Received: 07 July 2025 | Revised: 05 August 2025 | Accepted: 27 August 2025 | Published: 30 September 2025

RESEARCH ARTICLE
DOI: 10.47703/ejebs.v69i3.539

Spatial Zoning of Agrotechnological Hubs in Kazakhstan: Developing a Methodological Framework

Nurbakhyt N. Nurmukhametov[1] | **Alexander A. Tsoy**[2] | **Meiirzhan Abdykadyr**[2]*

[1] Korkyt Ata Kyzylorda University, Kyzylorda, Kazakhstan
[2] University of International Business named after K. Sagadiyev, Almaty, Kazakhstan

Corresponding author:
*Meiirzhan Abdykadyr – Researcher, University of International Business named after K. Sagadiyev, Almaty, Kazakhstan.
Email: meiirzhanabdykadyr@gmail.com

How to cite this article:
Nurmukhametov, N.N., Tsoy, A.A. & Abdykadyr, M. (2025). Spatial Zoning of Agrotechnological Hubs in Kazakhstan: Developing a Methodological Framework. Eurasian Journal of Economic and Business Studies, 69(3), 36-50.

Conflict of interest:
author(s) declare that there is no conflict of interest.

ABSTRACT

The development of Kazakhstan's agro-industrial complex requires the search for practical tools for the territorial location of innovation infrastructure. The purpose of this study is to develop a methodology for spatial zoning of agro-technological hubs in Kazakhstan based on quantitative assessment of innovation and agricultural potential of regions. The study uses microdata from World Bank Enterprise Surveys for 2024 on the formal agroindustrial sector and related industries, including processing, production, agricultural machinery and services. Using ten indicators normalised using the min–max method and aggregated with equal weights, it was constructed integral indicators such as the Innovation Potential Index (IPI) and the Agricultural Production Potential Index (API). The average values for these indices vary from IPI=0.052 to API=0.240 for the least developed regions and IPI=0.231 to API=0.413 for the most developed ones. The results showed that areas with high potential require consolidation of hubs, development of applied research, and development; territories with medium potential need technology transfer mechanisms, management practices; and regions with low potential need basic competencies formation, digitalization and modernization of infrastructure. The method is replicable and transportable to future WBES waves; limitations include the focus on the formal sector (WBES does not cover primary farms and informal units), as well as the cross-sectional design. Overall, the methodology can be used to monitor the dynamics of regional development and inform strategic adaptation, and it can be applied to future waves of WBES and other countries' industries.

KEYWORDS: Agrotechnology, Agrohub, Agricultural Economy, Innovation, Spatial Zoning, Region, Regional Development

SCSTI: 06.71.07

JEL Code: Q16; Q18; R58

FINANCIAL SUPPORT: This research has been funded by the Science Committee MSHE RK (Grant "Research of new tools for the development of agro-technological hubs in the regions of Kazakhstan to increase the country's competitiveness in the Eurasia region" No. AP23489234).

EJEBS

1. INTRODUCTION

Kazakhstan's agri-food economy is undergoing gradual restructuring from land- and resource-intensive production to more value- and technology-driven activities. Realising this transition is more than a matter of firm-level upgrading; it relies on spatially coherent ecosystems, where producers, processors, service providers, universities, finance, and standards bodies interact at low coordination costs. In this setting, agrotechnological hubs present a pragmatic option to densify services such as testbeds, extension-like advisory services, quality infrastructure, and growth finance. The policy question is where to locate such hubs and how to specialise their functions across diverse regions.

This study addresses the aforementioned challenge by developing a replicable, empirically driven approach to spatial zoning for the agrotechnology industry in Kazakhstan. It was leveraged the 2024 World Bank Enterprise Survey (WBES), which provides nationally representative microdata for the formal, private, agri-adjacent economy mainly food and beverage processors, producers of agricultural equipment and machinery, as well as service companies supporting production and commercialisation. Although the WBES does not cover primary farms or informal firms, it successfully covers the segment where many practices regarding technology adoption, processing, logistics, and management that impact value addition and market access are standard.

Our key hypothesis is that hub readiness is determined together by two latent abilities: (i) innovation potential a company's inclination to launch new products or processes, invest in research and development, and take up external technologies; and (ii) agricultural/production potential the richness of operational and market capabilities manifested in scale, capacity utilization, managerial experience, and digital connectivity. Since the constructs are not directly measurable, it was approximated them with a concise, policy-relevant set of indicators that are regularly available in WBES. Standardising the indicators to a standard scale and consolidating them into two intuitive composite measures gives the establishment-level Innovation Potential Index (IPI) and Agricultural Potential Index (API), which together chart the technological and production locations of companies.

Building on these micro measures, it was proceeded in two steps. First, it used unsupervised classification in the IPI–API plane to characterise firm heterogeneity in an interpretable manner (innovation-led, production-anchored, baseline-low archetypes). This diagnostic is not an end in itself; instead, it provides a microeconomic rationale for the types of services hubs will need to offer in various settings. Second, translated micro signals to space by aggregating IPI and API with WBES probability weights to the survey's seven regions and applying a parsimonious clustering to the set of regional points. The outcome is a three-tier zoning High, Medium, Low that summarises the joint profile of innovative and production capacity at the regional scale, suitable for targeting and sequencing hub interventions.

This paper makes three significant contributions. Conceptually, it codifies the idea that the location of agri-tech hubs must be predicated on the interaction between knowledge absorption and generation, as well as production depth, rather than considering them independently in either dimension. Methodologically, it offers an entirely transparent process including indicator selection based on a commonly accepted questionnaire, explicit normalisation, equal-weight indices, design-consistent territorial aggregation, and standard clustering that can be replicated or tested with alternative weights or scales. Practically, it produces a policy-relevant map that delineates areas for hub concentration (intense production deepness and satisfactory innovation), diffusion poles (urban knowledge centers with mid-level production depth), and foundation areas (areas where basic capabilities need to be established before advanced tools can be effectively applied).

