International Journal of Horticulture, Agriculture and Food Science (IJHAF)
ISSN: 2456-8635
[Vol-9, Issue-4, Oct-Dec, 2025]
Issue DOI: https://dx.doi.org/10.22161/ijhaf.9.4
Peer-Reviewed Journal

Construction and Analysis of a Binary State–Crop Availability Matrix for Indian Agricultural Data

Bharat Khushalani*, Sama Shaik, Deepika Ponakala

Shri Vishnu Engineering College for Women, Bhimavaram, India
*Corresponding Author

Received: 31 Aug 2025; Received in revised form: 28 Sep 2025; Accepted: 04 Oct 2025; Available online: 11 Oct 2025

Abstract— *Accurate and comprehensive agricultural data is essential for research, planning, and policymaking, forming the foundation for evidence-based decisions at both state and national levels. This study presents the construction of a binary state–crop availability matrix (A) for India, representing data coverage for 28 states and Union Territories over a five-year period from 2020 to 2024. In this matrix, each row corresponds to a year, each column corresponds to a specific crop, and each element is encoded as 1 if reliable data for that crop in the corresponding year and state exists, and 0 if data is missing or incomplete. By capturing the presence or absence of crop production data in a structured binary format, the A matrix provides a systematic overview of data availability across the country. The matrix reveals significant heterogeneity in reporting patterns across states, reflecting differences in the scale of agricultural activity, crop diversity, and administrative capacity. Larger states with diversified cropping systems, such as Karnataka, Tamil Nadu, and Madhya Pradesh, tend to exhibit a higher proportion of 1's, indicating comprehensive data coverage and robust reporting infrastructure. Conversely, smaller Union Territories such as Lakshadweep, Chandigarh, and Daman and Diu display larger proportions of 0's, highlighting gaps due to limited cultivation, fewer resources, or lower prioritization of statistical reporting. These systematic differences underscore the structural nature of data disparities and emphasize the need for targeted interventions to improve data collection in under-represented regions. Beyond identifying missing records, the A matrix provides a versatile foundation for a wide range of data-driven agricultural analyses. It enables quantitative assessment of regional reporting completeness, informs prioritization of capacity-building initiatives, and supports resource allocation to states or crops where data gaps are most pronounced. Furthermore, the matrix serves as a replicable framework for other countries or sectors seeking to evaluate the quality and coverage of their datasets. By combining this binary representation with analytical methods such as matrix algebra, similarity analyses, and multivariate techniques, researchers and policymakers can derive insights into inter-state crop overlaps, co-occurrence patterns, and regional specialization, ultimately contributing to more efficient planning, equitable resource distribution, and strategic interventions in agricultural development.*

Keywords— ***Agricultural Data, Data Availability Matrix, State-wise Analysis, Data Gaps, Evidence-based Policy***

I. INTRODUCTION

India's agricultural sector is highly diverse, both in terms of the variety of crops cultivated and the regional patterns of production across states and Union Territories. Reliable, crop-level data spanning multiple years is essential for monitoring crop performance, analyzing temporal and spatial trends, and informing evidence-based agricultural policies and planning. Accurate data enables assessment of regional productivity, identification of emerging trends in crop diversification, and development of interventions to improve food security and market efficiency. However, significant disparities exist in the availability and completeness of crop data across states due to multiple factors, including state size, the extent of agricultural

diversification, administrative efficiency, and the capacity of local agricultural departments.

To systematically capture this variability, we constructed a binary state–crop availability matrix, referred to as the A matrix. The matrix employs a simple but powerful coding system: each element is assigned a value of 1 if crop data for a given year and state exists, and 0 if the data is missing, incomplete, or unreliable. Rows correspond to the years 2020 through 2024, while columns represent individual crops officially reported by state agricultural departments. States are indexed sequentially from A1 to A28, beginning with Andaman and Nicobar Islands and concluding with West Bengal, allowing for standardized referencing and comparison across the dataset.

