INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 3, MARCH 2026 | ISSN: 2456-4184 | IJNRD.ORG

From Data to Knowledge: The Role of Chemometrics in Chemical Analysis and Herbal Standardization

[1]Sanket Chavan, [2]Kamlesh Bhati, [3]Atharv Mengade, [4]Rohan Suryawanshi, [5]Sneha Modi

[1]Student, [2]Student, [3]Student, [4]Student, [5]Student
[1]Alard College of Pharmacy, Pune, India

Abstract : Chemometrics is a multidisciplinary field that applies mathematical, statistical, and computational methods to extract meaningful information from complex chemical and biological data. With the increasing use of high-throughput analytical techniques, chemometrics has become indispensable in areas such as metabolomics, herbal drug standardization, pharmaceutical analysis, and food quality assessment. This review presents a comprehensive discussion of chemometrics based exclusively on published review literature, covering its theoretical foundations, multivariate data analysis techniques, role in metabolomics, and applications in the standardization and quality control of herbal medicines. Current challenges and future prospects of chemometrics-driven analytical science are also discussed.

INTRODUCTION

Modern analytical technologies such as nuclear magnetic resonance (NMR), mass spectrometry (MS), infrared spectroscopy, and chromatography generate large volumes of complex, multivariate data. Interpretation of such datasets using classical univariate methods is inadequate due to overlapping signals, matrix effects, and variable correlations. Chemometrics addresses these limitations by providing tools for data preprocessing, dimensionality reduction, pattern recognition, and predictive modelling. [1].

Chemometrics was introduced in early 1970s by Swedish scientist Svante Wold. He described it as the application of statistical and mathematical techniques to chemical data in order to obtain meaningful information. The word chemometrics originates from the Swedish term “Kemometri,” where “Kemo” refers to chemistry and “metri” means measurement. This concept highlights the importance of measuring, analyzing and interpreting chemical data in a systematic and scientific way [1].

Chemometrics has evolved into a core component of analytical chemistry and is now widely applied in metabolomics studies, pharmaceutical quality control, herbal drug standardization, and food analysis, where reliable interpretation of complex chemical fingerprints is essential [2].

Definition and Scope of Chemometrics

Chemometrics can be defined as the application of mathematical and statistical techniques to chemical and biological measurements in order to extract relevant information and establish relationships between measured variables and system properties [3].

According to the International Chemometrics Society founded in 1974 and its founders (Svante Wold and Bruce Kowalski), Chemometrics is defined as “ the chemical discipline that uses mathematical and statistical methods to (i) design or select optimal measurement procedures and experiments and (ii) provide maximum chemical information by analyzing chemical data.”

Chemometrics is essential to analytical chemistry because it helps in extracting meaningful information from complex chemical data. It uses a combination of statistical and mathematical techniques to analyze large and complex datasets generated during analytical procedures. By applying chemometric tools, analytical methods can be optimized, errors can be reduced, and the reliability and accuracy of results can be improved. Chemometrics also makes it easier to interpret hidden patterns and relationships within data, which may not be visible through conventional analysis. Overall, it plays a key role in managing massive datasets and converting raw chemical data into useful and reliable information [1]. The scope of chemometrics includes experimental design, signal preprocessing, exploratory data analysis, calibration, classification, and model validation. The central objectives of chemometrics is to maximize chemical information while minimizing experimental effort and analytical error. [4].

Evolution and Importance of Chemometrics

The development of chemometrics was driven by the rapid advancement of analytical instrumentation and computational capabilities. Early chemometric studies focused on regression analysis and experimental optimization. Over time, multivariate techniques such as principal component analysis and partial least squares regression became dominant tools for handling complex datasets [5].

Today, chemometrics is essential for managing high-dimensional datasets in metabolomics, herbal analysis, and systems chemistry, enabling holistic interpretation rather than single-compound evaluation [6].

