



UNIVERSITY OF NAIROBI

**DIRECT DETERMINATION OF MERCURY IN SKIN LIGHTENERS USING
CHEMOMETRIC ASSISTED ENERGY DISPERSIVE X-RAY FLUORESCENCE
SPECTROSCOPY**

BY

KIILU JULIUS MUSAU

B.Ed. (Hons)

I56/80960/2015

A Thesis Submitted in Partial Fulfillment of the Requirements for the Award of the
Degree of Master of Science in Physics of the University of Nairobi.

© August, 2022

Declaration

I declare that this thesis is my original work and has not been submitted elsewhere for examination, award of a degree or publication. Where other people's work or my own work has been used, this has properly been acknowledged and referenced in accordance with the University of Nairobi's requirements.

Kiilu Julius Musau

I56/80960/2015

Department of Physics

University of Nairobi

Signature.....

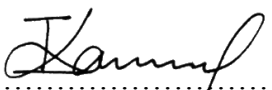
Date: 10th August 2022

This thesis has been submitted for examination with our approval as research supervisors:

Dr. M. I. Kaniu

Department of Physics

University of Nairobi

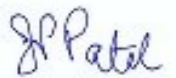
Signature.....

Date: 11th August 2022

Prof. J. P. Patel

Department of Physics and Space Sciences

Technical University of Kenya

Signature.....

Date: 11th August 2022

Acknowledgements

Firstly, my overwhelming gratitude goes to the Almighty God for giving me strength and grace to accomplish this noble task. My heartfelt appreciation is to my first supervisor, Dr. M. I. Kaniu for his tireless encouragement, support and commitment towards ensuring that I realize my long awaited and farfetched dream. Many thanks also to Prof. J. P. Patel for supervision and mentorship. Dr. J. M. Wanjohi of the Department of Chemistry cannot go unmentioned for the role he played in helping me identify the right base matrix for the simulate samples, and Dr. A. Dehayem-Massop for providing the real samples used in this work. A big thumps up to Mr. J. J. Okonda of the Department of Physics for being my instructor and mentor in the use of the EDXRF spectrometer for spectral data analysis of the samples. The success of this work could not have been realized without Mr. B. M. Muthoka, Department of Physics who came in handy when I was in dire need of a micropipette and other materials including some apparatus. Much appreciation to Mr. E. Mwangi, senior technologist, Department of Chemistry for his technical advice, assistance and support with apparatus during the simulate sample preparation. Lots of appreciation to my fellow M.Sc. students: Nancy M. David for her moral support and induction in spectral data preparation, C. N. Ndung'u who came in handy whenever I was in predicament especially during spectral data analysis with the R code; R. Ikedi can never miss in my list for his assistance in the PCA analysis. Lastly, special thanks to my beloved and dear wife for her unwavering moral support, encouragement and prayers.

Abstract

The use of cosmetic products for skin colour lightening or bleaching is a common practice across the globe. However human health effects associated with mercury in these products such as skin cancer and kidney failure have been reported. Conventional techniques used in the analysis of mercury such as atomic absorption spectroscopy and mass spectroscopy are destructive, time consuming and expensive. Although the energy dispersive X-ray fluorescence (EDXRF) spectroscopy method is rapid, non-destructive and requires minimal or no sample preparations, it has a high detection limit for mercury and therefore quantification of trace levels of mercury below 1ppm is challenging. This is mainly due to spectral overlaps, weak mercury fluorescence signals and extreme matrix effects. In this work, a novel chemometrics-assisted EDXRF spectroscopy method was utilized to realize rapid, direct detection and quantification of both low (< 1 ppm) and high mercury levels in skin whitening creams and lotions. 50 simulate (mixture of distilled water and pure glycerol) samples spiked with mercury concentrations ranging from a blank sample to 500 ppm were used for method development. The samples were then analyzed in triplicates for 900 seconds using the NeX EDXRF spectrometer set-up. Two chemometric techniques namely, principal component analysis (PCA) and artificial neural networks (ANNs) were used to perform exploratory analysis of the measured EDXRF spectra and quantification of the mercury levels. From PCA, it was found that the spectral data forms distinct clusters for both the low and high mercury concentration levels in the 9.6-10.4 keV Hg L_{α} and 11.2-12.4 keV Hg L_{β} regions. Two ANN Hg concentration models were developed, one for ppb concentrations (0-1000 ppb; 30 samples) and the other for ppm concentrations (0-500 ppm; 17 samples). The R^2 , RMSEP, LOD and LOQ values for the two models were 0.72, 20.6%, 527 ppb, 819 ppb and 0.98, 4.6%, 3 ppm and 11 ppm respectively. The ppm model was used to ascertain sample results acquired by utilizing the conventional EDXRF method, most of which were closely matching while the ppb model established that two of the real samples registered a Hg content equal to 731 ± 151 ppb. It may therefore be concluded that the chemometrics – EDXRF approach has potential for rapid, non-destructive and trace quantitative analysis of mercury compared to traditional measurement approaches. The technique is recommended for quality control and assurance of consumer products by the relevant regulatory authorities such as the Kenya Bureau of standards.

TABLE OF CONTENT

Declaration.....	i
Acknowledgements	ii
Abstract.....	iii
TABLE OF CONTENT.....	iv
LIST OF FIGURES	vii
LIST OF TABLES	viii
List of Abbreviations	ix
CHAPTER ONE	1
INTRODUCTION.....	1
1.1 Background to the Study.....	1
1.2 Statement of the Problem	3
1.3 Objectives of the study.....	3
1.3.1 General objective:	3
1.3.2 Specific objectives:	3
1.4 Justification and Significance of the Study.....	4
CHAPTER TWO	5
LITERATURE REVIEW	5
2.1 Overview	5
2.2 Assessment of mercury levels in skin lighteners around the world.....	5
2.3 Spectroscopy techniques used in the analysis of skin lighteners.....	7
2.3.1 A Survey of conventional spectroscopy methods	7
2.3.2 X-ray fluorescence spectroscopy.....	9
2.4 Utility of chemometrics in analytical spectroscopy.....	10
2.4.1 Exploratory analysis of spectral data.....	11
2.4.2 Preprocessing of spectral data	11
2.4.3 Quantification modeling of Spectral data	12
CHAPTER THREE	14
THEORETICAL BACKGROUND	14
3.1 Overview	14
3.2 The Sim Base Matrix Identification	14
3.3 Utility of chemometrics in EDXRF spectroscopy.....	14
3.3.1 Principal Component Analysis (PCA).....	15
3.3.2 Artificial Neural Networks (ANNs)	16

3.4 The X-Ray Fluorescence (XRF) Spectrometer	21
3.5 Energy Dispersive X-Ray Fluorescence Spectrometry	22
3.6 Interaction of X-rays with matter	24
3.6.1 Radiation scattering	25
CHAPTER FOUR.....	28
MATERIALS AND METHODS	28
4.1 Overview	28
4.2 Instrumentation of EDXRF Spectroscopy	28
4.3 Base Matrix candidate identification	29
4.3.1 Spectra analysis for candidate skin lightening base matrices	31
4.4 Simulate samples preparation and analysis	33
4.5 Sample volume for analysis	36
4.6 Real Samples Preparation and Analysis	36
4.7 Multivariate Chemometric analysis of EDXRF spectra	37
4.7.1 Data Preprocessing	37
4.7.2 Spectra Analysis by PCA	38
4.7.3 Training ANN Models for Hg concentrations in ppb and ppm.	38
CHAPTER FIVE	40
RESULTS AND DISCUSSION	40
5.1 Overview	40
5.2 EDXRF Spectra Analysis Results for Simulate Samples	40
5.2.1 Hg ROI Results for simulate samples	40
5.3 Results of EDXRF Spectra Data Analysis and Modelling Using Multivariate Chemometric Techniques	42
5.3.1 PCA Results for Hg concentrations (ppb)	42
5.3.2 The ANN modeling results for the ppb Data	43
5.3.3 PCA analysis Results for Hg concentrations (ppm)	46
5.3.4 The ANN modeling results for the ppm dataset	47
5.3.5 ANN Model Results for Real Samples	49
5.4 EDXRF Results for Simulate and Real Samples	50
5.4.1 Analysis results of the simulate samples	50
5.4.2 Analysis results of real samples (pastes)	50
5.4.3 Analysis results of real samples (liquids)	51
5.5 Comparison Between Conventional EDXRF and ANN model Results	53
CHAPTER SIX	55

CONCLUSIONS AND RECOMMENDATIONS.....	55
6.1 Conclusions.....	55
6.2 Recommendations.....	56
REFERENCES.....	57
APPENDICES.....	62
Appendix 1: Simulate sample preparation pictograms.....	62
Appendix 2: Real samples.....	63
Appendix 3: R-Scripts for the Multivariate Analysis Techniques.....	64
R-Script for PCA analysis of Hg concentrations (ppb).....	64
Training Algorithm for the ANN Model with Hg concentrations (ppb).....	67
R Script for PCA analysis of Hg Concentrations (ppm).....	72
Training Algorithm for ANN model with Hg concentrations (ppm).....	75
R script for calculating the LoD and LoQ for the ANN model with Hg concentrations (ppb)..	78

LIST OF FIGURES

Fig. 3.1: Flow chart showing modeling steps.....	18
Fig. 3.2: Model architecture.....	19
Fig. 4.1: EDXRF Schematic Instrumentation.....	30
Fig. 4.2: The base matrix samples.....	32
Fig. 4.3: Glycerol structure.....	33
Fig. 4.4: EDXRF scatter plot for the base matrices.....	33
Fig. 4.5: PCA analysis results for candidate matrices.....	34
Fig. 4.6: Spectral data acquisition time.....	37
Fig. 4.7: Curve for optimum sample volume.....	38
Fig. 5.1: Spectra plot for simulate samples.....	42
Fig. 5.2: Plot for spectral intensity Vs Hg concentration (ppm).....	43
Fig. 5.3: PC scores plot for Hg concentration (ppb).....	44
Fig. 5.4: The loadings plot for Hg concentrations (ppb).....	45
Fig. 5.5: Trained model architecture used in this work.....	46
Fig. 5.6: The ANN model used for quantification of Hg concentrations (ppb).....	46
Fig. 5.7: PC scores plot for Hg concentrations (ppm).....	48
Fig. 5.8: The loadings plot for Hg concentrations (ppm).....	48
Fig. 5.9: Trained model architecture for the ppm concentrations.....	49
Fig. 5.10: The ANN model used for quantification of Hg concentrations (ppm).....	50
Fig. 5.11: PC scores plot for real samples.....	53
Fig. 5.12: Loadings plot for real samples.....	54

LIST OF TABLES

Table 2.1: Conventional techniques for mercury identification and quantification.....	9
Table 4.1: Simulate sample concentrations	36
Table 4.2: Classed Hg concentrations (ppb).....	39
Table 4.3: Classed Hg concentrations (ppm).....	40
Table 5.1: EDXRF Simulate samples results (ppm).....	47
Table 5.2: Sample results by using the developed ANN model for Hg concentrations (ppb).....	47
Table 5.3: ANN model statistics for Hg concentrations (ppb).....	50
Table 5.4: ANN model statistics for Hg concentrations (ppm).....	51
Table 5.5: Sample results by using the developed ANN model for Hg concentrations (ppm)...	51
Table 5.6: EDXRF paste samples results.....	52
Table5.7: Real samples analysis results by using the ANN model for Hg conc. (ppb).....	53
Table 5.8: Comparison between results by conventional EDXRF and ANN model for sim samples.....	55
Table 5.9: Comparison between conventional EDXRF and ANN model results for real samples.....	56

List of Abbreviations

AAS	Atomic Absorption Spectroscopy
AFS	Atomic Fluorescence Spectroscopy
ANNs	Artificial Neural Networks
CV AAS	Cold Vapour Atomic Absorption Spectroscopy
EDA	Exploratory Data Analysis
EDXRF	Energy Dispersive X-Ray Fluorescence
EXAFS	Extended X-ray Absorption Fine Structure
FIAS	Flow Injection Atomic Spectroscopy
GEXRF	Grazing Emission X-Ray Fluorescence
GI XRF	Grazing Incident X-Ray Fluorescence
ICP OES	Inductively Coupled Plasma Optical Emission Spectroscopy
ICP MS	Inductively Coupled Plasma Mass Spectroscopy
LCLS	Linac Coherent Light Sources
PCA	Principle Component Analysis
PLSR	Partial Least Squares Regression
SVMs	Support Vector Machines
SDD	Silicon Drifted Detector
TXRF	Total Reflection X-Ray Fluorescence
WDXRF	Wavelength dispersive X-Ray Fluorescence
WHO	World Health Organization
XRFEL	X-Ray Free Electron Lasers
XANES	X-Ray Absorption Near Edge Spectroscopy
XRF	X-Ray Fluorescence
LOD	Limit of detection
LOQ	Limit of quantification
LLD	Lower limit of detection
RMSEP	Root mean square error of prediction
R^2	Square of the regression value

CHAPTER ONE

INTRODUCTION

1.1 Background to the Study

The use of skin lightening products in the African continent is common because a lot of the women crave for a fair complexion free from spots, specks and imperfection, which is associated with beauty and youthfulness (Kamakshi, 2011). Most of these skin lighteners contain mercury as one of the ingredients but mostly with unspecified quantification and labeling (Al-Saleh, 2016). The presence of mercury in skin lighteners has been found to suppress the enzyme that produces melanin in the human body (Murphy *et al.*, 2009). Thus, prolonged and extensive exposure to mercury poses a health risk to consumers and has been found to cause skin disorders, and in some cases infection to the brain, nervous system and kidneys (Mahé *et al.*, 2007).

A WHO publication on mercury in skin lighteners showed that mercury concentrations ranged from 1% to 10%, while other beauty products such as facial lightening creams contained concentrations of up to 33%, without the companies selling the products listing mercury as one of the ingredients (Campbell *et al.*, 2003; Dlova *et al.*, 2014). The CV AAS method has been used to detect and quantify mercury with a detection limit of a few ppt (Višnjevec *et al.*, 2014). However, positive interference of spectral signals remains a challenge to this technique. ICP MS having an LOD of up to 1 ppt has also been used for analysis of mercury but requires special sample preparation and is also expensive (Bailey *et al.*, 2003). ICP OES has also been used to detect and quantify mercury in skin lightening lotions (Nguyen *et al.*, 1998). However, the method is destructive, has poor detection limit and low sensitivity (E Schnauer *et al.*, 1989).

The XRF spectroscopy method in its various modalities i.e. the EDXRF and the WDXRF is a rapid, non-destructive method with minimal or without preparation of sample and can be applied in a wide ranging concentration of elements (solid, powder and liquid). In XRF spectrometry, the spectral lines are used to identify the elements. Such lines, grouped in the series K, L, M, N, etc are the characteristic line series for every element (Antwerpen *et al.*, 2006). Light elements produce K line series, middle-ranging elements give rise to K as well as L lines, with the heavy ones emitting K, L and M line series (Bote *et al.*, 2009).

The choice of an analysis line depends on the sample type, elements present in the sample, range of elemental concentration and the conditions for excitation. The L shell emits an electron which fills a vacancy in the K shell and thereby producing a K_{α} line radiation. When an electron from the M shell fills the same vacancy in the K shell, a K_{β} line radiation is obtained, while K_{γ} line radiation is given out when an electron from the N shell fills the vacancy.

X-ray fluorescence is capable of analyzing quantitatively multi-elemental composition of samples in which the detection limits are in ppb range, depending on the sample form and spectrometer excitation conditions such as tube conditions and strength of the radioisotope source. During analysis, high energy photons strike the target material, exciting electrons in the core levels of atoms in the material (Shibata *et al.*, 2009). This causes de-excitation through characteristic fluorescent radiation whose energy is used for identification based on intensity and elemental concentration in the sample. In XRF spectroscopy, incident X-ray photons produce scatter radiation when they interact with electrons in atoms of the target element (Wobrauschek *et al.*, 2010). This radiation is either Rayleigh (coherent) scattering or Compton (incoherent) scattering. Coherent scattering occurs when energy is conserved during collision between the incident beam and sample while incoherent scattering is produced when some energy is lost by the scattered photons (Marguí *et al.*, 2009).

However, detection limit, spectral overlaps produced by the L and M series as a result of inter- element effects (matrix effects) and weak fluorescence signals for light elements with atomic number below sodium (i.e. $Z < 11$) still remain a challenge in this technique (Nguyen *et al.*, 1998).

Spectral complexity caused by matrix effects in the XRF analysis of complex samples make spectrum evaluation difficult as well as deconvolution of resultant fluorescence intensities into respective concentrations (Wobrauschek *et al.*, 2010). In classical EDXRF, the concentration of all elements in the sample must be known in order to deal with matrix effects challenge. The use of chemometrics-assisted EDXRF spectroscopy method in this work overcame the challenges and attained a direct detection and quantification of trace level mercury in varying sample concentrations. Chemometric techniques such as PCA and ANNs are robust analytical tools which utilize mathematical, statistical and computational methods to reveal hidden relationships in data sets (Luo, 2006; Reinholds *et al.*, 2015; Worley *et al.*, 2013).

1.2 Statement of the Problem

Prolonged exposure to unregulated and unquantified mercury levels in creams and lotions used for skin lightening poses a health hazard to the product consumers. There is therefore need to perform quality assurance tests of the skin lighteners in order to ascertain that the mercury content in them does not exceed the World Health Organization limit of 1 ppm, (Bose-O'Reilly *et al.*, 2010). Current spectrometric techniques that have been used in the detection and quantification of mercury include AAS, AFS and ICP- MS. They have had limitations in that they involve wet – chemistry, are destructive to the samples, have poor limit of detection and high cost (von Burg, 1995). EDXRF spectroscopy which is rapid and nondestructive to the samples has also been used, though challenged by spectral overlaps associated with K_{α} and L_{β} lines of trace level elements as well as a LLD (Melquiades *et al.*, 2015), (Antwerpen *et al.*, 2006). In this research a chemometric assisted EDXRF technique will be used to overcome the aforementioned challenges and also address the problem of unspecified and unquantified mercury levels in skin lighteners.

1.3 Objectives of the study

1.3.1 General objective:

To perform rapid detection and quantification of mercury levels in skin lighteners utilizing a chemometric-assisted EDXRF spectroscopy approach.

1.3.2 Specific objectives:

- (i) Identify an appropriate skin lightening base matrix material and acquire EDXRF spectra from simulate samples spiked with a wide range of mercury concentrations.
- (ii) Develop a multivariate chemometric calibration model for mercury quantification using the EDXRF spectra obtained from specific objective (i) above.
- (iii) Perform the analysis of a wide variety of skin lighteners obtained from the local markets using the chemometric-assisted EDXRF technique to determine their mercury content.
- (iv) Compare the predicted mercury concentrations in both the simulate and real samples using the developed model with those obtained using conventional EDXRF.

1.4 Justification and Significance of the Study

Negative effects caused by the wide spread use of skin lightening products containing unspecified and unquantified mercury levels has raised a global health concern (Hamann *et al.*, 2014). Poor quality control and lack of proper labeling of the contents in these products has been the major challenge. Although a number of spectrometric techniques including EDXRF spectroscopy have been used in the determination of mercury, rapid and direct analysis of low concentrations i.e. less than 0.1 ppm is still a challenge (Orisakwe *et al.*, 2013). In the case of EDXRF, the spectral overlap of the Hg energies (9.6 – 12.4 keV) and other trace elements such as Pb and As, complicate the quantification of low mercury levels.

This research work utilizes chemometric techniques namely, the principal component analysis (PCA) and artificial neural networks (ANNs) to perform EDXRF spectral deconvolution and exploratory data analysis in the model development for rapid, direct detection and quantification of mercury levels of both high (> 3 ppm) and low (< 1 ppm) mercury levels in skin lighteners. The technique has also overcome the challenge of extreme matrix effects and spectral overlaps associated with the conventional EDXRF spectroscopy.

CHAPTER TWO

LITERATURE REVIEW

2.1 Overview

The chapter outlines four areas which entail the backbone of this research work. Section 2.2 unearths prevalence of mercury toxicity in the world while section 2.3 discusses some of the spectrometric techniques that have been used previously to analyze skin lighteners. XRF spectrometer and analytical chemometric spectroscopy are tackled in section 2.4 and 2.5 respectively. Quantification modeling of spectral data is finally handled in section 2.6.

2.2 Assessment of mercury levels in skin lighteners around the world

Elemental mercury, a heavy silvery odorless and volatile element at room temperature occurs naturally in uncombined state but can also be obtained mercuric sulfide ore. Its relative atomic mass (RAM) is 200.59 g while its atomic number $Z=80$ and has different solubility rates in water at different temperatures in its various forms. Mercury toxicity is the human poisoning which depends on the Hg form, exposure rate and the dosage (Langford and Ferner, 1999). When absorbed into the body, mercurous and mercuric salts damage the gut lining and kidney while methylmercury affects the entire body. Against the health risks posed by mercury exposure, a number of investigations to ascertain the mercury concentrations in skin lightening creams and lotions as well as other beauty products have been carried globally.

Several studies in Cambodia have found that skin lighteners used contain mercury concentrations which exceed 2% (Murphy *et al.*, 2009, 2013). Nine out of the samples investigated contained mercury concentrations beyond the recommended 1 ppm limit, while five others had concentrations of more than 2%. Those with the label “For export only” had much higher mercury concentrations. The same research indicated that 41 other skin whiteners were investigated in 2008 by using the same technique and revealed that 11 out of the 41 samples analyzed contained more than 2% of mercury (Murphy *et al.*, 2009). Furthermore, labeling of the products ranged from detailed to slight, with some without instructions.

In America, 45% of skin lighteners collected from 32 countries were analyzed and found and found to contain mercury concentrations exceeding 10% (Peltzer *et al.*, 2016), while in Europe, 13.8% of the skin lighteners contain mercury concentrations ranging from 30 µg/g (Peltzer *et al.*, 2016). In Africa, it was established that 100% of all the investigated brands in Ghana contained mercury concentrations from 0.01µg/g to 0.549µg/g (Amponsah *et al.*, 2014), while 25% of such products in the same country had concentrations ranging from 0.02µg/g to 25.7µg/g (Fang, 1995). In Sudan, Ahmed and Hamid (2016) confirmed that 52% of the population used the skin lighteners, while in Kenya, 70% of the skin lightening creams and lotions (70%) were found to contain mercury concentrations ranged 3.7% to 121% (Maina, 1997).