The resulting matrix provides a structured framework for analyzing differences in data completeness and reporting quality between states. It highlights patterns such as higher completeness in larger, agriculturally diverse states and more frequent data gaps in smaller Union Territories with limited cultivation. Beyond identifying missing data, the A matrix facilitates comparative analyses across regions, enabling researchers and policymakers to pinpoint under-represented crops and regions, evaluate the effectiveness of reporting systems, and design strategies to improve data collection. Ultimately, this approach offers a replicable method for assessing agricultural data coverage over time and can serve as a model for similar studies in other countries or sectors [1,2,4].

Monitoring crop diversity and availability has been a focus of agricultural research, as incomplete datasets can hinder research and policy decisions. Studies have examined crop diversification and its economic impact on Indian farming practices [1], as well as the use of varietal threat indices to monitor crop diversity at the farm level [3]. While prior research has primarily concentrated on yield predictions, crop suitability, and AI-based assessment techniques [7,9], systematic evaluation of data coverage across states and years remains limited. The construction of a binary availability matrix, such as the A matrix, provides a structured approach to quantify gaps in data reporting. This method complements existing studies by providing a state-wise, year-wise perspective on the completeness of crop data, which is essential for downstream analyses such as machine learning-based yield prediction and geospatial crop modeling [8,9].

II. METHODOLOGY

Matrix of Andhra Pradesh

1011010101111011101111101000010110101000111111111101000101111011001010001101011111110111011011110111101

1011010001111011101111001000010110101000111111111101000101111011001010001101011110110110011011111110101

1000010001111011100111001100011110101000010111111100000100111010000110001001011110110110011011101100101

1001010101111011100111101000011110101000010111111101000100011010001110001001011111110011011011111101101

1001010111110111100111101000011110101001010111110101000111011010011110001101011111100011111011111001101

Above matrix illustrates a state–crop availability matrix representing agricultural data for multiple crops over the five-year period from 2020 to 2024. In this figure, each row corresponds to a year, and each column corresponds to a specific crop. The entries are coded as 1 or 0, where 1 indicates that data for the crop in the corresponding year is available, and 0 indicates that the data is absent. This structure allows for a clear visualization of which crops are consistently reported across the years and which have gaps in reporting. Rows with a higher number of 1's reflect years with more comprehensive data coverage, while rows with more 0's indicate years with significant missing records. Similarly, columns with predominantly 1's represent crops that are well-monitored and consistently reported, whereas columns with scattered 0's highlight crops with irregular or incomplete reporting. By capturing these patterns, Fig. 1 provides valuable insights into the reliability and completeness of agricultural data, enabling researchers and policymakers to identify gaps, assess reporting efficiency, and make informed decisions regarding crop monitoring, resource allocation, and planning. Furthermore, this representation can be extended to advanced analyses, such as state–state or crop–crop similarity, to explore crop diversity, regional specialization, and opportunities for cooperative crop-sharing and redistribution strategies.

The A matrix was meticulously constructed using official datasets sourced from state agricultural departments and national agricultural repositories, covering the years 2020–2024. Each row in the matrix corresponds to a distinct state or Union Territory, while each column represents a specific crop cultivated across India. For every state, the presence of reliable and verifiable crop production data was encoded as 1, whereas missing, incomplete, or inconsistent records were represented as 0. This binary coding allows for a clear and systematic identification of data availability across the country. The resulting matrix spans 28 rows—one for each state or Union Territory—and includes columns representing all major crops officially reported nationwide. To maintain consistency and facilitate analysis, states were

indexed sequentially from A1 (Andaman and Nicobar Islands) to A28 (West Bengal).

The variation in the distribution of 1's and 0's across different states reflects multiple underlying factors. Primarily, it illustrates the scale and diversity of agricultural activity within each region, as well as the effectiveness and efficiency of local data collection and reporting mechanisms. Larger states such as Karnataka, Tamil Nadu, and Telangana tend to display a higher number of 1's, indicating not only a wider variety of crops cultivated but also more robust and comprehensive agricultural reporting systems. In contrast, smaller Union Territories like Lakshadweep, Chandigarh, and Daman and Diu exhibit fewer 1's, a pattern attributable to limited land availability for cultivation, constrained resources, and relatively lower priority in agricultural data collection.