The evaluation of chemometric model is a central aspect of chemometrics. Model evaluation ensures that the mathematical representation of chemical data is valid, reliable, and meaningful. A chemometric model must not only fit the available data but also possess predictive ability and chemical clarity. [14]

The importance of chemometrics, lies in its ability to manage and analyze complex chemical data in a systematic and objective manner. As chemical measurements increasingly involve multiple correlated variables, chemometrics provides the necessary framework to extract useful information that would otherwise remain unnoticed.[14]

Chemometrics significantly improves the quality and dependability of chemical analysis by reducing noise, experimental variability, and collinearity among variables. By using multivariate approaches, chemometrics allows more accurate characterization of chemical systems compared to classical single-variable methods.[14]

Data preprocessing and signal treatment

Analytical data often contain unwanted variation arising from instrumental noise, baseline drift, and sample heterogeneity. Data preprocessing is therefore a critical step in chemometric analysis [7].

Common preprocessing methods include:

- Baseline correction
- Smoothing and noise reduction
- Normalization and scaling
- Mean centering

Data preprocessing steps:

- **Missing value handling:**
  Missing data is managed by removing variables with too many missing values or by replacing missing entries with appropriate estimates.
- **Outlier detection:**
  Outliers are identified using methods such as PCA, clustering techniques, or distance measures to find unusual data points.
- **Scaling:**
  Data is scaled using the mean and standard deviation so that all variables are on a comparable scale.
- **Transformation:**
  Transformations are applied to reduce skewness and stabilize variance, helping models perform better.[12].

These techniques enhance data quality and improve the robustness of multivariate models [8].

Exploratory Multivariate data analysis

Exploratory data analysis allows visualization and understanding of underlying data structure without predefined hypotheses.

- **Principal Component Analysis (PCA)**

PCA is the most widely applied unsupervised chemometric technique. It reduces dimensionality by transforming correlated variables into orthogonal principal components that explain the maximum variance in the dataset. PCA is extensively used for sample clustering, trend analysis, and outlier detection in chemical and biological studies [6].

Principal Component Analysis (PCA) reduces data complexity by creating new components that capture the maximum variance in the dataset. It simplifies analysis and is commonly visualized using score plots and loading plots to show data patterns and variable contributions.[12].

Principal Component Analysis helps visualize complex data by representing variables (X) in different positions within a reduced space. It is useful for detecting clusters, identifying outliers and understanding the impact of anomalies in the dataset. PCA also compresses large datasets while retaining important information and helps reduce noise, making the data easier [7].

Multivariate Calibration and Regression Techniques

Multivariate calibration involves developing a model that relates a set of measured response variables (such as spectral intensities at different wavelengths) to known values of analyte concentrations or physical properties. Unlike **univariate calibration**, which uses a single measured signal at one wavelength or variable, multivariate calibration uses the **full dataset** to extract more meaningful information and improve prediction accuracy. This is particularly important when signals overlap or when variables are highly correlated.

- **Multiple Linear Regression (MLR)**

Multiple Linear Regression (MLR) serves as a foundational chemometric tool designed to model the relationship between a single dependent response variable and several independent predictors . In analytical chemistry, this typically involves correlating the concentration of an analyte with absorbance values at specific wavelengths.[15]

MLR relates a dependent variable to multiple predictors but becomes unreliable when predictors are highly correlated, as is comm on in spectral data [3].

Despite its mathematical simplicity, MLR is often unsuitable for raw spectral data due to several integral limitations:

- **Multicollinearity:** Spectral data is characterized by high inter-correlation between neighbouring wavelengths. This "redundancy" violates the MLR assumption of predictor independence, leading to unstable coefficient estimates and high sensitivity to small changes in input data.
- **Overfitting:** Because MLR attempts to capture all variance in the provided predictors, it often interprets random instrumental noise as meaningful signal, resulting in models that perform well on training data but fail on external validation.[15]

- **Partial Least Squares (PLS) Regression**

PLS (Partial Least Squares) regression is a method used when many predictor variables are strongly correlated with each other. It works by transforming both the predictor variables and the response variable into new underlying factors (called latent variables) that are designed to maximize the relationship between them. Because of this, PLS can handle multicollinearity effectively. It is widely used in spectroscopic, chromatographic, and metabolomic studies for making quantitative predictions.[12]

PLS is commonly applied in techniques such as NIR and UV calibration, where the data can be complex and overlapping. It builds reliable prediction models even when variables are highly correlated. Compared to PCR (Principal Component Regression), PLS often provides better accuracy because it directly focuses on linking the measured signals to the target values..[12]

Chemometrics in Metabolomics

Metabolomics aims to comprehensively analyze small-molecule metabolites in biological systems. Analytical platforms such as NMR and MS generate complex datasets containing thousands of variables per sample. Chemometrics is essential for extracting biologically meaningful information from such data [2].