Two particular boundary conditions are stated. First, because the WBES focuses on the formal sector and excludes primary agriculture, the zoning demarcates the formal agro-industrial and agri-tech ecosystem rather than the entire range of agriculture. Second, the study is cross-sectional; it clarifies patterns and informs targeting but does not determine causal links. However, these limitations are offset by the portability of the method. As future waves of WBES or new administrative and remote-sensing datasets become available, the indices and zoning can be re-assessed to gauge progress and refine hub strategy as necessary.

Against this background, the objectives of the research are fourfold: (1) to construct establishment-level indices of innovation and agricultural/production potential using WBES microdata; (2) to classify firms to clarify capability archetypes relevant to hub services; (3) to synthesize and aggregate regional profiles to obtain a brief and understandable spatial zoning; and (4) to obtain actionable implications for the strategic design, placement, and sequencing of agrotechnology hubs in Kazakhstan.

2. LITERATURE REVIEW

Innovation and spatial concentration research offer the conceptual basis for zoning agrotechnological hubs. Classic cluster work suggests that co-location enhances firm productivity, stimulates innovation through knowledge spillovers, and accelerates new business creation (Porter, 1998). Later empirical research demonstrates that areas of complementary, specialised activity grow faster and upgrade technologically (Delgado, Porter, & Stern, 2014), while the geography of innovation is strongly correlated with localised R&D spillovers (Audretsch & Feldman, 1996). Parallel literatures stress regional innovation systems the institutional and network structure through which firms, universities, finance, and government collectively produce and disseminate knowledge (Cooke, 2001; Cooke, 1997) and sectoral systems of innovation, which emphasise technology-, demand-, and actor-specific processes in industries like agrifood (Malerba, 2002). Collectively, these literatures provide a rationale for a spatial focus to targeting agri-tech initiatives.

In agriculture, the Agricultural Innovation Systems (AIS) framework reimagines innovation as a problem-solving, multi-actor process, rather than a linear R&D pipeline. The World Bank's AIS sourcebook codifies design principles for managing research, extension, finance, and market linkages (World Bank, 2012), while OECD guidance outlines the state's enabling role (OECD, 2013). Methodologically, recent studies take participatory and systems-oriented approaches, including Delphi-based consensus building for AIS diagnostics and FAO training manuals that operationalise AIS into tools (Toillier et al., 2022; FAO, 2022, 2021). This literature underpins composite, multi-indicator measurement of regional agri-tech readiness and diffusion capacity.

An emerging practice literature discusses innovation hubs and digital innovation hubs (DIHs) as place-based institutions that orchestrate services such as testbeds, brokerage, and skills especially for rural SMEs and agrifood chains. There is evidence that DIHs can reduce adoption costs for digital process and market innovations, enhance local entrepreneurial ecosystems, and deliver better sustainability outcomes (Stojanova et al., 2022). These findings lend policy relevance to zoning as a means of aligning hub functions with regional capability profiles.

The operationalisation of systems-related concepts into quantifiable frameworks typically relies on the use of composite indicators and unsupervised classification methods. The OECD–JRC Handbook outlines the best practices for selecting, normalising, weighting, and aggregating indicators, as well as conducting robustness tests (OECD & JRC, 2008/2005). Subsequent methodological advancements emphasise the impact of weights and aggregation methods on outcomes, promoting transparency and rigorous stress-testing procedures (Greco, Ishizaka, Tasiou, & Torrisi, 2019; Becker, Saisana, Paruolo, & Saltelli, 2017). These recommendations align with the construction of simple, reproducible indices that quantify innovation and production potential, then classify regions based on their joint distribution.

For zoning, K-means clustering remains a popular and interpretable method for dividing observations by reducing within-cluster dispersion (MacQueen, 1967; Lloyd, 1957, as cited in Jin, 2011). Internal validity is routinely evaluated using the Calinski–Harabasz criterion and silhouette coefficients (Caliński & Harabasz, 1974; Rousseeuw, 1987), both of which are implemented in standard statistical software and widely applied in spatial analysis.

Applied agronomic research shows that multi-indicator panel clustering can produce actionable management zones and spatial stratifications for agronomy and value-chain policy. Such examples as fuzzy or hard K-means clustering on agro-ecological and management factors, with or without preceding dimensionality reduction (Yuan et al., 2022; Reyes et al., 2023), and landscape-metric clustering with silhouette diagnostics to map out homogeneous intervention zones (Fang et al., 2025) inspire our two-dimensional zoning in innovation–production space.

The use of comparable high-quality microdata is central to successful zoning. The World Bank Enterprise Surveys (WBES) provide a nationally representative dataset that yields firm-level information regarding innovation, management practices, infrastructure, and performance in the formal sector for more than 160 countries (Enterprise Surveys, 2024). The innovation modules of WBES tested through dedicated methodological studies successfully measure both product and process innovations as well as research and development activities using brief yet informative questionnaires (Cirera, Fattal, & Maemir, 2016). The methodological design and exhaustive documentation of the survey enable probability-weighted regional aggregation and cross-wave portability, making it suitable for spatial analysis targeting agri-adjacent ecosystems.

For Kazakhstan, the nascent literature discusses innovation management in agriculture, cluster policy for the agro-industrial complex, and value chain governance. Research highlights the necessity of coordinated technology transfer, managerial upgrading, and institutional support to convert innovation inputs into productivity and diversification (Taishykov, 2024; Manatovna et al., 2023; Tkacheva et al., 2024). Previous policy critiques warn that cluster initiatives require realistic diagnostics of regional capabilities and linkages to achieve success (Wandel, 2010). Complementary World Bank operations emphasise instruments for commercialisation and applied research as components of national innovation policy (World Bank, 2013; World Bank, 2020). This literature inspires a measurement classification strategy that is transparent, survey-anchored, and specific to Kazakhstan's formal agro-industrial sector.