In a another study conducted by Peltzer *et al.*, (2016) investigating skin lightening samples originating from 26 countries found that China, Thailand, Vietnam and Mexico contained mercury concentrations ranging 0.01ppm to 1259%, 0.01µg/g to 8,578 ppm, 0.02 ppm to 355%, 878 ppm to 3,600 %, respectively. In Nepal, beauty products contained mercury concentrations of up to 0.112 ppm (Sah and Charitra, 2012), which is much lower compared to other Asian countries. Despite a European Union ban on skin lighteners Maneli *et al.*, (2016) showed that a third of the tested skin lightening products originated from Europe, with 40 percent of the samples containing mercury concentrations > 30 ppm . However, none of the investigated products indicated the detected Hg content as one of the ingredients. Additionally, a pilot survey of mercury in drugs, cosmetics and household products revealed that the cosmetics had high mercury concentrations with a large number of them being manufactured in developed countries and distributed worldwide via the black market routes, making it hard to monitor and control the import and sale of the products (Liang *et al.*, 2013).

From the surveyed literature, it is clear that skin lighteners can have extreme mercury levels, much higher than the US-FDA limit of 1 ppm. Continued use and long exposure to these products is unsafe due to mercury accumulation in the kidney and liver. Therefore, assessment of mercury and other toxic elements such as arsenic and cadmium in skin lighteners and other beauty products is necessary and should be done regularly and intensified in especially African countries where the practice is rampant. For this to be effective, rapid and accurate methods that are portable are required to assist regulatory authorities in the inspections (i.e. quality assurance and control) of these products.

2.3 Spectroscopy techniques used in the analysis of skin lighteners

2.3.1 A Survey of conventional spectroscopy methods

The use of skin lighteners containing unquantified and unregulated levels of mercury as one of the ingredients has raised a global concern due to the health complications associated with prolonged exposure to mercury (Chan, 2011; Murphy *et al.*, 2015). A number of spectrometric techniques have been used to assess the extent of the human health effects in regard to the wide spread and the prolonged use of these products. Table 2.1 highlights the conventional techniques, as well as their advantages and limitations.

Sah and Charitra (2012) used ICP-MS to investigate the health challenge associated with mercury in skin whiteners. It was found that the advantage is that the technique can be used for very fine metal detection ranging from lithium to uranium. However the limitations are that it requires special sample preparation and it is more expensive compared to the other techniques (Hattendorf and Günther, 2000). Voegborlo, *et al.*, (2014) used a CV AAS with an automatic mercury analyzer to determine the mercury levels as well as the hydroquinone contained in some skin lightening creams. The advantage of CV AAS is that it requires little operator interaction, offers detection limit of a few parts per trillion, fast, has simplicity and robustness. However, positive interference of spectral intensities and low sensitivity remain a challenge in this technique (Nguyen *et al.*, 1998).

Table 2.1: Conventional techniques for mercury identification and quantification.

TECHNIQUE	ADVANTAGES	LIMITATIONS	REFERENCE
CV AAS	<ul style="list-style-type: none"> -Offers detection limit of a few parts per trillion -Requires little operator interaction -Has robustness 	<ul style="list-style-type: none"> -Positive interference is a problem 	(Jacimovic and Horvat, 2004)
XRF	<ul style="list-style-type: none"> -Requires limited solid sample preparation -Has low running costs -Easily portable 	<ul style="list-style-type: none"> -Has low detection limit -The analyzer and substrates are affected by humidity or temperature -Has interferences caused by elements other than the metal of interest 	(Kalnicky and Singhvi, 2001)
ICP MS	<ul style="list-style-type: none"> -Has detection limit of up to 1ppt -Used for very fine metal detection ranging from Lithium to Uranium 	<ul style="list-style-type: none"> -Require special sample preparation -More expensive compared to other techniques -Has poor sensitivity 	(Bailey <i>et al.</i> , 2003)
ICP OES	<ul style="list-style-type: none"> -Has a variety of sample introduction methods -Has a high atomizing temperature ranging from 6000K to 10,000K 	<ul style="list-style-type: none"> -Detects just one emission frequency at a time -Has poor detection limits -Has low sensitivity 	(Eschnauer <i>et al.</i> , 1989)

2.3.2 X-ray fluorescence spectroscopy

The XRF spectroscopy method offers good precision and accuracy and can be applied over a wide range of elements (Na to U) (Brouwer, 2006; Revenko, 2018). In addition, it is an atomic spectroscopy technique where the type and content of the elements in a variety of samples (liquid, solid and gas) can be fingerprinted and determined by recording and measuring the energy and intensity of the generated characteristic X-ray fluorescence due to the irradiation of the sample under study by a primary X-ray source.

However, XRF spectra have spectral overlaps, weak fluorescence signals and extreme matrix effects (Jacimovic and Horvat, 2004; Pessanha *et al.*, 2009; Rousseau, 2006). The invention of modern ways of detection, optics as well as X-rays sources has revolutionized EDXRF spectroscopy with respect to their accurate measurements and precision. An example is the micro-x-ray fluorescence based on single or polycapillaries which are optical elements. Internal reflection of x-rays takes place in a tube with thin walls, resulting in a focused beam of high intensity (Wobrauschek *et al.*, 2010). Excitation geometries such as GIXRF, TXRF, GIXRF and GEXRF are commonly applied in layer and impact characterization (Grinyer *et al.*, 2007; Wobrauschek *et al.*, 2010a). Other sources are XRFEL, LCL, XANES utilized in identifying the elemental state in terms of its chemical composition and EXAFS useful in the establishment of the number of coordination as well as the atomic distance in the neighborhood (Mahé, 2014).

During the analysis of samples by the EDXRF spectroscopy for example, both full and partial spectral peaks are displayed. The full ones represent the elemental concentration and composition of the samples while the partial ones give the concentration of a particular element (Bennun *et al.*, 2002). The various peak heights or intensities comprising the full spectra represent the respective elements contained in the samples (qualitative analysis) while the peak areas give the elemental concentrations in the analyte (quantitative analysis). Given that each element depicts a unique characteristic spectral intensity (signatures), it is therefore easy to identify the element of interest from a given sample (Kamagaju *et al.*, 2016). The count rate in background and peak radiations is taken into account due to the fact that the peak varies directly with time while background counts increase with time. MDLs are lowered as analysis time is increased. For laboratory instruments, analysis time per sample ranges from 200 seconds to 1000 seconds in direct correlation with the improvement on the MDLs (Wobrauschek *et al.*, 2010). Thus,

improvement on MDLs for the EDXRF is achieved by increasing the irradiation time per sample particularly for low concentration samples of the element of interest.

The method has been used widely in the analysis of mercury in skin lighteners (Hamann *et al.*, 2014; Murphy *et al.*, 2015); in both studies, a handheld XRF analyzer was used to realize the analysis of bulk samples (549 and 676, respectively).

This indicates the applicability of the method in rapid assessments of skin lighteners, especially in its portable form. However, the detection limit of the EDXRF technique has been reported to be 3 ppm (Maina, 1997), making it difficult for the detection of mercury in samples with lower concentrations such as those with concentrations ranging 1 – 2.9 ppm (> 1 ppm regulatory limit).

2.4 Utility of chemometrics in analytical spectroscopy

Chemometrics, a chemical method based on statistics and mathematics gives complete information through the analysis of acquired data and is suitable for multivariate analysis of data obtained from complex sample matrices. These methods are capable of extracting relevant chemical properties from spectral data and can be utilized for quick and reliable analyses (Luo, 2006). The main reason for using chemometrics in analytical spectroscopy is to reduce spectral noise, identify outliers and deal with spectral interferences to attain multivariate calibration (Einax, 2005; Nolan, 2005; Rutan, 1996). The common chemometrics models used in data analysis are PCA, PCR, ANNs, GA, SVMs and PLSR (Luo, 2006; Reinholds *et al.*, 2015). These models are mainly used for curve resolution, prediction, pattern recognition and image analysis.

In recent work based on extending the capabilities of the conventional EDXRF approach utilizing multivariate chemometric techniques, rapid and direct analysis of various complex matrices (i.e. soil, lubricating oils and human tissues) has been achieved with varied success (Angeyo *et al.*, 2012; Kaniu *et al.*, 2012; Okonda *et al.*, 2017; Sichangi *et al.*, 2018). In 2011 Kaniu *et al* used chemometrics assisted EDXRFS spectroscopy approach for direct and quick analysis of macro and micro soil nutrients otherwise known as soil quality indicators. In their method, they used the EDXRF spectrometry technique to investigate the x- ray scatter peaks which were non-invasively acquired from soils to develop a calibration strategy for quantitative analysis of SQIs in modeled clay soils.

The study utilized PCA for compression of spectral data and pattern recognition as well as PLSR and ANNs to build a calibration and quantitative analysis strategy. This approach enabled the XRF method to display its potential in spectral data analysis and assessment of chemical soil indicators in a fast and reliable way.

Angeyo *et al.*, (2012) dealt with one of the greatest challenges associated with XRF technique in direct trace analysis of complex matrices. Using PLS together with EDXRF spectrometry lubricating oils were analyzed rapidly (200 s).

The method managed to give quality assurance in the analysis of liquid samples with complex matrices given that the outcome for those containing heavy as well as low-Z metal ingredients was good and promising. Okonda *et al.*, (2017) used calibration strategy involving ANN model and PCR to directly and rapidly analyze samples containing trace bio metals in model soft tissues. This hybrid nested approach gave a better and reliable outcome in the determination of trace bio-metals.

Sichangi *et al.*, (2018) used a robust chemometrics assisted EDXRF spectroscopy approach to accurately determine in a direct and rapid way, the cancer biomarker trace metals contained in soft tissues of the body. This method overcame the challenge associated with the EDXRF spectroscopy technique in direct characterization and diagnosis cancerous tissues through the analysis of bio-metals. In the study, spectral data were preprocessed by using wavelet transform, independent component analysis and PCA before developing hybrid multivariate chemometric calibration models namely, ANN and PLS that were used to establish the bio-metals contained in thin body tissues.

2.4.1 Exploratory analysis of spectral data

Principal component analysis (PCA), one of the multivariate chemometric techniques gives a clear outline of spectral data in a reduced dimensional space by performing data dimension reduction, modeling, outlier detection, pattern recognition as well as calibration. It performs its tasks by grouping spectra using latent variables with the largest scores variance (Herve *et al*, 2018). The dataset for PCA comprise a matrix in which the rows represent channels for spectral energies while the columns give fluorescence corresponding to sample concentrations (Kaniu *et al.*, 2012).

2.4.2 Preprocessing of spectral data

It is the pretreatment of spectral data which involves mathematical manipulation of the data before it is analyzed. This is chiefly done in order to do away with unnecessary

variation and thus reduce the amount of data, prevent undesired effects brought about by varying scales in the data, change the data and make it useful for future analysis and finally keep adequate information in the data for attaining goals of interest. Data preprocessing is achieved via data reduction which involves smoothing, mean centering /auto scaling, data distribution change, normalization as well as data transformation (Kaniu *et al.*, 2011).

2.4.3 Quantification modeling of Spectral data

In addition, a number of neural network models of different types have been developed in recent qualitative and quantitative spectral analyses by using chemometric tools in XRF spectrometry such as PCA and ANNs (Büchele *et al.*, 2019). By imitating the biological neural network in the human brain ANNs can perform data calculation and at the same time carry out knowledge representation of the dataset. Common models include Hopfield networks, adaptive resonance theory (ART) networks, Kohonen networks and backward error propagation (BEP). Out of all, the mostly used is the backward error propagation algorithm, which has played a major role in utilization of neural networks to solve problems globally in real life situations (Luo, 2006). The backward error propagation model which is usually trained by making use of the supervised learning and considering the error-correction learning rule consists of three layers, namely the input, the hidden and the output layers.

In a neural net model, the transfer function usually used is a sigmoid function. In practice, during the multivariate calibrations, development of a reliable neural net structure has to be carried out, clearly outlining the hidden layers in terms of the number and size. The hidden layers, with their number and size are mainly useful in determining the model's properties and its ability to solve problems with nonlinearities (Gershenson, 2003). In fitting continuous functions, one hidden layer is sufficient whereas two of them are capable of addressing discontinuous functions (Büchele *et al.*, 2019). A neural net with limited hidden nodes is incapable of dealing with complicated systems due to its linear estimation while the one with a number of hidden nodes gives estimation by overfitting the results.

The problem is eliminated by reducing these nodes to a number equal to or less than one in excess of the input nodes. Finally, an artificial neural net model is usually developed by training it using three datasets, i.e. the training, the test and the validation data sets (Luo, 2006). The XRF technique (discussed in Section 2.3.2) has challenges of low resolution

and detection limits. EDXRF spectroscopy technique in particular has a challenge in the detection and quantification of mercury due to the line series emitted by other trace elements (Clevenger *et al.*,2017). Thus, mercury as a heavy element radiates line series L and M, which in some cases are found to overlap with similar spectra from arsenic and lead, making it difficult for the method to distinguish between the two elements. The present study seeks to overcome the challenges of low resolution, low detection limit (currently, 3 ppm for EDXRF) and spectral overlaps produced by the L and M series as a result of matrix effects by using a chemometrics-assisted EDXRF spectroscopy approach.

CHAPTER THREE

THEORETICAL BACKGROUND

3.1 Overview

In this chapter, the Sim Base Matrix identification in section 3.2 and Utility of Chemometrics in EDXRF Spectroscopy (Section 3.3) are discussed. Section 3.3 outlines Multivariate chemometric tools i.e PCA and ANN used in this work as well including their use in chemometric modeling of the EDXRF spectra. In addition, the XRF Spectrometer (Section 3.4), EDXRF Spectrometry (Section 3.5) and Interaction of X-ray with Matter (Section 3.6) are also tackled.

3.2 The Sim Base Matrix Identification

In a research on Transdermal Kinetics of a Mercurous Chloride beauty cream by *Palmer et al., 2000*, Cosmetic industry mainly utilizes glycerol-water mixture an aqueous base formulation for the products. Glycerol, a trihydroxy sugar alcohol with three carbon atoms and three hydroxyl groups makes it an organic polyol compound which is readily miscible with distilled water. The main role of the glycerol in the products is to disrupt lipid and protein structure by decreasing the barrier function of the stratum corneum resulting in increased skin permeability. Water, a commonly used ingredient in the base formulations enhances penetration on the skin and therefore increasing its permeability. Environmental conditions such as hydration among others, influences the water content in the skin, causing hydration.

3.3 Utility of chemometrics in EDXRF spectroscopy

Chemometrics is an analytical tool which utilizes mathematical, statistical and computational methods to reveal hidden relationships in data sets (*Reinholds et al., 2015*). In addition, it can also be used to obtain robust results from a large data set. The common chemometrics models used in data analysis are principle component analysis (PCA), Artificial Neural Networks (ANNs), Genetic Algorithms (GA), Support Vector Machines (SVMs) and partial least squares regression (PLSR) (*Luo, 2006*). These models are mainly used for curve resolution, prediction, pattern recognition and image analysis.

In EDXRF spectral analysis, PCA and ANNs as Multivariate chemometric tools for spectral data evaluation entailing data preprocessing, exploratory data analysis and supervised learning have been used to carry out data deconvolution, accurate prediction of new samples with high robustness and efficiency (Wold and Eriksson, 2001). Determination of linear relationships as well as nonlinear ones with good robustness is possible by the ANNs technique. (Nagata *et al*, 2006). The main purpose for multivariate calibration is to establish the relationships between a response y-variable and several x-variables.

3.3.1 Principal Component Analysis (PCA)

It is an unsupervised method used in exploratory data analysis (EDA) in which training data set is not required. PCA is a chemometric tool which gives a 2-D description of spectral data in a reduced multi-dimensional space by identifying outliers and reducing data dimension through extraction of the relevant information (Worley *et al.*, 2013). To transform a given data set, the equation below is used:

$$X = T.P' + E \quad (3.1)$$

where T is the scores matrix with n rows, which gives the relationship among the samples, calculated as (n x A). P' is the loading matrix with p columns, given as (A x p), where A is the hidden dimension for the principal components useful in complete description of the entire information contained in the given data.

During principal component analysis, the graphical interface (scores and loadings plots) and the 2-D scatter plot display covariance between the samples, providing a data overview and quantitative information from PC scores. Objects clusters and patterns including outliers are identified with ease in the PC scores plot enabling exploration of unexpected and expected trends in the data. The loadings line plots explain the essence of original variables in each PC which could be useful in the reduction of quantitative differences that lead to trends or clusters in the data and thereby identify elements that result in the spectra peaks. The PC scores give sample differences or similarities by outlining the patterns in terms of data structure. The sample location in terms of coordinates along every PC is given by scores for the corresponding samples. Interpretation of the information contained in a certain PC is explained by using the loadings plot while scores along the PC depicts the sample characteristics.

For example, scores closely positioned along the same PC describe sample similarity based on qualitative and quantitative information. On the other hand, loadings plot explains sample correlations in a data structure in such a way that every variable displays a loading on every PC. The variability of a variable over the data points as well as its contribution to a particular PC are both given by the loadings. Loadings range between -1 and +1 because geometrically, they are cosines of angles between the variables and corresponding PC. Large loadings are as a result of small angles, implying a strong relationship between the PC and variable.

3.3.2 Artificial Neural Networks (ANNs)

This is a chemometric model built by utilizing a given data set whose purpose is to predict the outcome of unknown samples. Before use, the ANN model is first of all trained, tested and validated by using an appropriate number of simulate samples (Luo, 2006). Training is the process by which a neural net model responds to the input information or its spectra and concentration values by modifying the weights and bias terms through unsupervised or supervised learning. In the unsupervised learning the model is unaware of the error it is bound to make due to the fact that it has no knowledge of the correctness of a given output for a particular input.

In supervised learning, the training process is controlled by an external agent which analyses the final output of the model and continues to change the weights in case it does not agree with the correct value (Reinholds *et al.*, 2015). To achieve a thoroughly trained model, tuning is done by adjusting the number of hidden layers as well as the nodes per layer. Validation dataset comprising new samples whose spectral data and concentration is already determined is used to assess the developed model's performance. However, the disadvantages are that it is time consuming due to its trial-and-error approach given that a nonlinear technique is applied to linear data.

A trained model, regarded as simulated structure or designed model is one with the least zero error, capable of giving solutions in cases where noise or nonlinearities are contained in a given data. Modeling steps include data preparation, model building, evaluation and prediction. In data preparation, the data is classified, loaded into the R software, preprocessed and then validation dataset created. The model is built by choosing the input layer, number of hidden layers including their respective nodes as well as the output layer (Jalali-Heravi and Kyani, 2004).

Evaluation involves assessing the accuracy in the predictions on the validation dataset by using the model. Accuracy, as used in multivariate calibration refers to the correlation between the measured and predicted values determined by the calibration model it is usually calculated as the root mean square error of prediction (RMSEP) using the equation

$$RMSEP = [\sum (y' - y_{reference})^2 / n]^{1/2} \quad (3.2)$$

where y' represents the predicted sample concentrations and n the number of calibration samples.

The modeling steps are outlined as shown in Fig. 3.1, while the neural net model architecture is shown in Fig. 3.2.

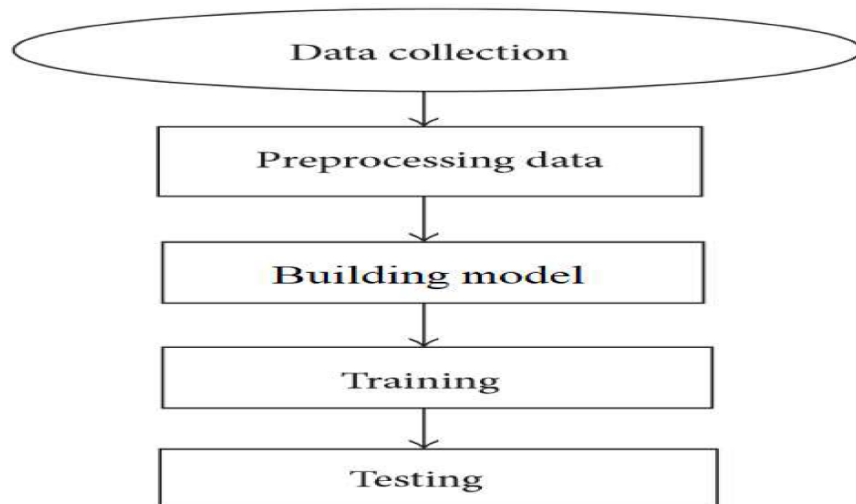


Figure 3.1: Flow chart showing modeling steps

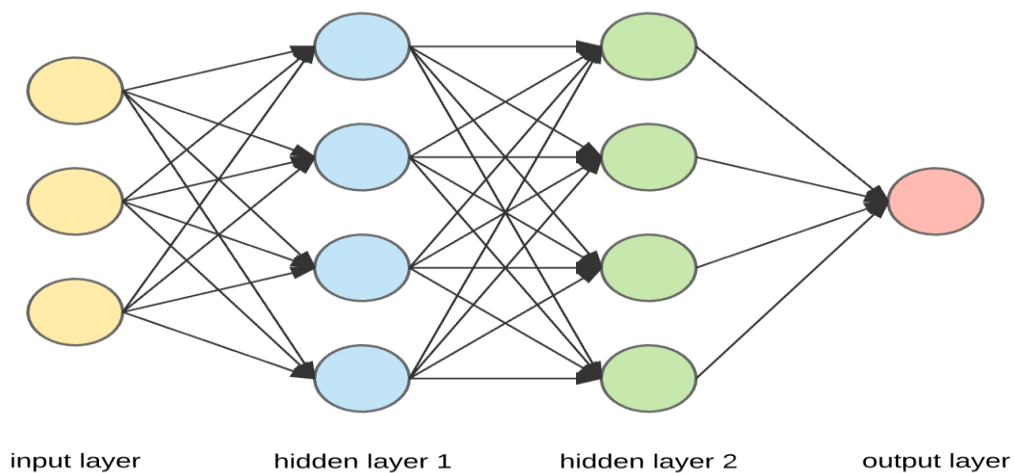


Figure 3.2: Neural network model architecture.

