International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 3, May.-June, 2024, pp: 1558-1572 www.ijhssm.org

Data Acquisition and Analysis of *Panax Notoginseng* Using Near-Infrared Spectroscopy

XuefengCheng[1,2], Abudhahir Buhari[2], HongmeiZhu[1],and Juan Liu[3*]
[1]*School of Big Data and Information Industry, Chongqing City Management College, Chongqing, China*
[2]*Faculty of Engineering, Science & Technology, Infrastructure University Kuala Lumpur, Malaysia*
[3]*School of Artificial Intelligence, Chongqing University of Education, Chongqing, China.*
**Corresponding Author*

Date of Submission: 08-06-2024 Date of Acceptance: 22-06-2024

ABSTRACT: This study explores the methods of data acquisition and analysis of *Panax Notoginseng* using near-infrared spectroscopy (NIRS). The roots of *Panax Notoginseng* contain various bioactive compounds such as saponins, flavonoids, polysaccharides, and amino acids, which are believed to have significant therapeutic effects. Ensuring the quality and authenticity of *Panax Notoginseng* products is of utmost importance. Traditional analytical methods such as thin-layer chromatography, high-performance liquid chromatography, and mass spectrometry, although accurate, are often time-consuming, labor-intensive, and require extensive sample preparation. In contrast, NIRS offers advantages such as minimal sample preparation, faster analysis times, and non-destructive testing. This study utilized a Bruker MATRIX-F FT-NIR spectrometer to collect representative samples from major producing areas in Yunnan Province, which were cleaned, dried, and ground to ensure uniform analysis. The results showed that NIRS can effectively identify and quantify key constituents of *Panax Notoginseng* and detect subtle differences in samples from different regions.

This study demonstrates the potential of NIRS in the quality control and authenticity verification of *Panax Notoginseng*, providing a new method for the rapid and non-destructive analysis of traditional Chinese medicinal materials.

KEYWORDS:*Panax Notoginseng*,data acquisition,feature analysis, near-infrared spectroscopy.

I. INTRODUCTION

The roots of *Panax Notoginseng* contain a variety of bioactive compounds, including saponins, flavonoids, polysaccharides, and amino acids, which are believed to contribute to its therapeutic effects. Due to its significant medicinal properties, ensuring the quality and authenticity of *Panax Notoginseng* products is of utmost importance. In the realm of traditional Chinese medicine, the quality control of herbal materials such as *Panax Notoginseng* involves assessing their chemical composition to verify the presence and concentration of active compounds.

Traditional methods of analysis, such as thin-layer chromatography, high-performance liquid chromatography, gas chromatography, mass spectrometry, Raman spectroscopy, and nuclear magnetic resonance, although accurate and reliable, are often time-consuming, labor-intensive, and require extensive sample preparation. Various high-performance thin-layer chromatography and thin-layer chromatography methods have been developed for the analysis and quality control of different compounds and products. For instance, a high-performance thin-layer chromatography method accurately estimated zingerone in ginger tablets [1]. Similarly, other studies have validated methods for different compounds, ensuring reliability and precision in quality control [2-7].

High-performance liquid chromatography and related techniques also play a crucial role in the analysis and quality control of various compounds and products. For example, carbohydrate compositions in lignocellulosic biomasses were compared using high-performance liquid chromatography and gas chromatography, revealing significant discrepancies in sugar content detection [8]. Other studies optimized high-performance liquid chromatography methods for analyzing synthetic pigments in food [9], detecting aflatoxins in coffee [10], and purifying taxanes from Taxus cuspidata [11-15].

Gas chromatography-mass spectrometry and high-performance liquid chromatography-mass spectrometry are versatile and effective for analyzing various compounds across diverse

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 3, May.-June, 2024, pp: 1558-1572 www.ijhssm.org

contexts. These techniques have been used to characterize the chemical composition of apple by-products [16], profile phenolic and volatile compounds in balsamic vinegars [17], and detect amino acids in Martian samples [18-20].

Advanced mass spectrometry techniques have been optimized for proteomic and microbial analysis. For instance, proteomic analysis of tear fluid using the Orbitrap Fusion Tribrid mass spectrometer has enhanced biomarker discovery for ocular diseases [21]. Other studies have utilized high-resolution mass spectrometry for identifying pathogenic yeast [22-24].

Raman spectroscopy and its enhanced variant, surface-enhanced Raman spectroscopy, have diverse applications across various fields. These techniques have been used for detecting water contaminants [25], differentiating glycosylation patterns in glioma samples [26], and studying metal-amino acid interactions [27-29].

Nuclear magnetic resonance techniques have been employed in various fields, including monitoring the cheese ripening process [30], measuring moisture content in coal powders [31], and identifying impurities in carbon nanodot samples [32]. These methods have also been used for verifying the authenticity of extra virgin olive oil [33-35].

NIRS has emerged as a promising alternative for the rapid and non-destructive analysis of medicinal herbs [36-37]. This technique measures the absorption of near-infrared light by the sample, providing information about its molecular composition based on the vibrational transitions of chemical bonds. NIRS offers several advantages over traditional methods, including minimal sample preparation, faster analysis times, and the ability to analyze multiple components simultaneously.

In the context of *Panax Notoginseng*, NIRS can be particularly beneficial for quality control and standardization. By capturing the spectral data of the herb, NIRS enables the identification and quantification of its key constituents, facilitating the assessment of its overall quality. Furthermore, this technique can help detect adulterants and contaminants, ensuring the authenticity and safety of *Panax Notoginseng* products in the market.

This paper focuses on the application of NIRS in the data acquisition and analysis of *Panax Notoginseng* roots. It explores the process of collecting and preparing samples, the methodology of NIRS, and the techniques for analyzing spectral data. Additionally, the paper compares NIRS with other conventional analytical methods, highlighting its advantages and potential limitations. By providing a comprehensive overview of NIRS's role in the analysis of *Panax Notoginseng*, this paper aims to underscore its significance in the quality control of traditional Chinese medicinal materials.