The analysis of the literature has shown that the concepts of cluster development, regional and sectoral innovation systems serve as the theoretical basis for the zoning of agrotechnological hubs. The existing methodology is based on proven practices of building composite indexes, using clustering methods and using microdata, which makes it possible to quantify innovation and production potential. Despite the extensive research on agro-innovation systems and digital hubs, their application in the context of Kazakhstan is limited to conceptual descriptions and individual cases without quantifying the potential based on representative microdata. In addition, little attention has been paid to adapting international methods to the specific needs of the formal agro-industrial sector in Kazakhstan, which accounts for a significant portion of production by small and medium-sized enterprises. Therefore, this research addresses this issue by developing a method for spatially dividing agrotechnological hubs, which involves assessing their innovation and production capabilities, categorizing them statistically, and creating regional profiles for informed strategic planning.

3. RESEARCH METHODS

This study formulates the identification of spatial areas for agrotechnological hubs in Kazakhstan using a two-stage measurement and classification approach, leveraging microdata from the 2024 World Bank Enterprise Survey (WBES). Since the WBES does not cover primary agricultural establishments, the analytical scope covers agri-adjacent firms in the agrotechnology ecosystem, mainly food and beverage processors, producers of agricultural machinery and equipment, as well as service activities supporting production and commercialisation. The establishment is the unit of analysis, with each record belonging to one of the stratified regions defined in the survey. All analysis is performed using Stata to ensure consistency of terminology with the WBES questionnaire, thereby enhancing replicability and adaptability in future iterations. The target construct consists of a duality of latent capacities, innovation potential, and agricultural/production potential, approximated through ten establishment-level indicators that are observable and relevant for policy purposes. Innovation potential is measured by indicators such as recent product and process innovations, the size of R&D expenditure, and the use of foreign-licensed technologies. Agricultural/production potential, on the other hand, is measured in terms of total annual sales, capacity utilisation compared to maximum possible output, employment levels at the time of start-up as an initial-scale indicator, sectoral experience of the lead manager, and having a website or social media presence, which is an indicator of digital connectivity.

The analysis takes over the seven WBES stratification regions Almaty City; Astana City; Centre (Karaganda, Ulytau); East (Abay, East Kazakhstan); North (Akmola, Kostanay, Pavlodar, North Kazakhstan); South (Almaty oblast, Jambyl, Zhetisu, Kyzylorda, Turkestan, Shymkent City); and West (Aktobe, Atyrau, West Kazakhstan, Mangistau) which group administrative units to obtain sufficient sample sizes and capture salient economic geography. Including Almaty and Astana as stand-alone regions captures the outsized contributions of these metropolitan knowledge and service centers, while the grouped Center, East, North, South, and West categories concatenate contiguous oblasts with broadly similar production structures (e.g., export-oriented hydrocarbons in the West; diversified crop–livestock systems in the North; higher population density and labor markets in the South). This stratification underlies the survey weights employed for regional aggregation and is the operative spatial scale for our zoning. All regional means and cluster assignments are for these composite regions, not individual oblasts, which is relevant for interpreting policy recommendations and benchmarking across territories. Table 1 presents definitions of indicators used in the innovation and agricultural potential indices.

TABLE 1. Definitions of indicators used in the innovation and agricultural potential indices

Block	Indicator (short name)	WBES item	Raw data type	Questionnaire wording (abridged)	How it enters the index	Interpretation
Innovation	New product/ service	H.1	Binary (0/1)	During the last three years, has the establishment introduced new or improved products or services?	As 0/1; no transformation beyond the normalisation step	1 ⇒ more vigorous product-side innovation activity
	New process	H.5	Binary (0/1)	During the last three years, has the establishment introduced any new or improved processes?	As 0/1; no transformation beyond the normalisation step	1 ⇒ stronger process-side innovation/operations upgrading
	Any R&D	H.8	Binary (0/1)	In the last fiscal year, did the establishment spend on	As 0/1; no transformation	1 ⇒ positive R&D

				R&D (in-house or contracted)?	beyond the normalisation step	effort/absorptive capacity
	R&D amount	H.9	Continuous (currency)	How much did the establishment spend on R&D in that year?	Min–max normalised to [0,1]	Higher ⇒ greater R&D intensity/resources
	Foreign-licensed tech	E.6	Binary (0/1)	Does the establishment use technology licensed from a foreign-owned company?	As 0/1; no transformation beyond the normalisation step	1 ⇒ stronger external technology adoption
Agri/Production	Total annual sales	D.2	Continuous (currency)	Establishment's total annual sales for all products/services in the last fiscal year	Min–max normalised to [0,1]	Higher ⇒ larger scale/market penetration
	Capacity utilization	F.1	Continuous (%)	Output produced as a % of the maximum feasible output using all physical capital	Min–max normalised to [0,1]	Higher ⇒ better utilisation/efficiency
	Start-up employment	B.6	Continuous (count)	Number of full-time workers when the establishment started operations	Min–max normalised to [0,1]	Higher ⇒ larger initial scale/growth headroom
	Manager's sector experience	B.7	Continuous (years)	Years of experience of the top manager in this sector	Min–max normalised to [0,1]	Higher ⇒ more substantial managerial human capital
	Digital readiness	C.22b	Binary (0/1)	Does the establishment have its website or social media page?	As 0/1; no transformation beyond the normalisation step	1 ⇒ better market connectivity/digital capability

Note: compiled based on WBES (2024)

A complete-case sample is used for the variables under study; binary items are inserted unchanged, while continuous items are standardized using min–max normalization applied within the analytical sample as per formula (1):

$$x_i^* = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (1)$$

where:

x_i – the original value for observation i;

$\min(x)$ – the smallest value in the dataset;

$\max(x)$ – the largest value in the dataset;

x_i^* – the idiosyncratic error term, assumed to be independent and identically distributed.

