According to t Critical two-tail at 95% confidence level, -0.03 is between ± 2.31, and the result shows no significant difference between SECV values obtained from the PLS model with raw data and PLS model-EMSC applied. For SEP values, t-stat is calculated as 1.84. According to t Critical two-tail at 95% confidence level, 1.84 is between ± 2.31. It can be said that there is no significant difference also SEP values obtained from validation sets of both two methods. On the other hand, mean of the SEP values is decreased from 1.18 to 0.82 to validation set with EMSC also variance of the SEP values is decreased from 1.26 to 0.59 with EMSC. Although the result was insignificant at the 95% confidence level, when examined numerically, it was seen that better SEP values were obtained by the PLS models with EMSC, with the decrease of the mean and variance in the SEP values.

5.2. The second scenario

The second scenario was that the CEM-I samples were used to construct a calibration model to predict significant oxides, constituting its chemical composition. The reason for constructing a calibration model with CEM-I types of cement samples was that every year in Turkey, a considerable amount of CEM-I is manufactured and used for local purposes. According to the Turkish cement Manufacturers' Association, in 2020, the sale capacity of CEM-I was approximately 30 million tons per year in domestic, and abroad sale capacity of CEM-I was approximately 10 million tons per year. (Türkiye Çimento Sanayicileri Birliği, n.d.-c) Each cement factory produces CEM-I type as the main product. For that reason, CEM-I quality assurance has a significant factor, and CEM-I

types of cement must have certain qualities. Changes in cement quality result from changes in the amounts of chemical components that affect the physical performance of cement. In the second scenario for a total of 56 CEM-I samples, a calibration model was tried to create to detect its chemical composition.

A total of 40 CEM-I samples were used as a calibration set to construct the PLS model, while the remaining 16 samples were used as an independent validation set to assess the model's predictive performance. The calibration data was left with at least three samples at the upper and lower limits to ensure that the resulting model covered boundary conditions and accounted for the majority of the variance, and the remaining samples were randomly distributed. In Table 5.6 the calibration data set samples' reference analyses of each component are given, and in Table 5.7, the validation set samples' reference analyses are given.

Table 5.6. Calibration data set for second scenario samples.

Calibration Data Set										
No	Sample	CaO (w/w%)	SiO_2 (w/w%)	Al_2O_3 (w/w%)	FE_2O_3 (w/w%)	MgO (w/w%)	SO_3 (w/w%)	Na_2O (w/w%)	K_2O (w/w%)	L.O.I (w/w%)
1	s44	63.75	20.29	5.26	2.18	1.3	3.28	0.17	0.75	3.01
2	s2	63.77	19.56	5.05	3.2	1.25	2.94	0.34	0.75	2.62
3	s14	63.46	19.68	5.3	3.17	1.28	2.9	0.35	0.75	2.6
4	s33	63.48	20.21	5.58	2.08	1.34	3.22	0.19	0.78	3.11
5	s47	63.48	20.21	5.58	2.08	1.34	3.22	0.19	0.78	3.11
6	s31	63.46	20.12	5.52	2.05	1.25	3.06	0.2	0.78	3.55
7	s55	62.27	18.94	4.89	2.86	2.48	3.39	0.49	0.79	3.42
8	s56	62.17	19.13	4.72	3.13	1.73	2.82	0.38	0.81	4.56
9	s29	63.34	20.24	5.4	2.12	1.67	3.24	0.16	0.82	3
10	s9	63.65	19.48	4.98	3.19	1.23	2.92	0.33	0.82	2.89
11	s24	63.21	19.35	5.13	3.27	1.23	3.05	0.34	0.82	3.09
12	s34	63.42	19.83	5.23	2.35	1.54	3.62	0.17	0.83	3
13	s35	63.02	19.99	5.54	2.12	1.59	3.57	0.18	0.83	3.15
14	s15	63.89	19.37	5.2	3.2	1.34	3.16	0.35	0.83	2.21
15	s13	63.8	19.02	5.42	3.26	1.28	2.74	0.38	0.83	2.76
16	s45	63.28	20.18	5.54	2.16	1.36	2.79	0.19	0.84	3.64
17	s50	62.5	20.32	5.69	2.28	1.46	3.55	0.2	0.84	3.15
18	s6	62.99	19.49	5.29	3.67	1.26	3.02	0.33	0.84	2.6
19	s54	63.13	18.78	5.16	2.88	1.98	3.25	0.38	0.84	3.03
20	s51	62.76	20.21	5.53	2.1	2.02	3.45	0.16	0.85	2.9
21	s41	63.25	20.36	5.41	2.17	1.67	3.21	0.17	0.85	2.9
22	s5	63.33	19.31	5.36	3.53	1.28	2.77	0.35	0.85	2.7
23	s57	63.1	18.98	5.34	2.84	1.36	3.47	0.37	0.85	3.17
24	s40	63.16	20.44	5.53	2.14	1.23	3.01	0.21	0.86	3.41
25	s37	62.46	20.24	5.62	2.26	1.52	3.32	0.21	0.86	3.57
26	s10	63.49	19.49	5.14	3.36	1.25	2.77	0.34	0.86	2.88
27	s17	65.01	19.96	5.27	3.57	1.28	3.2	0.34	0.86	2.43
28	s8	63.98	19.47	5.34	3.44	1.27	2.79	0.35	0.86	2.99
29	s36	64.89	19.16	5.26	2.85	1.3	2.82	0.2	0.87	2.63
30	s32	62.28	21.07	5.66	2.18	1.47	3.42	0.2	0.87	2.83
31	s53	62.33	20.82	5.76	2.32	1.47	3.11	0.22	0.88	3.07
32	s38	63.45	19.81	5.47	2.91	1.49	3.17	0.18	0.9	2.6
33	s43	61.96	21.23	5.76	2.1	1.9	3.07	0.21	0.9	2.86
34	s1	63.57	19.09	5.33	3.21	1.31	2.8	0.35	0.81	3.01
35	s28	63.12	19.58	5.38	3.32	1.29	2.99	0.34	0.81	2.76
36	s46	63.65	20.2	5.37	2.29	1.28	2.8	0.18	0.81	3.41
37	s26	63.43	19.5	5.16	3.42	1.25	2.79	0.34	0.82	2.78
38	s20	63.27	19.38	5.13	3.42	1.28	3.02	0.33	0.83	2.82
39	s16	63.62	19.4	5.11	3.11	1.28	3.09	0.34	0.84	2.69
40	s18	63.18	19.3	5.14	3.41	1.28	3.1	0.33	0.87	2.88