This is a neural net structure consisting of several nodes interconnected via directional links. Every node is a processing unit with the relationship represented by the links between connected nodes and can be changed, because the output depends on related parameters which are modifiable. An ANN model has power to compute and process information through its massive parallel distributed structure and also possesses the ability to learn and therefore to generalize. The capability of the ANN model to produce correct outputs for new inputs that were otherwise not encountered during the process of learning is referred to as generalization. For an ANN model the most important property is its environmental learning ability based on real-life experiences and aimed at improving the performance of the model through the process of learning. Iteration is the process through which an ANN model learns about its environment by applying adjustments to its connection weights. After every iteration during the learning process, it is important to note that the neural net becomes more acquainted to its environment.

Each sample data for the training of an ANN is made up of the input X (n) and the corresponding output y . The function which is used to calculate the output from the input vector comprise two parts in which the first one calculates the net input while the other one transforms this net input into a nonlinear output. Calculation of the net input (Net_j) is based on the equation below (Luo *et al.*, 1997):

$$Net_j = \sum_{i=1}^m w_{ji} x_i \quad (3.3)$$

The artificial neurons' weights w_{ji} are related to the actual synapse strengths of the neural network between the axons sending the signals and the projections of the neurons receiving them. The transfer function $f(u)$, which is basically a sigmoid function is given by:

$$f(u) = \frac{1}{(1 + e^{-u})} \quad (3.4)$$

This function node j to produce an output given as:

$$OUT_j = f(Net_j) \quad (3.5)$$

Another transformation is done to get to get an output (OUT_k), represented by:

$$OUT_k = f(Net_k) \quad (3.6)$$

This output of node k obtained from the last layer is then compared with various sample concentrations (Y_k) and the corresponding error in the training dataset using the equation:

$$Y_k = F(OUT_k) + E \quad (3.7)$$

Where F is the training function of the neural net model and E the calibration error. The complexity of actual neurons is never taken into account when modelling artificial neurons. This is because such neurons are basically made up of inputs which are multiplied by the respective signal strengths (weights) and then calculated by using a mathematical function which eventually determines the activation of the neuron. ANNs process information by combining artificial neurons given that higher input signals of an artificial neuron produce stronger input signals (Büchele *et al.*, 2019). For particular inputs, the desired output can be achieved by varying the weights of an artificial neuron through different computations based on the input signals.

When we have an ANN with many neurons to the tune of hundreds or thousands, it would complicate the process of determining the necessary weights by counting. However, methods which could be used to make adjustments to the ANN weights through learning or training may be utilized for the purpose of attaining the necessary output from the network (Reinholds *et al.*, 2015). ANNs differ in functions used, the values of acceptance, the structure, the learning algorithms such as the back propagation, etc. Backpropagation Algorithm is used in layered feed-forward ANNs in which the arrangement of the artificial neurons is in form of layers which fire their signals “forward”, and then transfer the errors backwards (Luo, 2006). The input signals received by the network are from neurons in the input layer while the neurons in the output layer produce the output signals.

The backpropagation method makes use of supervised learning by which the algorithm is provided with input and output signals of interest for the network to compute and finally, the error which is the difference between the actual and expected results is then computed. Supervised learning deals with training spectral data chosen from a grouped dataset to make a multilayer model with three clear characteristics. The first one comprising neurons in the hidden layer which are neither in the output nor in the input

layers but make the neural net model to learn and tackle problems with complications. Secondly, the nonlinearity displayed in the neural network which is differentiable while the third characteristic depicts the model's ability to display a higher order of connectivity due to its network. In supervised learning, an ANN model is trained and made to learn through back-error propagation based on the input and output data (Grieken *et al*, 2002). Learning through back error propagation involves forward and backward propagations. In forward propagation, an input vector is injected into the network and propagates the neurons as signals one at a time, making them emerge at the other end of the network as output signals (Carlos *et al*, 1996).

The error for a neuron is determined by comparing the calculated output and the expected outcome. During the backward propagation, the error obtained from the output neuron is fed into the network and propagated backwards. In the process, the slope for each neuron in a given layer is calculated and then the weights of the synapses allowed to change accordingly. The backpropagation algorithm reduces the error until the time when the ANN model learns about the data used for training. Training starts with weights which are randomly picked and repeatedly adjusted until a minimum error is attained (Gershenson, 2003). The backpropagation algorithm is generally taken as the summation of the product of the x- inputs and their respective weights (w_{jt}) as per equation (3.8) below and it is also the activation function of the artificial neurons in ANNs.

$$A_j(x, w) = \sum_{t=0}^n X_t W_{jt} \tag{3.8}$$

3.4 The X-Ray Fluorescence (XRF) Spectrometer

This is a tool used for the qualitative and quantitative analysis of a sample in order to establish the elements in the sample irrespective of the chemical composition. X-ray fluorescence spectroscopy is a method which is useful in the determination of different concentrations or chemical composition in various samples. (Marguí *et al.*,2014). It works on the principal of X-ray excitation in which high energy X-rays from a source irradiate elements in a sample making them emit characteristic X-rays, on which the radiation energy produced is dependent in accordance with the Mosley's equation,

$$E = K(Z - S) \quad (3.9)$$

In this equation, the constants K and S rely on spectral line series while the atomic number of the element is represented by Z. This method is fast, accurate and non-destructive which is usually used in determination of solid and liquid sample concentrations such as metals, cement, oils, polymer, food substances, pharmaceuticals, water and waste materials. This technique however has challenges associated with matrix effects and poor detection limits for low atomic elements ($Z < 13$), (Marguí *et al.*, 2009; 2014). Two XRF instruments in use are EDXRF and the WDXR spectrometers. The latter is based on Bragg's law

$$2d\sin\theta = n\lambda, \quad (3.10)$$

in which the diffraction of single crystal or synthetic multilayer uses a characteristic angle to detect X-rays of particular wavelength while the former uses a solid-state detector for detecting the fluorescence of the characteristic X-rays. EDXRF, for which the elemental range is from Na to U is more popular for laboratory applications as compared to WDXRFS whose range is from Be to U. However, count rate limitation and low energy resolution are the main disadvantages of the EDXRF, resulting in its reduced precision and accuracy (Antwerpen and Clapera, 2006).

In EDXRF spectroscopy, analysis is based on the principle of linear correlation that exists between sample concentrations and fluorescent X-ray intensities emitted by the present elements. In practice, there exists a nonlinear relationship between the concentrations and intensities for a specific sample due to the matrix or elements in the analyte. Since 1960, technological evolution has resulted into the present-day compact light weight EDXRF

designs which are electrically cooled. These, including the air-cooled low power X-ray tubes are suitably utilized as hand held spectrometer tools, an improvement from the former Liquid nitrogen cooled solid state detectors, nuclear electronics and small computers. More advanced spectrometers which have very low detection limits (MDLs), though very expensive, are the TXRF) and the Synchrotron radiation types (Wobrauschek *et al.*, 2010). In an ordinary EDXRFS, the crystal of the detector which is made of Silicon drifted with Lithium Si(Li) disperses the energy radiation and also counts the corresponding photons resulting into a spectrum consisting of intensity versus energy of the characteristic sample radiation (Çevik *et al.*, 2003).

3.5 Energy Dispersive X-Ray Fluorescence Spectrometry

It is a technique used to analyze a sample qualitatively and quantitatively in order to determine the elements present, ranging from sodium (Na) to uranium (U) in a wide range of sample concentrations. The versatility of this technique emanates from its ability to provide rapid, non-destructive multi-elemental analyses of low sample concentrations in parts per million (Wobrauschek *et al.*, 2010a). The fluorescent X-rays are analyzed qualitatively by using the Moseley's law/ equation

$$E_x = RhC(Z-\sigma)^2 \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right) \quad (3.11)$$

where E_x is the characteristic x-rays energy, R , which is equal to $1.09737 \times 10^7 \text{ m}^{-1}$, represents the Rydberg constant, h , the Planck's constant is equal to $6.6262 \times 10^{-34} \text{ J}\cdot\text{s}$ and C , the photons velocity is equal to $3.0 \times 10^8 \text{ m/s}$. Z and σ represent the atomic number and the shielding constant respectively while n_1 and n_2 are the corresponding energy series.

In this case, the detection of the fluorescent x-rays can be done and their spectrum of intensity against energy displayed, from which the peak locations are used to establish the various elements contained in the sample (Reinholds *et al.*, 2015). During quantitative analysis, the percentage number of energies or peak heights hitting the detector at equal energy level are utilized in the quantitative determination of each sample element.

When the incident radiation energy is very high, it ejects an innermost electron from its atomic orbital and as a result, an electron from the outer shell drops into the vacancy to fill the hole left behind. By this transition a radiation of characteristic energy (X-ray fluorescence) is emitted and as a result, a fluorescence detector detects it. The radiation

energy required to ouster an innermost electron corresponds to every element and therefore is the emitted energy resulting from the transition. A transition of an electron from the L shell into the K shell is termed a K_{α} transition, while an electron falling from the M shell into the K shell is a K_{β} transition. Line spectra which are categorized in series such as K, L, M, N etc., result from electron transitions released from different higher levels to the same shell (Wobrauschek *et al.*, 2010a).

When an empty space in a K shell is filled by an electron from the L shell K_{α} line radiation is obtained but when an electron from the M shell fills this space, it results in a K_{β} line radiation. In the same way, the L_{α} line radiation is obtained when the space in the L shell is filled by an electron from the M shell, and if filled by an electron from the N shell, we get the L_{β} line radiation. However, an L_{γ} line radiation is obtained if this vacancy is filled by an electron from the O shell (Melquiades *et al.*, 2015). The K, L and M series are the characteristics of every element.

Light elements produce K lines only, mid-range ones emit both K and L series while K, L and M series are produced by the heavy elements. EDXRF spectrometers convert characteristic X-rays into electrical signals by using a semiconductor material detector (Silicon-Lithium drifted detector). The signals produced are captured by the spectrometer's electronics and then transmitted to PC or internal electronics where they are analyzed and displayed.

The EDXRF analysis principal works on a process of direct excitation through which sample atoms are excited by incident photons from external sources, which may be a radioactive source, a synchrotron beam or an X-ray tube to produce primary fluorescence. Alternatively, indirect excitation, in which secondary fluorescence is produced by particles (electrons) emanating from direct excitation or any other secondary processes in the sample may also be used (Grieken *et al.*, 2002). X-ray, an electromagnetic radiation with wave-particle duality is emitted when high-energy particles strike the sample atoms. XRF spectrometry makes use of the primary X-ray emissions or other photonic particles for excitation of the sample atoms which results in secondary fluorescence useful for analyzing the composition of the material.

During direct or indirect excitation of a sample, the emitted electron is replaced by another electron from an outer atomic shell resulting in X-ray fluorescence, otherwise referred to as characteristic X-rays which correspond to particular energy level by a given element. For example, when an electron is ejected from the K shell of a manganese atom, another electron from the L shell of the atom moves to replace it, releasing 5.894 keV of energy. In an EDXRF spectrometer, the characteristic X-rays are converted into electrical signals by a detector. The signals are digitalized by electronics in the spectrometer and then sent to a PC for analysis and display (Yao *et al.*, 2015).

EDXRF spectrometers work on the fact that the energy of the X-ray photon directly correlates to the detector's signal pulse height which corresponds to the wavelength.

The fluorescent X-rays are detected and displayed as a spectrum of intensity versus energy, from which the peaks' location depict the elements present in the sample (qualitative analysis) while the peak heights reveal the quantity of every element in the sample (quantitative analysis).

3.6 Interaction of X-rays with matter

When an X-ray beam strikes a sample material it interacts with the sample atoms through photoelectric effect, coherent (Rayleigh) scattering or incoherent (Compton) scattering. In elements with high atomic number, photoelectric effect occurs when the photon energy is large enough to create a vacancy by ejecting an electron from one of the inner shells, resulting in characteristic fluorescence in the form of X-rays. The created vacancy may also de-excite an atom leading to ejection of Auger electrons, a common occurrence with low atomic number elements.

Coherent or incoherent scattering is dependent on composition of the sample and the photon energy. In a sample the target element is identified by the measurement of corresponding wavelength or energy. The attenuation of the X-rays beam intensity is via the sample in the aforesaid processes and takes place in accordance with the following law, expressed in exponential form:

$$I(E) = I_0(E) \exp(-[\mu(E)\rho x]), \quad (3.12)$$

from which $I_0(E)$ is the original beam intensity, $I(E)$ the beam intensity after travelling a sample distance x , ρ the sample density while $\mu(E)$ represents the attenuation coefficient of the sample mass. The properties influencing scattering processes as well as MDLs in an EDXRF are the sample thickness and the atomic number (Marguí *et al.*, 2014).

Three parts to consider for the proper working of an EDXRF spectrometer are the source of the beam of X-rays, the spectrometer's beam geometry and type of the detector. The source of the X-rays may be a tube of X-rays, a radioisotope emitting gamma rays from decaying radioisotope or fluorescence x-rays produced by the disintegrating material. Emission of a lower intensity beam than one from the X-ray tube as well as common isotopes with long half-lives are the disadvantages. The advantages are stable intensity source and that electric power source is not required.

However, radioisotope sources are commonly used in light instruments for field work. (Murphy *et al.*, 2012). Synchrotron radiation, another X-ray source is ideal for an EDXRF because the preferred energy with ideal brilliance can be obtained and confined as well, by focusing it in the electron plane or positron orbit, making it suitable for use in the reduction of scattered radiation which in turn minimizes background radiation (Wobruschek *et al.*, 2010).

In a beam spectrometer geometry, the triaxial arrangement of the beams (All at 90 degrees to each other) makes the x-ray beam linearly polarized, reducing scattering on the secondary target and thereby reducing background radiation. This arrangement minimizes the detection limits for a variety of elements. The third part is the type of detector, which may be liquid Nitrogen cooled or Silicon- Lithium drifted, the most widely used type. Other detectors are mercury-iodide detectors, high purity detectors and SDD, the latter being of poor efficiency due to its thin crystal dimension.

3.6.1 Radiation scattering

Scattering occurs when photons incident on a sample bounce away after hitting an electron. The photon loses some little energy to the loosely held electron which depends on the angle of incidence (Grieken *et al.*, 2002). This type of scattering is known as Compton (incoherent) scattering in which the energy of the scattered photon is given by:

$$h\nu = \frac{h\nu_o}{1 + \gamma(1 - \cos \theta)} , \quad (3.13)$$

where

$h\nu_o$ is the energy of the incident photon while

$$\gamma = \frac{h\nu_0}{M_0 C^2} \quad (3.14)$$

For very small energies, $h\nu_0 < M_0 C^2$

Thompson's scattering cross section for electromagnetic radiation on an electron is given by the equation:

$$\frac{d\sigma_{TH}}{d\Omega} = \frac{r_0^2}{2} (1 + \cos^2 \theta) \quad (3.15)$$

When photons bombard electrons strongly bound to the sample atoms, Rayleigh (coherent) scattering occurs. The electrons struck oscillate at the same frequency as that of the striking photon and thereby producing radiation at the same frequency (Yao, *et al.*, 2015). For this type of scattering, the differential equation for the cross-section is presented as:

$$\frac{d\sigma_R}{d\Omega} = \frac{d\sigma_{TH}}{d\Omega} \cdot F^2(q, z) \quad (3.16)$$

Where $\frac{d\sigma_R}{d\Omega}$ represents the differential Rayleigh cross-section while $\frac{d\sigma_{TH}}{d\Omega}$ is the

Thomson scattering cross-section per electron, referred to as the probability for scattering into unit solid angle per electron per unit incident flux. $F(q, z)$ is the factor for the atomic form of element Z , corresponding to the momentum transfer and $q = \sin(\theta_s/2)/\lambda$, the square form factor which is proportional to Z^2 for the atom. λ represents the incident photon wavelength while θ_s is the scattering angle.

The energy of the scattered photon is determined by the scattering angle and the incident photon's energy as per the equation below:

$$E = \frac{E_0}{1 + \gamma(1 - \cos \theta_s)} \quad (3.17)$$

E_0 accounts for the incident photon's energy, $\gamma = E_0 / M_0 C^2$ while θ_s represents the scattering angle which takes any value between 0 and 180° (Grieken *et al.*, 2002).

The radiation incident on the surface is slowly absorbed by the sample and at a layer of thickness t below the surface, the remaining fraction of the intensity $I_t(\lambda_0)$ is given by the Lambert-Beer law:

$$I_t(\lambda_o) = I_o(\lambda_o) \exp[-\mu_s(\lambda_o) \rho_s t (\csc \psi')] \quad (3.18)$$

where

$\mu_s(\lambda_o)$ is the specimen's mass attenuation coefficient

λ_o the specimen's wavelength and ρ_s the specimen's density.

For angle of incidence ψ' the path length travelled by the beam is given by

$$t \csc \psi' \quad (3.19)$$

The attenuation coefficient for the sample mass for elements j in the sample is the numerical sum of the product of each of the n elements present and their respective fractional mass W_j , according to the equation below (Grieken *et al.*, 2002):

$$\mu_s(\lambda_o) = \sum_{j=1}^n [\mu_j(\lambda_o) W_j] \quad (3.20)$$

CHAPTER FOUR

MATERIALS AND METHODS

4.1 Overview

In this chapter, instrumentation of EDXRF spectroscopy as well as the establishment of a suitable base matrix used in this work for simulate samples preparation has been discussed. Development of ANN models for chemometric assisted EDXRF spectroscopy has been outlined. Acquisition of optimum sample volume, analysis time and EDXRF spectra has also been handled in this chapter. The multivariate chemometric tools used to perform EDXRF spectral deconvolution and exploratory data analysis included PCA and ANNs (Luo, 2006; Reinholds *et al.*, 2015).

4.2 Instrumentation of EDXRF Spectroscopy

An EDXRF Spectrometer, one of the forms of XRF spectrometers is an analytical tool used for qualitative and quantitative analysis of the elements present in a sample. Figure 4.1 below illustrates the EDXRF schematic arrangement and instrumentation:

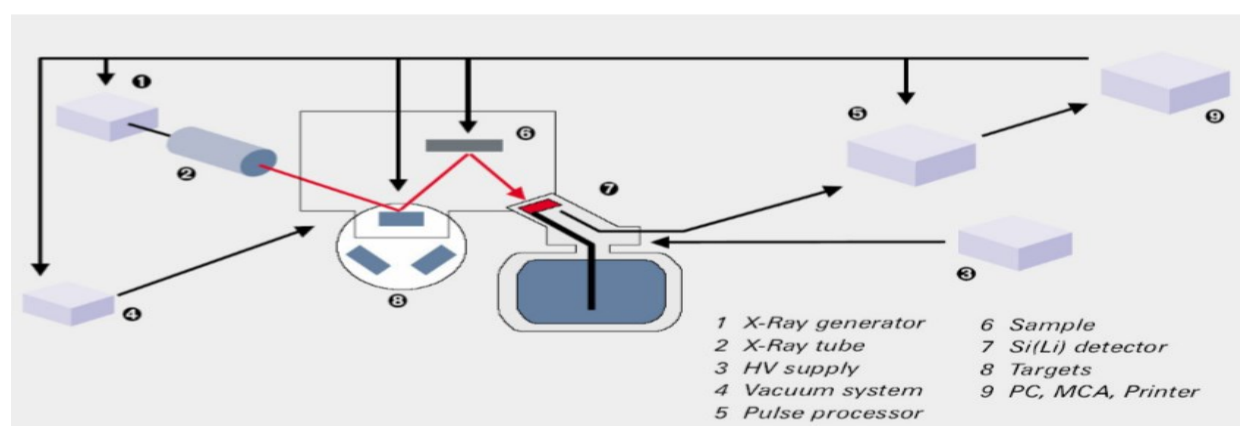


Figure 4.1: EDXRF Schematic Instrumentation

In this research, Rigaku NEXCG model of the EDXRFS was used to perform spectral analysis of liquid samples containing Hg concentrations from blank to 500000 ppb. The use of secondary target excitation rather than conventional direct excitation improved its sensitivity. Excellent spectral resolution and very high-count rate ability was achieved by using a high-performance SDD.

The system's hardware consists of the excitation source, basically an x-ray tube with 5 kV anode potential difference and a 50 W power, high performance SDD, a sample changer (38cm in diameter) having 15 position automatic sample changer external PC computer system.

The spectrometer operation was done procedurally by preheating it first for about 40 minutes before the experiment to increase the tube pressure, produce gradual flow to protect the light pipe and thereby achieving a higher stability. Initialization, necessary for calibrating channel magnification was the next step, during which a silver (Ag) sample was stimulated and the magnification adjusted to achieve the system's hardware best working condition. Air, the correct and suitable condition for this work such as air, was used. Finally, elemental determination through qualitative and quantitative analysis of the EDXRF sample spectral data from the PCs was carried out.

