DETERMINATION OF COLORING SUBSTANCES IN CLEANING PRODUCTS BY CHEMOMETRICS METHODS

A Thesis Submitted to
the Graduate School of
İzmir Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in Chemistry

by
Çetin TÖREMEN

July 2024
İZMİR

We approve the thesis of **Çetin TÖREMEN**

Examining Committee Members:

Prof. Dr. Durmuş ÖZDEMİR
Department of Chemistry, İzmir Institute of Technology

Prof. Dr. Talat YALÇIN
Department of Chemistry, İzmir Institute of Technology

Assoc. Prof. Dr. Füsun PELİT
Department of Chemistry, Ege University

12 July 2024

Prof. Dr. Durmuş ÖZDEMİR
Supervisor, Department of Chemistry,
İzmir Institute of Technology

Prof. Dr. Gülşah Şanlı Mohamed
Head of the Department of Chemistry

Prof. Dr. Mehtap EANES
Dean of the Graduate School

ACKNOWLEDGMENTS

I would like to express my deepest appreciation to my thesis advisor, Prof. Dr. Durmuş ÖZDEMİR for his constant support guidance and encouraging approach throughout the learning and writing process throughout my master's program.

I am deeply thankful to the members of my lab colleagues Cihan TEKİN and Dilek TEPELİ for their support.

I would also like to thank VIKING Cleaning and Cosmetics Inc. for providing materials and equipment for this project.

I would like to thank VIKING Research and Development Executive Murat DEMİR, for their assistance in using UV-Vis in the analysis of samples.

Words cannot express my gratitude to my lovely family Cennet TÖREMEN, Adnan TÖREMEN and Ezgi TÖREMEN for their love and support during this process. Without their encouragement and motivation, this journey would not have been possible.

ABSTRACT

DETERMINATION OF COLORING SUBSTANCES IN CLEANING PRODUCTS BY CHEMOMETRICS METHODS

In this thesis study, surface cleaner samples were studied. UV-Visible Spectrophotometry was applied to the samples prepared by adding dye amounts determined at various concentrations into these samples. Partial least squares (PLS), simple least squares regression (SLR) and genetic inverse least squares (GILS) methods have been successfully applied to the visible spectra of prepared samples for the quantitative determination of these coloring substances. A total of 35 samples were prepared as binary and ternary mixtures of these dyes along with single dye containing cleaning products. Among these 35 samples, 29 of them were chosen as calibration set and the remaining 6 samples were used as independent validation set. Absorption spectra between 400-700 nm were recorded on the prepared set and SLR, PLS and GILS methods were applied to the obtained spectral data and the resulting predictions were compared with actual values. The regression coefficients (R^2) and standard error cross-validation (SECV) values of these methods (PLS and GILS) and the standard error of prediction (SEP) were found, the results were evaluated, the accuracy of the chemometrics modelling was reviewed, and the predictions and references were compared. With these analysis methods, the concentration of dyes in the cleaning products can be determined by using chemometrics methods combined with UV-Visible Spectroscopy.

ÖZET

TEMİZLİK ÜRÜNLERİNDE RENKLENDİRİCİ MADDELERİN KEMOMETRİK YÖNTEMLERLE TAYİNİ

Bu tez çalışmasında yüzey temizleyici örnekleri incelenmiştir. Bu numunelerin içerisine çeşitli konsantrasyonlarda belirlenen boya miktarları eklenerek hazırlanan numunelere UV-Görünür Bölge Spektrofotometrisi uygulanmıştır. Kısmi en küçük kareler (PLS), basit en küçük kareler regresyonu (SLR) ve genetik ters en küçük kareler (GILS) yöntemleri, bu renklendirici maddelerin kantitatif tayini için hazırlanan numunelerin görünür spektrumlarına başarıyla uygulanmıştır. Bu boyaların ve tek boya içeren temizlik ürünlerinin ikili ve üçlü karışımları halinde toplam 35 adet numune hazırlandı. Bu 35 örnekten 29 tanesi kalibrasyon seti olarak seçilmiş, geri kalan 6 tanesi ise bağımsız validasyon seti olarak kullanılmıştır. Hazırlanan sete 400-700 nm aralığındaki absorpsiyon spektrumları kaydedilerek elde edilen spektral verilere SLR, PLS ve GILS yöntemleri uygulanarak ortaya çıkan tahminler gerçek değerlerle karşılaştırıldı. Bu yöntemlerin (PLS ve GILS) regresyon katsayıları (R^2) ve standart hata çapraz doğrulama (SECV) değerleri ile tahminin standart hatası (SEP) bulunmuş, sonuçlar değerlendirilmiş, kemometrik modellemenin doğruluğu gözden geçirilmiş, tahminler ve referanslar karşılaştırıldı. Bu analiz yöntemleri ile kemometrik yöntemlerin UV-Görünür Spektroskopi ile birleştirilmesiyle temizlik ürünlerindeki boyaların konsantrasyonu belirlenebilmektedir.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

Figure **Page**

LIST OF TABLES

Table **Page**

CHAPTER 1

INTRODUCTION

Cleaning industry products, colors that can be use in them not only for aesthetic purposes but also contribute to functionally. It helps to improve appearance of the products, where many times attracts consumer's attention generating brand awareness. This also makes it easier to differentiate between the different cleaning products that are used in our day-to-day activities, based on the color-coding system; kitchen cleaners generally have one code while bathroom cleaners may be coded with other colors. Dyes can indicate what a product might contain. For example, household cleaning products with antibacterial ingredients are frequently blue or green in color. Nevertheless, the dyes chemical composition used in cleaning products should be critically examined to assess their impact on safety and environment overall. However synthetic dyes on the other hand will provide bold colors and lasting results however some of them can be dangerous or/and toxic for nature and us. Water based paints are a great environmental option but can lack the same quality of some of our other options in terms of the long-term durability. While dyes act as a significant part of promoting products and steering consumer choice in what seems to be the most competent direction, it should not necessarily take place if their effects on health and environment are properly addressed. As such, the manufacturers of cleaning products must ensure that they choose the right types of paint which will help in maintaining aesthetics as well as avoid potential safety hazards.

The amount of dye used in the formulation of cleaning materials is an important factor to increase the aesthetic appeal of the product, consumer trust and brand recognition. The amount used ensures that the color of the product is distinct and consistent, giving the user a positive impression about the effectiveness and quality of the product, and the dye ratio is generally used at very low rates. This is enough to ensure that the coloring of that product is unique and uniform, making a good impression on the user regarding the efficiency and quality of it. This usually is reported in percentage by weight or volume, and when the formulation process occurs it is kept under meticulous control. Cleaning products are not just dyed to make them look pretty, though. The amount of dye in such a material is heavily regulated due to safety and functional elements. Excessive use causes increased skin irritation. Hence the right balance has to

be achieved in establishing correct quantum of dye. Quality control during laboratory testing ensures that the concentration of dyes in our products do not exceed acceptable levels. In turn, spectrophotometry and chromatography allow us to accurately determine the amount of dye and confirm that it is constant in each batch. In addition, devices that measure color (colorimeters) guarantee the brand consistency and each item will have exactly the same shade. Thus, the proper regulation of amount will ensure the competitive advantage while protecting consumers from health threat and safety concerns that may be associated with some of the products. That's why cleaning materials manufacturers are very careful to formulate the optimum by employing state-of-the-science methods and cutting-edge technology in how they use paint. In conclusion, while the utilization of dyes for presentation and guidance must be a significant sector in producing competition and influencing consumer selection when it comes to cleaning solutions, its use can only be limited due to certain environmental and health impacts. Hence, the cleaning product manufacturers should be careful in choosing suitable types of paint that fit both aesthetically and safety-wise.

If cleaning products containing dye do not have the desired color value and quality, various analytical methods determine what level of dye is present. Spectrophotometry provides sensitive and accurate information due to the measurement of light absorption properties of the dye, and HPLC assures precise quantification since they are separated. This is done by comparing the intensity of the colour to a standard reference. UV-Visible spectroscopy uses the information about how dyes (or colors) in them absorb radiation within the ultraviolet and visible spectrum to give high sensitivity over a wide concentration range. All these methods ensure accurate and homogenous estimates on the utilized dyes in cleaning items production; hence, it will enable optimization or enhance of processes of productions along with keeping product quality at all times.

Chemometric methods are useful in the evaluation and interpretation of data in chemical analysis, and increasingly essential for its verification. These methods help in extracting useful information for big and complex data, as well they also improve the accuracy, sensitivity, and reliability of the analysis. All of these explanations and analyses to be carried out will lead us to chemometric methods as a solution. Chemometrics is defined as the application of mathematical and statistical methods to various types of chemical information in order to obtain useful information on a large volume of processed data. These methods reveal previously unseen patterns and relationships from large,

complex and broad data such as spectra, chromatograms or any continuous measurement-based analytical data. In particular, it is widely used in all steps to build and validate calibration models or for noise reduction, signal separation and data correction. Chemometric techniques are developed to enhance the effectiveness of analytical methods, while it also ensures the accuracy and precision involved in data analysis. It can also be applied during process optimization, allowing to improve production processes and produce better products. It therefore appears that chemometrics has become part of modern chemical analysis.

CHAPTER 2

STATE OF THE ART

2.1. Historical Development of Cleaning Products

As the awareness of human beings becomes stronger and greater in terms of their hygiene and health, there have been a lot of evolutions made on cleaning products from time to time historically. Wherever people were scrubbing away dirt with basic water and sand mixtures, the Egyptians started using soap like substances of oil combined with alkaline salts. Soap production did not become widespread in Europe until the Middle Ages, with the introduction of animal fats and alkali made from ash. With the development of the modern chemical industry in the late 19th century, synthetic detergents appeared. Later, the detergent industry grew significantly using petroleum derivatives and household cleaning products began to proliferate.[1] At the end of the 20th century, environmental concerns along with other health considerations led to improved green cleaning products involving a greater amount of natural and biodegradable active ingredients that contain less dangerous chemicals. Nowadays, thanks to developing technologies, it is offered in a wide variety and increasingly effective with more efficacious cleaning products are safer and sustainable.

Figure 2.1. Cleaning Materials

2.2. Types of Cleaning Products

There are a few different kinds of cleanings and cleaning materials. The products used in this regard vary depending on their function, so they are categorized adequately as well. Detergent is one of the most common cleaning material uses for household chore such as washing clothes and dishwashing. Most detergent includes surfactants which lower the surface tension between water and dirt, allowing dirty to be wash away more easily. A disinfectant is a substance or chemical agent that kills microorganisms, especially those found only on living surfaces. These chemicals are trusted key agents in medical geographies and numerous houses; elsewhere, any cleaning resource that helps kill all harmful microbes can be said to be a powerhouse. What are surface cleaners? Surface cleaners are products used for cleaning floors, walls, glass and other surfaces. They generally contain water, surfactants (detergents), solvents and sometimes antiseptic agents. Toilet cleaners are used to remove dirt, detritus, and microorganisms (such as bacteria and viruses) in bathrooms and lavatories. Some cleaners kill pathogenic bacteria such as Enterococcus fecalis, Salmonella choleraesuis, Staphylococcus aureus, Candida albicans etc.[2] These products usually contain strong acids or alkalis. Multi-purpose cleaners are a type of general cleaning product with high versatility as it can be used on different surfaces such as home or even industrial use. 'Ammonia or alcohol' - These are the basic ingredients of glass cleaners that give your windows and mirrors a spotless clean. There are two primary types of detergent available: powder and liquid. Both powders and liquids specifically target skill levels, fabrics type or even the color of your clothing. Besides, for the tough stains and whitening procedures even the bleaches or stain removers are also utilized. Cleaning materials with eco-friendly alternatives have fewer harsh chemicals and are biodegradable, which makes it a better choice for environmental and health sustainability.

2.3. Cleaning Products Content

Cleaning materials are made up of a variety of raw materials, all of which play a different role in the effectiveness and safety of the final product. One of the main ingredients of cleaning products are surfactants, which are responsible for reducing the surface tension of a liquid thereby allowing it to mix with oils and dirt and greatly

increasing the effectiveness of the clean. Sodium lauryl sulfate belongs to the group of anionic surfactants, which are excellent for cleaning particulate soils and organic substances, and for the removal of grease, non-ionic surfactants are preferred, which have anti-grease properties. Water softening chelating agents, such as the builder sodium tripolyphosphate, can also be added to the wash water to bind metal ions which interfere with surfactants, and which prevent the formulation of soap scum.[3] Laundry detergents and dishwasher have a critical demand for enzymes, such as proteases, amylases, and lipases for cleaving protein-, starch- or fat-based stains.[4] These bleaching agents, such as sodium hypochlorite and hydrogen peroxide, are included for whitening fabrics and bactericidal activity by decomposing-colored organic compounds and microorganisms.[5] These contain other ingredients dissolved in water or appropriate kinds of solvents like ethanol or isopropanol to increase the cleaning power on greasy or sticky materials.[6] To avoid microbial contamination in the products, mainly formaldahyde donors and isothiazolinones are added as protective which guarantee the necessary stability of the products as well as their safety to the consumer during the shelf life. Fragrances and dyes are common in many household items to improve the appearance and odor of the product but can also be controlled because they have been shown to elicit allergenic responses.[7] If a different texture is desired for the detergent, thickening agents such as xanthan gum or cellulose derivatives may also be added. When precisely mixed, these raw materials give formulation to cleaning and cleansing products, which are efficient, safe, and easy to use.

2.4. Surface Cleaners and Its Content

Surface cleaners are general cleaning supplies that can be used to clean and sterilize numerous surfaces in large scale areas such as homes, workplaces or industrial units. They are made with water, surfactants, solvents, fragrances and may optionally contain antiseptic or antibacterial compounds. Surfactants help in detaching dirt and oil from the surface to extract them while cleaning. Hard-to-remove stains and oils are dissolved using solvents, guaranteeing a thorough clean on every surface. Every time an area is properly cleaned, fragrances impart a fresh feeling to the area as they always leave behind a nice and pleasant smell constitute hygiene with the help of antiseptic and antibacterial, components enabling death to microorganisms. Some were uniquely

formulated for certain types of surfaces such as wood, ceramics, glass and metal. Environment friendly surface cleaners are the ones composed of relatively less harmful chemicals and are also biodegradable hence a healthier as well as environmentally safer choice.

Figure 2.2. Surface Cleaner Applications

2.5. The Importance of Dye in Cleaning Products

Dye is essential for cleaning materials not only because they make it beautiful but also bring its functionality back to life. These dyes increase the aesthetic value found in products, allowing them to catch the customers' eye and connecting their visual appeal with brand recognition.[8] Certain colors mean certain things to the products' intended use or content. For example, blue is usually antibacterial, and green means there are green ingredients.[9] It also makes the various types of cleaners easier to identify, allowing consumers to more easily locate the product they are seeking. Dyed chemicals also enable the display and easier visibility of product instructions as well as warnings. On the other hand, there is a potential negative impact on the environment and human health from dyes applied to cleaning agents. Even though synthetic dyes help to achieve bright and long-lasting colors, they often include certain poisonous chemicals, which leads to water

pollution and harmful outcomes.[10] Hence, these days the manufacturers are more inclined towards using biodegradable and environment-friendly paints. These paints still provide the same visual appeal and advantages, but without further harm to our already ill environment. And most importantly, they do not affect consumer health as well. It is strict that the amount of dye involved in cleaning products should be regulated and assayed during formulation because a higher frequency utilization can disrupt the chemical equilibrium nature or result in unwanted reactions. The quantitative analysis of dyes and their optimal level for use is performed through a wide range of chemical methods such as spectrometry, chromatography, or colorimetric test. Such methods are essential for the optimization of production processes and ensuring that the quality of each product is consistently high. Proper and safe use of the paints will lead to the successful entry of the product into the market and also win consumer's trust.

Figure 2.3. Dyes in the Cleaning Industry

2.6. The Types of Dye in Cleaning Products

Dyes in cleaning products are examined under two groups. Firstly, they are synthetic dyes, and secondly, they are natural dyes. Azo dyes are widely used in laundry detergents and surface cleaners because of their bright, long-lasting color variety. Phthalocyanine dyes provide high stability in water-based products and are therefore preferred especially in blue and green colors. Natural dyes are environmentally friendly and biodegradable dyes obtained from biodegradable products. These include vegetable dyes and mineral pigment dyes with chlorophyll and anthocyanin derived from plants, such pigments as iron oxide and titanium dioxide derived from natural minerals and

offering stable coloring without toxicity.[11] Along with the dyes that are legal for use in the food, pharmaceutical, and cosmetic industries, food dyes are used in the cleaning and cosmetics industry. Optical brighteners are used, for example, in laundry detergents to give fabrics a whiter and brighter appearance. The bleaching effect is achieved when fluorescent substances absorb ultraviolet light and emit it back as blue light. Health and environmental safety, as well as aesthetics and functionality, are other equally important needs when choosing paint for cleaning products.[12]

Tartrazin	Food Yellow 4	19140	
Sunset Yellow FCF	Food Yellow 3	15985	
Carmoisine	Food Red 3	14720	
Ponceau 4R	Food Red 7	16255	
Amaranth	Food Red 9	16185	
Erythrosine	Food Red 14	45430	
Allura Red 40	Food Red 17	16035	
Chocolate Brown HT	Food Brown 3	20285	
Indigo Carmine	Food Blue 1	73015	
Brilliant Blue FCF	Food Blue 2	42090	
Quinoline Yellow	Food Yellow 13	47005	
Food Black PN	Food Black 1	28440	
Fast Green S	Food Green 4	44090	
Patent Blue V	Food Blue 5	42051	
Fast Red E	FastRed E	16045	
Food Red 2G	Food Red 2G	18050	
Fast Green FCF	Food Green 3	42053	

Figure 2.4. Food Dyes in the Cleaning Industry

2.7. Types of Dye

2.7.1. Allura Red

A synthetic azo dye, a category of synthetic dyes, Allura Red is especially distinctive by the intensely bright red hue it exhibits and its widespread utilization across several industries. Azo dyes are a major class of dyes known for their bright, highly colored, and long-lasting colors with excellent versatility.[13] Allura Red dyes find a

multitude of applications in the textile industry, where they are employed for dyeing textiles like cotton, wool, or silk to give colours that are fast to washing and light.[14] It is also employed in the leather industry as a dye for leather products, adds an appropriate aesthetic finishing tough and durable. Allura Red is used in the paper industry for manufacture of colored papers and of cardboard intended to contribute to the visual attractiveness of packaging materials and stationery.[12] In addition, the dye is incorporated into the food industry, subject to rules and regulations, and used in the marketing of candy, beverages, and processed foods to make them visually more beautiful for consumers. It is also use in the cosmetics industry (lipsticks, nail polish, hair dyes) to ensure shiny appealing colors. It is also used in pharmaceutical, where drugs and capsules are colored easily with the help of Allura Red and in that process helps to identify and differentiate drugs. For all these applications, synthetic azo dyes are used, such as Allura Red, however, there is also a great concern regarding the environmental and health consequences of the use of these dyes. Some azo dyes have been shown to can decompose to aromatic amines, some of which are known or suspected to be carcinogenic and regulated.[14] Consequently, research is in process to find safer and sustainable alternatives in the form of plant-based dyes and ecofriendly synthetic dyes.