Table 5.7. Validation data set for second scenario samples

Validation Data Set										
No	Sample	CaO (w/w%)	SiO_2 (w/w%)	Al_2O_3 (w/w%)	Fe_2O_3 (w/w%)	MgO (w/w%)	SO_3 (w/w%)	Na_2O (w/w%)	$_2O$ (w/w%	L.O.I (w/w%)
41	s12	62.9	19.5	5.26	3.14	1.24	2.99	0.35	0.8	3.31
42	s39	63.7	20.01	5.33	2.12	1.41	3.37	0.19	0.8	3.07
43	s19	63.71	19.26	5.15	3.38	1.27	2.95	0.34	0.81	2.61
44	s23	63.45	19.4	5.08	3.19	1.25	2.91	0.35	0.81	3.05
45	s3	63.1	19.58	5.19	3.39	1.27	3.08	0.34	0.81	2.72
46	s49	63.31	20.29	5.45	2.38	1.46	3.28	0.17	0.81	2.84
47	s25	63.18	19.05	5.4	3.26	1.32	2.93	0.38	0.83	3.14
48	s22	63.29	19.46	5.22	3.4	1.28	2.99	0.35	0.83	2.68
49	s30	63.38	19.82	5.14	2.66	1.43	3.52	0.19	0.83	3
50	s48	63.06	20.32	5.37	2.12	1.44	3.55	0.17	0.83	3.12
51	s42	63.53	20.41	5.27	2.15	1.38	3.22	0.17	0.83	3.03
52	s7	63.38	19.48	5.22	3.35	1.27	2.89	0.35	0.85	2.7
53	s21	63.35	19.34	5.33	3.22	1.28	2.96	0.36	0.86	2.79
54	s52	62.76	20.52	5.66	2.18	1.58	3.3	0.18	0.86	2.95
55	s4	63.13	19.38	5.42	3.12	1.25	3.07	0.38	0.87	2.88
56	s11	63.35	19.47	5.16	3.29	1.28	3.08	0.34	0.88	2.64

Figure 5.61, shows the raw FTIR-ATR absorbance spectra for a total of 56 CEM-I samples, which were recorded in the 4000–550 cm-1 wavenumber region.

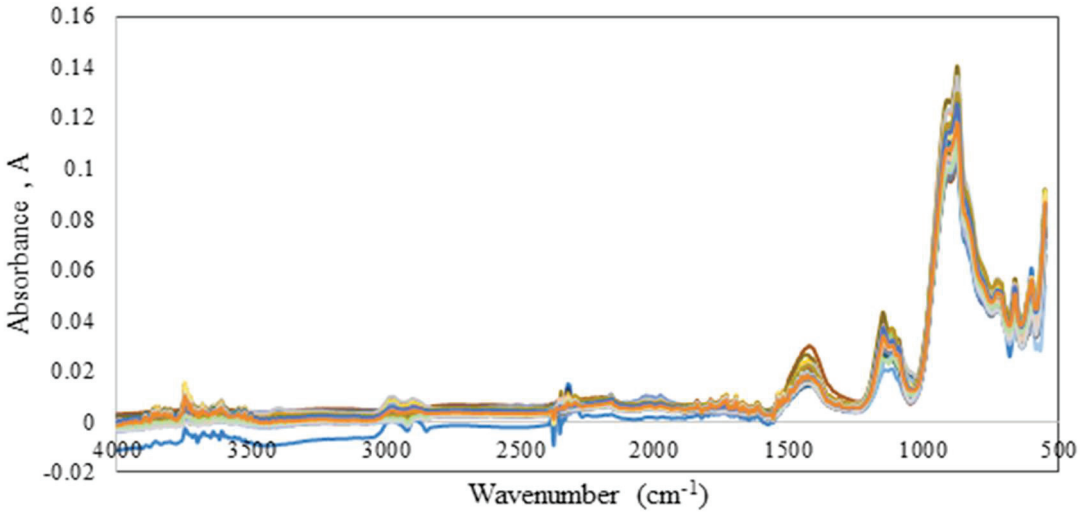


Figure 5.61. Raw FT-NIR spectra of a total of 56 CEM-I samples

The mean spectra of each sample were used to construct a PLS model, which was done in triplicate. The data was reduced by removing the sensitivity to atmospheric carbon dioxide area 2450-2350 cm^{-1}, and mean spectra were pre-processed with extended multiplicative scattering correction (EMSC). The PLS models of each oxide and loss of ignition component were constructed after mean-centering all spectra. The optimal number of latent variables was calculated using the leave one out cross-validation method. In the future, the best models will be given. The view of the EMSC applied spectra is given in Figure 5.62.