4.3 Base Matrix candidate identification

Preliminary investigation involved carrying out an EDXRF spectra for pure honey, glycerol, olive oil and glycerol-distilled water mixture in order to establish the suitable base matrix for use in this work. EDXRF spectra for two different skin lighteners was also taken and the dataset for the six spectra saved in an excel csv file. A directory containing this data was imported into the R software code for principal component analysis. Fig. 4.2 below captures the four samples used for this investigation:

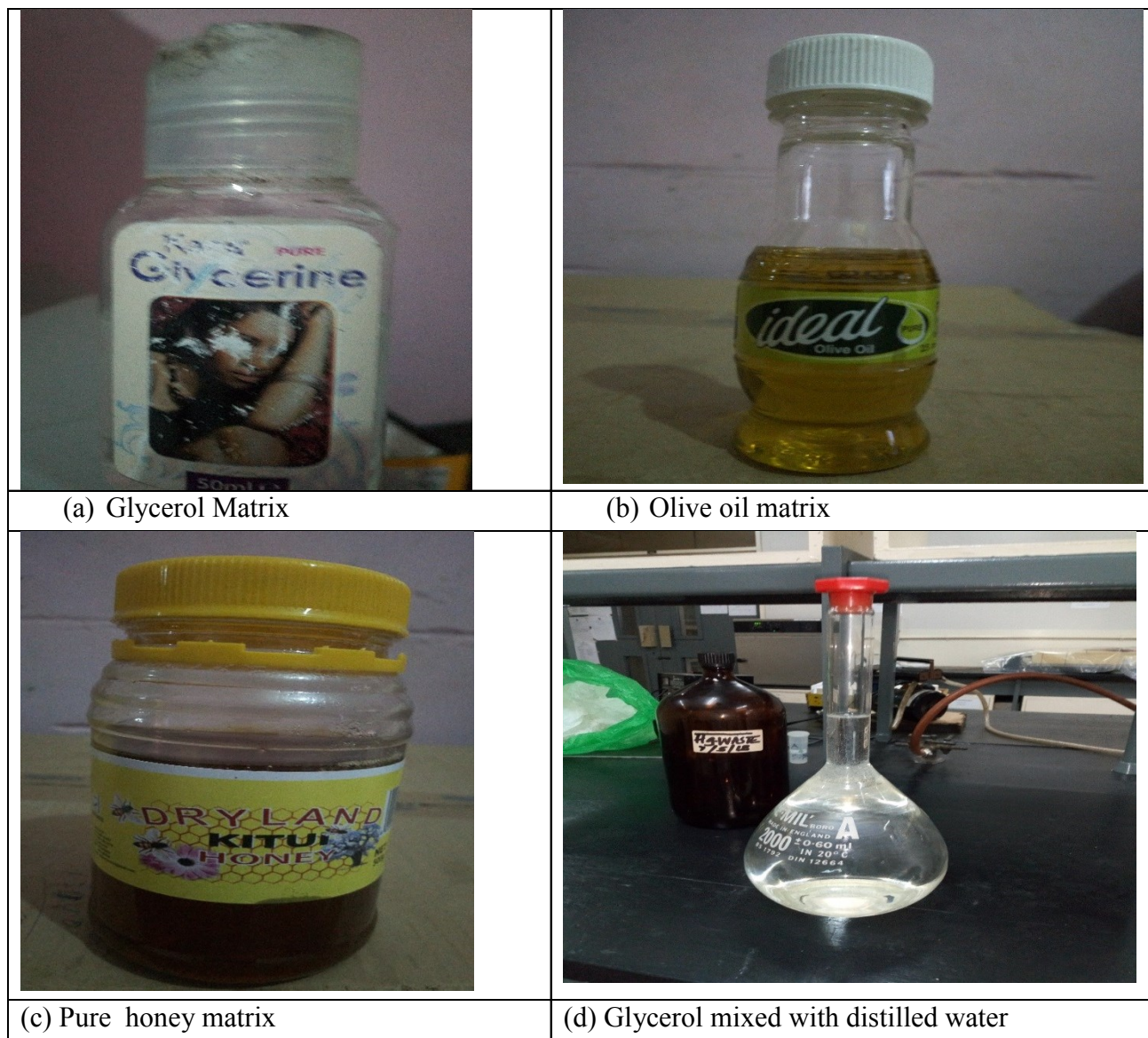


Figure 4.2: The base matrix samples

Glycerol, a colorless, odorless and viscous liquid at room temperature is non-toxic in low concentrations. Its other names are glycol alcohol, glycerin or glycerine (Quispe *et al.* 2013). The empirical formula is $C_3H_8O_3$ while its structural formula is $C_3H_5(OH)_3$. Glycerol, a trihydroxy sugar alcohol with three carbon atoms and three hydroxyl groups makes it an organic polyol compound. It has several structures but the simplest is shown below:

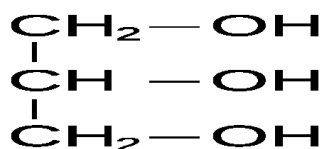


Figure 4.3: Glycerol structure

Pure glycerol has a melting point of 17.8° C and boiling point of 290° C. It is hygroscopic compound which makes it useful in cosmetics because it retains water and prevents the substance from drying out. Another property is that glycerol is miscible with water due to the fact that the polyol groups form hydrogen bonds with molecules of water(Sarkar *et al.* 2002).

4.3.1 Spectra analysis for candidate skin lightening base matrices

To identify the correct base matrix candidate for simulate samples preparation, EDXRF spectra were obtained from pure honey, glycerol and olive oil (see Fig. 5.1). It was noted that both the glycerine and pure honey have closely related characteristics different from those for olive oil. This is because they are equally viscous, organic in nature and are readily miscible with water.

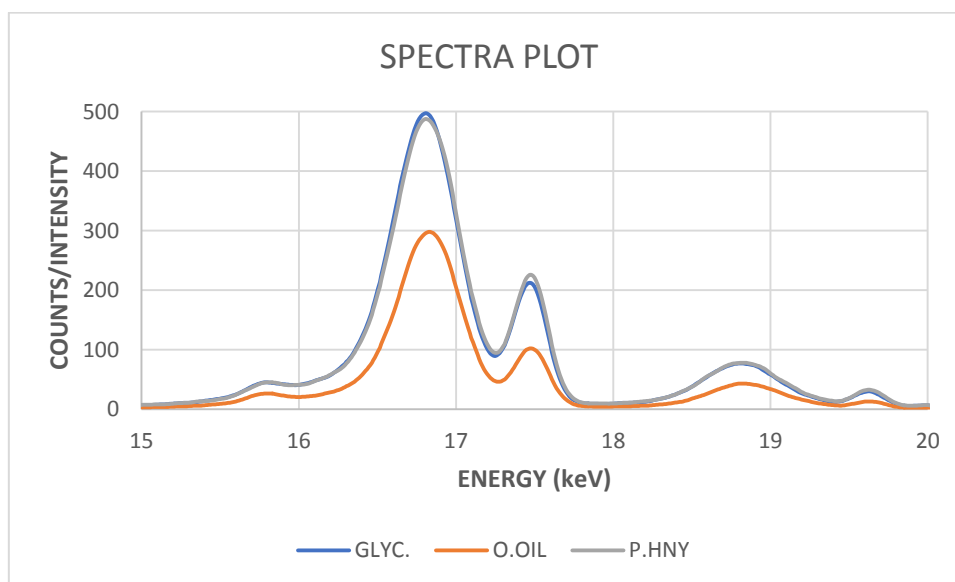


Figure 4.4: EDXRF scatter plot for the base matrices

The scatter plot was used in the description of the candidate skin lightening base matrices because it displayed discriminatory characteristics that distinguished the three matrices. The difference between olive oil and the other two matrices is confirmed in the PC plot shown in fig. 5.2 below.

A PCA for the EDXRF spectral data was also done for further investigation, incorporating glycerol and water mixture, as well as real skin lightening samples (see Fig. 4.5).

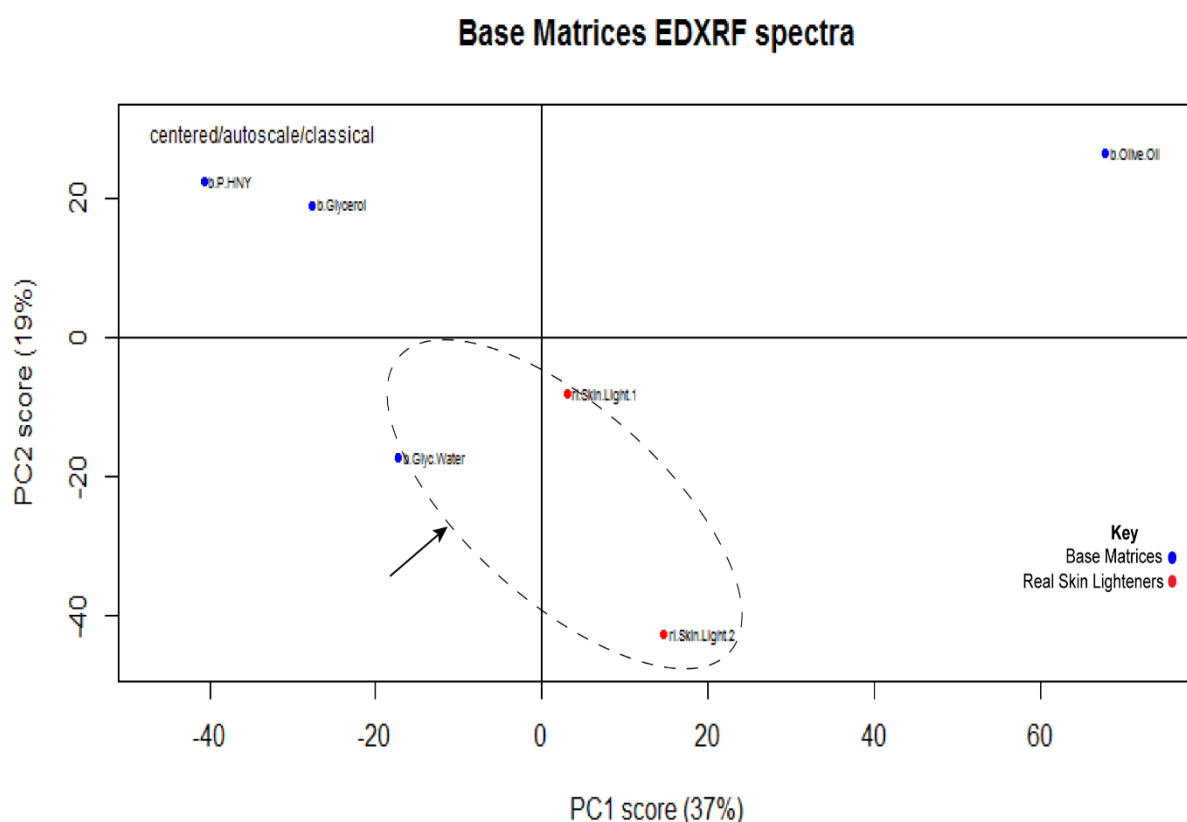


Figure 4.5: PCA analysis results for candidate base matrices

All the four base matrix samples and the two skin lighteners were fully described by a total variance of 56% in the two PCs. The positive side of PC1 as well as the negative side of PC2 fully explained the properties of the two skin lighteners while the glycerol-distilled water mixture was described by negative sides of both PC1 and PC2. Pure honey and glycerol were fully described by positive of PC2 and the negative of PC1. Olive oil was described by the positive of the two PCs. However, the clustering of the two skin lighteners and the glycerol-distilled water mixture on the negative side of PC2 unveiled lots of similar properties and characteristics found in them.

This confirms that the manufacture of skin lighteners makes use of base formulations which are mainly a mixture of glycerol and distilled water as already established previously (Breternitz *et al.*, 2008; Dnil *et al.*, 2019; Palmer *et al.*, 2000).

In the cosmetics industry, glycerol is used as a moisture control reagent and also to enhance the texture of creams and lotions. It is a widely used cosolvent and vehicle in aqueous formulations such as creams and lotions (Quispe *et al.*, 2013). One of its two major effects is that it solvates skin protein alpha-keratin and occupy the hydrogen bonding sites, causing a reduction in tissue binding of the penetrating substance. It also decreases the barrier function of the stratum corneum by disrupting lipid and protein structure, making the skin more permeable (Palmer *et al.*, 2000) However, the key function of glycerol in creams and lotions is to act as a viscosity enhancer and to retard evaporation.

Water, a commonly used penetration enhancer in beauty products increases skin permeability, making the absorption of mercury into the skin much easier. Humidity, among other conditions in the environment influences the water content in the skin, causing hydration. In this regard therefore a homogeneous solution, basically a glycerol-distilled water mixture was chosen and used in this study as the base matrix owing to its function in skin lightening creams and lotions.

4.4 Simulate samples preparation and analysis

Mercury chloride (HgCl_2) and glycerol ($\text{C}_3\text{H}_8\text{O}_3$) were used. Organic Analar grade glycerol was mixed with distilled water in a ratio of 1:1 and used as the base matrix (solvent) for the inorganic mercuric salt.

The relative formula mass for the mercuric salt is 271.50g while its percentage purity (Assay) is 99.5%. The relative atomic mass for mercury is 200.59g. Thus, the mass of 1g of mercury in the mercuric salt was given by:

$$\frac{271.50}{200.59} \text{g} = 1.353507153 \text{g}.$$

This mass corresponds to an assay of 99.5%. Hence, the mass corresponding to an assay of 100% is equal to:

$$\frac{100}{99.5} \times 1.3535071539 = 1.3603 \text{g}.$$

The 1.3603g was weighed in a chemical balance and dissolved in 1000ml solvent and the mixture shaken well to obtain a homogeneous mixture as the stock solution.

The formula below was used for serial dilution of the simulate samples:

$$C_1V_1 = C_2V_2 \quad (4.1)$$

Where C_1 was the concentration of the stock solution, V_1 the dilution volume from the stock solution, C_2 the required concentration in ppm and V_2 was the dilution volume taken from the glycerol/distilled water solution and put in a volumetric flask (VF).

To prepare 100ppm, 10ml was drawn from the stock solution and diluted with the solvent using the 100ml VF. Similarly, 10ppm was prepared by drawing 10ml from the 100ppm solution and diluted with the solvent using the 100ml VF. The Hg concentrations in the ppb level were prepared in a similar way by using some selected ppm concentrations to serve as the stock solutions. Fifty simulate samples were prepared and three EDXRF spectral data for each, abbreviated as SA, SB and SC obtained.

Table 4.1: Simulate sample concentrations

<i>S/No.</i>	Hg Concentrations(ppb)	<i>S/No.</i>	Hg Concentration (ppm)
<i>1</i>	5	<i>26</i>	0.1
<i>2</i>	10	<i>27</i>	0.5
<i>3</i>	15	<i>28</i>	0.87
<i>4</i>	20	<i>29</i>	0.925
<i>5</i>	25	<i>30</i>	1
<i>6</i>	35	<i>31</i>	2
<i>7</i>	40	<i>32</i>	3
<i>8</i>	50	<i>33</i>	5
<i>9</i>	62	<i>34</i>	8
<i>10</i>	70	<i>35</i>	10
<i>11</i>	80	<i>36</i>	12
<i>12</i>	87	<i>37</i>	15
<i>13</i>	120	<i>38</i>	17
<i>14</i>	130	<i>39</i>	20
<i>15</i>	148	<i>40</i>	26
<i>16</i>	180	<i>41</i>	30
<i>17</i>	220	<i>42</i>	38
<i>18</i>	300	<i>43</i>	45
<i>19</i>	366	<i>44</i>	50
<i>20</i>	400	<i>45</i>	60
<i>21</i>	440	<i>46</i>	100
<i>22</i>	550	<i>47</i>	200
<i>23</i>	630	<i>48</i>	250
<i>24</i>	710	<i>49</i>	300
<i>25</i>	800	<i>50</i>	500

Detection capability for most spectrometers is enhanced by using longer running times in order to improve on the excitation of electrons from the K and L shells in the element of interest (Wobrauschek, *et al*, 2010). In this regard, EDXRF spectra for samples with Hg concentrations of blank, 10 ppb, 35 ppb and 62 ppb was taken at different irradiation times ranging from 25 sec to 1500 seconds in order to establish the optimum analysis time for the samples.

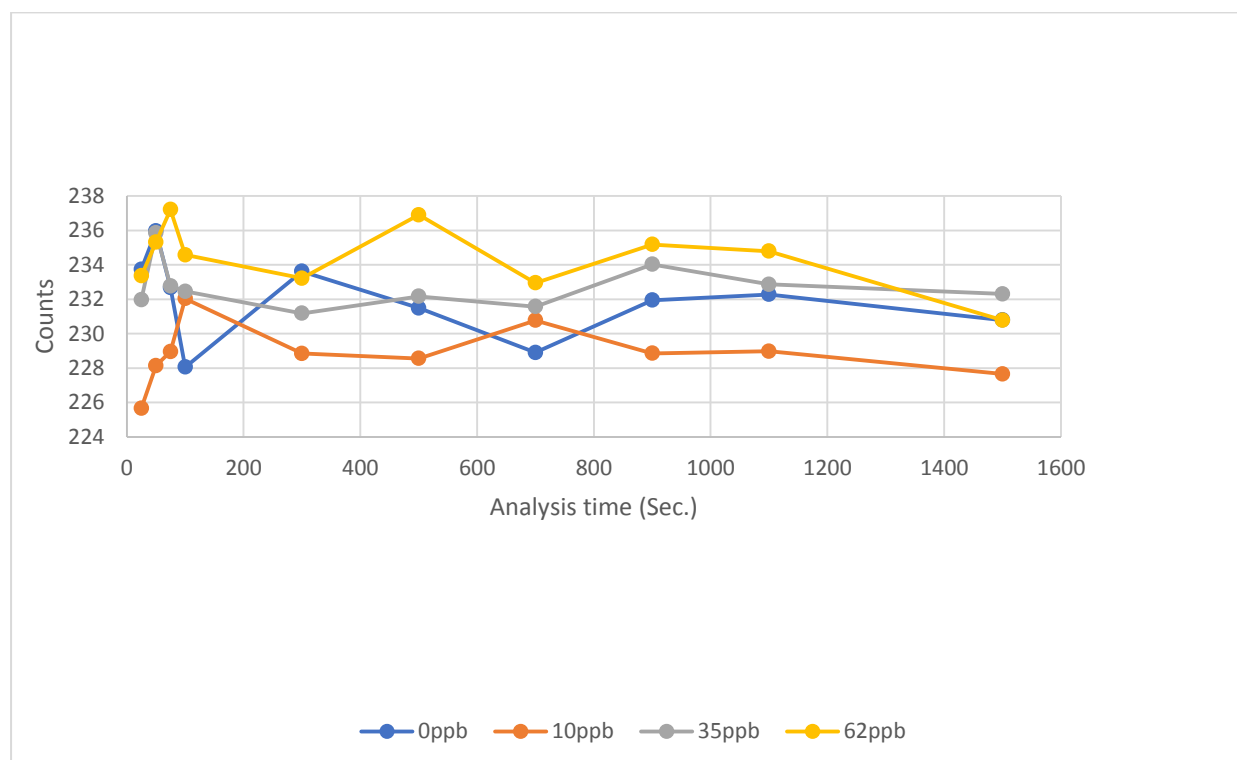


Figure 4.6: Spectral data acquisition time

The counts seem to correlate with the Hg concentration in a given sample, i.e, the higher the concentration the higher the counts. Thus, 62ppb sample displays the highest counts than the others. Below 650sec. the EDXRF detector is unable to give distinct counts for the samples. However, it is clear that from 700sec. onwards, all the samples give different counts at different corresponding times. Very distinct counts for all the concentrations is clearly noted at 900sec. The increased exposure time improved the detection capability of the spectrometer and therefore increasing the excitation of electrons from the k-shell in the element of interest (Wobrauschek, *et al*, 2010).

4.5 Sample volume for analysis

Sample volumes ranging from 0.5 ml to 2.5 ml were run in the EDXRF spectrometer and a graph of total counts versus the sample volume fitted in order to establish and optimize the sample volume for analysis.

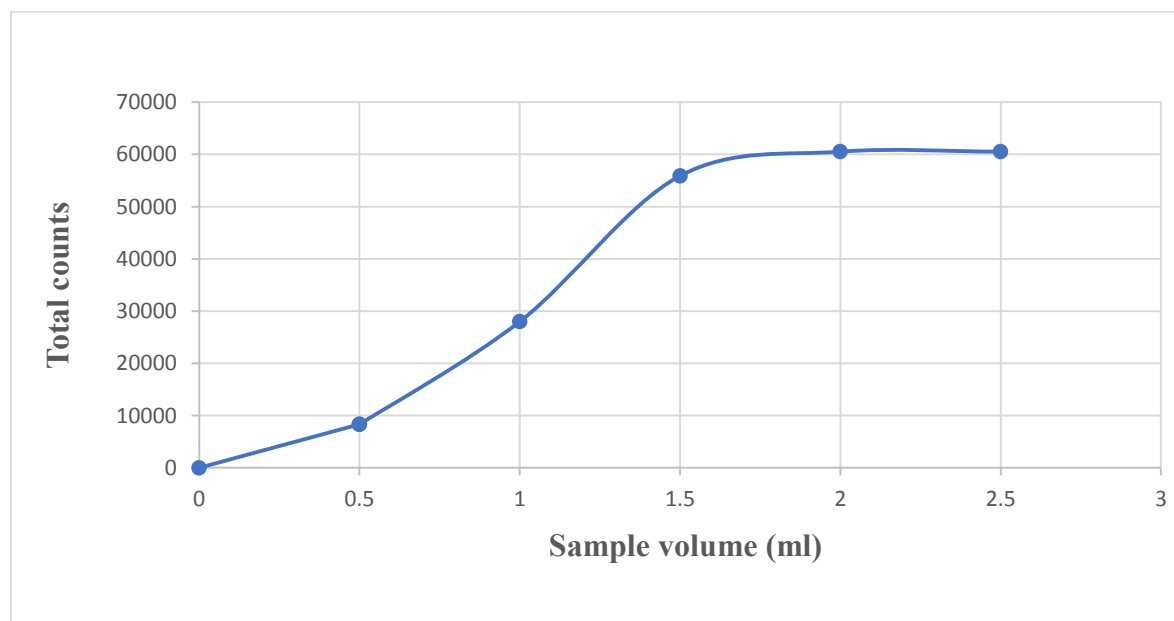


Figure 4.7: Curve for Optimum sample volume

From the plot therefore, it is clear that the optimum sample volume is 2ml. This was used during the spectral analysis of all the samples in the study.

4.6 Real Samples Preparation and Analysis

Eight different skin lighteners were sampled along River Road in Nairobi (Kenya) and old town in Mombasa (Kenya). One gram (1g) from each sample was weighed in the sample cups using a chemical balance and analyzed in the EDXRF spectrometer. Out of the eight samples, three of them registered high levels of Hg content while the rest had nothing detected, implying that their Hg concentrations were below the spectrometer's LOD of 3ppm.