To construct the latent constructs, indicators were pre-selected to correspond directly with the standard World Bank Enterprise Survey (WBES) items, thus ensuring conceptual consistency and allowing for cross-wave comparability. Raw variables were screened for obvious data-entry errors prior to normalisation, in line with questionnaire skip logic and internal consistency checks (e.g., positive sales for operating establishments and plausible bounds for capacity utilisation). Min–max scaling was conducted within the analytical sample so that all inputs fall in the range of [0,1] [0,1] [0,1], thus conserving the ordinal properties inherent in each indicator and enabling direct comparison of the two composite measures across different units. Where continuous indicators had long right tails, typical of financial and size variables, we took care that the min–max transformation did not introduce leverage through a small number of outliers; as a robustness check (reported in another context), it was replicated the procedure under mild winsorization and used a z-score transformation and obtained the same qualitative zoning results.

Two establishment-level composite indices are then created as transparent equal-weight averages of their normalised constituents the Innovation Potential Index (IPI) and the Agricultural Potential Index (API) to prevent the embedding of untestable priors regarding the relative significance of separate indicators, while leaving open the possibility of sensitivity analysis using alternative schemes as per formula (2):

$$IPI_i = \frac{1}{5}\sum_{j=1}^{5} z_{ij} \quad API_i = \frac{1}{5}\sum_{k=1}^{5} z_{ik} \qquad (2)$$

where:

IPI_i – Innovation Potential Index for establishment i;

API_i – Agricultural Potential Index for establishment i;

z_{ij} – the j-th normalized innovation indicator for establishment i;

z_{ik} – the k-th normalized agricultural/production indicator for establishment i.

Equal weights are favoured here since they (i) optimise transparency and reproducibility across users and waves; (ii) minimise the danger of over-fitting weights to an individual cross-section; and (iii) enable diagnostic decomposition, as each indicator enters additively and on the same scale. Be that as it may, the setup is modular: weighting schemes (e.g., information-theoretic or expert-elicited) can be replaced with alternative ones without altering the surrounding aggregation and classification logic, enabling simple stress tests of the zoning to alternative normative choices.

In an initial diagnostic microeconomic analysis, establishments are classified in a two-dimensional space defined by (IPI, API) via the K-means clustering algorithm with Euclidean distance and a large number of random initialisations. The number of clusters is set in advance to three, aligning with policy-relevant categories (High, Medium, Low), and labels are then assigned by ordering the centroids along each dimension. K-means is run with a fixed random seed and a large number of initial configurations to reduce the risk of local minima. It was assessed internal validity via conventional measures of separation and compactness. These diagnostics are used only to ensure that the indices capture significant heterogeneity and are not intended to serve as the zoning mechanism. The resulting three clusters are thus interpreted as archetypes: innovation-led, production-anchored, and baseline-low, which inform the types of services that hubs might need to provide in different contexts (e.g., testbeds, diffusion support, and foundational capability development).

In order to map micro signals into spatial hubs, establishment-level indicators are weighted up to the regions of the survey using WBES probability weights so that regional statistics correspond to the target population instead of the realised sample; the pair of design-consistent characteristics each region means, as per formula (3):

$$\overline{I_r} = \frac{\sum_{i \in r} w_i I_i}{\sum_{i \in r} w_i} \quad \text{for } I \in \{IPI, API\} \qquad (3)$$

where:

$\overline{I_r}$ – the weighted mean value of the index I (IPI or API) for region r;

w_i – the WBES probability weight for establishment i;

I_i – the index value (*IPI* or *API*) for establishment i.

Weights are directly derived from the WBES sampling design and reflect stratification by industry, size, and geography. Their use in the aggregation process preserves the representativeness of the survey and guards against potential bias arising from unequal selection probabilities or differential non-response rates across strata. Final zoning is achieved by applying K-means clustering to three groups, where clusters are labelled by ordered centroid values, allowing for consistent interpretation in terms of combined innovation and agricultural/production potential. Classification at the meso level is the sole instrument for spatial targeting. Its construction is deliberately reproducible and straightforward: any researcher with access to the same microdata can recalculate the indices, re-aggregate using the exact weights, and reapply the clustering algorithm. It was emphasised that the labels so assigned are algorithmically derived rather than normative; they capture joint positions in the IPI–API space and are intended to align policy bundles with capability profiles. Finally, it was noted that two boundary conditions on interpretation: the WBES framework covers formal, agri-adjacent players but not primary farms and informal micro enterprises; and the analysis is cross-sectional, yielding a snapshot for decision support rather than shedding light on causal relationships.

4. RESULTS AND DISCUSSION

The results transition from the micro to the mesoscale. It was first described firm heterogeneity in the joint innovation–production space by reporting the composition of the three unsupervised clusters in each WBES region (Table 2), followed by an interpretation of the cluster centroids that explains how the algorithm splits establishments along the Innovation Potential Index (IPI) and Agricultural/Production Potential Index (API) dimensions (Table 3). It was then visualised the distribution of establishments in the IPI–API plane with K-means assignments to visualise separation and within-cluster dispersion (Figure 1). Building on these diagnostics, it was possible to map micro signals to space by calculating design-consistent regional means of IPI and API and fitting a parsimonious three-way partition to the seven regional points, which provides the ultimate spatial zoning of agrotechnology potential (Table 4). Firm-level clustering is reported as a diagnostic to inform hub service design throughout, while regional clustering based on survey-weighted aggregates serves as the zoning tool for policy targeting. Table 2 shows Firm-level clusters by region (counts and row percentages).

