Near-infrared spectroscopy and chemometrics for rapid profiling of plant secondary metabolites*

Ntebogeng S. Mokgalaka‡, Sello P. Lepule, Thierry Regnier, and Sandra Combrinck

Department of Chemistry, Tshwane University of Technology, Private Bag X680, Pretoria, 0001, South Africa

Abstract: In this study, near-infrared (NIR) spectroscopy, in combination with chemometrics, was used as a rapid tool for determining if exposure to contamination from mine tailings influences the matrices of the specimens, compared to those from natural populations. Principal component analysis (PCA) plots were made from the chemometric models obtained to establish if season of harvest, geographical origin, and level of soil contamination play a determining role in the chemical profiles of the individual specimens harvested from mine sites or natural populations. The random distribution on PCA score plots corroborated the intraspecies variation of Lippia scaberrima previously observed by gas chromatography-flame ionization detection (GC-FID) and gas chromatography-mass spectrometry (GC-MS) essential oil profiles. Clustering according to the season and origin of the individual plants confirmed that the geographic location and the season of harvest influence the chemical profiles of L. scaberrima. The NIR data could not be correlated with the level of soil contamination to which the specimens were exposed. The PCA scores and loadings plots obtained from NIR data of Searsia pendulina suggest that the species is tolerant to pollution from mine tailings. Although separation was obtained in a three-component PCA model between specimens sampled during different seasons, some clustering was observed by specimens from the same geographical origin.

Keywords: analytical chemistry; chemometrics; mine tailings; near infrared spectroscopy; plants; secondary metabolites; vibrational spectroscopy.

INTRODUCTION

Vibrational spectroscopy techniques, particularly near-infrared (NIR), mid-infrared (MIR), and Raman, have gained momentum as analytical tools for rapid profiling of valuable plant chemical compounds [1–3]. Modern high-resolution spectrometers allow fast scanning over a wide wavelength range, thus increasing the sample throughput rate [4]. Vibrational spectroscopy techniques can be used for analyzing solid, liquid, or gaseous samples in a nondestructive manner [5]. These techniques can be considered as green tools for characterizing the chemical nature of the plant matrix, since they reduce or eliminate the use of hazardous solvents associated with extraction and metabolite profiling using...
chromatographic techniques [4,5]. Although NIR spectroscopy data is generally only meaningful after application of chemometric algorithms, MIR and Raman spectra present characteristic key bands that can be used as markers to discriminate different plant chemotypes [6].

Vibrational spectroscopy is well established as a valuable technique in a wide range of industries, including in the food [7,8], pharmaceutical [9,10], and petrochemical [11] industries, as well as in clinical [11] environmental [12] applications and in process control [13]. Recently, vibrational spectroscopy has been applied to the quality control and chemical profiling of medicinal plants [14,15].

Medicinal plants derive their unique properties from specific active secondary chemical constituents such as alkaloids [16], terpenoids [17,18], phenolic compounds [19,20], and sterols [21,22]. The profiles of secondary metabolites may vary, depending on the genetic traits, geographic origin, growing season, and the growing environment, which include climate and nutrient availability [23]. Therefore, rapid techniques for chemical profiling of secondary metabolites are crucial for determining optimum yield of desirable metabolites and for monitoring the quality of the plant-derived products. Essential oils, containing mainly volatile terpenoids, are produced world-wide by the distillation of aromatic plants. Attenuated total reflectance (ATR) Fourier transform infrared (FT-IR) [17], NIR [18], MIR [18,24], and Raman [17,25] spectroscopic methods have been established as acceptable methods for the rapid profiling, chemotaxonomy, identification and quality control of essential oils. Using Raman spectroscopy, seasonal, genetic variations and chemotypes were easily observed through characteristic key bands in essential oils isolated from basil [25,26], chamomile [26], thyme and oregano plants [26], without the need for chemometric manipulation. Portable FT-IR spectrometers are useful for rapid monitoring of wild populations of plants, because they allow in situ profiling of secondary metabolites, including volatile organic compounds, in the field [27].