From each of the five samples with undetected Hg content, 1g was weighed and put in separate test tubes containing 10 ml of the base matrix used during simulate sample preparation and thoroughly shaken to produce a homogeneous mixture. Two millilitres (2ml) from each sample were analyzed twice to produce ten different spectra.

Spectral data consisting of the ten spectra classed into groups A and B, colour coded blue and red respectively was saved in a csv file. A PCA analysis for this data was done on the R software code and the resulting PCs used to validate the developed ANN model.

4.7 Multivariate Chemometric analysis of EDXRF spectra

Simulate and real spectral data were auto-scaled, mean centered and normalized before chemometric analysis in order to remove unnecessary information from the original data matrix. Multivariate chemometric analysis i.e. PCA and ANNs were then performed on the data by utilizing the R-software code for both.

In this work, EDXRF spectral data matrix consisting of 4097 energy channels was used with back-error propagation to perform multivariate calibration of mercury concentrations. In order to eliminate some irrelevant information and reduce the number of variables, feature selection method was employed on the entire data in which only the Hg L_{α} (70 channels) and L_{β} (80 channels), within the spectra range 9.6 - 12.4 keV, were considered. After feature selection, the new data were orthogonalized by applying principal component analysis (PCA) whereby PC scores were obtained and optimized to get the first 2 PCs used as input data for the development of an artificial neural net model utilizing Hg concentrations (ppb). Another ANN model entailing Hg concentrations (ppm) was also developed in the same procedure by applying PCA on spectral data within the Hg L_{α} region (9.6 keV to 10.4 keV) and used to ascertain the results attained by using the conventional EDXRF spectrometer.

4.7.1 Data Preprocessing

A data frame consisting of 91 spectra out of which a blank sample and 30 simulate sample triplicates with Hg concentrations from 5 ppb to 1000 ppb within the ROI for the energies ranging from 9.606 keV to 12.4 keV was prepared and used to train and develop a neural net model for the Hg concentrations in ppb. The sample concentrations were classed into three groups, different colours assigned as shown below and the data saved in an excel csv file for principal component analysis.

Table 4.2: Classed Hg concentrations (ppb)

GROUP NO.	Hg CONCENTRATION (ppb)	COLOR
1	0-87	Blue
2	100-440	Green
3	500-1000	Red

Another data frame comprising 52 spectra for a blank sample and 17 simulate sample triplicates with Hg concentrations from 0.1 ppm to 30 ppm within the same ROI but in the L-alpha region for the spectral energies ranging from 9.6 keV to 10.4 keV was also prepared, used to train and develop a second neural net model for Hg concentrations (ppm). This model was used for the purpose of comparing its results and validating those obtained by utilizing the conventional EDXRF spectrometry. The sample concentrations were classed into three groups, different colours assigned as shown below and the data saved in an excel csv file for principal component analysis.

Table 4.3: Classed Hg concentrations (ppm)

GROUP NO.	Hg CONCENTRATIONS (ppm)	COLOR
1	0-1	Blue
2	2-12	Green
3	15-30	Red

4.7.2 Spectra Analysis by PCA

A directory containing a data made up of the 91 simulate sample spectra in the excel data file was imported into the R software code and used to perform PCA in R based on spectral and single value decomposition. Out of the 4097 energy channels in the matrix, the ROI for the energies ranging from 9.6 – 12.4 keV was chosen. This is the region depicting the Hg L-alpha line series (9.6 – 10.4 keV) and L-beta line series (11.4-12.4 keV). Data normalization, cleansing and standardization was also done in order to reduce matrix effects associated with the sample spectra.

4.7.3 Training ANN Models for Hg concentrations in ppb and ppm.

This was done by utilizing EDXRF spectral data with back-error propagation to perform multivariate calibration of mercury (Hg) concentrations. To eliminate some irrelevant information and reduce the number of variables, feature selection method was employed on the entire data in which only the Hg L_{α} (70 channels) and L_{β} (80 channels), within the spectra range 9.6 - 12.4 keV, were considered. After feature selection, the new data were orthogonalized by applying PCA whereby PC scores were obtained and optimized to get the first 2 PCs used as input data for the model development. Before utilizing the R software code to develop and train the ANN model from the Hg concentrations in ppb, a new column for the sample concentrations was created in the PC scores and the resulting data saved as a csv file.

The data used in this case comprised 91 spectra from which 64 were used for model calibration (training) and 27 as the test data frame. Modeling steps included data preparation, model building, evaluation and prediction. In data preparation, the data was classified, loaded into the R software, preprocessed and finally validation dataset created. The model was built by choosing the input layer, number of hidden layers including their respective nodes as well as the output layer. Evaluation involved getting the predictions on the validation dataset, using the model and calculating the RMSEP. Finally, prediction was done on the test data by using a code in the R software.

Some of the statistical information for the model were utilized in determining LOD and LOQ as per the equations below:

$$\text{LOD} = 3 \times (\text{Std error for slope}) / \text{slope} \quad (4.3)$$

$$\text{LOQ} = 10 \times (\text{Std error for slope}) / \text{Slope} \quad (4.4)$$

A second ANN model was built in the same procedure by using the 52 sample spectra for Hg concentrations (ppm). Out of this data, 36 spectra were used for training the model while 16 spectra for testing it. The developed model was utilized to validate the results obtained by the use of the conventional EDXRF spectroscopy.

CHAPTER FIVE

RESULTS AND DISCUSSION

5.1 Overview

This chapter discusses the principal component analysis for the samples used to establish the suitable base matrix used in this research work. It also handles EDXRF simulate samples results, multivariate chemometric analysis of the obtained results, development of the ANN models as well as EDXRF results for the real samples.

5.2 EDXRF Spectra Analysis Results for Simulate Samples

5.2.1 Hg ROI Results for simulate samples

Selected sample spectra were used to establish the region of interest (ROI) for Hg. Below are the results:

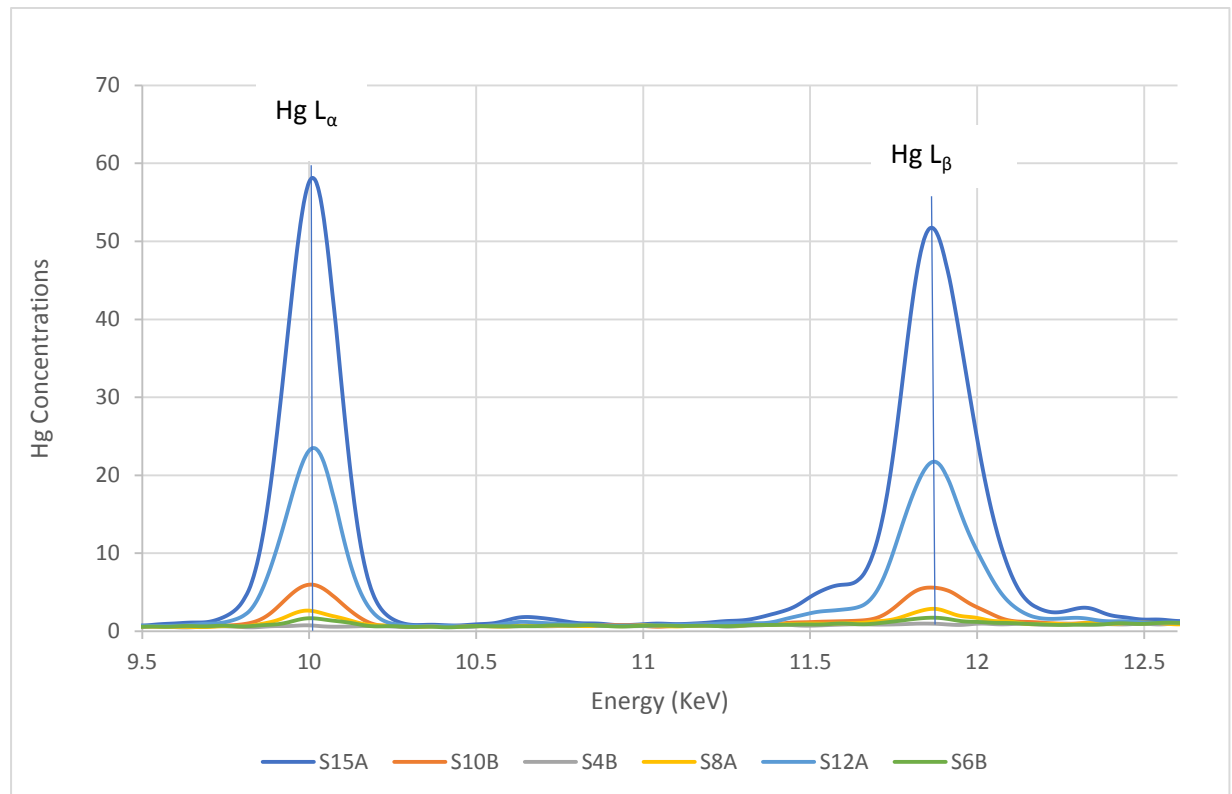


Figure 5.1: Spectra plot for simulate samples

Within the ROI, photon energy for L_{α} line series ranges from 9.598 keV to 10.4 keV while for the L_{β} line, it ranges from 11.2 keV to 12.6 keV. The above peaks, confirm that

samples with higher Hg concentrations display greater peak intensities for each of the line series. (Antwerpen, *et al*, 2006).

The EDXRFS results were analyzed by using a univariate approach from which an intensity versus Hg concentrations data plot was obtained as shown below:

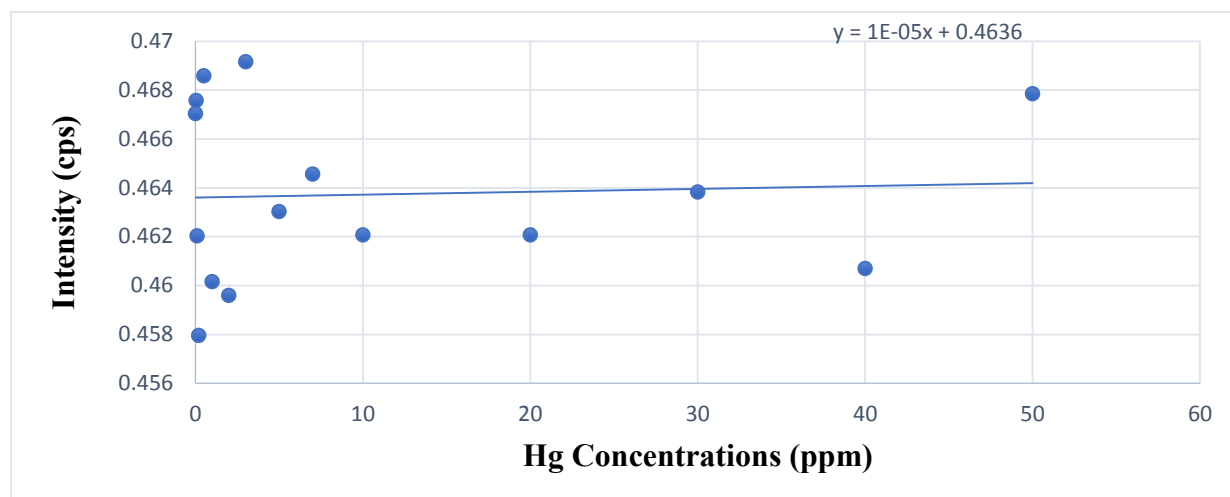


Figure 5.2: Plot for spectral Intensity Vs Hg concentrations (ppm)

The highly scattered points indicated that there was a very poor correlation between the spectral intensity and Hg concentrations proving the univariate approach to be an unsuitable method for analysis of the EDXRF spectra. In addition, detection of 2 ppm and below was a challenge to the conventional EDXRF spectrometer due to its limit of detection which was 3 ppm. Therefore, it was found necessary to employ a multivariate approach in the analysis of the non-linear data.

Consequently, two artificial neural network models were developed in this study and used for direct detection and quantification of Hg levels (ppb and ppm) both in simulated and real samples.

5.3 Results of EDXRF Spectra Data Analysis and Modelling Using Multivariate Chemometric Techniques

Full EDXRF sample spectra consisting of 4097 channels [0 - 40.8 keV] was obtained. However, feature selection method was applied in the reduction of number of variables, aimed at eliminating all the irrelevant information. Thus, the Hg L_{α} (70 channels) and L_{β} (80 channels) within the ROI spectra range 9.6 - 12.4 keV were considered. After feature selection, the reduced data were orthogonalized by principal component analysis (PCA) and four of the resulting principal component (PC) scores later used in this work as the input nodes for the ANN model.

5.3.1 PCA Results for Hg concentrations (ppb)

A directory containing data matrix made up of 91 simulate sample spectra for the sample concentrations (ppb) in excel data file was imported into the R software code and used to perform PCA in R based on spectral and single value decomposition. The results are displayed in figure 5.3 below:

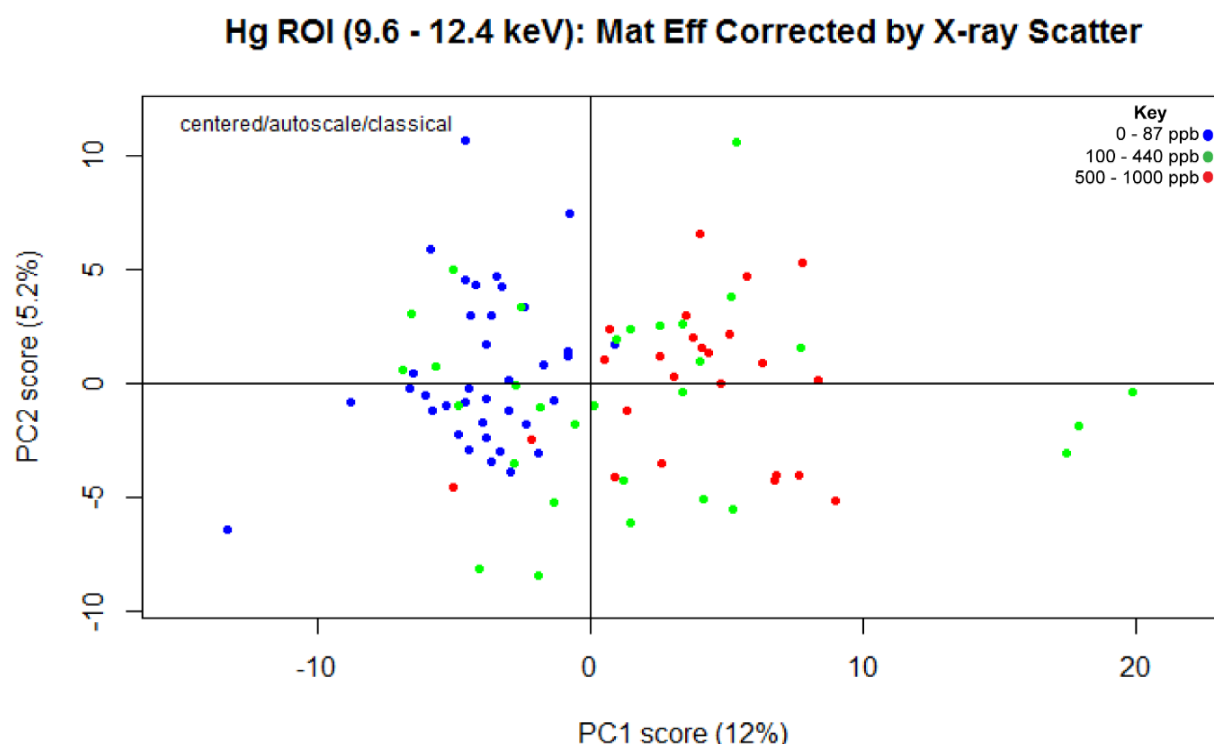


Figure 5.3: PC Scores plot for Hg concentrations (ppb)

Hg ROI (9.6 - 12.4 keV): Mat Eff Corrected by X-ray Scatter

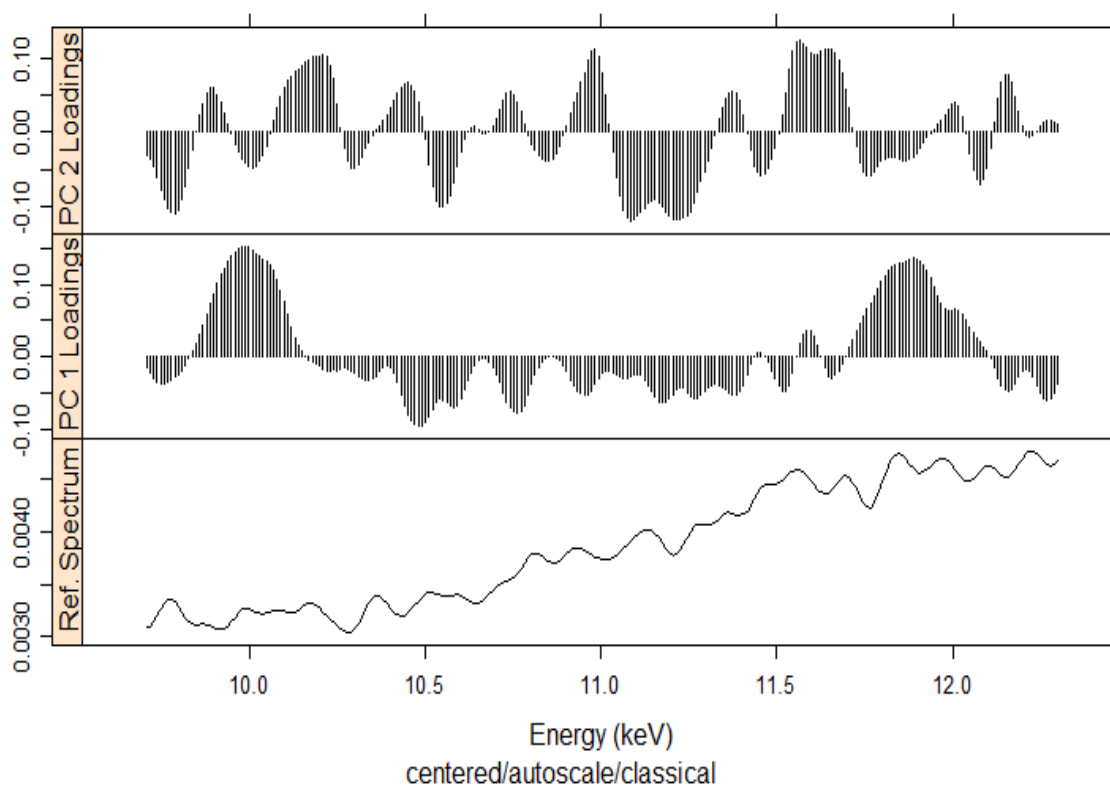


Figure 5.4: The loadings plot for Hg concentrations (ppb)

The two PC scores were fully explained by a total variance of 17.2% with the negative of PC1 describing most of the blue colored samples having Hg concentrations from 0-87 ppb. A mixture of most of the green and red cored groups were explained by the positive of PC2. Poor clustering of the samples was experienced during the analysis due to the matrix effect, a challenge which was corrected by using the X-ray scatter technique. A lot of spectral information for groups two and three is explained by the positive side of PC1 while the negative of the same PC contain some little information about the samples.

5.3.2 The ANN modeling results for the ppb Data

The model architecture consisted of 2 PCs as the input nodes, two hidden layers having 2 nodes and 3 nodes respectively and finally two output layers with 1 node each as shown below:

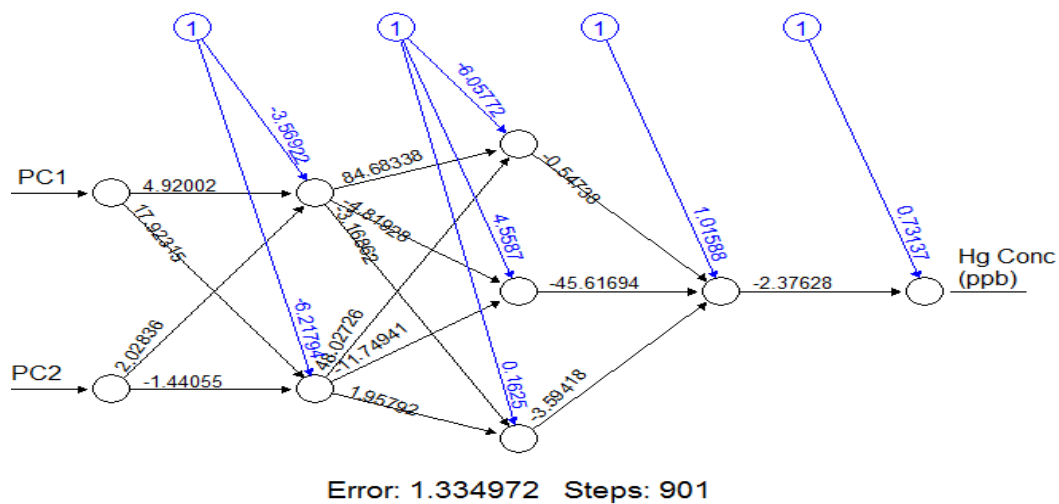


Figure 5.5: Trained Model architecture used in this work

The 2-PC scores as input data enhanced efficient reduction of the net architecture through lowering the number of nodes and weights in the input layer and thereby increasing the speed of the training phase. Optimization of the input layer was done by reducing the PCs associated with the smallest eigenvalues.