Multivariate techniques including PCA, PLS, and discriminant analysis are used in metabolomics to:

- Identify metabolic patterns
- Differentiate biological states
- Discover biomarkers
- Interpret system-level biochemical changes [2][10]

Chemometrics enables holistic interpretation of metabolomic fingerprints rather than focusing on individual metabolites, making it a cornerstone of systems biology [10].

Chemometrics in Herbal Drug Standardization

Herbal medicines consist of complex mixtures of bioactive compounds, and their quality cannot be adequately assessed using single-marker approaches. Chemometrics provides powerful tools for herbal drug standardization by analyzing complete chemical fingerprints [5].

Herbal standardization using fingerprinting, profiling and metabolomics ensures consistent quality and safety. These techniques identify unique chemical patterns and metabolite compositions, helping detect adulteration and batch variations in herbal products. [4].

Spectroscopic and chromatographic data combined with multivariate analysis allow:

- Authentication of herbal raw materials
- Detection of adulteration
- Batch-to-batch consistency evaluation
- Quality control of finished herbal products [5][6]

Chemometric fingerprinting approaches are increasingly recognized as reliable strategies for ensuring the safety, efficacy, and reproducibility of herbal medicines [8].

Classification and Pattern Recognition

Classification methods assign samples to predefined groups based on multivariate characteristics.

- **Linear Discriminant Analysis (LDA)**

LDA constructs linear combinations of variables that maximize separation between classes and is widely applied in metabolomics and herbal product classification [10].

LDA constructs linear combinations of measured variables that maximize separation between predefined classes, making it an effective supervised classification method in metabolomics and herbal product studies. Beyond basic classification, LDA can be combined with other chemometric techniques (e.g., PCA for dimension reduction followed by LDA for classification) to improve performance with highly multivariate spectral or chromatographic data.[16]

- **Soft Independent Modeling of Class Analogy (SIMCA)**

SIMCA builds class-specific PCA models and classifies samples based on their similarity to reference groups. It is particularly useful for herbal authentication and quality control [9].

SIMCA independently constructs a PCA model for each class using only its training samples, enabling classification based on similarity to these class models rather than a shared boundary. [16]

SIMCA can classify unknown samples into *one, multiple, or none* of the predefined categories based on their similarity to model spaces a useful feature for complex botanical authentication where samples may share traits with more than one group.[16]

Applications

Chemometrics is extensively used in pharmaceutical analysis for formulation assessment, impurity profiling, and quality control. Rapid, non-destructive spectroscopic methods combined with chemometric modeling improve analytical efficiency and regulatory compliance [7].

Applications of Chemometrics:

- Food science and technology
- Quality assessment and analysis of medicines
- Environmental analysis
- Green analytical chemistry
- Drug discovery
- Forensic science [1]

Chemometrics applications support drug design, development, and analysis. QSAR, SAR, and metabolomics help understand drug behavior and disease states, while powder flow, physical properties, and water content ensure product quality. Dissolution studies, tablet assays, polymorph analysis, and chromatographic optimization confirm drug performance, stability, and accurate measurement.[12].

In food science, chemometrics supports authenticity testing, adulteration detection, and quality evaluation by analyzing complex chemical profiles [8].

Challenges and Limitations

Despite its advantages, chemometrics faces challenges such as:

- Dependence on data quality
- Risk of model overfitting
- Complexity in model interpretation
- Requirement for proper validation strategies [10]
- Preserving data integrity
- Model transfer issues
- Complex information networks
- Improper use of tools
- Potential for misunderstanding & sample variables[1]

Addressing these challenges is essential for reliable application of chemometric methods in metabolomics and herbal analysis.

Future Perspectives

Future developments in chemometrics are expected to focus on integration with advanced computational methods, automated data analysis pipelines, and real-time monitoring systems. In metabolomics and herbal medicine research, chemometrics will continue to play a central role in holistic data interpretation and quality assurance [11].