Flavonoids are the most studied secondary metabolites and are of phenolic nature. Several authors have reported the use of NIRS to determine the antioxidant activity [21,22,28] and the amount of total polyphenols [28] in green tea leaves. Recently, NIR spectroscopy (NIRS) was used to determine the total flavonoid content of Ginkgo biloba, a medicinal plant used for the treatment of respiratory ailments and improvement of peripheral blood flow [20]. The harvest time influenced the flavonoid content in the leaves of G. biloba. Therefore, NIRS provided a rapid, non-destructive tool that is useful for determining when to harvest for the best-quality G. biloba leaves to ensure optimum yield of total flavonoids.

The multitude of analytical information contained in NIR spectra can be extracted by using multivariate analysis methods that relate analytical variables to analyte properties. Chemometric modeling has become a vital tool in vibrational spectroscopy measurements allowing the extraction of useful information from noisy signals [3,6,11]. Classification methods, such as principal component analysis (PCA), are used for qualitative multivariate data analysis methods for “pattern recognition”. These methods establish similarities and differences between samples expressed as the correlation coefficient (R) between samples. PCA was applied to NIR spectra of green tea leaves to discriminate tea leaves of different age [29], while NIR spectra of coffee beans originating from different geographical regions allowed regional classification [30].

Lippia scaberrima Sonder is a hardy aromatic shrub that naturally establishes on disturbed and contaminated soils. The plant produces a wide array of pharmacologically active volatile and non-volatile secondary metabolites. Combrinck et al. [31] studied the composition of L. scaberrima essential oil and found limonene and carvone to be the main components of the oil. The nonvolatile secondary metabolites in the infusions of aerial plant parts of L. scaberrima, prepared as a tonic or used for the treatment of stomach ailments, are more closely associated with its medicinal use [32]. These compounds include the phenyl ethanoidglycosides, verbascoside, and isoverbascoside [33], and an iridoidglycoside, known as theveridoside [34].

The compositions of the secondary metabolites of Lippia are highly variable. Lepule [35] identified specimens of L. scaberrima that produced high levels of isopiperitenone, rather than carvone. Both of these metabolites originate from limonene in the biosynthetic pathway, but prevailing conditions...
probably determine whether carvone is produced via carveol, or isopiperitenone via isopiperitenol [36]. Olivier et al. [33] reported that the levels of verbascoside and isoverbascoside were higher in leaves than other plant parts and that these values varied considerably, even within a single population of Lippia.

Searsia pendulina, formerly known as Rhus pendulina, is a popular garden shade tree. Very little information on the chemical constituents produced by the tree is available, although those from several other Searsia species have been thoroughly investigated [37]. The bark of S. pendulina is used for tanning, and a milk infusion of leaves is administered to children complaining of stomach ailments (www.plantzafrica.com).

In an effort to establish plants with valuable secondary metabolites on mine tailings, as part of a phytoremediation program, we studied the secondary metabolites profiles of acid-tolerant L. scaberrima ecotypes and S. pendulina growing on gold mine tailings. In this study, NIRS in combination with chemometrics was used for rapid profiling of secondary metabolites in L. scaberrima and S. pendulina leaf materials, without the need for solvent extraction. PCA of NIR profiles was used to determine if the time of harvest (season), the geographic location, and level of soil contamination influence the profiles of the individual specimens harvested from mine sites and natural populations.

**SAMPLING**

**L. scaberrima**

Specimens of L. scaberrima were collected in 2007 and 2008 from known populations [38] at the Vaal River and West Wits mining operations of AngloGold Ashanti Ltd. in the Free State and North West provinces of South Africa, as well as from sites far enough from the mine to exclude the possibility of mine contamination (natural populations). Five plant specimens were collected from each location. Witkowski and Weiersbye [39] characterized sites in the vicinity of tailings storage facilities based on their relative soil solution pH (aq) and total dissolved solids (TDS). Specimens of L. scaberrima were collected from some of these sites, which differed significantly in soil contamination status. Sites were regarded as having low contamination (pH > 6, TDS < 500 mg/L); moderate contamination (pH > 4 < 6, TDS > 1000 < 2000 mg/L); and high contamination (pH < 4, TDS > 3000 mg/L) as outlined by Weiersbye and Witkowski [39]. Aerial plant parts were harvested in late spring (November), late summer (February), and late autumn (May) with garden clippers. Care was taken not to damage the root systems of the plants so that they could continue to grow and could be revisited, to determine seasonal variation of the secondary metabolites. To enable sampling from the same plant, GPS coordinates were recorded.