Several runs were done on the R software code during the model development until a reliable model with an RMSEP of **20.6%** was obtained. The developed model is as displayed in figure 5.7 below;

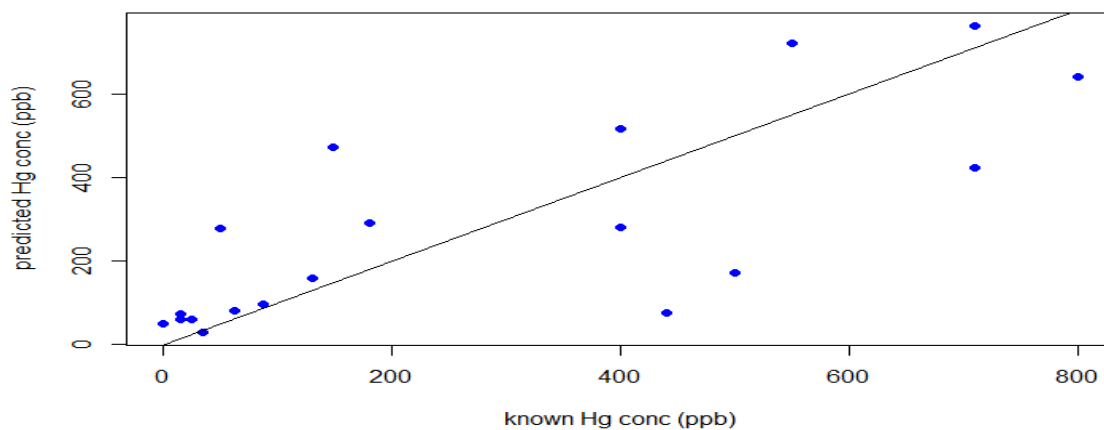


Figure 5.6: The ANN Model used for quantification of Hg concentrations (ppb).

Most of the points especially those above 200 ppb were scattered away from the linear fit implying that there was a high variability in the spectral data and a lower correlation.

Samples with Hg concentrations below 200 ppb showed little spread but most of them were highly correlated. However, the model had an R-squared value of 0.72, an RMSEP of 20.6% and an LOD equal to 527 ppb.

The developed ANN model was used to detect and quantify the Hg concentrations for the test data; table 5.1 below gives the results:

Table 5.1: Sample results by using the developed ANN model for Hg concentrations (ppb)

Known Hg conc (ppb)	Predicted Hg conc (ppb)
500	< 527
550	721±149
180	< 527
130	< 527
50	< 527
0	< 527
15	< 527
25	< 527
87	< 527
710	763±157
800	640±132
710	< 527
35	< 527
400	< 527
440	< 527
62	< 527
15	< 527
148	< 527
400	< 527

Hg concentrations for three samples were detected by the model while the rest had concentrations below the model's detection limit.

Table 5.2 below is a summary of the calculations which include other key statistical information:

Table 5.2: ANN Model statistics for Hg concentrations (ppb)

STRUCTURE	MODEL RMSE	R² VALUE	MODEL LOD (ppb)	MODEL LOQ (ppb)
2:2:3:1	20.6%	0.72	527	819

5.3.3 PCA analysis Results for Hg concentrations (ppm)

A data matrix made up of 52 simulate sample spectra for Hg concentrations (ppm) contained in a directory and stored in an excel data file was imported into the R software code, used to perform PCA in R and produced the following results:

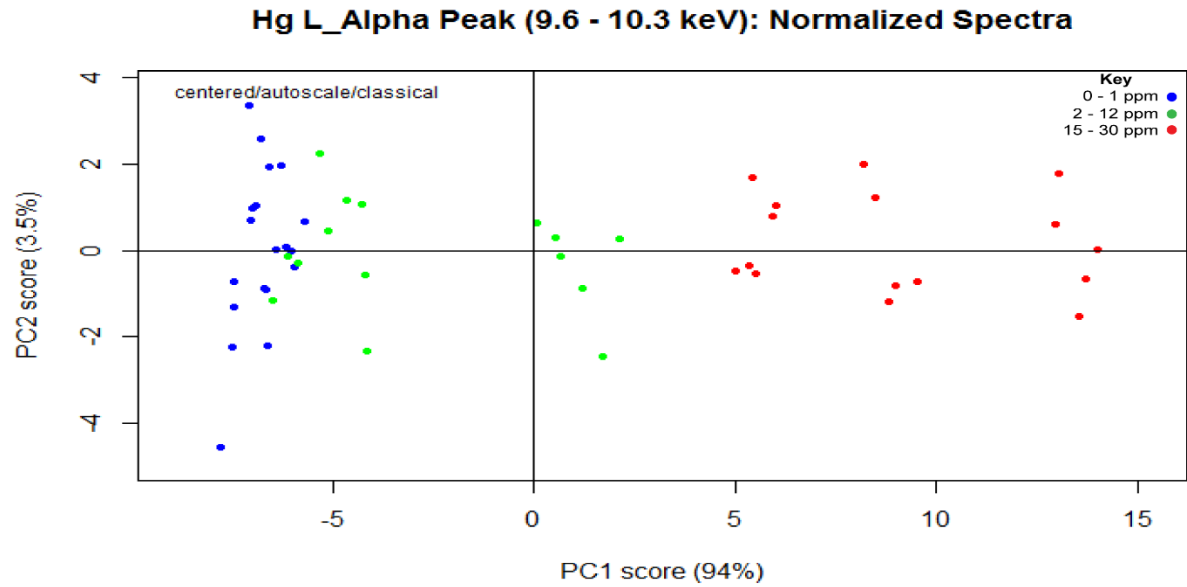


Figure 5.7: PC Scores plot for Hg concentrations (ppm)

The loadings plot obtained for the L-alpha peak region (9.6-10.3 keV) within the ROI is as per figure 5.8 below:

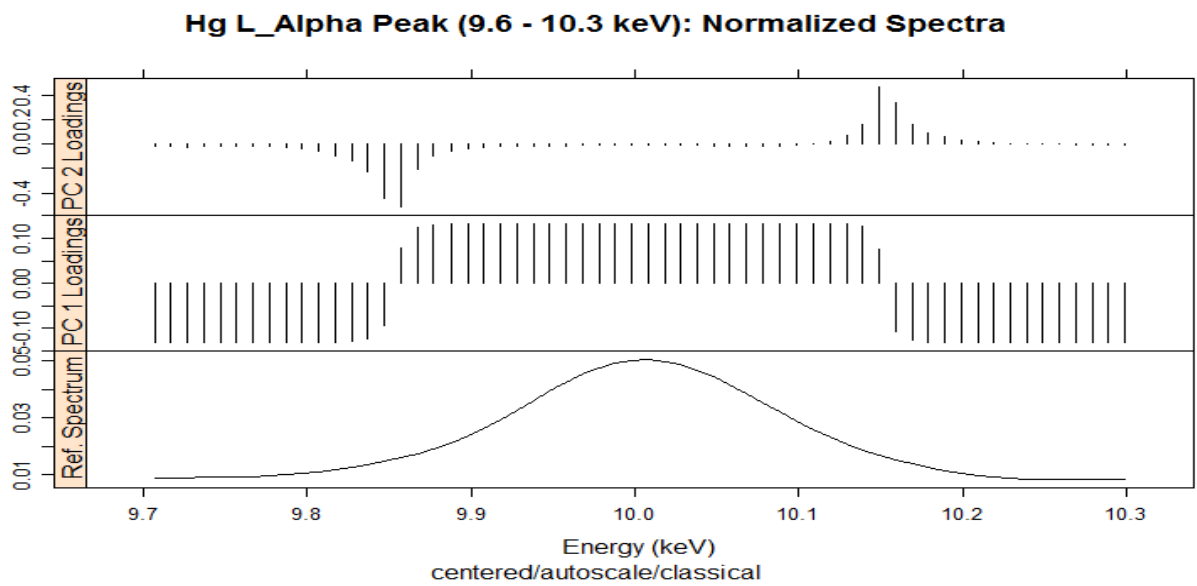


Figure 5.8: The Loadings plot for Hg concentrations (ppm)

There was a total variance of 97.5% for the two PCs which described correlation of the sample concentrations fully. Samples with the lowest Hg concentrations (0-1ppm) were clustered to the extreme end of negative PC1 while those with the highest concentrations (15-30ppm) on the extreme of the positive side of the same PC. The ones with Hg concentrations from 2-12 ppm scattered on both the positive and negative sides of the two PCs. Samples with 10 ppm and 12 ppm lay on the positive of PC1 while the rest in this group lay on the negative of PC1. However, all the sample concentrations were fully described by the positive and negative sides of both PC1 and PC2.

5.3.4 The ANN modeling results for the ppm dataset

The architecture used for model training and development consisted of a 2-PC input layer, one hidden layer with three nodes and a one node output layer as per the figure below:

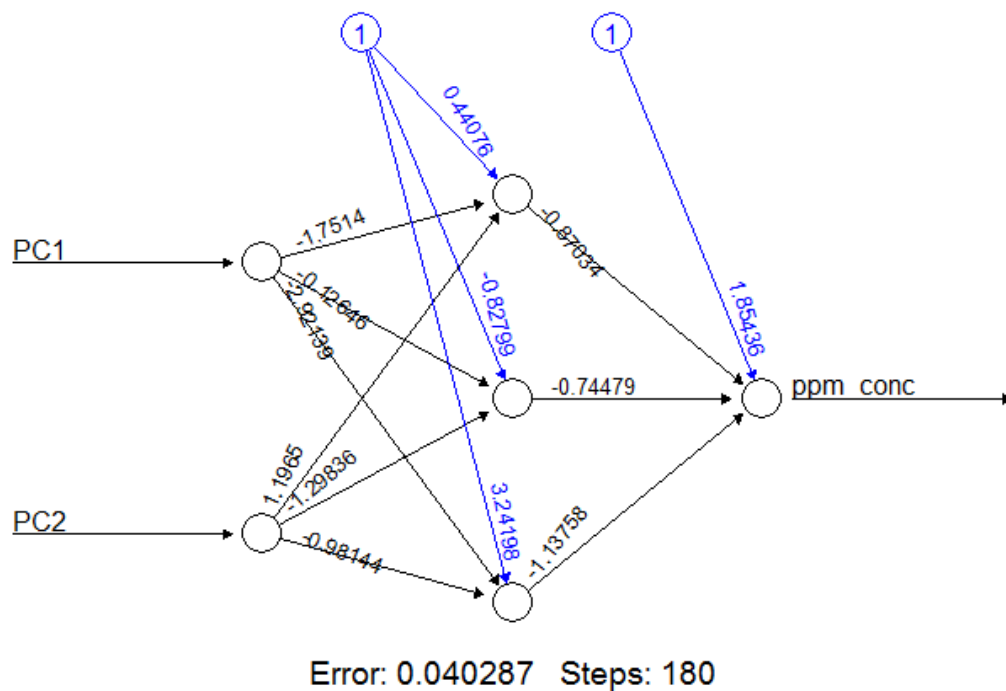


Figure 5.9: Trained model architecture for the ppm concentrations

The 2-PCs input data enhanced efficient reduction of the net architecture through lowering the number of nodes and weights in the input layer and thereby increasing the speed of the training phase. Several runs were done on the R software code during the model development until a reasonable model with an RMSEP of **4.6%** was realized.

The developed model is as shown in figure 5.10 below;

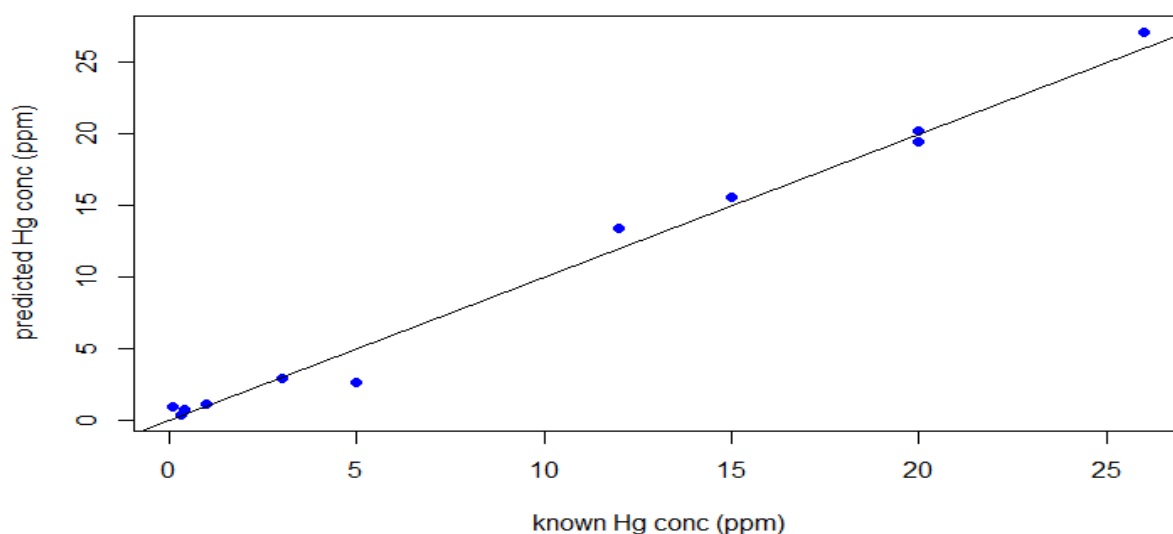


Figure 5.10: The ANN model used for quantification of Hg concentrations (ppm)

The high correlation between the predicted and known Hg concentrations depicted a good model based on the fact that it had an R squared value of 0.98, an RMSEP of 4.6% and a limit of detection equal to 3 ppm, as shown in Table 5.3.

Table 5.3: ANN model statistics for Hg concentrations (ppm)

STRUCTURE	MODEL RMSE	R² VALUE	MODEL LOD (ppm)	MODEL LOQ (ppm)
2::3:1	4.6%	0.98	3	11

Validation of the developed model for the predicted versus the measured Hg concentrations was done utilizing the test data set and the results displayed in Table 5.4 below were obtained:

Table 5.4: Sample results by using the developed ANN model for Hg concentrations (ppm)

known Hg conc (ppm)	predicted Hg conc (ppm)
0.4	< 3
20	19.5±0.9
0.3	< 3
15	15.5±0.7
12	13.4±0.6
26	27.2±1.3
1	< 3
3	2.9±0.13
20	19.5±0.16
0.1	< 3
5	< 3

5.3.5 ANN Model Results for Real Samples

Capability of the developed model was validated by using real samples spectra and the R software. The results obtained are as shown in Table 5.5.

Table 5.5: Real samples analysis results by using the ANN model for Hg concentrations (ppb)

S/NO	SAMPLE NAME	SAMPLE CODE	Hg CONC. (ppb)
1	BIO LIGHT CREAM	BIO. L	< 527
2	EXTRA WHITE	EW	< 527
3	QEI LOTION	QEI. L	< 527
4	GEL LOTION	GEL. L	731±151
5	REAL WHITE	R. W	731±151

Out of the five samples, had their Hg concentration below the detection limit for the model while two had their concentration detected within the model's quantification error of plus or minus 151 ppb. From the above results, it was noted that the developed neural net model was capable of detecting and quantifying very low Hg content in real samples (less than 2 ppb) as opposed to the conventional EDXRF method which could only detect a concentration of up to 3,000 ppb as the lowest.

5.4 EDXRF Results for Simulate and Real Samples

5.4.1 Analysis results of the simulate samples

Table 5.6 below gives EDXRF results for the 30 samples. Three spectra for each, including one from a blank sample were also obtained and later utilized in model development.

Table 5.6: EDXRF Simulate samples results (ppm)

<i>S/NO.</i>	SIMULATE CONC.	EDXRF RESULTS	<i>S/NO.</i>	SIMULATE CONC.	EDXRF RESULTS
<i>1</i>	0.1	NOT DETECTED	<i>12</i>	26	29.3
<i>2</i>	1	''	<i>13</i>	30	29.7
<i>3</i>	2	''	<i>14</i>	38	39
<i>4</i>	3	2.94	<i>15</i>	45	54
<i>5</i>	5	3.94	<i>16</i>	50	50.2
<i>6</i>	8	9.08	<i>17</i>	60	63.5
<i>7</i>	10	9.98	<i>18</i>	100	103
<i>8</i>	12	13.9	<i>19</i>	200	202
<i>9</i>	15	15.7	<i>20</i>	250	252
<i>10</i>	17	19.5	<i>21</i>	300	302
<i>11</i>	20	19.95	<i>22</i>	500	505

The detected Hg concentrations closely matched the simulate concentrations implying that stoke solution and series dilution was well done without ruling out the spectrometer's ability to accurately detect various concentrations. According to the results, the conventional EDXRF used had an LOD of 3ppm and therefore, none of the sample concentrations below this detection limit was detected by the spectrometer. However, all the spectral data for sample Hg concentrations in the ppb range were saved in a csv file and used for chemometric data analysis as well as in the development of ANN model for these concentrations.

5.4.2 Analysis results of real samples (pastes)

EDXRF results for the eight different skin lighteners analyzed are presented in Table 5.7.

Table 5.7: EDXRF paste samples results

S/NO	SAMPLE NAME	CODE	Hg CONC. (ppm)
1	BIO LIGHT CREAM	BIO. L	NOT DETECTED (ND)
2	EXTRA WHITE CREAM	EWC	ND
3	GEL LOTION	GL	ND
4	QEI LOTION	QL	ND
5	REAL WHITE	RW	ND
6	CHANDNI CREAM	CHAND.	393
7	FOREVER WHITENING CREAM	FOR. WC	2810
8	HIJAB CREAM	HIJJ.	8000

Out of the eight samples analyzed, three of them registered high levels of Hg content while the rest had nothing detected, implying that their Hg concentrations were below the spectrometer's LOD of 3 ppm.

5.4.3 Analysis results of real samples (liquids)

Spectral data consisting of ten spectra from the five samples with undetected Hg content, classed into groups A and B, colour coded blue and red respectively was saved in a csv file. A PCA analysis for this data was done on the R software code to get the following results:

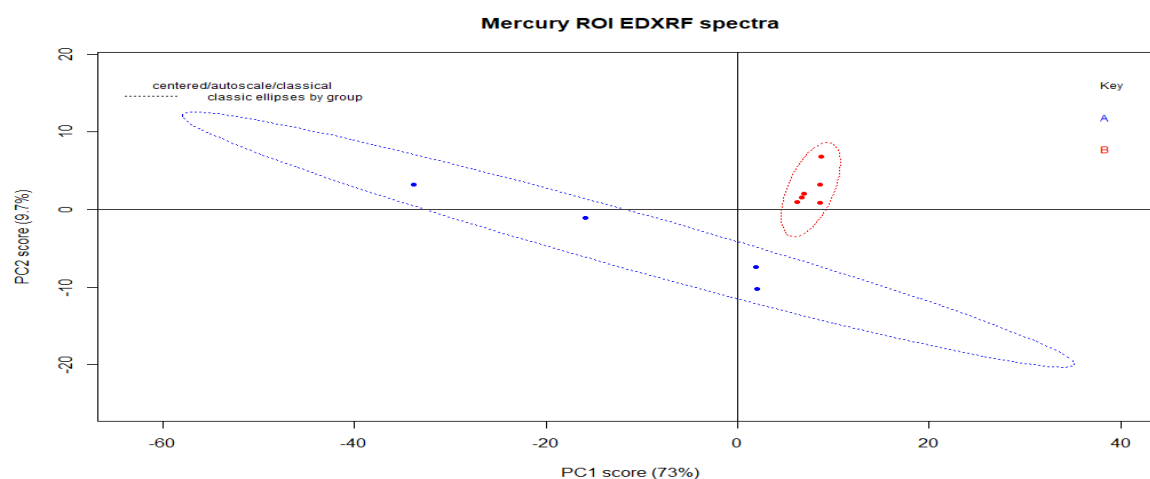


Figure 5.11: PC scores plot for real samples

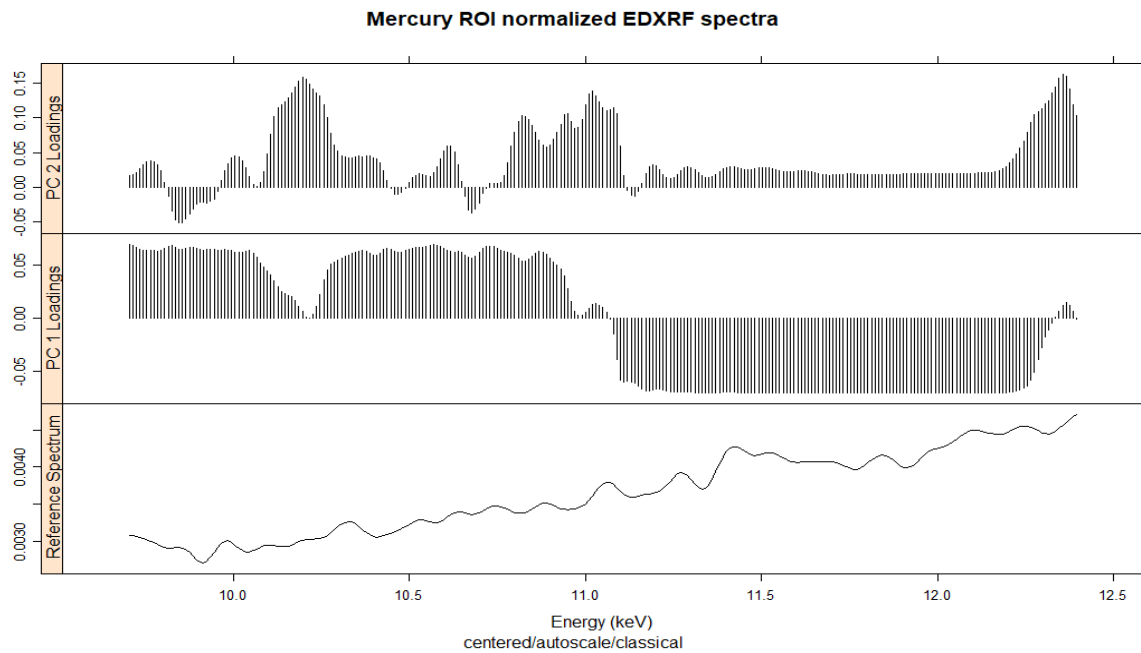


Figure 5.12: Loadings plot for real samples

It was found that 73% of the spectral energies regarding Hg concentrations for all group B samples (red) and some group A samples (blue) lying on the positive side of PC1 were fully explained. Only two samples from group A lay on the negative side of PC1. PC2 explained 9.7% of one group A sample and all group B samples lying on the positive side. Little information for three group A samples on the negative side of PC2 was given.