Figure 2.5. Powder Allura Red

2.7.2. Brilliant Blue

Brilliant Blue or Brilliant Blue FCF or FD&C Blue are synthetic dyes of blue, very used throughout the world in various sectors of the industry, being responsible for intense blue color and durability. The use of Brilliant Blue is in the food and beverage areas o is demonstrates loud and clear one form with the color to bring extensive food coloring such as candies, beverages, baked goods, and dairy products, attracting people

to see. The image is non-food dye approved for use in many countries including the United States and Europe with usage levels regulated to ensure consumer safety.[15] Pharmaceutical use Brilliant Blue is used for coloration in medicines as tablets and syrups helped in identification and differentiation since it fits in the pharmaceutical standards.[16] Brilliant Blue is also widely used in the cosmetic industry, in such products as shampoos, soaps, lotions, and makeup, to make the products more marketable and attractive.[17]

Along with being used in consumer goods, Brilliant Blue is also used in scientific research and medical diagnostics. Along with being used in consumer goods, Brilliant Blue is also used in scientific research and medical diagnostics. In addition, Brilliant Blue has been investigated for its neuroprotection properties, particularly after spinal cord diseases and nervous system-related disease states and has been found to have benefits in reducing inflammation and promoting healing synthetic colors such as Brilliant Blue are used worldwide but are suspected of having harmful effects on the environment and human health. Although generally considered safe, high doses of Brilliant Blue have been found to cause allergic reactions in sensitive individuals, leading regulatory bodies to strictly monitor its use.[18] Brilliant Blue has applications in addition to its use in consumer goods. in scientific research and medical diagnosis. It is used as a biological stain in microscopy and cell biology to visualize and distinguish cellular components and structures.

Figure 2.6. Powder Brilliant Blue

2.7.3. Tartrazine

Tartrazine known as FD&C Yellow No. 5 is a synthetic lemon-yellow azo dye

utilized in a variety of industries because of its properties as a dye. It is most used in the food and beverage sector to give candies, soft drinks, processed snacks, and baked goods a vibrant appearance. The dye is controlled by the U.S. Food and Drug Administration (FDA), as well as the European Food Safety Association (EFSA), which establish the maximum permissible concentration to avoid damage to consumers.[18] Its uses in the pharmaceutical field include that colouration of pills, tablets, and liquid medicine, to help people in the identification of the product and the differentiation of dosages in case of liquid medication.[16] Tartrazine is also used as a dye in the cosmetics industry for hair dyes, shampoos and lotions, where it gives a bright, inviting color.[17] A controversy and research regarding health risks of tartrazine has long been existing despite of enormous consumption of tartrazine. There have been occasional studies linking this compound to hyperactivity in children, bringing about some controversy about its safety. Labelling of foods as containing the dye has been required in some countries by a regulatory body. The widespread application of tartrazine in industrial processes is also associated with its wastewater pollution, because synthetic dyes are generally resistant to biodegradation. Therefore, efforts are always to explore other options of dyes removal from wastewater and advanced treatment measures. These efforts are focused on lowering the environmental footprint of synthetic dyes, without sacrificing their functional advantages.

Figure 2.7. Powder Tartrazine

2.8. Advantages and Disadvantages of Dyes Use in Cleaning Products

Table 2.1. Advantages and Disadvantages of Dyes

Types of Dye	Advantages	Disadvantages
Synthetic Dyes		
Azo Dyes	- Bright, vibrant colors - Long-lasting - Wide color variety	- Potential health risks - Environmental pollution due to non-biodegradability
Phthalocyanine Dyes	- High stability in water-based products - Strong colorfastness	- Environmental impact due to complex chemical structure
Optical Brighteners	- Enhance visual whiteness and brightness - Effective at low concentrations	- Can cause allergic reactions in sensitive individuals - Potential for environmental persistence
Natural Dyes		
Vegetable Dyes	- Environmentally friendly. - Biodegradable, perceived as safer by consumers	- Less color variety - Generally less stable and less vibrant
Mineral Pigment Dyes	- Stable coloring without toxicity - Derived from natural minerals	- Limited color range - Potential for heavy metal contamination

Each type of paint may differ depending on its intended use and environmental factors. Therefore, it is important to consider the location and conditions of use when choosing the most suitable type of color.

2.9. UV-Vis Spectroscopy

The most important and most widely used form of spectroscopy in analysis of the absorption of ultraviolet and visible light of chemical substances, UV-Visible spectroscopy (ultraviolet-visible spectroscopy). This method relies on the fact that

molecules absorb light at certain wavelengths, causing electronic transitions between orbitals.[19] Absorption spectra are a tool that measures the molecular structure of the analyte while providing information about its concentration and environment. For example, by applying the Beer-Lambert law (which relates absorption to concentration), concentrations of substances in solution can be measured by UV-Visible spectroscopy.[20] It is also used in investigation of reaction kinetics, for the change of absorbance as a function of time can tell the reaction rate and mechanism.[21] It is used in sectors such as pharmaceuticals for drug quality control, environmental science for monitoring pollutants, and biochemistry for the characterization of proteins, and nucleic acids etc.[22] It can say that UV-Visible spectrophotometers have diode array detectors and user-friendly software. Now available worldwide, UV-Visible spectroscopy is not without its disadvantages, such as interference from other absorbing species in the sample matrix, or the need for the sample to be in clear, colourless solution free from scattering effects.[23] However, its ability to perform fast, non-invasive quantitative analysis has made UV-Visible spectroscopy a ubiquitous analytical tool in virtually every scientific field.

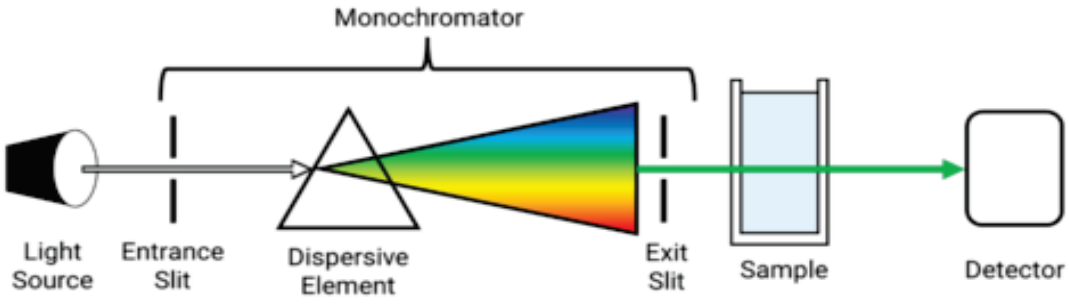


Figure 2.8. Schematic Summary of UV-Vis

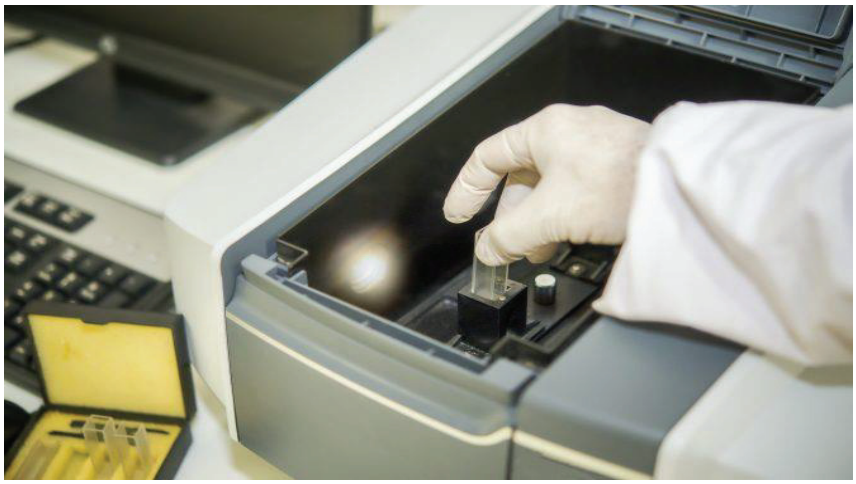


Figure 2.9. UV-Vis Usage and Interior View

2.9.1. UV-Vis Advantages

UV-Visible spectroscopy is one the most applied and versatile tool in the field of analytical chemistry to both qualitative and quantitative analysis. An important feature is that it is swift and precise way to analyse and to know the quantity of analytes in a solution knowing what the absorbance of ultraviolet and visible light is.[20] This technique, although being very pursed, allows for nanomolar concentrations to be detected, and therefore, it is invaluable for applications in pharmaceuticals, environmental monitoring, and biochemical research.[19] Furthermore, UV-Visible spectroscopy can be a non-destructive technique, so the sample can sometimes be tested without change or consumption of the original sample, which allows for additional tests on the sample, or further use of the sample.[24] Additionally, the technique benefits from relatively straightforward sample preparation when compared to other analytical methods, as it generally only requires that the sample be diluted in an appropriate solvent, making it both time and cost effective[24]. In addition, it is not widely used in the laboratories around the world at a relatively cheap price, since UV-Visible spectrometers are the most common and least expensive.[20] Data obtained using the UV-Vis spectroscopy are simple to analyze and interpret, absorbance spectra show well defined the specific peaks according to type of analyte, which allow quick identification and quantification.[19] Additionally, progress of instrumentation has allowed the production of small/portable UV-Visible spectrometers useful for field analysis and in situ testing, especially for environmental analysis and quality control.[24] Another character of UV-Visible absorption specifically that the UV-possibility expand

due to each foundation and band to country levels, similarly, to utilizing to UV-Visible spectroscopy in reading response kinetics and monitoring the course an artificial reaction in actual time, sighted at response mechanisms and quotes. UV-Visible spectroscopy is therefore a fundamental technique in modern analytical science, due to its simplicity, wide applicability and the fact that it is an extremely robust technique.

Table 2.2. Advantages and Disadvantages of UV-Vis

Advantages	**Disadvantages**
High Sensitivity	Limited to Transparent or Translucent Samples
Can detect very low concentrations of analytes, providing high precision in quantitative analysis.	Samples that absorb strongly in the UV-Vis region can lead to non-linear responses or complete absorption, complicating analysis.
Non-Destructive Analysis	Requires Calibration and Standards
Allows samples to be analyzed without being consumed, enabling further testing or use.	Accurate results depend on proper calibration with known standards, which can be time-consuming and require meticulous preparation.
Quick and Efficient	Susceptible to Matrix Effects
Provides rapid results with minimal sample preparation, making it time-efficient and cost-effective.	Sample matrix can affect the absorbance, leading to inaccuracies if not properly accounted for.
Versatile Applications	Limited Structural Information
Applicable to a wide range of substances, including organic and inorganic compounds, in various fields such as pharmaceuticals, environmental monitoring, and biochemical research.	UV-Vis spectra provide limited information about molecular structure compared to other techniques like NMR or mass spectrometry.

2.9.2. Beer-Lambert Law

The Beer-Lambert law is a fundamental principle in spectrophotometry that describes the linear relationship between the absorption of light by a substance and its concentration in solution. Lambert-Beer states that the absorbance of a sample (A) is equal to the path length of light in the sample (b) multiplied by the concentration of absorbent-containing substances (c) and the molar absorbance constant of proportionality (ε) as given in equation 2.1.

$$\mathbf{A = \varepsilon bc} \qquad \mathbf{(2.1)}$$

This equation is very important in quantitative chemical analysis; it is used to find out the concentration of a substance if the absorbance and molar absorptivity and path length is known.[20] The Beer-Lambert equation is incredibly important in molecular spectroscopy, including in analytical chemistry, biochemistry, and environmental science to determine the concentration of a solute's solution. For example, it is employed in clinical laboratories to measure the amount of glucose in blood, or the amount of hemoglobin. The law goes on to provide that the system is perfect, and hence absorbance is a consequence of only the absorbing species and that the light path is definite. Nonetheless, these deviations can happen as a result of the high concentration of solute giving rise to molecular interactions, the limitations of the instrument, or the stray light.[23] Nonetheless, even with these possibilities for inaccuracy, the principle of Beer-Lambert law continues to underpin our analytical techniques, providing a simple and self-contained method of concentration determination when correctly applied.[21] Instrumentation and data analysis has continued to increase the precision and utility of this cardinal law in most modern analytic methods.[22] Graphical illustration of Beer's law is given in Figure 2.10.

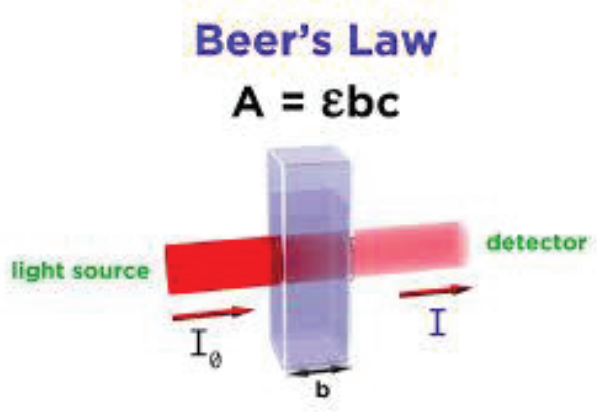


Figure 2.10. Beer's Law

2.10. Chemometrics Methods

The chemometric approach, by the application of mathematical and statistical methods on measurements, performs data processing to gain information from complex chemical problem solving, this approach will focus to increase performance and reliability of analytical results. They are necessary to work and interpret large data sets obtained by modern analytical techniques like spectroscopy, chromatography, or Mass spectrometry.[25] Principal component analysis (PCA) is a powerful chemometric method suitable for reducing the dimensionality of data sets, while preserving as many of the variances between samples as possible which simplifies easier analysis of data and discovery of patterns.[26] A second important approach is partial least squares (PLS) regression, which is commonly used to construct predictive models, often in cases in which the predictors are highly collinear or when the number of predictors outstrips the number of observations.[27] It also used chemometrics in pharmaceutical research to guarantee the quality of the product through the analysis of chemical composition and quality of the raw materials and final products.[28] By being combined with chemometric methods, these tools not only allow more accurate and efficient data interpretation but also greatly contribute to the expansion of analytical methodologies.[25] These tools allow the analyst to make decisions about the interpretation of data and system variations, therefore supporting several scientific and industrial fields with more robust data analysis available.

2.10.1. Comparison of Chemometrics Methods

Table 2.3. Comparison of Chemometrics Methods

Method	Advantages	Disadvantages
Principal Component Analysis (PCA)	- Reduces dimensionality of data, simplifying analysis - Identifies patterns and correlations among variables.	- Can be sensitive to scaling of variables. - Interpretation of principal components can be challenging.
Partial Least Squares (PLS)	- Handles collinear and noisy data well. - Provides both predictive models and variable selection.	- Requires careful validation to avoid overfitting. - Interpretation of results can be complex.
Genetic Inverse Least Squares (GILS)	- Efficiently selects relevant variables. - Robust against noisy data and overfitting.	- Computationally intensive due to genetic algorithm - Requires expertise to implement effectively.
Multiple Linear Regression (MLR)	- Simple to implement and interpret. - Provides direct relationships between variables.	- Assumes linear relationships which may not always be accurate. - Sensitive to multicollinearity among predictors.
Artificial Neural Networks (ANN)	- Can model complex non-linear relationships. - Adaptive learning capabilities.	- Requires large datasets for training. - Prone to overfitting if not properly regularized.

2.10.2. Principle Component Analysis (PCA)

The great advantage of Principal Component Analysis (PCA) is that each index calculated while studying our data represents the principal characteristics that distinguish our data from high dimensional data. PCA does this by converting your original variables into a new set of uncorrelated variables called principal components, which are ordered in such a way that the first few captures most of the variation present in the original dataset.[29] The main objective when using PCA is to project the data in question to a smaller set of axes defined in principal component space in order to allow us to visualize data, discover patterns, and to reduce noise for beneficial data interpretation and analysis. Especially in domains like spectroscopy and chromatography, this approach is useful as they have datasets with many correlations in between variables. In spectroscopy, for example, PCA can be used to recognize important spectral columns, in terms of which are the dominant features responsible for the largest fraction of the variance of the data, which makes easier the classification of samples by similarities or the detection of anomalous samples.[30] PCA is also widely used in exploratory data analysis to gain insight into the hidden patterns or underlying structure that exist in cases where the data is not directly accessible.[31] PCA has one of the advantages that it is non-parametric which does not require imposing any distribution for the data itself, making it very flexible.[29] But the interpretation of the principal components is difficult at times particularly with complex or non-linear data structures. Nonetheless, PCA is still an important tool in chemometrics and other data-rich disciplines, helping with finding valuable information within large data sets.