**S. pendulina**

Several experimental sites were established in 2001, as part of the Mine Woodlands Rehabilitation Programme, on and around the tailings dam complexes of different mining operations. Leaf samples of S. pendulina were collected from the Vaal River mine (VRM) and West Wits operations. The Vaal River mine and West Wits operations comprise four and three experimental sites, respectively. Each site is situated in a summer rainfall and frost area. Only the sites Moab Khotsong (Vaal River mine) Mponeng (West Wits Red Soil), both highly contaminated, and TauTona (West Wits Shallow Soil), less contaminated, were sampled. Trees, of approximately the same age, representing natural growing environments were sampled from gardens in Gauteng (Johannesburg and Pretoria).

Mine remediation sites, divided into plots at the start of the rehabilitation program, consist of 63 trees planted in a 9 × 7 m grid. In each plot, each tree is 2.5 m from the adjacent trees and 3 m away from trees in the next row. Trees were numbered according to their position on the plot. Adjoining plots were populated with different species to allow for differences in exposure to mine seepage. Twenty-three plots were sampled and leaves were collected from both mining operations in November,
February, and May, over the growing season 2007 to 2008. The natural populations were sampled for comparison in late summer (February 2008). Five individual specimens were sampled from each locality. Tree numbers were randomly selected for sampling, and trees with the same numbers were harvested from every site. Forty leaves were removed from the canopy, 1.5 m above the ground. Ten leaves were collected from each aspect of the tree to avoid variability resulting from differences in exposure to sunlight. The leaves were placed in labeled envelopes and sealed.

NIR SPECTRA ACQUISITION AND CHEMOMETRICS

Each specimen was rinsed gently under running tap water to remove surface dust. Samples were oven-dried at 35 °C and subsequently milled to a fine powder using a coffee grinder (Russell Hobbs, Model No. 9715), sieved (<500 μm particle size), and stored in glass vials until required. The NIR spectra of ground leaves obtained from L. scaberrima and S. pendulina, harvested from both natural population and mine areas, were collected using a NIRFlex N500 FT NIR spectrometer (Büchi, Labortechnik AG, Switzerland) with NIRWare software version 1.2.3000 advanced edition. Duplicate NIR reflectance spectra of the samples were collected between 10000 and 4000 cm⁻¹ (32 scans per sample) at a spectral resolution of 4 cm⁻¹. Spectral data were exported to Microsoft Excel® 2003, whereafter duplicated measurements were averaged for each data point.

PCA models were constructed for the NIR data using SIMCA-P+ Version 12.0 (Unimetrics, Sweden). Mean centering of the NIR spectra was applied throughout.

RESULTS AND DISCUSSION

L. scaberrima

Typical spectra of natural populations and specimens of L. scaberrima collected from highly contaminated mine sites are depicted in Fig. 1. These spectra appeared similar between 4000 and 7000 cm⁻¹, however, the specimen from the West Wits highly contaminated site indicated higher intensities between 7000 and 10000 cm⁻¹. To determine if the time of harvest (season) influences the chemical profiles of the individual L. scaberrima specimens, PCA models were constructed from the NIR spectral data. A two-component PCA model was constructed using the season (spring, summer, and autumn) as secondary observations (Fig. 2). The PCA scores plot indicates a random distribution of profiles for L. scaberrima specimens. Clustering is observed on the plot for only six specimens harvested in autumn. These specimens differed to such an extent from the other samples that five of these appear as outliers on the plot. Removal of these outliers did not improve the model substantially. The differences may be due to a number of unknown external factors such as precipitation patterns or temperature vari-

Fig. 1 Typical NIR spectra of leaf material from mine-contaminated (VRM HC and WW HC) and natural populations (Potch NP and VR NP) of L. scaberrima.
ables. Having reached no conclusion regarding the effect of the season on the secondary metabolites profiles of *L. scaberrima* specimens, we decided to investigate the effect of the geographic location and level of soil contamination during each season.