In this plot, all group B samples displayed high spectral energies as compared to most of group A samples, implying a high possibility of Hg content. A file containing two PCs from the ten spectra was imported into the R software. Detection and quantification of Hg present in the real samples was done by using the developed ANN model to assess and validate its performance. Below are the results.

5.5 Comparison Between Conventional EDXRF and ANN model Results

The Hg concentrations (ppm) for the simulate samples were analyzed utilizing the developed ANN model and the results compared with those acquired by direct detection using the conventional EDXRF spectrometer. Table 5.8 below shows the tabulated results.

Table 5.8: Comparison between conventional EDXRF and ANN model results for sim samples data (ppm)

SIMULATE Hg CONCENTRATION (ppm)	CONVENTIONAL EDXRF RESULTS (ppm)	ANN MODEL RESULTS (ppm)
0.1	NOT DETECTED	< 3
1	”	< 3
2	”	< 3
3	2.94	2.9±0.13
5	3.94	< 3
12	13.9	13.4±0.6
15	15.7	15.5±0.7
20	19.95	19.5±0.9
26	29.3	27.2±1.3

From the results, the model predictions closely match the ones attained through direct detection by the conventional EDXRF spectrometer. The samples which registered results less than 3ppm (the ANN model’s LOD) were those having Hg concentrations less than 3 ppm. However, the sim sample with Hg concentration of 5 ppm registered a result of < 3 ppm by the model, implying that it could have been an outlier.

Table 5.9 compares the real samples results obtained by using the ANN models with those by Conventional EDXRF.

Table 5.9: Comparison between Conventional EDXRF and ANN model results for real samples

S/NO	REAL SAMPLE NAME	EDXRF RESULTS	PPM MODEL RESULTS	PPB MODEL RESULTS
1	BIO LIGHT CREAM	NOT DETECTED	NOT DETECTED (ND)	< 527
2	EXTRA WHITE CREAM	NOT DETECTED	ND	< 527
3	GEL LOTION	NOT DETECTED	ND	731±151
4	QEI LOTION	NOT DETECTED	ND	< 527
5	REAL WHITE	NOT DETECTED	ND	731±151
6	CHANDNI CREAM	393 ppm	ND	NOT DETECTED
7	FOREVER WHITENING CREAM	2810 ppm	ND	NOT DETECTED
8	HIJAB CREAM	8000 ppm	ND	NOT DETECTED

Real samples were analyzed by using the conventional EDXRF and the developed models. Only samples with Hg concentrations above the spectrometer's LOD of 3 ppm were detected as per table 5.9 above. Samples having Hg concentrations far much below 3 ppm but above the ppb model's LOD of 527 ppb were also detected. These results display the model's capability and accuracy implying that it is a reliable tool for future utility in quantitative and qualitative analysis of beauty care products to ascertain the Hg content in them and give quality assurance.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

This research used suitably prepared simulate samples whose base matrix was identified by carrying out an EDXRF spectra analysis and PCA on several matrices. A mixture of glycerol and distilled water was chosen and used as the base matrix for this work after carrying out an investigation on several base candidate matrices; namely, pure honey, olive oil, glycerine and glycerine-distilled water mixture. Glycerol and water have commonly been used as the base formulations during the manufacture of the skin lightening creams and lotions due to the key functions they play in these beauty care products. 50 simulate samples spiked with varying Hg concentrations ranging from 5 ppb to 500 ppm were prepared using a mixture of glycerol and distilled water in a ratio of 1:1 as the base matrix. EDXRF spectra for the 50 samples was taken in triplicates, utilized in training and developing two artificial neural net models, one for Hg concentrations (ppb) and the other for ppm concentrations by employing PCA and ANNs as multivariate chemometric tools for exploratory data analysis.

The success of chemometric-assisted EDXRF spectroscopy was attained by developing the ANN models which were trained, tested and validated for rapid, direct detection and quantification of low mercury levels in skin whitening creams and lotions collected from various markets in Kenya. The ANN model developed for the ppb concentrations had an R-squared value of 0.72, an RMSEP of 20.6%, a limit of detection equal to 527 ppb and a limit of quantification of 819 ppb. When used for qualitative and quantitative analysis of the real samples, two out of the five samples tested were each found to contain 731 ± 151 ppb of mercury, a value much lower than the recommended value of 1000 ppb by the WHO. This value (731 ppb) was also lower than the LOD of 3 ppm for the conventional EDXRF spectrometer used in this research, confirming the model's reliability and accuracy for quality assurance and control of the beauty care products.

The second developed ANN model for Hg concentrations (ppm) was used to ascertain the sim samples results acquired through direct detection by using the conventional EDXRF tool for the purpose of comparing the results. It was found that most of the artificial neural net model results closely matched those by the aforesaid spectrometer given that the trained and developed ANN model had an R squared value of 0.98, an RMSEP of 4.6%, a limit of detection equal to 3ppm and a limit of quantification equal to 11 ppm.

6.2 Recommendations

The success and the results of this novel work proved that low Hg concentrations, much lower than the WHO's recommended value of 1ppm do exist in most of the beauty care products and therefore endangering the health of the users after a prolonged exposure due to continuous accumulation of small amounts of Hg in their bodies. Development of ANN models for chemometric assisted EDXRF spectroscopy based on multivariate chemometric approach in spectral data analysis is highly recommended for rapid, direct detection and quantification of low Hg content (ppb) in the beauty care products to give quality assessment, control and assurance to the users.

However, this research had challenges associated with exposure to Hg toxicity, acquisition of the beauty care products due to their expense and accessibility because they are sold in the black market. There was also a plan to ascertain the neural net model's results using the mass spectrometer which never materialized after failing to get one in good working condition. I recommend that the mass spectrometer be utilized for future validation of the capability and accuracy of such models developed in related research work.

REFERENCES

- Ahmed, A. and Hamid, M. (2016), A Survey of Female Sudanese College Students' Knowledge and Attitude towards Skin Lightening, *J Womens Heal. Issues Care* 5.
- Al-Saleh, I. (2016), Potential health consequences of applying mercury-containing skin-lightening creams during pregnancy and lactation periods, *Int. J. Hyg. Environ.*
- Amponsah, D., Sebiawu, G.E. and Voegborlo, R. (2014), Determination of amount of mercury in some selected skin-lightening creams sold in the Ghanaian market, *Int. J. Eng.*, **Vol. 3 No. 6**.
- Amponsah, D., Voegborlo, R. and Sebiawu, G.E. (2014), Determination of Amount of Hydroquinone in some selected Skin-lightening Creams sold in the Ghanaian Market., **Vol. 5 No. 6**, pp. 544–550.
- Angeyo, K.H., Gari, S., Mangala, J.M. and Mustapha, A.O. (2012), Principal component analysis-assisted energy dispersive X-ray fluorescence spectroscopy for non-invasive quality assurance characterization of complex matrix materials, *X-Ray Spectrom.*, **Vol. 41 No. 5**, available at:<https://doi.org/10.1002/xrs.2405>.
- Antwerpen, K.D.G. and Clapera, R.S. (2006), Energy Dispersive X-Ray Fluorescence : Measuring Elements in Solid and Liquid Matrices, **No. June**.
- Bailey, R.M., Stokes, S. and Bray, H. (2003), Inductively-coupled plasma mass spectrometry (ICP-MS) for dose rate determination: some guidelines for sample preparation and analysis, *Anc. TL*, **Vol. 21 No. 1**, pp. 11–15.
- Bennun, L., Greaves, E.D. and Blostein, J.J. (2002), New procedure for intensity and detection limit determination in spectral trace analysis : application for trace mercury by TXRF, pp. 289–295.
- Bose-O'Reilly, S., McCarty, K.M., Steckling, N. and Lettmeier, B. (2010), Mercury exposure and children's health, *Curr. Probl. Pediatr. Adolesc. Health Care*, **Vol. 40 No. 8**, pp. 186–215.
- Bote, D., Salvat, F., Jablonski, A. and Powell, C.J. (2009), Cross sections for ionization of K, L and M shells of atoms by impact of electrons and positrons with energies up to 1 GeV: Analytical formulas, *At. Data Nucl. Data Tables*, **Vol. 95 No. 6**
- Breternitz, M., Kowatzki, D., Langenauer, M., Elsner, P. and Fluhr, J.W. (2008), Placebo-controlled, double-blind, randomized, prospective study of a glycerol-based emollient on eczematous skin in atopic dermatitis: Biophysical and clinical evaluation, *Skin Pharmacol. Physiol.*, **Vol. 21 No. 1**
- Brouwer, P. (2006), Theory of XRF, *Almelo, Netherlands PANalytical BV*.
- Büchele, D., Chao, M., Ostermann, M., Leenen, M. and Bald, I. (2019), Multivariate chemometrics as a key tool for prediction of K and Fe in a diverse German agricultural soil-set using EDXRF, *Sci. Rep.*, **Vol. 9 No. 1**, pp. 1–11.
- von Burg, R. (1995), Inorganic mercury, *J. Appl. Toxicol.*, **Vol. 15 No. 6**, pp. 483–493.
- Campbell, L., Dixon, D.G. and Hecky, R.E. (2003), A review of mercury in Lake Victoria, East Africa: implications for human and ecosystem health, *J. Toxicol.*

Environ. Heal. Part B Crit. Rev., **Vol. 6 No. 4**, pp. 325–356.

- Çevik, U., Ergen, E., Budak, G., Karabulut, A., Tiraşoğlu, E., Apaydin, G. and Kopya, A.I. (2003), Elemental analysis of Akcaabat tobacco and its ash by EDXRF spectrometry, *J. Quant. Spectrosc. Radiat. Transf.*, **Vol. 78 No. 3**, pp. 409–415.
- Chan, T.Y. (2011), Inorganic mercury poisoning associated with skin-lightening cosmetic products, *Clin. Toxicol.*, **Vol. 49 No. 10**, pp. 886–891.
- Clevenger, W.L., Smith, B.W. and Winefordner, J.D. (2017), Trace Determination of Mercury : A Review Trace Determination of ' Mercury : A Review, **Vol. 8347 No. May**
- Dlova, N., Hamed, S.H., Tsoka-Gwegweni, J., Grobler, A. and Hift, R. (2014), Women's perceptions of the benefits and risks of skin-lightening creams in two South African communities, *J. Cosmet. Dermatol.*, **Vol. 13 No. 3**
- Dnil, E., Moldovan, Z., Albu Kaya, M.G. and Ghica, M.V. (2019), Formulation and characterization of some oil in water cosmetic emulsions based on collagen hydrolysate and vegetable oils mixtures, *Pure Appl. Chem.*
- Einax, J.W. (2005), M.J. Adams: Chemometrics in Analytical Spectroscopy, 2nd edition, *Anal. Bioanal. Chem.*, **Vol. 382 No. 4**
- Eschnauer, H., Jakob, L., Meierer, H. and Neeb, R. (1989), Use and limitations of ICP-OES in wine analysis, *Microchim. Acta*, **Vol. 99 No. 3–6**, pp. 291–298.
- Fang, Z. (1995), *Flow Injection Atomic Absorption Spectrometry*, John Wiley & Son Ltd.
- Gershenson, C. (2003), Artificial Neural Networks for Beginners, *Networks*, **Vol. cs.NE/0308**, p. 8.
- Grinyer, J., Popovic, M. and Chettle, D.R. (2007), Detection of mercury in the kidney via source-excited X-ray fluorescence †, pp. 99–103.
- Hamann, C.R., Boonchai, W., Wen, L., Sakanashi, E.N., Chu, C.Y., Hamann, K., Hamann, C.P., *et al.* (2014), Spectrometric analysis of mercury content in 549 skin-lightening products: Is mercury toxicity a hidden global health hazard?, *J. Am. Acad. Dermatol.*, **Vol. 70 No. 2**, pp. 281-287.e3.
- Hattendorf, B. and Günther, D. (2000), Characteristics and capabilities of an ICP-MS with a dynamic reaction cell for dry aerosols and laser ablation Presented at the 2000 Winter Conference on Plasma Spectrochemistry, Fort Lauderdale, FL, USA, January 10–15, 2000., *J. Anal. At. Spectrom.*, **Vol. 15 No. 9**, pp. 1125–1131.
- Jacimovic, R. and Horvat, M. (2004), Determination of total mercury in environmental and biological samples using k0-INAA, RNAA and CVAAS/AFS techniques: Advantages and disadvantages, *J. Radioanal. Nucl. Chem.*, **Vol. 259 No. 3**, pp. 385–390.
- Jalali-Heravi, M. and Kyani, A. (2004), Use of computer-assisted methods for the modeling of the retention time of a variety of volatile organic compounds: a PCA-MLR-ANN approach, *J. Chem. Inf. Comput. Sci.*, **Vol. 44 No. 4**, pp. 1328–1335.
- Kalnicky, D.J. and Singhvi, R. (2001), Field portable XRF analysis of environmental samples, *J. Hazard. Mater.*, **Vol. 83 No. 1**, pp. 93–122.

- Kamagaju, L., Morandini, R., Gahongayire, F., Stévigny, C., Ghanem, G., Pirotte, G. and Duez, P. (2016), Survey on skin-lightening practices and cosmetics in Kigali, Rwanda, *Int. J. Dermatol.*, **Vol. 55 No. 1**
- Kamakshi, R. (2011), Fairness via formulations: a review of cosmetic skin-lightening ingredients., *J. Cosmet. Sci.*, **Vol. 63 No. 1**, pp. 43–54.
- Kaniu, M.I., Angeyo, K.H., Mangala, M.J., Mwala, A.K. and Bartilol, S.K. (2011), Feasibility for chemometric energy dispersive X-ray fluorescence and scattering (EDXRFS) spectroscopy method for rapid soil quality assessment, *X-Ray Spectrom.*, **Vol. 40 No. 6**, pp. 432–440.
- Kaniu, M.I., Angeyo, K.H., Mwala, A.K. and Mwangi, F.K. (2012), Energy dispersive X-ray fluorescence and scattering assessment of soil quality via partial least squares and artificial neural networks analytical modeling approaches, *Talanta*, **Vol. 98**, pp. 236–240.
- Langford, N.J. and Ferner, R.E. (1999), Toxicity of mercury, *J. Hum. Hypertens.*, **Vol. 13 No. 10**, pp. 651–656.
- Liang, L., Gilkeson, J., Swain, E., Bennett, E., Li, M., Deng, M. and Pang, P. (2013), A Pilot Survey of Mercury in Drugs, Cosmetics and Household Products Using Reliable Analytical Methods, *J. Cosmet. Dermatological Sci. Appl.*, **Vol. 3 No. December**, pp. 256–262.
- Luo, L. (2006), Chemometrics and its applications to X-ray spectrometry, **No. May**, pp. 215–225.
- Mahé, A. (2014), The practice of skin-bleaching for a cosmetic purpose in immigrant communities, *J. Travel Med.*
- Mahé, A., Perret, J.L., Ly, F., Fall, F., Rault, J.P. and Dumont, A. (2007), The cosmetic use of skin-lightening products during pregnancy in Dakar, Senegal: a common and potentially hazardous practice, *Trans. R. Soc. Trop. Med. Hyg.*, **Vol. 101 No. 2**, pp. 183–187.
- Maina, C. (1997), *A Simple Non-Destructive Technique for the Analysis of Mercury in Creams*, University of Nairobi.
- Maneli, M.H., Wiesner, L., Tinguely, C., Davids, L.M., Spengane, Z., Smith, P., Van Wyk, J.C., *et al.* (2016), Combinations of potent topical steroids, mercury and hydroquinone are common in internationally manufactured skin-lightening products: A spectroscopic study, *Clin. Exp. Dermatol.*, **Vol. 41 No. 2**, pp. 196–201.
- Marguí, E., Queralt, I. and Hidalgo, M. (2009), Application of X-ray fluorescence spectrometry to determination and quantitation of metals in vegetal material, *TrAC Trends Anal. Chem.*, **Vol. 28 No. 3**, pp. 362–372.
- Marguí, E., Zawisza, B. and Sitko, R. (2014), Trends in Analytical Chemistry Trace and ultratrace analysis of liquid samples by X-ray fluorescence spectrometry, **Vol. 53**, pp. 73–83.
- Melquiades, F., Parreira, P., Endo, L. and Santos, G. (2015), Portable EDXRF for Quality Assurance of Cosmetics, *Cosmetics*.
- Murphy, T., Guo, J., Darell, K.I., Slotton, K.W., Lean, D. and Lim, S. (2013), Emerging

- Problems with Mercury in Cambodia, *Glob. Heal. Perspect.*, **Vol. 1 No. 2**, pp. 113–134.
- Murphy, T., Irvine, K., Bayen, S., Kelly, B.C. and Murphy, T. (2012), Application of Handheld X-Ray Fluorescence Analyzers to Identify Mercury in Skin-Whitening Creams in Cambodia, **Vol. 2 No. 3**, pp. 14–21.
- Murphy, T., Kim, S., Chanra, P., Lim, S., Wilson, K., Irvine, K.N., Slotton, D.G., *et al.* (2015), Mercury Contamination of Skin-whitening Creams in Phnom Penh, Cambodia, *J. Heal. Pollut.*, **Vol. 5 No. 9**, pp. 33–46.
- Murphy, T., Slotton, D.G., Irvine, K., Sukontason, K. and Goldman, C.R. (2009), Mercury Contamination of skin whiteners in cambodia, *Hum. Ecol. Risk Assess.*, **Vol. 15 No. 6**, pp. 1286–1303.
- Nagata, N., Peralta-zamora, P.G., Poppi, R.J., Perez, C.A. and Bueno, M.I.M.S. (2006), Multivariate calibrations for the SR-TXRF determination of trace concentrations of lead and arsenic in the presence of bromine, **No. October 2005**, pp. 79–84.
- Nguyen, T.H., Boman, J., Leermakers, M. and Baeyens, W. (1998), Mercury Determination in Environmental Samples Using EDXRF and CV-AAS, **Vol. 27 No. March 1997**, pp. 277–282.
- Nolan, A. (2005), Chemometrics in Analytical Spectroscopy by M. J. Adams, *Environ. Chem.*, **Vol. 2 No. 2**
- Okonda, J.J., Angeyo, K.H., Mangala, J.M. and Kisia, S.M. (2017), A nested multivariate chemometrics based calibration strategy for direct trace biometal analysis in soft tissue utilizing Energy Dispersive X-Ray Fluorescence (EDXRF) and scattering spectrometry, *Appl. Radiat. Isot.*
- Orisakwe, O.E. and Otaraku, J.O. (2013), Metal concentrations in cosmetics commonly used in Nigeria, *Sci. World J.*, **Vol. 2013**
- Palmer, R.B., Godwin, D.A. and McKinney, P.E. (2000), Transdermal kinetics of a mercurous chloride beauty cream: An in vitro human skin analysis, *J. Toxicol. - Clin. Toxicol.*, **Vol. 38 No. 7**
- Peltzer, K., Pengpid, S. and James, C. (2016), The globalization of whitening: Prevalence of skin lighteners (or bleachers) use and its social correlates among university students in 26 countries, *Int. J. Dermatol.*, **Vol. 55 No. 2**
- Pessanha, S., Guilherme, A. and Carvalho, M.L. (2009), Comparison of matrix effects on portable and stationary XRF spectrometers for cultural heritage samples, *Appl. Phys. A Mater. Sci. Process.*, **Vol. 97 No. 2**
- Quispe, C.A.G., Coronado, C.J.R. and Carvalho, J.A. (2013), Glycerol: Production, consumption, prices, characterization and new trends in combustion, *Renew. Sustain. Energy Rev.*, Elsevier, **Vol. 27 No. October**, pp. 475–493.
- Reinholds, I., Bartkevics, V. and Silvis, I. (2015), Analytical techniques combined with chemometrics for authentication and determination of contaminants in condiments: A review, *Anal.*
- Revenko, A.G. (2018), X-Ray fluorescence analysis in biology and medicine, *Anal. i Kontrol*, **Vol. 24 No. 4**

- Roldán, C., Murcia-Mascarós, S., Ferrero, J., Villaverde, V., López, E., Domingo, I., Martínez, R., *et al.* (2010), Application of field portable EDXRF spectrometry to analysis of pigments of Levantine rock art, *X-Ray Spectrom.*, **Vol. 39 No. 3**, pp. 243–250.
- Rousseau, R.M. (2006), Corrections for matrix effects in X-ray fluorescence analysis-A tutorial, *Spectrochim. Acta - Part B At. Spectrosc.*
- Rutan, S.C. (1996), Chemometrics in analytical spectroscopy, *Anal. Chim. Acta*, **Vol. 336 No. 1–3**
- Sah, R.C. and Charitra, R. (2012), Poisonous Cosmetics: The problem of mercury in skin whitening creams in Nepal, Sah, Ram Charitra. 2012. Poisonous Cosmetics, the Problem of Mercury in Skin Whitening Creams in Nepal, vi+10. Kathmandu: CEPHEDE.
- Sarkar, R., Kaur, C., Bhalla, M. and Kanwar, A. (2002), The Combination of Glycolic Acid Peels With a Topical Regimen in the Treatment of Melasma in Dark-Skinned Patients: A Comparative Study, *Dermatologic Surg.*
- Shibata, Y., Suyama, J., Kitano, M. and Nakamura, T. (2009), X-ray fluorescence analysis of Cr , As , Se , Cd , Hg , and Pb in soil using pressed powder pellet and loose powder methods, **No. April**
- Sichangi, E.K., Angeyo, H.K., Dehayem-Kamadjeu, A. and Mangala, M. (2018), Hybridized robust chemometrics approach for direct rapid determination of trace biometals in tissue utilizing energy dispersive X-ray fluorescence and scattering (EDXRFS) spectrometry, *Radiat. Phys. Chem.*, **Vol. 153**
- Višnjevec, A.M., Kocman, D. and Horvat, M. (2014), Human mercury exposure and effects in Europe, *Environ. Toxicol. Chem.*, **Vol. 33 No. 6**, pp. 1259–1270.
- Wobrauschek, P., Strelj, C. and Lindgren, E.S. (2010a), Energy Dispersive, X-ray Fluorescence Analysis, *Encycl. Anal. Chem. (R.A. Meyers Ed.)*, **Vol. 2000 No. 206**, pp. 1–17.
- Wobrauschek, P., Strelj, C. and Lindgren, E.S. (2010b), Fluorescence Analysis, pp. 1–17.
- Wold, S.S.M. and Eriksson, L. (2001), PLS-Regression: A Basic Tool of Chemometrics, pp. 109–130.
- Worley, B., Halouska, S. and Powers, R. (2013), Utilities for quantifying separation in PCA/PLS-DA scores plots, *Anal. Biochem.*, **Vol. 433 No. 2**, pp. 102–104.