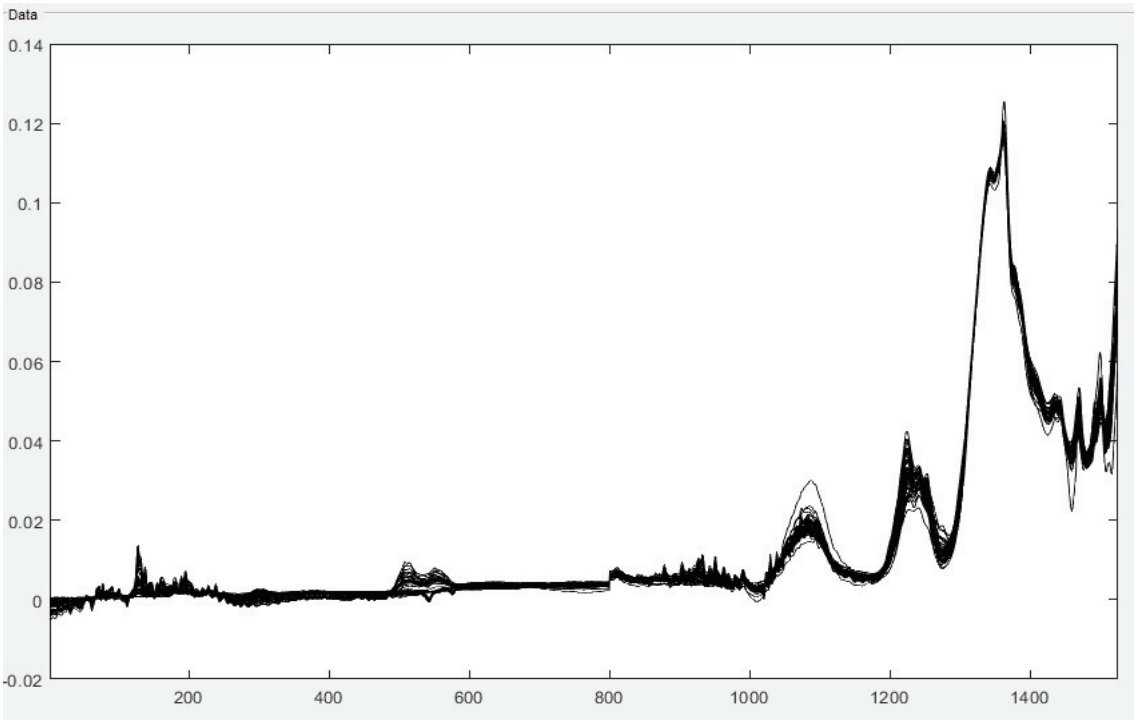


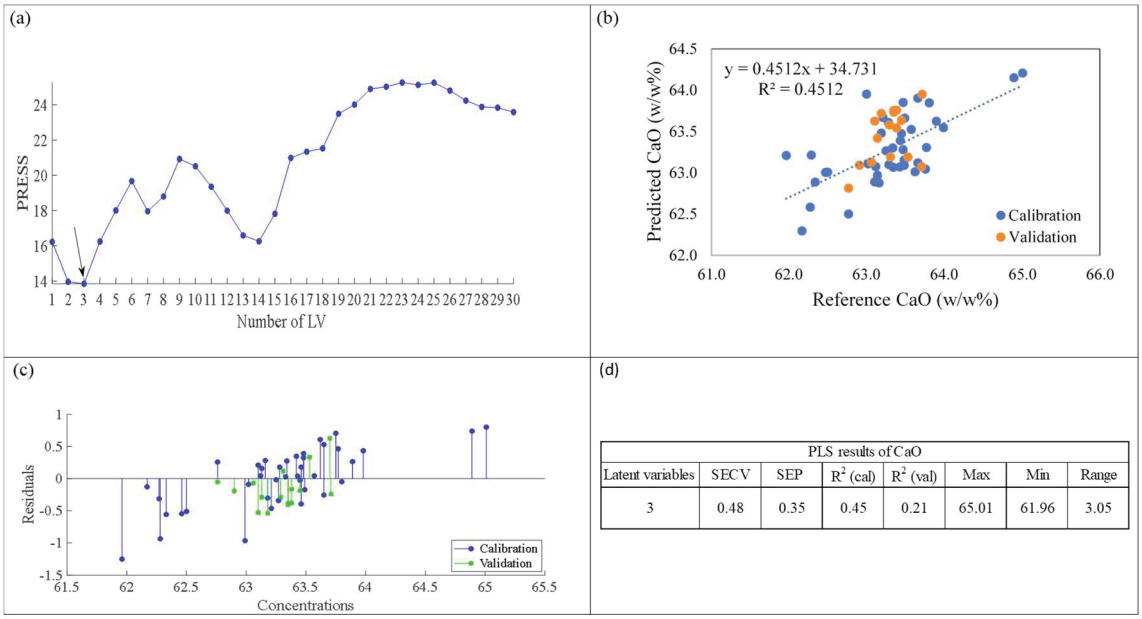
Figure 5.62. The view of spectra of CEM-I samples after EMSC pre-processed method applied

In Figure 5.62, the x-axis represents the wavenumber (cm^{-1}) from 550 cm^{-1} to 4.000 cm^{-1}, and the y axis represents the absorbance values. The program that was applied the EMSC has its own scale for wavenumber. Therefore, the numbers on the x-axis are different from the x axis of Figure 5.61.

The PLS was performed on the data set of 56 CEM-I samples. Also, the EMSC pre-processing was applied for FTIR-ATR spectra to reduce baseline shifts caused by the particle size of cement samples. Major and minor oxides were examined with the constructed PLS model to predict concentrations that constitute CEM-I chemical composition. In addition, the loss of ignition analysis is performed to control the humidity of the cement samples, and PLS also examined the estimation of the loss of ignition concentration of the studied samples. The regression coefficients (R^2) of the calibration and validation data sets, as well as the standard error of cross-validation (SECV) and standard error of prediction (SEP), were used to assess the prediction performance of the developed models. Models with high regression coefficients and low SECV and SEP values were favored. The PLS results were given for each component in the below.

5.2.1. The PLS results of CaO

Prediction of CaO concentration of studied samples was done by PLS method, the results of constructed PLS model were given in Figure 5.63.



PLS results of CaO							
Latent variables	SECV	SEP	R^2 (cal)	R^2 (val)	Max	Min	Range
3	0.48	0.35	0.45	0.21	65.01	61.96	3.05

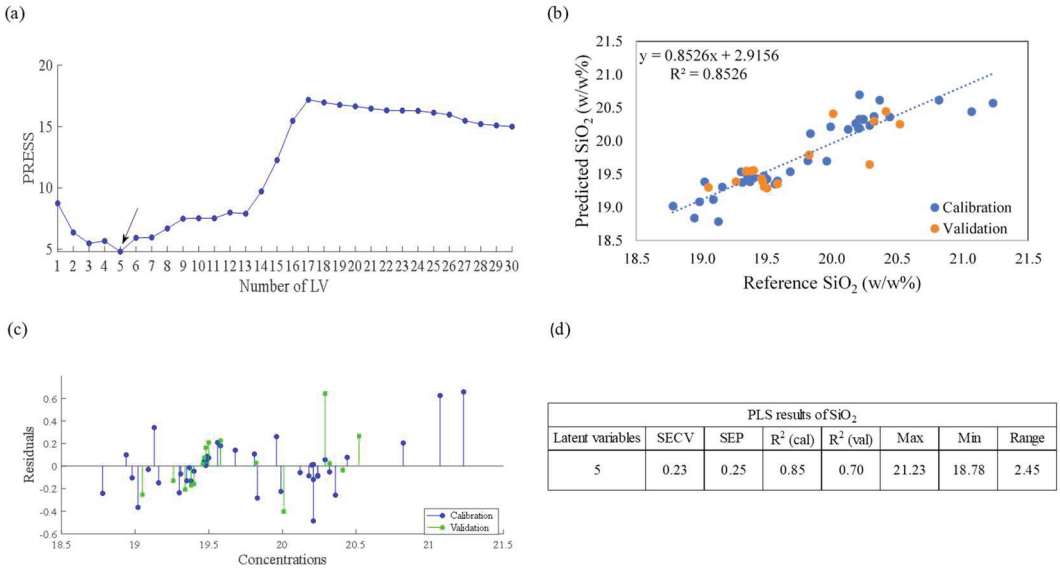
Figure 5.63. a) Number of PCs vs. PRESS plot for selecting the optimal number of PCs. b) Reference CaO concentrations vs. Predicted CaO concentrations c) Reference CaO concentrations vs. corresponding PLS prediction residuals d) Statistical parameters of CaO.

For determining the optimal number of latent variables, PRESS values were calculated for the first thirty latent variables, and results were given in Figure 5.63a. Considering Figure 5.63a, the first three latent variables were used to construct the PLS model to predict CaO concentration (w/w%). The PLS model performance was evaluated by plotting the correlation graph of the calibration and validation sets. Reference CaO concentrations against predicted CaO concentrations (w/w%) were shown in Figure 5.63b. The regression coefficient (R^2) of the calibration set was found as 0.4512, and the regression coefficient (R^2) of the validation set was found as 0.21. The model performance is not close for calibration and validation. When the range of CaO concentration of samples is examined, it is observed that the range is so narrow, making prediction more difficult for the PLS model. By adding more samples to the data set, the predictability of the model can be improved. Standard error of cross-validation (SECV) and standard error of prediction (SEP) were calculated as 0.48, 0.35 (w/w%),

respectively. The reference CaO concentrations vs. corresponding PLS prediction residuals were given in Figure 5.63c. Most of the residuals are in the range between ±1. Finally, all statistical parameters were shown in Figure 5.63d.

5.2.2. The PLS results of SiO_2

Prediction of SiO_2 concentration of studied samples was done by PLS method, the results of constructed PLS model were given in Figure 5.64.