To reveal the influence of soil contamination on the chemical profiles of *L. scaberrima*, individual PCA models were constructed from NIR spectra of plants harvested in autumn, spring, and summer, respectively. The PCA plot for autumn (Fig. 3) revealed the tightest clustering of population groups compared to those for spring (Fig. 4) and summer (Fig. 5). The natural population from Vaal River separated in the first component from the others, and the Vaal River specimens (natural and contaminated) separated from the others in the second component. It appears that geographical origin, rather than level of contamination, plays a role in the differences observed. This was confirmed by a distinction on the PCA scores plot (Fig. 3) between the West Wits and the Vaal River mine samples. This observation confirms our earlier findings [40] that *L. scaberrima* plants harvested from West Wits sites produced highly variable essential oil compositions, even within a single population. The plants were all healthy specimens and did not appear to be detrimentally affected by the conditions of the highly contaminated sites. Some researchers have reported that exposure to metal contamination may change the chemical composition of a plant, thereby affecting the bioactivity [41].

![PCA scores plot of NIR data obtained from *L. scaberrima* specimens harvested from mine and natural populations in autumn (red), spring (green), and summer (blue).](image1)

**Fig. 2** PCA scores plot of NIR data obtained from *L. scaberrima* specimens harvested from mine and natural populations in autumn (red), spring (green), and summer (blue).

![PCA scores plot indicating variations in the secondary metabolites profiles of *L. scaberrima* harvested in autumn from mine and natural areas. The origin of each specimen is indicated by a triangle: red = natural population (West Wits), black = natural population (Vaal River), blue = (West Wits mine) and green = (Vaal River mine).](image2)

**Fig. 3** PCA scores plot indicating variations in the secondary metabolites profiles of *L. scaberrima* harvested in autumn from mine and natural areas. The origin of each specimen is indicated by a triangle: red = natural population (West Wits), black = natural population (Vaal River), blue = (West Wits mine) and green = (Vaal River mine).
The PCA model of NIR spectra of the samples harvested in spring is illustrated in Fig. 4. Both natural populations clustered together, overlapping with some of the Vaal River mine samples, while the West Wits population was separated in the first component from the Vaal River population. Specimens from the Vaal River sites exhibited the highest variability of secondary metabolites profiles and agreed with the variations observed in the essential oil profiles reported in our earlier work [40]. The clustering observed appears to be linked to the geographical location, rather than to the level of contamination, since the West Wits samples, originating from highly contaminated soils, are clustered with the natural populations.

For summer (Fig. 5), the two natural populations were separated from each other in the first component. Despite exhibiting the greatest variation, the Vaal River specimens were separated from the other samples in the second component. Once again, the West Wits mine and natural populations clustered together, confirming that clustering appears to be linked to the geographical location, rather than to the level of contamination.

**Fig. 4** PCA scores plot indicating variations in the secondary metabolite profiles of *L. scaberrima* harvested in spring from mine and natural areas. The origin of each specimen is indicated by the triangle: red = natural population (West Wits), black = natural population (Vaal River), blue = (West Wits mine) and green = (Vaal River mine).

**Fig. 5** PCA scores plot indicating variations in the secondary metabolite profiles of *L. scaberrima* harvested in summer from mine and natural areas. The origin of each specimen is indicated by a triangle: red = natural population (West Wits), black = natural population (Vaal River), blue = (West Wits mine) and green = (Vaal River mine).
The profiles of the two natural populations displayed similarities in spring, but they were completely different in summer and autumn. This was ascribed to the different growth phases of the plants. These plants originated from different areas and, depending on the prevailing temperatures and rainfall, may have been at different growth stages. These stages include the active growth phase, the flowering and seeding stages or winter dormancy. Schulz et al. [29] used PCA of NIR data to discriminate between tea leaves of different ages. They attributed the separation obtained to differences in the antioxidant flavanol (epigallocatechin gallate and epicatechin gallate) levels of the leaves. The younger leaves contained higher concentrations of both compounds than mature leaves. The variability observed in the current study confirms that growth stage plays a more important role in determining the plant matrix of *L. scaberrima* plants than the level of contamination that the plant is exposed to. The three PCA plots (Figs. 3–5) indicate little clustering within contamination groups, but more clustering associated with the geographical origin of the plants. In some cases, specimens from highly contaminated areas clustered with specimens exposed to low or no contamination.