TABLE 2. Firm-level clusters by region (counts and row percentages)

Region	High Pot.	Medium Pot.	Low Pot.	Total	High %	Medium %	Low %
Almaty	8	3	7	18	44.4	16.7	38.9
Astana	3	5	6	14	21.4	35.7	42.9
Center	8	4	7	19	42.1	21.1	36.8
East	6	5	9	20	30.0	25.0	45.0
North	6	3	11	20	30.0	15.0	55.0
South	5	4	7	16	31.3	25.0	43.8
West	2	9	12	23	8.7	39.1	52.2
Total	38	33	59	130	29.2	25.4	45.4
*Cluster labels reflect the K-means output at the firm level; percentages are within rows							

Note: compiled by the authors

Table 2 summarises the structure of establishments over three unsupervised clusters in the IPI–API plane by WBES region. The labels are to be read as algorithmic, not normative: the centroid diagnostics indicate that the cluster labeled "Medium Potential" clusters the innovation-intensive firms (highest IPI, mid-range API), "Low Potential" clusters production-anchored firms (higher API, moderate IPI), and "High Potential" includes baseline-low firms (low on both indices). The table presents raw counts and row percentages (shares within each region).

The regional profiles are quite different. West is the most innovation-intensive composition, with the highest proportion of "Medium Potential" companies and the lowest presence of baseline-low companies. This suggests a relatively dynamic agri-tech sector with opportunities for expanding R&D and R&D-commercialisation connections (e.g., pilot testbeds, supplier development, growth financing). Astana also has a relatively high percentage of innovation-intensive establishments, in line with an urban knowledge base that can act as a source of technology diffusion to the surrounding production systems.

In contrast, North has the largest share of production-anchored firms and a relatively low fraction of innovation-intensive firms. This structure suggests aggressive technology transfer and adoption initiatives mechanization upgrading, process quality regimes, and digital operations more than frontier R&D. East and South exhibit mixed structures with large production-anchored fractions and modest innovation-intensive fractions; in these cases, balanced policies that blend diffusion (managerial and process upgrading, digital market access) with selective innovation partnerships will likely generate the highest marginal returns.

Almaty and the Centre show comparatively high baseline-low segments together with non-negligible production-anchored shares and lower innovation-intensive proportions. In practical terms, these areas may require a two-phase strategy: capability building (labour force qualifications, lean/process routines, basic digitalisation) to shift companies out of the baseline-low group; second, selective diffusion mechanisms to link promising manufacturers with urban innovation resources.

Collectively, the cross-regional differences in cluster composition suggest a microeconomic basis for differentiated hub strategies. More innovation-intensive regions (West, Astana) are candidates for agri-tech hub consolidation that prioritises commercialisation channels and growth capital. Production-anchored dominated regions (North, East, South) should emphasise diffusion extension-like services for tech adoption, vendor development, and logistics/digital market connectivity. Those with larger baseline-low segments (Almaty, Centre) require foundational capability building prior to which advanced instruments will be effective. Since Table 2 presents unweighted counts and within-region shares, these patterns should be interpreted as compositional signals rather than population totals. In the analysis that follows, survey weights are used to aggregate to the region and derive the final spatial zoning.

Table 3 shows firm-level cluster centroids (mean IPI and API).

TABLE 3. Firm-level cluster centroids (mean IPI and API)

Cluster label	Mean IPI	Mean API
High Potential	0.051	0.248
Medium Potential	0.567	0.442
Low Potential	0.136	0.479
*Means are from the establishment-level indices used in clustering		

Note: compiled by the authors

The centroids in Table 3 summarise the locations of each cluster in the two-dimensional space of the normalised indices (IPI, API). The separation is driven by the innovation dimension: the "Medium Potential" cluster has the highest IPI (innovation-intensive firms), the "Low Potential" cluster has a modest IPI but the highest API (production-anchored firms), and the "High Potential" cluster is low on both indices (baseline-low firms). Since the labels come from the unsupervised solution rather than a normative ranking, their practical meaning is: (i) an innovation-led group with above-average innovation and mid-range agricultural potential; (ii) a production-anchored group with relatively strong agricultural/production capability but only moderate innovation; and (iii) a baseline-low group with weak scores on both dimensions. Policy implications follow directly from these centroid positions. Firms in the innovation-led cluster are candidates for scaling and commercialisation instruments (testbeds, growth finance, IP/standards support) that translate innovative effort into market penetration and supply-chain depth. Firms in the production-anchored cluster are the natural targets for technology diffusion and adoption (process upgrading, quality certification, digital operations, equipment modernisation) to improve their innovation capacity without compromising production strength. Firms in the baseline-low cluster require foundational capability building namely, managerial training, basic digitalisation, and access to extension-like services before more advanced instruments can be effective. In short, the centroid geometry suggests a trade-off across regions between innovation intensity and production depth, with a third group that performs poorly on both; aligning instruments to these differential positions should deliver the most significant marginal gains.

Figure 1 shows establishment-level innovation potential (*IPI*) vs. agricultural/production potential (*API*) with k-means cluster assignment.