PLS results of SiO_2							
Latent variables	SECV	SEP	R^2 (cal)	R^2 (val)	Max	Min	Range
5	0.23	0.25	0.85	0.70	21.23	18.78	2.45

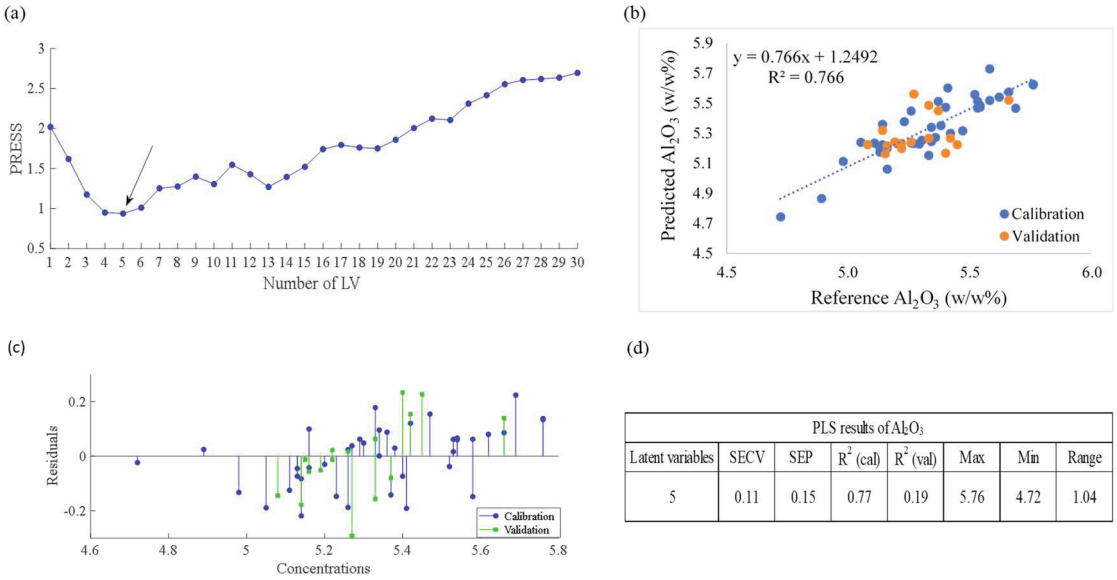
Figure 5.64. a) Number of PCs vs. PRESS plot for selecting the optimal number of PCs. b) Reference SiO_2 concentrations vs. Predicted SiO_2 concentrations c) Reference SiO_2 concentrations vs. corresponding PLS prediction residuals d) Statistical parameters of SiO_2.

As shown in Figure 5.64 a, the PRESS values were calculated to find the optimal number of latent variables to construct a PLS calibration model to estimate the concentration of the SiO_2 content in the samples. By taking into consideration Figure 5.64 a, five latent variables were used to create the PLS model. Then the performance of the PLS model was evaluated by finding the regression coefficients of the calibration model and independent validation set to test our calibration model predictability. The correlation graph was shown in Figure 5.64b. The regression coefficients of the calibration and validation set were found as 0.85, 0.70, respectively. If considered to the concentration range of the samples, these regression coefficients are not too low but not high enough to trust the model's predictability. SECV and SEP values were calculated as 0.23 and 0.25 (w/w%), respectively. Reference concentrations vs. corresponding PLS prediction

residuals were given in Figure5.64c. As shown in Figure 5.64c, most residuals are in the range of ±0.6. Finally, all statistical parameters are shown in Figure 5.64d.

5.2.3. The PLS results of Al_2O_3

Prediction of Al_2O_3 concentration of studied samples was done by PLS method, the results of constructed PLS model were given in Figure 5.65.



PLS results of Al_2O_3							
Latent variables	SECV	SEP	R^2 (cal)	R^2 (val)	Max	Min	Range
5	0.11	0.15	0.77	0.19	5.76	4.72	1.04

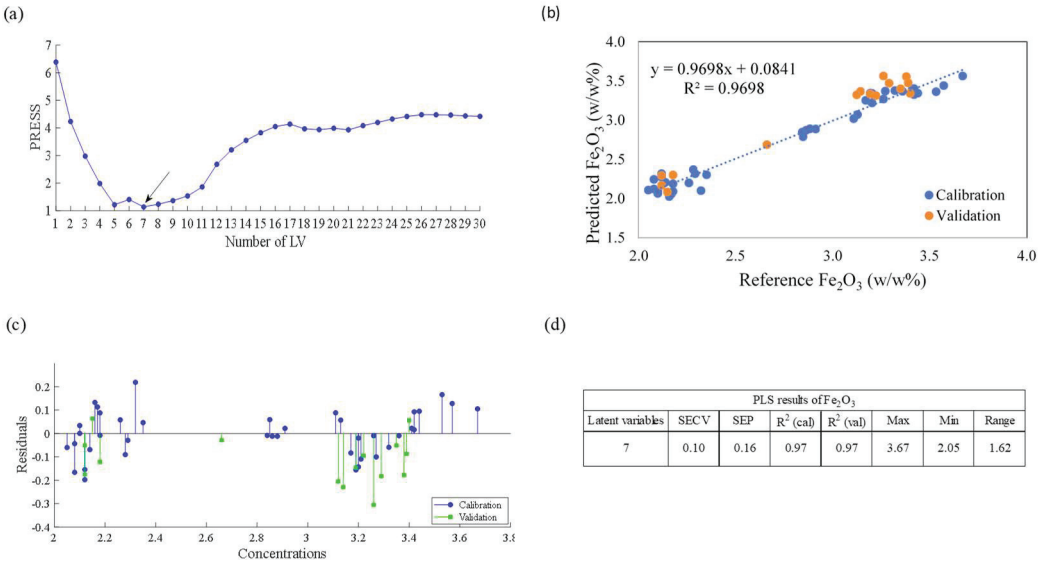
Figure 5.65. a) Number of PCs vs. PRESS plot for selecting the optimal number of PCs. b) Reference Al_2O_3 concentrations vs. Predicted Al_2O_3 concentrations c) Reference Al_2O_3 concentrations vs. corresponding PLS prediction residuals d) Statistical parameters of Al_2O_3.

For the construction of a PLS model, the optimal numbers of latent variables were decided by calculating the PRESS values, then, using five latent variables, PLS model was constructed. The performance evaluation of the model was done by plotting the correlation graph of calibration and independent validation set. The correlation graph of the model was given in Figure 5.66b. Correlation coefficients (R^2) were found as 0.77, 0.20 for calibration and validation sets, respectively. Correlation coefficients show that the PLS model failed to predict the Al_2O_3 concentration of the samples because of the having low regression coefficients. If considered to Al_2O_3 concentration of the validation set samples are distributed in the narrow concentration compared to the calibration set samples, that might be the reason to have a low regression coefficient for the validation set. SECV and SEP values were calculated as 0.11 and 0.15 (w/w%), respectively. The

Al_2O_3 concentration range of all samples is 1.04 (w/w%), and the residuals are in the range between ±0.2. Finally, all statistical parameters were given in Figure 5.65d.

5.2.4. The PLS results of Fe_2O_3

Prediction of Fe_2O_3 concentration of studied samples was done by PLS method, the results of constructed PLS model were given in Figure 5.66.