The PLS model equation described as:

$$\mathbf{A}_{mxn} = \mathbf{T}_{mxh} \,.\, \mathbf{B}_{hxn} + \mathbf{E}_{\mathrm{A}mxn} \qquad \textbf{(2.2)}$$

A is an *mxn* matrix of spectral absorbance, B is a *hxn* matrix of loading vectors or loading spectra, and T is an *mxh* matrix of intensities or scores in the new coordinate system of the *h* loading vectors for the *m* sample' spectra. E_A is the *mxn* matrix residuals not fit by the model.

2.10.3. Simple Least Squares Regression (SLR)

Simple Least Squares Regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables by minimizing the sum of the squared differences between the observed and predicted

values.

This technique is represented by the equation:

$$y=\beta0+\beta1x+\epsilon y \quad (2.3)$$

Where y is the dependent variable, x is the independent variable, β0 is the y-intercept, β1 is the slope, and ϵ is the error term.[32] The primary goal of Simple Least Squares is to find the line that best fits the data points, providing a clear interpretation of how changes in the independent variable affect the dependent variable.[33] This method assumes a linear relationship between the variables and that the residuals errors are normally distributed with a mean of zero and constant variance, known as homoscedasticity.[34] Simple Least Squares is widely used due to its simplicity and effectiveness in providing valuable insights into data relationships, despite being sensitive to outliers and assuming linearity, which may not always be the case in real-world data.[35]

2.10.4. Partial Least Squares (PLS)

Partial Least Squares (PLS) regression is a chemometric method that is used to model complex relationships of large number of observations with high multicollinearity and less objects between dependent and independent variables. PLS consists in projecting the predictor variables as well as the response variables in a new space defined by latent variables, the latter are linear combinations of the original variables.[36] This approach is specifically significant in cases where traditional regression techniques taking into account linear regressions experience the anxiety over Multicollinearity or high-dimensionality sort of issues in its work.[27] PLS has been widely used in spectroscopy where it is used to correlate the spectral data to the chemical concentrations of a sample, allowing for the development of predictive models able to deal with highly complex and noisy data .The PLS method further decomposes both the predictors and the responses such that the covariance between these matrices are maximized so that the latent variables have the most information relevant for prediction.[36] PLS is also widely used in calibration and validation of drug formulations, in pharmaceutical research, and in process control that is used to monitor and optimize industrial processes.[37] Although PLS offers significant strengths (such as those related to fitting nonlinear multivariate models with many correlated x-variables and strong retention of variance between variables), the interpretability of PLS models is inherently difficult as the latent variables are abstract

quantities as such, sample-based estimates) with no direct physical counterpart or connection that must be carefully validated and cross-validated to assure model validity.[38] However, PLS is still one of the major methods in chemometrics, and it leads to essential findings and predictions in many scientific and industrial contexts.[36]

The PLS model equation described as:

$$\mathbf{C}_{mxn} = \mathbf{T}_{mxh} \,.\, \mathbf{R}_{hxn} + \mathbf{E}_{Cmxn} \qquad \textbf{(2.4)}$$

C is an *mxn* matrix of dependent variables, T is a *hxn* matrix of scores matrix, and R is a regression coefficients matrix. E_C is error matrix. This matrix represents the remaining variance or prediction errors that the model cannot explain.

2.10.5. Genetic Inverse Least Squares (GILS)

This method demonstrates that GILS, an advanced chemometric approach that combines the robust nature of genetic algorithms with the accuracy of inverse least squares (ILS) regression, is a powerful method for optimizing predictive models for complex chemical analysis data. GILS is particularly good at model identification given a large number of variables in the data because it selects the most capable features and avoids using too many features. It begins with a genetic algorithm that simulates the process of natural evolution to find the best solution under the premise of individually selecting, homogenizing, and modifying the population of potential solutions. This method allows GILS to efficiently explore large search spaces and avoid local minima that often trap traditional optimizers. After the most appropriate subset of variables is created, the prediction model is created by adding ILS regression to the area where at least one variable contributes significantly to reducing the prediction error. It is particularly useful in the case of spectroscopy and chromatography, where the identification of important spectral bands or chromatographic peaks responsible for changes in chemical properties of interest is of great importance. GILS is highly advantageous in improving the interpretability and generalizability of models by reducing noise floors and hence can be used in environmental monitoring, pharmaceutical, food quality control, etc. It plays an important role in these fields in this case, the versatility and reliability of the method is strengthened by its ability to handle noisy data and, in particular, collinear data. Although hardly comprehensive, benefits from improvements in model accuracy and reduction of overfitting make GILS a preferred option for complex

analytical challenges. The nature of modern chemometrics suggests that GILS can support significant advances as computing power and algorithm design improve.

Inverse Least Square (ILS) method, which is based on the multivariate extension of inverse Beer-Lambert's law and assumed that concentration is a function of absorbance.

$$\mathbf{C}_{mxl} = \mathbf{A}_{mxn} \,.\, \mathbf{P}_{nxl} + \mathbf{E}c_{mxl} \qquad \textbf{(2.5)}$$

In this Equation, C is the concentration matrix, and A is the absorbance matrix. E is the matrix of errors in concentrations not fitted by the model. P is the nxl matrix of regression coefficients associated with the absorbance values to the concentrations of the components in the calibration set.

2.11. Literature Examples

In a study by Ping Qi, eleven synthetic color additives (Allura red, amaranth, azo rubin, brilliant blue, erythrosin, indigotin, ponceau 4R, new red, sunset yellow, quinoline yellow and tartrazine) in flour and meat foods were developed and validated using HPLC combined with DAD and MS/MS. The color additives were extracted with ammonia-methanol and further purified by the SPE procedure. These dyes were then analyzed and confirmed.[39]

Pekcan Ertokuş applied partial least squares and principal components regression methods. She examined various mixtures of Allura Red and Brilliant Blue to determine the concentrations of colorants, it was also analyzed by UV-spectrophotometry in chemical separation. The obtained experimental data were evaluated with chemometric methods as Partial Least Squares (PLS). As a result, the predictions were compared with the reference values.[40]

Food coloring mixtures containing Yongnian Ni, tartrazine, sunset yellow, ponceau 4R, amaranth and brilliant blue provided analysis by spectrophotometry. Data obtained from experiments. It was processed with chemometric approaches such as classical least squares (CLS) and principal component regression (PCR), and the analysis was made as a result of these methods.[41]

CHAPTER 3

MATERIALS AND METHODS

3.1. Materials

Surface cleaning product was prepared by adding certain percentages of soft water, protective raw materials, chelators, perfumes with nonionic active ingredients and finally 3 different color compounds to be determined in different percentages. All these raw materials were supplied from Viking Cleaning and Cosmetics company. The trade names of the three color compounds to be selected for this product are as follows: R1: Tartrazine (Yellow), R2: Brilliant Blue (Blue) and R3: Allura Red (Red). Figure 3.1 shows colorless surface cleaner product which is used as backgrand measuments in the UV-Visible spectrophotometric data collection part.

Figure 3.1. Colorless Surface Cleaner

As seen from Figure 3.1, it has a slightly yellow appearance due to its colorless chelating and protective raw material content which produces slightly yellow.

3.2. Experimental Setup

The surface cleaner choosen for this study contains only 0.1% (w/w %) dye compound in the final product. In a standard product, R1 is added about 0.055%, R2 is 0.017% and R3 is about 0.028% by mass. The colored surface cleaners that are studied in this thesis were cmposed of either binary or ternary mixtures of dyes alon with single component products by varying the dye concentrations as plus and minus 25% of the standard formula as shown in Table 3.1

Table 3.1. Percentages (w/w %) of Dyes in the Formulation

Color	Formulation (A)	A x 0.25= (B)	Low(A-B)	High(A+B)	Range
R1	0.055	0.01375	0.04125	0.06875	0.02750
R2	0.017	0.00425	0.01275	0.02125	0.00850
R3	0.028	0.00700	0.02100	0.03500	0.01400

While preparing the surface cleaner product, it was prepared with the help of Thermo brand mechanical mixer as shown in Figure 3.2. During mixing, all raw materials and color compounds at specified concentrations were added respectively.

Figure 3.2. Mechanical Stirrer (Thermomac)

Table 3.2 shows the dye composition (w/w %) of the 35 surface cleaning samples used in this study.

Table 3.2. Dye Composition (w/w %) of the 35 Surface Cleaning Samples

No	%R1	%R2	%R3
1	0.06875	0	0
2	0.05500	0	0.00700
3	0.05500	0.00425	0
4	0.04125	0	0.01400
5	0.04125	0.00425	0.00700
6	0.04125	0.00850	0
7	0.02750	0	0.02100
8	0.02750	0.00425	0.01400
9	0.02750	0.00850	0.00700
10	0.02750	0.01275	0
11	0.01375	0	0.02800
12	0.01375	0.00425	0.02100
13	0.01375	0.00850	0.01400
14	0.01375	0.01275	0.00700
15	0.01375	0.01700	0
16	0	0	0.03500
17	0	0.00425	0.02800
18	0	0.00850	0.02100
19	0	0.01275	0.01400
20	0	0.01700	0.00700
21	0	0.02125	0
22	0.05500	0.01700	0.02800
23	0.05500	0.01700	0.02800
24	0.05500	0.01700	0
25	0.05500	0.01700	0
26	0.05500	0	0.02800
27	0.05500	0	0.02800
28	0	0.01700	0.02800
29	0	0.01700	0.02800
30	0.05500	0	0
31	0.05500	0	0
32	0	0.01700	0
33	0	0.01700	0
34	0	0	0.02800
35	0	0	0.02800

Concentrations were enriched by the mixture design and different dye types were added to the products in different proportions. The appearance of these products after

preparation shows that although similar colors may appear, each color is different due to the addition of different concentrations of each dye. The final form of the 35 prepared products is given in Figure 3.3.

Figure 3.3. The Final Form of the 35 Prepared Products

3.3. Determining the Factors and Parameters of the Experiment

Three different dye types were determined as parameters and their concentrations were determined in different percentages. Additionally, our data was enriched with the help of mixture design. All of these 35 different products have different formulations and the percentages of the 3 dyes change each time. The prepared products are subjected to UV-Vis spectroscopy. This is how absorbance values are achieved by scanning the colorless version of the surface cleaner as a blank and then scanning the colored final product as sample. The raw materials in the formulations are generally absorbing in the UV part whereas the spectra of the dyes are seen in the visible part of the spectrum.

3.4. Analysis Method of Experimental Results

The final data was created by scanning the products on the UV-Vis region, and since the product to be analyzed would be in the visible region, the spectral data were recorded between 400 to 700 nm wavelength range. As a result of the spectra obtained from the UV-vis spectrophotometer, it is closely determined which absorbance values of

the dyes represent which dye in this study. Figure 3.4 shows the Thermo brand UV-Visible spectrophotometer used in this study.

Figure 3.4. UV-Vis Spectrophotometer (Thermo)

Mixture design method is a statistical design and optimization method for combining different components into a mixture containing certain ratios to obtain a mixture with the desired properties. This method is used widely for product formulation and for product improvement, particularly in the field of chemistry, food and agriculture, pharmaceuticals, cosmetics and materials science. In other words, the primary objective of mixture design is to determine how the performance, product stability, and other key properties of the product depend on the mixing ratios of the components. That way, you can get low-cost high-quality products by finding most suitable component combinations. The approach is based on the assumption that each component can vary within a certain range and that the total quantity of components is constant. Mixture models provide a more systematic approach than the traditional trial-and-error method in product design stage. and then it is an effective way to implement. This method helps in concluding the most performant formulations in fewer trials, thus less time and resources.

Chemometrics methods are a way to model relationships between dependent variables (responses) and independent variables (predictors) with the goal of obtaining the best-fitting regression model for both sets of variables. This is very useful when having large numbers of predictors, high-dimensional datasets, and multicollinearity

issues. The versatility and highly adaptable structure of chemometric methods are useful and are actively used in different fields, especially chemistry and economics, where they are used as an important modelling tool. More focus was placed on the analysis of chemical components. It is aimed to do this especially for the analysis of chemical components.

After completing the spectral measurements, data were transferred to MS Excel for the pre-processing of raw spectra. The moving average method is a technique used to analyze and interpret time series data. Its purpose is to reveal underlying trends and patterns more clearly by reducing random fluctuations or noise in data series. This technique is widely applied in finance, economics, meteorology, engineering and many other applications. Smoothing the data makes it more interpretable and allows analysts to better predict their expectations of future trends and behaviour. The goal of migration is to reduce the noise in the graph and reveal the main data signal by getting rid of random noise and data. On high-frequency datasets this can lead to significant advantages.

MS Excel was also used to generate simple least squares regression (SLR) models after applying five point moving average smoothing of the raw spectral data. Following this, Minitab statistical software (Minitab 16 Statistical Software. Minitab, Inc., State College, PA) were used in order to develop multivariate partial least squares (PLS) regression models. Finally, genetic inverse least squares (GILS) and PLS models were also generated by using algorithms developed in Matlab (MATLAB R2022b, Natick, Massachusetts: The MathWorks Inc.) programming environment.

CHAPTER 4

RESULTS AND DISCUSSION

The dye components which are coded as R1, R2 and R3 were analyzed by adding 3 different paint types to surface cleaner products at various concentrations. A total of 35 samples were prepared with various different concentrations of these compounds using the mixture design method. UV-Visible spectroscopy was used in the analysis and the range in the visible region (400 - 700 nm) was taken. A wide variety of calibration methods (SLR, PLS and GILS) were used to prepare univariate and multivariate calibration models for the three dye components in order to compare performance of univariate and multivariate calibration methods.

In the first scenario, the spectra in the 400-700 nm range of 35 different products were converted into separate data set and the wavelength of maximum absorbance values corresponding to each dye component were determined. Figure 4.1 shows the spectra of the three samples (s1, s16 and s21). These samples are single dye containing samples in the data set of 35 samples in order to determine maximum absorbing wavelengths of the dyes R1 (443 nm), R2 (635 nm) and R3 (569 nm).

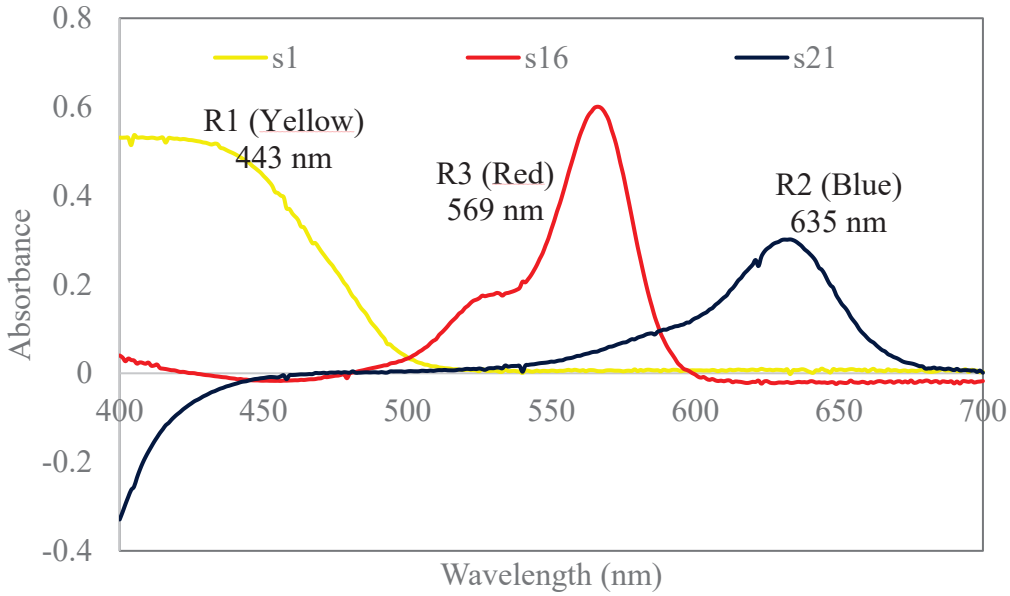


Figure 4.1. UV-Vis. Spectra of the Samples s1, s16 and s21

As can be seen from figure 4.1, it is clear that each dye has a distinct maximum

absorbing wavelengths although R2 and R3 has some overlap and R1 shows a flat trend down to 400 nm. These spectral features of the three single dye component containing samples suggest that even simple least squares may be applied to develop calibration models based on Beer's Law. On the other hand, it would be useful to compare the performance of the multivariate calibration methods such as PLS and GILS with SLR due to the partial overlap in the spectral data. The following section illustrates the results of the univariate and multivariate calibration modelling of the data after moving average smoothing application in order to remove partial noise fluctuations in the spectra.

4.1. Analysis Results of R1, R2, R3 Dyes in Surface Cleaner

4.1.1. Moving Average Smoothing of Raw Spectral Data

A total of 35 different samples were prepared as cleaning products with various concentrations of colouring dyes. Figure 4.2 shows UV-Vis. Spectra of the 35 samples from 400 to 700 nm wavelength interval by taking the colourless cleaning product as blank.