APPENDICES

Appendix 1: Simulate sample preparation pictograms



Figure A 1.1: Pure glycerol and flasks



Figure A 1.2: Sample cups and prepared samples



Figure A 1.3: The EDXRFS sample changer

Appendix 2: Real samples



Figure A 2.1: The analyzed real samples



Figure A 2.2: Some diluted real samples

Appendix 3: R-Scripts for the Multivariate Analysis Techniques

R-Script for PCA analysis of Hg concentrations (ppb)

```
library(ChemoSpec)
library("FactoMineR")
library(dplyr)
## Load dataset
# spec2=read.csv("musau_900s_ppb_simulates_classed.csv",check.names=FALSE)
## Subtraction of blank matrix
# analspec2 <- sweep(spec2[,3:91],1,spec2[,2])
# write.infile(analspec2,file="musau_900s_ppb_simulates_classed_wblank.csv", sep =
",")

# Reading the matrix data file stored in the working directory
# sample coding: ppb -> simulated samples with ppb concentrations, ppm -> simulated
samples with ppm concentrations,
# real -> skin lightening products bought from the market, bl -> blank samples
spec <- matrix2SpectraObject(gr.crit = c("ppb1", "ppb2", "ppb3"), gr.cols = c("blue",
"green", "red"),
                             freq.unit = "Energy (keV)",
                             int.unit = "Photons count",
                             descrip = "Mercury Spectra", in.file =
"musau_900s_ppb_simulates_classed_ver2_matrix_cor.csv",
                             out.file = "mydata", chk = TRUE, sep = ",", dec = ".")
# Summarizing the data
sumSpectra(spec)
# Plotting the full EDXRF spectra
plotSpectra(spec,
             main = "Full EDXRF Spectra",
             which = c(1,14,37),
```

```

yrange = c(-0.25,3600),
offset = 800,
lab.pos = 10)

# Feature selection: Select the mercury (Hg) peaks region of interest (ROI): 9.6 - 12.4 keV
roispec <- removeFreq(spec, rem.freq = spec$freq > 12.3
| spec$freq < 9.7)

# Feature selection: Mercury (Hg) L_alpha peak 9.6 - 10.3 keV
roispec1 <- removeFreq(spec, rem.freq = spec$freq > 10.3
| spec$freq < 9.7)

# Feature selection: Mercury (Hg) L_beta peak 11.5 - 12.3 keV
roispec2 <- removeFreq(spec, rem.freq = spec$freq > 12.3
| spec$freq < 11.5)

# Feature selection: Mercury (Hg) L_alpha & L_beta peaks 9.6 - 10.3 & 11.5 - 12.3 keV
roispec3 <- removeFreq(roispec, rem.freq = roispec$freq > 10.3
& roispec$freq < 11.5)

# Plotting the Hg ROI spectra for ppb samples
plotSpectra(roispec3,
main = "Mercury ROI Spectra - ppb concentrations",
which = c(10,30,50,60,80),
yrange = c(0.4,1.5),
offset = 0.15,
lab.pos = 1.8)

# Normalizing the Hg ROI spectra
roispec_norm<-normSpectra(roispec)
roispec1_norm<-normSpectra(roispec1)
roispec2_norm<-normSpectra(roispec2)
roispec3_norm<-normSpectra(roispec3)

# Plotting the normalized partial (ROI) spectra
plotSpectra(roispec3_norm,

```

```

    main = "Normalized ROI Spectra - ppm",
    which = c(10,30,50,60,80),
    yrange = c(0,1),
    offset = 0.2,
    lab.pos = 10.5)
# PCA analysis of the Hg ROI spectra spectra
pca<-c_pcaSpectra(roispec_norm,
    choice = "autoscale",
    cent = TRUE)
plotScores(roispec_norm, pca,
    main = "Hg ROI (9.6 - 12.4 keV): Mat Eff Corrected by X-ray Scatter",
    pcs = c(1,2),
    ellipse = "none",
    tol = "none")
abline(h=0,v=0)
plotLoadings(roispec_norm, pca,
    main = "Hg ROI (9.6 - 12.4 keV): Mat Eff Corrected by X-ray Scatter",
    loads = c(1, 2),
    ref = 89,
    tol = "none")
# To check pca outliers
diagnostics <- pcaDiag(roispec1, pca,
    pcs = 2,
    plot = "OD")
# Scree plot
plot(pca, type = "l")

#Saving the PCA Scores Data
pca_scores <- pca[["x"]]
write.csv(pca_scores, "./pca_results/pca
results_musau_ppb_data_HgLa&Lb_matrix_corrected_30.03.21.csv")

```

```

# summary & output of pca results
summary(pca)
res <- summary(pca)
write.infile(res,
             file="./pca_results/all_pca
results_musau_ppb_data_HgLa&Lb_matrix_corrected_30.03.21.csv", sep = ",")

```

Training Algorithm for the ANN Model with Hg concentrations (ppb)

```

# Script for performing ANN for Julius Musau EDXRF Spectral dataset for Hg Peaks
(PCA results)

```

```

# Spectra range: 9.6 - 12.4 keV

```

```

# #evtools::install_github("bips-hb/neuralnet")

```

```

# library(devtools)

```

```

source_url('https://gist.githubusercontent.com/fawda123/7471137/raw/466c1474d0a505ff
044412703516c34f1a4684a5/nnet_plot_update.r')

```

```

library(neuralnet)

```

```

library(dplyr)

```

```

library(caret)

```

```

#grid::current.viewport()

```

```

##Setting the seed so that we get same results each time we run the neural nets again

```

```

set.seed(100)

```

```

# Load dataset

```

```

data = read.csv("2PCs_NN_data_pca_results_matrix_effects
corrected.csv",check.names=TRUE)

```

```

data <- data[sample(nrow(data)), ] # Shuffle data

head(data, 3)

##### For looking at Structure of EDXRF spectra data

str(data)

#apply(data,2,function(x) sum(is.na(x)))

# EDA to help understand how data is distributed

#par(mfrow=c(2,2))

#plot(data$PC1, data$ppb_conc, cex=1)

#plot(data$PC2, data$ppb_conc, cex=1)

#plot(data$PC3, data$ppb_conc, cex=1)

#plot(data$PC4, data$ppb_conc, cex=1)

# Random sampling

samplesize = 0.80 * nrow(data)

#set.seed(100)

index = sample( seq_len ( nrow ( data ) ), size = samplesize )

# Create training and test set

datatrain = data[ index, ]

datatest = data[ -index, ]

## Scale data for neural network

maxValue <- apply(data, 2, max)

```

```

minValue <- apply(data, 2, min)

data_scaled<-as.data.frame(scale(data,center = minValue,
                                scale =maxValue-minValue))

# creating scaled training and test set

trainDF<-data_scaled[index,]

testDF<-data_scaled[-index,]

#trainDF <- as.matrix(trainDF, byrow=TRUE)

#trainDF <- t(trainDF)

allVars<-colnames(data)

predictorVars<-allVars[!allVars%in%"ppb_conc"]

predictorVars<-paste(predictorVars,collapse = "+")

form=as.formula(paste("ppb_conc~",predictorVars,collapse = "+"))

neuralModel<-neuralnet(formula =form,

                        data =trainDF,

                        hidden = c(2,2),

                        lifesign.step = 1000, algorithm = "rprop+", err.fct = "sse",

                        act.fct = "logistic",

                        linear.output = TRUE)

plot(neuralModel,rep = "best")

# plot(neuralModel,col.hidden = 'darkgreen',

```

```

# col.hidden.synapse = 'darkgreen',

# show.weights = T,

# information = T,

# fill = 'lightblue')

#Neural Network optimization

ctrl <- trainControl(method="cv", number=10) #Cross validation method, other methods
available

gridNN <- expand.grid(layer1 = c(2:7),

                      layer2 = c(2:6),

                      layer3 = c(0:2)) #Change to suitable number of layers and neurons

#Train the NN - Caret training function using the NN algorithm in Neuralnet

NN <- train(form,

            data = trainDF,

            trControl = ctrl,

            method = "neuralnet",

            tuneGrid = gridNN)

#summary of training, the best model being the one with the lowest RMSE

print(NN)

predNN <- predict(NN, newdata = testDF[,c(2:3)]*(max(data$ppb_conc) -
min(data$ppb_conc)) +

min(data$ppb_conc)

```



```

plot(datatest$ppb_conc, predNN, col='blue', pch=16, ylab = "predicted Hg conc (ppb)",

      xlab = "known Hg conc (ppb)")

abline(0,1)

Hg_values <- cbind(datatest$ppb_conc, predNN)

colnames(Hg_values) <- c("known Hg conc (ppb)", "predicted Hg conc (ppb)")

write.csv(Hg_values, file = "./Results/2pcs_NN results_pca matrix effects corrected.csv")

RMSE.neuralModel = (sum((datatest$ppb_conc - predNN)^2) / nrow(datatest)) ^ 0.5

save(NN, file = "./Results/2pcs_NN model_ppb matrix corrected.rda")

## Prediction using neural network

predict_testDF <- neuralnet::compute(neuralModel, testDF[,c(2:3)])

predict_testDF <- (predict_testDF$net.result * (max(data$ppb_conc) -
min(data$ppb_conc))) +

min(data$ppb_conc)

plot(datatest$ppb_conc, predict_testDF, col='blue', pch=16, ylab = "predicted Hg conc
(ppb)",

      xlab = "known Hg conc (ppb)")

abline(0,1)

Hg_values <- cbind(datatest$ppb_conc, predict_testDF)

colnames(Hg_values) <- c("known Hg conc (ppb)", "predicted Hg conc (ppb)")

RMSE.neuralModel = (sum((datatest$ppb_conc - predict_testDF)^2) / nrow(datatest)) ^
0.5

```

R Script for PCA analysis of Hg Concentrations (ppm)

```
library(ChemoSpec)
library("FactoMineR")
library(dplyr)
# Load dataset
# spec2=read.csv("musau_900s_ppb_simulates_classed.csv",check.names=FALSE)
## Subtraction of blank matrix
# analspec2 <- sweep(spec2[,3:91],1,spec2[,2])
# write.infile(analspec2,file="musau_900s_ppb_simulates_classed_wblank.csv", sep =
",")
# Reading the matrix data file stored in the working directory
# sample coding: ppb -> simulated samples with ppb concentrations, ppm -> simulated
samples with ppm concentrations,
# real -> skin lightening products bought from the market, bl -> blank samples
spec <- matrix2SpectraObject(gr.crit = c("ppm1", "ppm2", "ppm3"), gr.cols = c("blue",
"green", "red"),
                             freq.unit = "Energy (keV)",
                             int.unit = "Photons count",
                             descrip = "Mercury Spectra", in.file =
"musau_900s_ppm_simulates_classed_0 - 30ppm.csv",
                             out.file = "mydata", chk = TRUE, sep = ",", dec = ".")
# Summarizing the data
sumSpectra(spec)
# Plotting the full EDXRF spectra
plotSpectra(spec,
             main = "Full EDXRF Spectra",
             which = c(1,14,37),
             yrange = c(-0.25,3600),
             offset = 800,
             lab.pos = 10)
```

```

# Feature selection: Select the mercury (Hg) peaks region of interest (ROI): 9.6 - 12.4 keV
roispec <- removeFreq(spec, rem.freq = spec$freq > 12.3
                    | spec$freq < 9.7)
# Feature selection: Mercury (Hg) L_alpha peak 9.6 - 10.3 keV
roispec1 <- removeFreq(spec, rem.freq = spec$freq > 10.3
                    | spec$freq < 9.7)
# Feature selection: Mercury (Hg) L_beta peak 11.5 - 12.3 keV
roispec2 <- removeFreq(spec, rem.freq = spec$freq > 12.3
                    | spec$freq < 11.5)
# Feature selection: Mercury (Hg) L_alpha & L_beta peaks 9.6 - 10.3 & 11.5 - 12.3 keV
roispec3 <- removeFreq(roispec, rem.freq = roispec$freq > 10.3
                    & roispec$freq < 11.5)
# Plotting the Hg ROI spectra for ppb samples
plotSpectra(roispec3,
            main = "Mercury ROI Spectra - ppb concentrations",
            which = c(10,30,50,60,80),
            yrange = c(0.4,1.5),
            offset = 0.15,
            lab.pos = 1.8)
# Normalizing the Hg ROI spectra
roispec_norm<-normSpectra(roispec)
roispec1_norm<-normSpectra(roispec1)
roispec2_norm<-normSpectra(roispec2)
roispec3_norm<-normSpectra(roispec3)
# Plotting the normalized partial (ROI) spectra
plotSpectra(roispec3_norm,
            main = "Normalized ROI Spectra - ppm",
            which = c(10,30,50,60,80),
            yrange = c(0,1),
            offset = 0.2,

```

```

lab.pos = 10.5)
#-----
# PCA analysis of the Hg ROI spectra spectra
pca<-c_pcaSpectra(roispec1_norm,
                 choice = "autoscale",
                 cent = TRUE)
plotScores(roispec1_norm, pca,
           main ="Hg L_Alpha Peak (9.6 - 10.3 keV): Normalized Spectra",
           pcs = c(1,2),
           ellipse = "none",
           tol = "none")
abline(h=0,v=0)
plotLoadings(roispec1_norm, pca,
             main = "Hg L_Alpha Peak (9.6 - 10.3 keV): Normalized Spectra",
             loads = c(1, 2),
             ref = 51,
             tol = "none")
# To check pca outliers
diagnostics <- pcaDiag(roispec_norm, pca,
                      pcs = 2,
                      plot = "OD")
# Scree plot
plot(pca, type = "l")
#Saving the PCA Scores Data
pca_scores <- pca[["x"]]
write.csv(pca_scores, "./pca_results/pca
results_musau_ppm_data_HgLa_normalized_25.04.21.csv")
# summary & output of pca results
summary(pca)
res <- summary(pca)
write.infile(res,

```

```
file="./pca_results/all_pca
results_musau_ppm_data_HgLa_normalized_25.04.21.csv", sep = ",")
```

Training Algorithm for ANN model with Hg concentrations (ppm)

```
# Script for performing ANN for Julius Musau EDXRF Spectral dataset for Hg Peaks
(PCA results)

# Spectra range: 9.6 - 12.4 keV

#devtools::install_github("bips-hb/neuralnet")

#library(devtools)

#source_url('https://gist.githubusercontent.com/fawda123/7471137/raw/466c1474d0a505f
f044412703516c34f1a4684a5/nnet_plot_update.r')

library(neuralnet)

library(dplyr)

library(caret)

#grid::current.viewport()

##Setting the seed so that we get same results each time we run the neural nets again
set.seed(100)

# Load dataset

data = read.csv("2PCs_NN_data_pca_results_ppm_la
normalized.csv",check.names=TRUE)

data <- data[sample(nrow(data)), ] # Shuffle data

head(data, 3)

##### For looking at Structure of EDXRF spectra data
str(data)

#apply(data,2,function(x) sum(is.na(x)))

# EDA to help understand how data is distributed

#par(mfrow=c(2,2))

#plot(data$PC1, data$ppm_conc, cex=1)

#plot(data$PC2, data$ppm_conc, cex=1)

#plot(data$PC3, data$ppm_conc, cex=1)
```

```

#plot(data$PC4, data$ppm_conc, cex=1)

# Random sampling
samplesize = 0.80 * nrow(data)
#set.seed(100)
index = sample( seq_len ( nrow ( data ) ), size = samplesize )
# Create training and test set
datatrain = data[ index, ]
datatest = data[ -index, ]
## Scale data for neural network
maxValue <- apply(data, 2, max)
minValue <- apply(data, 2, min)
data_scaled<-as.data.frame(scale(data,center = minValue,
                                scale =maxValue-minValue))
# creating scaled training and test set
trainDF<-data_scaled[index,]
testDF<-data_scaled[-index,]

#trainDF <- as.matrix(trainDF, byrow=TRUE)
#trainDF <- t(trainDF)

allVars<-colnames(data)
predictorVars<-allVars[!allVars%in%"ppm_conc"]
predictorVars<-paste(predictorVars,collapse = "+")
form=as.formula(paste("ppm_conc~",predictorVars,collapse = "+"))
neuralModel<-neuralnet(formula =form,
                        data =trainDF,
                        hidden = c(3),
                        lifesign.step = 1000, algorithm = "rprop+", err.fct = "sse",
                        act.fct = "logistic",
                        linear.output = TRUE)

```

```

plot(neuralModel,rep = "best")
# plot(neuralModel,col.hidden = 'darkgreen',
#   col.hidden.synapse = 'darkgreen',
#   show.weights = T,
#   information = T,
#   fill = 'lightblue')
#Neural Network optimization

ctrl <- trainControl(method="cv", number=10) #Cross validation method, other methods
available

gridNN <- expand.grid(layer1 = c(1:3),
                      layer2 = c(0:2),
                      layer3 = c(0:1)) #Change to suitable number of layers and neurons

#Train the NN - Caret training function using the NN algorithm in Neuralnet
NN <- train(form,
            data = trainDF,
            trControl = ctrl,
            method = "neuralnet",
            tuneGrid = gridNN)

#summary of training, the best model being the one with the lowest RMSE
print(NN)

predNN <- predict(NN, newdata = testDF[,c(2:3)]*(max(data$ppm_conc) -
min(data$ppm_conc)) +
min(data$ppm_conc)

plot(datatest$ppm_conc, predNN, col='blue', pch=16, ylab = "predicted Hg conc (ppm)",
      xlab = "known Hg conc (ppm)")
abline(0,1)
Hg_values <- cbind(datatest$ppm_conc, predNN)
colnames(Hg_values) <- c("known Hg conc (ppm)", "predicted Hg conc (ppm)")
RMSE.neuralModel = (sum((datatest$ppm_conc - predNN)^2) / nrow(datatest)) ^ 0.5
#save(NN, file = "./Results/2pcs_NN caret model.rda")
#Prediction using neural network

```

```

predict_testDF <- neuralnet::compute(neuralModel, testDF[,c(2:3)])

predict_testDF <- (predict_testDF$net.result * (max(data$ppm_conc) -
min(data$ppm_conc))) +
  min(data$ppm_conc)

plot(datatest$ppm_conc, predict_testDF, col='blue', pch=16, ylab = "predicted Hg conc
(ppm)",
  xlab = "known Hg conc (ppm)")

abline(0,1)

Hg_values <- cbind(datatest$ppm_conc, predict_testDF)

colnames(Hg_values) <- c("known Hg conc (ppm)", "predicted Hg conc (ppm)")

write.csv(Hg_values, file = "./Results/2pcs_NN results_ppm la peak normalized.csv")

RMSE.neuralModel = (sum((datatest$ppm_conc - predict_testDF)^2) / nrow(datatest)) ^
0.5

save(neuralModel, file = "./Results/2pcs_NN model_ppm la peak normalized.rda")

```

R script for calculating the LoD and LoQ for the ANN model with Hg concentrations (ppb).

```

# Script for Calculating LoD & LoQ on Measurements Predicted by ANN Model

data<- read.csv("2pcs_NN results_ppm la peak normalized_lod.csv")

library(dplyr)

library(chemCal)

m <- lm(predicted ~ actual, data = data)

## Limit of detection

lod.din <- lod(m, alpha = 0.01, beta = 0.5, method = "din")

round(lod.din$predicted, 2)

## Limit of quantification

loq <- loq(m, alpha = 0.01)

round(loq$predicted, 4)

##### Plotting Measurements with Error Bars

stats1 <- group_by(data, actual) %>%

```



```

summarise(
  count = n(),
  mean = mean(predicted , na.rm = TRUE),
  sd = sd(predicted , na.rm = TRUE)
stats1 <- as.data.frame(stats1)
x<-stats1$actual
y<-stats1$mean
y.sd <- stats1$sd
plot(x, y, ylim=c(0,25), xlab="Predicted Hg (ppm) Concentrations",
  ylab="Actual Hg (ppm) Concentrations",
  main="ANN Model Performance on Simulate Samples: Test Set",
  pch=16, cex=1.1)

# Add error bars
arrows(x0=x, y0=y-y.sd, x1=x, y1=y+y.sd, code=3, angle=90, length=0.1)
abline(0,1)

```

Script for calculating LoD and LoQ for the ANN model with Hg concentrations (ppm)

```

# Script for Calculating LoD & LoQ on Measurements Predicted by ANN Model
data<- read.csv("2pcs_NN results_ppm la peak normalized_lod.csv")
library(dplyr)
library(chemCal)

m <- lm(predicted ~ actual, data = data)

## Limit of detection
lod.din <- lod(m, alpha = 0.01, beta = 0.5, method = "din")
round(lod.din$predicted, 2)

## Limit of quantification

```

```

loq <- loq(m, alpha = 0.01)
round(loq$predicted, 4)

#### Plotting Measurements with Error Bars
stats1 <-group_by(data, actual) %>%
  summarise(
    count = n(),
    mean = mean(predicted , na.rm = TRUE),
    sd = sd(predicted , na.rm = TRUE)
  )

stats1 <- as.data.frame(stats1)
x<-stats1$actual
y<-stats1$mean
y.sd <- stats1$sd

plot(x, y, ylim=c(0,25), xlab="Predicted Hg (ppm) Concentrations",
      ylab="Actual Hg (ppm) Concentrations",
      main="ANN Model Perfomance on Simulate Samples: Test Set",
      pch=16, cex=1.1)
# Add error bars
arrows(x0=x, y0=y-y.sd, x1=x, y1=y+y.sd, code=3, angle=90, length=0.1)
abline(0,1)

```