PLS results of Fe_2O_3							
Latent variables	SECV	SEP	R^2 (cal)	R^2 (val)	Max	Min	Range
7	0.10	0.16	0.97	0.97	3.67	2.05	1.62

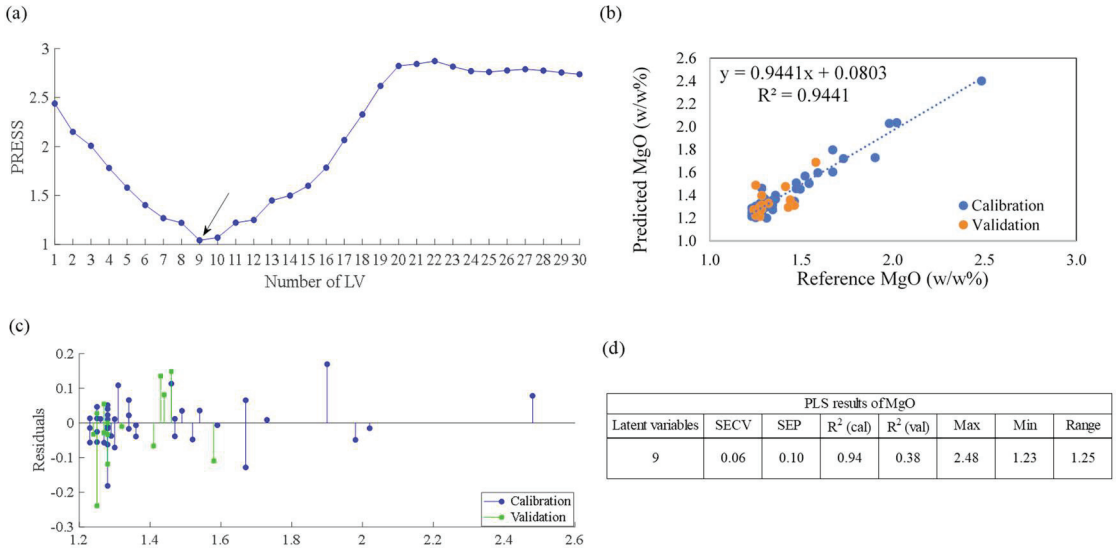
Figure 5.66. a) Number of PCs vs. PRESS plot for selecting the optimal number of PCs. b) Reference Fe_2O_3 concentrations vs. Predicted Fe_2O_3 concentrations c) Reference Fe_2O_3 concentrations vs. corresponding PLS prediction residuals d) Statistical parameters of Fe_2O_3.

As shown in Figure 5.66a, the PRESS values were calculated to find the optimal number of latent variables. By considering the corresponding figure, the first seven latent variables were used to form a PLS model. The performance evaluation of the model was done by plotting the reference Fe_2O_3 concentration against predicted Fe_2O_3 concentrations for both calibration and validation sets in Figure 5.66b. The regression coefficients of the model were found as 0.97, and 0.97 for calibration and validation sets, respectively. The SECV and SEP values were calculated as 0.10, 0.16, respectively. As shown in Figure 5.66b, the calibration set covered as much variance as possible for Fe_2O_3 concentrations, and the validation set represents a realistic scenario that overfitting could not be observed. The regression coefficients of the Fe_2O_3 concentrations were found high; also, SECV and SEP values were found small and close shows that the prediction ability of the model was reliable. The possible error range and residual trends were shown in

Figure 5.66c that apparently, most residuals are in the range between ±0.2. Lastly, the results obtained from the PLS model were shown in Figure 5.66d.

5.2.5. The PLS results of MgO

Prediction of MgO concentration of studied samples was done by PLS method, the results of constructed PLS model were given in Figure 5.67.



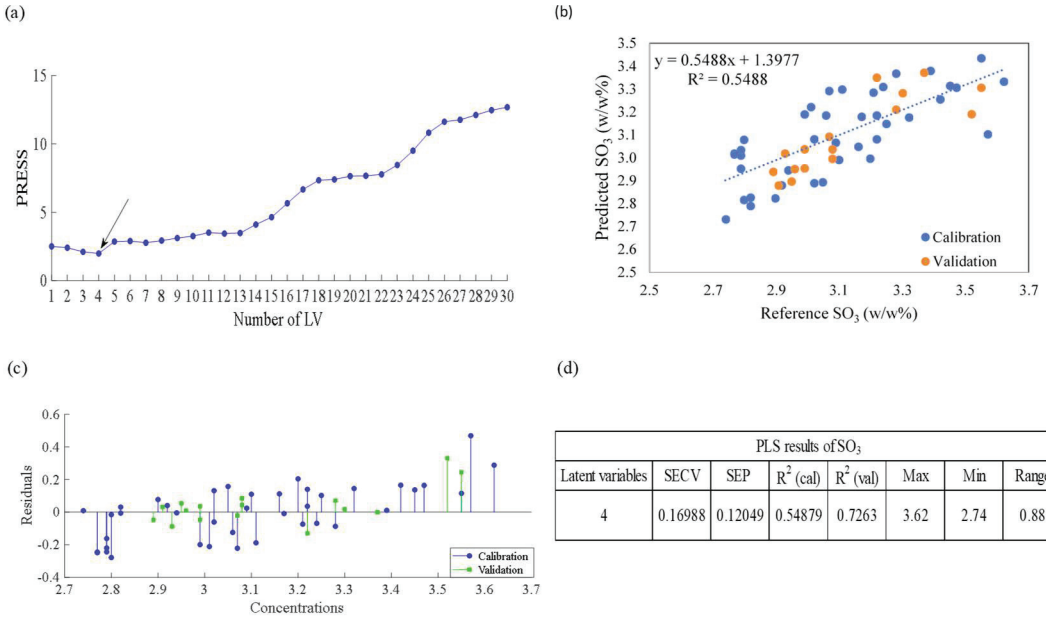
PLS results of MgO							
Latent variables	SECV	SEP	R^2 (cal)	R^2 (val)	Max	Min	Range
9	0.06	0.10	0.94	0.38	2.48	1.23	1.25

Figure 5.67. a) Number of PCs vs. PRESS plot for selecting the optimal number of PCs. b) Reference MgO concentrations vs. Predicted MgO concentrations c) Reference MgO concentrations vs. corresponding PLS prediction residuals d) Statistical parameter of MgO

For the determination of the optimal number of latent variables, the PRESS values were calculated for the first thirty latent variables shown in Figure 5.67a. By considering the corresponding Figure 5.67a, the nine latent variables were used to create a PLS model. The model's performance was obtained by plotting a correlation graph for the calibration and validation set, which was given in Figure 5.67b. The regression coefficient (R^2) of the PLS model was found as 0.94, and the regression coefficient (R^2) of the validation set was found as 0.38 shows that the model performance at the prediction stage against a data that is not used for modeling (independent validation set) was not reasonable. The SECV and SEP values were calculated as 0.06 and 0.10 (w/w%), respectively. The error range and the residuals trend were shown in Figure 5.67c. As shown in Figure 5.67c, most residuals were in the range between ±0.2. Lastly, all results are represented in Figure 5.67d.

5.2.6. The PLS results of SO_3

Prediction of SO_3 concentration of studied samples was done by PLS method, the results of constructed PLS model were given in Figure 5.68.