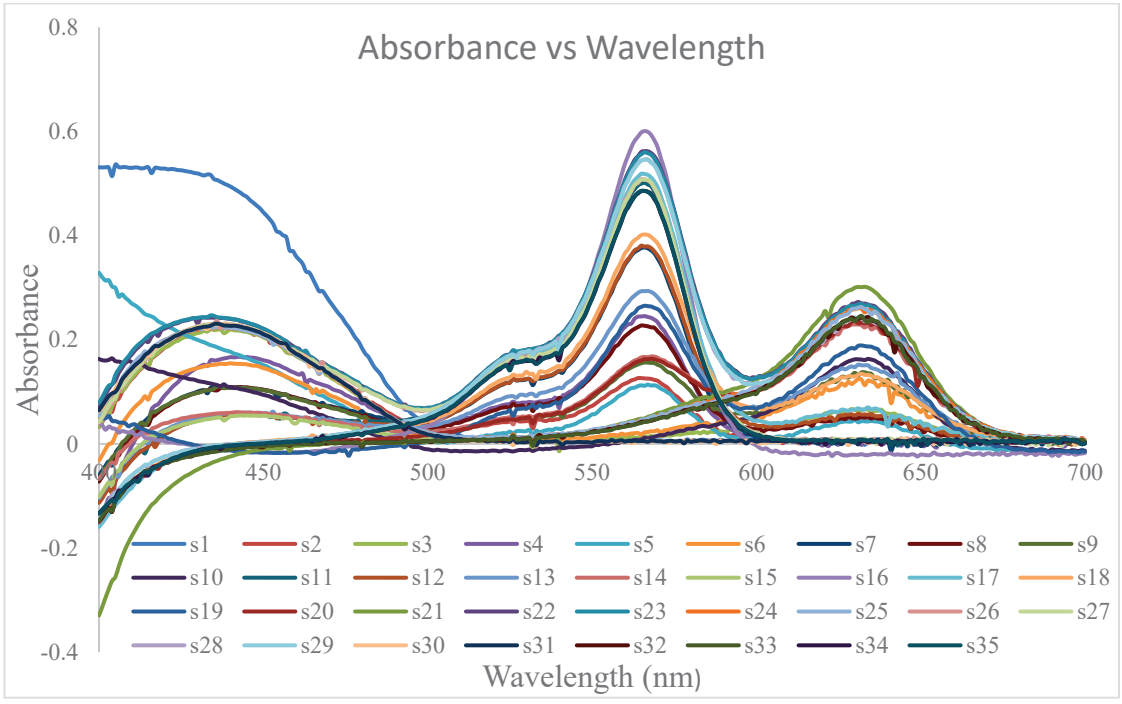


Figure 4.2. Graph of Absorbance Versus Wavelength (nm)

As can be seen in Figure 4.2, three peaks corresponding dyes R1, R2 and R3 show changes in absorbance values depending on the concentration changes of these dyes in the samples. On the other hand, there are some noise fluctuations in various parts of the spectra. These fluctuations negatively affect the data, and therefore the spectral data must be smoothed before further analysis. Here, the moving average method was applied to with a five consecutive data point as window size. Figure 4.3 shows the smoothed spectra of the 35 samples by the moving average smoothing method.

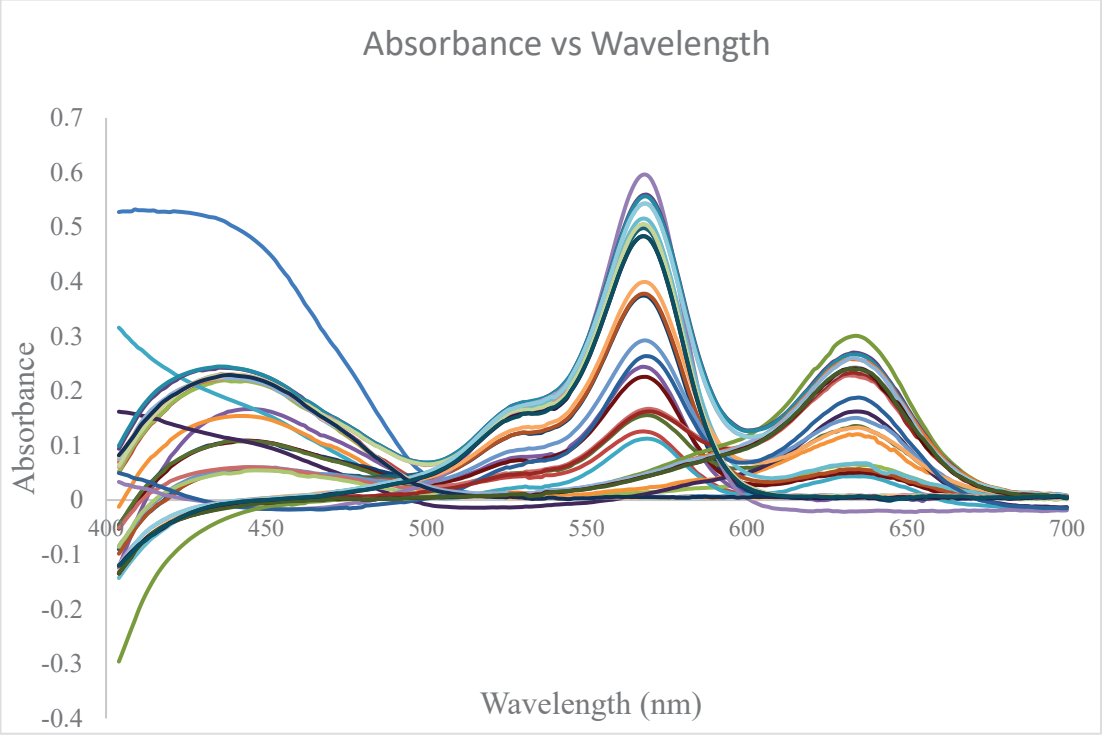


Figure 4.3. Smoothed Spectra of the 35 Samples by the Moving Average Smoothing

As seen from the smoothed spectra, significant portion of the fluctuations on the raw spectral data is corrected. This smoothed spectral data was used for further univariate and multivariate calibration modelling by using SLR, PLS and GILS methods. The following section shows the results of these methods.

4.2. The SLR Results of R1 (Tartrazine, Yellow)

First of all, the data set containing visible spectra (from 400 to 700 nm) of the 35 samples were divided into two subsets as calibration and independent validation sets. Among the 35 samples, 6 of them were chosen as independent validation where those

samples that are mostly the duplicate given in Table 3.2 were selected. The remaining 29 samples were used as the calibration set. Comparison was then made between the applied methods namely SLR, PLS and GILS to see how close the predictions with actual values.

Figure 4.4 shows the plot of concentration vs. absorbance at 443 nm for R1 by using simple least squares.

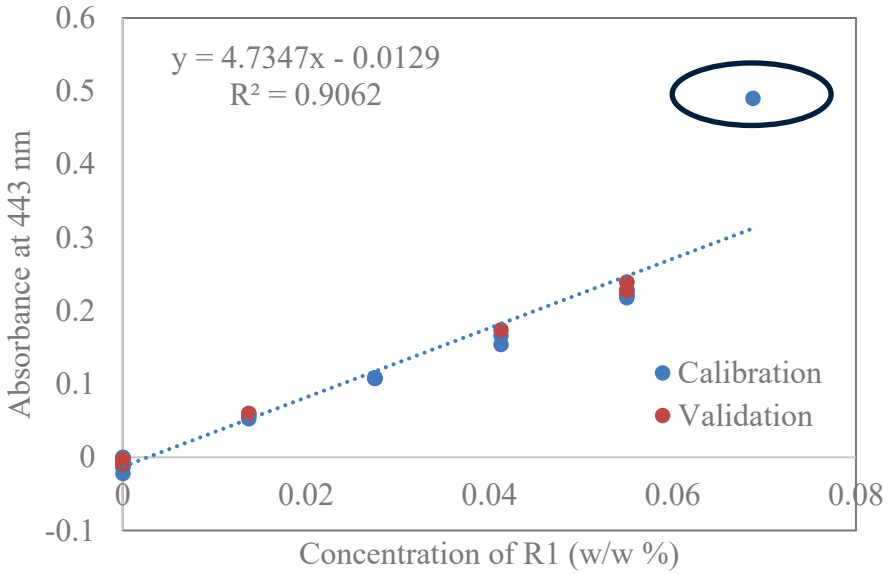


Figure 4.4. Plot of Concentration vs. Absorbance for R1 by Using Simple Least Squares

As seen from Figure 4.4, the sample with the highest R1 content causes a significant deviation from linearity and distorts the success of the calibration plot. Therefore, this sample is taken out from the calibration set only for R1 and the result of new SLR calibration model for R1 is redrawn (Figure 4.5).

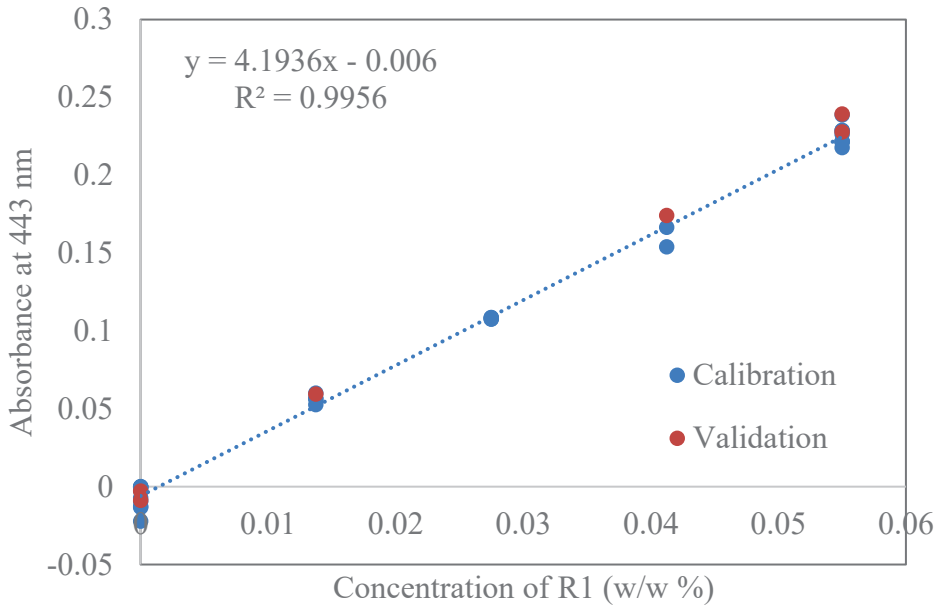


Figure 4.5. Plot of Concentrations of R1 vs. Absorbances at 443 nm by SLR

After removing the sample causing nonlinearity in the calibration line, it is clear that be method was successfully applied, and as a result of this new calibration plot, the regression coefficients of the model were improved from 0.9062 to 0.9956. The standard error of calibration (SEC) and the standard error of prediction (SEP) values were calculated as 0.00966 and 0.00183 (w/w %), for calibration and independent validation sets, respectively. These results demonstrated that even with simple least squares method based on Beer's law could be sufficient for the prediction of the Tartrazine yellow (R1) in the cleaning products for not only single component mixtures but also the mixtures contain binary and ternary mixtures of these three dyes (Tartrazine yellow, Brilliant blue and Allura red).

4.3. The SLR Results of R2 (Brilliant Blue, Blue)

The results of the SLR calibration model for R2 (Brilliant Blue) was plotted and these results were examined. Absorbance (635 nm) vs. concentration plot of R2 obtained from SLR is given in Figure 4.6.

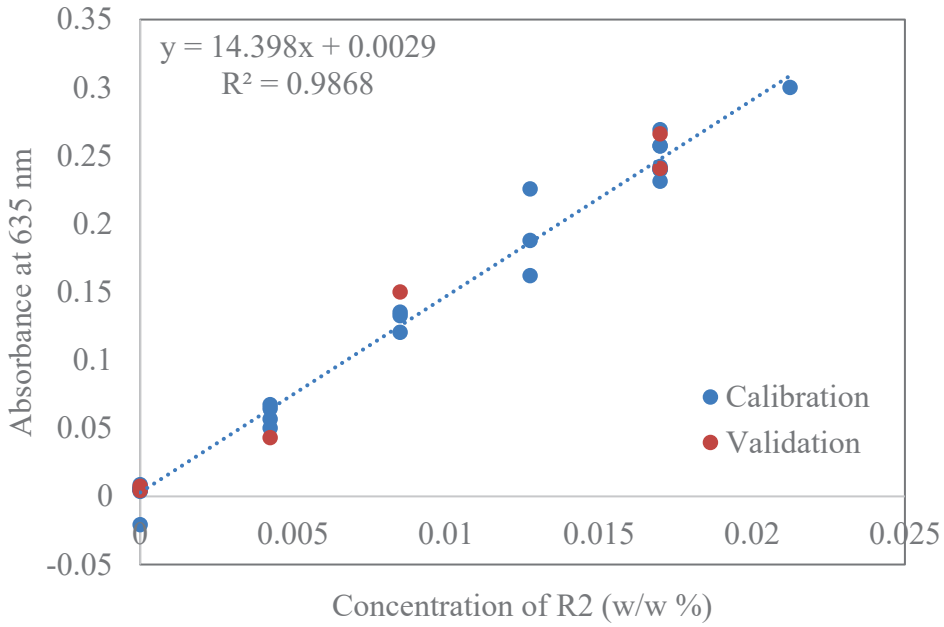


Figure 4.6. Plot of Concentrations of R2 vs. Absorbances at 635 nm by SLR

The regression coefficient of the model was found to be 0.9868 for the calibration set. The standard error of calibration (SEC) and the standard error of prediction (SEP) values were calculated as 0.00089% and 0.00109% (w/w %), for calibration and independent validation sets, respectively. The SLR method was successfully applied, and the error ranges were determined and estimated.

4.4. The SLR Results of R3 (Allura Red, Red)

The result of the SLR calibration model for R3 (Allura Red) was plotted (Figure 4.7) and these results were examined.

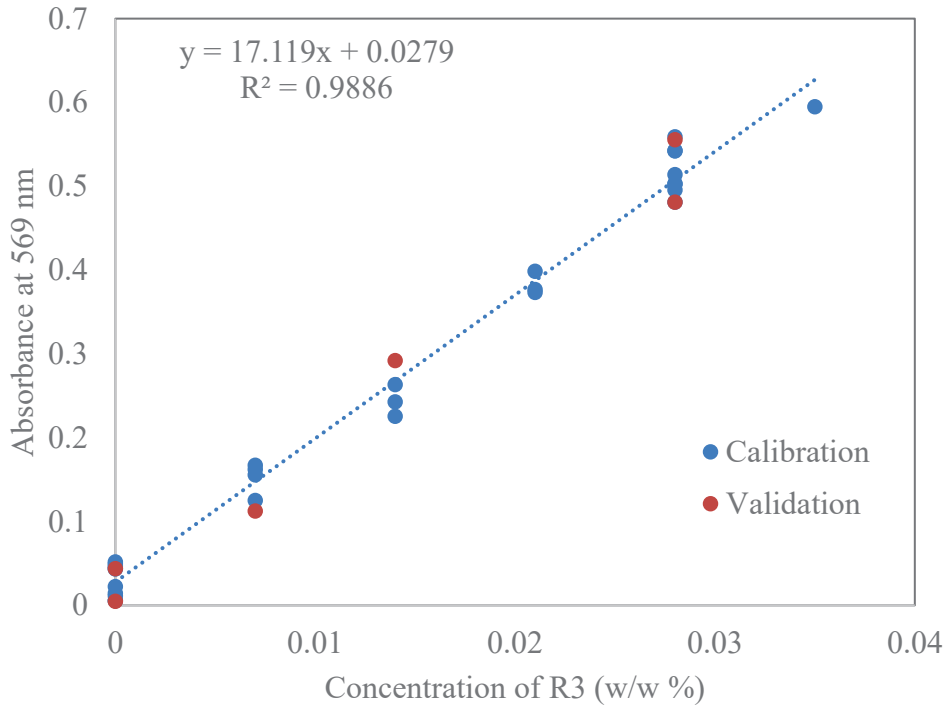


Figure 4.7. Plot of Concentrations of R3 vs. Absorbance at 569 nm by SLR

The regression coefficient of the model was found to be 0.9886 for the calibration set. The standard error of calibration (SEC) and the standard error of prediction (SEP) values were calculated as 0.00135 and 0.00179 (w/w %), for calibration and independent validation sets, respectively.

The results of the SLR method have demonstrated that even with a univariate calibration approach, it is possible to quantitatively determine these color compounds in the cleaning products which contain not only the single dye but also for the binary and ternary mixtures of these color compounds. However, application of multivariate calibration methods such as PLS and GILS could improve the performance of the quantitative determination of these dye compounds as they are able to use whole spectral data instead of focusing only a single maximum absorbing wavelength. The following part gives results of the multivariate calibration studies carried out with PLS and GILS.

4.5. The PLS Results of the R1 (Tartrazine, Yellow)

In the second part of this study, the PLS method was applied to the same the datasets as calibration and independent validation containing the visible spectra (400 to 700 nm) of 35 samples. Predictions were made using the PLS method for R1(Tartrazine)

and absolute error values were given for both calibration and independent validation set in Table 4.1 and 4.2, respectively.

Table 4.1. Results of Calibration Set for R1 by using PLS

Calibration Data			
No	**References R1**	**Predicted R1**	**Absolute Error**
s1	0.069	0.085	0.016
s2	0.055	0.050	0.005
s3	0.055	0.049	0.006
s4	0.041	0.051	0.010
s6	0.041	0.038	0.003
s7	0.028	0.027	0.001
s8	0.028	0.027	0.001
s9	0.028	0.027	0.001
s10	0.028	0.024	0.004
s11	0.014	0.014	0.000
s12	0.014	0.015	0.001
s14	0.014	0.012	0.002
s15	0.014	0.013	0.001
s16	0.000	0.002	0.002
s17	0.000	0.000	0.000
s18	0.000	-0.001	0.001
s19	0.000	0.000	0.000
s20	0.000	0.002	0.002
s21	0.000	0.013	0.013
s22	0.055	0.054	0.001
s24	0.055	0.050	0.005
s25	0.055	0.050	0.005
s26	0.055	0.051	0.004
s27	0.055	0.052	0.003
s28	0.000	0.001	0.001
s29	0.000	0.001	0.001
s31	0.055	0.049	0.006
s33	0.000	0.001	0.001
s34	0.000	0.000	0.000

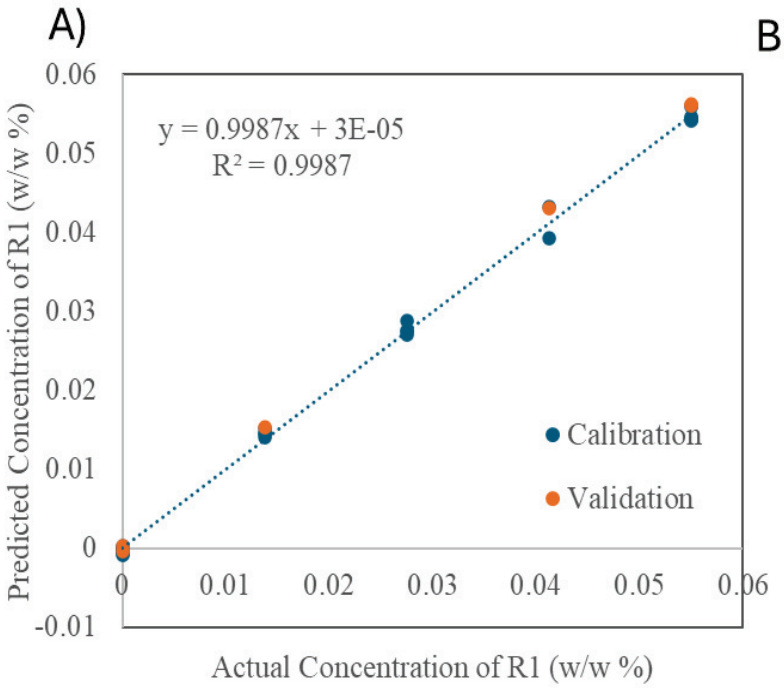
As can be seen in Table 4.1, the reference and predicted values in the calibration set were compared, and then the absolute error values were calculated for each sample in the calibration set. The difference between the estimates and the reference values shows the absolute error value; the smaller this difference is, the better the estimate shows the

result.