*S. pendulina*

A three-component PCA model was constructed after removal of 7 outliers (total *n* = 210) of which 50.02 % (R2X = 0.5002) of the variation in the data was explained by the first component, 44.64 % (R2X = 0.4464) by the second component and only 3.37 % (R2X = 0.0337) by the third. The cumulative Q value (94.4 %) reflects the percentage variation in the data that can be predicted by the model. These values indicate that the model is valid. The PCA scores plot (Fig. 6A) shows some separation of the summer and winter samples collected from mine sites. However, the spring samples reflected greater variability in matrix composition. These observations are emphasized by the loadings plot obtained (Fig. 6B). The loadings plot indicates clear separation and negative correlation between samples collected in February and May, while those harvested in November show a positive and negative spread when compared to the others. Seasonal changes, such as temperature and precipitation, clearly determine the chemical make-up of these samples.

When PCA models were constructed for each individual harvesting season, the influence of season was more evident. Clustering of samples collected from the same geographic area is observed on the PCA scores plot for November (Fig. 7A). The Vaal River and West Wits Red Soil samples are separated by the second component, yet they were all exposed to high levels of contamination. This result infers that the clustering and separation is probably due to differences in their environments, rather than to exposure to contamination. The loadings plot (Fig. 7B) clearly demonstrates the similarity of the chemical compositions of the West Wits plants, irrespective of soil contamination levels, and the differences between these plants and those from the Vaal River mine.

To establish the role of soil contamination, a PCA model including natural populations (no exposure to contamination) was constructed for samples collected in February. The similarities in the matrices of specimens from West Wits Red Soil (high contamination) is evident from the PCA scores and loadings plots (Figs. 8A and B). A large degree of variation is observed in the natural populations as reflected by the loadings plot. Specimens exposed to high levels of contamination (West Wits Red Soil and Vaal River samples) did not cluster, indicating that exposure to contamination did not trigger the enhanced production of specific metabolites. *S. pendulina* appears to be resilient in the presence of contaminated soils. This tree is therefore well adapted to polluted soils, making it a good candidate for remediation purposes.

It is therefore concluded that the geographic origin, and to a lesser extent, the season of harvest, influence the chemical profiles of *L. scaberrima* plants. PCAs were useful for comparing the chemical profiles of *S. pendulina* from mine tailings and those from natural populations. Geographic origin appeared to have the biggest influence on sample matrices, rather than level of exposure to contamination. PCA is an unsupervised method that provides a useful tool to elucidate similarities and differences in the NIR data [42]. Without meticulously scrutinizing the data, NIRS-PCA provided an overview of
Fig. 6 (A) PCA scores plot indicating variations in the secondary metabolite profiles of *S. pendulina* harvested from sites contaminated by mine tailings, as reflected by NIR data, in summer (red triangles), autumn (green triangles), and spring (blue triangles). (B) Loadings plot corresponding to (A).
Fig. 7 (A) PCA scores plot indicating the clustering of NIR data from *S. pendulina* samples originating from the same geographical location. All samples were harvested in spring (November). Vaal River mine (highly contaminated) = red, West Wits Red Soil (highly contaminated) = green, and West Wits Shallow Soil (moderately contaminated) = blue. (B) Loadings plot corresponding to (A).
ACKNOWLEDGMENTS

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the authors and not necessarily to be attributed to the NRF. The authors also thank Isabel Weiersbye and AngloGold Ashanti; the data and indicated that the chemical profiles of specimens from highly contaminated and uncontaminated areas could not be distinguished.

FIG. 8 (A) PCA score plot indicating the clustering of NIR data from *S. pendulina* samples originating from contaminated mine tailings and natural populations. All samples were harvested in summer (February). Vaal River mine (highly contaminated) = dark green; West Wits Red Soil (highly contaminated) = light blue, and West Wits Shallow Soil (moderately contaminated) = brown; Natural population 1 = red, Natural population 2 = bright green, and Natural population 3 = yellow, Natural population 4 = pink, and Natural population 5 = blue. (B) Loadings plot corresponding to (A).
Department of Trade and Industry-THRIP for funding the project; and Ben Oageng and Jacob Mahlangu for field assistance.

REFERENCES