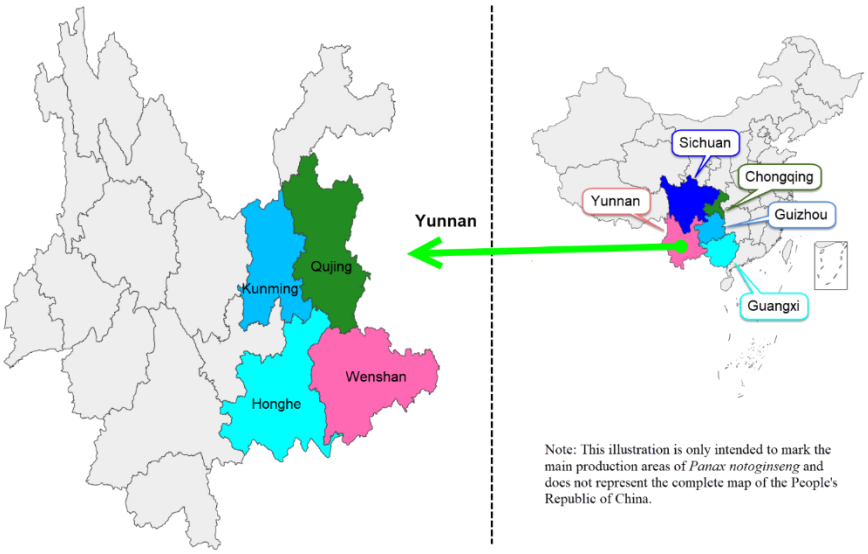

FIGURE 1: THE DISTRIBUTION OF *PANAX NOTOGINSENG* PRODUCTION AREAS IN CHINA

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 3, May.-June, 2024, pp: 1558-1572 www.ijhssm.org

II. SAMPLE COLLECTION AND PREPARATION

In China, *Panax Notoginseng* cultivation is widely distributed in provinces such as Yunnan, Guangxi, Sichuan, Guizhou, and Chongqing, the distribution of *Panax Notoginseng* production areas in China are shown in right side of Figure 1. The regions of Wenshan, Honghe, Kunming, Qujing, and Yuxi in Yunnan province are recognized as the main production areas of *Panax Notoginseng*, contributing to 98% of the total national production, Wenshan in Yunnan is renowned for producing the highest quality and accounts for 60% of these five regions, left side of Figure 1 shows the distribution of the five majors *Panax Notoginseng* production areas in Yunnan province.

TABLE 1: ROOT SAMPLES OF PANAX NOTOGINSENG FROM DIFFERENT REGIONS

REGIONS	NUMBER OF ROOT SAMPLES
Wenshan	25
Honghe	23
Kunming	18
Qujing	16
Yuxi	20

Samples of *Panax Notoginseng* were collected from five major production regions in Yunnan Province: Wenshan, Honghe, Kunming, Qujing, and Yuxi. These regions are renowned for their unique environmental conditions that influence the quality and chemical composition of *Panax Notoginseng*. The collection involved gathering the main roots of *Panax Notoginseng* from multiple locations within each region to ensure a representative sample set. The specific quantities collected from each region are as following Table 1.

The samples were then carefully prepared for NIRS analysis. Preparation involved cleaning the roots to remove soil and debris, followed by drying at a controlled temperature to prevent any alteration in chemical composition. The dried samples were ground into a fine powder to ensure uniformity in the near-infrared measurements.

III. DATA ACQUISITION PROCESS

Common Data Acquisition Methods. In the identification of the geographical origin of *Panax Notoginseng*, various analytical methods are utilized for data acquisition:

Thin-layer chromatography: *Panax Notoginseng* root samples are pulverized and extracted with a suitable solvent such as methanol or ethanol. The extract is filtered, and a small amount of the supernatant is spotted onto a thin-layer chromatography plate. The separation is achieved using a mobile phase, and spots are visualized under ultraviolet light or with a staining reagent for qualitative analysis.

High-performance liquid chromatography: The powdered *Panax Notoginseng* samples are subjected to ultrasonic extraction using high-purity solvents like a water-acetonitrile mixture. After centrifugation, the supernatant is collected and injected into the high-performance liquid chromatography system. The components are separated on a chromatographic column and detected using an ultraviolet or mass spectrometry detector for quantitative analysis of active compounds like saponins.

Gas chromatography: The *Panax Notoginseng* samples are pulverized and extracted with an organic solvent. The extract is then concentrated and derivatized if necessary to improve volatility. The prepared sample is injected into the gas chromatography system, where it is vaporized and separated on a gas chromatography column. Detection is typically performed using a flame ionization detector or mass spectrometry for the analysis of volatile compounds.

Mass spectrometry: The powdered *Panax Notoginseng* samples are extracted with a suitable solvent and, if required, subjected to chromatographic separation prior to mass spectrometry analysis. The extract is introduced into the mass spectrometer, where the compounds are ionized, fragmented, and their mass-to-charge ratios are measured. This provides detailed structural information about the chemical constituents.

Raman spectroscopy: The *Panax Notoginseng* samples are dried and ground into a fine powder to ensure uniformity. The powdered sample is placed on a Raman microscope stage, and Raman spectrum are collected directly without further preparation. This method allows for the non-destructive identification of molecular vibrations and chemical composition.

Nuclear magnetic resonance: *Panax Notoginseng* samples are dried, ground into a fine powder, and dissolved in an appropriate deuterated solvent. The solution is transferred to a nuclear magnetic resonance tube for analysis. Nuclear magnetic resonance spectroscopy provides detailed

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 3, May.-June, 2024, pp: 1558-1572 www.ijhssm.org

information on the molecular structure and composition by measuring the interaction of nuclear spins with an external magnetic field.

TABLE 2: COMPARISON OF DIFFERENT DATA ACQUISITION METHODS FOR *PANAX NOTOGINSENG*

METHODS	COMPLEXITY AND TIME REQUIRED	COST	SENSITIVITY AND SPECIFICITY	INFORMATION PROVIDED
TLC	Simple, quick; minimal equipment needed	Low	Lower sensitivity and specificity	Primarily qualitative
HPLC	High resolution and quantitative, but time-consuming	High	High sensitivity and specificity	Detailed compositional profiles
GC	Effective for volatile compounds; complex for non-volatiles	High	High sensitivity and specificity	Detailed compositional profiles
MS	Complex sample preparation; time-consuming	Very High	Extremely high sensitivity and specificity	Detailed structural information
RS	Minimal preparation; non-destructive; quick	Moderate	Moderate sensitivity; affected by fluorescence	Molecular vibrations and structure
NMR	Extensive preparation; time-consuming	Very High	High specificity; moderate sensitivity	Detailed molecular structure
NIR	Minimal preparation; rapid; non-destructive	Moderate to High	Moderate sensitivity and specificity	Broad compositional data

Near-infrared spectroscopy:*Panax Notoginseng* samples are dried and ground into a fine powder to ensure uniformity. The powdered sample is placed in a sample holder, and the NIRS spectrometer collects spectrum by shining near-infrared light on the sample. The absorption of specific wavelengths by the sample provides information about its chemical composition, including moisture content, carbohydrate structures, and other organic compounds. This method is non-destructive and requires minimal sample preparation, making it suitable for rapid and comprehensive analysis of the samples' constituents.