Table 4.2. Results of Validation Set for R1 from PLS

Validation Data			
No	**References R1**	**Predicted R1**	**Absolute Error**
s5	0.041	0.029	0.013
s13	0.014	0.016	0.002
s23	0.055	0.053	0.002
s30	0.055	0.049	0.006
s32	0.000	0.001	0.001
s35	0.000	-0.001	0.001

Calibration data set is used as the training set to produce multivariate PLS model with leave one out cross validation. Then independent validation set is used as the test set to verify the performance of the model in the training process, and during the testing process, it can be tested how accurately the model can perform or how well it can generalize across all observations.The result of the PLS calibration model for R1 (Tartrazine) was plotted in Figure 4.8 as actual vs predicted plot for both calibration and validation data sets along with standard error of cross validation (SECV) and standard error of prediction (SEP) for calibration and independent validation sets in addition to the R2 value of the calibration data.



SECV	0.00086
SEP	0.00132
R^2	0.9987

Figure 4.8. A) Actual R1 Concentrations vs. Predicted R1 Concentrations by PLS B) Statistical Parameters of R1

The regression coefficient of the model was found to be 0.9987 for the calibration and validation set, respectively. The standard error of cross validation (SECV) and the standard error of prediction (SEP) values were calculated as 0.00086 and 0.00132 (w/w %), for calibration and independent validation sets, respectively. The PLS method was successfully applied.

Residual plots graph for R1 (Tartrazine) is shown in Figure 4.9 Residual plot graphics were drawn using the PLS method on Minitab16. Residual plots are graphical representations used to analyze the residuals in a regression analysis. These plots help in assessing the goodness of fit of a regression model and in diagnosing potential problems with the model.

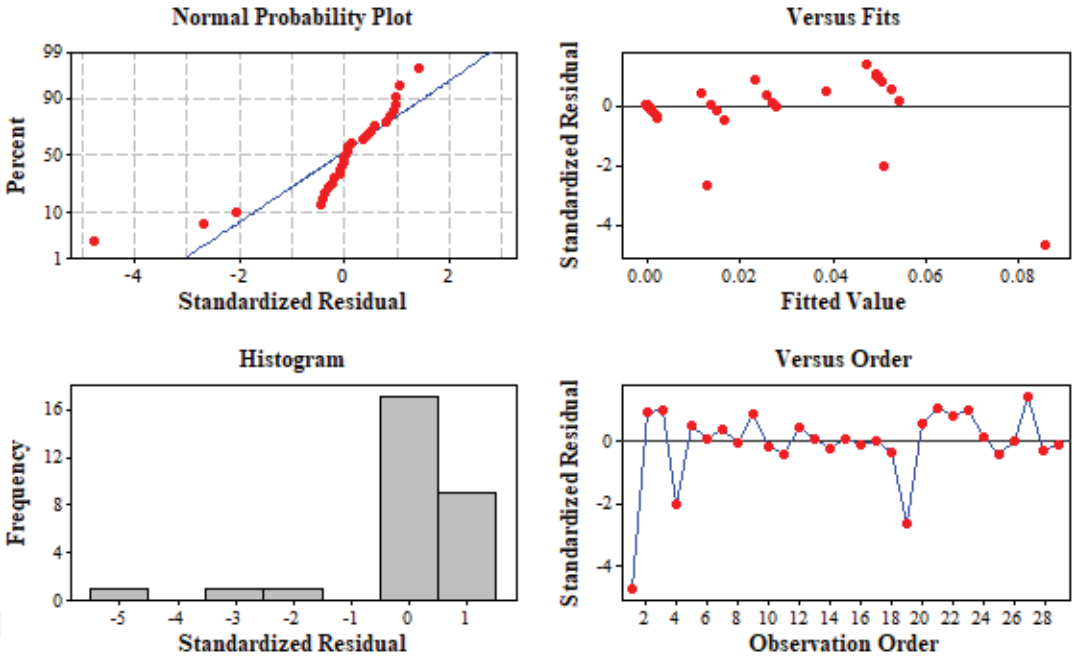


Figure 4.9. Residual Plots for R1

As can be seen Figure 4.9, there are four different residual plots showing the error (residual) analysis of the model. These residual plots show that there are some systematic errors in the residuals of the model and that the residuals do not fully comply with the normal distribution assumption. This indicates that the model needs to be improved or further investigation should be done to determine the source of the errors. Either more variables can be included to improve predictive capability of the model, the parameters are retuned or different modelling is carried out.

PLS Model Selection Plot for R1 (Tartrazine) is shown in Figure 4.10 PLS Model Selection were drawn using the PLS method on Minitab16. The PLS model selection plot is a visual tool used in the context of Partial Least Squares Regression to help determine the optimal number of components to include in the model. The goal is to balance model complexity and predictive accuracy, avoiding overfitting or underfitting the model.

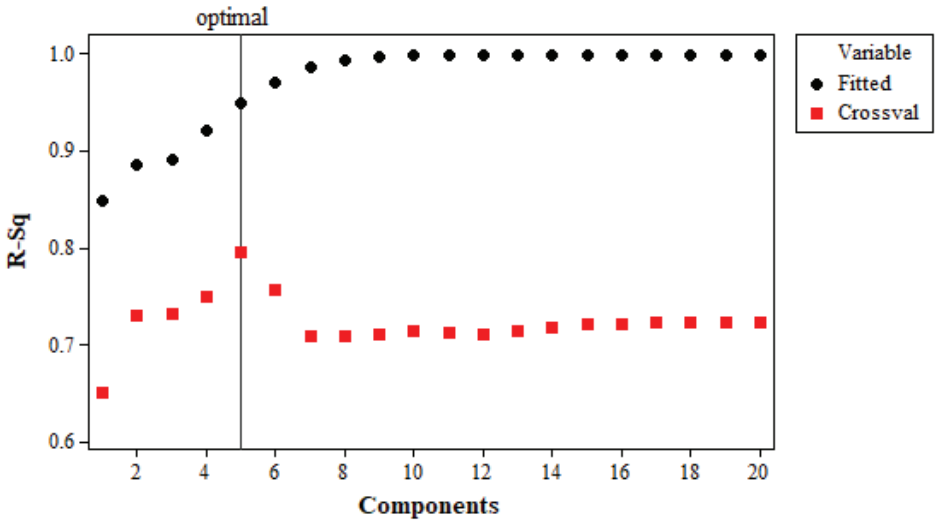


Figure 4.10. PLS Model Selection Plot R1

In this Figure 4.10, PLS shows (R^2) values by number of components for model selection. Black dots represent R^2 values in the model's fitted set, and red squares represent R^2 values in the model's cross-validation set. R^2 values in the training set approach 1 almost linearly as the number of components increases, indicating that the model performs quite well on the training data. However, R^2 values in the cross-validation set reach a constant level after a certain number of components and do not increase. The optimal number of components is approximately 6, because at this point the R^2 value in the cross-validation set reaches its highest level and the addition of additional components does not contribute to the generalizability of the model. This shows that the number of components that need to be taken into consideration to prevent overfitting of the model is 6. This analysis highlights that the number of components must be carefully selected to optimize the model's performance on both training data and new data.

PLS Coefficient Plot for R1 (Tartrazine) were drawn and examined, shown in Figure 4.11 PLS Coefficient plot were drawn using the PLS method on Minitab16. The PLS coefficient plot is a graphical representation used in Partial Least Squares Regression to visualize the importance and direction of each predictor variable's contribution to the model. This plot helps in understanding which variables have the most significant impact on the response variable and whether their relationship is positive or negative.

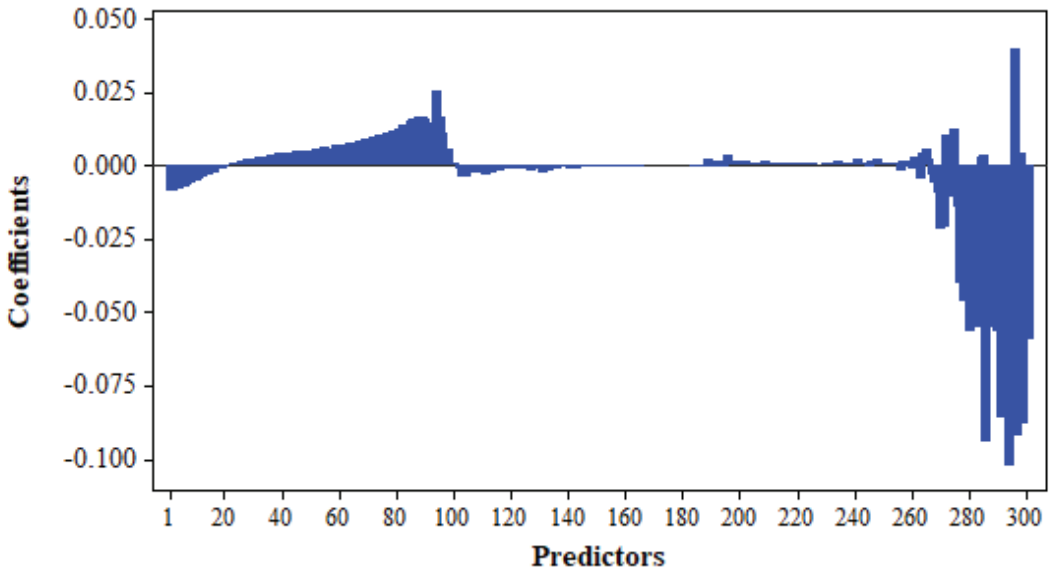


Figure 4.11. PLS Coefficient Plot R1

As can be seen in Figure 4.11, the coefficients that the PLS model estimates using various components for the %R1 response. The 300 predictions on the x-axis represent independent variables; On the y axis are the coefficient values of these predictions. As seen in the graph, coefficient values for most of the predictions are close to zero, indicating that these variables do not play a significant role in the model. However, the coefficient values of some predictions show significant positive or negative deviations. For example, the coefficients of predictions in the ranges 1-100 and 260-300 have higher absolute values, indicating that these predictions have a significant effect in the model. Such higher coefficients indicate that the relevant predictions have a stronger impact on the dependent variable (%R1). Positive coefficients mean that the dependent variable will increase as the predictions increase, and negative coefficients mean that the dependent variable will decrease as the predictions increase. Overall, this graph visually reveals which predictions of the model have a greater impact on the independent variable and which predictions are less significant in the model. This type of analysis increases the interpretability of the model and helps identifies important predictions.

After applying the PLS coefficient plot, the PLS Standard Coefficient plot was drawn, the standardized version of the graph minimizes the error and increases the applicability, this increases the examinability of the data. (Figure 4.12) The PLS standard coefficient plot is a variation of the PLS coefficient plot where the coefficients are standardized. This means that the coefficients are adjusted to have a mean of zero and a

standard deviation of one, making them comparable across different scales of predictor variables. This standardization allows for better interpretation of the relative importance of each predictor variable in the model.

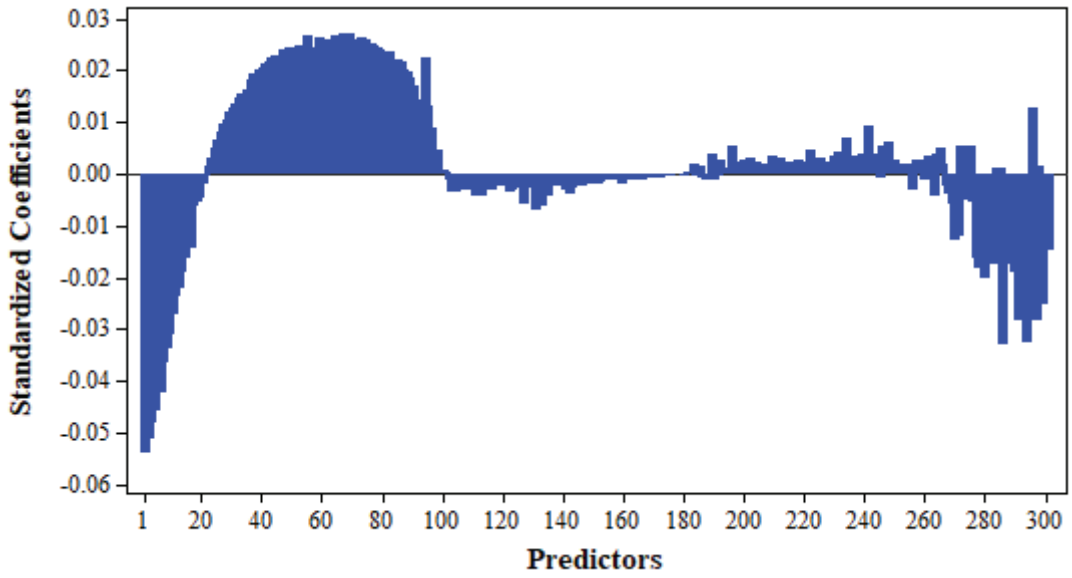


Figure 4.12. PLS Std. Coefficient Plot R1

As can be seen in Figure 4.12, the standardized coefficients of the PLS model for the %R1 response and reveals the effects of the predictions according to the 5 components used in the model. For example, it can be seen that predictions in the range of 20-80 mostly have positive coefficients and these predictions have a positive effect on the dependent variable (%R1). On the other hand, the fact that the predictions in the range of 250-300 are represented by negative coefficients indicates that these predictions have a negative effect on the dependent variable. These deviations in the graph clearly show which predictions the model emphasizes the effect on the independent variable. The surrounding almost zeros for the middle part predictions (80-250) informs that these predictions have little effect in the model and would not consequently induce a significant impact on the dependent variable. As a result, the model becomes more interpretable as it can assess how these predictions may affect the independent variable and the weight of these predictions within the model.

4.6. The PLS Results of the R2 (Brilliant Blue, Blue)

The PLS method was also applied for R2 (Brilliant Blue) using the data set containing the visible spectra (400 to 700 nm) of 35 samples as calibration and independent validation sets, 6 samples were selected as independent validation, and 29 samples were selected for the calibration set. Estimates were made using the PLS method for R2 (Brilliant Blue) and absolute error values were examined (Table 4.3).

Table 4.3. Results of Calibration Set for R2 by using PLS

Calibration Data			
No	**References R2**	**Predicted R2**	**Absolute Error**
s1	0.000	0.000	0.000
s2	0.000	0.000	0.000
s3	0.004	0.004	0.000
s4	0.000	0.000	0.000
s6	0.009	0.008	0.001
s7	0.000	0.000	0.000
s8	0.004	0.003	0.001
s9	0.009	0.009	0.000
s10	0.013	0.012	0.001
s11	0.000	0.000	0.000
s12	0.004	0.004	0.000
s14	0.013	0.016	0.003
s15	0.017	0.017	0.000
s16	0.000	-0.001	0.001
s17	0.004	0.005	0.000
s18	0.009	0.009	0.001
s19	0.013	0.013	0.001
s20	0.017	0.016	0.001
s21	0.021	0.020	0.001
s22	0.017	0.017	0.000
s24	0.017	0.017	0.000
s25	0.017	0.017	0.000
s26	0.000	0.000	0.000
s27	0.000	0.000	0.000
s28	0.017	0.016	0.001
s29	0.017	0.016	0.001
s31	0.000	0.000	0.000
s33	0.017	0.017	0.000
s34	0.000	0.001	0.001

As seen in Table 4.3, the reference and predicted values in the calibration set were

compared, and then absolute error values were calculated for each sample in the calibration set R2 (Brilliant Blue). The difference between the estimates and the reference values shows the absolute error value; the smaller this difference is, the better the estimate shows the result. The absolute error values here show that the estimate was made well.

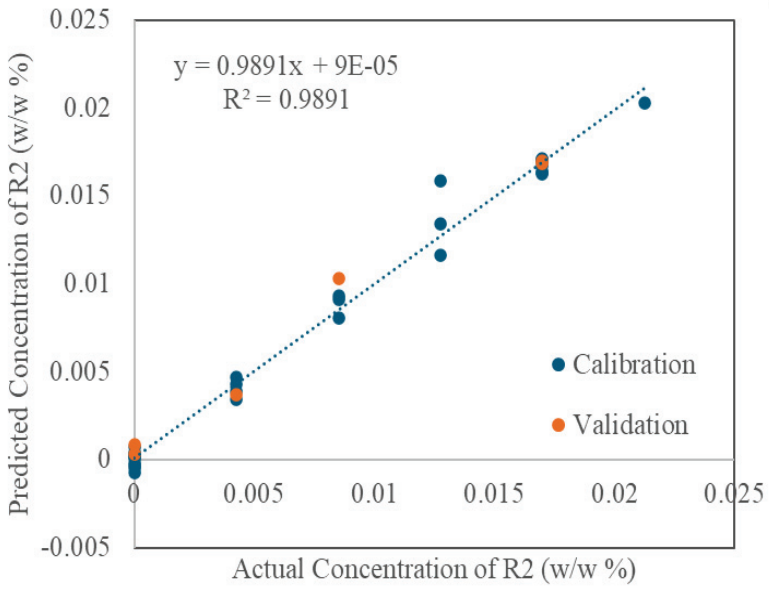
Table 4.4. Results of Validation Set for R2 from PLS

Validation Data			
No	**References R2**	**Predicted R2**	**Absolute Error**
s5	0.004	0.004	0.001
s13	0.009	0.010	0.002
s23	0.017	0.017	0.000
s30	0.000	0.000	0.000
s32	0.017	0.017	0.000
s35	0.000	0.001	0.001

Predictions and references for validation data were compared as seen in Table 4.4. Validation data is generated to evaluate the model's generalization ability, assist in model selection and fine-tuning, perform realistic performance analysis, and provide feedback for improvement. This is a critical step in the modelling process and is a must-use tool to obtain an accurate, reliable and generalizable model.