By systematically mapping the presence and absence of crop data, the A matrix provides a comprehensive framework for evaluating nationwide agricultural data coverage. It allows researchers and policymakers to identify critical gaps in reporting, prioritize areas for capacity-building in data collection, and support targeted interventions aimed at improving the completeness and reliability of agricultural statistics. Moreover, the matrix serves as a valuable analytical tool for longitudinal studies, enabling the tracking of trends in data availability and the assessment of state-level efforts to strengthen agricultural information systems over time. In essence, the AAA matrix not only captures the current state of crop data coverage but also lays the foundation for more informed decision-making and strategic planning in Indian agriculture.

III. RESULTS ANALYSIS

Analysis of the A matrix reveals distinct and informative patterns in data availability across Indian states and Union Territories. States characterized by extensive cultivated areas and a diverse range of crop portfolios, such as Karnataka, Madhya Pradesh, and Tamil Nadu, exhibit a high prevalence of 1's in the matrix. This pattern reflects comprehensive and systematic agricultural reporting, highlighting the presence of robust data collection mechanisms and effective administrative oversight. In contrast, Union Territories and smaller states, including Lakshadweep, Daman and Diu, and Chandigarh, display a predominance of 0's, which points to limited agricultural activity, constrained resources for data collection, or incomplete statistical reporting. Such disparities underscore the unevenness in the quality and completeness of crop production datasets across the country.

The variability in 1's and 0's further provides insight into the role of administrative practices, infrastructure, and crop prioritization in shaping data availability. States that cultivate fewer crops or focus predominantly on staple crops often report only selected datasets, leaving entries for minor or less commercially significant crops unrecorded. This selective reporting introduces gaps in the national agricultural dataset and highlights the importance of considering both cultivation scale and administrative emphasis when interpreting matrix patterns. Additionally, the binary representation of data facilitates both visual inspection and quantitative analysis, enabling researchers to identify under-represented regions, crops with limited reporting, and potential inconsistencies in historical datasets. By providing a clear, structured overview of data coverage, the A matrix serves as a critical tool for identifying deficiencies, guiding resource allocation for data collection, and informing the development of targeted, data-driven agricultural policies. Moreover, this structured binary framework supports comparative analyses over time, allowing policymakers and researchers to track improvements or declines in reporting completeness, assess the impact of interventions in specific states, and prioritize regions requiring enhanced agricultural monitoring systems.

IV. DISCUSSION

The A matrix underscores the significant heterogeneity in agricultural data availability across Indian states, revealing patterns that carry important implications for both research and policy-making. Regions exhibiting higher completeness of crop production records enable more precise modeling of agricultural systems, accurate yield prediction, and effective planning of crop-specific interventions. For instance, states like Karnataka and Madhya Pradesh, with extensive and diversified reporting, can support detailed analyses of crop rotations, input efficiency, and regional food security. In contrast, states and Union Territories with lower completeness present substantial challenges for accurate agricultural assessment, often necessitating interpolation, assumptions, or exclusion of certain crops from analyses, which can reduce the reliability of policy recommendations and economic planning.

The observed discrepancies in data availability are influenced by multiple interrelated factors, including state size, the diversity of crops cultivated, administrative efficiency, and the capacity of local agricultural departments to collect, verify, and report data systematically. Smaller regions or those with limited cultivation areas, such as Lakshadweep or Chandigarh, frequently display gaps in reporting, particularly for minor crops, highlighting the need for tailored data collection

strategies. By systematically converting data availability into a binary format, the A matrix not only facilitates quantitative comparison across states but also allows researchers to track improvements in reporting quality over time, assess the effectiveness of administrative reforms, and benchmark states against national standards.

Moreover, this binary framework supports the integration of agricultural data with complementary datasets, including yield records, market prices, and economic indicators, enabling comprehensive assessments of productivity, profitability, and food security. The matrix also provides a practical tool for targeting interventions aimed at enhancing data coverage, such as prioritizing regions for capacity-building, designing standardized reporting protocols, and identifying under-reported crops that may require more focused monitoring. Ultimately, by providing a clear, structured representation of data availability, the A matrix serves as both a diagnostic and planning instrument, guiding policymakers, researchers, and stakeholders toward more evidence-based, data-driven strategies for agricultural development across India.