The integration of chemometrics with artificial intelligence (AI), machine learning (ML) and deep learning (DL) has significantly improved data analysis in analytical chemistry. These advanced techniques enhance statistical analysis, increase accuracy and improve overall effectiveness of analytical methods. AI tools can interpret hidden patterns and accurately predict molecular behaviour. Modern approaches such as artificial neural networks (ANNs), support vector machines (SVMs) and chemometric methods like partial least squares- discriminant analysis (PLS-DA) provide powerful and reliable solutions for handling complex chemical datasets. [1]

Conclusion

Chemometric has become an important part of modern analytical science. It helps scientists study and understands complex data that contain many variables at the same time. This is especially useful in fields like metabolomics and herbal drug standardization, where it is important to examine the complete chemical profile rather than just one or two components. As new methods and technologies continue to develop, chemometrics will play an even bigger role in improving the accuracy of analysis, maintaining product quality, and increasing scientific knowledge.

Acknowledgment

We would like to express our sincere gratitude to Mrs. Ashwini Kshirsagar for valuable guidance and continuous support throughout the entire process of preparing and publishing this review article on chemometrics. The insightful suggestions, encouragement and constructive feedback provided by Mrs. Kshirsagar greatly contributed to the successful completion of this work. We are truly thankful for the mentorship and support received during this research and publication process.

REFERENCES

1. Saha P, Pandit B, Pramanik S, Shrestha B. A comprehensive review on the applications of chemometrics in analytical chemistry. **J Appl Pharm Res.** 2025;13(3):1-16. doi:10.69857/joapr.v13i3.861.
2. Trygg J, Holmes E, Lundstedt T. Chemometrics in metabonomics. **J Proteome Res.** 2007;6(2):469-479. doi:10.1021/pr060594q.
3. Esbensen KH. The philosophy and fundamentals of handling, modelling and interpreting large data sets—the multivariate chemometrics approach. In: Bakeev KA, editor. **Multivariate analysis in the pharmaceutical industry.** 2nd ed. Amsterdam: Elsevier; 2018. p.13-34.
4. Rebiai A, Hemmami H, Zeghoud S, Ben Seghir B, Kouadri I, Eddine LS, et al. Current application of chemometrics analysis in authentication of natural products: a review. **Comb Chem High Throughput Screen.** 2022;25(6):945-972. doi:10.2174/1386207324666210309102239.
5. Kusumadewi R, Kusumadewi A, Suryanti V. Chemometrics and multivariate analysis: applications in food and product analysis. **My Food Res.** 2021.
6. Singh I, Juneja P, Kaur B, Kumar P. Pharmaceutical applications of chemometric techniques. **ISRN Anal Chem.** 2013;2013:795178. doi:10.1155/2013/795178.
7. Kumari MV, Asif D, Aasik S, Karishma S, Mehar S. A review on chemometric techniques. **Int J Pharm Sci Res.** 2024;9(4):1-8.
8. Massart DL, Kaufman L, editors. **Chemometrics: A textbook.** Amsterdam: Elsevier; 1983.
9. Otto M. **Chemometrics: statistics and computer application in analytical chemistry.** 2nd ed. Weinheim: Wiley-VCH; 2007.
10. Kowalski BR, Bender CF. Pattern recognition: a powerful approach to interpreting chemical data. **J Am Chem Soc.** 1972;94(16):5632-5639.
11. Li Y, Shen Y, Yao CL, Guo DA. Quality assessment of herbal medicines based on chemical fingerprints combined with chemometrics approach: a review. **J Pharm Biomed Anal.** 2020;185:113215. doi:10.1016/j.jpba.2020.113215.
12. González-Domínguez R, Sayago A, Fernández-Recamales Á. An overview on the application of chemometrics tools in food authenticity and traceability. **Foods.** 2022;11(23):3940. doi:10.3390/foods11233940.
13. Wold S, Sjöström M, Eriksson L. Chemometrics in chemistry. **Chemom Intell Lab Syst.** 2001;58:109-130.
14. Saeys W, Do Trong NN, Dahwadar R, Nicolaï BM. Multivariate calibration in spectroscopy: from linear to nonlinear methods. **TrAC Trends Anal Chem.** 2019;118:510-522.
15. Abraham EJ, Kellogg JJ. Chemometric-guided approaches for profiling and authenticating botanical materials. **Front Nutr.** 2021;8:780228. doi:10.3389/fnut.2021.780228