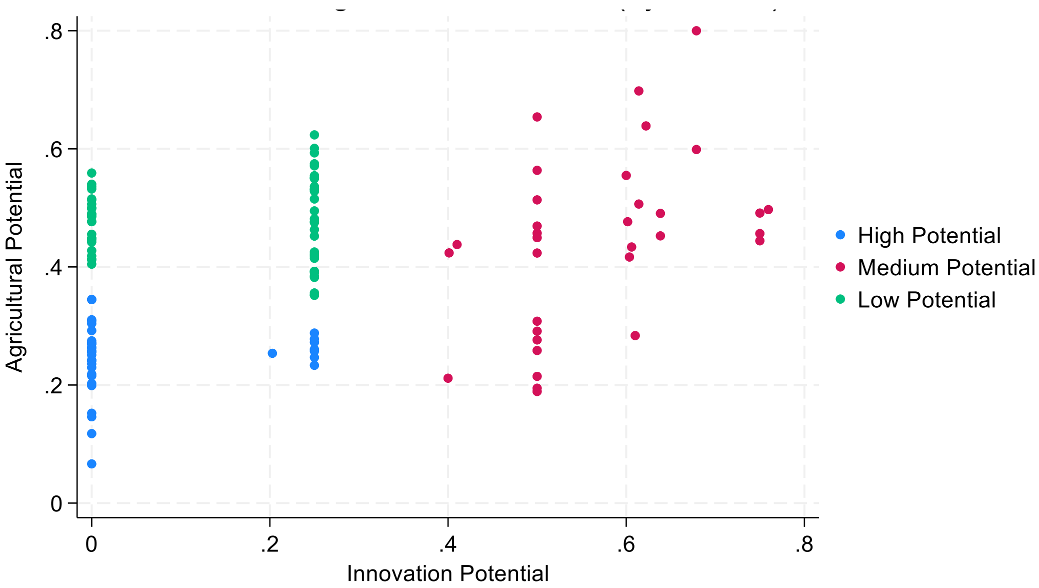

FIGURE 1. Clusters of innovation and agricultural potential

The scatter plot unlocks three statistically distinct groups of establishments within the normalised *IPI–API* plane, separated along the innovation axis (x) initially. Red-coded points "Medium Potential" per the algorithm fill out the right side with the highest IPI and broad vertical dispersion in API, signifying innovation-intensive companies whose production depth varies from modest to very strong; for them, scaling and commercialization tools (e.g., testbeds, standards support, growth finance) need to be combined with supply-chain and process improvement for those lower down on API. The green cluster ("Low Potential") is centred around moderate IPI but relatively high API, characterising production-anchored companies that prioritise technology diffusion and adoption process innovation, quality systems, and digital operations to enhance IPI without compromising their production strengths. Blue points "High Potential" per the unsupervised label but empirically baseline-low are positioned near very low IPI and low-to-mid API, indicating foundational capability gaps. Such firms need basic managerial training, lean/quality practices, entry-level digitalisation, and access to working capital before higher-order innovation instruments are effective.

Two structural aspects are notable. First, tri-modality along IPI with partial overlap in API suggests that innovation capacity is the primary stratifier in this sample, with production potential differing within clusters. Second, API variance increases with IPI (greater vertical scatter at higher x), indicating that innovation alone is not a guarantee of strong production performance some innovation-active companies have yet to succeed in converting knowledge inputs into operational depth. Practically, this map substantiates a staged pathway: shifting baseline-low companies toward production-anchored performance (blue→green) through capability building, then from green to red through focused diffusion and co-development that elevates innovation intensity.

Table 4 presents the regional means of IPI and API, along with the assigned zone.

TABLE 4. Regional means of IPI and API and assigned zone

Region	Mean IPI	Mean API	Assigned zone
Almaty	0.241	0.324	Region Medium
Astana	0.198	0.351	Region Medium
Center	0.185	0.323	Region Medium
East	0.126	0.358	Region High
North	0.166	0.360	Region High
South	0.052	0.240	Region Low
West	0.150	0.431	Region High

*Regional means are survey-weighted averages of establishment indices.
**Zoning is obtained by applying K-means clustering to the seven regional points in the IPI–API plane and labelled by ordered centroid values.

Note: compiled by the authors

Table 4 embeds establishment-level signals in a regional zoning by averaging the innovation (IPI) and agricultural/production (API) indices with WBES probability weights and then clustering the seven regional points in the IPI–API plane. The resulting High / Medium / Low tiers represent ordered centroid positions of these regional means and thereby capture joint innovative and agricultural potential more than either dimension in isolation.

High zone – East, North, West

These territories strike a balance between relatively good agricultural/production potential and sufficient innovation. West has the deepest production profile of all territories, and East and North have balanced profiles with good API and medium IPI. Practically, these lands are poised for hub consolidation: tools that expand and formalise value chains (quality infrastructure, cold chain and logistics, supplier development) supplemented by applied R&D and technology demonstration (pilot testbeds, mechanisation and process improvement, digital operations) should deliver quick productivity and market dividends.

Medium zone - Almaty, Astana, Centre.

These areas are innovation-biased compared to their agricultural foundation: they are home to companies with superior IPI but just mid-level API. They are diffusion nodes by nature urban knowledge and service hubs from which technology, managerial methods, and digital market access can be transferred to nearby production systems. Policy priorities must focus on the linkage mechanisms (university–industry collaboration, extension-like services adapted to processors and logistics providers, standards and certification assistance) that translate innovative efforts into broader supply-chain upgrading.

Low zone – South

This area is consistently weak across both indices, indicating that companies face limitations in both capability and scale simultaneously. The policy sequence is accordingly foundational capacity building: workforce skills, lean/quality management, basic digitalisation, access to working capital, and core infrastructure. Only once these foundations are established will more sophisticated innovation tools be practical.

Two further points are worth noting. First, the fact that Table 3 reports survey-weighted regional means implies that the zoning corresponds to the expected demographic of formal agri-adjacent firms rather than simply the realised sample taken. Second, the tiers are the outcome of a joint assessment of the IPI and API; a region can therefore gain entry to the High zone either through great production depth in combination with satisfactory innovation (as in the West) or through a balanced, above-average performance on both dimensions (as in the East and North). This combined perspective provides a coherent, policy-relevant map: concentrate centres where production depth is already consolidated and innovation is satisfactory; spread innovation from urban centres where the knowledge base exceeds that of agriculture; and provide support to areas where both competences are weak.

5. CONCLUSIONS

This paper crafts and implements an open, survey-based pipeline to map spatial zones for agrotechnology hubs in Kazakhstan. With World Bank Enterprise Survey microdata, two latent capacities —innovation potential and agricultural/production potential —were operationalised using ten establishment-level indicators mapped directly onto standard WBES items. Following harmonisation of heterogeneous measures through min–max scaling and the construction of equal-weight composite indices, it was (i) diagnosed firm heterogeneity in the IPI–API plane and (ii) decoded micro signals into region-level zoning through probability-weighted aggregation and K-means clustering. The emergent three-tier map is interpretable and policy-ready: East, North, and West are revealed as consolidation candidates with relatively strong production depth and sufficient innovation; Almaty, Astana, and the Centre serve as diffusion nodes with higher innovation compared to agricultural depth; and South shows foundational gaps on both dimensions.