The result of the PLS calibration model for R2 (Brilliant Blue) was plotted and these results were examined. Calibration and validation data are shown (Figure 4.13).

A)



B)

SECV	0.00080
SEP	0.00093
R2	0.9891

Figure 4.13. A) Actual R2 Concentrations vs. Predicted R2 Concentrations by PLS B) Statistical Parameters of R2

As can be seen Figure 4.13 the relationship between the predicted R2 Brilliant Blue (w/w %) values and the actual R2 (w/w %) values obtained using the calibration and validation sets. The regression coefficient of the model was found to be 0.9891 for the calibration and validation set, respectively. The standard error of calibration (SEC) and the standard error of prediction (SEP) values were calculated as 0.00080 and 0.00093 (w/w %), for calibration and independent validation sets, respectively. The PLS method was successfully applied.

Residual plots graph for R2 (Brilliant Blue) were drawn and examined, shown in Figure 4.14 Residual plot graphics were drawn using the PLS method on Minitab16.

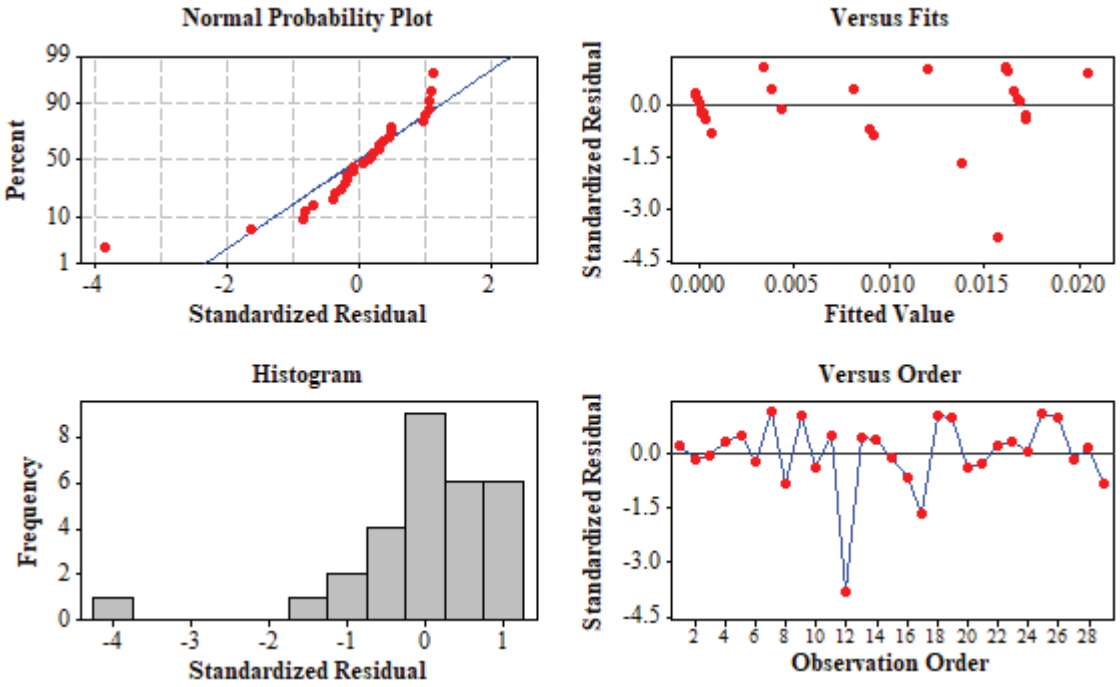


Figure 4.14. Residual Plots for R2

In this Figure 4.14, standardized residuals are listed on the y-axis and plotted on the normal distribution curve on the x-axis. In the top left graph, the majority of the points are aligned along the blue line, indicating that the residuals are normally distributed and comply with the model's assumptions. However, a few extreme points are observed, and these points indicate that the model deviates from the normal distribution in some data.

PLS Model Selection Plot for R2 (Brilliant Blue) were drawn and examined, shown in Figure 4.15 PLS model selection were drawn using the PLS method on Minitab16.

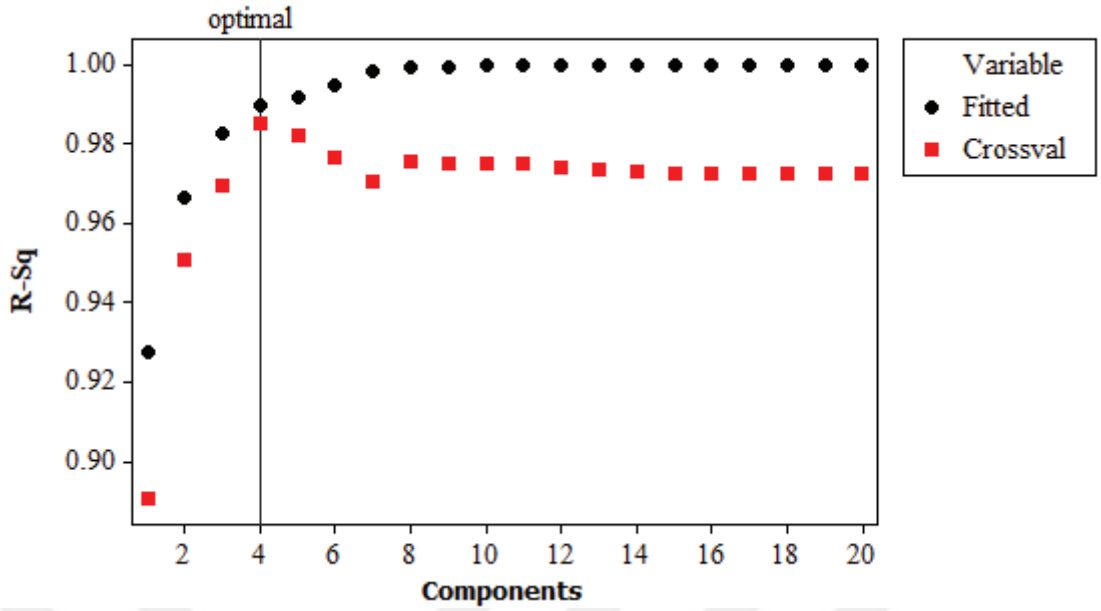


Figure 4.15. PLS Model Selection Plot R2

Shown in this Figure 4.15 are the PLS model's R-squared (R^2) values for R2%. The black dots in the graph represent the R^2 values in the fitted set of the model, and the red squares represent the R^2 values in the cross-validation set of the model. R^2 values in this set rise rapidly as the number of components increases and almost reach 1, indicating that the model provides an excellent fit on the training data. However, R^2 values in the cross-validation set stabilize after approximately 4 components and adding more components does not improve performance. The R^2 value here is approximately 0.9891. This graph shows that the number of components is 4, very good results are observed in the fitted data of the model, and the fact that it shows a more realistic performance in the cross-validation data compared to the fitted data shows that the generalization ability of the model is good and reliable predictions can be made on new data. The result shows that the model exhibits high performance in both calibration and validation data and that the optimal number of components is selected correctly.

PLS Coefficient Plot for R2 (Brilliant Blue) were drawn and examined, shown in Figure 4.16 PLS Coefficient plot were drawn using the PLS method on Minitab16.

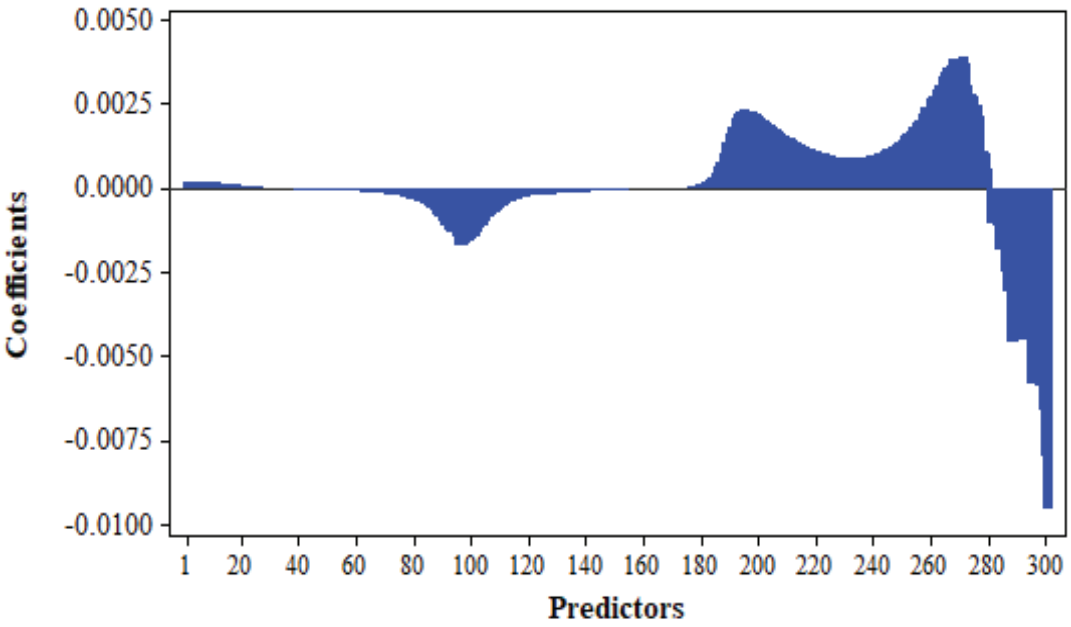


Figure 4.16. PLS Coefficient Plot R2

The majority of the estimates approach zero on a coefficient basis and are shown in the Figure 4.16, but some estimates show significant positive or negative deviations. More precisely, the negative coefficient for interest estimates over time between 80 and 120 is −0.0025, and the other the negative coefficient for interest estimates over time between 280 and 300.

PLS coefficient graph and PLS Standard Coefficient graph are drawn, the standardized version of the graph minimizes the error and increases the applicability, which increases the tractability of the data. As can be shown in Figure 4.17 for R2 (Brilliant Blue).

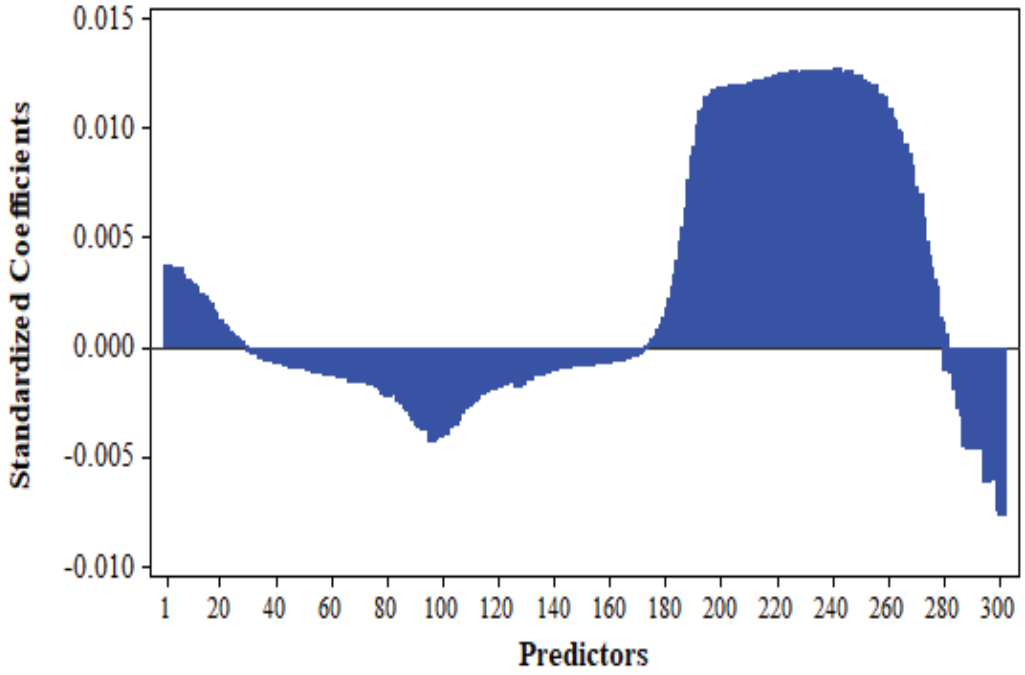


Figure 4.17. PLS Std. Coefficient Plot R2

The PLS standard coefficient plot is a variation of the PLS coefficient plot where the coefficients are standardized. This means that the coefficients are adjusted to have a mean of zero and a standard deviation of one, making them comparable across different scales of predictor variables. This standardization allows for better interpretation of the relative importance of each predictor variable in the model. Data from the graph shows that estimates between 20 and 160 have uniformly negative coefficients, with the lowest coefficient in this range being around -0.0075. On the other hand, there are estimates with positive coefficients in the range of 160-260, the largest of these coefficients being around 0.15. This shows that it has a positive effect on the dependent variable. This also shows that Figure 4.17. the low values in the negative coefficient estimates are between 280 and 300, around -0.01.

4.7. The PLS Results of the R3 (Allura Red, Red)

The PLS method was also applied for R3 (Allura red) using the data set containing visible spectra (400 to 700 nm) of 35 samples as calibration and independent validation sets, 6 samples were selected as independent validation, and 29 samples were selected for the calibration set. Predictions were made using the PLS method for R3 (Allura Red) and

compared with the reference values and absolute error values were examined (Table 4.5).

Table 4.5. Results of Calibration Set for R3 by using PLS

Calibration Data			
No	**References R3**	**Predicted R3**	**Absolute Error**
s1	0.000	0.000	0.000
s2	0.007	0.007	0.000
s3	0.000	0.000	0.000
s4	0.014	0.014	0.000
s6	0.000	0.000	0.000
s7	0.021	0.021	0.000
s8	0.014	0.012	0.002
s9	0.007	0.007	0.000
s10	0.000	0.000	0.000
s11	0.028	0.028	0.000
s12	0.021	0.021	0.000
s14	0.007	0.007	0.000
s15	0.000	0.000	0.000
s16	0.035	0.035	0.000
s17	0.028	0.028	0.000
s18	0.021	0.021	0.000
s19	0.014	0.014	0.000
s20	0.007	0.007	0.000
s21	0.000	0.000	0.000
s22	0.028	0.028	0.000
s24	0.000	0.000	0.000
s25	0.000	0.000	0.000
s26	0.028	0.028	0.000
s27	0.028	0.028	0.000
s28	0.028	0.028	0.000
s29	0.028	0.028	0.000
s31	0.000	0.000	0.000
s33	0.000	0.000	0.000
s34	0.028	0.027	0.001

As seen in Table 4.5, the reference and predicted values in the calibration set were compared and then the absolute error values were calculated for each sample in the calibration set R3 (Allura Red). Here the absolute error values show better results compared to other colors (Tartrazine(R1) and Brilliant Blue(R2)).

Table 4.6. Results of Validation Set for R3 from PLS

Validation Data			
No	**References R3**	**Predicted R3**	**Absolute Error**
s5	0.007	0.007	0.000
s13	0.014	0.015	0.001
s23	0.028	0.028	0.000
s30	0.000	0.000	0.000
s32	0.000	0.000	0.000
s35	0.028	0.027	0.001

The estimates and references for the validation data were compared as seen in Table 4.6 The difference between the estimate data and the reference data was examined and the absolute error values were shown. The absolute error value was found to be low for R3 (Allura Red).

The result of the PLS calibration model for R3 (Allura Red) was plotted and these results were examined. The calibration and validation data are shown (Figure 4.18).

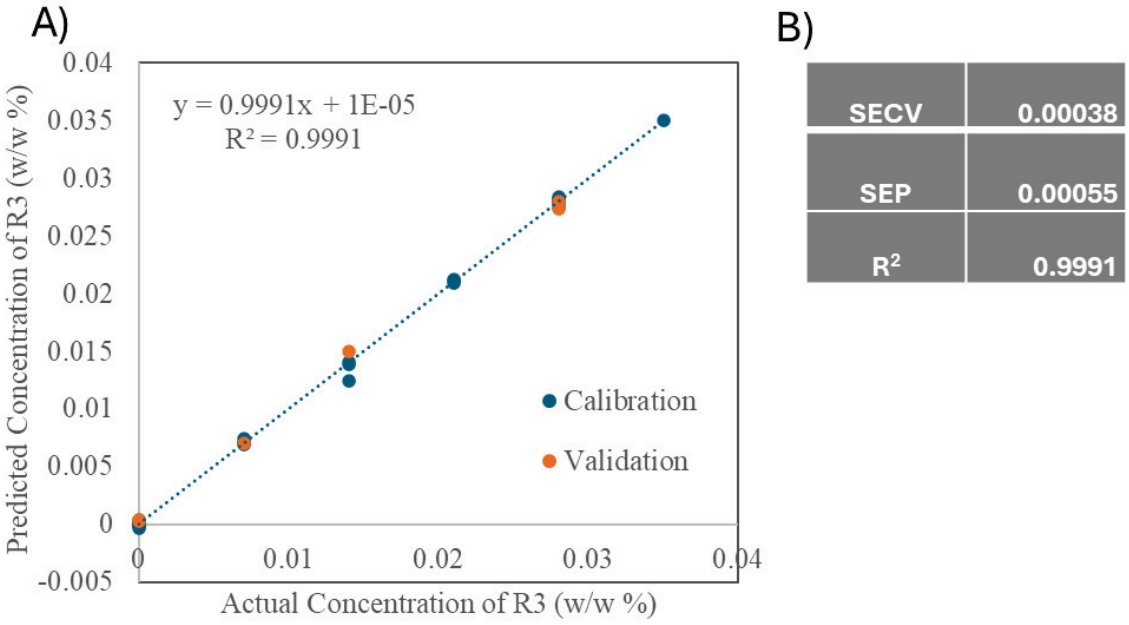


Figure 4.18. A) Actual R3 Concentrations vs. Predicted R3 Concentrations by PLS B) Statistical Parameters of R3

As can be seen Figure 4.18 the relationship between the predicted R3 Allura Red (w/w %) values and the actual R3 (w/w %) values obtained using the calibration and

validation sets. The regression coefficient of the model was found to be 0.9991 for the calibration and validation set, respectively. The standard error of calibration (SEC) and the standard error of prediction (SEP) values were calculated as 0.00038 and 0.00055 (w/w %), for calibration and independent validation sets, respectively. The PLS method was successfully applied.