Compared to other methods like thin-layer chromatography, high-performance liquid chromatography, gas chromatography, mass spectrometry, Raman spectroscopy, and nuclear magnetic resonance, NIRS offers several advantages as shown in Table 2.

NIRS stands out due to its minimal sample preparation and rapid, non-destructive analysis. It provides broad compositional data efficiently and cost-effectively, making it ideal for routine analysis. Additionally, NIRS is environmentally friendly, as it does not require solvents or chemicals. This method's ability to deliver quick results with minimal effort and expense, while preserving sample integrity, highlights its significant advantages in various analytical applications.

Principle and Experimental Equipment of Near-Infrared Spectroscopy Detection Technology. Near-infrared light is one of the first types of non-visible light understood by humans and is an electromagnetic wave with outstanding analytical properties. Classified according to wavelength, it occupies the region between visible light and mid-infrared light. The American Society for Testing and Materials (ASTM) defines the spectral range of near-infrared light as 780nm to 2526nm (12820-3959cm^{-1}). Based on this, researchers often further divide it into two regions: the near-infrared shortwave region from 780nm to 1100nm (12820-9090cm^{-1}) and the near-infrared longwave region from 1100nm to 2526nm (9090-3959 cm^{-1}). The generation of near-infrared spectrum mainly relies on the non-harmonic vibration of molecules, which

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 3, May.-June, 2024, pp: 1558-1572 www.ijhssm.org

drives molecular vibrations to transition from the ground state to higher energy levels.

TABLE 3: THE APPROXIMATE POSITION OF THE ABSORPTION BAND IN THE NEAR-INFRARED SPECTRAL REGION

HYDROGEN-CONTAINING GROUPS	COMBINED FREQUENCY	DOUBLE FREQUENCY	TRIPLE FREQUENCY	QUADRUPLE FREQUENCY	FIVE TIMES FREQUENCY
C-H	2350nm	1720nm	1180nm	903nm	750nm
N-H	2150nm	1500nm	1050nm	800nm	
O-H	2000nm	1430nm	950nm	740nm	
H_2O	1940nm	1440nm	960nm	750nm	

During testing and analysis, researchers rely on the information of the overtone and combination frequencies of single chemical bond vibrations recorded by near-infrared spectrum of the measured samples. These frequencies are primarily dominated by the overlap of overtones and combination frequencies of hydrogen-containing groups (X-H) in the measured sample molecules, where X can be elements such as C, O, N, and S. Therefore, when the measurement range is confined to the near-infrared spectral region, it mainly quantifies the absorption of overtones and combination frequencies of hydrogen-containing group vibrations and outputs them. Specific absorption characteristics can be referenced in Table 3.

TABLE 4: SPECIFICATIONS OF COMMON NEAR-INFRARED SPECTROMETERS

SPECIFICATIONS	DESCRIPTION
Bruker MATRIX-F FT-NIR	The Bruker MATRIX-F Near-Infrared Spectrometer, produced by Bruker Optics GmbH in Germany, is a high-performance Fourier Transform Near-Infrared (FT-NIR) spectrometer designed for process analysis and quality control. It features high resolution and sensitivity, with a robust and durable design suitable for industrial environments, supporting real-time online analysis. Operating over a wide wavelength range (800-2500 nm), it is equipped with a halogen lamp light source to ensure stable long-term operation, with the probe capable of emitting and receiving diffuse reflectance spectra, requiring contact with the sample during measurement.
INSION NIR-NT-spectrometer-OEM-system	The NIR-NT-spectrometer-OEM-system, a compact and portable near-infrared spectrometer produced by INSION GmbH in Germany, and the NIRQuest512 Near Infrared spectrometer manufactured by Ocean Optics Inc. in the United States, both feature light sources to maintain stable emission frequencies. Their probes are capable of emitting and receiving diffuse reflectance spectra, requiring contact with the sample during measurement.
ASD FieldSpec 4Hi-Res	The FieldSpec 4Hi-Res portable spectrometer from ASD Inc. in the United States does not have an emission light source. Instead, its probe collects spectrum of natural light or light from other sources reflected from the sample. It is recommended to maintain a probe distance of 10cm~15cm from the sample collection surface and to keep the probe as vertical as possible, requiring the sample to have a certain size and to be spread out as evenly as possible.

Common brands and models of near-infrared spectrometers include Bruker MATRIX-F FT-NIR, INSION NIR-NT-spectrometer-OEM-system, and ASD FieldSpec 4Hi-Res. These

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 3, May.-June, 2024, pp: 1558-1572 www.ijhssm.org

instruments are widely used in fields such as food, pharmaceuticals, chemicals, and agriculture for component analysis, quality control, and process monitoring. They are characterized by high sensitivity, high resolution, portability, and rapid detection capabilities. Their specifications are shown in Table 4.

FIGURE 2: BRUKER MATRIX-F FT-NIR SPECTROMETER

In the experimental process of this study, minimizing all sources of interference is crucial. Contact-based measurements reduce external light interference from a hardware perspective and allow for easier control of measurement distances, aligning well with laboratory testing environments. Considering the collection of spectral data from a limited number of samples and the need for precise detection, the first two types of spectrometers are more suitable for the detection objects of this study. Considering various engineering specifications, the Bruker MATRIX-F FT-NIR spectrometer was ultimately selected for this research due to its compact size, strong adaptability to working environments, ease of operation, high reliability, and calibration-free advantage, as shown in Figure 2.

Obtaining Panax Notoginseng near-infrared spectral data. A total of 102 batches of *Panax Notoginseng* medicinal samples were collected. The samples were pulverized and sieved through a 60-mesh to obtain uniform powder. Approximately 2g of powder from each sample was weighed and placed in a glass weighing bottle, ensuring the powder surface was flat. The following parameters were set for the NIRS: resolution of 1 cm^{-1}, 64 scans per sample, with air as the reference, using a diffuse reflectance scanning mode, and a spectral range of 4000-12000 cm^{-1}. Each powder sample's spectrum was collected three times, and the average spectrum was used as the near-infrared spectrum for the *Panax Notoginseng* sample. In total, 488 sets of balance dataset and 376 sets of imbalance dataset were obtained. The AppendixTable A1 shows the near-infrared spectral samples corresponding to the main roots of *Panax Notoginseng* from different production regions.