Three substantive contributions ensue. First, the measurement approach is replicable: indicators, normalisation, index construction, and aggregation are completely specified and portable to future WBES waves, allowing time-consistent updates without remaking the method. Second, the classification is joint in innovation and production, sidestepping the usual trap of ranking regions on one dimension and instead acknowledging that hub readiness necessitates both absorptive capacity and operational depth. Third, the pipeline is diagnostic at two levels: it brings to the surface establishment-level archetypes (innovation-led, production-anchored, baseline-low) and indicates how their mix differs across regions, furnishing microeconomic rationale for differentiated spatial policy.

Policy implications are immediate. In High zones (East/North/West), instruments ought to prioritise hub consolidation and scaling: applied R&D and demonstration testbeds linked to priority value chains; supplier-development programs and quality infrastructure (standards, certification, metrology); logistics and cold-chain upgrades; and blended finance to crowd in private investment for scale-up. In Medium zones (Almaty/Astana/Centre), priority is technology diffusion and linkage formation: university–industry partnerships, extension-like services for processors, managerial upgrading (lean/quality/digital operations), and market-access platforms connecting urban knowledge assets to proximate production. In the Low zone (South), the sequence should prioritise foundational capability building, including workforce skills, entry-level digitalisation, production planning and quality systems, access to working capital, and core infrastructure, before introducing more advanced innovation instruments. Throughout all zones, inclusion and resilience are crucial: SMEs, women-led enterprises, and climate-smart practices should be integrated into program design to prevent exclusion and mitigate vulnerability to climate and market shocks.

Limitations imply a straightforward research agenda. WBES spans the formal sector and excludes primary farms; zoning thus describes the formal agro-industry and agri-tech subsector, not the entire agriculture sector or the informal economy. The regional sample size is moderate (seven strata), and the cross-sectional data preclude causal inference. Equal weighting, though transparent, may misrepresent accurate marginal contributions of indicators in every context. Spatial dependence is only indirectly addressed through regional aggregation, rather than explicit spatial econometrics. Follow-on work should incorporate administrative and remotely sensed data (e.g., yield proxies, water stress, logistics accessibility), as well as agricultural census or firm registry coverage to capture micro and informal units, and network measures of buyer–supplier relationships. Longitudinal analysis with future WBES waves would allow for difference-in-differences or synthetic control assessments of hub interventions. Methodological refinements could include confirmatory factor analysis/SEM) to test the two-construct measurement model, spatial lag/error models to estimate spillovers, and multi-criteria decision analysis to introduce policy weights explicitly. Lastly, careful cost–benefit and distributional analyses should accompany the rollout of hubs to ensure additionality, prevent enclave development, and align incentives between public and private stakeholders. For implementation, it was suggested a practical roadmap: (1) take the current zoning as a targeting screen for pilot hubs; (2) undertake rapid value-chain diagnostics in each high-tier area to choose two to three anchor chains; (3) devise instrument bundles tailored to zone type (consolidation/diffusion/foundation), with clear eligibility and performance criteria; (4) put in place a monitoring system keyed to our indices e.g., proportions of firms reporting product/process innovation, incidence and intensity of R&D, capacity utilization, digital presence, and export or certification take-up so IPI and API can be recalculated every year; and (5) insert review points (e.g., every 18–24 months) to re-estimate the indices with fresh data and rebalance hub location or instrument mix as necessary.

Overall, the analysis presents a rigorous yet feasible approach to measuring, mapping, and prioritising agrotechnology development in Kazakhstan's regions. By combining an open indicator system with survey-weighted aggregation and frugal clustering, it translates dispersed micro evidence into a consistent spatial strategy. The framework does not replace in-depth project design. However, it offers a lasting foundation for where to intervene and what to prioritise, setting the stage for iterative learning as policies are implemented and new data become available.

AUTHOR CONTRIBUTION

Writing – original draft: Nurbakhyt N. Nurmukhametov, Alexander A. Tsoy, Meiirzhan Abdykadyr.
Conceptualization: Nurbakhyt N. Nurmukhametov, Alexander A. Tsoy, Meiirzhan Abdykadyr.
Formal analysis and investigation: Alexander A. Tsoy, Meiirzhan Abdykadyr.
Funding acquisition and research administration: Nurbakhyt N. Nurmukhametov, Meiirzhan Abdykadyr.
Development of research methodology: Nurbakhyt N. Nurmukhametov, Alexander A. Tsoy.
Resources: Nurbakhyt N. Nurmukhametov, Alexander A. Tsoy, Meiirzhan Abdykadyr.
Software and supervisions: Alexander A. Tsoy, Meiirzhan Abdykadyr.
Data collection, analysis and interpretation: Alexander A. Tsoy, Meiirzhan Abdykadyr.
Visualization: Alexander A. Tsoy, Meiirzhan Abdykadyr.
Writing review and editing research: Nurbakhyt N. Nurmukhametov, Alexander A. Tsoy, Meiirzhan Abdykadyr.