PLS results of SO_3							
Latent variables	SECV	SEP	R^2 (cal)	R^2 (val)	Max	Min	Range
4	0.16988	0.12049	0.54879	0.7263	3.62	2.74	0.88

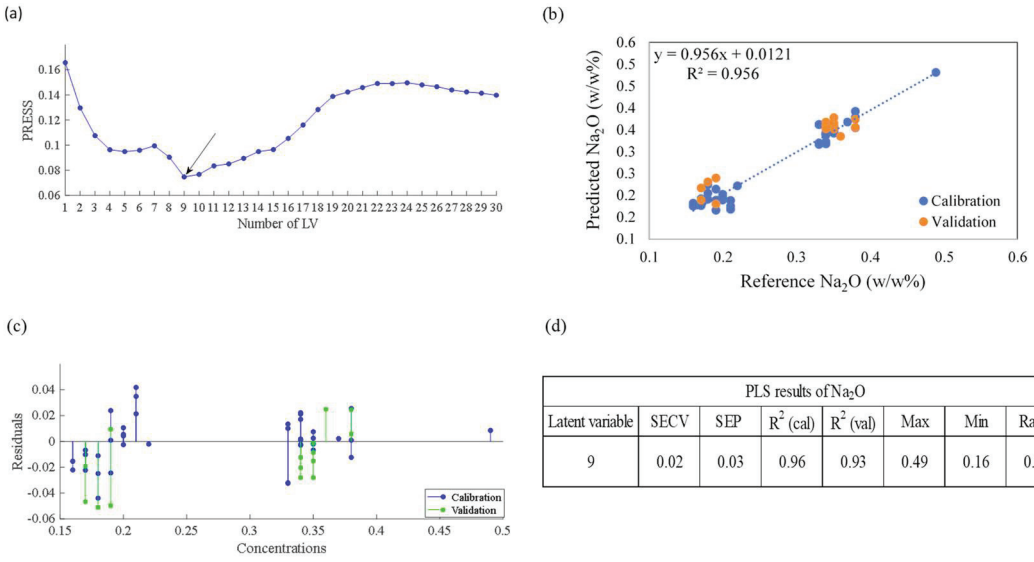
Figure 5.68. a) Number of PCs vs. PRESS plot for selecting the optimal number of PCs. b) Reference SO_3 concentrations vs. Predicted SO_3 concentrations c) Reference SO_3 concentrations vs. corresponding PLS prediction residuals d) Statistical parameters of SO_3.

For the estimation of SO_3 concentration, a PLS model was constructed with the optimal number of latent variables founded by calculating the PRESS values for the first 30 LVs then, and the first four latent variables were used to construct the PLS model. The model's performance was examined, by plotting reference SO_3 concentrations against predicted SO_3 concentrations by the PLS model for both calibration and validation sets. The regression coefficient of the PLS model was found as 0.58, and the validation set created to test the PLS model's prediction regression coefficient was found 0.76. The validation set's performance was found higher than the calibration set. Also, SECV and SEP values were calculated as 0.169 and 0.12, respectively. SEP value is lower than the SECV value. The selected validation samples were in the range of intermediate reference concentrations for the SO_3 component. The lowest and highest values in the studied reference concentration range were often included in the calibration set. If the validation set had a wider concentration range comparable to the calibration set, the errors might

have been similar. The residuals graph is given in Figure 5.68c shows that residuals are in the range between ±0.2. Lastly, all results are given in Figure 5.68d.

5.2.7. The PLS results of Na_2O

Prediction of Na_2O concentration of studied samples was done by PLS method, the results of constructed PLS model were given in Figure 5.69.



PLS results of Na_2O							
Latent variable	SECV	SEP	R^2 (cal)	R^2 (val)	Max	Min	Range
9	0.02	0.03	0.96	0.93	0.49	0.16	0.33

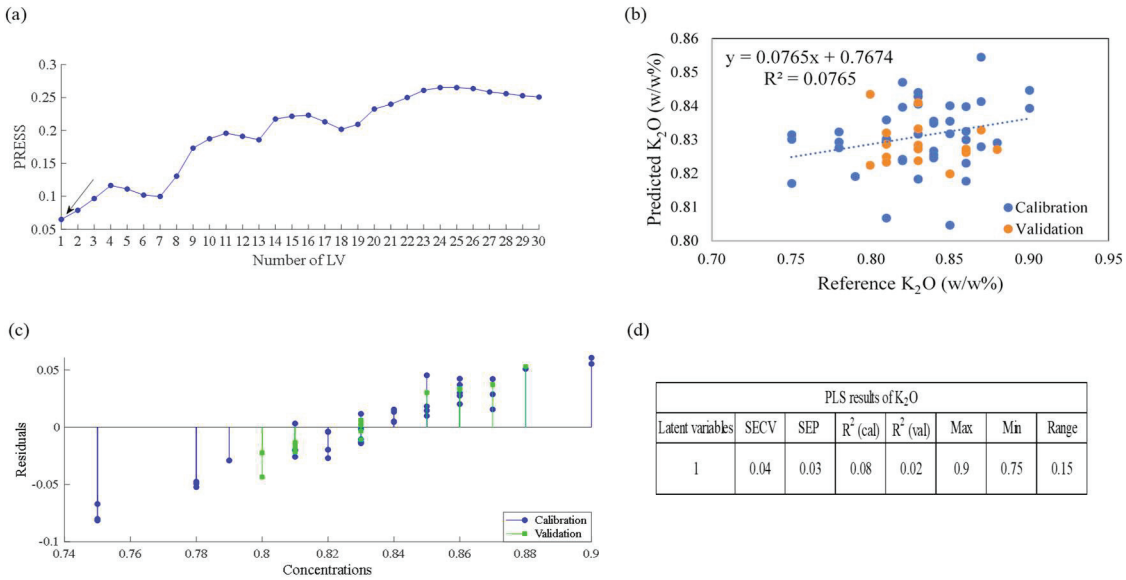
Figure 5.69. a) Number of PCs vs. PRESS plot for selecting the optimal number of PCs. b) Reference Na_2O concentrations vs. Predicted Na_2O concentrations c) Reference Na_2O concentrations vs. corresponding PLS prediction residuals d) Statistical parameters of Na_2O

PRESS values were calculated for the first thirty latent variables to find the optimal number of latent variables shown in Figure 5.69a. By considering Figure 5.69a, the first nine latent variables were used to create a PLS model for estimating the concentration of Na_2O in the samples. The model performance was evaluated by taking into consideration the correlation graph of calibration and validation sets. The correlation graph was given in Figure 5.69b. The regression coefficient of the model was found as 0.96, and the regression coefficient of the independent validation set was found as 0.93. The SECV and SEP values were calculated as 0.02 and 0.03 (w/w%), respectively. The regression coefficients of the two sets were found high, and SECV, SEP values were found low. The reference Na_2O concentrations vs. corresponding PLS prediction residuals were given in Figure 5.69c. As shown in Figure 5.69c, most residuals are in the range between ±0.04, and it is seen that the reference concentration values are divided into two groups, and there are many reference concentrations at the same value. The

concentration of Na_2O in the samples is extremely low when compared to other major oxides. As a result, XRF measurements for Na_2O were not precise enough, and several samples had the same concentration values, leading to unsatisfactory PLS results. As a result, to assure XRF measurement precision, each sample must be measured multiple times. Lastly, all statistical parameters obtained from the PLS model were given in Figure 5.69d.

5.2.8. The PLS results of K_2O

Prediction of K_2O concentration of studied samples was done by PLS method, the results of constructed PLS model were given in Figure 5.70.