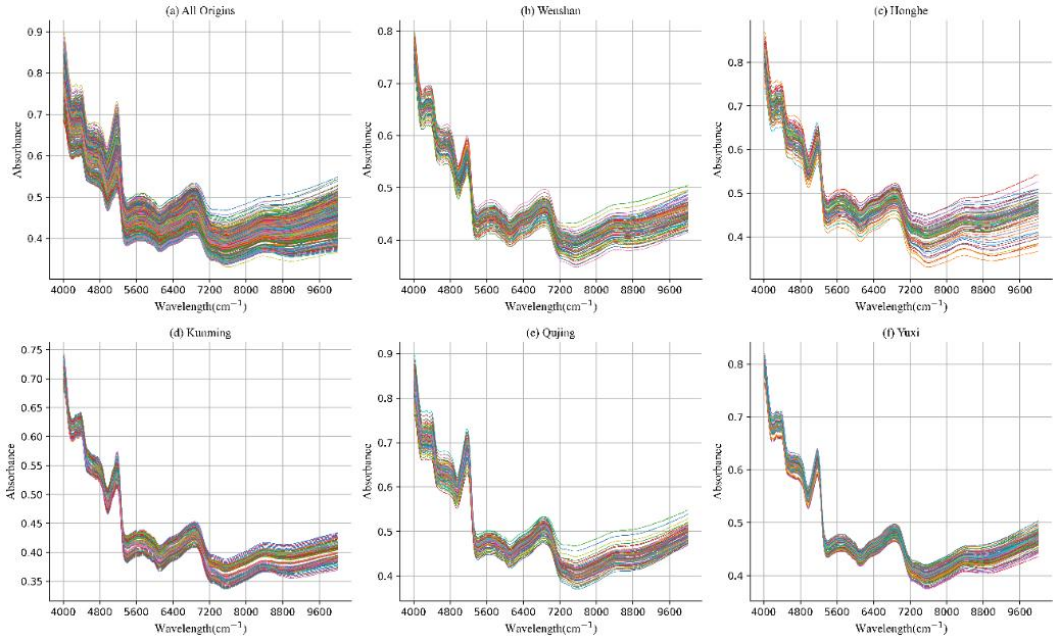

FIGURE 3: NEAR-INFRARED SPECTRUM OF THE BALANCED *PANAX NOTOGINSENG* SAMPLES

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 3, May.-June, 2024, pp: 1558-1572 www.ijhssm.org

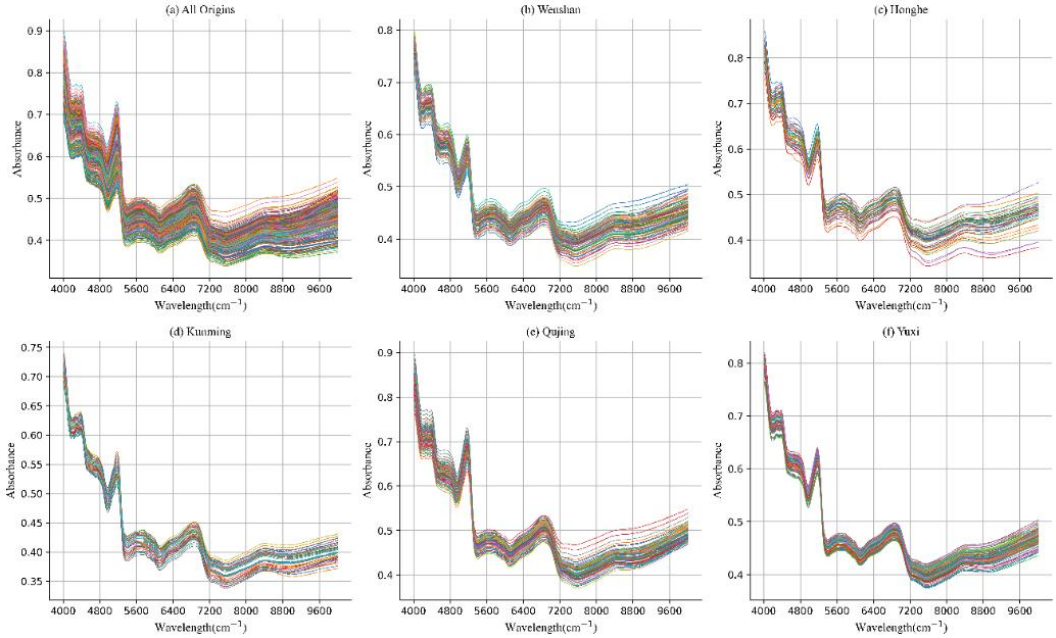

FIGURE 4: NEAR-INFRARED SPECTRUM OF THE IMBALANCED *PANAX NOTOGINSENG* SAMPLES

Figure 3 showcases near-infrared spectrum of samples from each region with a balanced distribution. Specifically, Figure 3(a) displays near-infrared spectrum of samples from all regions, while Figure 3(b) to Figure 3(f) respectively depict near-infrared spectrum of samples from Wenshan, Honghe, Kunming, Qujing, and Yuxi.

Figure 4 illustrates near-infrared spectrum of samples from each region with an imbalanced distribution. Similar to Figure 4, Figure 4(a) shows near-infrared spectrum of samples from all regions, and Figure 4(b) to Figure 4(f) display near-infrared spectrum of samples from Wenshan, Honghe, Kunming, Qujing, and Yuxi.

IV. FEATURE ANALYSIS OF NEAR-INFRARED SPECTRAL DATA

To provide a more intuitive analysis of the near-infrared spectral data of *Panax Notoginseng* from five different regions, a representative set of data from each region was selected for detailed analysis. The specific steps are as follows: first, representative samples were chosen from the 205 total samples from each region, ensuring that the selected samples accurately reflect the characteristics of *Panax Notoginseng* from that region through strict selection criteria. Then, the near-infrared spectral data of these samples were processed and compared to identify and analyze the differences and similarities between the regions.

As shown in Figure 5, all spectral curves exhibit similar shapes across the entire wavelength range, indicating a high degree of consistency in the chemical composition of *Panax Notoginseng* from the five regions. Each curve has multiple peaks and troughs, reflecting the absorbance variations at different wavelengths. In the lower wavelength range (4000-5000 cm^{-1}), all curves show significant fluctuations with distinct absorption peaks and troughs. As the wavelength increases (above 5000 cm^{-1}), the curves become relatively smooth, and the fluctuation amplitude decreases.