Residual plots graph for R2 (Allura Red) were drawn and examined, shown in Figure 4.19 Residual plot graphics were drawn using the PLS method on Minitab16.

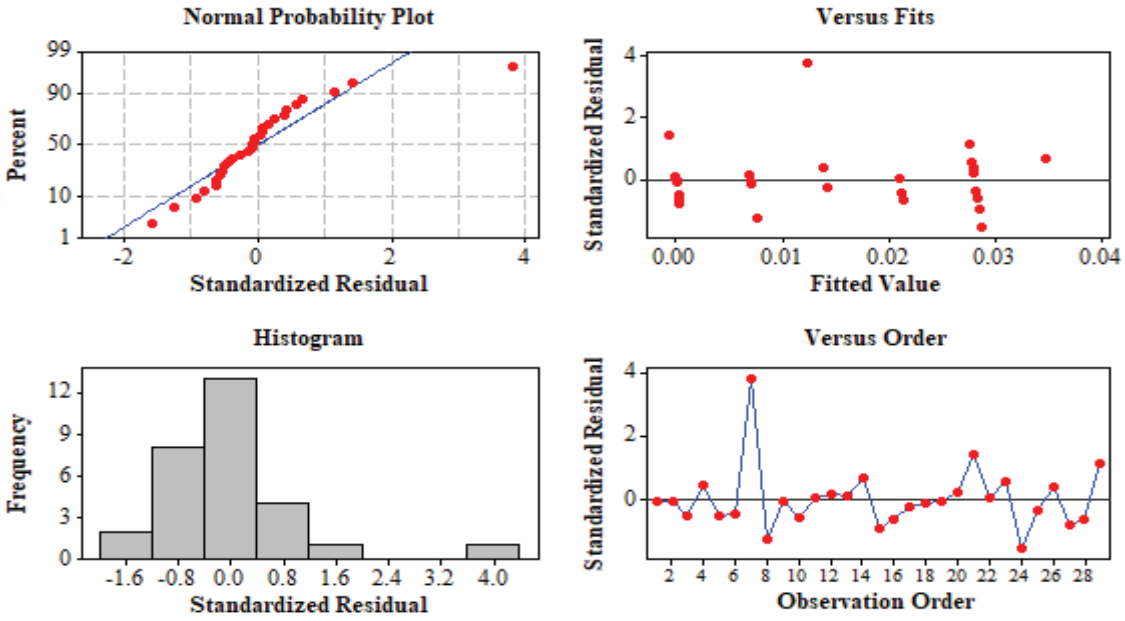


Figure 4.19. Residual Plots for R3

In this Figure 4.19. contains four different plots showing the residual analysis for the model's %R3 response. The performance of the model was evaluated by examining the normal probability plot in the graph, the residuals against the predicted values (versus fits), the residual histogram and the residuals against the observation order. For versus fits and versus order, the residuals are positive or negative, especially in certain observation intervals. It was observed that it concentrated in this direction. This may indicate the presence of some systematic errors in the model depending on time or observation order. But as a result, the accuracy and generalization ability of the model seem to be successful.

PLS Model Selection Plot for R3 (Allura Red) were drawn and examined, shown in Figure 4.20. PLS model selection were drawn using the PLS method on Minitab16.

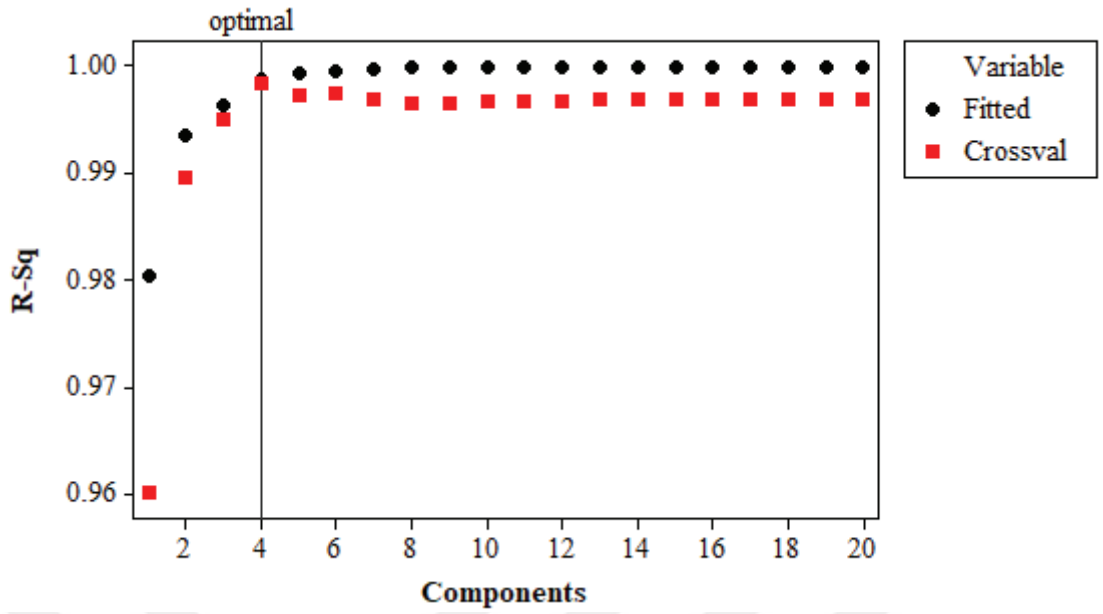


Figure 4.20. PLS Model Selection Plot for R3

This Figure 4.20. shows R-squared (R^2) values by number of components for the R3% response of the PLS (Partial Least Squares) model. The R^2 values in this set rise rapidly as the number of components increases, reaching almost 1, and the model closest to 1 is included in this graph. The model appears to fit the training data perfectly. R^2 values in the cross-validation set reach their highest value of approximately 0.9991 when the number of components reaches 4, and adding more components does not improve performance. At this point, the optimal number of components was determined to be 4 because at this point the model exhibits the best overall performance by preventing overlearning. The graph shows that the model exhibits high performance on both training and validation data and that the optimal number of components is selected correctly.

PLS Coefficient Plot for R3 (Allura Red) were drawn and examined, shown in Figure 4.21 PLS Coefficient plot were drawn using the PLS method on Minitab16.

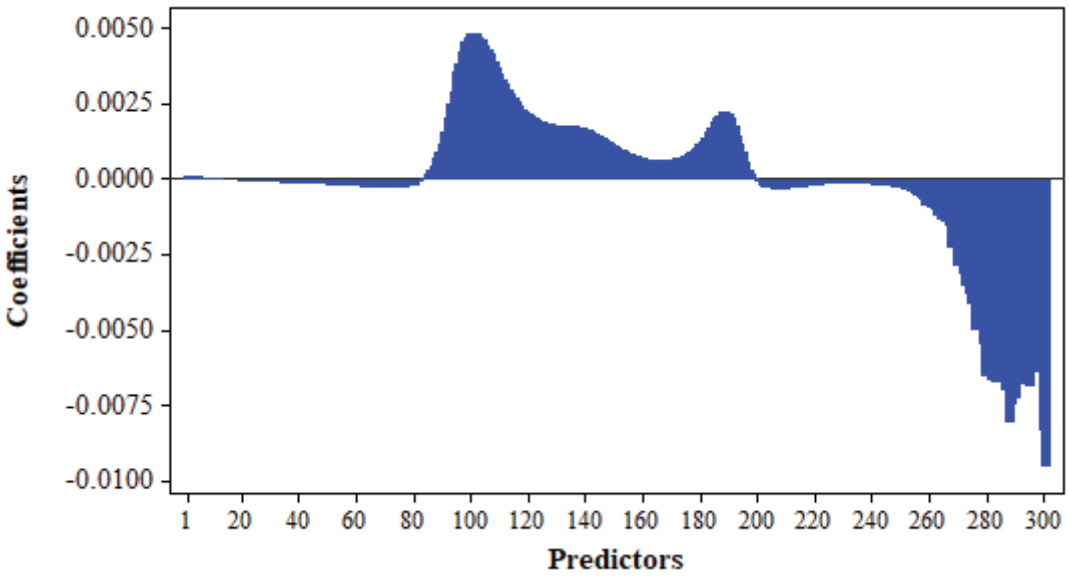


Figure 4.21. PLS Coefficient Plot for R3

In this Figure 4.21, there are 300 predictions on the X axis, and the coefficient values of these predictions are on the Y axis. For most of the estimates in the graph, the coefficients are close to zero, but the coefficients of some estimates show significant positive or negative deviations. In particular, estimates in the range 100-160 have positive coefficients, with the highest value in this range being approximately 0.005. This shows that these estimates have a positive impact on the dependent variable (%R3). Predictions in the range of 260-300 are marked with negative coefficients, and the lowest value in this range is approximately -0.01. This shows that these predictions have a negative impact on the dependent variable. Positive coefficients indicate that the dependent variable will increase as the estimates increase, and negative coefficients indicate that the dependent variable will decrease as the estimates increase.

PLS coefficient graph and PLS Standard Coefficient graph are drawn, the standardized version of the graph minimizes the error and increases the applicability, which increases the tractability of the data. As can be shown in Figure 4.22 for R3 (Allura Red).

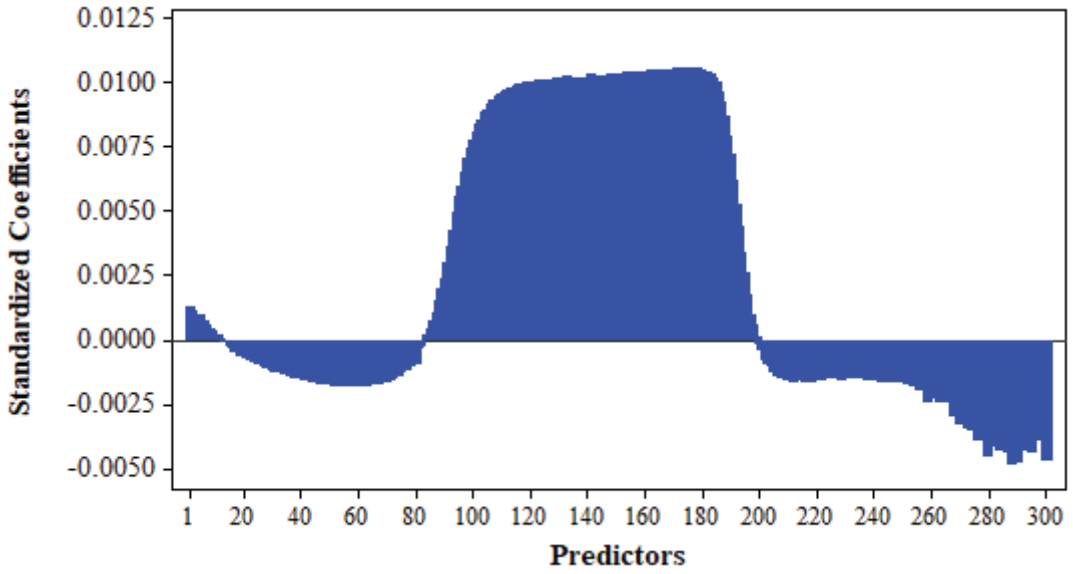


Figure 4.22. PLS Std. Coefficient Plot for R3

In this Figure 4.22, the %R3 response of the PLS model is observed. For most of the predictions in the chart, the coefficients are close to zero, but the coefficients of some predictions show significant positive or negative deviations. In particular, predictions in the range 100-200 have positive coefficients, with the highest value in this range being approximately 0.0125. This shows that the predictions have a positive effect on the dependent variable (%R3). On the other hand, predictions in the ranges 1-80 and 220-300 are marked with negative coefficients, and the lowest value in these ranges is approximately -0.005. This indicates that these predictions have a negative effect on the dependent variable. The general structure of the graph shows that certain predictions have a greater impact in the model and that these predictions play a critical role in predicting the dependent variable. Positive coefficients indicate that the dependent variable will increase as the predictions increase, and negative coefficients indicate that the dependent variable will decrease as the predictions increase. This analysis allows the model to visually evaluate which predictions have a stronger effect on the independent variable and the contribution of these predictions in the model.

4.8. The GILS Results of the R1 (Tartrazine, Yellow)

Thirdly, GILS method was applied, the data set containing visible spectra (400 to 700 nm) of 35 samples was divided into two subsets as calibration and independent

validation sets. Among the 35 samples, 6 were selected as independent validation, mostly repeated samples as given in Table 3.2 the remaining 29 samples were used as calibration sets. Estimates were made for R1 (Tartrazine) using GILS method. (Figure 4.23)

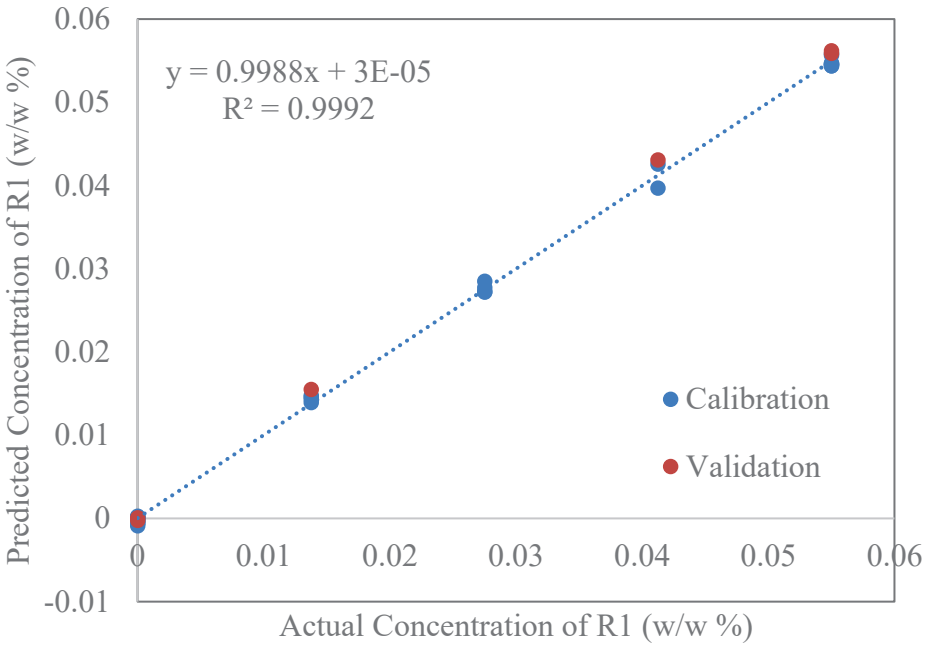


Figure 4.23. Actual R1 Concentrations vs. Predicted R1 Concentrations by GILS

As can be seen Figure 4.23, the regression coefficient of the model for R1 (Tartrazine) was found to be 0.9992. Calibration and validation data for respectively the standard error of calibration (SEC) and the standard error of prediction (SEP) values were calculated as 0.00956 and 0.00121 (w/w %), respectively. The GILS method was successfully applied, and the error ranges were determined and predicted. The most efficient prediction was made with the GILS method. The created model shows good efficiency. High performance is observed compared to other methods. These results also showed that mixtures containing binary and ternary mixtures of these three dyes (Tartrazine yellow, Brilliant blue and Allura red) may be sufficient for the prediction of Tartrazine yellow (R1) color in cleaning products.

4.9. The GILS Results of the R2 (Brilliant Blue, Blue)

The results of the GILS calibration model for R2 (Brilliant Blue) were plotted and examined. Calibration and validation data are shown (Figure 4.24). Predicted values are

plotted versus actual values.