REFERENCES

Abdullah, A. J., Doucouliagos, H., & Manning, E. (2015). *Does education reduce income inequality? A meta-regression analysis*. https://doi.org/10.1111/joes.12056

Audretsch, D. B., & Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. *American Economic Review, 86*(3), 630–640. https://www.jstor.org/stable/2118216

Becker, W., Saisana, M., Paruolo, P., & Saltelli, A. (2017). Weights and importance in composite indicators: Closing the gap. *Ecological Indicators, 80*, 12–22. https://doi.org/10.1016/j.ecolind.2017.03.056

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics, 3*(1), 1–27. https://doi.org/10.1080/03610927408827101

Cirera, X., Fattal, R., & Maemir, H. B. (2016). *Measuring firm-level innovation using short questionnaires*. World Bank Policy Research Working Paper. https://openknowledge.worldbank.org/server/api/core/bitstreams/ 1f1b3ecd-731e-57af-be8f-e1cf95695116/content

Cooke, P. (1997). Regional innovation systems: Institutional and organisational dimensions. *Research Policy, 26*(4–5), 475–491. https://doi.org/10.1016/S0048-7333(97)00025-5

Cooke, P. (2001). Regional innovation systems, clusters, and the knowledge economy. *Industrial and Corporate Change, 10*(4), 945–974. https://doi.org/10.1093/icc/10.4.945

Delgado, M., Porter, M. E., & Stern, S. (2014). Clusters, convergence, and economic performance. *Research Policy, 43*(10), 1785–1799. https://doi.org/10.1016/j.respol.2014.05.007

Enterprise Surveys. (2024). *Home Enterprise Surveys*. World Bank. Retrieved from https://www.enterprisesurveys.org/en/enterprisesurveys

Fang, G., Sun, D., Yu, Y., & Zhang, Z. (2025). A landscape-clustering zoning strategy to map multi-functional cropland. *Agriculture, 15*(2), 186. https://doi.org/10.3390/agriculture15020186

Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2019). On the methodological framework of composite indicators. *Social Indicators Research, 141*, 61–94. https://doi.org/10.1007/s11205-017-1832-9

Jin, X. (2011). K-means clustering. In *Encyclopedia of Machine Learning*. Springer. https://doi.org/10.1007/978-0-387-30164-8_425

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (pp. 281–297). University of California Press. https://matteucci.faculty.polimi.it/Clustering/tutorial_html/ kmeans.html?utm_

Malerba, F. (2002). Sectoral systems of innovation and production. *Research Policy, 31*(2), 247–264. https://doi.org/10.1016/S0048-7333(01)00139-1

Manatovna, T. A., Dabyltayeva, N. E., Ruziyeva, E. A., Sakhanova, G., & Yelubayeva, Z. M. (2023). Unlocking intersectoral integration in Kazakhstan's agro-industrial complex: Technological innovations, knowledge transfer, and value chain governance as predictors. *Economies, 11*(8), 211. https://doi.org/10.3390/economies11080211

OECD & Joint Research Centre. (2008/2005). *Handbook on constructing composite indicators: Methodology and user guide*. Paris: OECD Publishing. https://doi.org/10.1787/9789264043466-en

OECD. (2013). *Agricultural Innovation Systems: A framework for analysing the role of the government*. Paris: OECD Publishing. https://doi.org/10.1787/9789264200593-en

Porter, M. E. (1998). Clusters and the new economics of competition. *Harvard Business Review, 76*(6), 77–90.

Reyes, F., et al. (2023). Soil properties zoning of agricultural fields based on a K-means clustering analysis. *European Journal of Agronomy, 150*, 126930. https://doi.org/10.1016/j.eja.2023.126930

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Stojanova, S., et al. (2022). Rural Digital Innovation Hubs as a Paradigm for Sustainable Business Models in Europe's Rural Areas. *Sustainability, 14*(21), 14620. https://doi.org/10.3390/su142114620

Taishykov, Z. (2024). Management of innovation processes in agriculture. *World Development Perspectives, 33*, 100509. https://doi.org/10.1016/j.wdp.2024.100566

Tkacheva, A., et al. (2024). Problems and Prospects for the Development of Cluster Structuring in the Economy of Kazakhstan's Agricultural Sector: Theory and Practice. *Economies, 12*(7), 185. https://doi.org/10.3390/economies12070185

Toillier, A., Mathé, S., Saley Moussa, A., & Faure, G. (2022). How to assess agricultural innovation systems in a transformation perspective: A Delphi consensus study. *The Journal of Agricultural Education and Extension, 28*(2), 163–185. https://doi.org/10.1080/1389224X.2021.1953548

Wandel, J. (2010). The cluster-based development strategy in Kazakhstan's agro-food sector: A critical assessment from an Austrian perspective, Discussion Paper, No. 128, Leibniz Institute of Agricultural Development in Central and Eastern Europe (IAMO), Halle (Saale). https://nbn-resolving.de/urn:nbn:de:gbv:3:2-10641

World Bank. (2012). *Agricultural Innovation Systems: An Investment Sourcebook*. Washington, DC: World Bank. https://documents1.worldbank.org/curated/en/140741468336047588/pdf/672070PUB0EPI0067844B09780821386842.pdf?utm_

World Bank. (2013). *Kazakhstan Fostering Productive Innovation Project*. Washington, DC: World Bank. http://documents.worldbank.org/curated/en/410881468039550416

World Bank. (2020). *Innovation in Kazakhstan: From ideas to impact* [Video]. Washington, DC: World Bank. https://www.worldbank.org/en/news/video/2020/04/14/innovation-in-kazakhstan-from-ideas-to-impact?utm_

Yuan, Y., Shi, B., Liu, X., Tian, Y., Zhu, Y., Cao, W., & Cao, Q. (2022). Optimization of management zone delineation for precision crop management in an intensive farming system. *Plants, 11*(19), 2611. https://doi.org/10.3390/plants11192611

AUTHOR BIOGRAPHIES

Nurbakhyt N. Nurmukhametov – Cand. Sc. (Econ.), Associate Professor, Korkyt Ata Kyzylorda University, Kyzylorda, Kazakhstan. Email: nyrbahit73@mail.ru, ORCID ID: https://orcid.org/0000-0002-8551-0573

Alexander A. Tsoy – Researcher, University of International Business named after K. Sagadiyev, Almaty, Kazakhstan. Email: alt-kct@mail.ru, ORCID ID: https://orcid.org/0000-0002-7054-6063

***Meiirzhan Abdykadyr** – Researcher, University of International Business named after K. Sagadiyev, Almaty, Kazakhstan. Email: meiirzhanabdykadyr@gmail.com, ORCID ID: https://orcid.org/0009-0009-1974-7095