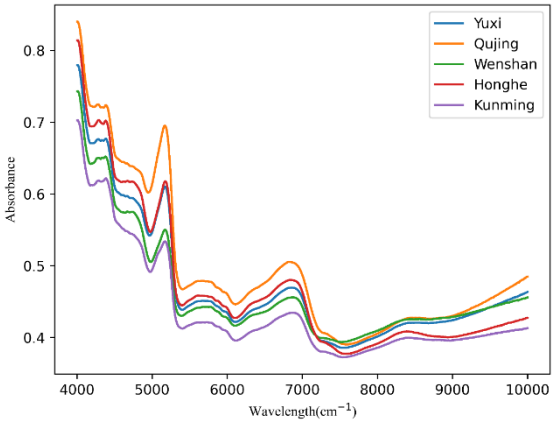

FIGURE 5: NEAR-INFRARED SPECTRUM OF THE *PANAX NOTOGINSENG* SAMPLES

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 3, May.-June, 2024, pp: 1558-1572 www.ijhssm.org

In the 4000-5000 cm^{-1} range, each spectrum exhibits multiple distinct absorption peaks. Although the peak positions slightly differ for each region, the overall trend is similar. For example, Qujing's spectrum has a prominent peak around 4200 cm^{-1}, which is higher than those of other regions. Yuxi and Honghe also show noticeable peaks around 4700 cm^{-1}, while Kunming's peak around 4300 cm^{-1} is relatively lower.

There are multiple troughs between the absorption peaks, and the variation trend of these troughs is also relatively consistent. For instance, Qujing and Wenshan show distinct troughs around 4400 cm^{-1}. Yuxi has more prominent troughs around 4600 cm^{-1} compared to other regions, while Kunming's troughs around 4500 cm^{-1} are relatively shallow.

In the lower wavelength range (4000-4500 cm^{-1}), the absorbance values of each spectrum are higher. For example, Qujing's spectrum shows significantly higher absorbance across the entire low-wavelength range, particularly around 4200 cm^{-1}, where it reaches its maximum absorbance. Yuxi and Honghe also have high absorbance around 4200 cm^{-1}, though lower than Qujing.

In the higher wavelength range (above 6000 cm^{-1}), the absorbance values of each spectrum are relatively low and tend to be consistent. For instance, Kunming's spectrum has the lowest absorbance values across the entire wavelength range, especially in the 5000-7000 cm^{-1} range. Wenshan and Honghe have relatively higher absorbance in the higher wavelength range, but the differences are not significant. Overall, Qujing's spectrum shows higher absorbance at multiple wavelengths compared to other regions, while Kunming's spectrum has relatively lower absorbance. These differences suggest that certain chemical components may vary between regions.

However, the spectral curves of *Panax Notoginseng* from different regions are very similar. The positions of the main absorption peaks and troughs are nearly identical across all regions, indicating no significant differences in the major chemical components at these wavelengths. For example, all regions' spectrum exhibits multiple distinct absorption peaks and troughs in the 4000-5000 cm^{-1} range, with consistent positions, showing a high degree of chemical composition consistency.

Additionally, the overall trend of the spectrum across the entire wavelength range is very similar for each region. From low to high wavelengths, the trend of all curves is almost identical. This high consistency in overall trends makes it very challenging to distinguish the regions based solely on spectral shapes.

Measurement noise is inevitably introduced during the measurement process, such as from instrument instability and environmental changes. This noise may cause slight fluctuations in the spectral curves, making it more difficult to capture subtle differences. In the figure, the slight fluctuations observed in the spectral curves at certain wavelengths may be manifestations of noise, which can affect the precise analysis of the spectral data.

Natural variation in the samples is another important factor. Due to differences in growth conditions, maturity, and harvest time, the chemical composition of *Panax Notoginseng* from different regions can vary. Although these variations manifest as small differences in the spectra, subtle absorbance differences at certain key wavelengths may be masked by noise and other factors, making it more challenging to distinguish between regions.

Overall, the original spectral data that has not been preprocessed may make it difficult to capture subtle but crucial differences when analyzed directly. It is necessary to use data preprocessing techniques (such as smoothing, normalization, denoising, etc.) and feature extraction techniques (such as principal component analysis, partial least squares, etc.) to enhance the differences between regions.

V. CONCLUSION

This paperexplored the application of NIRS in the analysis of *Panax Notoginseng*. It began with an overview of the medicinal properties of *Panax Notoginseng* and the importance of quality control in traditional Chinese medicine. Various conventional analytical methods were discussed, highlighting their accuracy and reliability but also their drawbacks, such as being time-consuming and requiring extensive sample preparation.

The process of sample collection and preparation was detailed, emphasizing the significance of selecting representative samples from major production regions in Yunnan Province. The samples underwent cleaning, drying, and grinding to ensure uniformity for NIR spectroscopy analysis.

Different data acquisition methods, such as thin-layer chromatography, high-performance liquid chromatography, gas chromatography, mass spectrometry, Raman spectroscopy, and nuclear magnetic resonance, were compared against NIR spectroscopy. NIR spectroscopy was noted for its advantages, including minimal sample preparation,

rapid and non-destructive analysis, and the ability to provide broad compositional data efficiently and cost-effectively.

The principle and experimental equipment of NIRS were described, focusing on the spectral range and the absorption characteristics of hydrogen-containing groups. The specifications of common near-infrared spectrometers were highlighted, with the Bruker MATRIX-F FT-NIR spectrometer selected for this study due to its high resolution, sensitivity, and adaptability to various working environments.

The paper concluded with the data acquisition process, detailing the collection of spectral data from the prepared samples. The near-infrared spectrum of samples from different regions were presented and analyzed, showing high consistency in chemical composition but also subtle regional variations. These differences underscore the importance of using advanced data preprocessing and feature extraction techniques, which will be further elaborated in the subsequent paper.

ACKNOWLEDGEMENTS

This work was supported by Science and Technology Project of Chongqing Municipal Education Commission (Grant No. KJZD-K202303301, KJQN202203309 and KJQN202203308),Planning Project of Chongqing Municipal Education Science (Grant No. K22YG313315) and School-Level General Project of Chongqing City Management College (Grant No. 2022SZZX11).

REFERENCES

[1] A. H. Rajput, R. D. Gavali, and A. P. Jadhav, “Development and validation of a novel high‐performance thin‐layer chromatography method for the quantitative estimation of zingerone”, *JPC-J Planar Chromat*, vol. 37, no. 1, pp. 87–93, Feb. 2024, doi: 10.1007/s00764-023-00268-7.

[2] M. Gangwar, J. Khan, M. S. Alam, B. P. Panda, and A. Ahamad, “Development of a high-performance thin-layer chromatography method for the rapid quantification of S-equol in biological samples of albino Wistar rats”, *JPC-J Planar Chromat*, vol. 37, no. 1, pp. 57–67, Feb. 2024, doi: 10.1007/s00764-024-00287-y.