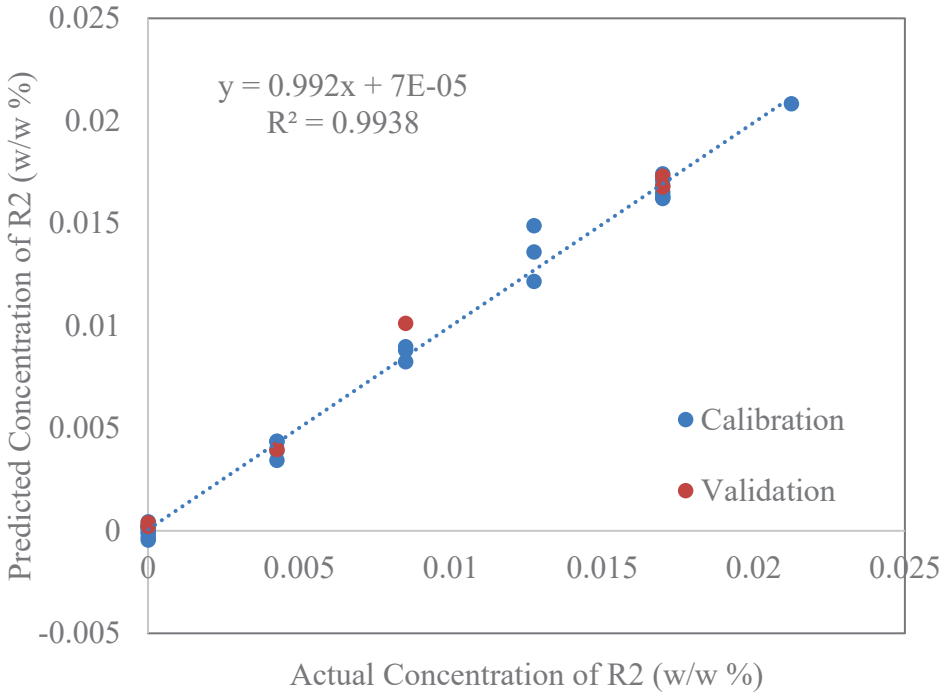


Figure 4.24. Actual R2 Concentrations vs. Predicted R2 Concentrations by GILS

As can be seen Figure 4.24, the regression coefficient of the model for R2 (Brilliant Blue) was found to be 0.9938. Calibration and validation data for respectively the standard error of calibration (SEC) and the standard error of prediction (SEP) values were calculated as 0.00060 and 0.00072 (w/w %), respectively. The GILS method was successfully applied, and the error ranges were determined and predicted. The most efficient prediction was made with the GILS method.

4.10. The GILS Results of the R3 (Allura Red, Red)

The results of the GILS calibration model for R3 (Allura Red) were plotted and examined. Calibration and validation data are shown (Figure 4.25). Predicted values are plotted versus actual values.

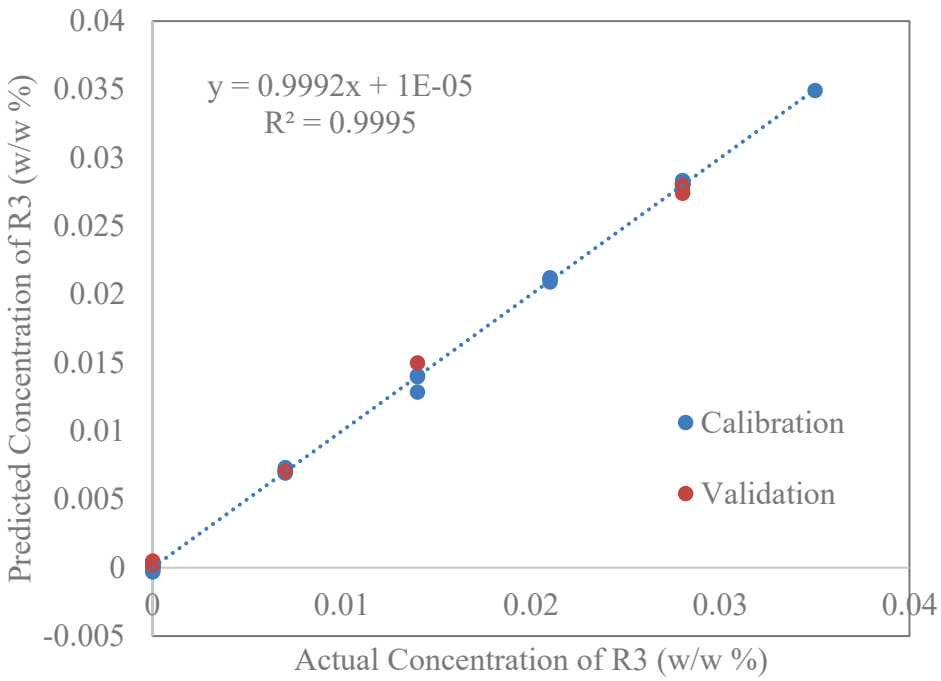


Figure 4.25. Actual R3 Concentrations vs. Predicted R3 Concentrations by GILS

As can be seen Figure 4.25, the regression coefficient of the model for R3 (Allura Red) was found to be 0.9995. Calibration and validation data for respectively the standard error of calibration (SEC) and the standard error of prediction (SEP) values were calculated as 0.00029 and 0.00052 (w/w %), respectively. The GILS method was successfully applied, and the error ranges were determined and predicted. The most efficient prediction was made with the GILS method.

The most efficient prediction was made with the GILS method. The model created shows good efficiency. High performance is observed compared to other methods. As a result of all analyses, an almost perfect prediction was made as a result of the modelling applied according to the results for R3 (Allura Red). GILS made as a result of comparisons, it creates the most efficient model and provides predictions.

4.11. Summary of Values for R1, R2 and R3

As can be seen in Table 4.7, the results of the applied SLR, PLS and GILS methods are analyzed. According to the results, the most efficient prediction method was determined and the Standard Error of Cross-Validation (SEC), Standard Error of

Prediction (SEP) and Regression Coefficient (R^2) values were compared among them.

Table 4.7. Summary of Values for R1, R2 and R3

COLOR	Simple Least Squares Regression Calibration Model Results			Partial Least Squares Calibration Model Results			Genetic Inverse Least Squares Calibration Model Results		
	SECV w/w%	SEP w/w%	R^2	SECV w/w%	SEP w/w%	R^2	SECV w/w%	SEP w/w%	R^2
R1 (Yellow)	0.00966	0.00183	0.9956	0.00086	0.00132	0.9987	0.00956	0.0012	0.9992
R2 (Blue)	0.00089	0.00109	0.9868	0.00080	0.00093	0.9891	0.00060	0.0007	0.9938
R3 (Red)	0.00135	0.00179	0.9886	0.00038	0.00055	0.9991	0.00029	0.0005	0.9995

Standard Error of Cross-Validation, Standard Error of Prediction and R^2 values of the methods used were compared among themselves. To calculate the SEC value, it is necessary to calculate the standard deviation of the differences between the model's predictions and the actual values. The smaller this value, the better the predictive ability of the model. SEP, on the other hand, tests the overall performance and reliability of the model, and these two values serve the same purpose. The R^2 value is an important measure of performance that indicates how well a model explains the variance of the dependent variable. An R^2 value close to 1 indicates that the model explains the dependent variable well and can make predictions with high accuracy. It is seen that GILS gives the healthiest results for these three methods. SEP and SEC values are the ones with low values and R^2 value is the highest among these values.

CHAPTER 5

CONCLUSIONS

As a result, three different dye amounts (R1, R2 and R3) were estimated in the surface cleaning product in this study. These colors are yellow (R1, Tartrazine), blue (R2, Brilliant Blue) and red (R3, Allura Red) to obtain various absorbance values at different wavelengths. Using the mixture design method, 35 different surface cleaner samples containing various concentrations of dyes were prepared. The composition of these samples was determined in the visible region (400-700 nm). The dye amounts of the prepared samples were taken as reference and these references were estimated using the chemometric modelling methods SLR, PLS and GILS. Data were generated using the raw UV-Visible spectra of the samples. Calibration and validation data were generated for each dye before applying the methods. Several graphs were plotted and interpreted based on each chemometric modelling. These interpretations were used to compare how close the predictions were to the reference values. Among the three methods, the GILS method appears to be the most efficient modelling and provides the best prediction. The regression coefficients (R^2) of the GILS model ranged from 0.9938 to 0.9995, the standard error of the cross-validation (SEC) values ranged from 0.00029 to 0.00956 and the standard error of the prediction (SEP) ranged from 0.00052 to 0.00121 (%w/w). Finally, the models developed in this study perform well, with the GILS model performing well, making the best prediction and also providing very good results.

REFERENCES

1. Long, D. C. Greening of Consumer Cleaning Products. In *Green Techniques for Organic Synthesis and Medicinal Chemistry*, 2018; pp 91-115. DOI: 10.1002/9781119288152.ch5

2. Otter, J. A.; Vickery, K.; Walker, J. T.; deLancey Pulcini, E.; Stoodley, P.; Goldenberg, S. D.; Salkeld, J. A. G.; Chewins, J.; Yezli, S.; Edgeworth, J. D. Surface-attached cells, biofilms and biocide susceptibility: implications for hospital cleaning and & disinfection. *J. Hosp. Infect.* **2015**, *89* (1), 16-27. DOI: 10.1016/j.jhin.2014.09.008

3. Gerster, F. M.; Vernez, D.; Wild, P. P.; Hopf, N. B. Hazardous substances in frequently used professional cleaning products. *Int. J. Occup. Med. Environ. Health* **2014**, *20* (1), 46-60. DOI: 10.1179/2049396713Y.0000000052

4. Van der Maarel, M. J. E. C.; van der Veen, B.; Uitdehaag, J. C. M.; Leemhuis, H.; Dijkhuizen, L. Properties and applications of starch-converting enzymes of the α-amylase family. *J. Biotechnol.* **2002**, *94* (2), 137-155. DOI: 10.1016/S0168-1656(01)00407-2

5. Dufossé, L. Chapter 19 - Current and potential natural pigments from microorganisms (bacteria, yeasts, fungi, and microalgae). In *Handbook on Natural Pigments in Food and Beverages (Second Edition)*, Schweiggert, R. Ed.; Woodhead Publishing, 2024; pp 419-436. DOI: 10.1016/B978-0-323-99608-2.00001-X

6. Teasdale, S. M.; Kademi, A. Quality challenges associated with microbial-based cleaning products from the Industry Perspective. *Food Chem. Toxicol.* **2018**, *116*, 20-24. DOI: 10.1016/j.fct.2017.10.029

7. Nazaroff, W. W.; Weschler, C. J. Cleaning products and air fresheners: exposure to primary and secondary air pollutants. *Atmos. Environ.* **2004**, *38* (18), 2841-2865. DOI: 10.1016/j.atmosenv.2004.02.040

8. Primary Colors Red, Yellow, and Blue Use as Self-Cleaning Paints with Layer of Titanium Dioxide. *Baghdad Sci. J.* **2019**, *16* (4), 0903. DOI: 10.21123/bsj.2019.16.4.0903

9. Ismadji, S.; Soetaredjo, F. E.; Ayucitra, A. Natural Clay Minerals as Environmental Cleaning Agents. In *Clay Materials for Environmental Remediation*, Ismadji, S., Soetaredjo, F. E., Ayucitra, A. Eds.; Springer International Publishing, 2015; pp 5-37. DOI: 10.1007/978-3-319-16712-1_2

10. Shahid, M.; Shahid ul, I.; Mohammad, F. Recent advancements in natural dye applications: a review. *J. Clean. Prod.* **2013**, *53*, 310-331. DOI: 10.1016/j.jclepro.2013.03.031

11. Haji, A.; Naebe, M. Cleaner dyeing of textiles using plasma treatment and natural dyes: A review. *J. Clean. Prod.* **2020**, *265*, 121866. DOI: 10.1016/j.jclepro.2020.121866

12. Christie, R. *Colour Chemistry*; The Royal Society of Chemistry, 2014. DOI: 10.1039/9781782626510

13. European Food Safety, A. Refined exposure assessment for Allura Red AC (E 129). *EFSA J.* **2015**, *13* (2), 4007. DOI: 10.2903/j.efsa.2015.4007

14. Ullah, R.; Iftikhar, F. J.; Ajmal, M.; Shah, A.; Akhter, M. S.; Ullah, H.; Waseem, A. Modified Clays as an Efficient Adsorbent for Brilliant Green, Ethyl Violet and Allura Red Dyes: Kinetic and Thermodynamic Studies. *Pol. J. Environ. Stud.* **2020**, *29* (5), 3831-3839. DOI: 10.15244/pjoes/112363

15. Scotter, M. J. Methods for the determination of European Union-permitted added natural colours in foods: a review. *Food Additives & Contaminants: Part A* **2011**, *28* (5), 527-596. DOI: 10.1080/19440049.2011.555844

16. Pifferi, G.; Restani, P. The safety of pharmaceutical excipients. *Il Farmaco* **2003**, *58* (8), 541-550. DOI: 10.1016/S0014-827X(03)00079-X

17. Adeel, S.; Abrar, S.; Kiran, S.; Farooq, T.; Gulzar, T.; Jamal, M. Sustainable Application of Natural Dyes in Cosmetic Industry. In *Handbook of Renewable Materials for Coloration and Finishing*, 2018; pp 189-211. DOI: 10.1002/9781119407850

18. Additives, E. P. o. F.; Nutrient Sources added to, F. Scientific Opinion on the re-evaluation of Brilliant Blue FCF (E 133) as a food additive. *EFSA J.* **2010**, *8* (11), 1853. DOI: 10.2903/j.efsa.2010.1853

19. Akash, M. S. H.; Rehman, K. Ultraviolet-Visible (UV-VIS) Spectroscopy. In *Essentials of Pharmaceutical Analysis*, Akash, M. S. H., Rehman, K. Eds.; Springer Nature Singapore, 2020; pp 29-56. DOI: 10.1007/978-981-15-1547-7_3

20. Talsky, G.; Mayring, L.; Kreuzer, H. High-Resolution, Higher-Order UV/VIS Derivative Spectrophotometry. *Angew. Chem., Int. Ed. Engl.* **1978**, *17* (11), 785-799. DOI: 10.1002/anie.197807853

21. Mantele, W.; Deniz, E. UV–VIS absorption spectroscopy: Lambert-Beer reloaded. *Spectrochim Acta A Mol Biomol Spectrosc* **2017**, *173*, 965-968. DOI: 10.1016/j.saa.2016.09.037

22. Kaur, G.; Singh, H.; Singh, J. Chapter 2 - UV-vis spectrophotometry for environmental and industrial analysis. In *Green Sustainable Process for Chemical and Environmental Engineering and Science*, Inamuddin, Boddula, R., Asiri, A. M. Eds.; Elsevier, 2021; pp 49-68. DOI: 10.1016/B978-0-12-821883-9.00004-7

23. Wamsley, M.; Zou, S.; Zhang, D. Advancing Evidence-Based Data Interpretation in UV–Vis and Fluorescence Analysis for Nanomaterials: An Analytical Chemistry Perspective. *Anal. Chem.* **2023**, *95* (48), 17426-17437. DOI: 10.1021/acs.analchem.3c03490

24. Prajapati, P. Chemometrics and its Applications in UV Spectrophotometry. *Int. J. Pharm. Chem. Anal.* **2016**, *3*, 43-48. DOI: 10.5958/2394-2797.2016.00005.8

25. Brown, S. D.; Blank, T. B.; Sum, S. T.; Weyer, L. G. Chemometrics. *Anal. Chem.* **1994**, *66* (12), 315-359. DOI: 10.1021/ac00084a014

26. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2* (1), 37-52. DOI: 10.1016/0169-7439(87)80084-9

27. Geladi, P.; Kowalski, B. R. Partial least-squares regression: a tutorial. *Anal. Chim. Acta* **1986**, *185*, 1-17. DOI: 10.1016/0003-2670(86)80028-9

28. Lavine, B. K.; Workman, J., Jr. Chemometrics. *Anal. Chem.* **2013**, *85* (2), 705-714. DOI: 10.1021/ac303193j

29. Principal Component Analysis for Special Types of Data. In *Principal Component Analysis*, Jolliffe, I. T. Ed.; Springer New York, 2002; pp 338-372. DOI: 10.1007/0-387-22440-8_13

30. Abdi, H.; Williams, L. J. Principal component analysis. *WIREs Comp Stats* **2010**, *2* (4), 433-459. DOI: 10.1002/wics.101

31. Jolliffe, I. T.; Cadima, J. Principal component analysis: a review and recent developments. *Philos. Transact. A Math. Phys. Eng. Sci* **2016**, *374* (2065), 20150202. DOI: 10.1098/rsta.2015.0202

32. Hawkins, D. L.; Gallant, A. R.; Fuller, W. A simple least squares method for estimating a change in mean. *Commun. Stat. Simul. Comput.* **1986**, *15* (3), 523-530. DOI: 10.1080/03610918608812531

33. de Jong, S. SIMPLS: An alternative approach to partial least squares regression. *Chemom. Intell. Lab. Syst.* **1993**, *18* (3), 251-263. DOI: 10.1016/0169-7439(93)85002-X

34. Hubert, M.; Branden, K. V. Robust methods for partial least squares regression. *J. Chemom.* **2003**, *17* (10), 537-549. DOI: 10.1002/cem.822

35. Linder, M.; Sundberg, R. Second-order calibration: bilinear least squares regression and a simple alternative. *Chemom. Intell. Lab. Syst.* **1998**, *42* (1), 159-178. DOI: 10.1016/S0169-7439(98)00032-X

36. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58* (2), 109-130. DOI: 10.1016/S0169-7439(01)00155-1

37. Haaland, D. M.; Thomas, E. V. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Anal. Chem.* **1988**, *60* (11), 1193-1202. DOI: 10.1021/ac00162a020

38. Abdi, H.; Williams, L. J. Partial Least Squares Methods: Partial Least Squares Correlation and Partial Least Square Regression. In *Computational Toxicology: Volume II*, Reisfeld, B., Mayeno, A. N. Eds.; Humana Press, 2013; pp 549-579. DOI: 10.1007/978-1-62703-059-5_23

39. Qi, P.; Lin, Z.-h.; Chen, G.-y.; Xiao, J.; Liang, Z.-a.; Luo, L.-n.; Zhou, J.; Zhang, X.-w. Fast and simultaneous determination of eleven synthetic color additives in flour and meat products by liquid chromatography coupled with diode-array detector and tandem mass spectrometry. *Food Chem.* **2015**, *181*, 101-110. DOI: 10.1016/j.foodchem.2015.02.075

40. Pekcan Ertokus, G. Determination of the Colorants in Various Samples by Chemometric Methods Using Statistical Chemistry. *Iran. J. Chem. Chem. Eng.* **2018**, *37* (3), 127-134. DOI: 10.30492/ijcce.2018.34168

41. Ni, Y.; Gong, X. Simultaneous spectrophotometric determination of mixtures of food colorants. *Anal. Chim. Acta* **1997**, *354* (1), 163-171. DOI: 10.1016/S0003-2670(97)00297-3