PLS results of K_2O							
Latent variables	SECV	SEP	R^2 (cal)	R^2 (val)	Max	Min	Range
1	0.04	0.03	0.08	0.02	0.9	0.75	0.15

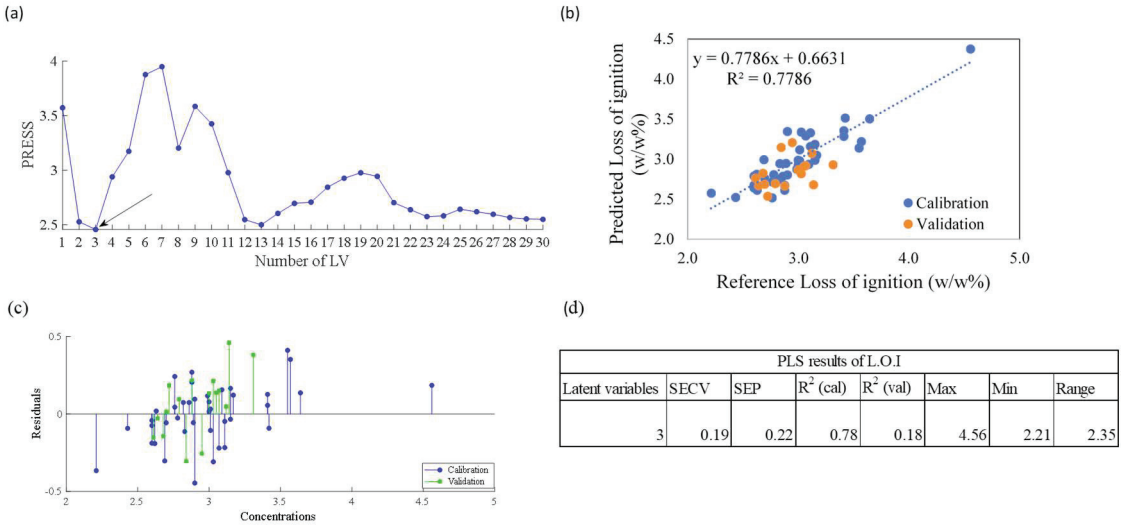
Figure 5.70. a) Number of PCs vs. PRESS plot for selecting the optimal number of PCs. b) Reference K_2O concentrations vs. Predicted K_2O concentrations c) Reference K_2O concentrations vs. corresponding PLS prediction residuals d) Statistical parameters of K_2O

The PRESS values were calculated for finding the optimal number of latent variables, and the results were shown in Figure 5.70a. The optimal number of the latent variable was chosen as one show that the K_2O content in the samples was not modeled with PLS. Other latent variables were tried to improve model performance, but this time model has an overfitting problem because the SEP value is increased. Also, regression coefficients were found at lower values. The PLS model is not working because the samples have a low concentration of K_2O, and the PLS model fails to capture relevant

spectral information for K_2O. Therefore, it causes poor fit and low prediction. Also, there are many reference concentrations at the same value, as shown in Figure 5.70c. Therefore, XRF measurement precision could not be achieved for K_2O. After replicating measurements for each sample, the PLS model can be created again. Finally, all statistical parameters for the PLS model are shown in Figure 5.70d.

5.2.9. The PLS results of Loss of ignition

Prediction of Loss of ignition estimation of studied samples was done by PLS method, the results of constructed PLS model were given in Figure 5.71.



PLS results of L.O.I							
Latent variables	SECV	SEP	R^2 (cal)	R^2 (val)	Max	Min	Range
3	0.19	0.22	0.78	0.18	4.56	2.21	2.35

Figure 5.71. a) Number of PCs vs. PRESS plot for selecting the optimal number of PCs. b) Reference L.O.I concentrations vs. Predicted L.O.I concentrations c) Reference L.O.I concentrations vs. corresponding PLS prediction residuals d) Statistical parameters of L.O.I.

Loss of ignition values of samples was estimated by constructing a PLS model. PLS model was constructed by finding the optimal number of latent variables for that the PRESS values were calculated for the first thirty latent variables, and the results were shown in Figure 5.71a. By considering 5.71a, the first third latent variable was used for constructing a PLS model. The model's performance was evaluated by considering the regression coefficients of the calibration and validation sets. The correlation graph was given in Figure 5.71b. The calibration coefficient was found as 0.77. The validation set's regression coefficient was found as 0.182 and seen that the predictability of the PLS model for L.O.I was not found reliable because of the enormous difference between the calibration and validation regression coefficients. The SECV and SEP values were calculated as 0.19 and 0.22, respectively. The possible error range and the residuals trend

were given in Figure 5.71c. Lastly, all results obtained from the PLS model were given in Figure 5.71d.

In the second scenario, PLS models were tried to be constructed to predict concentrations of minor and major oxides also loss of ignition component for CEM-I type of cement samples. PLS regression models with high predictability were not obtained excepts for SiO_2, Fe_2O_3 components. The reason for that, CEM-I must have certain limits for oxides content; therefore, the studied concentration range is so low this situation decreases the model's predictive ability. To improve the model's predictability, the number of studied samples can be increased, or other cement types such as type II, III, IV can be added to PLS models to increase the concentration range of major and minor oxides. Due to the large number of CEM-I samples compared to other types of cement that we have easy to accessibility, a PLS study was also carried out for 56 CEM-I samples that a specific type of cement.

XRF measurements were accomplished using two-sample preparation techniques, pressed pellet, and glass bead, as the first scenario included samples of various types. Since all samples were Cem-1 in the second scenario, XRF measurements were carried out using only the pressed pellets sample preparation technique. In the glass bead preparation technique, the samples are melted to at high temperatures, which allows us to obtain more homogeneous analysis surfaces and eliminates the measurement errors that may arise from the particle size of the sample. Due to XRF measurement errors caused by the pressed pellet sample preparation technique in the second scenario, Constructed PLS models to estimate major oxides and minor oxides in CEM-I samples may not have well predictive power.

CHAPTER 6

CONCLUSION

In conclusion, the present study has demonstrated a new analytical method for the quantitative determination of composition for cement raw materials, intermediate products, and finished products by using Fourier Transform Infrared Spectroscopy coupled with Chemometrics multivariate calibration method as an alternative for the current XRF technique. The developed new method ensures that much faster than XRF because there is no need for additional sample preparation. Also, the developed new method provides much more safety than XRF because of the usage of radioactive sources in XRF. Furthermore, the developed new method has a lower instrumentation cost than other techniques used for the chemical composition of cement. In the study, two different scenarios are created. In the first scenario, the broadened samples from the raw materials to finished products were used for the construction of PLS models to estimate major and minor oxides and loss of ignition parameter in samples. The PLS models were created using raw FTIR-ATR spectra of the samples and were created by applying a proper pre-processing method that was decided as extended multiplicative scatter correction (EMSC). The higher quality PLS models have obtained with the EMSC-applied spectra thanks to the reduction of scattering caused by the particle size of the samples. The regression coefficients(R^2) of the PLS model were found to varies from 0.95 to 0.99, and the standard error of cross-validation (SECV) values were founded as from 0.21 to 1.42 (w/w%). In the second scenario, the PLS method was applied for only Cem-I samples to estimate the chemical composition of the samples. However, the Cem-I samples must have certain limits in terms of the concentration of each oxide. Therefore, the concentration range of the studied samples was found narrow. The high-quality PLS models were not obtained in the second scenario. If the addition of other types of cement (Cem-II, Cem-III, Cem-IV, CemV) to Cem-I samples, the concentration range would be more expansive. Perhaps, with all types of cement and with a more extensive sample set collected over an extended time to improve variability, the models could provide better predictive performance.