[3] K. Makay, C. Griehl, and C. Grewe, “Development of a high-performance thin-layer chromatography-based method for targeted glycerolipidome profiling of microalgae”, *Anal Bioanal Chem*, vol. 416, no. 5, pp. 1149–1164, Feb. 2024, doi: 10.1007/s00216-023-05101-y.

[4] M. Starek, K. Homa, J. Stępińska, and M. Dąbrowska, “Development of thin-layer chromatography–densitometry for the quantification of lecithin in dietary supplements”, *JPC-J Planar Chromat*, vol. 36, no. 2, pp. 99–110, Aug. 2023, doi: 10.1007/s00764-023-00234-3.

[5] J. Żandarek, M. Starek, and M. Dąbrowska, “Development of Thin-Layer Chromatography–Densitometric Procedure for Qualitative and Quantitative Analyses and Stability Studies of Cefazolin”, *Processes*, vol. 12, no. 3, p. 591, Mar. 2024, doi: 10.3390/pr12030591.

[6] F. Ravat *et al.*, “Phytochemical analysis, isolation and quantitative estimation of karanjin in the stem bark of Millettia pinnata by a validated high-performance thin-layer chromatography method”, *JPC-J Planar Chromat*, vol. 37, no. 1, pp. 11–20, Feb. 2024, doi: 10.1007/s00764-023-00270-z.

[7] D. M. Milojković-Opsenica, J. Đ. Trifković, P. M. Ristivojević, and F. Lj. Andrić, “Thin-layer chromatography in the authenticity testing of bee-products”, *Journal of Chromatography B*, vol. 1188, p. 123068, Jan. 2022, doi: 10.1016/j.jchromb.2021.123068.

[8] J. Y. Jung, S. Y. Ha, and J.-K. Yang, “Comparison of carbohydrate composition in lignocellulosic biomass by high performance liquid chromatography and gas chromatography analysis”, *BioRes*, vol. 17, no. 1, pp. 1454–1466, Jan. 2022, doi: 10.15376/biores.17.1.1454-1466.

[9] Y. Lin and W. Yan, “Deep learning-based high performance liquid chromatography for food analysis”, *Applied Mathematics and Nonlinear Sciences*, vol. 9, no. 1, p. 20231671, Jan. 2024, doi: 10.2478/amns.2023.2.01671.

[10] M. A. Al-Ghouti, A. AlHusaini, M. H. Abu-Dieyeh, M. Abd Elkhabeer, and M. M. Alam, “Determination of aflatoxins in coffee by means of ultra-high performance liquid chromatography-fluorescence detector and fungi isolation”, *International Journal of Environmental Analytical Chemistry*, vol. 102, no. 18, pp. 6999–7014, Dec. 2022, doi: 10.1080/03067319.2020.1819993.

[11] H. N. P. Dang and J. P. Quirino, “High Performance Liquid Chromatography versus Stacking-Micellar Electrokinetic

Chromatography for the Determination of Potentially Toxic Alkenylbenzenes in Food Flavouring Ingredients", *Molecules*, vol. 27, no. 1, p. 13, Dec. 2021, doi: 10.3390/molecules27010013.

[12] Y. Zhang, Z. Zhao, W. Li, Y. Tang, H. Meng, and S. Wang, "Purification of Two Taxanes from Taxus cuspidata by Preparative High-Performance Liquid Chromatography", *Separations*, vol. 9, no. 12, p. 446, Dec. 2022, doi: 10.3390/separations9120446.

[13] A. Pašalić, S. Šegalo, D. Maestro, A. Čaušević, and A. Suljović, "Quantification of some additives in energy drinks using high-performance liquid chromatography", *JHSCI*, Sep. 2022, doi: 10.17532/jhsci.2022.1925.

[14] H. Ye *et al.*, "Simultaneous Determination of Tetrodotoxin in the Fresh and Heat-Processed Aquatic Products by High-Performance Liquid Chromatography–Tandem Mass Spectrometry", *Foods*, vol. 11, no. 7, p. 925, Mar. 2022, doi: 10.3390/foods11070925.

[15] O. R. Adianingsih, B. R. P. Ihsan, O. E. Puspita, and K. S. Maesayani, "Validation of High-Performance Liquid Chromatography (HPLC) Method for Quantification of Ethyl p-Methoxycinnamate in Kaempferia galanga Extract: http://www.doi.org/10.26538/tjnpr/v7i8.39", *Tropical Journal of Natural Product Research (TJNPR)*, vol. 7, no. 8, Art. no. 8, Aug. 2023.

[16] A. C. Fărcaș, S. A. Socaci, M. S. Chiș, F. V. Dulf, P. Podea, and M. Tofană, "Analysis of Fatty Acids, Amino Acids and Volatile Profile of Apple By-Products by Gas Chromatography-Mass Spectrometry", *Molecules*, vol. 27, no. 6, p. 1987, Mar. 2022, doi: 10.3390/molecules27061987.

[17] M. Kašpar, T. Bajer, P. Bajerová, and P. Česla, "Comparison of Phenolic Profile of Balsamic Vinegars Determined Using Liquid and Gas Chromatography Coupled with Mass Spectrometry", *Molecules*, vol. 27, no. 4, p. 1356, Feb. 2022, doi: 10.3390/molecules27041356.

[18] R. Fkiri *et al.*, "Gas Chromatography Fingerprint of Martian Amino Acids before Analysis of Return Samples", *Chemosensors*, vol. 11, no. 2, p. 76, Jan. 2023, doi: 10.3390/chemosensors11020076.

[19] R. P. Dias, T. A. Johnson, L. F. V. Ferrão, P. R. Munoz, A. P. De La Mata, and J. J. Harynuk, "Improved sample storage, preparation and extraction of blueberry aroma volatile organic compounds for gas chromatography", *Journal of Chromatography Open*, vol. 3, p. 100075, Nov. 2023, doi: 10.1016/j.jcoa.2022.100075.

[20] Z. Zakwan, Z. Lubis, and E. Julianti, "The analysis of the glyceride components on the treatment variation of refinedbleached deodorized palm oil by gas chromatography method", *Food Res.*, vol. 6, no. 1, pp. 164–167, Feb. 2022, doi: 10.26656/fr.2017.6(1).773.

[21] G. Jones *et al.*, "Comparison of Different Mass Spectrometry Workflows for the Proteomic Analysis of Tear Fluid", *IJMS*, vol. 23, no. 4, p. 2307, Feb. 2022, doi: 10.3390/ijms23042307.

[22] A. Jamalian, J. Freeke, A. Chowdhary, G. S. De Hoog, J. B. Stielow, and J. F. Meis, "Fast and Accurate Identification of Candida auris by High Resolution Mass Spectrometry", *JoF*, vol. 9, no. 2, p. 267, Feb. 2023, doi: 10.3390/jof9020267.