V. CONCLUSION

The binary state–crop availability matrix A provides a robust and systematic framework for evaluating the completeness and reliability of agricultural data across India over the five-year period from 2020 to 2024. By converting the presence or absence of crop production records into a binary format, the matrix allows for clear visualization and quantitative assessment of data coverage across all states and Union Territories. The matrix highlights distinct differences in reporting patterns, with larger, crop-diverse states such as Karnataka, Madhya Pradesh, and Tamil Nadu exhibiting consistently high completeness due to extensive agricultural activity and well-established reporting systems. In contrast, smaller or less agriculturally intensive regions, including Union Territories like Lakshadweep, Chandigarh, and Daman and Diu, tend to display a predominance of missing entries, reflecting limited cultivation, resource constraints, or inconsistencies in administrative reporting.

This structured representation provides researchers, policymakers, and agricultural planners with a powerful tool for identifying gaps in data coverage, prioritizing efforts to improve collection protocols, and ensuring more equitable and comprehensive representation of all crops and regions. By highlighting under-reported crops or states with incomplete records, the matrix can inform targeted interventions, capacity-building initiatives, and standardization of reporting practices. Moreover, the A matrix can serve as a foundation for integrating additional layers of information, such as crop yields, economic indicators, climate data, and remote-sensing observations, enabling more sophisticated analyses of productivity, regional food security, and policy impacts. Future extensions of the matrix may include a broader range of crops, finer temporal resolution (such as quarterly or seasonal data), and direct integration with predictive crop yield models. Such enhancements would allow for more precise, data-driven decision-making and strategic planning, ultimately strengthening the ability of Indian agricultural authorities, researchers, and policymakers to monitor, manage, and optimize crop production systems nationwide.

REFERENCES

[1] S. Neogi and B. K. Ghosh, "Evaluation of crop diversification on Indian farming practices: A Panel Regression Approach," *Sustainability*, vol. 14, no. 24, p. 16861, 2022. https://doi.org/10.3390/su142416861

[2] A. Kumar, S. Nayak, and M. R. Pradhan, "Status and determinants of crop diversification: Evidence from Indian states," Letters in Spatial and Resource Sciences, vol. 17, no. 1, 2024. https://doi.org/10.1007/s12076-023-00366-4

[3] M. E. Dulloo et al., "Varietal Threat Index for Monitoring Crop Diversity on Farms in Five Agro-Ecological Regions in India," *Diversity*, vol. 13, no. 11, p. 514, 2021. https://doi.org/10.3390/d13110514

[4] J. Fennell, "A long-term perspective on collecting Indian agricultural statistics: reviewing the purposes, methods, and implications for Indian development policy," Indian Economic Review, vol. 59, p. 309-325, 2024. https://doi.org/10.1007/s41775-024-00219-x

[5] A. K. Basantaray, R. R. Acharya, and R. K. Patra, "Crop diversification and income of agricultural households in India: An empirical analysis," *Discover Agriculture*, vol. 2, no. 8, 2024. https://doi.org/10.1007/s44279-024-00019-0

[6] FAO, *The State of Food and Agriculture 2021: Making agri-food systems more resilient to shocks and stresses*, Rome: Food and Agriculture Organization of the United Nations, 2021. [Online]. Available: https://doi.org/10.4060/cb4476en

[7] R. N. Mandapati et al., "Crop Yield Assessment Using Field-Based Data and Crop Models at the Village Level: A Case Study on a Homogeneous Rice Area in Telangana, India," *AgriEngineering*, vol. 5, no. 4, pp. 1909–1924, 2023. https://doi.org/10.3390/agriengineering5040117

[8] S. Sathiyamurthi et al., "Assessment of crop suitability analysis using AHP-TOPSIS and geospatial techniques: A case study of Krishnagiri District, India," Environmental and Sustainability Indicators, vol. 24, p. 100466, 2024. https://doi.org/10.1016/j.indic.2024.100466

[9] R. N. V. Jagan Mohan, P. S. Rayanoothala, and R. Praneetha Sree, "Next-gen agriculture: Integrating AI and XAI for precision crop yield predictions," *Frontiers in Plant Science*, vol. 15, p. 1451607, 2025. https://doi.org/10.3389/fpls.2024.1451607