REFERENCES

Arslanhan, M. (2016). "Çimento Üretim Proseslerinde Eritiş Metoduyla Hazırlanmış Numunelerin WDXRF Analizleri ve Hata Etkilerinin İncelenmesi". Yüksek Lisans Tezi. Bursa Teknik Üniversitesi Fen Bilimleri Enstitüsü.Afseth, N. K., & Kohler, A. (2012). Extended multiplicative signal correction in vibrational spectroscopy, a tutorial. Chemometrics and Intelligent Laboratory Systems, 117, 92–99.

Attenuated Total Reflectance (ATR) | For FTIR Applications. (n.d.). Retrieved December 23,2021,from https://www.mt.com/tr/tr/home/products/L1_AutochemProducts/ReactIR/attenuated-total-reflectance-atr.html

Bağcıoğlu, M. (2011). Development of chemometric multivariate calibration models for spectroscopic quality analysis of biodiesel blends. İzmir Institute of Technology.

Brereton, R. G. (2000). Introduction to multivariate calibration in analytical chemistryElectronic Supplementary Information available. See http://www. rsc. org/suppdata/an/b0/b003805i. Analyst, 125(11), 2125–2154.

Choi, K.-K., Lam, L., & Luk, S.-F. (1994). Analysis of cement by atomic absorption spectrophotometry and volumetric method. Talanta, 41(1), 1–8.

De Schepper, M., De Buysser, K., Van Driessche, I., & De Belie, N. (2013). The regeneration of cement out of Completely Recyclable Concrete: Clinker production evaluation. Construction and Building Materials, 38, 1001–1009.

Dunstetter, F., De Noirfontaine, M.-N., & Courtial, M. (2006). Polymorphism of tricalcium silicate, the major compound of Portland cement clinker: 1. Structural data: review and unified analysis. cement and Concrete Research, 36(1), 39–53.

Ferrari, L., Kaufmann, J., Winnefeld, F., & Plank, J. (2012). Reaction of clinker surfaces investigated with atomic force microscopy. Construction and Building Materials, 35, 92–96. https://doi.org/10.1016/j.conbuildmat.2012.02.089

Gastaldi, D., Canonico, F., Irico, S., Pellerej, D., & Paganini, M. C. (2010). Near-infrared spectroscopy investigation on the hydration degree of a cement paste. Journal of Materials Science, 45(12), 3169–3174.

Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. Analytica Chimica Acta, 185, 1–17.

Haaland, D. M., & Thomas, E. V. (1988). Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. Analytical Chemistry, 60(11), 1193–1202.

Hughes, T. L., Methven, C. M., Jones, T. G. J., Pelham, S. E., Fletcher, P., & Hall, C. (1995). Determining cement composition by Fourier transform infrared spectroscopy. Advanced cement Based Materials, 2(3), 91–104.

Kapkaç, F. (2013). Çimento Çeşitleri, Özellikleri, Hammaddeleri ve Üretim Aşamaları. MTA Doğal Kaynaklar ve Ekonomi Bülteni, 223.

Khelifi, S., Ayari, F., Tiss, H., & Chehimi, D. B. H. (2017). X-ray fluorescence analysis of Portland cement and clinker for major and trace elements: accuracy and precision. Journal of the Australian Ceramic Society, 53(2), 743–749.

Kowalski, B., Gerlach, R., & Wold, H. (1982). Chemical systems under indirect observation. Systems under Indirect Observation, 191–209.

Lindgren, Fredrick, Geladi, P., Rännar, S., & Wold, S. (1994). Interactive variable selection (IVS) for PLS. Part 1: Theory and algorithms. Journal of Chemometrics, 8(5), 349–363.

Lindgren, Fredrik, Geladi, P., Berglund, A., Sjöström, M., & Wold, S. (1995). Interactive variable selection (IVS) for PLS. Part II: Chemical applications. Journal of Chemometrics, 9(5), 331–342.

Marjanovic, L., McCrindle, R. I., Botha, B. M., & Potgieter, J. H. (2000). Analysis of cement by inductively coupled plasma optical emission spectrometry using slurry nebulization. Journal of Analytical Atomic Spectrometry, 15(8), 983–985.

Martens, H., Nielsen, J. P., & Engelsen, S. B. (2003). Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures. Analytical Chemistry, 75(3), 394–404.

Nasrazadani, S., & Springfield, T. (2014). Application of Fourier transform infrared spectroscopy in cement Alkali quantification. Materials and Structures, 47(10),

1607–1615.

Neville, A. M. (1995). Properties of concrete (Vol. 4): Longman London.

Otto, M., & Wegscheider, W. (1985). Spectrophotometric multicomponent analysis applied to trace metal determinations. Analytical Chemistry, 57(1), 63–69.

Rebouças, J. P., Rohwedder, J. J. R., & Pasquini, C. (2018). Near infrared emission spectroscopy for rapid compositional analysis of Portland cements. Analytica Chimica Acta, 1024, 136–144.

Schramm, R. (2016). Use of X-ray fluorescence analysis for the determination of rare earth elements. Physical Sciences Reviews, 1(9).

Skoog, D. A., Holler, F. J., & Crouch, S. R. (2017). Principles of instrumental analysis. Cengage learning.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society: Series B (Methodological), 36(2), 111–133.

Suart, B. (2004). Infrared spectroscopy: Fundamental and applications. Google Scholar.

TAYLOR, H. F. W. (1964). The Chemistry of cements. Edited by HFW Taylor. Academic press.

Teknik Bilgiler - Batıçim. (n.d.). Retrieved May 14, 2021, from https://www.baticim.com.tr/cimento-hakkinda-bilgiler/cimento-nedir

Titus, D., Samuel, E. J. J., & Roopan, S. M. (2019). Nanoparticle characterization techniques. In Green Synthesis, Characterization and Applications of Nanoparticles (pp. 303–319). Elsevier.

Türkiye Çimento Sanayicileri Birliği. (n.d.-a). Retrieved May 14, 2021, from https://www.turkcimento.org.tr/tr/cimento_nasil_uretilir

Türkiye Çimento Sanayicileri Birliği. (n.d.-b). Retrieved May 14, 2021, from https://www.turkcimento.org.tr/tr/istatistikler/aylik-veriler

Türkiye Çimento Sanayicileri Birliği. (n.d.-c). Retrieved June 3, 2021, from https://www.turkcimento.org.tr/en/statistics/capacity

Türkiye Çimento Sektörü 100 milyon ton üretim kapasitesine ulaştı. (n.d.). Retrieved

May 14, 2021, from https://www.yapimalzeme.com.tr/turkiye-cimento-sektoru-100-milyon-ton-uretim-kapasitesine-ulasti/

Wold, S., Albano, C., Dunn III, W. J., Esbensen, K., Hellberg, S., Johansson, E., & Sjöström, M. (1983). Pattern recognition: finding and using regularities in multivariate data. Food Research and Data Analysis, 3, 183–185.

X-ray Fluorescence (XRF) | Protocol. (n.d.). Retrieved May 15, 2021, from https://www.jove.com/v/5498/x-ray-fluorescence-xrf

Zaini, N., Van der Meer, F., Van Ruitenbeek, F., De Smeth, B., Amri, F., & Lievens, C. (2016). An alternative quality control technique for mineral chemistry analysis of Portland cement-grade limestone using shortwave infrared spectroscopy. Remote Sensing, 8(11), 950.