[23] A. De Girolamo, V. Lippolis, and M. Pascale, "Overview of Recent Liquid Chromatography Mass Spectrometry-Based Methods for Natural Toxins Detection in Food Products", *Toxins*, vol. 14, no. 5, p. 328, May 2022, doi: 10.3390/toxins14050328.

[24] Z. Ghafoori, T. Tehrani, L. Pont, and F. Benavente, "Separation and characterization of bovine milk proteins by capillary electrophoresis- mass spectrometry", *J of Separation Science*, vol. 45, no. 18, pp. 3614–3623, Sep. 2022, doi: 10.1002/jssc.202200423.

[25] S. Almaviva, F. Artuso, I. Giardina, A. Lai, and A. Pasquo, "Fast Detection of Different Water Contaminants by Raman Spectroscopy and Surface-Enhanced Raman Spectroscopy", *Sensors*, vol. 22, no. 21, p. 8338, Oct. 2022, doi: 10.3390/s22218338.

[26] A. Quesnel *et al.*, "Glycosylation spectral signatures for glioma grade discrimination using Raman spectroscopy", *BMC Cancer*, vol. 23, no. 1, p. 174, Feb. 2023, doi: 10.1186/s12885-023-10588-w.

[27] P. Pál, M. Veres, R. Holomb, M. Szalóki, A. G. Szöllősi, and I. Csarnovics, "Identification of histidine- Ni (II) metal complex by Raman spectroscopy", *J Raman Spectroscopy*, vol. 54, no. 3, pp. 278–287, Mar. 2023, doi: 10.1002/jrs.6490.

[28] D. Jonker *et al.*, "Low-Variance Surface-Enhanced Raman Spectroscopy Using Confined Gold Nanoparticles over Silicon

Nanocones", *ACS Appl. Nano Mater.*, vol. 6, no. 11, pp. 9657–9669, Jun. 2023, doi: 10.1021/acsanm.3c01249.

[29] S. Parlamas, P. K. Goetze, D. Humpal, D. Kurouski, and Y.-K. Jo, "Raman Spectroscopy Enables Confirmatory Diagnostics of Fusarium Wilt in Asymptomatic Banana", *Front. Plant Sci.*, vol. 13, p. 922254, Jun. 2022, doi: 10.3389/fpls.2022.922254.

[30] Y. Kharbanda, S. Mailhiot, O. Mankinen, M. Urbańczyk, and V.-V. Telkki, "Monitoring cheese ripening by single-sided nuclear magnetic resonance", *Journal of Dairy Science*, vol. 106, no. 3, pp. 1586–1595, Mar. 2023, doi: 10.3168/jds.2022-22458.

[31] Y. Nakashima, T. Sawatsubashi, and S. Fujii, "Nondestructive quantification of moisture in powdered low-rank coal by a unilateral nuclear magnetic resonance scanner", *International Journal of Coal Preparation and Utilization*, vol. 42, no. 5, pp. 1421–1434, May 2022, doi: 10.1080/19392699.2020.1722656.

[32] B. Bartolomei, A. Bogo, F. Amato, G. Ragazzon, and M. Prato, "Nuclear Magnetic Resonance Reveals Molecular Species in Carbon Nanodot Samples Disclosing Flaws", *Angew Chem Int Ed*, vol. 61, no. 20, p. e202200038, May 2022, doi: 10.1002/anie.202200038.

[33] V. Maestrello, P. Solovyev, L. Bontempo, L. Mannina, and F. Camin, "Nuclear magnetic resonance spectroscopy in extra virgin olive oil authentication", *Comp Rev Food Sci Food Safe*, vol. 21, no. 5, pp. 4056–4075, Sep. 2022, doi: 10.1111/1541-4337.13005.

[34] S. Tang, Y. Zhang, W. Li, X. Tang, and X. Huang, "Rapid and Simultaneous Measurement of Fat and Moisture Contents in Pork by Low-Field Nuclear Magnetic Resonance", *Foods*, vol. 12, no. 1, p. 147, Dec. 2022, doi: 10.3390/foods12010147.

[35] H. Jin *et al.*, "Rapid Detection of Avocado Oil Adulteration Using Low-Field Nuclear Magnetic Resonance", *Foods*, vol. 11, no. 8, p. 1134, Apr. 2022, doi: 10.3390/foods11081134.

[36] M. Kamruzzaman, D. Kalita, Md. T. Ahmed, G. ElMasry, and Y. Makino, "Effect of variable selection algorithms on model performance for predicting moisture content in biological materials using spectral data", *Analytica Chimica Acta*, vol. 1202, p. 339390, Apr. 2022, doi: 10.1016/j.aca.2021.339390.

[37] M. Wu, Y. Li, Y. Yuan, S. Li, X. Song, and J. Yin, "Comparison of NIR and Raman spectra combined with chemometrics for the classification and quantification of mung beans (Vigna radiata L.) of different origins", *Food Control*, vol. 145, p. 109498, Mar. 2023, doi: 10.1016/j.foodcont.2022.109498.

regions.

APPENDIX

TABLE 5: THE NEAR-INFRARED SPECTROSCOPY SAMPLES CORRESPONDING TO THE MAIN ROOTS OF *PANAX NOTOGINSENG* FROM DIFFERENT PRODUCTION REGIONS

REGIONS	NUMBER OF ROOT SAMPLES	SAMPLE DISTRIBUTION (BALANCED DATASET)	SAMPLE DISTRIBUTION (IMBALANCED DATASET)
Wenshan	W1	S139, S292, S386	S2, S104, S195, S319
	W2	S3, S141, S296, S389	S9, S113, S202, S321
	W3	S9, S143, S297, S391	S10, S124, S217, S327
	W4	S10, S144, S298, S392	S11, S134, S218, S333
	W5	S8, S155, S300, S393	S17, S136, S226, S336
	W6	S21, S161, S310, S396	S20, S137, S227, S337
	W7	S24, S166, S315, S400	S23, S138, S230, S342

	W8	S25, S173, S317, S407	S25, S140, S231, S344
	W9	S27, S181, S318, S412	S27, S146, S233, S349
	W10	S32, S185, S320, S425	S36, S150, S234, S359
	W11	S42, S198, S322, S429	S40, S152, S239, S361
	W12	S45, S237, S332, S432	S42, S166, S240, S364
	W13	S49, S239, S339, S444	S45, S170, S243, S365
	W14	S60, S242, S345, S447	S49, S171, S245, S368
	W15	S61, S246, S356, S449	S53, S174, S249, S375
	W16	S69, S250, S357, S451	S59, S175, S252
	W17	S70, S255, S359, S485	S72, S177, S254
	W18	S80, S259, S361	S75, S180, S264
	W19	S87, S267, S363	S78, S181, S272
	W20	S92, S275, S365	S79, S183, S274
	W21	S102, S280, S367	S80, S184, S275
	W22	S114, S282, S371	S88, S185, S280
	W23	S116, S283, S372	S98, S186, S292
	W24	S119, S285, S381	S101, S188, S10
	W25	S120, S288, S385	S102, S191, S317
Honghe	H1	S4, S128, S236, S358	S5, S164
	H2	S15, S131, S238, S376	S12, S178
	H3	S28, S146, S258, S378	S16, S196
	H4	S35, S153, S261, S382	S22, S213
	H5	S47, S154, S262, S383	S26, S219
	H6	S50, S160, S264, S402	S37, S221
	H7	S51, S164, S265, S403	S66, S223
	H8	S56, S168, S273, S409	S70, S225
	H9	S72, S170, S286, S421	S76, S248
	H10	S75, S178, S305, S422	S85, S253

	H11	S83, S186, S308, S423	S87, S258
	H12	S84, S193, S309, S424	S91, S262
	H13	S89, S194, S311, S436	S97, S270
	H14	S90, S197, S316, S438	S116, S284
	H15	S91, S199, S324, S454	S122, S286
	H16	S93, S201, S325, S461	S127, S291
	H17	S97, S204, S326, S468	S129, S297
	H18	S105, S206, S328, S470	S135, S302
	H19	S109, S207, S335, S479	S141, S339
	H20	S110, S223, S344, S482	S143, S345
	H21	S113, S224, S348	S148, S377
	H22	S121, S228, S349	S155, S378
	H23	S125, S235, S350	S160
Kunming	K1	S6, S112, S219, S284, S380, S450	S6, S161, S244
	K2	S16, S118, S222, S287, S384, S452	S18, S187, S366
	K3	S17, S124, S227, S290, S395, S455	S19, S194, S370
	K4	S20, S126, S229, S291, S398, S465	S33, S198, S371
	K5	S29, S127, S230, S302, S404, S474	S38, S201, S379
	K6	S36, S129, S231, S314, S405, S76	S63, S206
	K7	S44, S134, S245, S319, S406, S484	S77, S207
	K8	S57, S140, S247, S321, S408	S83, S210
	K9	S59, S149, S251, S333, S411	S86, S228
	K10	S71, S158, S254, S338, S415	S103, S242
	K11	S73, S169, S263, S340, S416	S118, S256
	K12	S81, S174, S266, S347, S418	S120, S269
	K13	S82, S188, S268, S353, S419	S121, S273
	K14	S85, S191, S269, S355, S420	S123, S288
	K15	S88, S196, S270, S360, S430	S133, S340

	K16	S98, S213, S276, S364, S434	S145, S343
	K17	S106, S215, S278, S375, S435	S147, S352
	K18	S108, S218, S279, S379, S443	S151, S356
Qujing	Q1	S2, S99, S176, S243, S334, S440, S487	S1, S54, S114, S212, S294, S328
	Q2	S5, S100, S182, S248, S337, S445	S8, S57, S115, S214, S296, S329
	Q3	S8, S103, S184, S249, S341, S446	S14, S58, S117, S216, S298, S330
	Q4	S12, S111, S187, S256, S342, S448	S15, S60, S131, S224, S299, S331
	Q5	S37, S117, S189, S271, S343, S453	S21, S62, S139, S232, S305, S335
	Q6	S41, S123, S190, S281, S346, S456	S28, S64, S149, S236, S306, S338
	Q7	S62, S145, S195, S289, S366, S458	S30, S69, S156, S241, S308, S347
	Q8	S63, S147, S205, S293, S368, S463	S31, S71, S159, S246, S309, S350
	Q9	S64, S148, S209, S299, S369, S464	S35, S73, S168, S250, S311, S354
	Q10	S66, S150, S210, S306, S374, S469	S39, S84, S169, S259, S312, S357
	Q11	S68, S151, S211, S307, S377, S471	S43, S94, S197, S271, S314, S358
	Q12	S76, S156, S212, S312, S397, S472	S44, S96, S199, S278, S322, S362
	Q13	S77, S157, S216, S323, S399, S475	S46, S107, S200, S281, S323, S367
	Q14	S79, S159, S220, S329, S413, S480	S47, S110, S203, S282, S324, S372
	Q15	S94, S165, S232, S330, S427, S481	S51, S111, S204, S283, S325, S376
	Q16	S96, S172, S240, S331, S433, S484	S52, S112, S208, S290, S326
Yuxi	Y1	S1, S54, S138, S2, S327, S351, S442	S3, S89, S154, S220, S285, S351
	Y2	S11, S55, S152, S2, S336, S352, S457	S4, S90, S157, S222, S287, S353
	Y3	S13, S58, S162, S226, S354, S459	S7, S92, S158, S229, S289, S355
	Y4	S14, S65, S163, S233, S362, S460	S13, S93, S162, S235, S293, S360
	Y5	S19, S67, S167, S234, S370, S462	S24, S95, S163, S237, S295, S363
	Y6	S22, S74, S171, S241, S373, S466	S29, S99, S165, S238, S300, S369
	Y7	S23, S78, S175, S244, S387, S467	S32, S100, S167, S247, S301, S373
	Y8	S26, S86, S177, S252, S388, S473	S34, S105, S172, S251, S303, S374
	Y9	S30, S95, S179, S253, S390, S477	S41, S106, S173, S255, S304

	Y10	S31, S101, S180, S257, S394, S478	S48, S108, S176, S257, S307
	Y11	S33, S104, S183, S260, S401, S486	S50, S109, S179, S260, S313
	Y12	S34, S107, S192, S272, S410	S55, S119, S182, S261, S315
	Y13	S38, S115, S200, S274, S414	S56, S125, S189, S263, S316
	Y14	S39, S122, S202, S277, S417	S61, S126, S190, S265, S318
	Y15	S40, S130, S203, S294, S426	S65, S128, S192, S266, S320
	Y16	S43, S132, S208, S295, S428	S67, S130, S193, S267, S332
	Y17	S46, S133, S214, S301, S431	S68, S132, S205, S268, S334
	Y18	S48, S135, S217, S303, S437	S74, S142, S209, S276, S341
	Y19	S52, S136, S221, S304, S439	S81, S144, S211, S277, S346
	Y20	S53, S137, S225, S313, S441	S82, S153, S215, S279, S348